

Lecture Notes in Civil Engineering

Scott Walbridge · Mazdak Nik-Bakht ·
Kelvin Tsun Wai Ng · Manas Shome ·
M. Shahria Alam · Ashraf El Damatty ·
Gordon Lovegrove *Editors*

Proceedings of the Canadian Society of Civil Engineering Annual Conference 2021

CSCE21 Hydrotechnical and
Transportation Track

 Springer

Lecture Notes in Civil Engineering

Volume 250

Series Editors

Marco di Prisco, Politecnico di Milano, Milano, Italy

Sheng-Hong Chen, School of Water Resources and Hydropower Engineering,
Wuhan University, Wuhan, China

Ioannis Vayas, Institute of Steel Structures, National Technical University of
Athens, Athens, Greece

Sanjay Kumar Shukla, School of Engineering, Edith Cowan University, Joondalup,
WA, Australia

Anuj Sharma, Iowa State University, Ames, IA, USA

Nagesh Kumar, Department of Civil Engineering, Indian Institute of Science
Bangalore, Bengaluru, Karnataka, India

Chien Ming Wang, School of Civil Engineering, The University of Queensland,
Brisbane, QLD, Australia

Lecture Notes in Civil Engineering (LNCE) publishes the latest developments in Civil Engineering - quickly, informally and in top quality. Though original research reported in proceedings and post-proceedings represents the core of LNCE, edited volumes of exceptionally high quality and interest may also be considered for publication. Volumes published in LNCE embrace all aspects and subfields of, as well as new challenges in, Civil Engineering. Topics in the series include:

- Construction and Structural Mechanics
- Building Materials
- Concrete, Steel and Timber Structures
- Geotechnical Engineering
- Earthquake Engineering
- Coastal Engineering
- Ocean and Offshore Engineering; Ships and Floating Structures
- Hydraulics, Hydrology and Water Resources Engineering
- Environmental Engineering and Sustainability
- Structural Health and Monitoring
- Surveying and Geographical Information Systems
- Indoor Environments
- Transportation and Traffic
- Risk Analysis
- Safety and Security

To submit a proposal or request further information, please contact the appropriate Springer Editor:

- Pierpaolo Riva at pierpaolo.riva@springer.com (Europe and Americas);
- Swati Meherishi at swati.meherishi@springer.com (Asia - except China, and Australia, New Zealand);
- Wayne Hu at wayne.hu@springer.com (China).

All books in the series now indexed by Scopus and EI Compendex database!

More information about this series at <https://link.springer.com/bookseries/15087>

Scott Walbridge · Mazdak Nik-Bakht ·
Kelvin Tsun Wai Ng · Manas Shome ·
M. Shahria Alam · Ashraf El Damatty ·
Gordon Lovegrove
Editors

Proceedings of the Canadian Society of Civil Engineering Annual Conference 2021

CSCE21 Hydrotechnical and Transportation
Track

 Springer

Editors

Scott Walbridge
University of Waterloo
Waterloo, ON, Canada

Mazdak Nik-Bakht
Concordia University
Montreal, QC, Canada

Kelvin Tsun Wai Ng
University of Regina
Regina, SK, Canada

Manas Shome
Matrix Solutions Inc.
Edmonton, AB, Canada

M. Shahria Alam
University of British Columbia - Okanagan
Campus
Kelowna, BC, Canada

Ashraf El Damatty
University of Western Ontario
London, ON, Canada

Gordon Lovegrove
University of British Columbia - Okanagan
Campus
Kelowna, BC, Canada

ISSN 2366-2557

ISSN 2366-2565 (electronic)

Lecture Notes in Civil Engineering

ISBN 978-981-19-1064-7

ISBN 978-981-19-1065-4 (eBook)

<https://doi.org/10.1007/978-981-19-1065-4>

© Canadian Society for Civil Engineering 2022

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Contents

| | |
|---|----|
| Operating Speed Prediction Models for Urban Streets in Residential Areas: A Case Study in the City of Saskatoon | 1 |
| Hassan Hamad and Emanuele Sacchi | |
| Exploring the Linkage Between Human Factors and Road Geometric Elements Influencing the Road Traffic Accidents on the National Roads of South Africa | 15 |
| D. Das | |
| Development of an AI Tool to Identify Reference Reaches for Natural Channel Design | 31 |
| C. Kupferschmidt and A. Binns | |
| Stability Analysis of an Overtopped Spillway Using Computational Fluid Dynamics | 43 |
| M. Freitas, P. Léger, and L. Pedroso | |
| The Effects of Climatic and Design Variables on Evapotranspiration in Bioretention Systems | 57 |
| R. Nasrollahpour, A. Skorobogatov, J. He, C. Valeo, A. Chu, and B. van Duin | |
| Wetted Area in Comparison to Habitat Suitability Criteria Based Methodologies for Sustainable Water Resources Management | 63 |
| A. Kneale and H. Ghamry | |
| Influence of Surge Waves on the Transport of Macroplastics | 69 |
| P. Patel, M. Krol, and S. Karimpour | |
| Establishing Acceleration Profiles of Light-Duty Vehicles Departing in a Straight Path from Two-Way Stop-Controlled Intersections | 81 |
| Essam Dabbour and Olaa Dabbour | |

| | |
|---|-----|
| Dynamic Interaction of Twin Particle Clouds in Stagnant Water | 89 |
| M. Janati and A. H. Azimi | |
| Experimental Study of Sand-Water Coaxial Jets with Low Velocity Ratio | 103 |
| F. Sharif and A. H. Azimi | |
| Design of Roundabout to Replace All-Way Stop Controlled Intersection: Case Study in Niagara-on-the-Lake | 117 |
| N. Farag, F. T. Choudhury, Y. Cao, D. Ponce, C. Del Rosario, S. Alkarawi, M. Rataul, S. Arkatkar, and S. M. Easa | |
| Improving Selected Intersections Across Niagara Region for Miovision’s Smart City Program | 133 |
| Neil Gobin, Michelle Nguyen, Jasmine Puthoor, Rahmah Tariq, Michelle Yip, Sulaf Alkarawi, Manny Rataul, Said M. Easa, and Essam Dabbour | |
| Exploring the Use of Artificial Neural Networks for Scour Prediction | 149 |
| M. Marrocco, P. Williams, R. Balachandar, and R. Barron | |
| Valuing Demand Response on Electric Vehicles Charging | 163 |
| H. Yang and Y. H. Kim | |
| Advancing the Practice of Soil Bioengineering in Alberta—The Bow River Bioengineering Demonstration and Education Project | 177 |
| M. Gallant, N. Posada, J. Slaney, and P. Raymond | |
| Case Study of Flood Profile Hydraulic Modeling Uncertainty | 183 |
| Nathan A. Valsangkar and David G. McLean | |
| Surrogate Model Development for Bioretention Cell Simulation-Optimisation Applications | 197 |
| R. Khalid, E. Snieder, and U. T. Khan | |
| Segregation of Particles in Multi Size and Density Beds by Circular Wall Jets | 211 |
| H. Hernandez, A. Mostaani, and A. H. Azimi | |
| A Multiple Linear Regression (MLR) Model for the Application of Electrical Vehicles in the United States | 223 |
| L. Navarro and B. Bathaei | |
| Travel Time Reliability with the Presence of Connected and Autonomous Vehicles | 239 |
| F. Hosseinzadeh, M. Ansari Esfeh, and L. Kattan | |
| The Role of Microplastics’ Size and Density on Their Vertical Turbulent Mixing and Transport | 253 |
| Arefeh Shamskhany and Shooka Karimpour | |

Establishing Workflow Processes for Bridge Inspections Using RPAS 267
 C. Schmelzle and M. Paulsen

Numerical Investigation of Turbulent Structures and Air Entrainment in Positive Surge Waves 281
 Z. Li and S. Karimpour

A Review of the Literature on Design and Performance of Multi Lane Roundabouts in Canada: The Case for Turbo Roundabouts 291
 Juniper Scott and Lovegrove Gordon

Smartphone: A Source for Transit Service Planning and Management Using Wifi Sensor Data 305
 Muhammad Arslan Asim, Lina Kattan, and S. C. Wirasinghe

Assessment of Adaptation Solutions to Floods with PCSWMM and a Multicriteria Analysis for a Very Small Watershed 321
 Audrey Coulombe, Jean-Luc Martel, Annie Poulin, Mathias Glaus, Geneviève Audet, and Steve Girard

Hydrail One: A Standard Gauge Low Power Fuel Cell/battery Hybrid Hydrail Vehicle 335
 M. Hegazi and G. Lovegrove

Suppression of Vortex Shedding from Bridge Pier Using Attached Splitter Plate 345
 S. Samuel Li and S. Kheshtgar Darvazeh

A Comparative Study of Mean and Turbulent Flow Fields Around a Boulder Within Boulder Array and an Individual Boulder 357
 R. Pierce, A. Golpira, and A. B. M. Baki

Predicting the Geometrical Characteristics of an Inclined Negatively-Buoyant Jet for Angles from 30° to 60° Using GMDH Neural Network 369
 Hassan Alfaifi, Abdulmajid Mohammadian, and Hossein Bonakdari

Influence of Rainfall on the Probability of Red-Light Running at Signalised Intersections 379
 J. Oyaro and J. Ben-Edigbe

Pressure Distribution in a Dam-Break Flow Simulated by a Mesh-Free Method 393
 Tibing Xu, Jinlong Zuo, and Zhijian Huang

Estimates of River Bathymetry from Satellite Images: A Case Study of the Nicolect River in Quebec 401
 Shayan Salavitarab and S. Samuel Li

| | |
|---|-----|
| Numerical Simulation of Turbulent Offset Dense Jet Flow | 413 |
| Mostafa Bigdeli and Abdolmajid Mohammadian | |
| Comparison of Three Minimization Theories for River Morphological Adjustments | 423 |
| Eman AlQasimi and Tew-Fik Mahdi | |
| Investigating Event Selection for GA-Based SWMM Rainfall-Runoff Model Calibration | 429 |
| E. Snieder and U. T. Khan | |
| How Can a State-of-the-Art Prediction Technique for Random Parameter Models Improve the Construction Work Environment in Ontarian Highways? | 443 |
| Seyedata Nahidi and Susan Tighe | |
| Quality of Life Outcomes of the Smarter Growth Neighborhood Design Principles: Case Study City of Kelowna | 457 |
| Abdul Rahman Masoud, Ahmed Idris, and Gordon Lovegrove | |
| Reducing the Risk of Basement Flooding Through Building- and Lot-Scale Flood Mitigation Approaches: Performance of Foundation Drainage Systems | 471 |
| B. Kaur, A. Binns, D. Sandink, B. Gharabaghi, and E. McBean | |
| Predicting Navigability in the Lower Athabasca River System Through Numerical Modelling | 479 |
| S. Kashyap, E. Kerkhoven, Z. Islam, A. Petty, and S. Depoe | |
| Water Distribution System Leak Detection Using Support Vector Machines | 491 |
| Z. Cai, R. Dziedzic, and S. S. Li | |
| Comparison of Machine Learning Classifiers for Predicting Water Main Failure | 501 |
| M. Amini and R. Dziedzic | |
| New Hybrid Search Algorithm for the Capacitated Vehicle Routing Problem | 513 |
| Nayera Elgharably, Ashraf Nassef, Said Easa, and Ashraf El Damatty | |
| Comparison of Evaporative Losses in Alberta Based on Five Evapotranspiration Models | 529 |
| Zahidul Islam, Shalini Kashyap, and Michael Seneka | |
| Numerical Modelling of Variable Density Shallow Water Flows with Friction Term | 543 |
| Amine Hanani, Abdelaziz Beljadid, and Driss Ouazar | |
| A Short-Term Flood Forecasting Model Using Markov Chain | 555 |
| Arpita Islam, Maysara Ghaith, Sonia Hassini, and Wael El-Dakhkhni | |

Modelling of Unsaturated Flow Through Porous Media Using Meshless Methods 565
Mohamed Boujouadar, Abdelaziz Beljadid, and Ahmed Taik

Digital Twin: A City-Scale Flood Imitation Framework 577
Maysara Ghaith, Ahmed Yosri, and Wael El-Dakhakhni

Modelling of Coupled Surface and Subsurface Water Flows 589
Hasan Karjoun and Abdelaziz Beljadid

Development of New Extreme Rainfall Maps for Urban Infrastructure Design in Canada Using the Scale-Invariance Generalized Extreme Value Distribution 603
Truong-Huy Nguyen and Van-Thanh-Van Nguyen

Assessing Drainage System Impacts Due to Urban Intensification in Rurally-Serviced Residential Areas 615
M. Senior and R. Scheckenberger

Review and Updated Guidance for Embankment Overtopping Dam Breach Dimensions 631
Mayari Bernard-Garcia and Tew-Fik Mahdi

Robust Cross-Validated Feature Selection for Machine Learning Streamflow Forecasting Models: A Case Study 643
A. Gharib, E. G. R. Davies, and N. Ilich

A Multiyear Infrastructure Planning Framework for Connected and Automated Vehicles 649
Fehintola Sanusi, Juyeong Choi, and Ren Moses

The Influence of Socioeconomic Variables on Cycling Infrastructure Preferences in Calgary 663
H. Smith and F. Sadeghpour

Riverine Hydrokinetic Energy Extraction: Investigation into a Location’s Suitability for Turbine Deployment 677
K. Kirby, C. Rennie, J. Cousineau, S. Ferguson, and I. Nistor

Integrating Activity-Based Modelling with Agent-Based Transit Assignment 689
Patrick Meredith-Karam, Siva Srikukenthiran, James Vaughan, Eric J. Miller, and Amer Shalaby

About the Editors

Dr. Scott Walbridge Ph.D., P.Eng. (Alberta) is a Professor at the University of Waterloo, where he has worked since completing his Ph.D. at the EPFL in Lausanne, Switzerland, in 2005/2006. Dr. Walbridge currently serves as Technical Committee (TC) Vice-Chair for the CSA S157 Design Code (Strength Design of Aluminum Structures) and TSC Chair for CSA S6 (Canadian Highway Bridge Design Code) Section 17 (Aluminum Structures). He is also a member of CSA W59 and W59.2 (Canadian Structural Steel and Aluminum Welding Codes). In addition to his work on various structural design codes, he also currently serves as the Program Director for the University of Waterloo's new undergraduate program in architectural engineering. His research has investigated topics in various areas including fatigue of steel and aluminum welded connections, performance of steel and aluminum mechanical connections, vibration of aluminum pedestrian bridges, and stability of aluminum structures. He was awarded an Alexander von Humboldt Fellowship related to his research on metal fatigue in 2016–2017.

Dr. Mazdak Nik-Bakht is an Associate Professor of Construction Engineering and Management in the Department of Building, Civil and Environmental Engineering at Concordia University, Montreal. He is the director of Compleccity Lab and a co-director of Centre for Innovation in Construction and Infrastructure Engineering and Management (CICIEM) at Gina Cody School of Engineering and Computer Science. He has a Ph.D. in construction engineering and management from the University of Toronto (2015), as well as a Ph.D. in structural engineering from Iran University of Science and Technology (2011). Dr. Nik-Bakht is a licensed Professional Engineer with Professional Engineers Ontario and has years of professional experience as a structural designer, structural engineering division head, and project manager in structural & infrastructure rehabilitation projects. A new course called 'Big Data Analytics for Smart City Infrastructure' to train the next generation of construction and infrastructure engineering professionals was developed by him in 2019 and has been delivered annually ever since.

Dr. Kelvin Tsun Wai Ng is a Professor of Environmental Systems Engineering at University of Regina, Canada. His major fields of interest are in waste management and environmental sustainability, particularly in development of (i) sustainable waste management system, (ii) evidence-based waste policy, and (iii) data-driven waste collection and disposal methods. His projects have been funded by a number of provincial and federal agencies, including NSERC, Mitacs, Communities of Tomorrow, Innovation Saskatchewan, Ministry of Environment, City of Regina, etc. Dr. Ng is a Professional Engineer with Permission to Consult in Saskatchewan, Canada. He is also the instructor for a number of popular courses at the Faculty of Engineering and has received the 2017 President's Award for Teaching Excellence.

Manas Shome Ph.D., P.Eng. is a Principal Engineer at Matrix Solutions, a Calgary-based environmental engineering consulting firm. Based in Edmonton, he provides technical leadership on water resources engineering projects. He obtained a B.Sc. and M.Sc. in civil engineering at the Bangladesh University of Engineering and Technology and earned his Ph.D. in civil engineering from the University of Alberta in 1995. His professional experience includes more than 30 years in sustainable water resources management, river engineering, computational modeling of rivers and drainage systems, design of hydraulic structures, and environmental impact assessment studies. An expert in the field of hydrology and hydraulic engineering, he frequently serves as an expert witness at regulatory hearings. Outside his consulting practice, Manas enjoys research, teaching, and watching sports. His research outcomes have been published in prestigious peer-reviewed journals and in conference proceedings. He is a co-recipient of the 2010 Keefer Medal award and recipient of the 2015 Excellence in Teaching award from the University of Alberta's Faculty of Extension. He has been elected to the grade of Fellow by the CSCE in 2019. Manas has been involved with CSCE since 2000 serving on the Edmonton section in various roles including Chair and on the organizing committee of the 2012 Annual Conference in Edmonton, Alberta.

Dr. M. Shahria Alam is a Professor of Civil Engineering at the University of British Columbia (UBC)'s Okanagan Campus and the Director of Green Construction Research & Training Center (GCRTC) at UBC. He is currently serving as the Chair of the Engineering Mechanics and Materials (EMM) division of the Canadian Society for Civil Engineering (CSCE). His research interests include applications of smart materials like shape memory alloys and sustainable construction materials, performance-based design, repair/retrofit of structures, and self-centering structures. Dr. Alam is the Director of the Applied Laboratory for Advanced Materials and Structures (ALAMS) at UBC. He has authored over 250 peer-reviewed articles and holds several patents. He is the recipient of several national and international awards including CSCE's Pratley Award. He is also an active member of several ACI and ASCE code committees.

Dr. Ashraf El Damatty is a Professor and Chair of the Department of Civil and Environmental Engineering at the University of Western Ontario, London, Ontario, Canada. He is a Fellow of the Canadian Society of Civil Engineering and Fellow of the Engineering Institute of Canada. He is a Research Director at the WindEEE Research Institute. He holds the title of High-End Expert at Tongji University and Sichuan Universities, China. He obtained a BSc. and M.Sc. from Cairo University in 1986 and 1991, Ph.D. in structural engineering from McMaster University, Canada, in 1995, and an MBA in 2016 in higher education management from University College, London, UK. He is the founder of the Canadian Society of Civil Engineering (CSCE) Steel Structures Committee and serves currently as the Chair of the CSCE Structures Division. Dr. El Damatty has consulted on many projects in North America and the Middle East. He has written about 250 publications, supervised more than 30 Ph.D. and 20 M.Sc. students, and has been invited as keynote speaker in many countries around the globe. He received several awards including the Alan Yorkdale Award by ASTM, Honorable Mention in 2014 Casimir Gzowski Medal Competition, 2015 CSCE Whitman Wright Award, 2016 CSCE Horst Leipholtz Medal and Western University Faculty Scholar Award, and the 2018 Professional Engineers of Ontario Engineering Medal of Research of Development. His research work has influenced the international codes and engineering practice worldwide.

Dr. Gordon Lovegrove is an MBA/Civil Engineer with over 30 years of experience in smarter growth, housing, sustainable transportation, safety, economics, and project management. Dr. Lovegrove was the point man that got the award-winning Universal Transportation Pass (U-Pass) for UBC that tripled transit use at both campuses. His current research focus is on zero-emission, hydrogen-powered rail (hydrail), including several industry partnerships to develop Canada's first hydrail switcher locomotive. Dr. Lovegrove is founding partner of the SMARTer Growth Partnership (smartergrowthpartnership.ca), an Associate Professor/founding member of UBCO's School of Engineering (ok.ubc.ca), VP Technical Programs for the Canadian Society of Civil Engineering (csce.ca); and, co-author of several best practice guides on sustainable development, health, and safety. He brings 20 years of practical industry development experience plus 14 years of engineering research and design expertise. His research team includes electrical, mechanical, and civil engineering graduate students, plus computer science coders and research assistants.

Operating Speed Prediction Models for Urban Streets in Residential Areas: A Case Study in the City of Saskatoon



Hassan Hamad and Emanuele Sacchi

1 Introduction

The prediction of the speed at which drivers operate their vehicles (operating speed) is an important area of research in highway engineering. Operating speed prediction models can be used by highway engineers to verify, at the design stage, whether the desired (designed) speed of a facility will be likely assumed by drivers.

Based on the Geometric Design Guide for Canadian Roads [8], three speed values have to be considered in highway design: the design speed, the posted speed and the operating speed. The design speed is the speed value employed to determine the geometric elements of a road, such as superelevation, radius of a circular curve, sight distance, and the length of crest and sag vertical curves. The posted speed is the speed limit introduced to encourage drivers to travel at an appropriate speed considering the design speed of the geometric elements and surrounding conditions. The posted speed is set to be lower than the design speed limit to ensure a safety interval for drivers and account for possible human errors. Finally, the operating speed is the speed at which drivers are observed traveling under unimpeded (free-flow) conditions. Measuring the operating speed is conducted with a speed survey under uncongested (free-flow) conditions, so that the motion of a vehicle is not interrupted by other vehicles in the traffic flow or by traffic control devices (e.g., traffic signals, stop or yield signs). The 85th percentile of the distribution of speeds collected is widely considered to be an acceptable measurement of the operating speed for a given roadway section. Therefore, when the 85th percentile speed is aligned with the posted speed limit, drivers travel below the design speed and it can be inferred that the design has produced the intended travel speeds set by the designer.

H. Hamad · E. Sacchi (✉)
University of Saskatchewan, Saskatoon, Canada
e-mail: emanuele.sacchi@usask.ca

© Canadian Society for Civil Engineering 2022
S. Walbridge et al. (eds.), *Proceedings of the Canadian Society of Civil Engineering Annual Conference 2021*, Lecture Notes in Civil Engineering 250,
https://doi.org/10.1007/978-981-19-1065-4_1

In order to measure the operating speed (85th percentile) for urban streets, speed measurements are usually collected mid-segment in free-flow conditions, away from traffic interruptions caused by intersection controls. Speed measurements can be collected using radar guns, pneumatic tubes, inductive loops or microwave sensors, among others. The speed measurement collection at a specific spot along the roadway is referred to as spot speed measurement (or time-mean speed). Spot speed measurements can be either aggregated or disaggregated measurements meaning that speeds are grouped in speed bins for the former and ungrouped for the latter. Spot speed is defined as the arithmetic mean of speeds of vehicles passing a point.

In urban residential areas, operating speeds often exceed the intended design speeds creating speeding concerns among residents. Speeding occurs when drivers tend to travel at speeds above the posted speed limit. Speeding is a serious issue because it increases mean speeds and speed variation among drivers, which has been demonstrated to compromise the level of safety [4]. One reason for speeding is the lack of performance-based design procedures that incorporate expected operating speeds during the planning and design stages of urban roadways. In recent years, the City of Saskatoon has conducted several speed surveys within its neighborhood traffic reviews (NTRs) to gain an understanding of the traffic patterns at the neighborhood level opposed to the case-by-case analysis prior to that [15]. Around eight neighborhoods are selected each year for the analysis based on a prioritization criteria that considers the residents' concerns, collision history, the stage of development and the age of the neighborhoods. Within this context, the goal of this research was to understand the relationship between operating speeds and various road characteristics for the sites reviewed within NTRs, with the ultimate goal of promoting performance-based design procedures in highway design. Since the 85th percentile speed is widely used to model operating speeds, 85th-percentile free-flow speed prediction models were developed based on 140 residential streets in Saskatoon.

2 Literature Review

Operating speed predictors have been discussed in a large body of literature over the years. Researchers have investigated those variables that could influence driver's speed choice, their level of significance in a model, their magnitude and positive or negative association to operating speeds. Some of the more prominent studies conducted in the past two decades were reviewed in this section.

In 1996, Tarris et al. considered aggregate and disaggregate 85th percentile speeds for 27 urban collectors in Pennsylvania [17]. These streets were characterized by different roadway alignment, cross-sectional elements, roadside elements and land use. The analysis was focused on curves with radii values between 11 and 230 m. The results showed that the 85th percentile speed depends linearly to the degree of curvature and other roadway geometric design parameters. Fitzpatrick et al. [6] analyzed eight suburban arterials in Texas. Free flow speeds were collected at several

sections with varying geometric design characteristics. Two operating speed prediction models for tangents and curves, respectively, were derived using approach density (frequency of all approaches like driveways and intersections over the road segment) and horizontal radius (for horizontal curves) as covariates. In the NCHRP Report 504 [5], posted speed limit was introduced as covariate for urban-road speed prediction models and it was demonstrated that speed limits progressively decrease their effect on operating speeds when other variables like access density, parking, and pedestrian activity level become more relevant.

Figuroa et al. [12] developed a speed prediction model able to estimate different percentile speeds. Ordinary least-squares regression was applied to panel data structure. Factors affecting mean speed for tangent and curve sections included percent of trucks, speed limit, grade, residential development indicator, sight distance, intersection presence, roadway width, shoulder widths, curve presence, degree of curvature and superelevation. Factors affecting standard deviation of speeds for tangent and curve sections included speed limit, grade, intersection presence, roadway width, roadside clear zone, degree of curvature and superelevation. The linear combination of both mean and standard deviation in the prediction model allowed to estimate speed percentiles.

In a study conducted by Wang et al. [19], in-vehicle GPS devices were used to collect continuous vehicle activity data along low-speed urban streets. Operating speed models were developed for tangent sections of urban roads with speed limits ranging from 30 to 40 mph. The factors used for the prediction of the operating speed (85th and 95th percentile speeds) were: number of lanes, density and offsets of road-side objects, density of T-intersections and driveways, raised curb presence, sidewalk presence, on-street parking, and land uses. The study concluded that the inclusion of the posted speed limit as a factor of operating-speed prediction is not recommended due to the strong correlation between the posted speed limit and the observed operating speed. This study attempted to develop operating speed models with the consideration of driver and vehicle effects as random effects which allows for the intercepts of those groups to vary independently from the population intercept. 34.9% of the unexplained variance was attributed to individual differences among drivers and vehicle characteristics.

Bassani and Sacchi [1] developed multiple linear regression models to predict operating speeds (85th percentile) at urban arterials and collectors in the city of Torino, Northern Italy. Three models were developed: one for urban arterials and collectors (aggregated analysis), and two for divided and undivided roads (disaggregated analysis). The variables which were found to affect significantly the operating speed were the posted speed limit, lane position, lane width, median width, roadway width and pavement surface condition. A recalibration study of operating speed models from other countries using the Italian data set was conducted at the end of the study. Moses et al. [14] evaluated the 2010 HCM urban-street free-flow speed prediction model. The model uses the posted speed limit and eight additional variables to predict the free-flow speed. The variables associated with the prediction of the free-flow speed on urban roads are: the proportion of segment length with restrictive median, the proportion of segment with curb on the right-hand side, the

number of access point approaches on the right side in the subject direction of travel, the number of access point approaches on the right side in the opposing direction of travel, the segment length, the width of the signalized intersection, the number of through lanes, and the distance between intersections. The 2010 HCM speed prediction methodology applied to the study data underpredicted free-flow speed when the segment posted speed limit was 35 mph or higher. The use of the predicted free-flow speeds in determining the level of service of arterial segments yielded lower levels of service for 10 of the 20 arterials in the study. The researchers concluded that the prediction of the HCM model can be improved by decreasing the influence of cross section and access density factors and by increasing the posted speed limit coefficient.

Free-flow speed data was collected at tangential urban sections (arterials and collectors) in the city of Edmonton [18]. Variables with significant effects on the operating speed at arterial and/or collector roads were: segment length, median width, posted speed limit, lane configuration, access density, tree and pole densities and their offsets, sidewalk presence, roadside treatment, road width, pedestrian crossing, bus stops, service road presence, bike route presence, traffic composition and end treatment. Variables with significant effects on the speed variability at arterial and/or collector roads were: posted speed limit, lane configuration, access density, tree maturity, road width, average vehicle length, road class and roadside treatment. Different models were developed for the different road classes. The parameters that predict the operating speed and speed variation differed among different road classes.

Overall, research has showed that posted speed limit, number of lanes, access points density, presence of sidewalks and roadside treatment were some of the most common predictors of operating speed [3, 12, 14, 18, 19]. Other less common factors included segment length, road width, median indicators, on-street parking, bike lane presence, bus stop presence, grade, density and offsets of road-side objects and land use.

3 Data Collection

Speed data to develop operating (85th percentile) speed prediction models was obtained from the City of Saskatoon within NTRs. Data were collected at 140 locations in 7 neighbourhoods—suburban development areas (SDAs) (i.e., Core Neighborhood SDA East, Core Neighborhood SDA West, Lakewood SDA, Nutana SDA, University Heights SDA, Confederation SDA, and Lawson SDA). Speeds were measured between the months of April and October for the years 2017 and 2018 using pneumatic tubes placed at mid-segments. 24 h of data were recorded at all sites for an average of 6 days per location. In order to develop operating speed prediction models, speed measurements needed to reflect free-flow conditions. Therefore, all speed measurements for vehicles with headway less than 6 s were removed which is

equivalent to volumes of less than 600 veh/h [7, 9]. Nighttime and weekend measurements were also removed from speed data to eliminate biases due to visual impairment in low-light environments or unusual travel patterns/behavior in non-weekdays [14].

Regarding site-related characteristics, data was obtained from the Google Maps and Street View tool. This latter survey was based on the conditions in 2018. Built-in measuring tools were used to compile the data set along with satellite images. Length and width measurements were directly made on the map while density variables were processed after counting was conducted. The specifications of the speed variable and site-related characteristics were based on the literature of speed prediction model for urban streets, following also the work of Moses and Mtoi in evaluating the free-flow speeds at interrupted flow facilities [10, 13]. Table 1 presents a description of the variables employed whereas summary statistics of them is reported in Table 2.

Table 1 Description of speed variable and site-related characteristics

| Variable | Description |
|----------|--|
| V85 | 85th percentile speed; aggregated spot speed measurements in free flow-conditions were used to generate speed distributions, sample mean, standard deviation and the 85th percentile for each site |
| L | Length of the road segment, measured between two traffic control devices interrupting the flow (signal, yield or stop sign) |
| AccPD | Density of access points: the count of stop or yield-controlled accesses along the road segment per unit of length; an access point has a minimum flow rate of 10 vehicles/h |
| DWD | Density of driveways: the count of accesses to properties along the road segment per unit length; a driveway has a maximum flow rate of 10 vehicles/h |
| PedXD | Density of pedestrian crossings: the count of pedestrian crossings per unit of length; pedestrian crossing points are only counted when a sign or pavement marking is present |
| BSD | Bus stop density: the count of bus stops per unit of length |
| Tree | Tree density: the count of roadside trees per unit of length |
| Pole | Pole density: the count of roadside utility poles (light poles, electricity and communication poles) per unit length |
| TWW | Travelled-way width: the cross-sectional width of the paved surface (curb to curb) |
| OSP | On-street parking presence: 0 = not present on both sides; 1 = all other cases |
| SW | Sidewalk presence: 1 = present on both sides; 0 = all other cases |
| BL | Bike lane presence: 0 = not present on both sides; 1 = all other cases |
| CLP | Centerline presence: centerline marking dividing two opposing way of traffic. 0 = not present; 1 = present |
| SchZn | School zone presence: accounts for the presence of a school zone sign indicating a lower speed limit (30 km/h). 0 = school zone not present; 1 = school zone present |

Table 2 Summary statistics of variables

| Variable | Unit of measure | Maximum | Minimum | Average | Std. dev |
|----------|-----------------|---------|---------|---------|----------|
| V85 | km/h | 85.59 | 24.63 | 46.82 | 8.63 |
| L | m | 1863.00 | 100.00 | 702.03 | 412.33 |
| AccPD | Access/km | 18.13 | 0.00 | 7.43 | 4.46 |
| DWD | Driveway/km | 132.40 | 0.00 | 47.02 | 27.24 |
| PedXD | Crossing/km | 11.36 | 0.00 | 1.90 | 2.18 |
| BSD | Stop/km | 20.51 | 0.00 | 3.17 | 4.08 |
| Pole | Pole/km | 85.05 | 13.19 | 26.49 | 11.81 |
| Tree | Tree/km | 310.00 | 0.00 | 76.08 | 45.49 |
| TWW | m | 25.00 | 6.75 | 12.91 | 3.04 |
| OSP | 0/1 | 1.00 | 0.00 | 0.96 | 0.19 |
| SW | 0/1 | 1.00 | 0.00 | 0.89 | 0.32 |
| BL | 0/1 | 1.00 | 0.00 | 0.06 | 0.25 |
| CLP | 0/1 | 1.00 | 0.00 | 0.38 | 0.49 |
| SchZn | 0/1 | 1.00 | 0.00 | 0.20 | 0.40 |

3.1 Multicollinearity Analysis

Multicollinearity analysis was conducted as preliminary step to develop operating speed prediction models (Table 3). Multicollinearity refers to the fact that one independent variable in a multiple regression model can be linearly predicted from the others and is measured through the Pearson correlation coefficient (ρ). The null hypothesis of multicollinearity analysis is that there is no correlation between independent variables. The t-statistic in Table 3 (shown below Pearson correlation coefficient) demonstrates the significance of the relationship. Pearson's correlation coefficient values range from -1 to 1 , where a value of -1 represents total negative correlation, a value of 1 represents total positive correlation and a value of 0 represents no correlation. Table 2 demonstrated that most of the independent variables showed low ($-0.5 < \rho < 0.5$) or insignificant correlation. The only variables with ρ significant and higher than 0.5 were TWW and Pole. Therefore, Pole was discarded from the analysis because TWW was more highly and significantly correlated to V85 than Pole (see column/row in bold in Table 3).

4 Methodology

Two multiple linear regression models were developed in this study in order to predict V85 using site-related (road and traffic) characteristics, i.e., fixed-effects and mixed-effects (random intercept) models. For the fixed-effects model, parameter

Table 3 Pearson correlation coefficients and their significance for the independent variables

| Variable | V85 | L | AccPD | DWD | PedXD | BSD | Tree | Pole | TWW | OSP | SW | BL | CLP | SchZn |
|----------|-------------------|--------------|---------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| V85 | 1.000 | 0.234 | -0.048 | -0.062 | 0.343 | 0.151 | -0.352 | 0.071 | 0.453 | -0.150 | -0.019 | 0.313 | 0.374 | -0.105 |
| L | 0.234 | 0.005 | 0.573 | 0.466 | <0.0001 | 0.076 | <0.0001 | 0.402 | <0.0001 | 0.078 | 0.827 | 0.000 | <0.0001 | 0.217 |
| AccPD | 0.005 | 1.000 | 0.311 | 0.283 | -0.023 | 0.195 | -0.201 | -0.236 | 0.086 | 0.050 | 0.075 | -0.043 | 0.265 | 0.410 |
| DWD | 0.005 | 0.000 | 0.000 | 0.001 | 0.786 | 0.021 | 0.017 | 0.005 | 0.312 | 0.554 | 0.376 | 0.618 | 0.002 | <0.0001 |
| PedXD | -0.048 | 0.311 | 1.000 | 0.008 | 0.151 | 0.278 | -0.264 | -0.157 | -0.178 | 0.074 | 0.093 | -0.125 | 0.195 | 0.255 |
| BSD | 0.573 | 0.000 | 0.074 | 0.925 | 0.001 | 0.001 | 0.002 | 0.064 | 0.035 | 0.385 | 0.272 | 0.143 | 0.021 | 0.002 |
| Tree | -0.062 | 0.283 | 0.008 | 1.000 | -0.248 | -0.102 | -0.077 | -0.383 | -0.140 | 0.116 | 0.231 | -0.177 | -0.128 | 0.107 |
| Pole | 0.466 | 0.001 | 0.925 | 0.003 | 0.003 | 0.232 | 0.363 | <0.0001 | 0.098 | 0.174 | 0.006 | 0.037 | 0.133 | 0.207 |
| TWW | 0.343 | -0.023 | 0.151 | -0.248 | 1.000 | 0.248 | -0.360 | 0.157 | 0.133 | -0.059 | 0.008 | 0.400 | 0.352 | 0.161 |
| OSP | <0.0001 | 0.786 | 0.074 | 0.003 | 0.003 | 0.003 | <0.0001 | 0.065 | 0.116 | 0.487 | 0.925 | <0.0001 | <0.0001 | 0.058 |
| SW | 0.151 | 0.195 | 0.278 | -0.102 | 0.248 | 1.000 | -0.067 | 0.277 | 0.368 | -0.065 | 0.194 | 0.156 | 0.457 | 0.108 |
| BL | 0.076 | 0.021 | 0.001 | 0.232 | 0.003 | 0.003 | 0.429 | 0.001 | <0.0001 | 0.446 | 0.022 | 0.065 | <0.0001 | 0.203 |
| CLP | -0.352 | -0.201 | -0.264 | -0.077 | -0.360 | -0.067 | 1.000 | 0.162 | 0.038 | 0.016 | 0.101 | -0.170 | -0.163 | -0.006 |
| SchZn | <0.0001 | 0.017 | 0.002 | 0.363 | <0.0001 | 0.429 | 0.055 | 0.055 | 0.659 | 0.849 | 0.235 | 0.044 | 0.055 | 0.947 |
| OSP | 0.071 | -0.236 | -0.157 | -0.383 | 0.157 | 0.277 | 0.162 | 1.000 | 0.581 | 0.012 | 0.007 | 0.251 | 0.038 | -0.076 |
| SW | 0.402 | 0.005 | 0.064 | <0.0001 | 0.065 | 0.001 | 0.055 | <0.0001 | 0.886 | 0.886 | 0.935 | 0.003 | 0.655 | 0.371 |
| BL | 0.453 | 0.086 | -0.178 | -0.140 | 0.133 | 0.368 | 0.038 | 0.581 | 1.000 | -0.021 | 0.112 | 0.185 | 0.161 | -0.027 |
| CLP | <0.0001 | 0.312 | 0.035 | 0.098 | 0.116 | <0.0001 | 0.659 | <0.0001 | 0.805 | 0.805 | 0.188 | 0.029 | 0.058 | 0.749 |
| SchZn | -0.150 | 0.050 | 0.074 | 0.116 | -0.059 | -0.065 | 0.016 | 0.012 | -0.021 | 1.000 | 0.415 | -0.106 | 0.071 | 0.096 |
| OSP | 0.078 | 0.554 | 0.385 | 0.174 | 0.487 | 0.446 | 0.849 | 0.886 | 0.805 | <0.0001 | <0.0001 | 0.211 | 0.406 | 0.258 |
| SW | -0.019 | 0.075 | 0.093 | 0.231 | 0.008 | 0.194 | 0.101 | 0.007 | 0.112 | 0.415 | 1.000 | -0.089 | 0.142 | -0.045 |

(continued)

Table 3 (continued)

| Variable | V85 | L | AccPD | DWD | PedXD | BSD | Tree | Pole | TWW | OSP | SW | BL | CLP | SchZn |
|----------|-------------------|---------|--------|--------|---------|---------|--------|--------|--------|---------|--------|--------|-------|--------|
| | 0.827 | 0.376 | 0.272 | 0.006 | 0.925 | 0.022 | 0.235 | 0.935 | 0.188 | <0.0001 | | 0.296 | 0.095 | 0.598 |
| BL | 0.313 | -0.043 | -0.125 | -0.177 | 0.400 | 0.156 | -0.170 | 0.251 | 0.185 | -0.106 | -0.089 | 1.000 | 0.276 | -0.131 |
| | 0.000 | 0.618 | 0.143 | 0.037 | <0.0001 | 0.065 | 0.044 | 0.003 | 0.029 | 0.211 | 0.296 | | 0.001 | 0.123 |
| CLP | 0.374 | 0.265 | 0.195 | -0.128 | 0.352 | 0.457 | -0.163 | 0.038 | 0.161 | 0.071 | 0.142 | 0.276 | 1.000 | 0.088 |
| | <0.0001 | 0.002 | 0.021 | 0.133 | <0.0001 | <0.0001 | 0.055 | 0.655 | 0.058 | 0.406 | 0.095 | 0.001 | | 0.299 |
| SchZn | -0.105 | 0.410 | 0.255 | 0.107 | 0.161 | 0.108 | -0.006 | -0.076 | -0.027 | 0.096 | -0.045 | -0.131 | 0.088 | 1.000 |
| | 0.217 | <0.0001 | 0.002 | 0.207 | 0.058 | 0.203 | 0.947 | 0.371 | 0.749 | 0.258 | 0.598 | 0.123 | 0.299 | |
| Variable | V85 | L | AccPD | DWD | PedXD | BSD | Tree | Pole | TWW | OSP | SW | BL | CLP | SchZn |

Pearson Correlation Coefficient (ρ)Prob > |r| under $H_0: \rho = 0$

estimates were constrained to be fixed among different sites/neighbourhoods. In the mixed-effects model, the intercept of the model was estimated as a random variable varying among the 7 different neighbourhoods. This latter modeling framework was employed to account for possible unobserved heterogeneity at the neighbourhood level. Unobserved heterogeneity refers to the fact that many components affecting speed levels are not easily available to the analyst. The use of a random intercept implies an additional variance component representing the variation of speeds among different neighbourhoods. More details related to the modeling are provided in the following two subsections.

4.1 Fixed-Effects Regression Model

Let Y_i denote the 85th percentile speed observed at site i ($i = 1, \dots, n$). It is assumed that speed observations at the n sites are independent and:

$$Y_i = \beta_0 + \sum \beta_j X_j + \epsilon_i \quad (1)$$

where X_j is the j -th independent variable, β_0 and β_j are model parameters and ϵ_i is a normally distributed random error term with mean 0 and variance equal to σ^2 .

4.2 Mixed-Effects Regression Model

In this study, the n segments under consideration belong to 7 mutually exclusive neighbourhoods. In such cases, an additional variance component can be included in the model to allow for the possibility that different neighbourhoods have different speed levels because of different local conditions. Suppose that the i th segment belongs to neighbourhood $h(i) \in \{1, 2, \dots, 7\}$. Let Y_i denote the 85th percentile speed observed at site i ($i = 1, \dots, n$). It is assumed that speed observations at the n sites are independent and:

$$Y_i = \beta_0 + \sum \beta_j X_j + u_{h(i)} + \epsilon_i \quad (2)$$

where $u_{h(i)}$ is a normally distributed error term of the random intercept with mean equal to zero and variance equal to σ_h^2 .

4.3 Bayesian Estimation

Bayesian estimation of model parameters was adopted in this study. In classical (frequentist) inference, the parameters of the regression model are fixed quantities which are obtained, for instance, by maximizing the likelihood function [11]. The formulation of Bayesian models has the additional feature of requiring the formulation of a set of prior distributions for any unknown parameter. A prior distribution summarizes any knowledge about the parameters that may be available before observing any data. Then a posterior distribution is estimated with sampling techniques. The specification of a set of prior distributions for a problem usually involves hyper-parameters (i.e., parameters of the prior distribution).

In this study, prior distributions for the whole set of parameters (β_0 and β_j) and hyper-parameter, σ^2 , were assumed as non-informative to reflect the lack of precise knowledge of their value. In details, following [2], the regression parameter were chosen as diffused normal distributions, with zero mean and large variance, i.e. normal $(0, 10^3)$, and as inverse-gamma $(0.001, 0.001)$ for variances. Afterwards, the posterior distributions were sampled using Markov Chain Monte Carlo (MCMC) techniques so that it was possible to obtain approximate quantities of the posterior mean and variance of the parameters.

The Bayesian analysis software WinBUGS was selected as the modeling platform for model estimations [16]. The code produced draws from the posterior distribution of the parameters, and given those draws, the MCMC technique was used to approximate the posterior mean and standard deviation of the parameters. therefore, the posterior summaries in this study where computed by running two independent Markov chains for each of the parameters in the models for 20,000 iterations to reach ratios of the Monte Carlo errors relative to the standard errors for each parameter around or less than 5%. Chains were thinned using a factor of 10 and the first 10,000 iterations in each chain were discarded as burn-in runs. The convergence was also monitored using the BGR statistics and visual approaches such as observing trace plots.

The deviance information criterion (DIC) was used as a relative goodness-of-fit measure for model selection. In common with Akaike information criteria (AIC), DIC penalizes the complexity of a model (i.e., models with a higher number of parameters are penalized more then those with a lower number) and, given two candidate models for the data, the preferred model is the one with the minimum DIC. It is assumed that a difference of more than 10 in the value of DIC might rule out the model with higher DIC. Differences between 5 and 10 are substantial. However, it could be misleading to report the model with the lowest DIC if the difference is less than 5 and the models make very different inferences [16].

5 Results and Discussion

All candidate variables in Table 1, apart from Pole which was found highly correlated to TWW, were tested and backward elimination was used to achieve the lowest DIC and retain only significant variables at the 95% confidence level. Table 4 shows the parameter estimates for the fixed-effects model in Eq. 1: the estimates were all significant as the 95% credible intervals were bounded away from zero. The significant predictors (X_j in Eq. 1) of the operating speed (Y_i) were length of the roadway segment, density of pedestrian crossings, density of bus stops, tree density, travelled way width, presence of on-street parking, presence of a centerline marking, and presence of a school zone. The presence of bus stops, trees, on-street parking and school zone had a negative association with the 85th percentile speed. On the contrary, road segment length, pedestrian crossing points, travelled way width and presence of centerline marking were positively associated with the 85th percentile speed.

Afterwards, the mixed-effects model in Eq. 2 was developed using the same covariates to predict the 85th percentile speed. Inference was similar in terms of magnitude and sign of parameter estimates but a lower deviance information criteria (DIC) value of 915.90 was found compared to the fixed-effects model (DIC = 919.93). This indicated a better goodness-of-fit of the mixed-effects model supporting the hypothesis of accounting for neighbourhood variation through a random intercept (mean of $\sigma_h = 1.792$) (Table 5).

Overall, the results appeared in line with the literature reviewed in Sect. 2. The positive association of the segment length (higher operating speeds for longer segment without traffic controls) aligns with the findings related to the recalibration of the HCM2010 urban operating free-flow speed model by Moses et al. and the operating speed model for urban collectors developed by Thiessen et al. [14, 18]. Regarding travelled way width, the positive association found in this research

Table 4 Parameter estimates of the fixed-effects operating-speed-prediction model

| Variable | Mean | Std. dev | 2.50% percentile | 97.50% percentile |
|-----------|--------|----------|------------------|-------------------|
| Intercept | 36.260 | 3.967 | 28.480 | 44.080 |
| L | 0.005 | 0.002 | 0.001 | 0.008 |
| PedXD | 0.758 | 0.298 | 0.175 | 1.343 |
| BSD | -0.495 | 0.158 | -0.807 | -0.183 |
| Tree | -0.043 | 0.013 | -0.069 | -0.017 |
| TWW | 1.280 | 0.189 | 0.904 | 1.654 |
| OSP | -6.986 | 2.936 | -12.840 | -1.239 |
| CLP | 4.882 | 1.330 | 2.262 | 7.529 |
| SchZn | -4.241 | 1.526 | -7.196 | -1.257 |
| σ | 6.251 | 0.391 | 5.542 | 7.071 |
| DIC | 919.94 | | | |

Table 5 Parameter estimates of the mixed-effects operating-speed-prediction model

| Variable | Mean | Std. dev | 2.50% percentile | 97.50% percentile |
|------------|--------|----------|------------------|-------------------|
| Intercept | 36.810 | 3.937 | 29.170 | 44.760 |
| L | 0.004 | 0.002 | 0.001 | 0.007 |
| PedXD | 0.793 | 0.293 | 0.217 | 1.362 |
| BSD | -0.426 | 0.163 | -0.743 | -0.109 |
| Tree | -0.039 | 0.014 | -0.066 | -0.013 |
| TWW | 1.254 | 0.189 | 0.882 | 1.622 |
| OSP | -7.265 | 2.739 | -13.030 | -1.996 |
| CLP | 4.705 | 1.331 | 2.078 | 7.306 |
| SchZn | -4.036 | 1.548 | -7.114 | -0.976 |
| σ | 6.064 | 0.398 | 5.341 | 6.881 |
| σ_h | 1.792 | 1.412 | 0.049 | 5.060 |
| DIC | 915.90 | | | |

(larger travelled way width associated to higher operating speeds) was also found in Moses et al., Thiessen et al. and in the work by Bassani and Sacchi [1]. Density of pedestrian crossings was found to be positively associated to the 85th percentile speed but this finding is debated in the literature. In the study by Eluru et al. [3], the presence of sidewalks (proxy for the presence of pedestrian traffic) on local urban sections contributed to higher operating speeds. However, in other studies on urban tangents, the presence of sidewalks or pedestrian crossings was associated to reduced operating speeds [18, 19]. The reason for this difference can be related to different adjacent land use, cultural differences in drivers' population or, simply, issues related to model specification. Higher bus stop density was associated to lower operating speeds in the literature and this finding was confirmed in this study. This outcome was also supported by the study carried out by Thiessen et al. in 2017 where the density of bus stops was shown to contribute to lower operating speeds on urban segments [18]. Similarly, in the studies by Wang et al. and Eluru et al., the presence of on-street parking contributed to a similar behavior leading to lower operating speed on urban road segments [3, 19]. The findings of Wang et al. and Thiessen et al. were confirmed for the case of roadside tree density. The models showed a negative association meaning that the higher the roadside tree density, the lower the operating speed [18, 19]. Regarding the presence of school zones, lower speed limits are set for drivers, i.e., 30 km/h instead of the 50 km/h speed limit across the city, to lower operating speed levels around schools. This outcome was confirmed by the models in this research which showed a negative association between the presence of school zones and the 85th percentile speed (lower operating speeds when school zones are present). Finally, the presence of centerline marking was found to be positively associated to 85th percentile speed. In other studies, the presence of raised median was explored and a positive association was observed. Therefore, the common result

between this research and the literature is that dividing the two way of a road can increase operating speeds (i.e., drivers tend to drive faster) [1, 14, 18].

6 Conclusions

Understanding factors that affect driver's speed choice on urban streets is a topic of primary importance in highway engineering. Prediction of operating speeds at the design stage, is essential to determine an appropriate design speed and the corresponding speed limit. In this work, two operating speed prediction models were developed in order to understand the effect of roadway factors to higher or lower operating speeds in Saskatoon's residential streets. Several significant predictors of the 85th percentile speeds were found in this research: segment length, pedestrian crossing density, bus stop density, tree density, traveled-way width, on-street parking presence, presence of centerline marking and presence of school zones. It has been demonstrated that the predictors of the 85th percentile were mostly in line with the literature, with the exception of density of pedestrian crossings which requires more evidence in terms of sign associated with 85th percentile speeds. Furthermore, accounting for mixed effects contributed to a better model fit while demonstrating that local speed variation related to drivers' population, traffic and environment of different neighbourhoods have an effect on the operating speeds and should be considered in future evaluations.

As for the limitations of this study, one predictor of operating speeds not included was the posted speed limit which is a variable commonly used in the literature of speed prediction models. This was caused by the fact that almost all streets in the dataset had a speed limit of 50 km/h.

The ultimate goal of this work was to provide tools to decision makers and practitioners for performance-based design of urban streets and to promote a more quantitative approach in speed studies. Also, the significance of the modeling effort for the city of Saskatoon can be used by other jurisdictions to predict operating speeds. This will allow, in turn, to test these models on different environments to confirm or recalibrate the parameter estimates.

References

1. Bassani M, Sacchi E (2012) Calibration to local conditions of geometry-based operating speed models for urban arterials and collectors. *Procedia Soc Behav Sci* 53:821–832. <https://doi.org/10.1016/j.sbspro.2012.09.931>
2. El-Basyouny K, Sayed T (2011) A full Bayes multivariate intervention model with random parameters among matched pairs for before-after safety evaluation. *Accid Anal Prev* 43(1):87–94. <https://doi.org/10.1016/j.aap.2010.07.015>

3. Eluru N, Chakour V, Chamberlain M, Miranda-Moreno LF (2013) Modeling vehicle operating speed on urban roads in Montreal: a panel mixed ordered probit fractional split model. *Accid Anal Prev* 59:125–134. <https://doi.org/10.1016/j.aap.2013.05.016>
4. Elvik R, Christensen P, Amundsen A (2004) Speed and road accidents: an evaluation of the power model. TOI report 740, vol 740, Issue December. <http://www.trg.dk/elvik/740-2004.pdf>
5. Fitzpatrick K, Carlson P, Brewer MA, Wooldridge MD, Miaou SP (2003) NCHRP report 504: design speed, operating speed and posted speed practices. Transportation Research Board of the National Academies. <http://www.national-academies.org/trb/bookstore>
6. Fitzpatrick K, Shamburger CB, Krammes RA, Fambro DB (1997) Operating speed on suburban arterial curves. *Transp Res Rec* 1579:89–103. <https://doi.org/10.3141/1579-11>
7. Gargoum SA, El-Basyouny K (2016) Exploring the association between speed and safety: a path analysis approach. *Accid Anal Prev* 93:32–40. <https://doi.org/10.1016/j.aap.2016.04.029>
8. Geometric design guide for Canadian roads (2017). Transportation Association of Canada
9. Hassan Y, Sarhan M (2011) Modeling operating speed. <http://doi.org/10.17226/22864>
10. HCM 2010: highway capacity manual (2010). National Research Council (U.S.). Transportation Research Board
11. McCullagh P, Nelder JA (1989) Monographs on statistics and applied probability. Generalized linear models, 37
12. Medina AMF, Tarko AP (2005) Speed factors on two-lane rural highways in free-flow conditions. *Transp Res Rec* 1912(1):39–46. <https://doi.org/10.3141/1912-05>
13. Moses R, Mtoi E (2013) Evaluation of free flow speeds on interrupted flow facilities. In: FAMU-FSU College of Engineering Tallahassee (ed) BDK83 977-18. Issue May. http://books.google.co.in/books/about/Evaluation_of_Free_Flow_Speeds_on_Interr.html?id=VceNnQEACAAJ&pgis=1
14. Moses R, Mtoi ET, Ozguven EE (2014) Evaluation of the 2010 highway capacity manual urban street free-flow speed prediction model. *Transp Res Rec* 2461:1–8. <https://doi.org/10.3141/2461-01>
15. Neighbourhood traffic reviews (2018). City of Saskatoon. <https://www.saskatoon.ca/moving-around/driving-roadways/managing-traffic/traffic-studies/neighbourhood-traffic-reviews>
16. Spiegelhalter D, Thomas A, Best N, Lunn D (2005) WinBUGS user manual. MRC Biostatistics Unit, Cambridge
17. Tarris JP, Poe CM, Mason JM, Goulias KG (1996) Predicting operating speeds on low-speed urban streets: regression and panel analysis approaches. *Transp Res Rec* 1523:46–54. <https://doi.org/10.3141/1523-06>
18. Thiessen A, El-Basyouny K, Gargoum S (2017) Operating speed models for tangent segments on urban roads. *Transp Res Rec J Transp Res Board* 2618(1):91–99. <https://doi.org/10.3141/2618-09>
19. Wang J, Dixon KK, Li H, Hunter M (2006) Operating-speed model for low-speed urban tangent streets based on in-vehicle global positioning system data. *Transp Res Rec J Transp Res Board* 1961(1):24–33. <https://doi.org/10.1177/0361198106196100104>

Exploring the Linkage Between Human Factors and Road Geometric Elements Influencing the Road Traffic Accidents on the National Roads of South Africa



D. Das

1 Introduction

Road traffic accidents across the world particularly in developing countries is a serious and recurring challenge. Over the years it has become very critical on the national roads/highways [17]. Some of the important challenges attributed to the occurrence of enormous traffic accidents in developing countries in contrast to developed countries are the lack of advanced infrastructure, lack of stronger legislation, and poor road safety campaigns [22]. Further, human factors such as driver and road user behaviour and errors as well as geometric elements of the roads are recognised as significant reasons for the occurrence of accidents. Developed countries have designed and implemented different strategies to reduce the scale and severity of this problem through education, enforcement of regulations and engineering solutions [16]. However, it remains a major concern in developing countries, because of the severity of the ever-growing challenge, which warrants the development of feasible countermeasures within the limited resources [3].

From the point of view of road infrastructure, appropriate design of roads is argued to be one of the most essential requirements to improve road safety [3, 17, 20]. According to Dwikat [10], consistency in design and user-friendliness is necessary to improve safety on roadways. Similarly, human factors influence the safe and efficient movement of vehicles on the roads, which have a huge impact on the occurrence of road accidents [23]. So, it is argued that road traffic accidents remain a serious challenge as a result of both design inconsistencies in road geometric parameters and errors emanated from the human factors and driver behaviour, although other

D. Das (✉)

Civil Engineering, Sustainable Transportation Research Group, University of KwaZulu Natal,
Durban 4041, South Africa
e-mail: dasd@ukzn.ac.za

vehicular, traffic, land use and road-related factors can contribute to the occurrence of traffic accidents.

In the context of South Africa, this challenge has risen to a significant proportion in recent years. A National Household Transport Survey (NHTS) conducted in 2013 indicated that the percentage of car ownership had risen from 22.9% in 2003 to 32.6% in 2013. Subsequently, there were more vehicles and more drivers, which contributed to the increasing traffic accidents on South African roads. For instance, a considerable portion of the high mortality rate (31.79 per 100,000 people) that is being experienced in the country is contributed by road traffic accidents (9.3%) [19]. It is also found that young people ranging from the age group 15–44 years are the major casualties [26]. National Roads such as the N1, N3, and N4 roads in South Africa are found to contribute significantly to this high accident and mortality rates in the country [6]. Thus, there is a need to explore why such a high rate of accidents occur on the national roads of the country. In this context, the road users and drivers often allege the poor design of the roads, which influence driving and sometimes lead to the occurrence of accidents, particularly on the national roads, although all the national roads have been presumed to be designed by following design standards and norms. Apparently, there seems to be a gap between the design of roads and the perceptions of the road users (drivers) and perhaps there might be a linkage between the road geometric elements and human factors including driver behaviour on the roads, which leads to road traffic accidents [14, 24].

Therefore, by using the study context of a national road (N1) and perceptions of the road users and drivers of South Africa, the study examined the various human and road geometric factors that influence the occurrence of accidents on the national roads and explored linkages between human factors and road geometric elements.

A survey research method was used to collect both quantitative and qualitative primary data on the human factors, and their influence on the road geometric parameters and the occurrence of accidents from the road users and drivers. The study revealed that the major human factors that influence the occurrence of accidents include the driver reaction time, the influence of alcohol and nausea, driving competency, awareness about traffic rules and regulations, awareness about the road features, and awareness about the traffic conditions. These human factors perceived to cause certain driver behaviours such as speeding, recklessness and overconfidence, uncertain driving, poor control of vehicles, lack of respect towards traffic and rule and regulations, which cause accidents on the national roads. Furthermore, shoulder width and condition, merging of lanes, alignment, decision sight distance, stopping sight distance and curvature, are the influential road geometric elements that are the cause of concern for the drivers. As perceived by the drivers and road users, these geometric elements are linked to different driver behaviour and contribute to the occurrence of accidents on the national roads.

The findings of the paper will enable understanding of the relationship between human factors and geometric elements of national highways from the road user and drivers perspectives, which will assist in taking remedial measures in terms of improving human behaviour and design of geometric elements of the roads to reduce traffic accidents.

2 Lessons from Literature

According to the World Health Organisation, road accidents are among the leading causes of mortality for people between the ages of 15–44 [13, 26]. Given the current trend, accident fatalities are projected to become the fifth leading cause of death worldwide by 2020, resulting in 2.4 million deaths worldwide per year [9, 26]. However, the causes of road traffic accidents are related to humans, road and vehicles. At the design stage of a road, it is important from the road safety point of view to establish a harmony between the human factors and the road and vehicle-related factors [14, 21, 24]. Scholars have argued that though, human factors are more dominant than the road and vehicle factors, by making the geometrically acceptable design of roads, it is possible to compensate for the other factors and thus decrease the number of road traffic accidents [9, 12]. This makes it apparent that to establish road safety, it is important to design the geometrical parameters of the roads appropriately.

Watters [25] suggested that although several factors influence the occurrence of traffic accidents that need to be considered while developing strategies for improving road safety, the relationships between road safety and road geometric design elements can be seen intuitively as the first approach. However, the important point is to determine the level of these relationships quantitatively. Although the relationships generally show the same tendency, their levels vary according to road characteristics and conditions.

Studies by Gomes [12] and Dinga [8] have also shown that an improvement of infrastructure can also lead to an increase in road traffic accidents. This may be due to road users believing that better-equipped roads are inherently safer, thus, they become careless and less adherent to road operatives [12]. However, the inadequate provision of road infrastructure leads to an even greater increase in road accidents [8].

Literature also suggests that specific geometric factors have been found to influence road users/drivers differently, especially in relation to their age and driving manner [2, 5]. For example, younger and older drivers are most likely to be involved in accidents than middle-aged drivers when experiencing heavy traffic volume, smaller lane widths and multiple lanes [2]. Similarly, alignment consistency represents a key issue in modern highway geometric design [1, 11]. A consistency in alignment would allow most drivers to operate safely at their desired speed along the entire alignment. Additionally, it is also found that in some cases existing design speed-based alignment policies permit the selection of a design speed that is less than the desired speeds of the majority of drivers [15]. Thus, the two foremost factors that play a role in the occurrence of traffic accidents are related to human factors such as driver behaviour, and the road geometric parameters. It is also suggested that the road safety improvement approaches should be composed of three key aspects: safe drivers, safe vehicles and safe roads [22]. Thus, if a hazardous location is to be identified and develop effective strategies to improve road safety, there is a need to consider all three of the above-mentioned aspects [17].

Geometric design consistency evaluations are a widely used method of determining sections of roads that require improvement. This method identifies geometric inconsistencies on roads by means of design evaluation criteria to conduct a geometric design consistency evaluation [1, 25]. A road with acceptable geometric design elements such as the average degree of horizontal curvature, increase in the number of lanes, increase in lane and shoulder width, appropriate of intersections etc., is important to reduce traffic accidents [4, 17, 20]. While the development of safe vehicles remained an exogenous factor for road safety being looked after by automobile engineers and companies, the human factor remains another important parameter in achieving road safety [9, 18]. Although some researches have been conducted with regards to examining the causes of road traffic accidents on the roads of South Africa [7], not many significant researches have been conducted with regards to the national roads. Also, literature revealed that the factors are mostly examined in isolation, for example, the influence of geometric parameters or human factors on road safety. However, no interlinkage among the factors such as between the geometric factors and human factors have been established in the South African context. Besides, unpacking the behaviour of the road users and drivers and their relationship with other factors is not an easy task, so it is argued that correcting those inconsistencies and deficiencies based on the interlinkages remains an important step for road accident reductions [2]. So, exploring the interlinkage between the road geometric parameters and human factors should provide a basis for the development of remedial measures. Therefore, analyses of perceptions of drivers and road users who have driven through the national road N1 was conducted to explore what road geometric and human factors and how they are interlinked and contribute to the occurrence of accidents on the national roads.

3 Study Context and Research Methods

3.1 Study Context

The section of the National Road N1 between the city of Bloemfontein and Johannesburg has considered for the purpose of this investigation, specifically for the collection of data. The N1 is a national route that connects the city of Cape Town of South Africa and the Zimbabwean border at Beit Bridge passing through important cities such as Bloemfontein, Johannesburg, Pretoria, Polokwane, etc. It covers a distance of 1929 km (Kms). Generally, it constitutes a single carriageway with two or more lanes, however, in certain sections of the road, it becomes a dual carriageway road. Also, certain parts of the road have been developed as Freeway with two or more lanes. Specifically, the chosen section for investigation constitutes both single carriageway and dual carriageway at different sub-sections. The maximum speed limit on the road is 120 km per hour in general although the speed limits could range from 60 to 80 km at different road sections depending on the road or the surrounding

conditions. The road is fully paved with asphalt, although some stretches are made up of concrete pavements. All the road elements, and signage and road markings are observed on the road. However, the road experiences a very high accident rate. The various types of accidents include but not limited to rear-end crashes, overturns, rollovers, falling in the side drains or ditches, side-on collisions between two vehicles, etc. According to the Road Traffic Management Corporation (RTMC), at a national level, human factor contributed to about 91% of the accidents, and 7.2% of the accidents have occurred because of environmental factors. Also, male drivers contribute to about 72% of accidents, which is also on the rise. Therefore, it is ascertained that the human factor and driver behaviours are the major reasons for the accidents on the roads of South Africa. However, arguments have emerged that the road geometric factors engender certain kind of driver behaviour, and thus this study.

3.2 Research Methods

A mixed-method of research that includes both quantitative and qualitative methods were used for this study. The mixed-method was adopted for two reasons. First, not much structured statistical data was available relating to the occurrence of accidents and different causes of the accidents on the national roads. Secondly, the study focused on the human factors and driver behaviour, which need opinions and perceptions of the drivers and road users and their experiences. A questionnaire survey was used to collect both quantitative and qualitative primary data on the human factors, and their influence on the road geometric parameters and the occurrence of accidents from the drivers and road users. The questionnaire survey was conducted among the drivers and road users who have driven or travelled on the road section between Bloemfontein and Johannesburg on the N1 route. The respondents were selected randomly based on their willingness to participate in the survey. A total number of 210 (125 from Bloemfontein and 85 from Johannesburg) respondents participated in the survey. Of the total respondents 82% (172 respondents) have driven on the road and the rest 18% were travellers, who have passed through the road. Key attributes on which the respondents' perception and opinion were sought included, different parameters which cause accidents, different road parameters which influence driver behaviour while driving on the road, the mental and physical conditions that influence driving and decision making, the relationship between different road parameters and driver behaviour, in other words, how do they react or behave under different road conditions, etc. The data on the perceptions of the respondents were collected through a five-point Likert scale ranging between 1 and 5 in which 1 indicate not influential, 2 indicates marginally influential, 3 indicates fairly influential, 4 indicates significantly influential and 5 is most influential. The respondents were asked to respond based on their own experiences than on theoretical perspectives.

Further, qualitative discussion with 30 people (identified as respondent 1–30) who took part in the survey was conducted, to provide their opinion on the practical occurrences on the road, traffic safety, status of geometric elements and human factors and

driver behaviour related to traffic accidents. This information was collected to corroborate the findings from the survey. Also, physical road geometric data were collected from different sections of the road through physical survey and observations.

The major human factors and road geometric elements that influence traffic accidents on the national roads were established using the perception index method and significant tests (Z test). Further, the linkages between various human factors and road geometric elements were established using significance tests (Z test) and narrative analyses of the opinions collected from qualitative discussions.

The survey data was used to evaluate the perception index (PI) of various parameters that cause accidents. The PI was developed by use of the mean score of the Likert scale. PI was obtained by using Eq. (1).

$$PI = \frac{\sum_1^n LI}{N} \quad (1)$$

where

LI Likert index assigned by each respondent.

N Total number of respondents.

However, standard deviation (σ) was used to check the consistency of the responses based on which PI was developed. The σ values were calculated by Eq. (2)

$$\sigma = \frac{\sqrt{\sum (LI - LI(M))^2}}{N} \quad (2)$$

where

LI Likert index assigned by each respondent.

LI (M) Mean value of all Likert indices assigned by each respondent.

N Total number of respondents.

Further, significant test (Z test) was conducted and z probability values were obtained to check the veracity and acceptability of the PI and establish the influence of different road geometric factors and human factors on the accidents. The Z score was obtained by use Eq. 3. The z probability values were obtained from Z probability tables.

$$Z_s = \frac{PI - LI(X)}{\sigma} \quad (3)$$

where

Zs Z score.

PI Perception index of each factor.

LI(X) Mean value of all Likert indices in the Likert scale.

σ Standard deviation.

4 Results and Findings

The question under investigation is whether human (driver) behaviour is linked to the geometric parameters of the road that lead to the occurrence of accidents. In order to examine the interlinkage, this study first examined the various human factors and driver behaviours and geometric parameters that influence the occurrence of accidents independently and then explore the linkage between different geometric parameters and human behaviour, which influence the occurrence of the accidents on the national road N1. The results of the three aspects are discussed in the following subsections.

4.1 Influence of Major Human Behavioural Factors Concerning the Occurrence of Accidents on N1

Table 1 presents the major human behavioural factors, which influence the occurrence of traffic accidents on the N1 road. According to the respondents, with perception indices greater than 4, the reaction time of the drivers (PI = 4.25), and influence of alcohol and nausea (PI = 4.12) are the major human factors that influence the occurrence of the accidents. Also, the high z probability values (z probability > 0.96) indicate that these factors are the most likely causes of accidents from the human factor point of view. While health condition of the drivers or road users do not influence the occurrence of accidents, with moderate PI ($3 \leq PI \leq 4$) and z probability

Table 1 Influence of major human behavioural factors concerning the occurrence of accidents on N1

| Human factors | PI | SD | Z score | Z probability values |
|--|------|------|---------|----------------------|
| Driver reaction time | 4.25 | 0.67 | 1.87 | 0.969 |
| Driving competency | 3.84 | 0.52 | 1.62 | 0.947 |
| Awareness about traffic rules and regulations | 3.64 | 0.54 | 1.19 | 0.882 |
| Awareness about the road features | 3.23 | 0.56 | 0.41 | 0.659 |
| Awareness about the traffic conditions | 3.62 | 0.51 | 1.22 | 0.888 |
| Health condition | 2.76 | 0.42 | -0.57 | 0.284 |
| Under the influence of alcohol and nausea | 4.12 | 0.58 | 1.93 | 0.973 |
| <i>Driver behaviour</i> | | | | |
| Speeding | 4.30 | 0.67 | 1.94 | 0.973 |
| Recklessness and overconfidence | 4.24 | 0.62 | 2.00 | 0.977 |
| Uncertain driving | 4.03 | 0.63 | 1.63 | 0.948 |
| Lack of respect towards traffic and rule and regulations | 3.42 | 0.54 | 0.78 | 0.782 |
| Poor control of vehicles | 3.82 | 0.58 | 1.41 | 0.920 |

(≥ 0.6) values, competency of drivers (PI = 3.84, z probability = 0.947), awareness about traffic rules and regulations (PI = 3.64, Z probability = 0.882), awareness about the traffic conditions (PI = 3.62, Z probability = 0.888), and awareness about the road features (PI = 3.23, z probability = 0.659) influence the occurrence of accidents to a moderate extent. Similarly, from the driver behaviour point of view, speeding (PI = 4.30), recklessness and overconfidence (PI = 4.24), and uncertain driving (PI = 4.03) are the most influential behavioural aspects which are perceived to cause accidents. The high Z probability values (z probability > 0.94) support the findings. However, lack of respect towards traffic and rule and regulations (PI = 3.42, Z probability = 0.782) also contribute to the occurrence of accidents to a certain extent.

These findings are also corroborated by the findings of the narrative analysis. For example, respondent 7, who is an experienced driver opine that:

...I have seen that people particularly some young people drive very fast without giving any respect to the speed limit and surrounding conditions and lose control and meet accidents.

Another respondent (respondent 12) said that:

.... Some drivers react late to the situation, for example, while overtaking a number of vehicles or large vehicles or while being overtaken by a very high speeding vehicle and end up moving sideways or breaking...consequently either hit other vehicles on the sides or slide and end up in the ditch.

On the influence of health condition on accidents, a number of respondents (13,17, 21) opined that:

... Most people generally do not drive when they are sick and when they do, they are perhaps very careful.

4.2 Design Elements Concerning the Occurrence of Accidents

The influence of the road geometry factors on the occurrence of accidents as obtained from the survey is presented in Table 2. The design of road elements has been done in concurrence with the standards and norms and design criteria used by South African National Roads Limited (SANRAL manual for the national roads). It was observed that the majority of the sections of N1 are of the single carriageway and with more than 2 lanes. Some of the sections have dual carriageways without medians, however, demarcated by pavement markings. Besides, the major intersections are grade-separated but the minor intersection, i.e., the intersections joining N1 and minor roads are at the same level. However, adequate pavement markings are provided for the flow of traffic. One of the major observations is that large sections of the road have straight and long alignment, offering adequate stopping sight distance. The curves are observed to be mostly gentle and according to the design standards. Reserve lanes indicated by yellow lines ranging between less than 0.5 m (m) to

Table 2 Influence of major road design elements concerning the occurrence of accidents on N1

| Design elements | Status | PI | SD | Z score | Z probability values |
|-----------------------|---|------|------|---------|----------------------|
| Road width | 7.0–7.5 m (two lanes) 7.0–10.5 (three lanes) | 2.12 | 0.43 | −2.05 | 0.0201 |
| Lane width | 3.5–3.75 m | 2.74 | 0.51 | −0.51 | 0.305 |
| Number of lanes | 2–3 | 2.76 | 0.48 | −0.50 | 0.308 |
| Shoulder width | 2.0–3.0 m | 4.11 | 0.68 | 1.63 | 0.948 |
| Reserve lanes | 1.5–3.0 m | 2.45 | 0.56 | −0.98 | 0.163 |
| Type of intersections | Grade separated: At the crossing of the major roads | 1.55 | 0.45 | −3.22 | 0.0006 |
| | At the level: At the crossing of minor roads (road marking provided) | 3.12 | 0.56 | 0.21 | 0.583 |
| Kerbs | | 2.83 | 0.47 | −0.28 | 0.389 |
| Medians | Lack of medians in dual carriageways | 3.83 | 0.54 | 1.54 | 0.938 |
| Sight distance | Stopping sight distance: adequate: according to design standards on straight alignments | 3.78 | 0.62 | 1.26 | 0.916 |
| | Decision sight distance: uncertain according to the drivers | 4.01 | 0.69 | 1.46 | 0.927 |
| Curvature | Generally moderate to low curvature | 3.58 | 0.56 | 1.04 | 0.854 |
| Merging of lanes | On specific intervals of about 1–3 km distance | 4.05 | 0.58 | 1.81 | 0.964 |
| Alignment | Generally long and straight | 3.86 | 0.52 | 1.65 | 0.950 |

more than 2.0 m widths (at different sections) are provided throughout the road. The merger of lanes and divergence of lanes have been created at a distance of 1.5–3.0 km to (approximate in general) to facilitate overtaking by fast vehicles. The distance provided for merger and divergence of lanes is about 400. Despite the various provisions and adherence to design standards, the road users and drivers perceived that certain geometric elements of the road contribute to the occurrence of the accidents both directly and indirectly. According to the respondents' perceptions, road width, lane width, number of lanes, reserve lanes, kerbs, and grade-separated intersections ($1.5 < PI < 3.0$; z probability < 0.5) do not influence the occurrence of accidents to a large extent. However, shoulder-width ($PI = 4.11$), merging of lanes ($PI = 4.05$) and decision sight distance (particularly at high speed) ($PI = 4.01$) perceived to influence the occurrence of accidents significantly. Followed by, very long and straight alignment, ($PI = 3.86$), lack of medians in dual carriageway sections ($PI = 3.83$),

stopping sight distance (PI = 3.78), curvature (PI = 3.58) and intersections (at the same level) (PI = 3.12) are perceived to influence accidents fairly. The findings are supported by the high z probability values (>0.885). Also, the narrative analysis of the qualitative discussions corroborated these findings.

As three respondents (3, 20, 24) pointed out that:

...The roads are wide with many lanes and reserve lanes.... drivers [we] do not face any kind of problems while driving. However, at some sections, shoulders are dangerous, do not have enough width, not well maintained... in some places, there are obstacles, if one skid towards shoulders, may fall in a ditch if not careful.

According to respondents 12, 27, 29, and 30 stopping sight distance and curvature can contribute to the occurrence of accidents to a certain extent. As per their opinion:

...some drivers drive very fast and at a high speed.... maybe beyond 120 km/hour, it very difficult to stop a car suddenly. Also, in the curved sections, it is difficult to control a car at a very high speed and may meet accidents.

Similarly, about 7 of the respondents (4, 9, 12, 15, 16, 23, 28) argued that the merger of lanes is the most dangerous and likely cause of accidents. According to them:

...Merging of lanes creates uncertainty among the drivers [us] when driving parallel or alongside large vehicles such as large goods trucks. [We] are not sure ...to pass the truck or wait as trucks travel at a particular speed. In case the trucks do not allow us to move ahead near the mergers, there is a possibility of ending on the shoulders, which can be dangerous. A bad and narrow shoulder with an inconsiderate large truck driver may become dangerous for the smaller vehicles like cars particularly at sections leading to merging of lanes.

4.3 The Perceived Linkage Between Road Geometric Elements on Human Factors and Driver Behaviour

The perceived interlinkages between the road geometric elements and human factors and behaviour are presented in Table 3. In the absence of statistical data to establish the empirical relationship, the linkages were established based on perceptions index and agreements among the majority of the respondents. It is observed that with high perception indices, straight alignment is highly linked to recklessness and speeding, which are found to be amongst two major human behavioural reasons for accidents on the road. Similarly, long sight distance on account of straight alignment also causes recklessness among the drivers to a large extent. Besides, it can cause complacency and relaxation moderately among the drivers, which may lead to the occurrence of accidents. Long sight distance in straight alignments is perceived to be linked to the recklessness of the drivers highly and complacency moderately. On the other hand, sight distance at the curved sections of the road can cause uncertain driving and poor vehicle control to a moderate extent. Merging of lanes perceived to engender uncertain driving, speeding and poor reaction time at high speeds to a moderate extent. Curved sections also engender uncertain driving and can be attributed to poor

Table 3 Linkage of geometric elements with human factors and driver behaviour

| Design elements | Human factors and behaviour | Number of respondents agreed on the linkage | Rating score (1–5) | Significance |
|---------------------------------------|-----------------------------|---|--------------------|-------------------|
| Alignment | Recklessness | 91 | 4.21 | Highly linked |
| | Complacency | 72 | 3.98 | Moderately linked |
| | Speeding | 102 | 4.32 | Highly linked |
| Merging of lanes | Uncertain driving | 83 | 3.69 | Moderately linked |
| | Speeding | 62 | 3.24 | Moderately linked |
| | Poor reaction time | 72 | 3.66 | Moderately linked |
| Curvature | Uncertain driving | 76 | 3.45 | Moderately linked |
| | Poor reaction time | 67 | 3.24 | Moderately linked |
| Sight distance on straight alignments | Recklessness | 84 | 4.03 | Highly linked |
| | Complacency | 79 | 3.86 | Moderately linked |
| Sight distance on curves | Uncertain driving | 62 | 3.12 | Moderately linked |
| | Poor control vehicles | 64 | 3.27 | Moderately linked |
| Shoulder condition | Uncertain driving | 78 | 3.84 | Moderately linked |
| | Poor control of vehicles | 76 | 3.82 | Moderately linked |
| Shoulder width | Uncertain driving | 87 | 3.98 | Moderately linked |
| | Poor control of vehicles | 92 | 4.26 | Highly linked |
| Number of lanes | Recklessness | 82 | 3.66 | Moderately linked |
| | Speeding | 76 | 3.61 | Moderately linked |
| | Unnecessary overtaking | 68 | 3.44 | Moderately linked |
| Type of intersections: at the level | Uncertain driving | 69 | 3.32 | Moderately linked |

(continued)

Table 3 (continued)

| Design elements | Human factors and behaviour | Number of respondents agreed on the linkage | Rating score (1–5) | Significance |
|---|-----------------------------|---|--------------------|-------------------|
| Reserve lanes | Speeding | 62 | 3.12 | Moderately linked |
| Combination of merging of lanes and poor shoulder width and condition | Uncertain driving | 82 | 4.08 | Highly linked |
| | Speeding | 79 | 4.03 | Highly linked |
| | Poor control of the vehicle | 76 | 4.01 | Highly linked |

reaction time. It is also found that shoulder width is related to poor vehicle control and uncertain driving moderately. However, while shoulder condition is moderately linked to uncertain driving, it is highly linked to poor vehicle control that might lead to accidents. The number of lanes is related to speeding, recklessness and unnecessary overtaking to a certain extent (moderately). Intersections at the level of roads are linked to uncertain driving moderately. Also, reserve lanes are linked to speeding to a moderate extent. However, most importantly a combination of merging of lanes and poor shoulder width and condition is highly linked to uncertain driving, speeding and poor vehicle control and could create lethal conditions for the occurrence of accidents.

5 Discussion and Conclusions

Traffic accidents on the national roads of South Africa is a huge challenge. A lot of critical accidents and fatalities are being experienced every year. The section of the national road N1 between Bloemfontein and Johannesburg also experiences a significant number of accidents. Apparently, the road has been designed according to the design standards of the national roads and care is taken to provide all pavement and roadside information, through adequate and appropriate signage, and pavements markings. The carriageway is also well maintained. Despite the various provisions that have been made, accidents are observed to be regular phenomena. Although several reasons for the accidents have been identified at the national level, no specific study regarding the causes of the accidents on the national roads has been conducted so far. At the national level, human factors have been attributed as the major causes of accidents. However, arguments have emerged that there might be a relationship between the road geometric factors, which engender a certain driver (human) behaviour that leads to accidents. Therefore, this study examined the linkage between geometric elements and human factors and driver behaviour, which influence the occurrence of accidents. In the absence of structured statistical data, the study relied

on primary data obtained from a survey among the road users including the drivers and both qualitative and quantitative interpretation of the data.

Findings suggested that driver reaction time, the influence of alcohol and nausea, driving competency, awareness about traffic rules and regulations, awareness about the road features, awareness about the traffic conditions are the major human factors that cause accidents. The human factors apparently engender certain driver behaviour such as speeding, recklessness and overconfidence, uncertain driving poor control of vehicles, lack of respect towards traffic and rule and regulations, which cause accidents on the national roads. Similarly, shoulder width, merging of lanes, alignment, decision sight distance, stopping sight distance and curvature, are the major geometric elements of the roads that are causes of concern for the drivers. As perceived by the drivers, these geometric elements engender different driver behaviour. For example, a very long and straight alignment makes the driver complacent and speeding. Also, at a higher speed, the reaction time reduces, leading to a higher likelihood of the occurrence of accidents. Similarly, sight distance on acutely curved sections is a cause of concern and lead to uncertain driving and poor vehicle control. Moreover, the shoulder width and shoulder condition play critical roles in driver behaviour in terms of uncertain driving and poor vehicle control. Merging of lanes perceived to engender uncertain driving, sometimes speeding and poor reaction time. Specifically, car drivers tend to become uncertain whether to move past or wait when a larger vehicle (large goods trucks) is moving alongside the car. This situation occurs as the car driver becomes uncertain about the behaviour of the truck drivers as to whether the truck driver would hold the lane and do not allow the fast car driver to pass or slow down to allow the cars to pass. In this regard, no specific rules of right of way are seemed to be adhered to by any of the drivers concerned in such situations. Consequently, the drivers become uncertain and sometimes try to speed up. In such a scenario, if one of the drivers do not yield, the driver on the outer lane swerve towards the shoulder. However, if the shoulder is not wide enough and in bad condition, then it becomes a lethal combination leading to loss of control and accidents.

Thus, it is observed that according to the perceptions of the road users, geometric elements and human (driver) behaviour are interlinked, which cause accidents on the national roads. However, due to limitations of availability of structured statistical data and long-term practical observation, those linkages could not be empirically established, which is the further scope of the research. However, this study suggests that while designing the national roads the linkage between human factors and driver behaviour and critical geometric elements of roads needs to be considered while designing the road elements to improve road safety.

References

1. Abdel-Aty M, Chen C, Radwan E, Brady R (1999) Analysis of accident-involvement trends by driver's age in Florida. *ITE J Web* 96:69–74
2. Abdel-Aty M, Radwan E (2000) Modelling traffic accident occurrence and involvement. *Accid Anal Prev* 32:633–642
3. Berhanu G (2004) Models relating traffic safety with road environment and traffic flows on arterial roads in Addis Ababa. *Accid Anal Prev* 36:697–704
4. Bester CJ, Makunje JA (1998) The effect of rural road geometry on safety in Southern Africa. *Transportation research circular 1 EC003*, vol 15. Transportation Research Board, Washington DC, USA. See <http://worldcat.org/issn/00978515>, pp 1–10
5. Briz-Redón Á, Martínez-Ruiz F, Montes F (2019) Spatial analysis of traffic accidents near and between road intersections in a directed linear network. *Accid Anal Prev* 132:105252. <https://doi.org/10.1016/j.aap.2019.07.028>
6. CSIR (2016) Development of a methodology for the calculation of the costs of crashes in South Africa. Project report phase 2, built environment: transport management, design and systems
7. Das D, Burger E (2017) Appraisal of urban road safety factors in South Africa. In: *Proceedings of the Institution of Civil Engineers—municipal engineer*, vol 170, pp 6–15. <https://doi.org/10.1680/jmuen.15.00022>
8. Dinga N (2014) Blackspot analysis: infrastructure impacts on the occurrence of blackspots on the N1 highway in the Western Cape. Masters Dissertation, University of Capetown
9. Dinu RR (2012) Studies on safety performance of two-lane rural highways in heterogeneous traffic. Ph.D. thesis, Indian Institute of Technology Madras, Chennai
10. Dwikat MG (2014) Modelling relationships between geometric design consistency and road safety for two-lane rural highways in West Bank. Master's degree dissertation, Najah National University
11. Gitelman V, Balasha D, Carmel R, Hendel L, Pesahov F (2012) Characterization of pedestrian accidents and an examination of infrastructure measures to improve road safety in Israel. *Accid Anal Prev* 44:63–73
12. Gomes S (2013) The influence of the infrastructure characteristic in urban road accidents occurrence. *Accid Anal Prev* 60:289–297
13. Gregoriades A, Mouskos KC (2013) Black spots identification through a Bayesian networks quantification of accident risk index. *Transp Res Part C* 28:28–43
14. Karlaftis MG, Golias I (2002) Effects of road geometry and traffic volumes on rural roadway accident rates. *Accid Anal Prev* 34:357–365
15. Mohammed HA (2013) The influence of road geometric design elements on highway safety. *Int J Civ Eng Technol* 44:146–162
16. Mohamed R, Vom Hofe R, Mazumder S (2014) Jurisdictional spillover effects of sprawl on injuries and fatalities. *Accid Anal Prev* 72:9–16
17. Nadesan Reddy N, Knight S (2013) The effect of traffic calming on pedestrian injuries and motor vehicle collisions in two areas of the eThekweni municipality: a before-and-after study. *Discipline of Public Health Medicine, School of Nursing and Public Health, University of KwaZulu-Natal, Durban, South Africa*, 103 No. 9 SAMJ
18. Noland R (2003) Traffic fatalities and injuries: the effect of changes in infrastructure and other trends. *Accid Anal Prev* 35:599–611
19. RTMC (2011) Road traffic report. Road Traffic Management Cooperation, December 2010
20. Thwala B (2012) Dancing with dirty road traffic accidents data: the case of Gauteng
21. Uzundu C, Jamson S, Lai F (2018) Exploratory study involving observation of traffic behaviour and conflicts in Nigeria using the traffic conflict technique. *Saf Sci* 110(Part A):273–284
22. Vanderschuren M, Jobanputra R (2009) Traffic calming measures: review and analysis
23. Vanderschuren M, Jobanputra R (2011) Hazardous road safety location analysis: a case study of the Western Cape. In: *Proceedings of the 30th Southern African transport conference. Document Transformation Technologies*

24. Vayalamkuzhi P, Amirthalingam V (2016) Influence of geometric design characteristics on safety under heterogeneous traffic flow. *J Traffic Transp Eng (English Edition)* 3:559–570
25. Watters P (2007) The relationship between geometric design consistency and safety on rural single carriageway in Ireland. The University of Dublin, Trinity College
26. World Health Organization (WHO) (2011) Global plan for the decade of action for road safety 2011–2020

Development of an AI Tool to Identify Reference Reaches for Natural Channel Design



C. Kupferschmidt and A. Binns

1 Introduction

River channel morphological patterns and classification schemes have long been the subject of investigation from academic and practical perspectives. While rivers are complex systems and no two rivers are identical, in many cases it is desirable to group rivers into similar categories that generalize their morphologies and behaviours. Numerous classification schemes have been developed for these purposes, with commonly used examples including the Brice [1] or modified Brice [19], Rosgen [15, 16], and Montgomery and Buffington [11] classification systems. Most existing classification schemes tend to be hierarchical in nature, with some systems categorizing rivers based predominantly on channel planform geometry, while others consider additional variables such as bed material, reach or valley slope, and channel cross-sectional geometry.

The use of classification systems is particularly important in the fields of geomorphology and natural channel design, where they are often used to identify reference reaches. According to Rosgen [17], a reference reach is "... a river segment that represents a stable reach within a particular valley morphology." The purpose of identifying one or more reference reaches is to allow existing channels or proposed designs to be quantitatively or qualitatively compared with the historical and current behaviour of the reference reaches via an analog comparison. Reference reaches are often used analytically for purposes such as assessing stream stability, inferring geomorphic processes, predicting future geomorphic responses, and guiding stream rehabilitation or restoration activities [7].

By the nature of their development, classification systems are subject to biases imparted by their creators and divisions between classes may be arbitrarily selected

C. Kupferschmidt (✉) · A. Binns
School of Engineering, University of Guelph, Guelph, Canada
e-mail: kupfersc@uoguelph.ca

based on characteristics that can be easily described using conventional subject area terminology or descriptors. While this makes existing classification systems easy to understand and apply, post-construction analyses of channel restoration projects completed using such techniques show that many fail within months or years of construction [10]. In fact, Juracek and Fitzpatrick [7] suggest that the widely used Rosgen Level II classification system [16] which categorizes streams into 41 categories using a two-tiered hierarchical system should not be used for any purposes beyond description and communication of stream characteristics. Juracek and Fitzpatrick [7] identify several limitations of existing classification systems, notably convergence, which they describe as an inability to distinguish between “the production of similar results from different processes and causes.”

Machine learning (ML) is a subset of artificial intelligence (AI) that uses training data to develop algorithms that can be used to generate decisions or predictions. Convolutional neural networks (CNNs) are commonly used in analyzing visual imagery to extract high-dimensionality features [8]. These models are commonly trained using labelled data (supervised learning) to develop tools that can be used for classification. However, if labelled data is not available, CNNs can also be used to extract features (unsupervised learning) for clustering similar examples. While the use of machine learning does not eliminate human bias, it can help to reduce its impact. Additionally, machine learning techniques have the potential to identify patterns or features that may not be intuitive to human users or domain area experts.

In recent years, high-quality digital channel planform geometry data have become increasingly available on national and global scales. In Canada, the National Hydrographic Network [12] contains digitized polylines of stream networks across the entire country, while the National Hydrography Dataset [20] provides similar information for the USA. Decades of raster data of river channels collected via satellite imagery are also available on a global scale through tools such as through the Global Surface Water Explorer [13]. These large datasets are excellent candidates for use with ML techniques such as CNNs. In the current study we present the development and implementation of an unsupervised AI-based tool that uses CNNs and channel planform geometry data to identify n potential reference reaches within a distance d of a given location. The current development version of the AI tool is implemented for a case study of the lower Albany River watershed in northern Ontario, Canada.

2 Methodology

2.1 Data Sources

The National Hydrographic Network [12] is a publicly available dataset that provides geospatial data on Canada’s inland waterways including lakes and rivers. This dataset is available in geodatabase (.gdb), shapefile (.shp), geography markup language

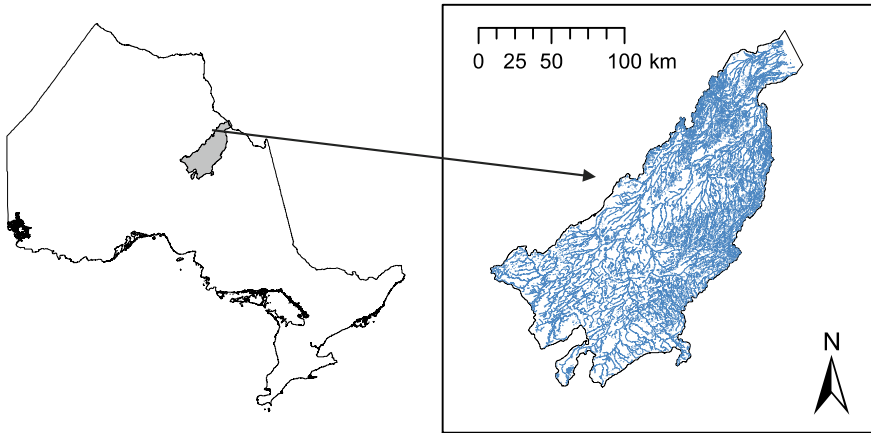


Fig. 1 Location and study area boundary for the lower Albany River watershed in Northern Ontario, Canada (04HA000)

(.gml) and keyhole markup language (.kmz) formats. Within the National Hydrographic Network river planform geometry is represented using polylines that follow the centre of the river channels. Surface water is represented using polygons that delineate the boundaries of water-covered areas.

For the current study we used data for a single watershed (04HA000) located in northern Ontario, shown in Fig. 1. This study area consists of the lower reach of the Albany River watershed, where it flows into James Bay. This watershed was selected because the region has experienced little historical development and there were expected to be minimal human impacts on channel morphology.

2.2 Data Pre-processing

Hydrographic data was downloaded in shapefile (.shp) format. The dataset was pre-processed to convert the data into a suitable format for input to the CNNs. The following steps were used for pre-processing:

- Perform a geospatial clip of the river geometry polyline dataset using the surface water polygon dataset to remove all lakes and other non-river segments.
- Re-project the clipped river geometry dataset into the local UTM coordinate system (EPSG:2958 NAD83 (CSRS) UTM zone 17N was used for the current study area).
- Remove any river polyline segments with fewer than 100 vertices from the dataset.
- Subdivide any river polyline segment with more than 224 vertices into multiple segments with approximately equal numbers of vertices, resulting in between 100 and 224 vertices per segment.

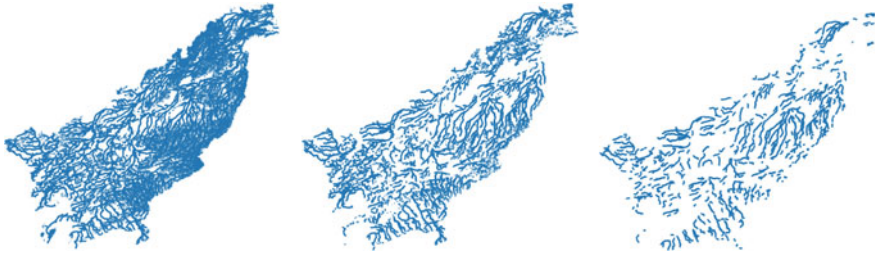


Fig. 2 Stream segment polylines for study area shapefile shown at input (left), after clipping waterbodies (middle) and after applying filtering criteria (right)

- Remove any river polyline segments with a length of less than 500 m.
- Rotate each river segment so that a line drawn between the first and last point is horizontal.
- Scale each river segment so that the maximum height and width do not exceed 224 pixels.
- Export each river segment as a .jpg image with a size of 224 pixels \times 224 pixels where the river segment is centered within the image.

The implementation of these steps on the study area watershed produced 712 input images that met the criteria for use in the current study. The clipped river geometry polyline dataset was then filtered so that only segments meeting the criteria for use in the study remain (Fig. 2).

2.3 *AI Model Development*

For the current study we used TensorFlow (2.4.0-rc0) through the Keras API in Python (3.8.6) to implement two different CNN model architectures: VGG16 [18] and ResNet50 [5], shown in Table 1. Both architectures take a 224 pixel by 224 pixel, 3-channel image as input, are relatively deep, and make use of small convolution filters.

Since the input images produced in the data pre-processing step of the current study are unlabelled, it was not possible to use traditional supervised learning techniques to calibrate the parameters of these models. Instead, we elected to use a technique known as transfer learning, where a model trained on one set of data is used to evaluate a second different set of data.

The models used in the current study were pre-trained on ImageNet, an online database of more than 1.2 million labelled images that represent 1000 distinct classes [4]. It has become common to implement transfer learning by training CNN models on ImageNet and adapting the output features for other tasks [6]. Models trained on ImageNet have been shown to produce good general-purpose features, which allows for these features to be used to evaluate classes not contained in the training dataset.

Table 1 Model architectures for ResNet50 and VGG16

| ResNet50 architecture | VGG16 architecture |
|---------------------------------|--|
| input (224 × 224 RGB image) | input (224 × 224 RGB image) |
| conv {7-64} × 1 | conv {3-64} × 2 |
| maxpool | maxpool |
| conv {1-64, 3-64, 1-256} × 3 | conv {3-128} × 2 |
| conv {1-128, 3-128, 1-512} × 4 | maxpool |
| conv {1-256, 3-256, 1-1024} × 6 | conv {3-256, 3-256, 1-256} × 1 |
| conv {1-512, 3-512, 1-2048} × 3 | maxpool |
| averagepool ^a | conv {3-512, 3-512, 1-512} × 1 |
| FC {1000} ^a | maxpool |
| softmax ^a | conv {3-512, 3-512, 1-512} × 1 |
| | maxpool ^a |
| | FC {4096, 4096, 1000} ^a × 1 |
| | softmax ^a |

^a Layers removed for model use in current study

For example, despite not containing medical images, ImageNet is commonly and successfully used for transfer learning-based medical image analysis [3]. Implementations of VGG16 and ResNet50 that have been pre-trained on ImageNet are available through Keras and can be implemented in an off-the-shelf manner.

Since the purpose of using these models was to perform feature extraction rather than classification tasks related to ImageNet, we removed the heads of the models that contained the fully-connected (FC) and pooling layers, leaving the model base layers behind. Each of the images in the input dataset were fed through the two models, and the output features were used as inputs to the AI tool.

2.4 Model Fine-Tuning

In addition to using the features generated by the pre-trained models, we elected to implement a fine-tuning process using a small subset of the input data, since fine-tuned models can typically achieve higher accuracies than classifiers that use off-the-shelf features [9]. To perform fine-tuning we froze the base layers of the CNN models and implemented new model heads. The outputs of the final convolution layer from the frozen model bases were flattened and fed into a $1 \times 1 \times 512$ fully-connected (FC) layer that uses the rectified linear unit (ReLU) activation function.

Table 2 Implemented model head architecture for fine-tuning

| Model head architecture |
|-------------------------|
| flatten |
| FC-512 |
| softmax ^a |

^a Layer used for training tasks only

This layer was then fed into a softmax layer used for classification. The implemented architecture for the model heads is shown in Table 2.

For the training task, 200 random images were selected from the input dataset and each image was assumed to each represent its own class (i.e. one class per image). Since only one image existed for each of the training classes, we implemented a self-supervised training approach by using a data generator to implement augmentation methods and produce a training set. The implemented augmentations were a combination of random flipping in both the horizontal and vertical directions, and random cropping of between 0 and 40% of the image dimensions. In addition to increasing the size of the training dataset by creating multiple examples per class, this process helps to prevent the model from clustering stream segments based on general features such as flow direction or overall stream segment convexity. The original images were used as a validation dataset.

The model heads were trained for 100 epochs using random initialization with a batch size of 64 images. The training optimizer used stochastic gradient descent and cross-entropy loss to maximize classification accuracy of original images with a momentum of 0.9 and a learning rates of $1e-3$ for ResNet50 and $1e-4$ for VGG16, respectively. After fine-tuning, the softmax layer used for classification was removed from the model, so that the features generated by the final fully-connected layer could be used for reference reach identification. Each of the images in the input dataset were fed through the fine-tuned models, and the output features were used as inputs to the AI tool.

2.5 Reference Reach Identification

To identify reference reaches, the AI tool uses the features output by the CNNs as inputs. For both the models with no fine-tuning and the fine-tuned models, principal component analysis (PCA) is used to reduce the number of features to 100 to reduce computational requirements. Distance matrices are calculated for both the feature-space and physical space for each of the input features. The AI tool uses these distance matrices as input. Please note that solving the distance matrix is an $O(n^2)$ problem that may not be practical to solve for very large study areas or without using PCA.

The developed AI tool allows users to input a list of k study stream segments and specify a number n of reference reaches to return for each input segment, along with the CNN architecture to use for identifying the reaches. Potential reference

reaches are then ranked and the n closest points in the model-space distance matrix are returned using a k -nearest neighbours approach. In the current version of the AI tool, limiting results to stream segments located within a distance d of the study site must be performed in a separate step, which can be implemented either prior to or after returning n ranked reference reaches.

For the current study we used the AI tool to identify reference reaches for six randomly selected study site stream segments within the study area, returning the seven closest reference reaches to each segment. The AI tool was run using four different models: ResNet50 without fine-tuning, VGG16 without fine tuning, ResNet50 with fine-tuning, and VGG16 with fine-tuning.

3 Results and Discussion

The training curves (Fig. 3) show that the training tasks in the fine-tuning process resulted in validation accuracies of about 40% for ResNet50 and 15–20% for VGG16 after 100 epochs. Since the models used 200 classes for training, these results suggest that the models were able to develop considerable skill beyond a random classifier, which would be expected to have an accuracy of approximately 0.5%. This suggests that the fined-tuned AI models were able to extract meaningful features from the stream segments that can be used to identify other similar segments.

Reference reaches identified using the AI tool for the six study stream segments are shown without fine-tuning (Fig. 4) and with fine-tuning (Fig. 5). For the CNNs that did not use fine-tuning, both ResNet50 and VGG16 identified at least one of the same stream segments as reference reaches for all six of the evaluated sites. The fact that both of these models were pre-trained to extract features used to identify ImageNet classes may explain why similar predictions were produced by both networks.

A qualitative comparison of the study sites with the reference reaches in Fig. 4 shows that in some cases the models appear to identify reference reaches with channel

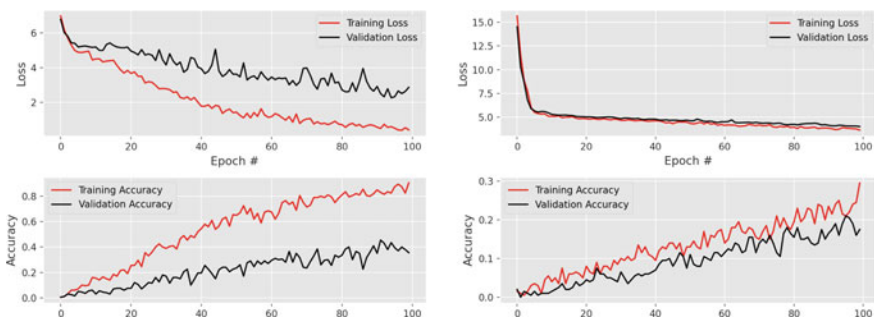


Fig. 3 Training curves for fine-tuning of the ResNet50 (left) and VGG16 (right) models. A learning rate of $1e-3$ was used for ResNet50 and a learning rate of $1e-4$ was used for VGG16. Both models used a batch size of 64 samples and were run for 100 epochs

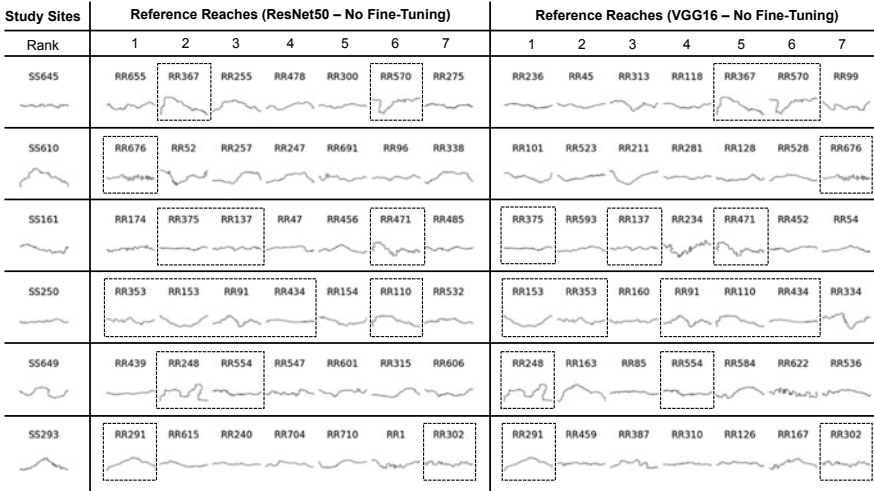


Fig. 4 Ranked reference reach results ($n = 7$) for six study sites identified using the AI tool with ResNet50 (left) and VGG16 (right) convolutional neural networks (CNNs) pre-trained on ImageNet without fine-tuning. Dashed boxes denote reference reaches identified by both CNNs

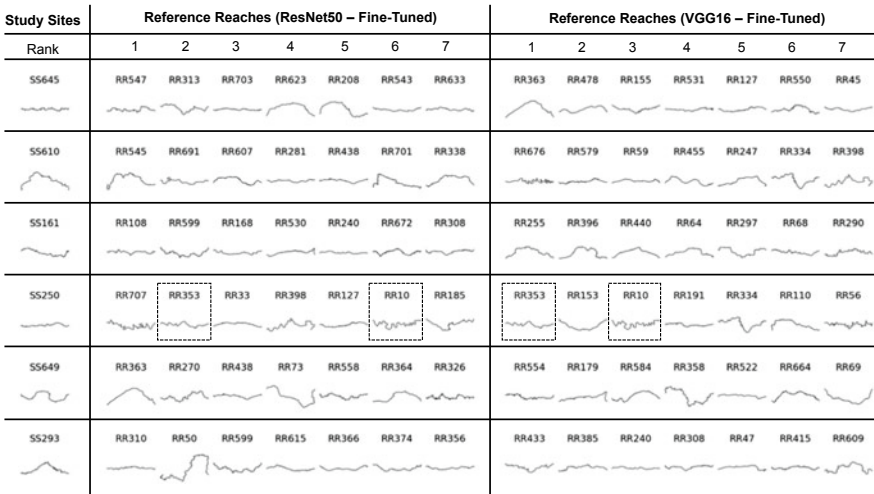


Fig. 5 Ranked reference reach results ($n = 7$) for six study sites identified using the AI tool with ResNet50 (left) and VGG16 (right) convolutional neural networks (CNNs) pre-trained on ImageNet with fine-tuning. Dashed boxes denote reference reaches identified by both CNNs

morphology characteristics similar to the study site that can be described using traditional terms from geomorphology. For example, for SS645 and SS649 the multiple wavelengths of meandering present in the study site morphology are also present in many of the reference reaches. However, for several sites tortuous meanders are present in some of the identified reference reaches but not the study site reaches. Unfortunately, due to the black-box nature of machine learning algorithms, it is difficult to know exactly why the algorithms selects specific reaches.

In contrast, the fine-tuned networks did not tend to yield similar predictions, and identified different reference reaches than the pre-trained networks. Subjectively, the fine-tuned ResNet50 model appeared to perform better than the fine-tuned VGG16 model, selecting reference reaches with more similar morphologies to the study site. We believe that fine-tuning the model heads reduced the generalizability of the models, and that identified features were not robust. We feel that the use of alternative techniques involving a small amount of manually labelled data such as semi-supervised learning may help to produce future models that are more robust. For the current time, we recommend implementing the AI tool using the models without fine-tuning to identify reference reaches.

The developed AI tool has potential uses to augment current industry practices for identifying reference reaches, which tend to use time-consuming, manual tasks. Even in its current state, the developed AI tool can be cautiously implemented as a pre-screening technique to rapidly identify potential reference reaches. We do not envision the current AI tool replacing the current expert-driven process, but rather helping to augment the workflow (Fig. 6), where pre-screened sites identified using the AI tool can be included as potential candidates for reference reaches along with other manually identified sites. As when using any channel classification system, a detailed review of reach morphology and characteristics to ensure suitability between the study site and potential reference reaches should be implemented prior to selecting one or more references for use in any type of analysis.

A potential use for the AI tool is in the rapid development of regional classification systems. By limiting the input stream segment shapefile to only include streams within a given watershed, and by using the processed stream segment image files, it is possible to use the AI tool to perform local clustering. Localized approaches to

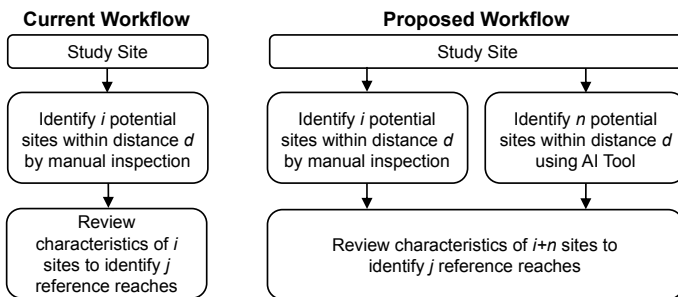


Fig. 6 Proposed workflow for reference reach identification using the developed AI tool

channel classification can mitigate many of the risks described by Kondolf [10] and Juracek and Fitzpatrick [7] that are associated with existing classification systems. Regional or reach-scale channel classification schemes have been successfully developed and used in several cases (e.g. Rinaldi [14], Brierley and Fryirs [2]). While the development of regional classification systems has historically been a time-consuming process that requires an expert understanding of processes governing regional channel morphology, the developed AI tool can be used to potentially reduce the time required for this process.

4 Next Steps

The AI tool presented in this paper is still under development. While masking for distance d must currently be done manually, future versions of the AI tool will take this distance as an input and perform this task automatically.

Although we feel that the current fine-tuning tasks were not successful in improving the performance of the model, we believe that it should be possible to improve the performance of future versions by implementing different training processes, including supervised, semi-supervised, or self-supervised learning. While the use of manually labelled data will increase human biases in the model, they may also help to produce findings that can be more easily explained using conventional technical terminology such as degree of sinuosity or tortuosity. For example, future data labels could include information such as bed slope, sinuosity, or even channel class identified using an existing classification system.

Kondolf [10] suggests that existing classification systems and natural channel designs may be biased towards single-thread channel geometries. In many existing classification systems (e.g. Rosgen), braided or multi-channeled rivers tend to be represented by a small number of classes relative to single-thread channels. While the current study only evaluated the use of the developed AI tool on rivers that can be represented using single channel polylines, there is no reason why the AI tool in its current form could not be implemented for braided or multi-channeled rivers.

The source code for the latest development version of the AI tool is available for download from GitHub at <https://github.com/codykupf/river-cnn>.

5 Conclusions

In the current study we present a case-study implementation of an AI tool for identifying reference reaches on the lower Albany River watershed in northern Ontario, Canada. We found that fine-tuning the CNNs was not effective in improving the model skill and suggest that features should be extracted by using the ResNet50 and VGG16 models pre-trained on ImageNet without fine-tuning for the time being. Although the AI tool presented in this study is still in the preliminary development

stage, even its current form it can potentially be used as a rapid assessment method to identify a number of potential reference reach sites for further manual review. We propose an augmented workflow that integrates the AI tool into existing workflows for identifying reference reaches.

Acknowledgements The authors of this study are funded by the Natural Sciences and Engineering Research Council of Canada through the Canada Graduate Scholarship and the Discovery Grant programs. The authors would like to thank the following individuals for their technical input on the machine-learning components of this project: Kristina Kupferschmidt, Graham Taylor, and Terrance Devries.

References

1. Brice JC (1975) Airphoto interpretation of the form and behavior of alluvial rivers. Washington University, St Louis, MO, USA
2. Brierley GJ, Fryirs K (2000) River styles, a geomorphic approach to catchment characterization: implications for river rehabilitation in Bega Catchment, New South Wales, Australia. *Environ Manag* 25(6):661–679
3. Cheplygina V, de Bruijne M, Pluim JPW (2019) Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Med Image Anal* 54:280–296. <https://doi.org/10.1016/j.media.2019.03.009>
4. Deng J, Dong W, Rocher R, Li L-J, Li K, Fei-Fei L (2009) ImageNet: a large-scale hierarchical image database. In: *IEEE conference on computer vision and pattern recognition*. IEEE, Miami, Florida, USA, pp 248–255. <http://doi.org/10.1109/CVPR.2009.5206848>
5. He K, Zhang X, Ren S, Sun J (2015) Deep residual learning for image recognition. [arXiv:1512.03385v1](https://arxiv.org/abs/1512.03385v1) [cs.CV], pp 1–12
6. Huh M, Agrawal P, Efros AA (2016) What makes imagenet so good for transfer learning? [arXiv:1608.08614v2](https://arxiv.org/abs/1608.08614v2) [cs.CV], pp 1–10
7. Juracek KE, Fitzpatrick FA (2003) Limitations and implications of stream classification 1. *JAWRA J Am Water Resour Assoc* 39(3):659–670
8. Khan S, Rahmani H, Shah SAA, Bennamoun M (2018) A guide to convolutional neural networks for computer vision. *Synth Lect Comput Vis* 8(1):1–207
9. Kornblith S, Shlens J, Le QV (2019) Do better imagenet models transfer better? In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp 2661–2671
10. Kondolf GM (2006) River restoration and meanders. *Ecol Soc* 11(2):42
11. Montgomery DR, Buffington JM (1997) Channel-reach morphology in mountain drainage basins. *Geol Soc Am Bull* 109(5):596–611
12. [NRC] Natural Resources Canada (2016) National hydrographic network. Government of Canada. Accessed Feb 4, 2021. <https://open.canada.ca/data/en/dataset/a4b190fe-e090-4e6d-881e-b87956c07977>
13. Pekel J-F, Cottam A, Gorelick N, Belward AS (2016) High-resolution mapping of global surface water and its long-term changes. *Nature* 540:418–422. <https://doi.org/10.1038/nature20584>
14. Rinaldi M (2003) Recent channel adjustments in alluvial rivers of Tuscany, Central Italy. *Earth Surf Process Land* 28:587–608
15. Rosgen DL (1994) A classification of natural rivers. *CATENA* 22(3):169–199. [https://doi.org/10.1016/0341-8162\(94\)90001-9](https://doi.org/10.1016/0341-8162(94)90001-9)
16. Rosgen DL (1996) Applied river morphology. Wildland hydrology. Pagosa Springs, Colorado
17. Rosgen DL (1998) The reference reach: a blueprint for natural channel design. In: *Engineering approaches to ecosystem restoration*, pp 1009–1016

18. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) [cs.CV], pp 1–14
19. [TRB] Transportation Research Board (2004) Methodology for predicting channel migration. National Academies of Sciences, Engineering, and Medicine, The National Academies Press, Washington, DC, USA. <http://doi.org/10.17226/23352>
20. [USGS] United States Geological Survey (2020) National hydrography dataset. Version 2.3

Stability Analysis of an Overtopped Spillway Using Computational Fluid Dynamics



M. Freitas, P. Léger, and L. Pedroso

1 Introduction

The overtopping of non-overflow gravity dam sections is a serious problem for the stability of the structure. Extreme floods, which have been made more frequent due to global warming, can cause overtopping of these structures, threatening their integrity (Fig. 1b, c). Overtopping is the cause of 43% of failures in masonry dams and 20% of failures in concrete dams [14]. Faulty gate operation can also contribute towards the occurrence of overtopping and the increase of the hydrodynamic loads [2, 9, 11, 19]. Likewise, the accumulation of floating debris is a problem during floods in several parts of the world. They produce additional thrusts on structures and clog spillway openings, thus reducing the discharge capacity while increasing the upstream water level [1, 8, 24].

When subject to these conditions, several failure modes in gravity dams can occur (Fig. 1a). The accurate determination of the flow velocity, pressure field and resulting hydrodynamic pressures are a major challenge to the evaluation of the structural stability. Traditionally, physical models and simplified analytical formulations have been used to estimate those quantities. More recently, CFD (Computational Fluid Dynamics) has been used as an alternative method that is more accurate than the analytical formulations as well as cheaper and faster than building and testing physical models. CFD has been used to model standard ogee spillways [12, 22], stepped spillways [25], as well as complex-shaped spillways [23]. CFD was also used to calculate the structural stability of spillways [7, 10].

M. Freitas (✉) · P. Léger
Polytechnique Montréal, Montreal, Canada
e-mail: mario-raul.freitas@polymtl.ca

L. Pedroso
Universidade de Brasília, Brasília, Brazil

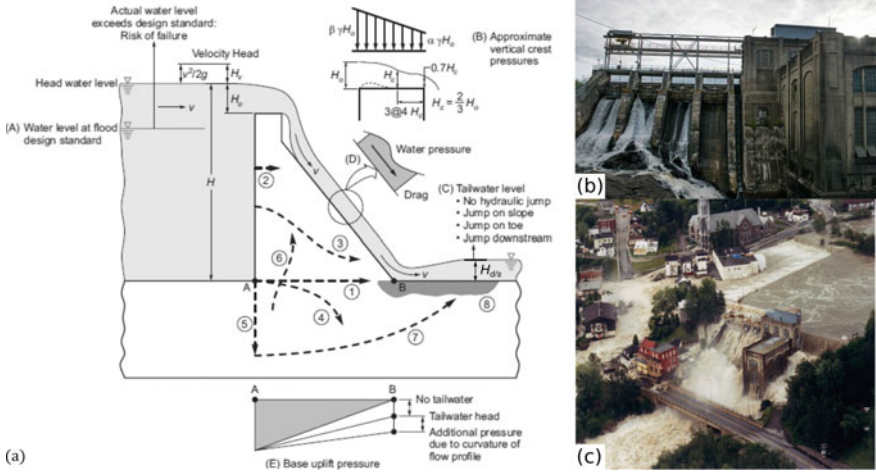


Fig. 1 Dam overtopping: **a** Failure mechanisms of an overtopped dam [17]; **b** Chicoutimi dam in normal operating conditions; **c** Chicoutimi dam during 1996 Saguenay flood

Herein, CFD is used to perform a back analysis of the Chute Garneau spillway, which was overtopped due to a flood in 1996. The hydrodynamic pressures are obtained with CFD using the software ANSYS Fluent. Then, the hydrodynamic forces are calculated by integrating the pressures and those are input into the software CADAM3D to perform the stability calculations using the gravity method. Then, the sensitivity of the structure to the tensile strength and cohesion in the rock-concrete interface is studied.

2 Stability of Overtopped Structures

2.1 Hydrostatic and Hydrodynamic Forces

The stability of dams and spillways is addressed in many guidelines [3, 6, 26]. Their stability is a function of the hydrostatic and hydrodynamic forces acting on them (Fig. 2). These guidelines suggest that the pressure on the upstream wall can be considered hydrostatic. The magnitude of the force on the crest F_v and its position x from the downstream crest face to upstream can be estimated as

$$F_v = 0.5(\alpha + \beta)\rho g B H_0 \tag{1}$$

$$x = \frac{\alpha + 2\beta}{3(\alpha + \beta)} B \tag{2}$$

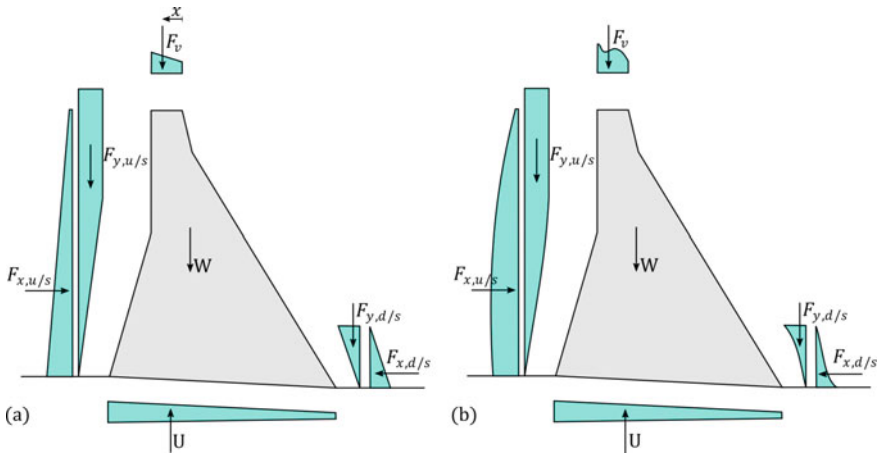


Fig. 2 Expected loads on a gravity dam: **a** guidelines estimation and **b** CFD pressure fields

where ρ is the water density, g is the gravity acceleration, B is the width of the crest, H_0 is the overtopping water height, α and β are constants, usually taken as 1 and 1/2, respectively [5].

The pressure on the downstream face can be estimated by assuming that the tailwater level $H_{d/s}$ is equal to H_0 (Fig. 1) if the tailwater condition significantly reduces or eliminates the hydraulic jump, or $0.6H_0$ otherwise. The amount of reduction in the effective tailwater depth may vary according to the degree of submergence of the crest and the backwater conditions in the downstream channel [26]. Thus, it is expected that the downstream force computed from CFD might be larger than that estimated by the guidelines. The forces on the upstream ($F_{x,u/s}$ and $F_{y,u/s}$) and downstream ($F_{x,d/s}$ and $F_{y,d/s}$) faces can be obtained by integrating the hydrostatic pressures on the wetted surfaces. To compute the uplift pressure, however, the tailwater level is always considered equal to H_0 .

Numerical studies with CFD showed that the hydrodynamic forces upstream are nearly hydrostatic. The hydrodynamic force on the crest can be better estimated with Eqs. 1 and 2 using $\alpha = \frac{2}{3}$ and $\beta = \frac{1}{3}$, rather than $\alpha = 1$ and $\beta = \frac{1}{2}$. The downstream hydrodynamic forces vary significantly in time due to the flow turbulence, but the guideline recommendations for them are conservative [7].

2.2 Stability Assessment

The stability of dams and spillways is generally assessed in terms of the position of the resultant, the sliding safety factor (SSF) and the stress state. For global stability, guidelines recommend that the resultant position shall be within the middle third of the structure base for usual load cases [3, 6, 26]. For unusual and extreme load cases,

it may be outside of the middle third and within the base if the other performance indicators satisfy acceptance criteria. This ensures that the structure will not fail by overturning.

The SSF acceptance criteria are dependent on the considerations of friction and cohesion. If only friction is considered, then the SSF must be equal to or higher than 1.5 for usual cases, 1.3 for unusual cases and 1.1 for flood cases [3]. However, if cohesion is also considered and no tests are available, then the acceptance criteria increase to 3.0 for usual cases, 2.0 for unusual cases and 1.3 for flood cases. The SSF can be calculated with the gravity method as

$$SSF = \frac{(\sum \bar{V} - U) \tan \phi + c A_c}{\sum H} \quad (3)$$

where $\sum \bar{V}$ is the sum of the vertical forces, excluding uplift pressures, U is the uplift force, ϕ is the friction angle, c is the cohesion, A_c is the area under compression, and $\sum H$ is the sum of horizontal forces.

Typically, cohesion and tensile strength of the concrete-foundation interface are considered zero. However, in situ testing performed in multiple dams in Ontario, Canada showed that the average tensile strength along the contact between rock and concrete in existing dams is 1.08 MPa with a minimum of 0.18 MPa [20]. Other studies presented data indicating that for the peak shear strength in concrete-granite-gneiss foundations, the best fit cohesion c is 1.30 MPa, the friction angle ϕ is 57° and the tensile strength σ_t is 0.83 MPa. The lower bound cohesion is 0.48 MPa, the friction angle is 57° , and the tensile strength is 0.31 MPa [4]. Herein, a peak friction angle $\phi = 55^\circ$ is assumed and a small amount of cohesion and tensile strength, within the experimental bounds observed in the literature, are used in some stability calculations.

3 CFD Modelling and Simulation

3.1 Governing Equations

Computational fluid dynamics (CFD) consists of a series of methods and techniques used to solve fluid flow problems numerically. Fluid flow problems are governed by Navier–Stokes and continuity equations. Assuming an incompressible and isothermal fluid, these equations can be written as [15]:

$$\nabla \cdot u = 0 \quad (4)$$

$$\frac{\partial u}{\partial t} + \nabla \cdot (uu) = -\nabla p + \nabla \cdot (\nu \nabla u) + g \quad (5)$$

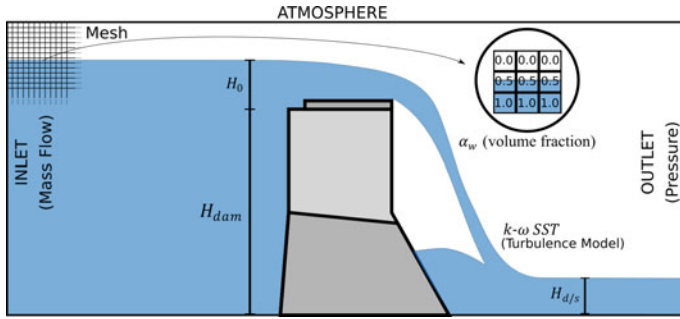


Fig. 3 Summary of CFD model boundary conditions, dimensions, mesh and flow properties

where u is the fluid velocity vector, p is the pressure, ν is the kinematic viscosity, g is the gravitational acceleration and t is the time. For multiphase flow, besides the continuity and Navier–Stokes equations, a multiphase model is needed. The volume of fluid model [13] is used here for its good performance with models where there is a clear separation between phases and no interpenetration. In this model, the volume fraction parameter α_w is used to track the motion of fluid (Fig. 3). This parameter corresponds to the percentage of water within each given cell. Cells with $\alpha_w = 0$ contain no water (full of air), cells with $\alpha_w = 1$ contain only water and cells with $0 < \alpha_w < 1$ contain a mix of air and water. The free surface is then computed with interpolation between cells. The fluid properties for each cell are taken as a weighted average of the properties of each phase using α_w as the weight. The flow turbulence, mainly present in the downstream section, is treated with the shear stress transport (SST) $k-\omega$ model [21]. This model combines the benefits of the $k-\epsilon$ model, which performs better in the free-flow region, and the $k-\omega$ model, which performs better in the boundary layer.

3.2 Boundary Conditions

In the CFD models developed herein, the upstream boundary is set as a mass-flow inlet, and the downstream and upper boundaries are set as pressure outlets. The boundaries corresponding to the floor and structures are set as no-slip walls (Fig. 3). The models studied are symmetrical, so only half of the geometry is modelled, and a symmetry boundary condition is set in the symmetry plane. The vertical boundaries parallel to the symmetry plane are set as no-shear walls. The downstream portion of the domain is broadened to allow for aeration of the overtopping nappe.

In addition, the open channel option present in the CFD software is enabled, which allows the specification of the inlet water level. The software then computes the related inlet pressure. With this option enabled, the pressure in the outlets can be

determined (i) by specifying the tailwater level, (ii) by interpolation from the neighbouring cells, or (iii) by specifying a gauge pressure. On the downstream boundary, the pressure is computed from the neighbouring cells; along the upper boundary, it is set to the atmospheric pressure.

4 Back Analysis of Chute Garneau Spillway

4.1 Spillway Description

In 1996, a major flood occurred in the Saguenay region (Québec, Canada). The rain-induced flow was much greater than the spillway capacity of the several gravity structures located in this area. During this flood, the gravity dams and spillways were subjected to intense hydrodynamic loading conditions, which resulted in overtopping of more than 2 m in some cases [18]. One of the structures affected by this flood was the Chute Garneau spillway, which is studied herein. This spillway is made of a series of piers that support a 6.30 m high concrete bridge. During the flood, a considerable amount of debris was accumulated between the gate lifting structures, blocking some of the flow (Fig. 4a). Despite the considerable hydrodynamic forces, the structure survived the flood event.

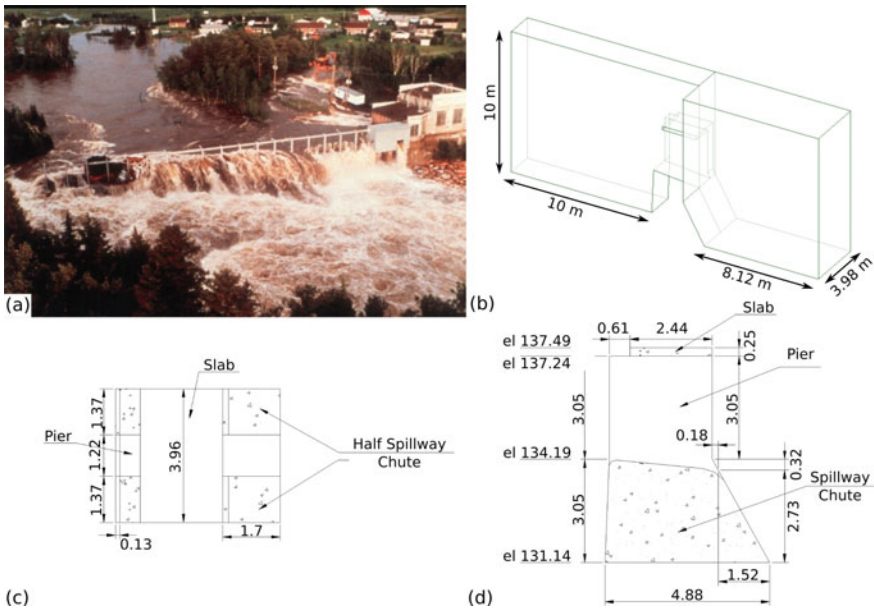


Fig. 4 Chute Garneau: a 1996 Saguenay flood, b 3D model geometry, c plan view, d elevation

4.2 Model Description

Considering the geometry of the spillway (Fig. 4c, d), a 3D symmetrical CFD model was developed to analyse the stability of this structure during the flood (Fig. 4b). The CFD model is limited to one section with a symmetry plane across one of the piers, with no normal flow along the lateral boundaries. Multiple CFD simulations with varying overtopping heights, with gates opened, closed, partially closed and with accumulation of floating debris were conducted. The hydrodynamic pressure fields were then obtained and, by integrating them, the hydrodynamic forces were calculated. Those forces were then input in the software CADAM3D [16] to assess the structure stability with the gravity method.

Figure 5a, b shows the profile, and the water volume fractions on a cross-section at the middle of the spillway chute and at the middle of the pier, respectively. Figure 5c, d show the velocity contour plots and Fig. 5e, f show the pressure contour plots in the same sections. All these results correspond to an overtopping height of 2.16 m.

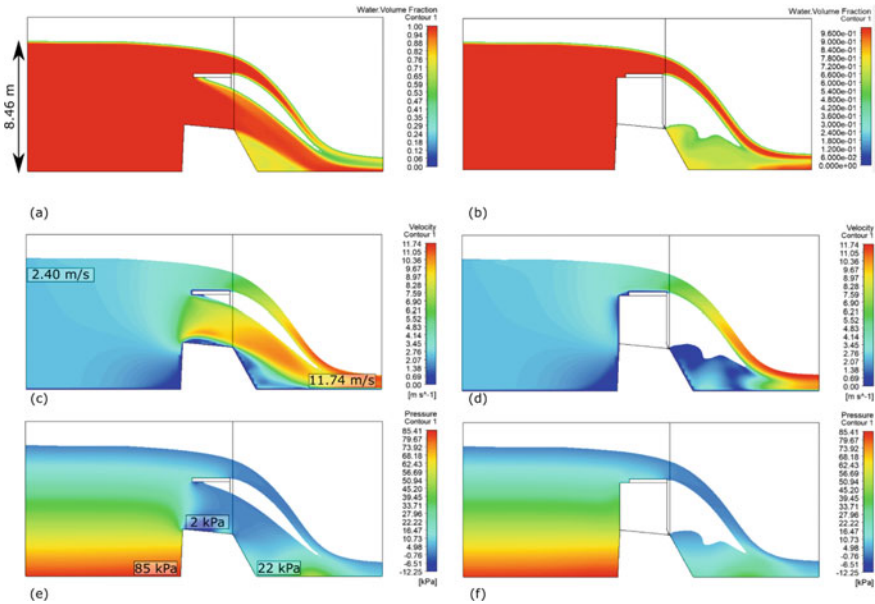


Fig. 5 Chute Garneau spillway CFD model for $H_0 = 2.16$ m: water volume fraction at **a** spillway section and **b** pier section, velocity contours at **c** spillway section and **d** pier section, pressure contours at **e** spillway section and **f** pier section [7]

4.3 Analysis of Spillway with Open Gates

The Chute Garneau spillway stability is evaluated under multiple scenarios: (i) with the gates completely opened, (ii) completely closed, (iii) partially closed, and (iv) with accumulation of floating debris. Table 1 summarizes the stability assessment with opened gates such that they do not interfere with the flow. The CFD stability indicators are obtained using the forces and uplift pressure obtained using CFD. The stability indicators for $\alpha = 1$ and $\beta = \frac{1}{2}$, and $\alpha = \frac{2}{3}$ and $\beta = \frac{1}{3}$ are obtained using the guideline recommendations for hydrodynamic forces, vertical force on the crest obtained from Eqs. 1 and 2 and uplift pressure taken from CFD. This is because the flow condition differs significantly from the overtopping of a gravity dam with classical geometry, increasing the uplift pressure significantly.

The values of $\alpha = 1$ and $\beta = \frac{1}{2}$ overestimate the sliding safety factors by as much as 34%. Using $\alpha = \frac{2}{3}$ and $\beta = \frac{1}{3}$ also overestimate the sliding safety factors in some cases, but only by 14% when $H_0 = 2.16$ m. This is consistent with the findings from the literature that indicate that $\alpha = \frac{2}{3}$ and $\beta = \frac{1}{3}$ better estimates the vertical force on the crest [7].

To demonstrate stability and avoid brittle failure by cracking of the base, a small amount of rock-concrete tensile strength (up to 70 kPa) was considered. Without this tensile strength, the structure fails by overturning. Moreover, a small amount of cohesion (up to 50 kPa) is needed to achieve the minimum safety factor of 1.3 in case of flood [3]. The spillway is expected to exhibit a very brittle failure mechanism, because it relies on the tensile strength to keep the structure stable. If a small amount of tensile strength cannot be mobilised, the structure would suddenly crack, the uplift pressure increases, and the structure fails. Both the 70 kPa of rock-concrete tensile strength and 50 kPa of cohesion assumed in the model are within the lower bound reported in the literature [4, 20], showing that these assumed values are reasonable and likely to have been mobilised during the flood.

4.4 Analysis of Spillway with Closed Gates

The closed gates scenario is reported in Table 2. The spillway opening is completely blocked by the gates, and water can only flow over the slab. This case is like the overtopping of a gravity dam. Thus, the uplift pressure is estimated using the full tailwater height. All forces are computed as in the opened gates scenario. For comparison, the sliding safety factors using a tailwater height of $0.6H_0$, as recommended by guidelines for the evaluation of the downstream force F_d^s , is also computed.

Closed gates are the worst-case scenario for stability (ignoring the presence of floating debris) because the destabilizing horizontal upstream force is significantly increased, while the stabilizing downstream force decreases, and the vertical force on the crest stays about the same. The use of $\alpha = 1$ and $\beta = \frac{1}{2}$ lead to a small overestimation of the SSF (approximately 7% when $H_0 = 2.18$ m), while $\alpha = \frac{2}{3}$

Table 1 Chute Garneau stability with opened gates

| H_0 (m) | σ_t (kPa) | c (kPa) | Sliding safety factor | | | Resultant position in % ^a | | |
|----------------|------------------|-----------|--------------------------------|---|-------------------------------------|--------------------------------------|---|-------------------------------------|
| | | | F_v^{CFD} — Reference | $F_v^{\alpha=\frac{2}{3}, \beta=\frac{1}{3}}$ | $F_v^{\alpha=1, \beta=\frac{1}{2}}$ | F_v^{CFD} — Reference | $F_v^{\alpha=\frac{2}{3}, \beta=\frac{1}{3}}$ | $F_v^{\alpha=1, \beta=\frac{1}{2}}$ |
| 0 ^b | 0 | 0 | 2.78 (1.00) | 2.74 (0.99) | 2.91 (1.05) | 57 (1.00) | 57 (1.00) | 56 (0.98) |
| | | | 3.35 (1.00) | 3.30 (0.99) | 3.51 (1.05) | | | |
| 1.10 | 33.7 | 0 | 1.34 (1.00) | 1.45 (1.08) | 1.64 (1.23) | 83 (1.00) | 78 (0.94) | 72 (0.88) |
| | | | 1.64 (1.00) | 1.77 (1.08) | 1.98 (1.21) | | | |
| 1.45 | 48.1 | 0 | 1.15 (1.00) | 1.31 (1.14) | 1.51 (1.32) | 82 (1.00) | 83 (1.01) | 76 (0.83) |
| | | | 1.44 (1.00) | 1.62 (1.13) | 1.84 (1.28) | | | |
| 1.80 | 56.9 | 0 | 1.13 (1.00) | 1.19 (1.05) | 1.38 (1.22) | 95 (1.00) | 87 (0.92) | 80 (0.83) |
| | | | 1.43 (1.00) | 1.48 (1.03) | 1.69 (1.18) | | | |
| 2.16 | 68.8 | 0 | 0.96 (1.00) | 1.09 (1.14) | 1.29 (1.34) | 105 (1.00) | 93 (0.89) | 83 (0.79) |
| | | | 1.24 (1.00) | 1.37 (1.10) | 1.58 (1.28) | | | |

Note The values in parenthesis represent the ratio between stability indicators using Eqs. 1 and 2 versus CFD, taken as reference. Reference values are highlighted in bold font

^a Expressed in percentage of base width from upstream

^b Normal operating level with no overtopping and total water height equals to 5.38 m

Table 2 Chute Garneau stability with closed gates

| | | Sliding safety factor | | | |
|----------------|------------------|-----------------------|--|--|------------------------------------|
| | | F_v^{CFD} | $F_v^{1.0H_0}$ | $F_v^{0.6H_0}$ | |
| H_0 (m) | σ_r (kPa) | c (kPa) | $F_v^{\alpha=\frac{2}{3},\beta=\frac{1}{3}}$ | $F_v^{\alpha=\frac{2}{3},\beta=\frac{1}{3}}$ | $F_v^{\alpha=1,\beta=\frac{1}{2}}$ |
| 0 ^a | 0 | 0 | 2.28 | | $F_v^{\alpha=1,\beta=\frac{1}{2}}$ |
| 1.06 | 94.6 | 0 | 1.02 (1.00) | 0.99 (0.97) | 1.03 (1.00) |
| | | 50 | 1.33 (1.00) | 1.30 (0.98) | 1.34 (1.01) |
| | | 110 | 1.70 (1.00) | 1.67 (0.98) | 1.71 (1.01) |
| 1.45 | 113.4 | 0 | 0.90 (1.00) | 0.88 (0.98) | 0.92 (1.02) |
| | | 50 | 1.19 (1.00) | 1.17 (0.98) | 1.22 (1.02) |
| | | 110 | 1.55 (1.00) | 1.53 (0.99) | 1.59 (1.03) |
| 1.78 | 128.2 | 0 | 0.81 (1.00) | 0.79 (0.98) | 0.85 (1.04) |
| | | 50 | 1.10 (1.00) | 1.09 (0.99) | 1.14 (1.04) |
| | | 110 | 1.44 (1.00) | 1.44 (1.00) | 1.50 (1.04) |
| 2.18 | 147.0 | 0 | 0.72 (1.00) | 0.71 (0.99) | 0.77 (1.07) |
| | | 50 | 0.99 (1.00) | 1.00 (1.01) | 1.06 (1.07) |
| | | 110 | 1.32 (1.00) | 1.34 (1.02) | 1.41 (1.07) |

Note The values in parenthesis represent the ratio between stability indicators using Eqs. 1 and 2 versus CFD, taken as reference. Reference values are highlighted in bold font

^a Normal operating level with no overtopping and total water height equals to 5.38 m

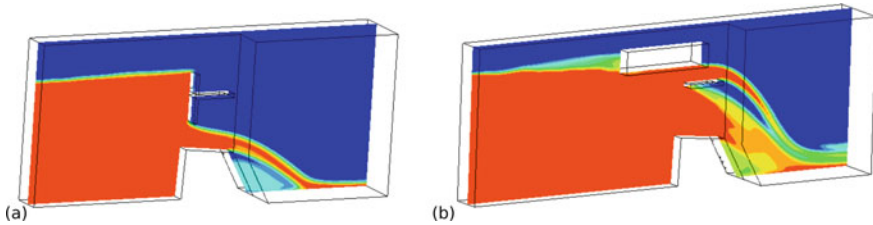


Fig. 6 Chute Garneau CFD model: **a** partially closed gates ($H = 7.42$ m), **b** floating debris ($H_0 = 2$ m)

and $\beta = \frac{1}{3}$ yields an accurate estimation of the SSF compared to CFD simulations. With gates closed, the amount of tensile strength to avoid cracking and cohesion to reach the minimum safety factor are increased from 70 to 150 kPa compared with the gates opened scenario. Likewise, the minimum cohesion necessary to ensure stability increased from 50 to 110 kPa compared with the gates opened scenario. Nevertheless, the assumed tensile strength and cohesion values are within the lower bound limits presented in the literature [4, 20]. Moreover, the resultant force is always positioned outside of the section, and the only mechanism that prevents the spillway from overturning is the tensile strength allowed.

4.5 Analysis of Spillway with Partially Closed Gates

Three cases with a partially closed gate were also modelled. In these cases, the gate is lifted halfway up the spillway opening, partially blocking the flow over the spillway chute and preventing overtopping on the bridge (Fig. 6a). For a flow of $12 \text{ m}^3/\text{s}$, the total upstream headwater was 7.42 m, using $c = 50$ kPa; the SSF was 1.51, and the maximum tensile stress was 81.6 kPa. These results can be compared with the opened gates case with $H_0 = 1.10$ m (where the total upstream headwater is 7.40 m) and the closed gates case with $H_0 = 1.06$ m (where the total upstream headwater is 7.36 m). In the equivalent opened gate case, the SSF obtained was 1.64 and the maximum tensile stress required was 33.7 kPa. In the equivalent closed gates case, the SSF obtained was 1.33, and the maximum tensile stress was 94.6 kPa. This indicates that the partially closed gates case is more critical in terms of the overturning moments developed on the gate than in terms of the sliding forces (Table 3).

4.6 Analysis of Spillway with Floating Debris

An opened gates case with floating debris was also modelled. The debris is considered to get stuck in the gate's lifting structure and accumulates for a length of 6 m, thus spanning the entire spillway width. The presence of floating debris is considered,

Table 3 Chute Garneau stability with partially closed gates

| H (m) | σ_t (kPa) | c (kPa) | SSF— F_v^{CFD} | Resultant position in % ^a — F_v^{CFD} |
|---------|------------------|-----------|------------------|--|
| 5.58 | 0 | 0 | 2.52 | 60.63 |
| | | 50 | 3.04 | |
| 6.4 | 24.4 | 0 | 1.72 | 76.47 |
| | | 50 | 2.10 | |
| 7.42 | 81.6 | 0 | 1.17 | 101.84 |
| | | 50 | 1.51 | |

^a Expressed in percentage of base width from upstream

in a simplified way, by modelling an obstacle to the flow at the water surface level (Fig. 6b). Only the additional inertial and drag force are considered. The force due to the impact of the floating debris against the structure is not studied here. Beyond adding the inertial and drag forces to the structure, the presence of debris also changes the flow conditions, causing a modification in pressure fields.

A single case with floating debris is modelled, where the overtopping height is 2 m and there was 1 m of submerged debris. The total thrust caused by the debris was computed with CFD as 16.8 kN and applied at an elevation of 7.75 m. The SSF obtained with $c = 50$ kPa was 1.23, and the maximum tensile stress was 72.3 kPa. An equivalent open gates case with no debris and 2 m of overtopping height was modelled for comparison. The SSF obtained in this case was 1.30 and the maximum tensile stress was 63.2 kPa. Similar to the partially closed gates case, the presence of floating debris significantly increases the overturning moment, even with a small increase in the sliding force.

5 Conclusions

This paper presented a back analysis of the Chute Garneau spillway under the 1996 Saguenay flood using CFD to estimate the hydrodynamic pressure acting on the structure. The analysis indicated that for both opened and closed gates scenarios, the minimum SSF acceptance criteria were met when 150 kPa of tensile strength and 110 kPa of cohesion were considered. The tensile strength and cohesion adopted in the stability analysis are well within the range found in the literature [4, 20], which present a lower bound of 310 kPa for tensile strength and 480 kPa for cohesion. This shows that although the guidelines do not recommend reliance on rock-concrete tensile strength to ensure stability, it may contribute significantly to stability during an extreme flood. This may make the difference between the structure failing or surviving the event. Both the partially closed gates and the floating debris scenarios mainly affected the structure stability by increasing the overturning moment with a relatively small increase in horizontal forces with a longer lever arm. Thus, in both cases, the increase in the tensile stress is more significant than the increase in the

SSF. The floating debris case should be further investigated by modelling the debris either as a single solid floating particle (more simplified) or as a collection of particles (more detailed) rather than an obstacle and by considering the impact force on the structure.

Acknowledgements This study was financed in part by the Coordination for the Improvement of Higher Education Personnel-Brazil (CAPES)—Finance Code 001. The authors also acknowledge that this study was made possible with the support of Global Affairs Canada and the NSERC (National Science and Engineering Research Council of Canada).

References

1. Abela C (2018) Recommendations for determining debris loads for tainter gates. *ASCE Pract Periodical Struct Des Constr* 23(2)
2. Allen P (2009) Operation of spillway gates—how to avoid the problems and pitfalls. In: 34th annual QLD water industry operations workshop, Caloundra, Queensland, Australia, 16–18 June
3. CDA (2013) Dam safety guidelines. Canadian Dam Association, Edmonton, Alberta, Canada
4. EPRI (1992) Uplift pressures, shear strengths, and tensile strengths for stability analysis of concrete gravity dams. Electric Power Research Institute, Palo Alto, California
5. FERC (1991) Engineering guidelines for evaluation of hydropower projects—chapter III gravity dams. Federal Energy Regulatory Commission, Washington, D.C.
6. FERC (2016) Engineering guidelines for the evaluation of hydropower projects—draft chapter III: gravity dams. Federal Energy Regulatory Commission, Washington D.C., USA
7. Freitas M, Favre E, Léger P, Pedroso LJ (2021) Stability evaluation of overtopped concrete hydraulic structures using computational fluid dynamics. *Can J Civ Eng* 48(3):341–346
8. Godtland K, Tesaker E (1994) Clogging of spillways by trash. In: 18th ICOLD congress, Durban, South Africa
9. Graham WJ, Hilldale RC (2001) Spillway gate failure or misoperation: representative case histories. U.S. Department of the Interior, Bureau of Reclamation
10. Griffith RA, Rutherford JH, Alavi A, Moore D, Groeneveld J (2007) Stability review of the Wanapum spillway using CFD analysis. *Can Dam Assoc Bull* 2016:16–26
11. Hartford DND, Baecher GB, Zielinski PA, Patev RC, Romanas A, Rytters K (2016) Operational safety of dams and reservoirs—understanding the reliability of flow-control systems. ICE Publishing, London, UK
12. Haun S, Olsen NRB, Feurich R (2011) Numerical modeling of flow over trapezoidal broad-crested weir. *Eng Appl Comput Fluid Mech* 5(3):397–405
13. Hirt CW, Nichols BD (1981) Volume of fluid (VOF) method for the dynamics of free boundaries. *J Comput Phys* 39(1):201–225
14. ICOLD (1995) Dam failures statistical analysis. Bulletin 99. International Commission on Large Dams, Paris, France
15. Jasak H (1996) Error analysis and estimation for the finite volume method with applications to fluid flows. Ph.D. thesis, Department of Mechanical Engineering, Imperial College
16. Leclerc M, Léger P (2003) Computer aided stability analysis of gravity dams—CADAM. *Int J Adv Eng Soft* 34(7):403–420
17. Léger P (2019) Dam compendium—hydrological safety and dam, spillways stability during flood, chap 6. Polytechnique Montreal
18. Léger P, Tinawi R, Larivière R (1998) The behaviour of gravity dams and spillways in extreme floods: Canadian experience. *J Hydropower Dams* 5(3):73–77

19. Lewin J (2001) Hydraulic gates and valves: in free surface flow and submerged outlets. Thomas Telford
20. Lo K, Grass J (1994) Recent experience with safety assessment of concrete dams on rock foundations. In: Dam safety, Canadian Dam Association annual conference, Winnipeg, Manitoba
21. Menter RF (1993) Zonal two equation $k-\omega$ turbulence models for aerodynamic flows. In: 24th fluid dynamics conference, Orlando, Florida, USA, 6–9 July
22. Olsen NBR, Kjellesvig HM (1998) Three-dimensional numerical flow modeling for estimation of spillway capacity. *J Hydraul Res* 36(5):775–784
23. Patarroyo J, Damov D, Shepherd D, Hiscock J (2015) Hydraulic design of the stepped overflow spillway of the muskrat falls hydroelectric development. *Can Dam Assoc Bull* 27(1):13–25
24. SCD (2017) Floating debris at reservoir dam spillways. Swiss Committee on Dams
25. Sweeney BP (2014) Converged stepped spillway models in OpenFOAM. M.Sc. thesis, Department of Computing and Information Sciences, Kansas State University
26. USACE (1995) Engineering and design: gravity dam design. US Army Corps of Engineers, Washington, D.C.

The Effects of Climatic and Design Variables on Evapotranspiration in Bioretention Systems



R. Nasrollahpour, A. Skorobogatov, J. He, C. Valeo, A. Chu, and B. van Duin

1 Introduction

Urbanization has led to the growth of impervious surfaces and in turn, resulted in increased stormwater runoff volume and peak flow. In addition, stormwater runoff, which mobilizes and transports a variety of pollutants deposited on the land surface, is known to degrade the quality of water receiving bodies. To attenuate the adverse impacts of urbanization, low impact development (LID) technologies are introduced to supplement traditional centralized drainage or conveyance systems [5]. The LID practices are green strategies to mimic the natural hydrology of a site using decentralized microscale control practices by enhancing evapotranspiration (ET), infiltration, and/or groundwater recharge. Bioretention systems are a typical LID practice with confirmed benefits in improving stormwater quality, apart from retaining stormwater runoff and decreasing peak flow rate [4]. In the urban hydrologic cycle, ET is one of the critical components [3]; however, in the context of bioretention systems, the understanding of ET is still limited and it is often unaccounted for in current practices [2]. Therefore, the primary objective of this paper is to advance the understanding of ET in bioretention systems, particularly to investigate the effects of design factors (vegetation and media) and the climatic variables on ET. Linking ET to the design

R. Nasrollahpour (✉) · A. Skorobogatov · J. He · A. Chu
Civil Engineering, Schulich School of Engineering, University of Calgary, 2500 University Drive
NW, Calgary, AB T2N 1N4, Canada
e-mail: reza.nasrollahpour@ucalgary.ca

C. Valeo
Mechanical Engineering, University of Victoria, 3800 Finnerty Road, Victoria, BC V8P 5C2,
Canada

B. van Duin
The City of Calgary, 625 - 25 Ave S.E., Calgary, AB T2G 4K8, Canada

variables presents an opportunity to improve the design of bioretention systems which could consequently have important practical implications.

2 Materials and Methods

A bioretention research facility was constructed in the Town of Okotoks, which is 45.3 km south of the City of Calgary, Alberta. The facility consists of 24 mesocosms, each of which is approximately $2 \times 2 \text{ m}^2$ in surface area. Each cell is framed by wooden walls and is 1.2 m in height, including 0.3 m ponding depth on the top, followed by a 0.6 m growing media layer and a 0.3 m gravel drainage layer. The mesocosms were individually lined at the bottom and sloped to a single corner to drain via a standpipe. The mesocosms were constructed with three different media types, each having a set of 8 cells (Fig. 1). Media 70 and 40, characterized by a high percentage of sand and low clay content, are designated media types for the City of Calgary [1]; media CL (clay-loam mixture) is commonly seen in the study region. The average particle sizes (D_{50}) of the media 40, 70, and CL are 0.58, 0.42, and 0.13 mm, respectively. In each set of 8 cells, cells 1, 4, 6, and 7 were planted with herbaceous mixes; cells 2 and 3 were planted with woody mixes; cells 5 and 8 were vegetated with turfgrass.

To examine the performance of the bioretention system, the mesocosms were monitored under 49 simulated events in the growing seasons (May–October) of 2018 and 2019. This included performing simulated events corresponding to runoff from 4 mm, 9 mm, 14 mm, and 24 mm storm events. A weather station was installed

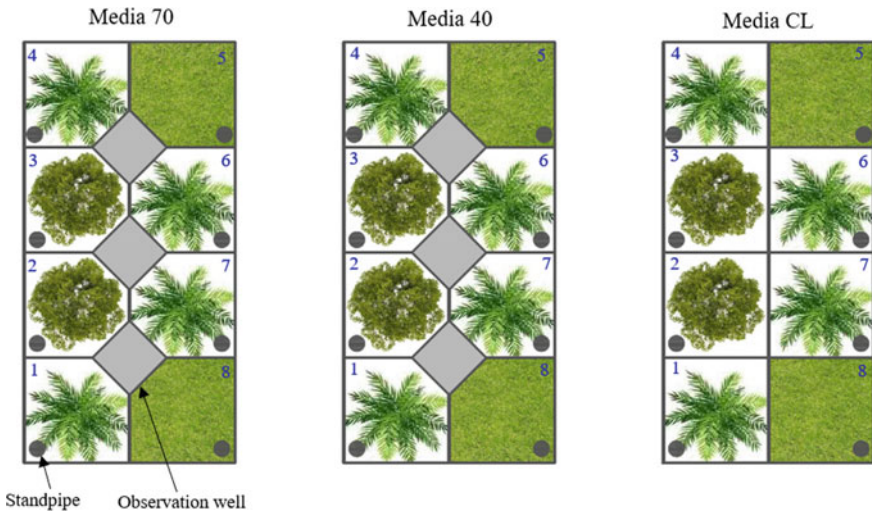


Fig. 1 The layout of the bioretention mesocosms

onsite to record climatic variables, and TERSO 12 sensors were deployed into the cells at two depths (20 and 40 cm) to measure soil moisture. In the field seasons, the average daily air temperature was 12.7 °C. As the focus of this paper is on ET, the data collected in between the simulated events (dry periods) were used herein. The variations in media moisture and field observations indicated that gravity drainage lasted no longer than 1.5 days following each simulated event. The dry periods were thus defined as the time from the end of gravity drainage to the beginning of the next simulated event. However, the days when the total rainfall amount exceeded 3 mm were excluded from the dry periods as this amount of rainfall was observed to produce a measurable change in the media moisture. In a dry day, ET was estimated using the daily changes in media moisture at the surface (20 cm) and deep (40 cm) layers. At each layer, ET (in mm/day) was quantified by multiplying the daily media moisture loss by the media depth (300 mm). The daily climatic variables were their average values over the day.

3 Results and Discussion

3.1 Effect of Climatic Variables

The role of several climatic parameters (i.e., solar radiation, wind speed, relative humidity (RH), and air temperature) was investigated through correlation analysis. As shown in Table 1, ET (at both depths) was positively correlated to air temperature, wind speed, and solar radiation, while it negatively depended on RH. Among these climatic variables, solar radiation, air temperature, and RH were found to be strongly associated with ET, especially at the surface layer, but the role of wind speed was not as prominent as others. The energy supplied by solar radiation is the primary driving force for the vaporization of water. However, ET can be affected by the combination of solar radiation and other climatic parameters. In humid conditions, for example, ET can be inhibited by a lack of notable vapor pressure differentials between plant

Table 1 Spearman’s correlation coefficients between climatic variables and ET at the surface and deep layers

| Parameter | Year | Depth 20 | Depth 40 |
|-----------------|------|---------------|---------------|
| Wind speed | 2018 | 0.258 | 0.164 |
| Air temperature | 2018 | 0.660 | 0.324 |
| RH | 2018 | -0.605 | -0.286 |
| Solar radiation | 2018 | 0.612 | 0.289 |
| Wind speed | 2019 | 0.266 | 0.208 |
| Air temperature | 2019 | 0.492 | 0.273 |
| RH | 2019 | -0.432 | -0.317 |
| Solar radiation | 2019 | 0.527 | 0.275 |

Significant correlation coefficients at the level of 5% are boldfaced

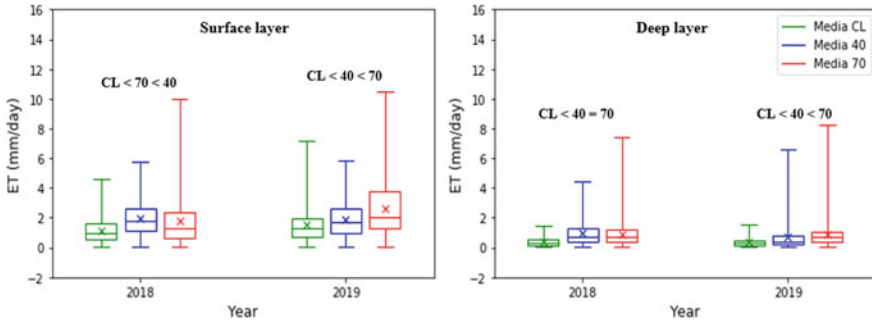


Fig. 2 Box-whisker plots of ET at the surface and deep layers for different growing media. The box covers the range of data between the 25th and 75th percentiles; whiskers extend to the minimum and maximum values; median and mean are indicated by the line and diagonal cross inside the box, respectively

stomata/moist substrate and ambient air. The obtained negative correlations between ET and RH reflect this concept in the context of bioretention systems.

3.2 Effect of Growing Media

The data collected from all the cells were grouped for each media, irrespective of the vegetation types, and compared using the Kruskal–Wallis tests followed by multiple pairwise comparisons at both depths. As seen in Fig. 2, the median of ET in media CL was significantly lower than that of the other two media types at both depths throughout the study period. To improve the infiltration of CL cells, wood chips (at 40% in volume) were mixed with CL, which can create large drainable pores and channels. This is supported by the fact that the CL cells had the greatest infiltration rate (in an average of 1251.3 mm/h) and the lowest amount of water retained (in an average of 8.5%). In 2019, when vegetation was more mature, the media 70 cells had a higher ET compared to the media 40 cells.

3.3 Effect of Vegetation

All the recorded data were grouped in terms of the vegetation, irrespective of the growing media, and compared using the Kruskal–Wallis tests followed by multiple pairwise comparisons. As shown in Fig. 3, the median of ET in woody cells was consistently higher than the control (grass) cells at both depths, whereas the median of ET in herbaceous cells was not always statistically higher than that of the control cells. In 2018, the absence of the significant differences between the woody and herbaceous cells at the surface layer can be ascribed to the negligible differences

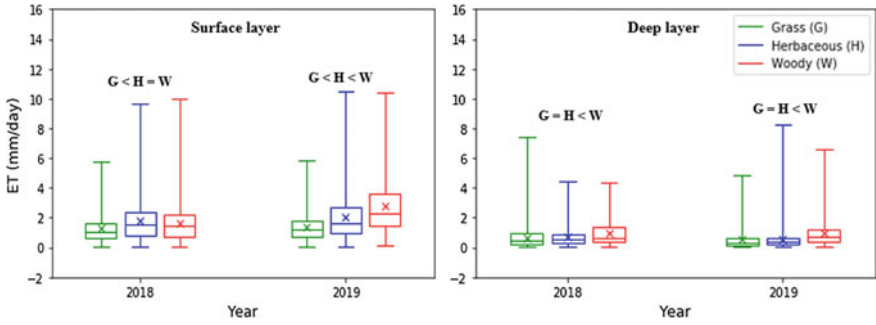


Fig. 3 Box-whisker plots of ET at the surface and deep layers for different vegetation types. The box covers the range of data between the 25th and 75th percentiles; whiskers extend to the minimum and maximum values; median and mean are indicated by the line and diagonal cross inside the box, respectively

in their roles at the early stage of vegetation development. At the deep layer, the median of ET in woody cells was consistently higher over the study period, while the significant difference in the median of ET between herbaceous and control cells was not presented. These results suggest that, compared to herbaceous vegetation, woody vegetation was more beneficial in improving water loss through ET, as this vegetation type increased ET at both layers. In addition, the increased ET in the vegetated cells throughout the study period demonstrates the temporal evolution of ET and its potential role on the performance of bioretention systems.

4 Conclusions

The findings from this two-year study confirmed the roles of media and vegetation, apart from the climatic variables, on ET in the bioretention systems. The results show that the media 70 cells vegetated with woody mixes outperformed (in terms of ET) all other cells. This calls for the consideration of these design variables, whose roles on ET are dynamic, when designing the bioretention systems for optimal performance.

Acknowledgements This research was funded by NSERC CRD, the Town of Okotoks, the City of Calgary, the Bow River Basin Council, and Source2Source. The authors also would like to acknowledge the funding from the federal and provincial governments and many others for constructing the bioretention research facility, and the involvement of many organizations including the Alberta Low Impact Development Partnership.

References

1. City of Calgary (2016) Low impact development guidelines: module 2—bioretention and bioswales. Final report
2. Ebrahimian A, Wadzuk B, Traver R (2019) Evapotranspiration in green stormwater infrastructure systems. *Sci Total Environ* 688:797–810
3. Hess AJ (2017) Rain garden evapotranspiration accounting. Villanova University
4. Kratky H, Li Z, Chen Y, Wang C, Li X, Yu T (2017) A critical literature review of bioretention research for stormwater management in cold climate and future research recommendations. *Front Environ Sci Eng* 11(4):1–15
5. Liu J, Sample D, Bell C, Guan Y (2014) Review and research needs of bioretention used for the treatment of urban stormwater. *Water* 6(4):1069–1099

Wetted Area in Comparison to Habitat Suitability Criteria Based Methodologies for Sustainable Water Resources Management



A. Kneale and H. Ghamry

1 Introduction

Sustainably managing the competing water needs of human societies and natural freshwater ecosystems continues to challenge a variety of water resources professionals [4, 8]. Over the years, varying levels of approaches were developed to address these environmental flow issues [3, 8]. One approach, known as the ecohydraulic assessment or hydraulic microhabitat assessment, attempts to describe suitable habitat at a spatial scale assumed to be consistent with the behaviour of fish through the use of hydrodynamic-habitat modelling software [8]. Habitat suitability criteria (HSC) are the core biological component of this approach [1]. HSC is based on the known selection of particular habitat conditions during specific periods of the species' life history and is developed from a variety of different abiotic habitat variables such as water depth, water velocity, bottom substrate, cover, water temperature, dissolved oxygen, and turbidity. The most commonly used hydraulic variables at a microhabitat scale are water velocity, depth, and substrate/cover [1, 8, 9]. For a particular habitat variable, the HSC represents the degree of preference displayed by the fish over a range of flows. Values of HSC generally range from zero (poor habitat) to one (excellent habitat). HSC curves are developed based on (I) judgement, (II) utilization (field studies), or (III) preference (utilization corrected by availability) [1]. Category I HSC curves developed via literature sources, professional judgement, and the Delphi technique are the most widely used.

To determine the quantity of available habitat for a species of fish and life stage based on its habitat preferences, HSC curves are applied to hydraulic information

A. Kneale · H. Ghamry (✉)
Department of Fisheries and Oceans, Winnipeg, Canada
e-mail: Haitham.Ghamry@dfo-mpo.gc.ca

A. Kneale
e-mail: andrea.kneale@dfo-mpo.gc.ca

associated with each point along one or more transects for a given discharge in the hydrodynamic-habitat modelling software. The resulting value is defined as the weighted usable area (WUA). Typically, WUA increases with flow until a peak flow is achieved, at which point the WUA decreases as flows become too high to sustain optimal habitat conditions [3, 11]. The habitat-flow relationship of the WUA curve gives users the opportunity to link changes in fish habitat to broader population-level processes, such as recruitment, growth, and migration [4, 7].

While HSC-based methodologies of aquatic assessment such as PHABSIM (Physical HABitat SIMulation system developed by the U.S. Fish and Wildlife Service) are widely used, they are also widely criticized for its focus of modelling velocity, depth, and substrate when different species of fish may cue on different hydraulic variables or habitat components [8]. HSC-based methodologies of aquatic assessment are further complicated by the lack of validation studies completed to illustrate the accurate representation of HSC and WUA curves in nature [10].

To find an alternative to HSC-based methodologies that may allow for more validation studies and avoid the use of data intensive models, we studied the potential for a wetted area (WA)-based methodology as a more simplified approach to inform management at larger temporal or spatial scales. Using HSC curves for five adult freshwater fish species inhabiting the Embarras reach of the Athabasca River, our objectives were to (a) build a hydrodynamic-habitat model to simulate changes in habitat availability across a range of flows and (b) examine the relationship between flow and HSC-based habitat availability relative to WA.

2 Study Site Description

Modelling was completed for the Embarras reach (58° 12' 15.9" N 111° 23' 41.4" W upstream to 58° 14' 05.4" N 111° 25' 33.4" W downstream) of the Athabasca River, which is situated in the northeast corner of Alberta, Canada (see Fig. 1 in [5]). The river reach at Embarras is located between 165 and 171 km downstream of the city of Fort McMurray, Alberta. The average width and length of the Embarras reach of the river is 480 m and 6.3 km, respectively. The bed material is mostly sand. Flows range from 75 to 2000 m³/s and the water depth ranges from 0.71 to 7.15 m. For more information regarding the study site and methods used to collect the point hydraulic variables see Ghamry and Katapodis [5].

3 Habitat Modelling

River2D (www.river2d.ca) was used to model the physical habitat (depth, velocity, and substrate) for a range of relevant flows at the Embarras reach of the Athabasca River. River2D is a finite-element model that, like other depth averaged models, solves the conservations of mass (continuity) and momentum equations. It is based

on the resolution of the De Saint–Venant equations using the finite-element method. The model can resolve flow details, which are generally larger than the depth of flow. It is capable of simulating the flow of natural streams and rivers by accommodating the supercritical and subcritical flow transitions. It also has established capabilities for wet and dry area solutions [12]. See Ghanem et al. [6] for a more detailed description of the model along with the solution of the 2D depth-averaged equations, underlying assumptions, and finite-element formulation. Reference Ghamry and Katapodis [5] for a detailed description of the calibration of the River2D model for the Embarras study site.

4 Habitat Availability

River2D was used to assess the variations in habitat availability at the study site for five adult freshwater fish species with distinct habitat preferences. Two area values were calculated: WUA and WA. In River2D, the WUA fish habitat component is based on the concept used in the PHABSIM family of fish habitat models [1, 2]. In this study, WUA values were calculated by multiplying the depth, velocity, and substrate HSC curves for the adult life stages of Goldeye, Longnose Sucker, Northern Pike, Walleye, and White Sucker against the simulated flow conditions [12]. WA values were calculated by multiplying a pseudo depth, velocity, and substrate HSC curve, where all habitat values were set to one (excellent habitat) above a minimum depth of 20 cm, against the simulated flow conditions. Hydraulic modelling was limited to the wetted channel, which corresponded to a maximum discharge of 2000 m³/s. The depth, velocity, and substrate HSC curves for adult Goldeye, Longnose Sucker, Northern Pike, Walleye, and White Sucker were provided by Trillium Engineering and Hydrographics Inc. [13] for the Embarras reach of the Athabasca River.

4.1 Analyses of WA Relative to WUA

To examine the relationship between flow and HSC-based habitat availability relative to WA we calculated the percentage of WUA in relation to WA and absolute percentage difference in optimal flow. Optimal flow is the point on the area-flow curve where habitat area is optimized (i.e., the peak area before the area-flow curve plateaus or recedes). Percent difference or change in optimal flow based on WUA and WA was calculated with reference to the optimal flow from the WA curve. The WA and WUA curves were plotted on the same figure for visual assessment.

Figure 1 presents the WA and WUA curves for adult Goldeye, Longnose Sucker, Northern Pike, Walleye, and White Sucker at the Embarras reach of the Athabasca River. The WA curve plateaued at 1000 m³/s, at which point area increased gradually with flow. The WUA curve for Goldeye, Longnose Sucker, Walleye, and White Sucker followed a similar area-flow relationship as the WA curve up to a flow of

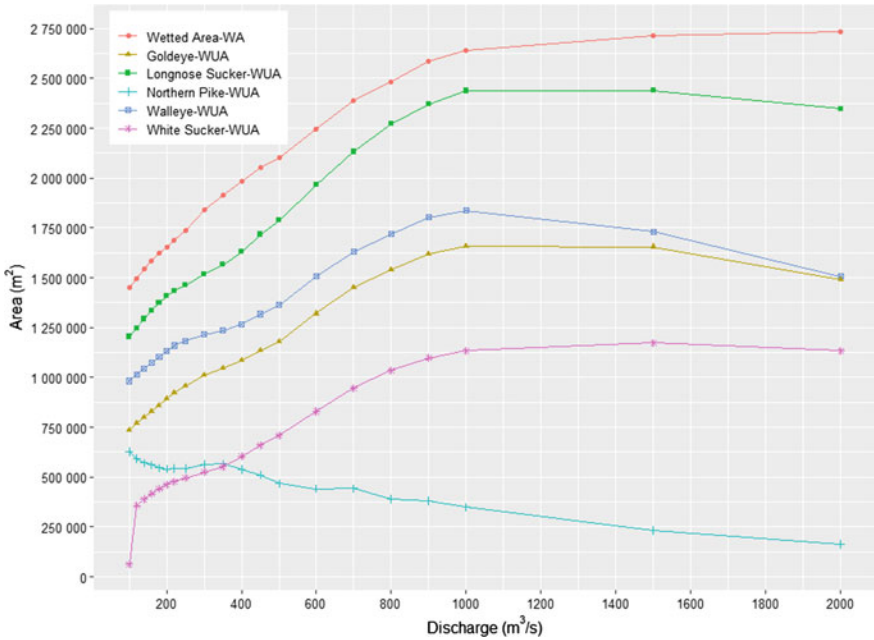


Fig. 1 Total wetted area (WA) and weighted usable area (WUA) curves for target species

1000 m³/s. At this flow point, the WUA curve for Longnose Sucker and Walleye began to decrease and the WUA curve for Goldeye and White Sucker plateaued. Therefore, the optimal flow for Goldeye, Longnose Sucker, Walleye, and White Sucker was 1000 m³/s (Table 1). The WUA curve for Northern Pike did not follow the same trend as the WA curve. This is expected as Northern Pike prefer to inhabit slow moving, heavily vegetated rivers. The largest available habitat based on HSC curves for Northern Pike occurred at 100 m³/s, which was the lowest flow simulated (Table 1).

The percent total habitat (WUA/WA) was calculated at low flow (100 m³/s) for each fish species (Table 1). WA over estimated the amount of habitat available in

Table 1 Summary of habitat availability analysis for target species

| Species | Optimal flow (m ³ /s) | Optimal flow percent difference (%) | Weighted usable area at 100 m ³ /s (m ²) | Percent total habitat at 100 m ³ /s (%) |
|-----------------|----------------------------------|-------------------------------------|---|--|
| Goldeye | 1000 | 0 | 740,000 | 51 |
| Longnose Sucker | 1000 | 0 | 1,200,000 | 83 |
| Northern Pike | 100 | 90 | 630,000 | 43 |
| Walleye | 1000 | 0 | 980,000 | 68 |
| White Sucker | 1000 | 0 | 64,000 | 4 |

comparison to WUA for each fish species. This was most notable for White Sucker, where the WUA was 4% of the WA at 100 m³/s.

5 Conclusion

The percent difference in optimal flow indicated the value gained by using HSC curves in aquatic assessments. The optimal flows based on WUA curves were equivalent to the optimal flow based on the WA curve for four of the five fish species studied. Therefore, no value was gained by using HSC curves to understand the optimization of habitat availability for these four adult fish species. Future research should focus on understanding the similarities in habitat preferences of the four fish species that had equivalent HSC-based optimal flow as the WA curve.

The mean value of WUA as a percent of total habitat at low flow was 50% (minimum and maximum were 4% and 83%, respectively). Based on this research, the WA approach does not provide a promising alternative to HSC-based methodologies for habitat quantity offsetting requirements.

Preliminary results showed that HSC-based area-flow curves followed a similar trajectory as the site-specific WA curve and that similar flows resulted in optimized habitat for a majority of the fish species investigated in this study. To confirm this study, we recommend expanding this research to other study sites, species of fish, and fish life stages.

Acknowledgements The authors gratefully acknowledge that the data used in this paper were the result of surveys funded by Department of Fisheries and Oceans Canada (DFO), and the Cumulative Environmental Management Association (CEMA), a non-profit multi-stakeholder organization researching and making recommendations on cumulative environmental impacts within the northeast region of Alberta.

References

1. Bovee KD (1986) Development and evaluation of habitat suitability criteria for use in the instream flow incremental methodology. Instream flow information paper 21. US Department of the Interior, Fish and Wildlife Service Biology reports, vol 86, no 7, 235 p
2. Bovee KD, Zuboy JR (1988) Proceedings of a workshop on the development and evaluation of habitat suitability criteria. US Department of the Interior, Fish and Wildlife Service Biology reports, vol 88, no 11, 407 p
3. Conallin J, Boegh E, Kroegsgaard Jensen J (2010) Instream physical habitat modelling types: an analysis as stream hydromorphological modelling tools for EU water resource managers. *Int J River Basin Manag* 8(1):93–107
4. Dunbar MJ, Alfredsen K, Harby A (2011) Hydraulic-habitat modelling for setting environmental river flow needs for salmonids. *Fish Manage Ecol* 19(6):500–517
5. Ghamry H, Katopodis C (2017) Computational optimization in simulating velocities and water-surface elevations for habitat–flow functions in low-slope rivers. *J Ecohydraulics* 2(2):99–121

6. Ghanem AM, Steffler P, Hicka F, Katapodis C (1996) Two-dimensional hydraulic simulation of physical habitat conditions in flowing streams. *Regul Rivers Res Manage* 12(2–3):185–200
7. Mochnaczn NJ, Ghamry HK, Enders EC, Watkinson DA, Gallagher CP, Reist JD (2019) Flow and spawning habitat relationships for Dolly Varden: understanding habitat–population dynamics in the Canadian Western Arctic. *River Res Appl* 36(1):68–78
8. Nestler JM, Milhous RT, Payne TR, Smith DL (2019) History and review of the habitat suitability criteria curve in applied aquatic ecology. *River Res Appl* 35(5):1–26
9. Rosenfeld J (2003) Assessing the habitat requirements of stream fishes: an overview and evaluation of different approaches. *Trans Am Fish Soc* 132:953–968
10. Rosenfeld JS, Leiter T, Lindner G, Rothman L (2005) Food abundance and fish density alters habitat selection, growth, and habitat suitability curves for juvenile Coho Salmon (*Oncorhynchus kisutch*). *Can J Fish Aquat Sci* 62:1691–1701
11. Stalnaker C, Lamb BL, Henriksen J, Bovee KD, Bartholow J (1995) The instream flow incremental methodology: a primer for IFM. United States Department of the Interior, National Biological Service Biology reports, vol 29, 45 p
12. Steffler P, Blackburn J (2002) Two dimensional depth averaged model of river hydrodynamics and fish habitat. Introduction to depth averaged modeling and user’s manual. University of Alberta, Edmonton, Canada
13. Trillium Engineering and Hydrographics Inc. (2005) Summer survey of the Athabasca River at Embarras (Reach #2). CEMA. Project Number: 04-568. Lower Athabasca River habitat surveys

Influence of Surge Waves on the Transport of Macroplastics



P. Patel, M. Krol, and S. Karimpour

1 Introduction

Marine plastics represent 70% [1] to 80% [2] of total marine waste from surface levels to deep-sea sediments. It is well understood that plastic waste has profound consequences on marine wildlife, environment, and the human food chain; plastic waste is still to blame for the death of 100,000 marine mammals and more than a million seabirds each year [3]. In addition, plastic waste has been ingested by 52% of sea turtles [4], and over 90% of seabirds [5]. Once consumed, microplastics (plastics sized smaller than 5 mm) are known to release toxic chemicals and these chemicals may also end up in the human food chain [6]. According to Plastics [7], 368 million tonnes of plastics were manufactured in 2019, of which 300 million tonnes turned into plastic waste. With the present trends, annual plastic production is expected to reach 1124 million tonnes by 2050 [8]. Polyethylene (low-density polyethylene and high-density polyethylene), a lightweight synthetic resin with countless applications was responsible for 30% of the total plastics market in 2019, followed by polypropylene accounting for 19.4% of global plastics [7]. These two polymers also happen to be predominant in aquatic environments. According to a recent study, about 200 million tonnes of plastic debris are currently circulating in the Atlantic Ocean [9]. Several studies carried out in multiple lakes, rivers, beaches and five subtropical gyres

P. Patel (✉) · M. Krol · S. Karimpour
Department of Civil Engineering, Lassonde School of Engineering, York University, Toronto,
Canada

e-mail: Preet26@my.yorku.ca

M. Krol

e-mail: Magdalena.krol@lassonde.yorku.ca

S. Karimpour

e-mail: Shooka.karimpour@lassonde.yorku.ca

concluded that polyethylene and polypropylene were the most abundant source of plastic waste in marine environment [10–15].

Each year, an estimated 12.7 million tonnes of plastic waste manages to seep into the marine environment through various pathways [16]. Land-based sources are responsible for approximately 80% of annual plastics inputs, while the remaining 20% originates from marine activities. Rivers and coastal events are responsible for transporting a significant portion of land-based plastics to the marine environment. In 2010, coastal activities (human activities within 50 km of the coastline) produced 100 million tonnes of plastic waste, of that, approximately 33 million tonnes were categorized as mismanaged plastic waste [16]. Inadequately managed waste near the coastline can end up in marine environment with the help of coastal events (such as tsunamis, surge, and flood events). For example, the most powerful tsunami ever recorded in Japan transported 5 million tonnes of debris to the marine environment [17]—this amount is within the range of annual global plastics input into the oceans. Based on a recent estimate, coastal activities drive approximately 8.2 million tonnes of plastic waste into the marine environment each year [18], and with increasing plastic production this estimate is expected to rise. In addition, riverine systems are analogous to highways for transporting plastic waste. World rivers transport 1.15–2.41 million tonnes of plastic waste from inland sources to the marine environment [19]. Regardless of the transport methods, accumulation and transport of plastic waste is greatly influenced by the density, size and surface area of plastics, as seen in numerical modelling of coastal waters [20]. Tidal bores have substantial impact on the river mouth, as debris are suspended due to wave motion, and bed erosion takes place beneath the bore front [21, 22]. Such phenomenon gives a sharp rise to sediment concentration in the water column and tidal bores then transport the suspended debris. Windage also has the potential to transport plastic litter. Macroplastics (>5 mm) are more susceptible to windage compared to microplastics, since buoyant macroplastics drift in the upper layer where the stokes drift is more rapid and the transportation is efficient [23]. Horizontal transport of plastic litter during a surge wave can be explained by surge wave Froude number and other surge wave properties. However, vertical behaviour of plastics in a surge/tsunami wave remains to be adequately determined.

Substantial research efforts have been devoted to examining the abundance, fate, and transport of microplastics and macroplastics in the marine environment. However, the transport of plastic debris in coastal events and riverine system due to surge waves have commonly been neglected. Low-density materials, such as polyethylene (PE) and polypropylene (PP), are susceptible to turbulent mixing which has a significant impact on the residence time, transport pathways, and accumulation of plastic debris. Intense turbulent kinetic energy generated in breaking bores and surge waves with a surge wave Froude number of (Fr_s) greater than 1.5 have substantial influence on the movement of plastic debris. Therefore, the motive behind this novel research is to comprehend the impact of the Froude number on the carrying capacity of surge waves and transport of plastic debris.

2 Methodology

2.1 Experimental Setup

For this research, a GUNT hydraulic flume with dimensions of 5000 mm by 309 mm by 450 mm (Fig. 1) was used. A breaking bore was generated by obstructing the incoming flow by a sudden closure of a sluice gate. Solid plastic balls (referred to hereafter as particles) made up of PP or ACR were introduced in a surge wave. These polymers were selected as they are the predominant plastic seen in marine environments [24, 25]; in addition, selected polymers have densities that are marginally higher and lower than water, increasing the involvement of these polymers in a surge wave. A GoPro Hero9 camera was utilized to observe the transport of plastic debris at a resolution of 2.7K with 120 frames per second. Transport of the particles (with a diameter of 1–2.5 cm) were analyzed with particle tracking velocimetry program. A camera was mounted on the side viewing panel of the hydraulic flume, as depicted in Fig. 1. A checkerboard with 2 cm grids was added to the rear of the viewing panelling, allowing the grid to be visible by the camera. This calibration grid allowed for lens correction during post processing.

Two Mic+35/IU/TC acoustic displacement sensors, recording at 50.00 ms (20 Hz), were mounted on top of the flume to measure surface water displacement (Fig. 1). The sensors were linked to the Catman Data Acquisition and Analysis program, with the use of HBM QuantumX MX840B, to digitalize water depth measurements.

Surge waves with various Fr_s were generated by governing flowrate and upstream water depth, d_1 :

$$Fr_s = \frac{u_1 + c}{\sqrt{gd_1}} \tag{1}$$

where c is the celerity and u_1 is the velocity upstream. Figure 2 illustrates positive bore propagating in the upstream direction, where P#1 and P#2 indicate downstream

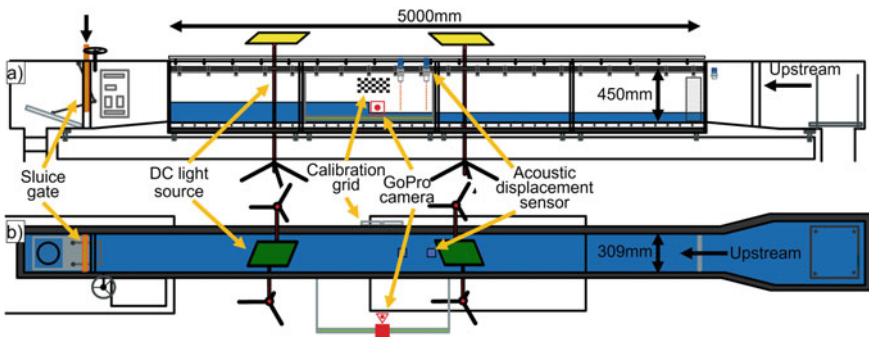


Fig. 1 Experimental setup schematic showing flume: a side view and b top view

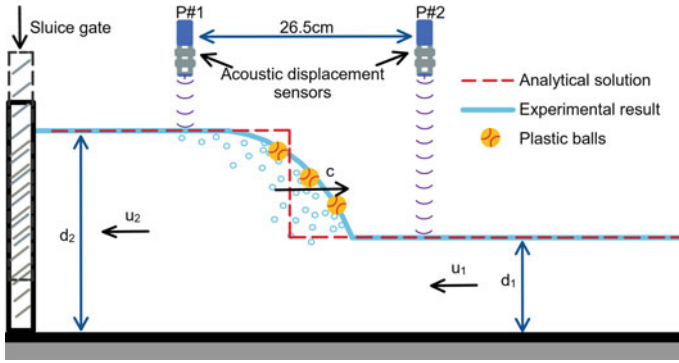


Fig. 2 Positive bore advancing upstream (observed from a stationary position)

Table 1 Summary of experimental parameters and results obtained at a flowrate of 25 m³/h. The density of freshwater is denoted by ρ_w

| Particle type | Test# | Polymer density, ρ_P , (g/cm ³) | Particle diameter (cm) | d_1 (cm) | u_1 (cm/s) | d_2 (cm) | c (cm/s) | Fr_s |
|------------------------------|-------|--|------------------------|------------|--------------|------------|------------|--------|
| ACR ($\rho_P > \rho_w$) | 1 | 1.17–1.20 | 2.54 | 3.90 | 57.62 | 7.72 | 53.00 | 1.79 |
| | 2 | 1.17–1.20 | 1.91 | 4.11 | 54.67 | 7.61 | 53.00 | 1.70 |
| | 3 | 1.17–1.20 | 1.11 | 4.01 | 56.06 | 7.74 | 53.00 | 1.74 |
| PP ($\rho_P < \rho_w$) | 1 | 0.90–0.93 | 2.54 | 3.97 | 56.55 | 7.65 | 53.00 | 1.75 |
| | 2 | 0.90–0.93 | 1.91 | 4.05 | 55.54 | 7.52 | 53.00 | 1.72 |
| | 3 | 0.90–0.93 | 1.11 | 3.90 | 57.62 | 7.14 | 48.18 | 1.71 |

and upstream acoustic displacement sensor locations, respectively. Particles with different diameters and densities were tested at a flowrate of 25 m³/h (Table 1).

2.2 Particle Tracking Velocimetry

Tracker, an open-source PTV program developed by the multiphase flow group at US Department of Energy, National Energy Technology Laboratory [26], was utilized to determine particle trajectories and velocities, and to examine the movement of plastic clusters, while to manually correct particle trajectory ImageJ was employed [27]. To minimize post processing effort, experimental equipment such as camera and lighting were carefully set up. Lastly, curvature on camera lenses produces undesirable distortion in a video signal and can generate inaccurate particle trajectories and velocities. There are several post-processing methods to alter distorted images. Widely used image manipulation software, Photoshop CS6, was used to correct distortion.

3 Analytical Solutions

The analytical solution for surge wave Froude number is based on the reflection of a bore on a sluice gate, where friction is usually neglected to simplify the analysis. Considering a positive surge wave seen by an observer traveling at celerity, c , where the upstream velocity is $(u_1 + c)$, and downstream velocity, u_2 , is $(u_2 + c)$. Figure 2 shows the parameters used in the analytical solutions where d_1 is the initial flow depth being impeded by a sluice gate closure, resulting in a positive bore with celerity, c , and downstream flow depth, d_2 , advancing towards the upstream direction. Downstream velocity is zero due to complete closure.

Considering the continuity equation, with a unit depth, in a moving frame of reference (Eq. 2).

$$d_1(u_1 + c) = d_2(u_2 + c) \quad (2)$$

Unit flow, q , is a product of upstream water depth and upstream velocity. This relation leads to an analytical solution for celerity, represented by Eq. 3.

$$c = \frac{q}{d_2 - d_1} \quad (3)$$

Considering the momentum equation in a moving frame of reference we can calculate the analytical upstream depth (Eq. 4).

$$\frac{1}{2}d_1^2 - \frac{1}{2}d_2^2 = \frac{c^2 d_2^2}{g} - \frac{(u_1 + c)^2 d_1}{g} \quad (4)$$

Replacing celerity with Eq. 3 and solving for q , yields Eq. 5, the analytical flow.

$$q^2 = \frac{1}{2} \frac{g}{d_2} (d_2 + d_1) d_1 (d_2 - d_1)^2 \quad (5)$$

Analytical downstream water depth can be obtained by solving Eq. 5 while the analytical surge wave Froude number may also be represented by water depth (Eq. 6).

$$Fr_s^2 = \frac{1}{2} \frac{d_2}{d_1} \left(\frac{d_2}{d_1} + 1 \right) \quad (6)$$

4 Results

Water depths before and after the surge wave were measured using acoustic sensors. Based on flowrate and upstream water depth, d_1 , upstream velocity, u_1 , was determined. Experimental celerity was estimated based on the time lag and the distance between two acoustic displacement sensors (26.5 cm) (Fig. 3). The resulting surge celerity is shown in Table 1. Experimental surge wave Froude number was obtained using Eq. 1, average experimental Fr_s are presented in Table 1. Using Eq. 3, analytical surge wave celerity was calculated. Equation 5 was solved to obtain water depth after the surge wave. The experimental surge wave Froude number was achieved with the use of Eq. 6, analytical celerity, and analytical downstream water depth.

Based on the analytical solution, if flowrate and upstream water depth remains constant, there would be no fluctuation in celerity and downstream water depth, and as a result no deviation in surge wave Froude number. However, due to the turbulent nature of the phenomena combined with the effect of environmental perturbations, acoustic displacement sensors reported water depth fluctuation upstream, across, and downstream of the surge wave. Surge wave Froude number is dependent on upstream water depth, celerity and flowrate. Compared to the analytical solution, the overall experimental mean downstream water depth, celerity, and surge wave Froude number have a discrepancy of 7.9%, 3.0%, and 1.5%, respectively. Furthermore, the identification of surge length based on depth profiles is not trivial, since it is challenging to locate surge toe.

Transport of plastic debris is greatly influenced by their density. In this study, two different polymer densities were examined: ACR and PP. The ACR plastic balls have densities greater than of water and hence are negatively buoyant and sink, whereas PP plastic balls have densities marginally lower than the density of water and are positively buoyant (Table 1). Figure 4 shows series of frames of particle trajectory for a particle diameter of 2.54 cm. As predicted, the low-density particle (PP) was buoyant and considerably involved with the highly aerated flow in a roller at the bore front (Fig. 4a). Conversely, the ACR particle settled on the flume bed and was indifferent to surge wave Froude number of as high as 1.79 (Fig. 4b).

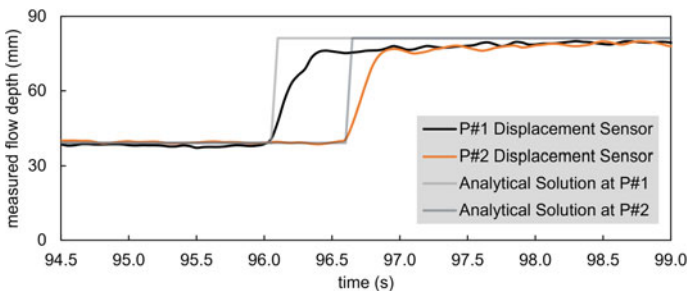


Fig. 3 Time lag observed between two acoustic displacement sensors (P#1 and P#2) during ACR-1 with an experimental Fr_s of 1.79. The grey lines denote the analytical solution with a Fr_s of 1.79

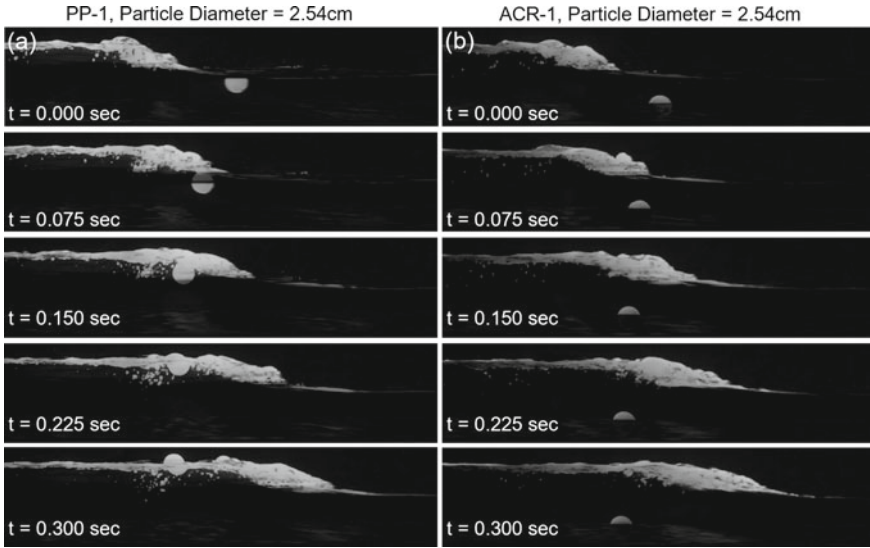


Fig. 4 Sequence of images showing the interaction of a surge wave on the transport of **a** PP, and **b** ACR particles of identical diameter. Particles were observed 0.042 s prior to impacting the surge wave. Surge wave was propagating upstream (left to right). Geometric image distortion was 2.5% and vertical perspective transformation was 8%

Figure 5 illustrates particles vertical displacement and shows that the PP particle had substantial movement in the vertical direction during the entrainment process, while the ACR particle displacement was minimal as the denser particle was not entrained in the surge wave. However, although the ACR particle had settled on the flume bed, it was still influenced by the upstream velocity and moved towards the downstream region as the transient recirculation behind the surge wave had some impact on the ACR particle’s trajectory and velocity. As a result, the ACR-1 particle’s downstream

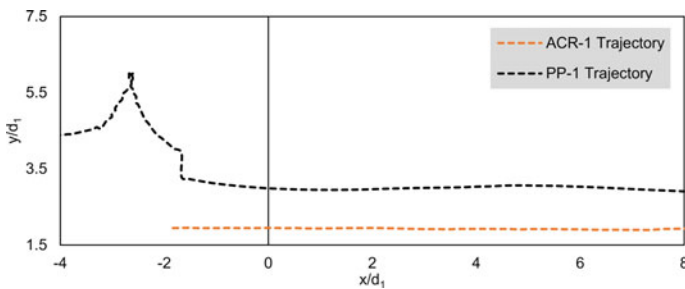


Fig. 5 Tracker and ImageJ were employed to determine particle trajectory for ACR-1 and PP-1. The particle encountered the surge wave (propagating upstream—left to right) at $x/d_1 = 0$. Horizontal and vertical trajectory of the particle are indicated by the x -position/ d_1 (x/d_1) and y -position/ d_1 (y/d_1), respectively; the flume bed is situated at $y/d_1 = 0$

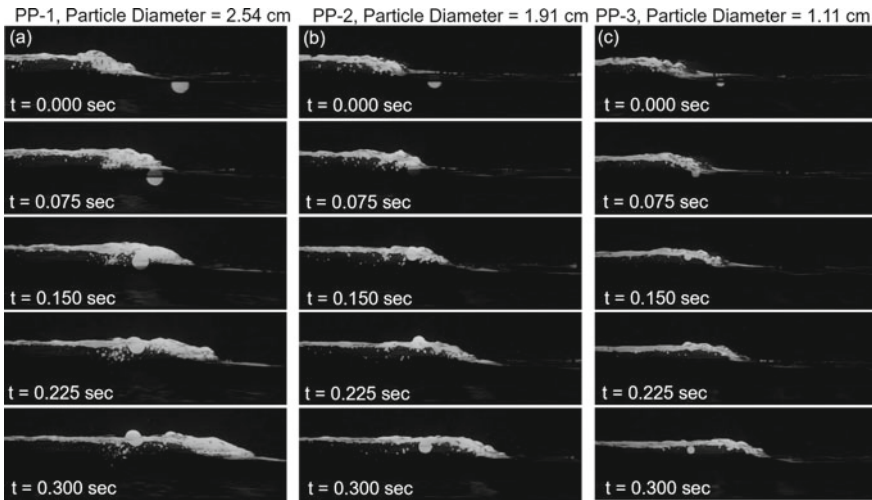


Fig. 6 Sequence of images showing the impact surge wave had on PP particle with a diameter of: **a** 2.54 cm, **b** 1.91 cm, and **c** 1.11 cm. Particle behaviour was observed 0.042 s prior to arriving at the bore front. Surge wave was propagating upstream (left to right). Geometric image distortion was 2.5% and vertical perspective transformation was 8%

velocity started to decelerate, and it gradually changed direction towards the upstream region during the passage of the surge wave (Fig. 5). Similar transport was observed for ACR-2 and ACR-3.

Transport of floating debris can also be dependent on particle size. To analyze the impact of particles with different sizes, tests were carried out with PP plastic particles with varying diameters (Table 1). As previously discussed, PP particle with a diameter of 2.54 cm was entrained in the surge wave. This is also shown in Fig. 6a. Similarly, 1.91 and 1.11 cm PP particles exhibited this behaviour where the particles were entrained in a roller at the bore front (see Fig. 6b, c). In addition, transport of PP particles in the vertical direction exhibited similar behaviour regardless of the particle size (Fig. 7). After closely analyzing the particle trajectories it was observed that entrainment decreased as the particle diameter decreased (Fig. 7). The velocity perturbation in the surge front and behind the surge have the ability to mobilize the plastic debris in the opposite direction of its buoyancy. In smaller particles the force required to overcome buoyancy is less and hence it is more likely for the smaller particles to engage in the highly turbulent wave front.

5 Conclusion

Limited information is available to date on the transport of plastic debris due to surge waves in rivers and coastal events. In this study, laboratory experiments were carried

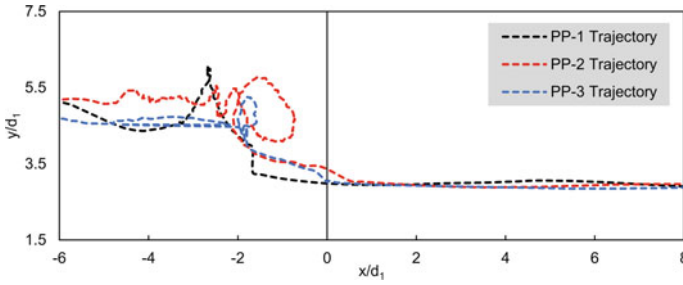


Fig. 7 Particle trajectory for PP-1, PP-2, and PP-3 was achieved using Tracker and ImageJ. The particle was impacted by the surge wave (propagating upstream—left to right) at $x/d_1 = 0$. Horizontal and vertical trajectory of the particle are indicated by the x -position/ d_1 (x/d_1) and y -position/ d_1 (y/d_1), respectively; the flume bed is located at $y/d_1 = 0$

out with varying polymer densities and sizes to examine this transport process. Surge waves with higher Froude number result in greater turbulent motion which suspends particles in the water column and the advected particles are then transported upstream with the breaking bore. An ideal breaking bore was generated with a Froude number of greater than 1.5 by regulating flowrate, gate closure, and upstream water depth. Particles with varying densities and sizes were introduced into the surge wave, the process was then recorded, and particle trajectories were obtained using PTV.

The effect of the surge wave on particles of two distinct polymer densities and two different diameters were examined in this study. Firstly, the influence of particle density was analyzed by using ACR and PP plastic particles. PP was entrained in the surge wave, with considerable vertical transport due to turbulent motion in the bore front and the lighter density of the particle. This was not observed with the ACR particle even when the surge wave Froude number was as high as 1.79. However, the ACR particle was still influenced by the transient recirculation in a surge wave. When particles with the same density but varying diameters were analyzed, vertical transport was quite evident for all PP particle sizes since PP is buoyant and fairly involved with the turbulent motion in a surge front. Overall, as the diameter of PP particles decreased, so did the involvement in the surge wave.

In future studies, several other polymer densities and sizes will be assessed with a wide range of surge wave Froude numbers in order to fully comprehend the transport of floating debris. In addition, 3D-PTV will be considered to examine the particle movement in the z -direction (along the flume width).

Acknowledgements We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

1. Kornei K (2017) Plastic makes up nearly 70% of all ocean litter. *Science* 363:6426
2. Thevenon F, Carroll C, Sousa J (eds) (2014) Plastic debris in the ocean: the characterization of marine plastics and their environmental impacts, situation analysis report. IUCN, Gland, Switzerland: 52
3. UNESCO (2017) Facts and figures on marine pollution. Retrieved from <http://www.unesco.org/new/en/natural-sciences/ioc-oceans/focus-areas/rio-20-ocean/blueprint-for-the-future-we-want/marine-pollution/facts-and-figures-on-marine-pollution/>
4. Schuyler QA, Wilcox C, Townsend KA, Wedemeyer-Strombel KR, Balazs G, van Sebille E, Hardesty BD (2016) Risk analysis reveals global hotspots for marine debris ingestion by sea turtles. *Glob Change Biol* 22(2):567–576
5. Wilcox C, Van Sebille E, Hardesty BD (2015) Threat of plastic pollution to seabirds is global, pervasive, and increasing. *Proc Natl Acad Sci USA* 112(38):11899–11904
6. Han E (2016) Microbeads are leaching toxic chemicals into fish, sparking public health fears. *The Sydney Morning Herald*, August 16, 2016. Retrieved from <https://www.smh.com.au/business/consumer-affairs/microbeads-are-leaching-toxic-chemicals-into-fish-sparking-public-health-fears-20160816-gqtlpk.html>
7. Plastics Europe (2020) Plastics—the facts 2020. PlasticEurope. Retrieved from <https://www.plasticseurope.org/en/resources/publications/4312-plastics-facts-2020>
8. Ellen MacArthur Foundation (2016) The new plastics economy: rethinking the future of plastics. Ellen MacArthur Foundation, Cowes, United Kingdom
9. National Oceanography Centre, UK (2020) There is at least 10 times more plastic in the Atlantic than previously thought. *ScienceDaily*, August 18, 2020. Retrieved from www.sciencedaily.com/releases/2020/08/200818114940.htm
10. Gómez V, Pozo K, Nuñez D, Přibylková P, Audy O, Bains M, Fossi MC, Klánová J (2020) Marine plastic debris in Central Chile: characterization and abundance of macroplastics and burden of persistent organic pollutants (POPs). *Mar Pollut Bull* 152:110881
11. Klein S, Worch E, Knepper TP (2015) Occurrence and spatial distribution of microplastics in river shore sediments of the Rhine-Main area in Germany. *Environ Sci Technol* 49(10):6070–6076
12. Lebreton L, Slat B, Ferrari F, Sainte-Rose B, Aitken J, Marthouse R, Hajbane S, Cunsolo S, Schwarz A, Levivier A, Noble K (2018) Evidence that the Great Pacific garbage patch is rapidly accumulating plastic. *Sci Rep* 8(1):1–15
13. Mason SA, Kammin L, Eriksen M, Aleid G, Wilson S, Box C, Williamson N, Riley A (2016) Pelagic plastic pollution within the surface waters of Lake Michigan, USA. *J Great Lakes Res* 42(4):753–759
14. Noik VJ, Tuah PM (2015) A first survey on the abundance of plastics fragments and particles on two sandy beaches in Kuching, Sarawak, Malaysia. *IOP Conf Ser Mater Sci Eng* 78:012035
15. Suaria G, Avio CG, Mineo A, Lattin GL, Magaldi MG, Belmonte G, Moore CJ, Regoli F, Aliani S (2016) The Mediterranean plastic soup: synthetic polymers in Mediterranean surface waters. *Sci Rep* 6(1):1–10
16. Jambeck JR, Geyer R, Wilcox C, Siegler TR, Perryman M, Andrady A, Narayan R, Law KL (2015) Plastic waste inputs from land into the ocean. *Science* 347(6223):768–771
17. Ministry of the Environment, Japan (2012) Estimated total amount of debris washed out by the Great East Japan Earthquake. Retrieved from <http://www.env.go.jp/en/focus/docs/files/20120901-57.pdf>
18. Parker L (2018) Fast facts about plastic pollution. *National Geographic*, December 20, 2018. Retrieved from <https://www.nationalgeographic.com/news/2018/05/plastics-facts-infographics-ocean-pollution/>
19. Lebreton LC, Van Der Zwet J, Damsteeg JW, Slat B, Andrady A, Reisser J (2017) River plastic emissions to the world's oceans. *Nat Commun* 8(1):1–10
20. van Utenhove E (2019) Modelling the transport and fate of buoyant macroplastics in coastal waters. Delft University of Technology

21. Docherty NJ, Chanson H (2012) Physical modeling of unsteady turbulence in breaking tidal bores. *J Hydraul Eng* 138(5):412–419
22. Koch C, Chanson H (2008) Turbulent mixing beneath an undular bore front. *J Coastal Res* 24(4):999–1007
23. Iiwasaki S, Isobe A, Kako SI, Uchida K, Tokai T (2017) Fate of microplastics and mesoplastics carried by surface currents and wind waves: a numerical model approach in the Sea of Japan. *Mar Pollut Bull* 121(1–2):85–96
24. Erni-Cassola G, Zadjelovic V, Gibson MI, Christie-Oleza JA (2019) Distribution of plastic polymer types in the marine environment; A meta-analysis. *J Hazard Mater* 369:691–698
25. Schwarz AE, Lighthart TN, Boukris E, van Harmelen T (2019) Sources, transport, and accumulation of different types of plastic litter in aquatic environments: a review study. *Mar Pollut Bull* 143:92–100
26. US Department of Energy National Energy Technology Laboratory (2018) Tracker. Retrieved from <https://mfix.netl.doe.gov/tracker/>
27. Rasband WS (1997) ImageJ. U. S. National Institutes of Health, Bethesda, Maryland, USA. Retrieved from <https://imagej.nih.gov/ij/>

Establishing Acceleration Profiles of Light-Duty Vehicles Departing in a Straight Path from Two-Way Stop-Controlled Intersections



Essam Dabbour and Olaa Dabbour

1 Introduction

At two-way stop controlled (TWSC) intersections, drivers on the minor road need to select adequate gaps for their safe departure. Those gaps are needed for any departure scenario, including right-turning departures, left-turning departures, or straight departures along the minor road (see Fig. 1).

It is established that humans cannot directly perceive time gaps; and therefore, departing drivers must instead perceive the distance and speed of the vehicle approaching in the cross-traffic stream to indirectly estimate the available gap, t_{app} . Furthermore, departing drivers must also estimate the time needed for their departing vehicles to clear the intersection, t_{dep} , which depends on the acceleration capabilities of the departing vehicle. Once departing drivers are able to estimate both t_{app} and t_{dep} , they must compare them together to decide whether to accept the gap. These complex mental tasks may result in some departing drivers failing to accurately perceive t_{app} and t_{dep} and colliding with an approaching vehicle if an inadequate gap was selected.

This paper establishes acceleration profiles of light-duty vehicles when departing in a straight path from rest at TWSC intersections by understanding how drivers depart those intersections based on their behavior patterns as well as the mechanical characteristics of their vehicles. By understanding how drivers accelerate when departing a TWSC intersection, we can understand how gap-acceptance decisions are made so that countermeasures may be implemented to reduce the probability of departing drivers accepting improper gaps. Furthermore, establishing an acceleration profile is also important when analyzing and investigating intersection accidents at a

E. Dabbour (✉)
Advantage Forensics Inc., Toronto, Canada
e-mail: edabbour@aforensics.ca

O. Dabbour
Ryerson University, Toronto, Canada

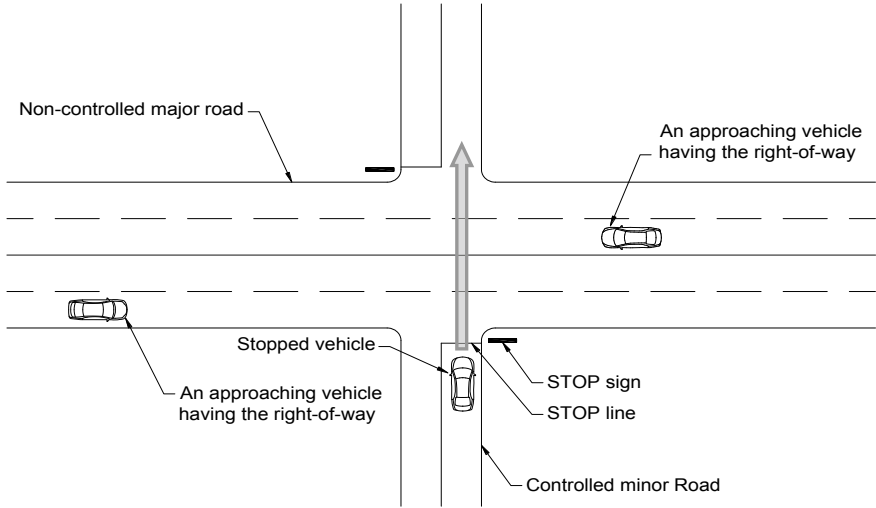


Fig. 1 Straight departure of a stopped vehicle at TWSC intersections

microscopic level to understand how an accident occurred and whether the departing driver was the one at fault (e.g., by selecting an improper gap), or the approaching driver was the one at fault (e.g., by speeding).

2 Proposed Method

It was previously found that there is an inverse relationship between the speed and the acceleration rate [8, 10]. This linear-decreasing model is one of the most common models utilized to describe how vehicles accelerate when starting from a rest position [2–7, 9, 10]. According to this model, the acceleration at any time is given by:

$$a = dv/dt = \alpha - \beta v \pm G_1 g \tag{1}$$

where a is the acceleration rate (m/s^2); v is the speed (m/s); α is the initial acceleration rate (m/s^2); β is the rate of change in acceleration with respect to speed (s^{-1}); G_1 is the grade of the minor road (m/m); and g is the gravity ($9.81 m/s^2$).

The speed at any time, v , may be obtained by integrating the above equation with respect to time, t :

$$v = [(\alpha \pm G_1 g)/\beta] - [(\alpha \pm G_1 g)/\beta - v_0]e^{-\beta t} \tag{2}$$

where v_0 is the initial speed (m/s). The traveled distance, d , is calculated by integrating the above equation with respect to time, giving:

$$d = [t(\alpha \pm G_1g)/\beta] - [(\alpha \pm G_1g)/\beta - v_0][(1 - e^{-\beta t})/\beta] \quad (3)$$

The following section provides a methodology to estimate the values of the coefficients α and β when a vehicle departs a TWSC intersection in a straight path.

3 Collecting and Preparing Data

Global Positioning System (GPS) data logging devices were utilized to collect data from different light-duty vehicles departing TWSC intersections in straight paths. The logging devices registered the positions and speeds of the vehicles, on which the logging devices were mounted onto, every second. The research team extracted the records collected by those devices, as spreadsheets, into a personal computer. In those records, each row represented the record collected at a certain second. The records were also extracted to a map, which was linked to the spreadsheet using the date and time stamps embedded in every record.

There were 32 participating drivers (18 male and 14 female) from Abu Dhabi (in the United Arab Emirates) who participated in this study from March 2018 to November 2019. The statistical models developed in this study were calibrated using data collected from those drivers. Furthermore, the models were validated using more data collected by 12 other drivers (7 male and 5 female) in the Greater Toronto Area (in Canada) who participated in this study from June to August 2018.

Participating drivers were provided with GPS data logging devices so that every driver would install a device on their private vehicle. Participating drivers used their equipped vehicles for their daily routine trips for four to five days; after which, they returned the devices to the research team to download the logged data. To maintain the naturalistic nature of the experiment, the research team explained to the drivers that the study aimed at generally understanding driver's behavior without discussing the exact details of the study. According to the approved research ethics application, drivers' identities were removed from the analyzed data to ensure their privacy.

The characteristics of the drivers and vehicles included in the study are shown in Table 1, which shows that 32 vehicles/drivers were used for model calibration, and 12 vehicles/drivers were used for model validation. Table 2 shows a sample of a spreadsheet that represents one acceleration profile with 15 data points, and Fig. 2 shows the location of that acceleration profile. A vehicle was assumed to complete its acceleration maneuver when the speed does not change by more than 1 km/h during three consecutive seconds.

The collected data records consisted of 548 acceleration profiles for model calibration, and other 174 acceleration profiles for model validation. The calibration records included 381 profiles related to drivers who crossed major roads that had four lanes, and 167 profiles related to drivers who crossed major roads that had two lanes. The validation records included 102 profiles related to drivers who crossed major roads that had four lanes, and 72 profiles related to drivers who crossed major roads that had two lanes.

Table 1 Drivers and vehicles participated in the study

| Characteristic | Calibration | Validation |
|-------------------------------|-------------|------------|
| <i>Driver's age (years)</i> | | |
| Mean (and standard deviation) | 36 (15.3) | 40 (17.6) |
| Minimum | 19 | 19 |
| Maximum | 71 | 70 |
| Median | 33 | 34 |
| <i>Vehicle age (years)</i> | | |
| Mean (and standard deviation) | 7 (3.8) | 7 (5.5) |
| Minimum | 1 | 1 |
| Maximum | 16 | 19 |
| Median | 6 | 7 |
| <i>Engine power (hp)</i> | | |
| Mean (and standard deviation) | 239 (81.8) | 214 (38.7) |
| Minimum | 106 | 172 |
| Maximum | 400 | 283 |
| Median | 237 | 195 |
| <i>Vehicle type</i> | | |
| Passenger car | 17 | 5 |
| SUV | 10 | 2 |
| Mini van | 4 | 3 |
| Pick-up truck | 1 | 2 |

The acceleration rate at the i th second, a_i , during departure is calculated as the difference between the speed during the previous second, v_{i-1} , and that during the next second, v_{i+1} , divided by the time needed to change the speed from v_{i-1} to v_{i+1} , which is typically 2 s:

$$a_i = \frac{(v_{i+1} - v_{i-1})}{2} \quad [0 < i < n] \quad (4)$$

where a_i is the acceleration rate at the i th second, v_{i+1} is the speed at the $(i + 1)$ th second, v_{i-1} is the speed at the $(i - 1)$ th second, and n is the total number of data points during the given acceleration profile. Based on the relationship represented in Eq. 4, a regression model may be developed by relating the acceleration rate, a_i , with the instantaneous speed, v_i , as represented by:

$$a_i = \alpha - \beta v_i + \varepsilon \quad (5)$$

In Eq. 5, the speed, v_i , is taken from the experimental data, and the acceleration rate, a_i , is estimated from Eq. 4. The maximum likelihood method is utilized to

Table 2 Sample data sheet collected by the GPS data logging device

| Index | Track number | Local date | Local time | Latitude | N/S | Longitude | E/W | Altitude | Speed (km/h) |
|-------|--------------|------------|------------|----------|-----|-----------|-----|----------|--------------|
| 1027 | 7 | 5/6/2019 | 15:55:06 | 24.32896 | N | 54.53313 | E | 8.564209 | 0.0 |
| 1028 | 7 | 5/6/2019 | 15:55:07 | 24.32894 | N | 54.53313 | E | 7.946899 | 0.1 |
| 1029 | 7 | 5/6/2019 | 15:55:08 | 24.32894 | N | 54.53313 | E | 7.981323 | 0.6 |
| 1030 | 7 | 5/6/2019 | 15:55:09 | 24.32891 | N | 54.53313 | E | 8.010742 | 1.3 |
| 1031 | 7 | 5/6/2019 | 15:55:10 | 24.32888 | N | 54.53313 | E | 8.042236 | 4.2 |
| 1032 | 7 | 5/6/2019 | 15:55:11 | 24.32885 | N | 54.53314 | E | 7.413208 | 13.1 |
| 1033 | 7 | 5/6/2019 | 15:55:12 | 24.32880 | N | 54.53314 | E | 7.391113 | 18.2 |
| 1034 | 7 | 5/6/2019 | 15:55:13 | 24.32873 | N | 54.53314 | E | 8.328125 | 26.7 |
| 1035 | 7 | 5/6/2019 | 15:55:14 | 24.32867 | N | 54.53314 | E | 9.501465 | 32.6 |
| 1036 | 7 | 5/6/2019 | 15:55:15 | 24.32861 | N | 54.53314 | E | 9.990479 | 39.9 |
| 1037 | 7 | 5/6/2019 | 15:55:16 | 24.32852 | N | 54.53314 | E | 10.38159 | 47.8 |
| 1038 | 7 | 5/6/2019 | 15:55:17 | 24.32844 | N | 54.53315 | E | 11.08057 | 54.2 |
| 1039 | 7 | 5/6/2019 | 15:55:18 | 24.32836 | N | 54.53315 | E | 11.41504 | 59.1 |
| 1040 | 7 | 5/6/2019 | 15:55:19 | 24.32827 | N | 54.53315 | E | 11.83789 | 61.7 |
| 1041 | 7 | 5/6/2019 | 15:55:20 | 24.32821 | N | 54.53315 | E | 11.97437 | 62.4 |

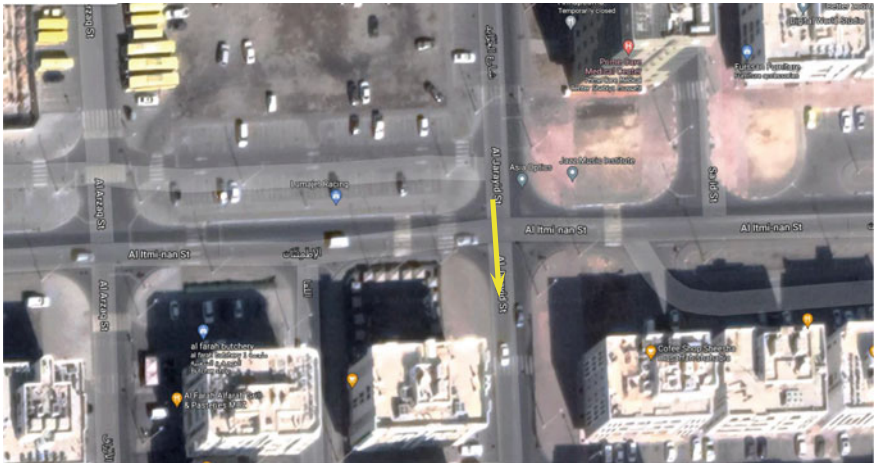


Fig. 2 Aerial image (from Google Earth ©2021 Maxar Technologies) showing the location of the acceleration profile described in Table 2 (marked by the thick light-colored line)

estimate the parameters α and β . Finally, the error term, ε , accounts for the variation in data that cannot be explained by the regression model.

4 Acceleration Models

It was found from the collected data that the acceleration rate initially increases with the increase in speed before it starts to decrease. Therefore, a data subset was extracted from the records, which includes records for the initial 4 s where the acceleration rate increases with the speed. The remainder of the data records were for acceleration after 4 s from departure. The acceleration models for crossing two-lane major roads are:

$$a = 1.1568 + 0.1285v \text{ (initial 4 s from the start of departure)} \quad (6)$$

$$a = 2.1994 - 0.0847v \text{ (after 4 s from the start of departure)} \quad (7)$$

The acceleration models for crossing four-lane major roads are:

$$a = 1.1743 + 0.1326v \text{ (initial 4 s from the start of departure)} \quad (8)$$

$$a = 2.2789 - 0.0853v \text{ (after 4 s from the start of departure)} \quad (9)$$

In Eqs. 6–9, the parameter a is the acceleration rate (m/s^2) and v is the speed (m/s). The p -values correspondent to the speed variable (v) in all the four models are less than 0.05, which indicates that the coefficients are significantly different from zero at the 95% confidence level. The calibrated models shown in Eqs. 6–9 have values of the coefficient of determination, R^2 , of 0.6743, 0.6381, 0.6579, and 0.6245, respectively.

5 Validation

There were 174 acceleration profiles reserved for validating the acceleration models. Those profiles were collected from participants in the Greater Toronto Area in Canada. It was found from the validation results that the mean squared errors were insignificant, and the root mean squared errors were also insignificant with their values close to the standard errors associated with model calibration. Based on these results, it can be concluded that the developed models are robust and stable.

6 Comparison with Current Design Guides

Figure 3 shows the distance travelled by a departing vehicle in relation to the time after departing a TWSC intersection. The distance was computed by applying Eq. 4

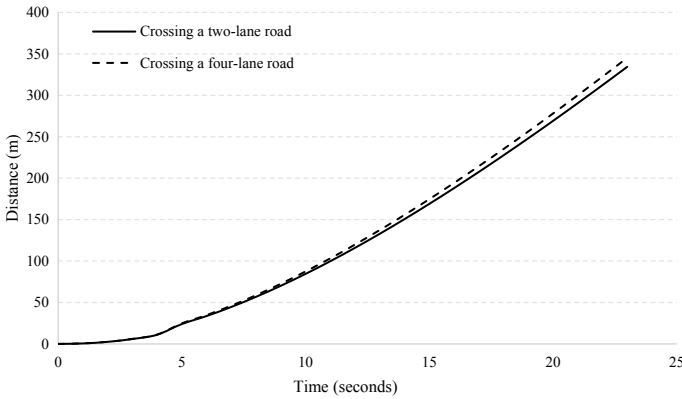


Fig. 3 Distance travelled by the departing vehicle in relation to the time spent after starting departure

with substituting the values of the coefficients α and β with those found in Eqs. 6–9. Assuming a lane width of approximately 3.75 m and a typical vehicle length of approximately 6.00 m, and assuming no median on the major road, a departing vehicle would fully clear a two-lane major road in approximately 4–5 s, or a four-lane major road in approximately 5–6 s. By assuming a minimum headway of 2 s as a safety margin between the clearance of the intersection by the departing vehicle and the arrival of the nearest approaching vehicle, the minimum safe gap, t_{app} , would be approximately 6–7 s to cross a two-lane road, or 7–8 s to cross a four-lane road. These values are mainly consistent with the suggested values found in the design guide [1] of 6.5 s (when crossing a two-lane road) or 7 s (when crossing a four-lane road).

7 Conclusions

In this paper, acceleration profiles were established for vehicles departing in a straight path at TWSC intersections. Those acceleration profiles were established using GPS data logging devices that recorded the positions and speeds of 32 vehicles operated by volunteer drivers. It was found that when departing a TWSC intersection in a straight path, the acceleration rate increases during the initial 4 s after the start of the departure; and afterward, the acceleration rate decreases. It was also found that the acceleration profile for crossing a two-lane major road is different from that for crossing a four-lane major road. Assuming no median on the major road, it was found that a typical vehicle would fully clear a TWSC intersection in approximately 4–5 s when two-lane roads or 5–6 s when crossing four-lane roads. By assuming a minimum headway of 2 s as a safety margin between the clearance of the intersection by the departing vehicle and the arrival of the nearest approaching vehicle, the minimum safe gap

would be approximately 6–7 s when crossing two-lane roads or 7–8 s when crossing four-lane roads. These values are mainly consistent with the suggested values found in the geometric design guide in the United States.

References

1. American Association of State Highway and Transportation Officials (2011) A policy on geometric design of highways and streets. American Association of State Highway and Transportation Officials (AASHTO), Washington, DC, USA
2. Dabbour E (2015) Design gap acceptance for right-turning vehicles based on vehicle acceleration capabilities. *Transp Res Rec J Transp Res Board* 2521:12–20
3. Dabbour E, Easa S (2021) Revised method for calculating departure sight distance at Two-Way Stop-Controlled (TWSC) intersections. *Transp Res Rec J Transp Res Board*. <http://doi.org/10.1177/036119812111031544>
4. Dabbour E, Easa S (2017) Sight-distance requirements for left-turning vehicles at two-way stop-controlled intersections. *J Transp Eng* 143(1). <http://doi.org/10.1061/JTEPBS.0000018>
5. Drew DR (1968) Traffic flow theory and control. McGraw-Hill Inc., New York, NY, USA
6. Long G (2000) Acceleration characteristics of starting vehicles. *Transp Res Rec J Transp Res Board* 1737:58–70
7. Perco P, Marchionna A, Falconetti N (2012) Prediction of the operating speed profile approaching and departing intersections. *J Transp Eng* 138(12):1476–1483
8. Rakha H, Snare M, Dion F (2004) Vehicle dynamics model for estimating maximum light-duty vehicle acceleration levels. *Transp Res Rec J Transp Res Board* 1883:40–49
9. Rao SK, Madugula KSM (1986) Acceleration characteristics of automobiles in the determination of sight distance at stop-controlled intersections. *Civ Eng Pract Des Eng* 5:487–498
10. Wang J, Dixon KK, Li H, Ogle J (2004) Normal acceleration behavior of passenger vehicles starting from rest at all-way stop-controlled intersections. *Transp Res Rec J Transp Res Board* 1883:158–166

Dynamic Interaction of Twin Particle Clouds in Stagnant Water



M. Janati and A. H. Azimi

1 Introduction

Particle clouds are formed when a finite mass of dense sand particles is released into a quiescent ambient, whereas the continuous disposal of sedimentary materials into a stagnant ambient forms a sand jet. The mixing of particles in ambient water has been observed in different engineering applications, including wastewater disposal, construction of artificial islands, dredging activities, and mining operations [1–3]. In such activities, the understanding of how sand particles can be placed precisely in designated areas and how sedimentary materials disperse in the ambient are important to either preserve the sand mass for bottom placement or decrease the turbidity in the aquatic environment, which can adversely affect the aquatic ecosystems. The initial condition of releasing dredge materials, the type of dredged materials, and the local conditions determine the level of environmental interferences. The understanding of particle dynamics and their controlling parameters are important for proper design, optimization, and monitoring of the dredging release processes.

Previous experimental studies have investigated the dynamic behavior of sand particles in stagnant water [4–9]. A particle cloud is formed when a finite amount of sand particles is released through a nozzle and dumped into the water. The particle cloud descends and particle evolution can be described by three phases of motion [10]. The initial acceleration phase (Phase 1) within which the particle cloud accelerates as a solid body until reaches a turbulent condition, the self-preserving or turbulent thermal phase (Phase 2) where the vortex ring is formed inside the cloud so that the cloud decelerates since the water entrains into the cloud from the edges and the

M. Janati · A. H. Azimi (✉)
Department of Civil Engineering, Lakehead University, Thunder Bay, ON, Canada
e-mail: azimi@lakeheadu.ca

M. Janati
e-mail: mjannati@lakeheadu.ca

dispersive or particle-settling phase (Phase 3) in which sand particles detached from the entrained fluid and descend individually at their settling velocities (u_∞).

A cloud number can be defined as the ratio of particle settling velocity of each particle u_∞ to the characteristic circulation velocity u_c in the particle thermal as: $Nc = u_\infty/u_c$ where $u_c = B^{1/2}/r_o$, B is buoyancy, and r_o is the bulk radius of the cloud [10]. Moghadaripour et al. [11] rearranged the cloud number equation and stated that particle cloud evolution depends on two main factors including the aspect ratio (L_o/d_o) and the Stokes number (S_t), where the aspect ratio is the ratio of the length of occupied sand in a pipe with a diameter of d_o as $L_o = 4m/\pi d_o^2 c_o \rho_s$, where c_o is the initial sand volume fraction (i.e., $c_o = 0.6$ volume/volume), ρ_s is the sand density, and Stokes number, $S_t = [(\rho_s - \rho_w)d_{50}^2 u_o]/18\mu d_o$ where d_{50} is the mean particle size and μ is the dynamic viscosity of water. Bond and Johari [12] found that the effect of L_o/d_o is very remarkable and controls the motion of particle clouds. Moghadaripour et al. [11, 13] conducted a wide range of experiments and found that the aspect ratio has a significant effect on the growth of cloud width; however, it has a small impact on the depth progression of particle clouds.

In recent years various studies were conducted regarding the behavior of particle cloud in the presence of viscous fluid [14–16]. The dynamic of particle clouds in a viscous fluid showed that the motion of particles and the number of frontal heads are correlated with the aspect ratio as well [17]. It was found that the width of particle clouds in viscous fluid increased with the aspect ratio, although it never exceeded the growth rate of particle clouds in the water. Mohammadidinani et al. [16] performed experiments on the dynamics of a sand jet passing through a layer of oil into the water. They observed that the existence of an oil layer decays the initial velocity of particle clouds. Consequently, the initial momentum of sand jets or particle clouds can be controlled with an oil layer which acts as a damper in the system.

The cloud motion characteristics such as cloud length, width, and velocity were evaluated to understand the motion of particle clouds dispersion. The particle grouping and wake formation behind particles increase the velocity of particles and these results in a bulk velocity higher than the individual settling velocity of particles. Azimi et al. [1] indicated a group settling velocity of $5u_\infty$ for sand jets front (i.e., $L_o/d_o \rightarrow \infty$) and a threshold particle cloud velocity of $1.4u_\infty$ was defined for particle clouds with a limited aspect ratio to categorize the thermal and dispersive phases [18]. The objective of the current study is to investigate the effect of releasing sand particles in form of twin particle clouds in stagnant water and compare the results with the corresponding single particle cloud.

2 Experimental Setup

Laboratory experiments were conducted in a 1.65 m (length) \times 0.85 m (width) \times 0.95 m (depth) glass-walled water tank in the Multiphase Flow Research Laboratory (MFRL) at Lakehead University. The water depth was set at 0.85 m in the tank with a temperature of $20^\circ\text{C} \pm 2^\circ\text{C}$ ($\mu = 9.77 \times 10^{-4}$ kg/m s). A schematic front

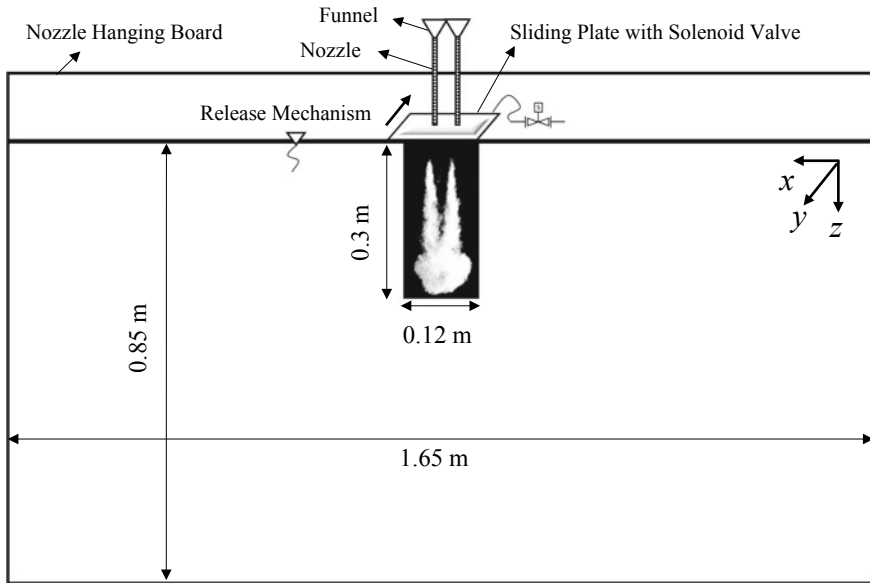


Fig. 1 A schematic front view of the experimental setup and coordinate system. The sample image is related to Test No. T11 ($L_o/d_o = 19.4$, $\varepsilon/d_o = 2.5$)

view of the presented setup is shown in Fig. 1. Moreover, Fig. 1 shows a snapshot images of twin particle cloud. Sand particles were released through two nozzles with the inside diameter of $d_o = 8$ mm (Test No. T11). Single and twin particle clouds with a relatively wide range of aspect ratios (i.e., $9.7 \leq L_o/d_o \leq 38.8$) were formed by releasing different mass of sand particles (i.e., $m = 6$ g, 12 g, 18 g, and 24 g) with a density of 2564 kg/m³ through vertical nozzles into water. It should be noted that the sand particles with a median sand size of $d_{50} = 0.507$ mm were carefully washed and dried before filling the nozzles. The releasing processes of sand particles with different aspect ratios and source separations were recorded by a high-speed camera (Photron-FASTCAM, 1024PCI-100KC, Japan) with a frequency of 30 Hz (30 frames per second). Particle clouds were illuminated with light sources (Woods L13, 1000-W Telescope Work Light, Canada) to eliminate background reflection. In order to minimize bubble formation due to air entrainment from water surface, a 5 mm distance was considered between the nozzles and water surface. In this study, the Stokes number was calculated as $S_t = 0.55$ which means that sand particles in this study are classified as hydrodynamically large particles [11].

To investigate the evolution of twin particle clouds, four tests (i.e., T1 to T4, see Table 1) were carried out with $\varepsilon/d_o = 0$, which indicates that the two nozzles are completely aligned and represents a single particle cloud. These tests are used as benchmark tests in this study. 24 experiments (i.e., T5 to T28, see Table 2) with various aspect ratios (L_o/d_o) and source separations (ε/d_o), were conducted and the

Table 1 Experimental details of the benchmark tests (single particle clouds), $\varepsilon/d_o = 0$

| Test | Mass (g) | L (cm) | L_o/d_o | N_c |
|------|---------------|----------|-----------|-------|
| T1 | 1×6 | 7.76 | 9.7 | 0.112 |
| T2 | 1×12 | 15.52 | 19.4 | 0.099 |
| T3 | 1×18 | 23.28 | 29.1 | 0.093 |
| T4 | 1×24 | 31.04 | 38.8 | 0.089 |

Table 2 Experimental details of the twin particle clouds with different source separations, $d_o = 8$ mm, $d_{50} = 0.507$ mm

| Test | ε/d_o | Mass per nozzle (g) | L (cm) | L_o/d_o | N_c |
|------|-------------------|---------------------|----------|-----------|-------|
| T5 | 2.5 | 3 | 7.76 | 9.7 | 0.112 |
| T6 | 5 | 3 | 7.76 | 9.7 | 0.112 |
| T7 | 7.5 | 3 | 7.76 | 9.7 | 0.112 |
| T8 | 10 | 3 | 7.76 | 9.7 | 0.112 |
| T9 | 12.5 | 3 | 7.76 | 9.7 | 0.112 |
| T10 | 15 | 3 | 7.76 | 9.7 | 0.112 |
| T11 | 2.5 | 6 | 15.52 | 19.4 | 0.099 |
| T12 | 5 | 6 | 15.52 | 19.4 | 0.099 |
| T13 | 7.5 | 6 | 15.52 | 19.4 | 0.099 |
| T14 | 10 | 6 | 15.52 | 19.4 | 0.099 |
| T15 | 12.5 | 6 | 15.52 | 19.4 | 0.099 |
| T16 | 15 | 6 | 15.52 | 19.4 | 0.099 |
| T17 | 2.5 | 9 | 23.28 | 29.1 | 0.093 |
| T18 | 5 | 9 | 23.28 | 29.1 | 0.093 |
| T19 | 7.5 | 9 | 23.28 | 29.1 | 0.093 |
| T20 | 10 | 9 | 23.28 | 29.1 | 0.093 |
| T21 | 12.5 | 9 | 23.28 | 29.1 | 0.093 |
| T22 | 15 | 9 | 23.28 | 29.1 | 0.093 |
| T23 | 2.5 | 12 | 31.04 | 38.8 | 0.089 |
| T24 | 5 | 12 | 31.04 | 38.8 | 0.089 |
| T25 | 7.5 | 12 | 31.04 | 38.8 | 0.089 |
| T26 | 10 | 12 | 31.04 | 38.8 | 0.089 |
| T27 | 12.5 | 12 | 31.04 | 38.8 | 0.089 |
| T28 | 15 | 12 | 31.04 | 38.8 | 0.089 |

results were compared with the benchmark tests to evaluate the performance of twin particle clouds.

Four main aspect ratios were defined in this study based on sand particles mass and nozzle diameter including $L_o/d_o = 9.7, 19.4, 29.1,$ and 38.8 . For twin particle

clouds experiments, the total mass was split into two equal portions and particles were released in water through two nozzles which were held by a nozzle board above the tank. In order to investigate the effect of source separation, the two nozzles were distanced from $\varepsilon = 2$ cm to 12 cm and the source separations were normalized with the nozzle diameter, ε/d_o (see Tables 1 and 2).

In order to analyze the recorded images, the MATLAB image processing toolbox [19] was used to detect the cloud’s edges, binarize the images, and extract the required data including cloud width, length, and frontal positions. Particle Image Velocimetry toolbox (PIVlab) from MATLAB was employed to extract the velocity data of individual particles inside the cloud along the centerline of particle clouds. The details on image preparation and image thresholding procedures can be found in [20].

3 Results

3.1 Twin Particle Cloud Evolution

Time evolution of single particle cloud and twin particle cloud with nozzle distance of $\varepsilon = 4$ cm in four different non-dimensional times (t/T), is illustrated in Fig. 2 with photo-illumination techniques. It should be noted that T is a characteristic time scale ($T = B^{0.5}/u_\infty^2$), which is $T = 1.38$ s, 1.95 s, 2.39 s and 2.76 s for $L_o/d_o = 9.7, 19.4, 29.1,$ and 38.8, respectively.

The illustrated single and twin particle clouds have the same aspect ratio of $L_o/d_o = 19.4$ so that the effect of dividing sand particles mass into two equal portions can be compared. As can be seen in Fig. 2, the most important effect of dividing the mass into two masses is that the trailing part of the cloud in the water column can be diminished quicker in twin particle clouds. As a result, the risk of turbidity and mass loss, which can cause severe pollutions in the water environment, can be reduced. Another benefit of splitting the total mass of sand particle into two parts is mixing

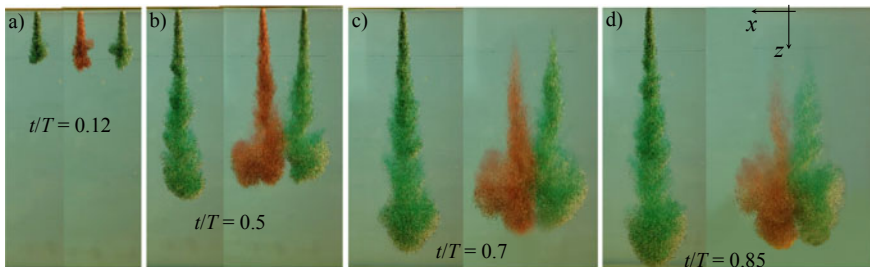


Fig. 2 Time evolution of twin and single particle clouds with the aspect ratio of $L_o/d_o = 19.4$ in different non-dimensional times. Test No. T2 ($L_o/d_o = 19.4, \varepsilon/d_o = 0$) for the single cloud and Test No. T12 for the twin cloud ($L_o/d_o = 19.4, \varepsilon/d_o = 5, T = 1.95$ s)

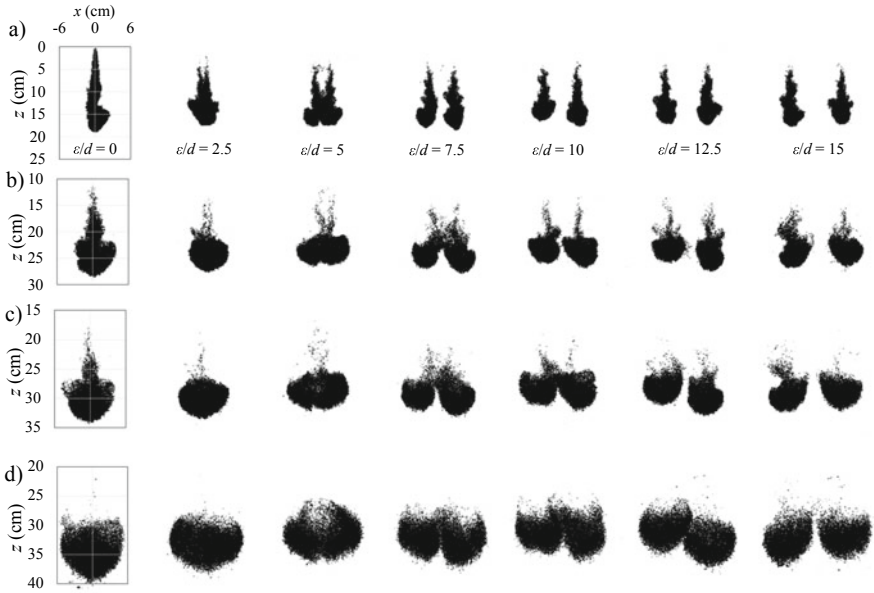


Fig. 3 Single and twin particle cloud development in a different phases of motion with $L_o/d_o = 9.7$ (1×6 g vs. 2×3 g, test No. T1 and T5 to T10): **a** initial acceleration phase $t = 0.85$ s, **b** turbulent thermal phase $t = 1.5$ s, **c** turbulent thermal phase $t = 2$ s and **d** dispersive phase $t = 2.5$ s

enhancement and dilution improvement. As shown in Fig. 2d ($t/T = 0.85$), while the single cloud is still in its first phase of cloud formation and is discharging into the water, the two adjacent clouds are fully developed, entangled with each other, and start to merge so that the dilution process initiates earlier in twin particle clouds than the corresponding single clouds. Such process is helpful in minimizing the duration of discharging process.

Figure 3 shows the evolution of twin particle clouds with $L_o/d_o = 9.7$ and $\epsilon/d_o = 0, 2.5, 5, 7.5, 10, 12.5, 15$ (i.e., test No. T1 and T5 to T10) in four different times of $t = 0.85$ s, 1.5 s, 2 s, and 2.5 s. The effect of increasing the distance between two nozzles on formation of twin particle clouds can be observed. As can be seen, the actual planar cloud areas, which were shaded by sand particles, become larger once two clouds are separated and distant from each other. Moreover, once the two clouds start to merge, the trailing part of each cloud becomes smaller after the interactions. This phenomenon can be useful in conditions where a vast area of the aquatic environment needs to be nourished with nutritional substances. However, when two clouds are released far from each other (even if they have no interactions), they still have impacts on one another’s motion. At the beginning of the turbulent thermal phase, because of the ambient entrainment between two clouds, twin particle clouds with large source separation can force the adjacent cloud to move faster or slower (see Fig. 3b, c; $\epsilon/d_o = 12.5, 15$).

Additionally, as shown in Fig. 3, the twin cloud with very close distances (i.e., $\epsilon/d_o = 2.5, 5$) evolved in a dense and highly symmetric shape that lookalikes a sphere with a sparse trailing section. However, by increasing the source separation (i.e., $\epsilon/d_o \geq 12.5, 15$), the two clouds are not fully merged even in the dispersive phase and they only have a very small attraction toward each other (see Fig. 3d). In this condition, the trailing part is wider than the twin particle clouds with smaller source separation.

3.2 Particle Cloud’s Length

In order to find out the percentage of the length reduction in twin particle clouds, the variations of cloud length in single and twin particle clouds with time are illustrated in Fig. 4. As mentioned before, the overall cloud’s length is diminished by dividing the sand mass into two parts. However, the amount of this change is not exactly equal to 50% of a single cloud’s length. As can be seen, by increasing the aspect ratio, the percentage differences between single and twin particle clouds’ lengths are the same in all aspect ratios except in $L_o/d_o = 38.8$. For example, in Fig. 4a–c with $L_o/d_o =$

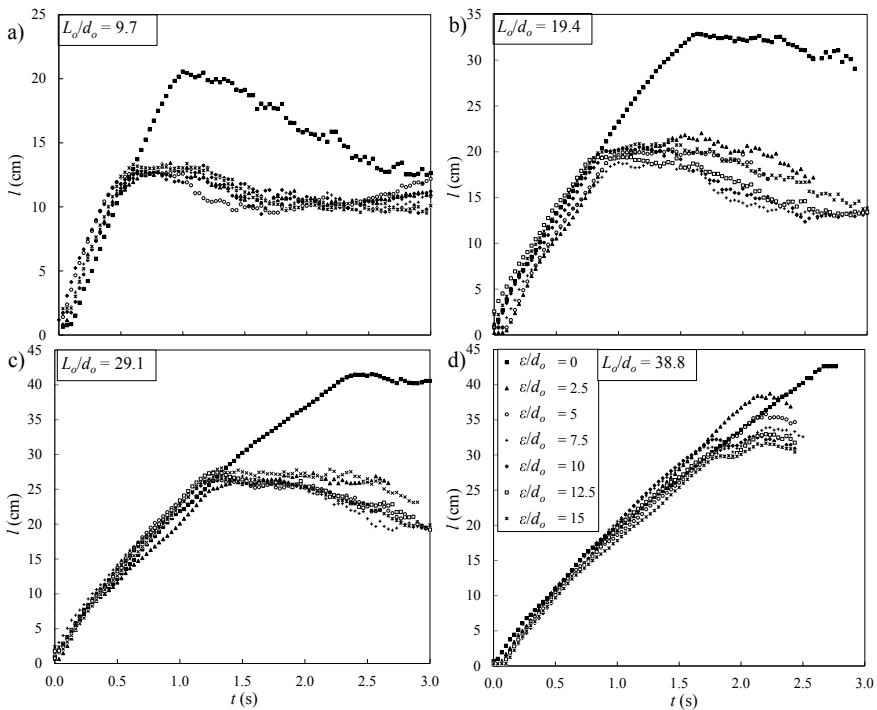


Fig. 4 Variations of single and twin particle clouds’ length for different source separations and aspect ratios; **a** $L_o/d_o = 9.7$, **b** $L_o/d_o = 19.4$, **c** $L_o/d_o = 29.1$, **d** $L_o/d_o = 38.8$. (Test No. T1 to T28)

9.7, 19.4, and 29.1, the maximum length of a single particle cloud is approximately 36% larger than the maximum length of twin particle clouds. However, in $L_o/d_o = 38.8$ due to the limitation of water depth in the tank, the single particle cloud acted as a sand jet [20]. In high aspect ratios, the particle cloud's length is not fully formed and the particles reached the bottom of the tank due to high aspect ratio compared to the tank depth. In all cases, the cloud's length started to decrease after reaching its maximum as the trailing part fell into the cloud's head. It should be mentioned that the cloud's length is considered from the frontal cloud's head to the back of the cloud's trailing part and calculated through image processing techniques in MATLAB software. Moreover, in twin particle clouds with large source separations (i.e., $\varepsilon/d_o \geq 10$), the length of the twin clouds with the same aspect ratios was not extremely affected by increasing the nozzle distances and the total differences between their lengths are approximately 10%.

3.3 Particle Cloud's Width

The cloud width was defined as the cloud's maximum horizontal extent [5]. Figure 5 shows the effects of source separation and aspect ratio on variations of cloud's width

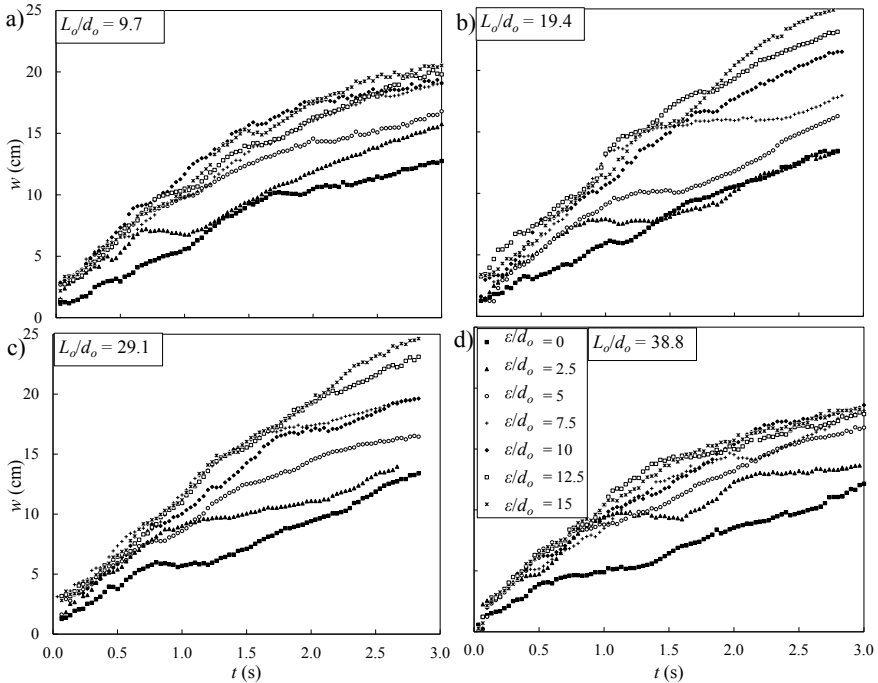


Fig. 5 Variations of twin clouds' width in different source separations for different aspect ratios; **a** $L_o/d_o = 9.7$, **b** $L_o/d_o = 19.4$, **c** $L_o/d_o = 29.1$, **d** $L_o/d_o = 38.8$ (Test No. T1 to T28)

(w) with time. The distance between the adjacent nozzles (ε) was included on calculation of the total cloud width. As can be seen in all four plots, the growth rate of a single particle cloud (i.e., $\varepsilon/d_o = 0$) is lower than other nozzle distances. However, in all other source separations with different aspect ratios, the cloud's width at the beginning of the releasing process remains the same until $t = 1$ s. After $t = 1$ s, the cloud's width begins to deviate from each other with a gradual accelerating slope for twin particle cloud with different source separations. However, for the minimum source separations (i.e., $\varepsilon/d_o = 2.5$), the twin cloud's width after $t = 1$ s started to incline toward the single cloud's width and even in small aspect ratios (i.e., $L_o/d_o = 9.7, 19.4$), their width in some stages is completely similar to each other. For example for $L_o/d_o = 19.4$, the cloud's width of a single cloud increased linearly from the start point of the releasing process, but after $t = 1$ s to $t = 1.5$ s, the twin clouds width did not change and had a constant value of approximately $w = 7.5$ cm. However, after crossing $t = 1.5$ s, both cloud's width increased with the same slope. Additionally, in other source separations cases, the cloud's width increased at a slow rate despite increasing the distance between adjacent nozzles. As it is illustrated in Fig. 5a–d, the width of the cloud in medium and large source separations (i.e., $\varepsilon/d_o \geq 7.5$) slightly increased by 15–20% as source separation increased.

3.4 Particle Cloud's Velocity

The frontal cloud's velocity was calculated based on the foremost position of the particle cloud's frontal head, which was obtained using the MATLAB programming from the recorded images in this study. Figure 6 illustrates the variations of the twin cloud's frontal speed for $L_o/d_o = 9.7$ and $\varepsilon/d_o = 2.5$ and the results were compared with the proposed equation of Moghadaripour et al. [13] for a single particle cloud with the same aspect ratio. The equation for particle clouds with a large particle size is $u_f/u_\infty = 9.9 (z/d_o(L_o/d_o)^{-0.4})^{-0.68}$ and this equation was proposed for particle clouds with an aspect ratio (L_o/d_o) between 1.5 and 19.6. As can be seen in Fig. 6, in a very close nozzles distance, (i.e., $\varepsilon/d_o = 2.5$), the results show promising compatibility with Moghadaripour's equation for $z \geq 20$ cm, when the two clouds were completely mixed and sand particles started to descend as a whole merged cloud instead of bursting out. However, from the location where two clouds were released at $z = 20$ cm, some frontal velocity fluctuations can be observed. Figure 6 shows that as soon as the two clouds are released into water, they tend to expand by the entrainment of the surrounding ambient. However, when two particle clouds are discharged at a very close distance (see Fig. 6), they have a common ambient and each cloud tries to entrain the common water volume into its domain. This may be the main reason why twin particle clouds were drawn toward each other and they have a tendency to evolve as one single particle cloud. After frontal head mearing, the frontal velocity gradually decreased until they reach the settling velocity of $u_\infty = 0.075$ m/s at the end of the tank.

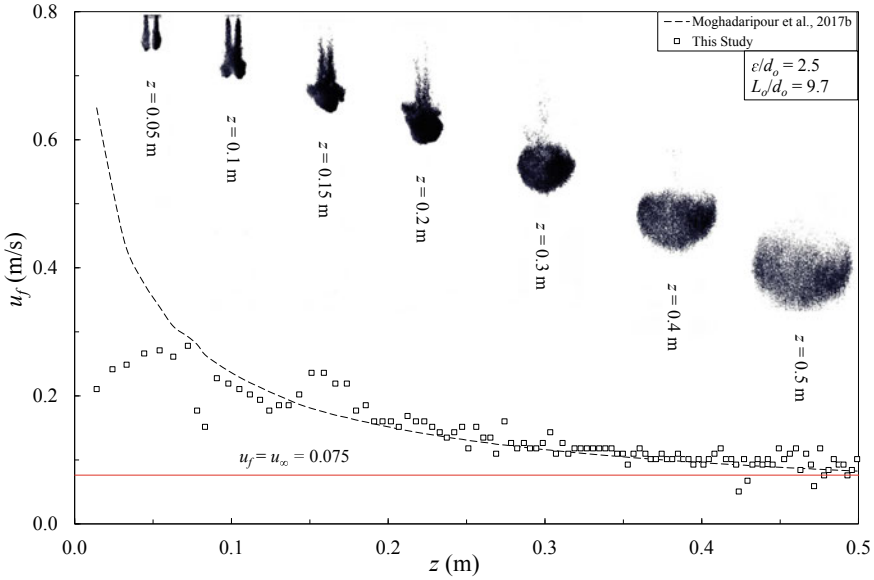


Fig. 6 Variations of cloud frontal velocity (u_f) along with the water depth (z) for the Test No. T5. The related images of the experiments are added in different positions

In order to measure the velocity of sand particles inside the cloud (u_z), the centerline velocity of the main twin cloud (Fig. 7a1) and from the centerline of left (Fig. 7b1) and right (Fig. 7c1) clouds were measured using the PIVlab toolbox (MATLAB), and the results were plotted for test No. T11 (2×6 g, $\epsilon/d_o = 2.5$, $L_o/d_o = 19.4$) in Fig. 7a–c. Three critical lines were selected and the axial velocity profiles were extracted from images. It should be noted that the PIVlab images on Fig. 7 (i.e., Fig. 7a1–c1), only show one frame of the test, however, the plots present the whole test duration (i.e., Fig. 7a–c). Figure 7a represents data on the centerline of the main twin cloud where $u_{z(\max)}$ started to increase to its maximum value of $u_{z(\max)} = 0.26$ m/s in the first 3 s of the test and after that, it quickly diminished in the second half of the test. This indicates that the plot has a sharp peak at approximately $t = 3$ s and the velocity fluctuation intervals is approximately 0.05 m/s at that time. Figure 7b, c show the maximum vertical velocity on the centerline of the left and right clouds. The peak velocity occurred at $t = 1.7$ s and the peak velocity values were similar in both cases (i.e., $u_{z(\max)} = 0.18$ m/s). As can be seen, on the centerline of both individual clouds on the left and right sides, the maximum centerline velocities increased until they reached the plateau at $u_{z(\max)} = 0.18$ m/s and remained at this stage for approximately three seconds. The velocity fluctuations were less than 0.007 m/s in the plateau. Velocity fluctuations decreased until the end of the test. Overall, the maximum velocity on the centerline of the main cloud (Fig. 7a) was 30% more than the velocities on the centerline of the single particle clouds (Fig. 7b, c). Consequently, once the velocities of the cloud in the middle of the main cloud became more than

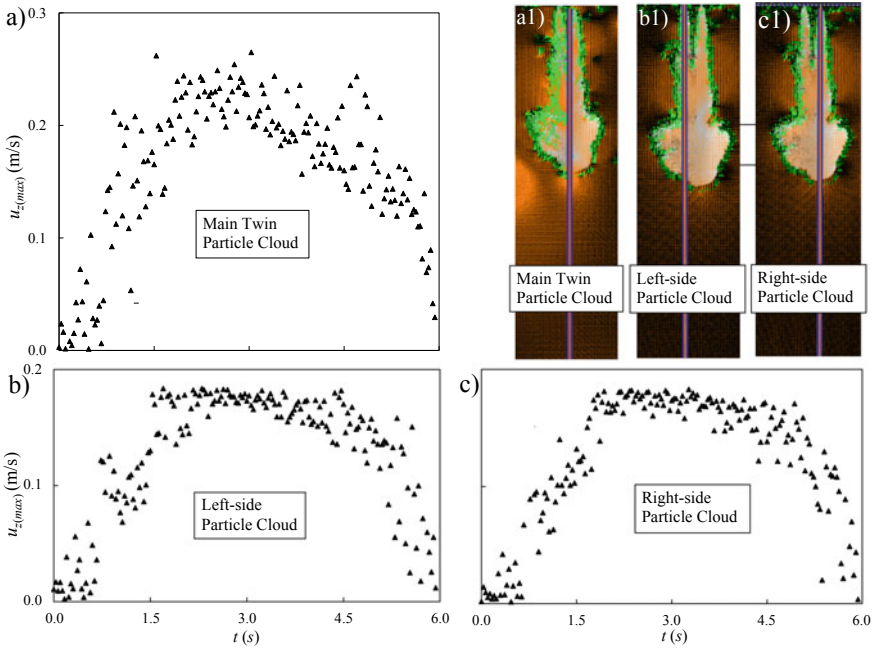


Fig. 7 Variations of cloud centerline velocity in a different part of twin particle clouds for the test No. T11 (2×6 g, $\varepsilon/d_o = 2.5$, $L_o/d_o = 19.4$). **a** Centerline velocity of the twin cloud. **b** Variations of velocity in the centerline of the left-side cloud. **c** Variations of velocity in the centerline of the right-side cloud

the velocities in the side clouds, the main cloud had a great tendency to develop from the frontal middle point which is why the particle clouds develop a mushroom-shape frontal head. Moreover, the reason why the velocity of individual particles inside the clouds remained the same for 3 s in the middle of the test was that they have smaller velocities therefore less interactions occurred in comparison with the main cloud's centerline. Therefore, it took more time for them to dissipate their kinetic energies in comparison to the case in which the cloud has 30% greater velocities.

4 Conclusion

The present study shows the results of laboratory investigation on the behavior twin sand particles in quiescent water. The main physical characteristics of twin particle clouds such as length, width, and velocity of the clouds were obtained through advanced image processing techniques and the effects of source separations and aspect ratio were studied. The results were also compared with the corresponding single particle clouds ($\varepsilon/d_o = 0$). It was found that, although the mass of sand particles

in twin particle clouds are divided into two parts and each of which contains 50% of the total mass, their lengths were 64% of the length of the corresponding single particle clouds. This outcome can be beneficial for the aquatic environments since the remaining sand particles in water column can cause more environmental pollutions. However, this is not applicable for particle clouds with large aspect ratios (i.e., $L_o/d_o = 38.8$), since the particle cloud acted as a sand jet due to the limited depth of the tank.

Moreover, twin particle clouds with source separation of $\varepsilon/d_o = 2.5$ have the smallest cloud width amongst other source separations of twin clouds. Moreover, in small aspect ratios (i.e., $L_o/d_o = 9.7, 19.4$) after passing one second from the beginning of the test, the twin clouds' width becomes as small as the cloud width of single particle cloud. For source separations greater than $7.5d_o$, increasing source separations by 50% make the clouds' width larger by approximately 15–20%, which may be useful in particular dredging activities, where a designated area is defined to dump dredged materials. For $\varepsilon/d_o = 2.5$, the cloud's frontal velocity reached the maximum amongst other source separations and this twin particle cloud had the smallest width amongst the twin clouds as well.

The last part of this study focused on the particle cloud velocity at both macro and micro levels. At the macro-level, the frontal velocity (u_f) of the whole cloud was investigated and at the micro-level, the inside clouds' maximum vertical velocities (u_z) were extracted with the PIVlab toolbox. It was found that the frontal velocity of twin particle clouds was compatible with the study of Moghadaripour et al. [13]. However, at the beginning of the test, the frontal velocities had fluctuations until they reached a certain position (i.e., for $\varepsilon/d_o = 2.5, L_o/d_o = 9.7; z = 20$ cm), where the two clouds were completely merged. Moreover, the velocity of particles on the centerlines of the individual clouds reached at most 70% of the ones on the centerline of the main merged cloud, which showed that there were severe particle interactions and collisions exactly on the centerline of the main cloud than the ones on the centerline of the side clouds. In addition, particle velocities inside the individual clouds were uniform with less fluctuation than the ones in the middle of the main cloud.

Acknowledgements The work presented here was supported by NSERC-Discovery grant No. 421785 and the Ontario Trillium Scholarship.

References

1. Azimi AH, Zhu DZ, Rajaratnam N (2012) Experimental study of sand jet front in water. *Int J Multiph Flow* 40:19–37
2. Azimi AH, Zhu DZ, Rajaratnam N (2012) Computational investigation of vertical slurry jets in water. *Int J Multiphase Flow* 47:94–114
3. Azimi AH, Qian Y, Zhu DZ, Rajaratnam N (2015) An experimental study of circular sand–water wall jets. *Int J Multiphase Flow* 74:34–44
4. Azimi AH, Zhu DZ, Rajaratnam N (2011) Effect of particle size on the characteristics of sand jet in water. *J Eng Mech* 137(12):822–834

5. Bush JWM, Thurber BA, Blanchette F (2003) Particle clouds in homogeneous and stratified environments. *J Fluid Mech* 489(489):29–54
6. Lai ACH, Wang RQ, Law AWK, Adams EE (2016) Modeling and experiments of poly-disperse particle clouds. *Environ Fluid Mech* 16(4):875–898
7. Lai ACH, Chan SN, Law AW, Adams EE (2016) Spreading hypothesis of a particle plume. *J Hydraul Eng* 142(12):1–12
8. Sharif F, Azimi AH (2020) Particle cloud dynamics in stagnant water. *Int J Multiph Flow* 125:103197
9. Zhao B, Law AWK, Adams EE, Er JW (2014) Formation of particle clouds. *J Fluid Mech* 746(1):193–213
10. Rahimpour H, Wilkinson D (1992) Dynamic behavior of particle clouds. In: Eleventh Australasian fluid mechanics conference, vols 1 and 2. University of Tasmania, Hobart, Australia, pp 743–746
11. Moghadaripour M, Azimi AH, Elyasi S (2017) Experimental study of oblique particle clouds in water. *Int J Multiphase Flow* 91:193–213
12. Bond D, Johari H (2005) Effects of initial geometry on the development of thermals. *Exp Fluids* 39(3):589–599
13. Moghadaripour M, Azimi AH, Elyasi S (2017) Experimental study of particle clouds in stagnant water. *J Eng Mech* 143(9):04017082-1-17
14. Manzouri M, Azimi AH (2019) A study of mound formation by discharging sand particles through oblique pipes in stagnant water. *Int J Sedim Res* 34(6):P564–576
15. Manzouri M, Azimi AH (2019) Effects on oily sand jet evolution from impact momentum and channelization of particles through an immiscible interface. *Int J Multiph Flow* 121:103124 (18 p)
16. Mohammadidinani N, Azimi AH, Elyasi S (2017) Experimental investigation of sand jets passing through immiscible fluids. *ASME, J Fluids Eng* 139(5):051303, 13p
17. Azimi AH (2019) Experimental investigation on the motion of particle cloud in viscous fluids. *J Fluids Eng* 141(3):031202(1–14)
18. Buhler J, Papantoniou DA (2001) On the motion of suspension thermals and particle swarms. *J Hydraulic Res* 39(6):643–653
19. MATLAB software (Computer software, Math Works, Inc., 2018), Natick, MA
20. Janati M, Azimi AH (2021) Mixing of twin particle clouds in stagnant water. *J Eng Mech* 147(5):04021022

Experimental Study of Sand-Water Coaxial Jets with Low Velocity Ratio



F. Sharif and A. H. Azimi

1 Introduction

Particle-laden turbulent jets are consist of a central main carrier phase and a discrete phase. They have a wide range of engineering applications such as in wastewater disposals and mixing processes [7, 8, 10, 13, 26, 27, 43]. Due to the particle inertia and the distinct behavior of the particles, the flow properties are more complex in comparison to single-phase turbulent jets [1, 23, 37]. Coaxial jets are formed when an annular fluid stream is added outside of a central jet flow to enhance the mixing capabilities of different outlet streams. Coaxial jets have been used in combustion chambers, industrial burners, air blast atomizers [10, 40].

Many research studies have analyzed the dynamics and the motion of particles in two-phase turbulent jets [2–5, 9, 11, 17, 19, 26, 28, 31]. The calculation of the particle Stokes number is important to study the dispersion and the response time of particles to the carrier fluid in turbulent jets. The Stokes number can be defined as the ratio of the particle momentum response time, $\tau_p = \rho_s D_{50}^2 / 18\mu$, and the time scale of the flow, $\tau_f = d_o / u_o$. where ρ_s is the density of solid particles, D_{50} is the mean particle size, d_o is the nozzle diameter, μ is the dynamic viscosity of the ambient fluid, and u_o is the jet initial velocity. Particles with Stokes number less than one act as flow tracers, if the Stokes number is close to unity particles are propelled off the vortex core and respond to the turbulent eddies, and for Stokes number greater than one, particles adopt limited properties of the carrier fluid and they are not responsive to the velocity fluctuations [15]. The Stokes number can affect mixing parameters of

F. Sharif · A. H. Azimi (✉)
Department of Civil Engineering, Lakehead University, Thunder Bay, ON, Canada
e-mail: azimi@lakeheadu.ca

F. Sharif
e-mail: fsharif@lakeheadu.ca

particle-laden jets such as entrainment, particle collision, spreading rate, and depth progression [3, 13, 27, 30, 34].

Many researchers have examined the dynamics and mixing capabilities of single-phase coaxial jets as a function of the velocity ratio of annular to central streams, Ru [10, 16, 47, 49]. Ko and Kwan [22] characterized the flow field properties in single coaxial jets into three different regions: the initial merging zone, the intermediate merging zone, and the fully developed zone. Champagne and Wygnanski [14] studied the effect of the velocity ratio on the length of the potential core in which the flow properties are similar to the initial nozzle condition. They found also a self-similar property of the jet flow in the downstream section due to entrainment. Dahm et al. [16] indicated that the flow fluctuation in the inner and outer shear layers are strongly paired together. Dinesh et al. [38] indicated that the vortex shedding led to velocity fluctuation in turbulent coaxial jets.

Introducing solid phase to the single-phase coaxial jets can modify the dynamical and mixing capabilities of solid–liquid jets [18, 21, 29, 32, 39]. Fan et al. [17] found that in poly-disperse particle coaxial jets the particle concentration had a self-similar behavior in the zone of the flow establishment. Furthermore, lighter particles are radially dispersed in comparison to heavier particles. Pedel et al. [36] reported an asymmetric pattern in initial flow zone of particle-laden coaxial jets. Mostafa et al. [32] found that the presence of particles attenuated the turbulence intensity and led to a higher energy reduction rate. The experimental results of [39] correlated the decay rate of the centerline velocity and particle dispersion as a function of particle-loading and velocity ratio respectively. A great reduction of turbulence intensity was indicated by Kannaiyan and Sadr [21] along the initial flow zone of two-phase coaxial jets.

Several statistical methods can be used to capture the influence of particles on the mixing and inter-scale coherent structures in turbulent flow fields. The most widely used statistical method to resolve turbulent data is the Proper Orthogonal Decomposition (POD) technique [24]. The POD technique decomposes the time-history data into its lower-ranked special components modes and the frequency-ranked dynamic modes. The Proper Orthogonal Decomposition (POD) technique has been extensively used to identify the most persistent structures in the flow fields [12]. This method was first proposed by Pearson [35] for the superlative description of multi-component data sequences. Most recently, Schmidt et al. [42] proposed an empirical method to extract coherent structures or modes in both space and time known as Spectral Proper Orthogonal Decomposition (SPOD) based on the POD method. Since coherent motions could occur at different scales of special, frequencies, and energy levels, the traditional phase averaging methods (i.e., POD and DMD) are not capable of obtaining coherent structures in both space-frequency. Therefore, SPOD can extract the energy content and frequency by combining POD and DMD capabilities. Schmidt et al. [42] indicated that the temporal coefficients of the orthogonal modes of turbulent jets can be characterized by Kelvin–Helmholtz (KH) type wave packets. Kadu et al. [20] employed the SPOD method to annotate the physical structures or modes in a coaxial jet. They found that the pairs of counter-rotating vortical structures can be identified by the first two modes in the inner mixing layer. Zhang

et al. [51] indicated that the different detected orthogonal modes and their frequencies can be connected to the coherent structures in the near-field region of turbulent jets.

The mentioned studies in the literature analyzed the characteristic of pre-mixed particle-laden coaxial jets with high impinging velocity ratios [18, 21, 32, 36, 39]. A series of laboratory experiments on sand-water coaxial jets were performed by a new coaxial jet set-up consists of two concentric nozzles to generate a central water jet combined with an outer annular sand jet. As the first objective of this research the axial and radial sand velocity and concentration profiles were examined and were characterized for different low-velocity ratios. The secondary objective of this paper is to study the low-ranked coherent flow configuration of the coaxial sand-water jet using the Spectral Proper Orthogonal Decomposition technique. The SPOD coefficients provide insight into the time scale associated with the spatial modes.

2 Experimental Setup

Laboratory experiments were carried out in the Multiphase Flow Research Laboratory (MFRL) at Lakehead University to study the dynamical and flow properties of coaxial sand-water jets with low-velocity ratios. Circular coaxial jet nozzles were used in the experiment with a core and a concentric annular nozzle with the diameters of $d_o = 6$ mm and $d_a = 12$ mm, respectively, and an annular gap with a thickness of 2.5 mm. The central jet was connected to the tap water through an accurate flow meter (LZT M-15, UXCEL, China) and it was issued from the inner nozzle. In addition, the sand particles were released through the surrounded annular nozzle into a rectangular tank of 1.65 m long, 0.85 m wide, and 0.85 m deep filled with tap water and the ambient temperature of $20\text{ }^\circ\text{C} \pm 1\text{ }^\circ\text{C}$. Sand particles with a median diameter of $D_{50} = 0.386$ mm, the density of $\rho_s = 2540\text{ kg/m}^3$, Stokes number of 0.7, initial concentration of 60% by volume, and initial mass flux of $\dot{m}_o = 10\text{ g/s}$ were implemented. The initial flow Reynolds of 6622 to 11,919 were chosen to provide velocity ratios in the range of $0 \leq Ru \leq 0.33$. Overall, four experiments were tested in this study to examine particle dynamical properties of sand-water coaxial jets. Experimental details and non-dimensional parameters are listed in Table 1.

Table 1 Initial experimental details of sand-water coaxial jets

| No. | Test No. | D_{50} (mm) | d_o (mm) | d_a (mm) | \dot{m}_o (g/s) | u_{wo} (m/s) | Re | Ru | S_t |
|-----|----------|---------------|------------|------------|-------------------|----------------|--------|------|-------|
| 1 | C1 | 0.386 | 6 | 12 | 10 | 0 | 0 | 0 | 0.7 |
| 2 | C2 | 0.386 | 6 | 12 | 10 | 0.75 | 6622 | 0.33 | 0.7 |
| 3 | C3 | 0.386 | 6 | 12 | 10 | 1.16 | 9270 | 0.21 | 0.7 |
| 4 | C4 | 0.386 | 6 | 12 | 10 | 1.49 | 11,919 | 0.16 | 0.7 |

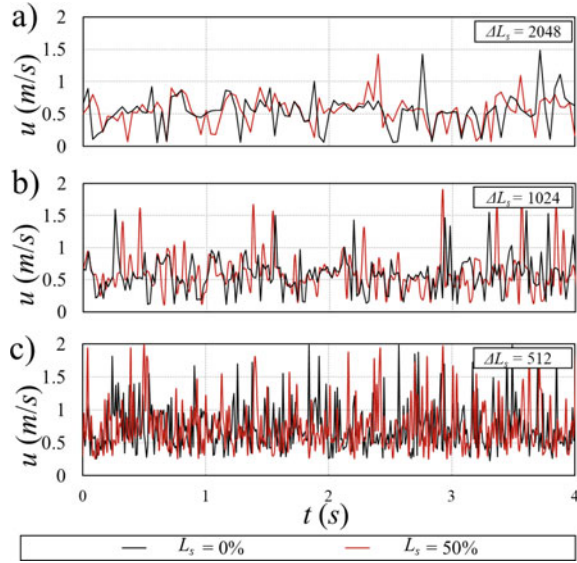
The axial and radial distributions of sand concentration and velocity were measured by an advanced optical fiber probe (PV6, Institute of Processing Engineering, Chinese Academy of Science, China). The probe is capable of measuring simultaneous time series of sand concentration and velocity. The probe tip has a diameter of 4 mm with two mutual 1 mm diameter light sources and light-refracting optical sensors. The distance between two fiber optics is 2.14 mm and transmits raw analog signals through a PV6 processing unit to a high-resolution data acquisition board (National Instrument, Austin, U.S.). The optical probe has indicated high performance in sand concentration and velocity and measurements in many experimental studies such as solid-particle turbulent jets and solid fluidization systems [6, 19, 43, 44, 46]. To investigate the flow properties such as mixing and periodic motions higher frame rate snapshots were recorded using a high-speed camera (Photron-FASTCAM, 1024PCI-100KC) with a resolution of 1024×1024 pixels. The camera was equipped with a 15–55 mm AF-S Nikkor, 1:3.5–5.6 GII (Nikon, Japan) lens to capture images. Raw images were captured with a frame rate of 250–500 frames per second and a shutter speed of 0.004 s.

3 Signal Processing

Filtering and noise removal techniques are commonly used in signal processing to filter out and enhance the oscillatory properties of raw signals. In this experiment, the eight-level and soft thresholding of the Wavelet noise reduction toolbox of MATLAB [25] was employed to eliminate noises from captured voltage signals. The wavelet denoising technique can analyze time and frequency domain at the same time and it already was tested in turbulent multiphase flows [6, 43, 48, 50]. The MATLAB signal processing toolbox was utilized to carry out the cross-correlation technique and analyzing different segment sizes of voltage signals to extract particles' instantaneous velocities. The effects of segment sizes, ΔL_s (i.e., $\Delta L_s = 2^9 = 512$, $2^{10} = 1024$ and $2^{11} = 2048$) and segment overlap of 50% [48] on instantaneous velocity data was studied and the results are shown in Fig. 1. The segment sizes correspond to sampling frequencies of 120 Hz, 60 Hz, and 30 Hz. As depicted in Fig. 1 the cross-correlations of voltage signals with a lower segment size have led to the high-velocity fluctuations. It was found that the data uncertainty associated with 2048, 1024, and 512 were $\pm 4\%$, $\pm 9\%$, and $\pm 13\%$, respectively. It was also found that the maximum data uncertainty of velocity between 0 and 50% overlap was $\pm 1.5\%$. Therefore, a segment length of 2048 with 0% overlap (i.e., no overlap) was selected for all data in this study.

Each time series of voltage signals recorded by the probe's tips can be attributed to the sand concentration of measured points through a calibration curve [19]. The acquired velocity and concentration data accuracy were examined by the integrated mass flux (i.e., the mass of a substance which passes per unit of time) at different cross-sections along with the sand-water coaxial jet as:

Fig. 1 The effect of data segmentation on instantaneous sand velocity data with $L_s = 0\%$ (i.e., no overlap), and 50% and segment lengths ΔL_s for $Ru = 0.16$, $x/d_0 = 20$. **a** $\Delta L_s = 2048$; **b** $\Delta L_s = 1024$; **c** $\Delta L_s = 512$



$$\dot{m} = \frac{dm}{dt} = 2\pi\rho_s \int_0^r c_o u r dr \quad (1)$$

where u is the particle velocity and r is the horizontal distance from the jet axis. The integration results demonstrated the mass conservation of above 90%.

4 Spectral Proper Orthogonal Decomposition (SPOD)

The Spectral POD method is conceptually and algorithmically (i.e., eigenvectors of a matrix) similar to the POD method. The main algorithm of SPOD is based on Welch's decomposition method [42, 41]. Welch [48] proposed an averaging technique to the converting time history data set from the time domain to the frequency domain known as the Power Spectrum Density (PSD). Therefore, temporal coefficients of the orthogonal modes can be explained as an increasing range of frequency spectrum. The initial step to perform the SPOD method is by adjusting a time series of snapshots to construct a single data matrix P using N_p snapshots and M_p pixels. The matrix P first needs to be segmented into N_b blocks that overlap by N_o snapshots. The size of the N_b block can be obtained as $N_b = (N_p - N_o)/(N_f - N_o)$ where N_f is the length of segmentation (i.e., 64, 128, 256, ...) corresponding to an overlap of 50%. However, by choosing larger block sizes and overlap them, the variance of data is increased accordingly. The Welch periodogram method was employed to construct an ensemble of realizations of the temporal Fourier transform of the data from a

single time series consisting of N_p snapshots by breaking it into N_b then, reordering it by frequency in each block. The POD of an ensemble of frequency-based blocks P_B leads to the extraction of the spatial modes such that any image of each block can be expressed as a linear weighted sum of the modes and can be described by Sirovich [45]:

$$P_{B_i}(x, t) = \sum_{j=1}^{r_p} \alpha_{ij} \phi_j(x) \quad (2)$$

where x and t are the independent variables denoting space and time, respectively, $i = 1$ to N_b , $j = 1$ to $r_p < N_b$. Besides, the coefficient $\alpha_{ij}(t)$ can be defined as:

$$\alpha_{ij}(t) = \beta_{ij}(t) \lambda_j^{0.5} \quad (3)$$

where β_{ij} is the temporal coefficients and λ_j is called the eigenvalue. The orthogonal eigenvectors or spatial modes, $\phi_j(x)$, represent the ensemble-averaged spatial features of the whole ensemble. The eigenvalues are arranged in decreasing order ($\lambda_1 > \lambda_2 > \lambda_3 \dots$) yields the SPOD of each block. Therefore, SPOD modes are the eigenvectors of a cross-spectral density tensor at each frequency corresponding to the modal kinetic energy.

5 Results

Figure 2a shows the effect of velocity ratio (i.e., $0.16 \leq Ru \leq 0.33$) on the variations of the normalized axial mean centerline sand concentration of coaxial sand-water jets along with the normalized distance. The uncertainty of sand phase concentration measurements was shown by the overbars indicators. The data uncertainty of concentration measurement was found to be $\pm 9\%$. The axial sand concentration of slurry jet and particle cloud were also added Fig. 2a [19, 43]. As can be seen, all given velocity ratios had a similar non-linear dissipation rate for $x/d_o \leq 16$, and the results are independent of Ru . Such correlations can be attributed to the formation of the intermediate mixing zone. However, for $x/d_o > 16$, the sand water coaxial jets with $Ru = 0.33$ had a smaller decay rate in comparison with $Ru = 0.16$ and 0.21 which decay at a faster rate close to the slurry jet model. For all cases, sand concentration in coaxial sand-water jets dissipated lower than particle clouds indicating the dissipation rate as a function of the velocity ratio further downstream of jets. The potential core (i.e., $c_m/c_o \approx 1$) of the sand-water coaxial jet was found to be five times the nozzle diameter which is $5.8d_o$ for slurry jets [19]. The axial normalized sand velocity with different velocity ratios (i.e., $0.16 \leq Ru \leq 0.33$) are depicted in Fig. 2b. The sand centerline velocity models for particle cloud [43] and water jet [23] were also added in Fig. 2b. The maximum uncertainty of sand velocity measurements was found to be approximately $\pm 11\%$. As shown in Fig. 2b the axial velocities decay rates

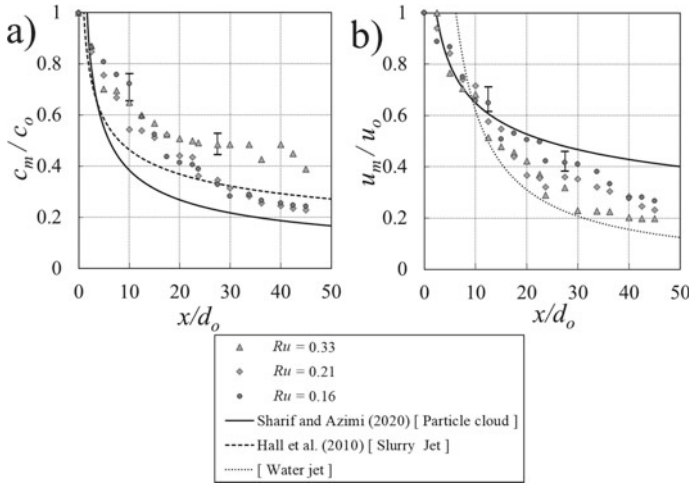


Fig. 2 **a** The variation normalized axial sand concentration, c_m/c_o , with normalized axial distance from the nozzle, x/d_o with different velocity ratios; **b** The variation normalized axial sand velocity, u_m/u_o , with normalized axial distance from the nozzle, x/d_o with different velocity ratios

are independent velocity ratio for $x/d_o \leq 16$. However, for $x/d_o > 16$ decreasing the velocity ratio (i.e., $Ru = 0.16$) caused a higher decay rate similar to the water jet and adopting the water phase flow velocity. However, coaxial jets with the velocity ratio of 0.33 and 0.21 had higher dissipation in comparison with particle clouds and the dissipation rate was smaller than slurry jets. Figure 3 shows the transverse profiles of the normalized sand concentration, c/c_m , with the normalized transverse distance from the nozzle, r/d_o , for sand-water coaxial jets with the velocity ratio of $Ru = 0.16$ and 0.21. As can be seen, all velocity ratios and the radial concentration exhibited a

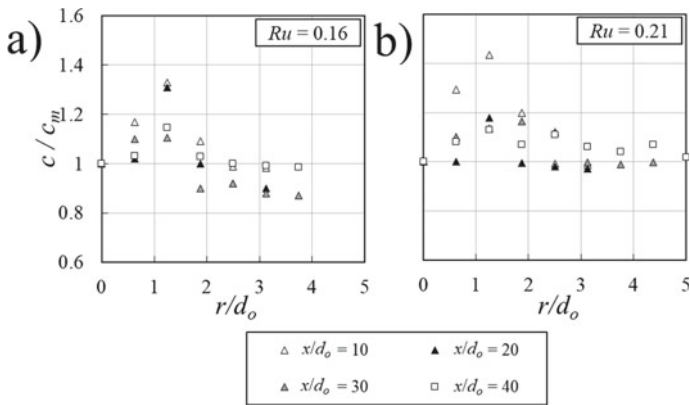


Fig. 3 The normalized radial distribution of sand particles concentration for $10 \leq x/d_o \leq 40$: **a** $Ru = 0.16$; **b** $Ru = 0.21$

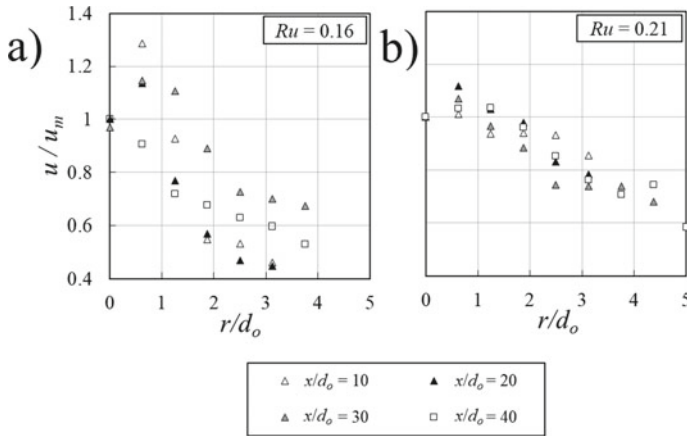


Fig. 4 The normalized radial distribution of sand particles velocity for $10 \leq x/d_o \leq 40$: **a** $Ru = 0.16$; **b** $Ru = 0.21$

near-field hump at $r/d_o = 1.25$ indicating the migration and mixing of particles jet towards the jet's axis and reduction in the intensity of hump further downstream in the flow development zone. It was found that by increasing the velocity ratio particle distribution in the radial direction was increased (see Fig. 3b).

Figure 4 shows the vertical velocities of solid particles in different normalized radial sections. As can be seen in Fig. 4a, the hump similar to the radial distribution of sand concentration exists at $r/d_o = 1.25$ and for $x/d_o \leq 30$. The normalized distance of $x/d_o \leq 30$ corresponds to the flow development zone. The intensity of the peak axial velocity decreased by decreasing the velocity ratio indicating a faster particle-flow mixing velocity ratio decreased (see Fig. 4b).

The high-speed images of sand water-coaxial jets with various velocity ratios (i.e., $0 < Ru < 0.33$) are shown in Fig. 5. As shown in Fig. 5a, b particle and flow oscillation was high for $Ru = 0$ and 0.33. However, by increasing the initial water phase Reynolds number (leading to smaller velocity ratios) more uniform flow and particle distribution were observed (see Fig. 5c, d).

The SPOD analysis was performed for the series of 1500 high-speed snapshots to investigate the low-rank dynamic of particles and flow properties. The block size was chosen as 256 data with a 50% overlap. The first SPOD mode visualization for the frequencies of 0, 3.91, and 7.81 Hz for the velocity ratio of 0, 0.16, and 0.33 are shown in Fig. 6. The Kelvin–Helmholtz (KH) type wave packs can be distinguished in all exhibited velocity ratios. The large size in wave packs for the velocity ratio of 0 and 0.33 with the frequency of 0 and 3.91 Hz can be distinguished at the downstream section for the frequency of zero (see Fig. 6a, d). Such wave packs indicate particle oscillation and shedding which was already shown in instantaneous high-speed images (see Fig. 5a, b). However, by increasing the Reynolds number of water and decreasing the velocity ratio (i.e., $Ru = 0.16$) wave packs turned to an axially elongated shape. Comparing the modal shape represented in Fig. 6g with the

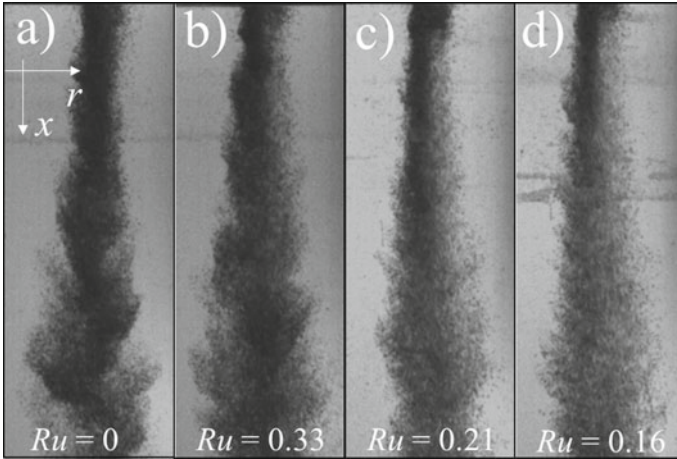


Fig. 5 The instantaneous high speed images of sand-water coaxial jets with different velocity ratios: **a** $Ru = 0$; **b** $Ru = 0.33$; **c** $Ru = 0.21$; **d** $Ru = 0.16$

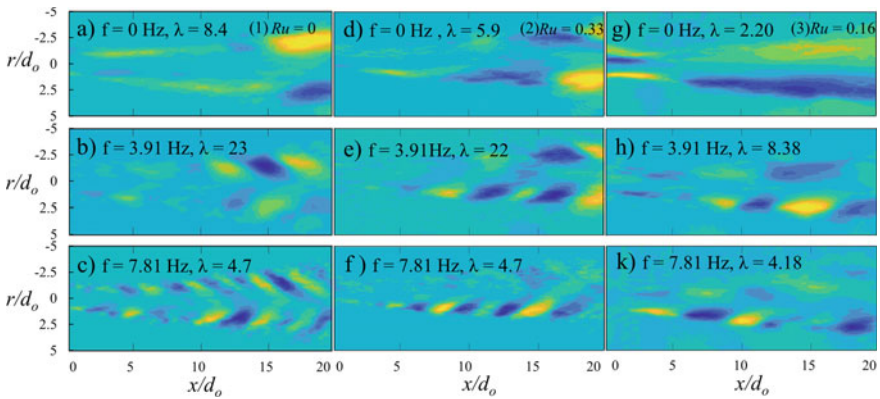


Fig. 6 Visualization of the SPOD mode number one for frequencies of 0, 3.91 and 7.81 Hz for the velocity ratios of $Ru = 0, 0.33$ and 0.16

high seed images demonstrated an almost uniform flow and particle pattern along the jet axis (see Fig. 5c, d). In all three cases of Ru for the frequencies of 3.91 and 7.81 Hz, the a-wave train dominant mode can be observed and localized further downstream.

Besides, the SPOD energy spectrum or eigenvalues, λ , the turbulent kinetic energy decreased by a reduction in the velocity ratio from 0.33 to 0.16. The SPOD energy spectrum, λ , for the first seven modes with the velocity ratio of 0.16 and 0.33 are displayed in Fig. 7. Large separation between the eigenvalues is a good indicator of dominant mode and energy contribution. As shown in Fig. 7a the first two modes had a larger separation due to higher energy levels. The separation between eigen spectra of the modes implies that vortex shedding dominated the dynamics of the

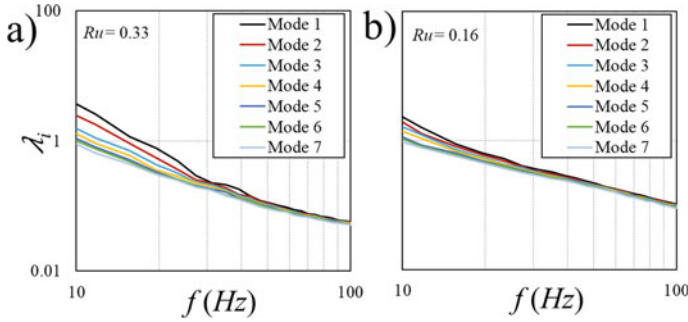


Fig. 7 The SPOD eigenvalue spectra of the first seven modes: **a** $Ru = 0.33$; **b** $Ru = 0.16$

connected modes [33]. However, by decreasing the velocity ratio, the separation distance decreased accordingly due to less vortex shedding intensity (see Fig. 7b). It was found that most of the turbulence kinetic energy was stored in the first ten modes especially for smaller velocity ratios (i.e., $Ru = 0.16$) in sand-water coaxial jets.

6 Conclusion

A series of laboratory experiments were carried out to investigate the effects of the velocity ratios smaller than 0.5 on the dynamics of sand-water coaxial jets in stagnant water. A light refractive optical probe was implemented to measure sand concentration and velocity of sand-water coaxial jets in axial and transverse directions. The probe accuracy and data measurements consistency were examined by conservation of mass in different cross-sections with an uncertainty level of 10%. The initial mixing zone was indicated around 5 times the nozzle diameter and the zone of intermediate flow establishment was found to occur at $x/d_o \leq 16$. The axial alternation of velocity and concentration were found to be self-similar in the intermediate zone of flow establishment. However, by decreasing the velocity ratio in the zone of flow development (i.e., $x/d_o > 16$), the decay rate increased accordingly and became similar to single-phase water jets. The concentration decay rate was found to be smaller than the decay rates of particle clouds and slurry jets, particularly for the high-velocity ratio ($Ru = 0.33$). The radial distributions of concentration and velocity profiles were also investigated. A near-field hump was observed at $r/d_o = 1.25$ for all cases in concentration and velocity profiles due to migration and mixing of particles with the central water jet. The presence of a hump indicated that a significant portion of the turbulence kinetic energy and flow vortex shedding is stored in low-frequency modes of coaxial sand-water jets with a higher velocity ratio (i.e., $Ru = 0.33$). Furthermore, for velocity ratios smaller than 0.16, most of the energy distribution occurred in the first ten modes. The Kelvin–Helmholtz (KH) type wave

packets were observed in low frequencies and the first ten modes associated with the annular nozzle.

References

1. Abramovich GN (1963) The theory of turbulent jets. MIT Press, 671 p
2. Azimi AH, Zhu DZ, Rajaratnam N (2011) Effect of particle size on the characteristics of sand jet in water. *J Eng Mech* 137:1–13
3. Azimi AH, Zhu DZ, Rajaratnam N (2012) Experimental study of sand jet front in water. *Int J Multiph Flow* 40:19–37
4. Azimi AH, Zhu DZ, Rajaratnam N (2012) Computational investigation of vertical slurry jets in water. *Int J Multiph Flow* 47:94–114
5. Azimi AH, Zhu DZ, Rajaratnam N (2014) An experimental study of sand deposition from slurry wall jets. *J Eng Mech* 140:296–314
6. Azimi AH, Zhu DZ, Rajaratnam N (2015) An experimental study of circular sand–water wall jets. *Int J Multiph Flow* 74:34–44
7. Azimi AH, Zhu DZ, Rajaratnam N (2012a) Experimental study of sand jet front in water. *Int J Multiph Flow* 40:19–37
8. Azimi AH, Zhu DZ, Rajaratnam N (2012b) Computational investigation of vertical slurry jets in water. *Int J Multiph Flow* 47:94–114
9. Azimi AH (2019) Experimental investigation on the motion of particle cloud in viscous fluids. *J Fluids Eng* 141:031202
10. Balarac G, Métais O, Lesieur M (2007) Mixing enhancement in coaxial jets through inflow forcing: a numerical study. *Phys Fluids* 19:075102
11. Brush LMJ (1962) Exploratory study of sediment diffusion. *J Geophys Res* 67:1427–1433
12. Butcher D, Spencer A (2019) Cross-correlation of POD spatial modes for the separation of stochastic turbulence and coherent structures. *Fluids* 4:134
13. Bühler J, Papantoniou D (2001) On the motion of suspension thermals and particle swarms. *J Hydraul Res* 39(6):643–653
14. Champagne FH, Wygnanski IJ (1971) An experimental investigation of coaxial turbulent jets. *Int J Heat Mass Transf* 14:1445–1464
15. Crowe CT, Gore RA, Troutt TR (1985) Particle dispersion by coherent structures in free shear flows. *Part Sci Tech Int J* 3:149–158
16. Dahm WJA, Clifford EF, Tryggvanson G (1992) Vortex structure and dynamics in the near field of a coaxial jet. *J Fluid Mech* 241:371–402
17. Fan J, Zhang L, Zhao H, Cen K (1990) Particle concentration and particle size measurements in a particle-laden turbulent free jet. *Exp Fluids* 9:320–322
18. Fan J, Zhao H, Jin J (1996) Two-phase velocity measurement in particle-laden coaxial jets. *Chem Eng J* 63:11–17
19. Hall N, Elenany M, Zhu DZ, Rajaratnam N (2010) Experimental study of sand and slurry jets in water. *J Hydraul Eng* 136:727–738
20. Kadu A, Sakai Y, Ito Y, Iwano K, Sugino M, Katagiri T, Hayase T, Nagata K (2020) Application of spectral proper orthogonal decomposition to velocity and passive scalar fields in a swirling coaxial jet. *Phys Fluids* 32:015106
21. Kannaiyan K, Sadr R (2013) Numerical simulation of particle-laden coaxial turbulent jets. *Int J Comput Methods Eng Sci Mech* 14:61–73
22. Ko NWM, Kwan ASH (1976) The initial region of subsonic coaxial turbulent jets. *J Fluid Mech* 73:305–332
23. Lee JHW, Chu VH (2003) Turbulent jets and plumes; A Lagrangian approach. Kluwer Academic Publishers Group, The Netherlands, p 390

24. Lumley JL (1967) The structure of inhomogeneous turbulence. In: Yaglom AM, Tatarski VI (eds) *Atmospheric turbulence and wave propagation*, pp 166–178
25. MATLAB [Computer software]. MathWorks, Natick, MA
26. Manzouri M, Azimi AH (2019) A study of mound formation by discharging sand particles through oblique pipes in stagnant water. *Int J Sediment Res* 34:564–576
27. Manzouri M, Azimi AH (2019) Effects on oily sand jet evolution from impact momentum and channelization of particles through an immiscible interface. *Int J Multiph Flow* 121:103124
28. Mazurek KA, Christison K, Rajaratnam N (2002) Turbulent sand jet in water. *J Hydraul Res* 40:527–553
29. Moghadaripour M, Azimi AH, Elyasi S (2017) Experimental study of particle clouds in stagnant water. *ASCE J Eng Mech* 143:04017082
30. Moghadaripour M, Azimi AH, Elyasi S (2017) Experimental study of oblique particle clouds in water. *Int J Multiph Flow* 91:101–119
31. Mohammadidinani N, Azimi AH, Elyasi S (2017) Experimental investigation of sand jets passing through immiscible fluids. *J Fluids Eng* 139:051303
32. Mostafa AA, Mongia HC, McDonnell VG, Samuelsen GS (1990) An experimental and numerical study of particle-laden coaxial jet flows. *Int J Heat Fluid Flow* 11:90–97
33. Nidhan S, Chongsiripinyo K, Schmidt O, Sarkar S (2020) Spectral POD analysis of the turbulent wake of a disk at $Re = 50,000$. *Phys Rev Fluids* 5(12):124606
34. Pakzad L, Azimi AH (2017) Investigations on the dynamics of particle clouds in stagnant water using response surface methodology. *Can J Civ Eng* 44(2):117–128
35. Pearson K (1901) On lines and planes of closest fit to systems of points in space. *Philos Mag* 2(6):559–572
36. Pedel J, Thornock JN, Smith ST, Smith PJ (2014) Large-eddy simulation of polydisperse particles in turbulent coaxial jets using the direct quadrature method of moments. *Int J Multiph Flow* 63:23–38
37. Rajaratnam N (1976) *Turbulent jets*. Elsevier, Netherlands, p 304
38. Ranga-Dinesh KKJ, Savill AM, Jenkins KW, Kirkpatrick MPA (2010) Study of mixing and intermittency in a coaxial turbulent jet. *Fluid Dynam Res* 42:025507
39. Sadr R, Klewicki JC (2005) Flow field characteristics in the near field region of particle-laden coaxial jets. *Exp Fluids* 39:885–894
40. Safer D, Beghidja A (2019) Mixing in turbulent coaxial jet. *J Mech Eng Res Dev* 42:110–119
41. Schmidt O, Colonius T (2020) Guide to spectral proper orthogonal decomposition. *AIAA J* 58:1–11
42. Schmidt O, Towne A, Colonius T, Cavalieri A, Jordan P, Brès G (2017) Wavepackets and trapped acoustic modes in a turbulent jet: coherent structure eduction and global stability. *J Fluid Mech* 825:1153–1181
43. Sharif F, Azimi AH (2020) Particle cloud dynamics in stagnant water. *Int J Multiph Flow* 125:101–119
44. Sharif F, Azimi AH (2021) Effects of velocity ratio on dynamics of sand-water coaxial jets. *Int J Multiph Flow* 140:103643
45. Sirovich L (1987) Turbulence and the dynamics of coherent structures. I. Coherent structures. *Q Appl Math* 45:561–571
46. Taofeeq H, Aradhya Sh, Shao J, Al-Dahhan M (2018) Advance optical fiber probe for simultaneous measurements of solids holdup and particles velocity using simple calibration methods for gas-solid fluidization systems. *Flow Meas Instrum* 63:18–32
47. Villermaux E, Rehab H (2000) Mixing in coaxial jets. *J Fluid Mech* 425:161–185
48. Welch PD (1967) The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Trans Audio Electroacoust* 15(2):70–73
49. Wicker RB, Eaton JK (1994) Near field of a coaxial jet with and without axial excitation. *AIAA J* 32:542 p

50. Wu HJ, Zhou FD, Wu YY (2001) Intelligent identification system of flow regime of oil-gas-water multiphase flow. *Int J Multiph Flow* 27:459-475
51. Zhang Y, Vanierschot M (2021) Determination of single and double helical structures in a swirling jet by spectral proper orthogonal decomposition. *Phys Fluids* 33:015115

Design of Roundabout to Replace All-Way Stop Controlled Intersection: Case Study in Niagara-on-the-Lake



N. Farag, F. T. Choudhury, Y. Cao, D. Ponce, C. Del Rosario, S. Alkarawi, M. Rataul, S. Arkatkar, and S. M. Easa

1 Introduction

Concerning future community growth, Niagara Region is undertaking various transportation improvement initiatives in order to prepare for the need for better connectivity and capacity as set out in their Transportation Master Plan (TMP), *How we Grow*. Furthermore, providing better and safer access to active transportation facilities is another primary consideration of future projects [9]. This study focuses on the intersection of York Road and Four Mile Creek Road, in the municipality of Niagara-on-the-Lake.

The existing four-legged intersection uses All-Way Stop Control (AWSC) to assign the right-of-way between approaching vehicles. However, the existing design poses several challenging movements, and the potential of insufficiency to address future traffic demands considering an anticipated 2% growth rate. To address these concerns, this report proposes three alternative improvements: (1) implementation of a roundabout, or (2) signalization of the intersection, as well as (3) the do-nothing alternative (i.e. maintaining AWSC). Other alternatives such as interchanges were explored in the initial stages but were ruled out in the preliminary stages considering the existing and projected traffic volumes, cost, and space requirements. The proposed alternatives were evaluated based on cost, safety, environmental impact and operational benefits to the study area.

The original project from the Municipality of Niagara included conducting an Environmental Assessment (EA). While many components of the EA process are omitted in this study, this report nonetheless strives to identify existing problems

N. Farag (✉) · F. T. Choudhury · Y. Cao · D. Ponce · C. Del Rosario · S. Arkatkar · S. M. Easa
Department of Civil Engineering, Ryerson University, Toronto, Canada
e-mail: natalie.farag@ryerson.ca

S. Alkarawi · M. Rataul
Transportation Planning, Regional Municipality of Niagara, Niagara, Canada

within the intersection and assess the impact and effectiveness of various traffic control measures in the study area, including environmental impacts. The critical component tasks to completing this project included a literature review, development of design alternatives, modelling and analysis, followed by evaluating the alternatives to determine the optimal choice, and a detailed design of the best alternative. The literature review consisted of relevant topics including a review of the Environmental Assessment process, geometric, safety and operational effects of roundabouts and signalization, relevant standards and guidelines, and case studies on similar projects. Further details about the findings in this paper can be found in the full capstone study.

The following sections discuss the study area and identified design alternatives. This is followed by the evaluation methodology, including weighting and sensitivity analyses. Lastly, a detailed design of the optimal alternative and operational modelling is presented (Fig. 1).

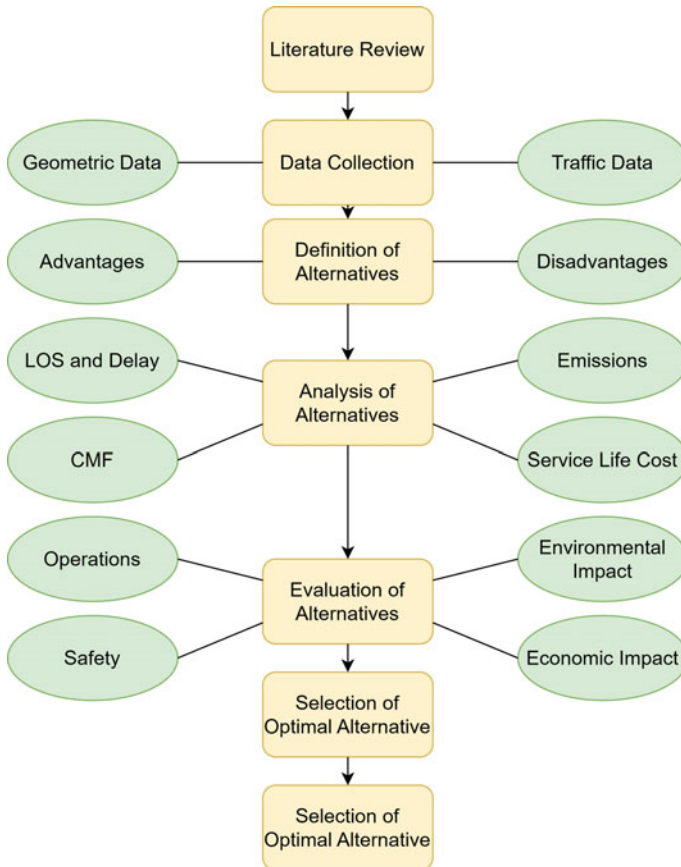


Fig. 1 Project tasks

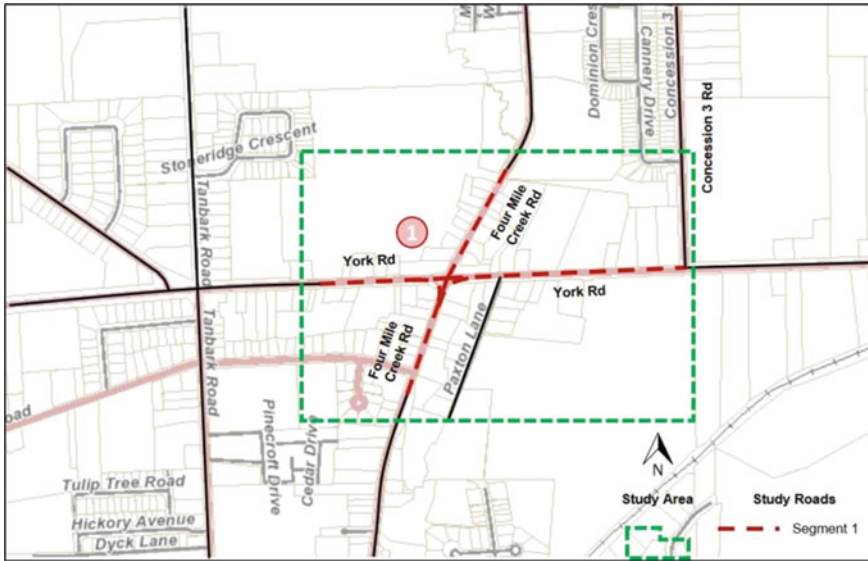


Fig. 2 Study roads for York Road and Four Mile Creek Road [9]

2 Study Area

The study area is a four-legged unsignalized intersection operating with All-Way Stop Control between York Road (RR81) and Four Mile Creek Road (RR100) located in Niagara-on-the-Lake, Ontario. The study area is bounded by York Road at distances of 250 m to the North and South, Concession 3 Road on the East, and 250 m to the West of Four Mile Creek Road as shown in Fig. 2.

Issues encountered at the existing unsignalized intersection primarily include complicated movements in the eastbound and northbound right-turn channels. Furtherly, the study area is connected to the Queen Elizabeth Way (QEW) in the west and the community of Queenston in the east, generating high traffic volumes through the intersection. Simulation of existing conditions produces intersection delays of 15.4 s with LOS C, with projected intersection delays for the year 2041 of 101.5 s, at LOS F.

3 Alternatives

The three alternatives proposed included a roundabout, signalization, and the do-nothing alternative (i.e. maintaining AWSC). Modelling of the current intersection after application of a 2% growth rate determined that the existing AWSC design would be insufficient to accommodate projected volumes given a LOS F. As such,

this alternative was eliminated from further consideration. The other two alternatives are explored in detail in the following section.

3.1 Roundabout

The first alternative considered was implementing a roundabout within the intersection to improve traffic operations, and improve safety for all road users, including vehicles, cyclists, and pedestrians. Overall, compared to unsignalized intersections, roundabouts are safer as they promote lower travelling speeds to reduce crash severities and probabilities [11]. Roundabouts can often also provide low delays and traffic control in moderate traffic conditions [11].

The majority of design guidance for preliminary evaluation of the roundabout were based on typical values from NCHRP 672 [11]. Based on traffic counts of the intersection from Niagara Region, the projected peak hour traffic would not exceed 1600 veh/h, which a typical single-lane roundabout would sufficiently serve. The preliminary design speed limit would be reduced from the posted limit by 10 km/h, to a design speed of 40 km/h. A preliminary design circulatory road width of 4.5 m was used to accommodate a theoretical S-BUS-40 design vehicle. The dimensions of the inscribed circle diameter would range between 27 and 55 m for a single lane roundabout and is dependent on the specific design vehicle. For preliminary evaluation design purposes, a typical roundabout alignment angle of 90° was assumed. As shown in Fig. 3, while a diameter of 27 m would result in the least disruption to the surrounding features in the study area, the larger diameter of 55 m would still be feasible.



Fig. 3 Inscribed circle diameter of 55 m (left) and 27 m (right) on study intersection

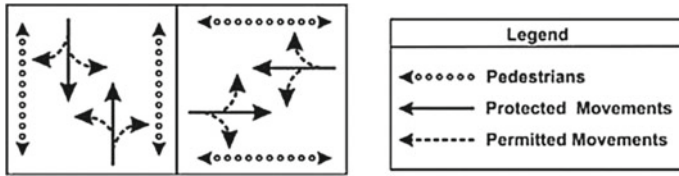


Fig. 4 Typical phase composition [5]

Table 1 Preliminary phase timing design [8]

| Phase element | Timing (s) |
|----------------------|------------|
| Cycle time | 62.0 |
| Minimum green time | 20 |
| Amber | 4.1 |
| All-red | 2.0 |
| Pedestrian walk time | 14.0 |

3.2 Signalized Intersection

The second alternative presented is the implementation of fully actuated traffic control signals. This alternative would pose minimal effects to existing infrastructure, as well as the potential to reduce vehicle delays and organize traffic movements for road users, including pedestrians and cyclists.

Actuated signal control is highly dependent on the sensors that make decisions regarding cycle lengths and other phasing elements [15]. As the signalization design is actuated, the signal would typically rest on green on the main road until a vehicle or pedestrian calls for movement on the minor road. However, for evaluation purposes, it was assumed that equal timing would be provided to both directions and a cycle length between 70 and 90 s would be provided. Using an iterative approach for phase composition, a typical 2-phase design as shown in Fig. 4 was used for analysis.

Furthermore, following Niagara Region and Ontario Traffic Manual (OTM) standards, the phase timing as shown in Table 1 was tentatively devised [8].

4 Evaluation Methodology

4.1 Evaluation Criteria

4.1.1 Operations and Mobility

Quality of service is a subjective observation on the performance of a transportation facility concerning road users [13]. Control delay is a measure of increased travel

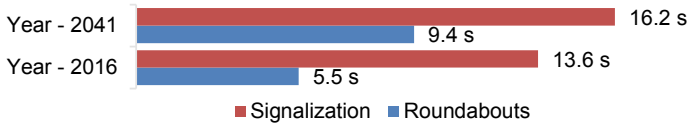


Fig. 5 Intersection delay (seconds) for alternatives for current and future traffic volumes

time, driver discomfort and frustration among other factors contributing to quality of service. Overall (intersection) delay is used for evaluation in this study due to its quantitative nature, compared to the Level of Service (LOS) metric being non-quantitative in nature.

Traffic Reports from Niagara Region in the year of 2016 for the PM period were used for analysis, where a 2% growth rate was used for a design period of 20 years to calculate the anticipated traffic demand and used for delay calculations. Forecasted volumes are calculated using Eq. 1, where AADT is the Annual Average Daily Traffic, AACR is the growth rate, and n is the number of forecasted years [14].

$$AADT_{\text{future}} = AADT_{\text{current}}(1 + AACR)^n \quad (1)$$

The All-Way Stop Module in Highway Capacity Software (HCS) 7 was used to construct a model of the existing AWSC intersection. The intersection was found to produce an LOS F, indicating the current design would fail to service future traffic demand. As such, the do-nothing alternative was eliminated from further consideration. Simulation models based on preliminary design parameters were constructed for the roundabout using the Roundabouts module in HCS7, and the signalization design in InterCalc software. Assumptions were made regarding geometric features, heavy vehicle characteristics, peak hour factors, and other parameters within reasonable tolerances and following design guides. Figure 5 demonstrates the intersection delays simulated by using alternatives to service existing and future traffic volumes. While both alternatives scored an overall LOS A, it is evident that the roundabout is more robust in reducing delays with a lower overall delay of 9.4 s.

4.1.2 Safety

Of the various criteria that can be used to evaluate the overall efficacy of transportation facilities, safety is the most critical factor [1]. Crash frequency is a fundamental indicator of safety in transportation-based decision making. CMFs are an excellent tool for evaluating the effectiveness of different treatments or designs and are calculated as the expected crash frequency ratio before and after treatment. In general, CMF factors between 0 and 1 will indicate a reduction in crash frequency, whereas a CMF greater than 1 indicates an unsuccessful treatment.

Before-after studies of converting five, rural, single-lane stop-controlled intersections to roundabouts in the U.S. found a CMF of 0.42 for overall crashes and

0.18 for injury crashes, respectively [10]. While later research has concluded that there are insignificant effects on safety when converting from AWSC to roundabouts [11], the former empirical report was used for further calculations. In the study by Persaud, Retting, Garder, and Lord, CMFs for converting signalized intersections to roundabouts were determined to be 0.65 and 0.26 for overall and injury crashes, respectively. Using the mathematical concept of CMFs, it can be inferred the CMF of converting AWSC to a signalized intersection would be approximated at 0.729. As such, the roundabout alternative is more effective in reducing crashes compared to the signalization alternative, the former with a CMF of 0.42 and the latter with a CMF of 0.729. As noted, while the later research on AWSC-roundabout CMFs would discredit the previous study by Persaud et. al, the values used from this only produce a more conservative comparison between the two proposed alternatives, and the difference would not impact the conclusion.

4.1.3 Environmental Impact

Traffic intersections are often characterized as significant contributors to air pollution; when vehicular traffic is forced to slow down and idle at intersections, there is a higher incidence of fuel consumed and consequently, higher vehicular emissions [7]. Long-term air quality is largely impacted by the pollutants found in vehicular emissions, including carbon monoxide (CO), carbon dioxide (CO₂), nitrogen oxides (NO_x), as well as hydrocarbons, and overall fuel consumption.

Due to their design, modern roundabouts significantly reduce vehicular emissions compared to other intersection control methods by reducing idling time. As traffic rarely stands still within a roundabout, the overall emissions produced by vehicles is minimized. Conversely, signalized intersections are considered significant proponents of causing vehicular emissions and adverse environmental impacts due to significant levels of idling at such intersections and higher levels of congestion [6]. For emissions at a roundabout to be equal to that of a signalized intersection, the average delay at the roundabout would need to be significantly larger than that of the signalized intersection. A model constructed by Várhelyi [16] using a before-after study comparing roundabout and signalization found significantly lower emissions at roundabouts as demonstrated in Fig. 6.

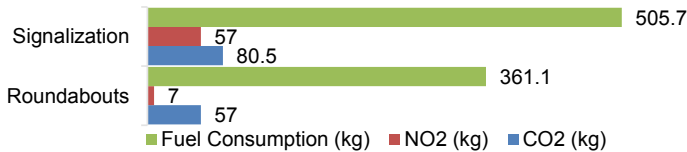


Fig. 6 Roundabout and signalization emissions [16]

4.1.4 Cost

Several factors that contribute to the budget required to install and maintain traffic facilities, including intersections. From a case study done for an intersection in the city of Hamilton, the total 20-year costs for roundabouts and traffic signals are summarized in Table 2. Typical costs for implementing an average roundabout include construction, land purchase, design and engineering costs, and operational and maintenance fees. Roundabout maintenance may consist of landscaping, restriping and repaving, and snow removal and storage during winter. For a signalized intersection, some fees to consider are the signal design costs, fixture installations, and annual maintenance of facilities. There is also a significant additional cost required which is the re-timing and replacement of the traffic signal. Injury crash costs are typically higher for signalized intersections than for roundabouts. This is due to the higher severity of crashes and incidences of crashes from more conflict points at signalized intersections.

4.2 Selection of Best Alternative

The proposed design alternatives, (1) roundabout and (2) signalization, were evaluated based on the four criteria outlined, and the results of the weighted evaluations are summarized in Table 3. Criteria scores were first normalized. The priority-weighted

Table 2 20-year costs of implementing roundabouts and signalization [2]

| | Roundabout | Signalization |
|-----------------------------------|------------|---------------|
| Total estimated capital cost | \$410,000 | \$426,000 |
| Property costs | \$29,500 | \$13,000 |
| Injury crash costs | \$314,850 | \$626,000 |
| Annual maintenance | \$40,000 | \$80,000 |
| Future traffic signal replacement | – | \$75,000 |
| Total | \$794,350 | \$1,220,300 |

Table 3 Results of evaluation

| Evaluation criteria | Non-weighted score | | Priority weighted score | | Equally weighted score | |
|---------------------|--------------------|--------|-------------------------|--------|------------------------|--------|
| | Alt. 1 | Alt. 2 | Alt. 1 | Alt. 2 | Alt. 1 | Alt. 2 |
| Safety | 4.2 | 7.29 | 1.47 | 2.55 | 1.05 | 1.82 |
| Operations | 0.94 | 3.52 | 0.25 | 1.23 | 0.23 | 0.88 |
| Cost | 4.25 | 6.43 | 0.63 | 2.25 | 1.06 | 1.61 |
| EI | 7.9 | 12.2 | 1.82 | 4.27 | 1.98 | 3.05 |
| Total | 17.3 | 29.4 | 4.18 | 10.30 | 4.32 | 7.36 |

scheme allocated weights as the following: safety (35%), operations (27%), environmental impact (15%) and cost (23%). This weighting scheme was largely based on consultation from industry clients due to limitations preventing public consultations. An additional equal weighting scheme was used, allocating a weight of 25% to each criterion. In this scheme, a lower score indicates higher success; thus, it is concluded that Alternative 1 with lower scores is the optimal solution. Additionally, it is noted that this conclusion is independent of the weights allocated to the evaluation criteria.

5 Detailed Design

Several design elements crucial to the safe and successful operation of a roundabout, including geometric design, illumination, and adequate landscaping, which help achieve overall speed control, mobility and visibility objectives.

5.1 Design Considerations

Before geometric design, several determinations must be made with regards to the design of the roundabout. First, based on existing and projected traffic counts, it was determined that with a 23% left-turning volume and overall AADT under 16,000 veh/day, it would be viable to go forward with a single-lane roundabout design based on NCHRP 672 recommendations [11]. As the existing area of the intersection is not adequate to accommodate a four-legged single lane roundabout, it would be expected that additional property may need to be purchased. A significant consideration in geometric design is accommodating the design vehicle, which depends on anticipated road users, and the needs of the surrounding road network. The most suitable design vehicle was determined to be the WB-20 (WB-67). Lastly, speed management is a significant factor to be considered to ensure safe operation of the intersection. With existing posted road speeds of 50 km/h, and due to the existing geometry of the approaches, it was determined that a posted intersection speed of 20 km/h would be required to ensure sufficient sight distance on all approaches. Using specific provincial standards in Ontario, lane widths used are a minimum of 3.5 m.

5.2 Geometric Design

The selection of geometric design elements for roundabouts are primarily dependent on area constraints, the requirements of the design vehicle, while also providing enough deflection to control speeds. The design parameters chosen are summarized in Fig. 7, a diagram of the final geometric design. This section discusses some of the

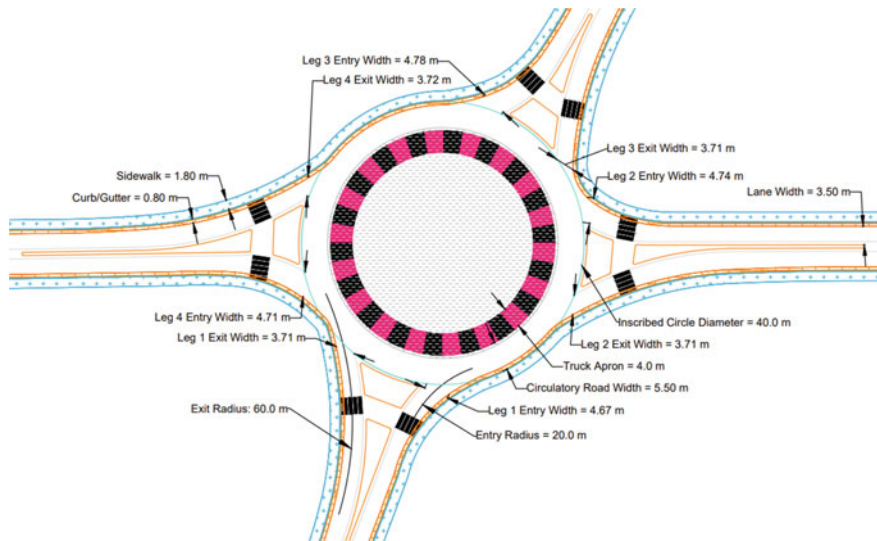


Fig. 7 Geometric design of roundabout

rationales behind many of the geometric design choices based on guidelines outlined in NCHRP 672 [11], while a comprehensive design and discussion can be found in the complete capstone research paper by the authors.

The inscribed circle diameter is the total distance across the circulatory segment of the roundabout, which is the sum of the central island diameter and double the circulatory roadway width. A suitable inscribed circle diameter should be large enough to accommodate the design vehicle, however, provide tight turning radii to slow down smaller vehicles. Through an iterative process to provide adequate movement for the design vehicle, while also considering space requirements, a circle diameter of 53.0 m was selected. Additionally, considering that circulatory roadway widths range between 4.8 and 6.1 m [11], a value of 5.5 m was used for this design in conjunction with the truck apron. As such, the central island would result in a diameter of 42 m given the other geometric design choices. Lastly, a truck apron with a width of 4.0 m was provided to allow for the movement of the design vehicle through the roundabout, while restricting the movement of smaller vehicles to prevent speeding.

While it is typically advantageous to employ an intersection aligning at 90° for all approaches, due to the existing alignment of the intersection, a non-standard alignment as shown in Fig. 7 was used to minimize disruptions to the existing land features. However, through manipulation of the entry and exit radii, it was ensured that approach angles were not less than 20° as recommended by the AASHTO Green Book [12]. As for entry design, entry widths were used within the typical range of 4.2–5.5 m. Large entry widths may confuse drivers who may consider the large entry widths to be two distinct lanes. Additionally, while entry radii may range from 15 to 30 m [11], typically entry radii exceeding 20 m has been found to be inconsequential

in improving roundabout capacity. As such, entry radii of 20 m were used on all approaches to ensure there is adequate deflection to prevent speeding. Lastly, exit radii of 60 m were provided on all exit legs. Exit radii should exceed that of the entry curb to prevent congestion or crashes, while providing enough deflection for vehicles to yield to pedestrians.

Lastly, while a detailed discussion on splitter design is not presented in this paper, the splitter islands were designed with lengths exceeding 30 m to provide adequate visibility and prevent entering or exiting traffic from entering conflicting traffic streams. Furthermore, typical nose and corner radii were used as provided in AASHTO and NCHRP 672 guidelines. Additionally, splitter islands were no less than 1.83 m wide to provide sufficient refuge to active transportation users. Sidewalk and curb widths were designed at 1.80 and 0.80 m according to Niagara Region Complete Streets guidelines.

5.3 Active Transportation

Data on movements in the existing intersection shows that there is minimal pedestrian and cyclist volumes. With this said, with community growth and improvements to transportation facilities, the increase in active transportation (pedestrians, cyclists) may increase. Traditionally, modern roundabouts do not implement traffic control devices. However, pedestrian signals may be applied should the need arise in the future. In the absence of additional control, adequate deflection in the circulatory roadway leading to the exit and sufficient visibility is provided to drivers in the roundabout to slow and yield for pedestrians. The pedestrian has the responsibility to determine an appropriate time to enter the crosswalk. For cyclists, there is no additional facility implemented as a dedicated bike lane on the circulatory roadway would be unsafe [11]. Thus, cyclists can choose to cross as pedestrians, or seasoned cyclists may traverse the circulatory roadway by merging with vehicle traffic.

5.4 Sight Distance and Visibility

As vehicles approach a roundabout, it is required to provide adequate visibility of vehicles already circulating within the roundabout, and vice-versa. As such, roundabout design requires the analyses of two types of standard sight distances: (1) intersection sight distance (ISD), and (2) stopping sight distance (SSD). Both are explored in detail for all approaches in the authors' full capstone paper. However, the following section summarizes ISD calculations for a single approach. It is noted, ISD, SSD, and additional checks for swept path and fastest path for mobility purposes were verified in the complete study.

Intersection Sight Distance (ISD) is expressed as the necessary distance for a driver to identify gaps in a conflicting traffic stream to complete safe movements [3, 4].

ISD should be checked at the entry each leg at two locations: (1) a specified distance (15 m) before the yield line, and (2) a vehicle at the yield line. From these points, ISD triangles are made to the entering and conflicting/circulating vehicles to determine if yielding vehicles have enough visibility to stop or slow down if necessitated. It should be noted that excessive ISD may result in compromised safety. Thus, strategic landscaping was employed to control unnecessary visibility. First, the length of the conflicting leg to a circulating vehicle (d_1) can be calculated from Eq. (2), where t_c is the critical headway for entry into the roundabout (seconds), and V_{cir} is the design speed for the conflicting circulating movement.

$$d_1 = 0.278t_c V_{cir} \quad (2)$$

Next, the length of the conflicting leg of the entering vehicle (d_2) may be calculated using Eq. (3), in which the V_{enter} is the design speed of the conflicting entry movement, and d_{cir} is the distance travelled along the circulatory roadway from the point of entry to the travelled along the circulatory roadway from the point of entry to the y -axis of the roundabout [3].

$$d_2 = 0.278t_c(V_{enter} + V_{cir})/2 - d_{cir}(V_{enter} - V_{cir})/2V_{cir} \quad (3)$$

For the south leg, the critical gap was assumed to be 5.00 s based on software defaults, and the design speed used for the circulating vehicle was 20 km/h, with a northbound entry speed of 35 km/h, producing a distance d_1 of 27.8 m, and d_2 of 48.7 m as demonstrated in Fig. 8. Figure 8 also demonstrates the section of non-commercial property that would require tree landscaping (or alternatives) that would restrict visibility past the sight leg to the conflicting circulating vehicle.

Fig. 8 ISD triangle for approach 1 (South Leg)

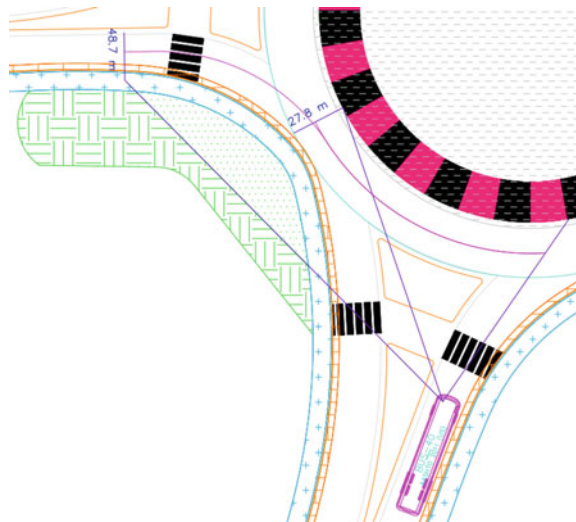


Table 4 Critical simulation results for future volumes

| Approach leg | Time interval (s) | Delay (s) | LOS |
|--------------------|-------------------|-----------|-----|
| NB | 800–900 | 15.96 | C |
| SB | 400–500 | 7.11 | A |
| WB | 200–300 | 12.19 | B |
| EB | 500–600 | 18.04 | C |
| Intersection total | 800–900 | 7.77 | A |

5.5 Simulation

5.5.1 Results

The roundabout design was simulated in VISSIM for an interval of 15-min using afternoon peak-hour volumes for the current and projected traffic demand, as discussed in Sect. 4.1.1. The roundabout was reconstructed based on the geometric design elements within the current iteration, and all other inputs were left as defaults in the software. The summary of the critical delays and LOS during the simulation of projected peak hour traffic summarized in Table 4. The overall intersection delays were minimal, producing the critical LOS A for the entire simulation interval. Intersection delays were reduced significantly from the existing AWSC intersection, at 93.73 s less delay for design year demand.

5.5.2 Limitations of Results

Due to the limitations of the software, it was not possible to input traffic volumes per turning movement. Therefore, volumes and heavy vehicle percentages were averaged for each approach leg for hourly volumes, given 3-h peak demands. Additionally, while there was minimal pedestrian volume in current traffic counts, pedestrian volumes were estimated to be 10 pedestrians per approach leg for the entire interval for future simulation.

6 Conclusions

With the growing population and the greater demands on transportation networks, there is a need to improve existing road facilities using improved intersections such as roundabouts. As such, in accordance with the Niagara Region's TMP to evaluate specific options to improve the existing AWSC intersection, two design alternatives, (1) signalization and (2) a roundabout were considered along with the do-nothing alternative (i.e. maintain AWSC). It was determined that maintaining the current intersection would fail to service the anticipated traffic volumes for the 20-year design

period. The evaluation of the alternatives showed that a roundabout would be the optimal choice to improve the existing intersection. A detailed design using NCHRP 672 recommendations and existing literature was constructed, and the roundabout design's operational abilities was simulated. From AWSC, the roundabout improved the overall LOS from F to A. Due to the level of intricacies to this project, several details and discussions have been excluded in relation to costs, grading, and splitter design, among other considerations.

Acknowledgements The authors would like to thank Dr. Bhagwant Persaud and other faculty members for providing their overall insight and advice on constructing the analytical and design models. The authors also are grateful for the opportunity provided by the team at Niagara Region, and appreciate the continued support.

References

1. AASHTO (2014) Highway safety manual, 1st edn. with supplement 2014. American Association of State Highway and Transportation Officials (AASHTO), Washington, DC, USA
2. City of Hamilton (2008) Use of roundabouts in the City of Hamilton. City of Hamilton, Hamilton, ON, Canada
3. Easa SM (2017) Design guidelines for symmetrical single-lane roundabouts based on intersection sight distance. *J Transp Eng Part A Syst* 143(10):1–11
4. Easa SM, Ma Y, Liu S, Yang Y, Arkatkar S (2021) Reliability analysis of intersection sight distance at roundabouts. *Infrastructures* 5(67):1–20
5. Gough JW (2008) Canadian capacity guide for signalized intersections. Institute of Transportation Engineers (ITE), Washington, DC, USA
6. Madireddy M, De Coensel B, Can A, Degraeuwe B, Beusen B, De Vlielier I, Botteldooren D (2011) Assessment of the impact of speed limit reduction and traffic signal coordination on vehicle emissions using an integrated approach. *Transp Res Part D Transp Environ* 16(7):504–508
7. Mandavilli S, Rys M, Russel ER (2007) Environmental impact of modern roundabouts. *Int J Ind Ergon* 38:135–142
8. Niagara Region (2014) Traffic signal standards. Niagara Region, Niagara, ON, Canada
9. Niagara Region (2020) Request for proposal for consulting assignment to undertake a municipal class environmental assessment for intersection improvements to York Road and Four Mile Creek Road. Niagara Region, Niagara, ON, Canada
10. Persaud BN, Retting RA, Garder PE, Lord D (2001) Safety effect of roundabout conversions in the united states: empirical Bayes observational before-after study. *Transp Res Rec* 1751(1):1–8
11. Rodegerdts L, Bansen J, Tiesler C, Julia Knudsen EM, Johnson M, Moule M et al (2010) Roundabouts—an informational guide, 2nd edn. (NCHRP Report 672). Transportation Research Board, Washington, DC, USA
12. Stančerić I, Ahac S, Bezina Š, Vlaović F (2019) Design limits for intersection angles between approach legs of suburban roundabouts. *GRAĐEVINAR* 71:389–399
13. Transportation Research Board (2016) Highway capacity manual—a guide for multimodal mobility analysis, 6th edn. Transportation Research Board, Washington, DC, USA
14. U.S. Department of Transportation (2018) Traffic data computation method. Department of Transportation, Washington, DC, USA

15. Urbanik T, Tanaka A, Lozner B, Lindstrom B, Lee K, Quayle S, Beaird K, Tsoi S, Ryus P, Gettman D, Sunkari S, Balke K, Bullock D (2015) NCHRP report 812: signal traffic manual, 2nd edn. Transportation Research Board, Washington, DC, USA
16. Várhelyi A (2002) The effects of small roundabouts on emissions and fuel consumption: a case study. *Transp Res Part D Transp Environ* 7(1):65–71

Improving Selected Intersections Across Niagara Region for Miovision's Smart City Program



Neil Gobin, Michelle Nguyen, Jasmine Puthoor, Rahmah Tariq, Michelle Yip, Sulaf Alkarawi, Manny Rataul, Said M. Easa, and Essam Dabbour

1 Introduction

Niagara Region is a developing region that is continuously expanding in population, employment, and business. By 2041, the population is expected to balloon by 168,000 people and add 80,000 jobs [4]. As a result, traffic and commuting infrastructure will require additional management to ensure existing facilities can meet the forecasted demand. To identify areas that may require an improved level of service (LOS), the Niagara Region, in collaboration with Miovision, has decided to identify the three most problematic intersections in the region and develop solutions that can accommodate the traffic demands of this growing region.

This project entails determining the top three intersections from the shortlist of ten historically problematic intersections that benefit the most from proposed solutions. The ten intersections are as follows:

1. Carlton Street/North Service Road and Geneva Street
2. Christie Street Clarke Street/South Service Road
3. Bunting Road and Queenston Street
4. Burgar Street and Division Street
5. Dunlop Drive and Geneva Street
6. Dieppe Road/North Service Road/Ramp and Niagara Street
7. Dunn Street and Stanley Avenue
8. Garner Road and Lundy's Lane
9. Geneva Street and Westchester Avenue
10. Niagara Street and Parnell Road.

N. Gobin · M. Nguyen · J. Puthoor · R. Tariq (✉) · M. Yip · S. M. Easa · E. Dabbour
Department of Civil Engineering, Ryerson University, Toronto, ON, Canada
e-mail: rahmah.tariq@ryerson.ca

S. Alkarawi · M. Rataul
Transportation Planning, Regional Municipality of Niagara, Niagara, ON, Canada

© Canadian Society for Civil Engineering 2022
S. Walbridge et al. (eds.), *Proceedings of the Canadian Society of Civil Engineering Annual Conference 2021*, Lecture Notes in Civil Engineering 250,
https://doi.org/10.1007/978-981-19-1065-4_12



Fig. 1 Top three most problematic intersections (Niagara 2020)

In selecting the three most high-risk intersections, a weighted scoring system was used to determine the relative safety compared to the other intersections. This analysis considers quantifiable data such as collision data and social impact, intersection LOS, and pedestrian and cyclists based on peak movement counts. In addition, qualitative data such as land use in the area, local businesses, and stakeholders were considered to gain a complete understanding of the area. A two-sided analysis was used to determine the three most problematic intersections: Carlton Street/North Service Road and Geneva Street, Geneva Street and Westchester Avenue, and Dieppe Road/North Service Road/Ramp and Niagara Street. Figure 1 illustrates the location of the top three intersections.

Moving forward, the objective of this report is to provide a concrete plan of action for each intersection to improve its overall performance. This will be determined after minor changes and software simulations are conducted. The team will first investigate a combination of signal timing optimization and geometric changes for each intersection until a significant improvement in LOS or delay timing is achieved based on these changes. Following this, minor changes that can be immediately implemented, such as improvements and additions in signage, will be considered. Lastly, a final analysis will be provided showcasing the necessary changes, compared to the “do-nothing” approach, that can be made to each intersection to improve its LOS and effectively conduct traffic flow and volume.

2 Methodology

2.1 Selecting the Top Three Intersections

In this project, a combination of weighted rankings and scaled criteria was used to perform the analysis to determine the top three intersections out of the ten intersections initially provided. The weighted ranking was used for quantitative analysis, which helped organize and analyze the dataset that factored in multiple layers of numerical data. A higher weight was assigned to a parameter that would potentially contribute to making the intersection less safe throughout all the weighted equations. This ensured that the intersections with the highest scores were the more problematic intersections. The scaled criteria were used for qualitative information, which helped analyze subjective characteristics of each intersection that cannot be expressed as a number. A three-part scoring system was also used in the analysis process: Score I, Score II, and Final Score. Score I consisted of the sum of all the individual parameters that made up the equation. Score II consisted of the sum of each parameter after being multiplied by their respective weight. The Final Score utilized a calibration method such that the Score II for each collision was divided by the highest Score II in that data set. This method gave the collision with the highest score a final score of 100, which helped to have a more organized and adequately calibrated data set.

To select the top three intersections, an objective evaluation criterion was developed to yield the intersections that are deemed unsafe and problematic. The objective evaluation assessed the various impacts of a single criterion and related this score to the objective function. The function considered both quantitative and qualitative information, including collisions, intersection LOS, pedestrian LOS, cyclist LOS, and social impact. The objective function was as follows:

$$I_{final} = 0.25C + 0.13S + 0.25ILOS + 0.20PLOS + 0.17CLOS \quad (1)$$

where

| | |
|-------------|---------------------------------------|
| I_{final} | final objective intersection function |
| C | collision analysis |
| S | social impact |
| $ILOS$ | intersection level of service |
| $PLOS$ | pedestrian level of service |
| $CLOS$ | cyclists analysis. |

Each parameter was evaluated individually and then combined into Eq. 1 using weightage. The weights chosen for the parameters were subjective and were solely based on the severity of their effects or indication of an intersection that may classify the intersection as unsafe or problematic. A sensitivity analysis was also performed with two additional trials using different weightage parameters to see if the chosen weights are suitable and if they have a significant effect on the top three intersections.

Subsequently, following the analysis and simulations of the intersections in the Niagara Region, the three intersections that deemed to be the most problematic regarding the above-mentioned methodology were:

1. Carlton Street/North Service Road and Geneva Street
2. Geneva Street and Westchester Avenue
3. Dieppe Road/North Service Road/Ramp and Niagara Street.

For a more detailed explanation on selecting the top three intersections, the reader is referred to [8].

2.2 Implementing Countermeasures

The chosen three intersections were then further analyzed to decide on what course of action should be taken for the individual intersections to optimize their functionality and improve their Level of Service. As per the requirements of the Niagara Region, the countermeasures recommended for the intersections were:

1. The “Do-Nothing” Approach
2. Level of Service Improvements
3. Safety Countermeasures.

To select the countermeasures for each intersection, a planning process utilized to produce the best overall results. The intersection analysis process evaluated “Do-Nothing”, LOS Improvements, and Safety Countermeasures. For the LOS Improvements, the signal timing optimization and the geometric design improvements were considered. If the intersection elapsed, further adjustments were made to these countermeasures until ideal results were obtained. Minor improvements, such as signage, were then added to the intersection to account for the intersection’s safety. The process also carefully considered the additional factors that would impact the intersection, such as cost, and environmental and social impacts.

2.2.1 The “Do-Nothing” Alternative

When solving traffic networks’ concerns, the “do-nothing” countermeasure involves no newly developed facilities. It is typically chosen due to multiple factors that outweigh the need for change, such as cost, construction process, or environmental factors (Municipal Engineers Association 2020). For this project, the “do-nothing” alternative was considered a reference point for future traffic projections and to understand how each of the intersections will be affected if no changes are made.

2.2.2 Level of Service

Signal Timing Optimization

At each intersection, an existing timing plan is designed and provided by the Niagara Region, responsible for managing and controlling the traffic flow. Depending on when the signal timing plan was created, it may be reaching the end of its design period, meaning it is not managing traffic adequately. This would lead to multiple failures per cycle, and there would be long delays and lower LOS as a result. Therefore, adjusting the timings of signals would seem to be the most cost-effective and instant boost to intersection performance.

In this project, Synchro 11 software was used in tandem with traffic counts provided by the Niagara Region. The timing plans produced in this report have a design period of 10 years and should meet traffic demands barring any sudden growth or shifts in population. Using the existing traffic counts, the counts ten years into the future were projected at a growth rate of 2% per year, leading to a cumulative growth figure of 22% of original volumes.

Using Synchro11, detailed reports of each movement and its corresponding LOS and the overall intersection delay, and the volume/capacity (v/c) ratios of the movements were determined. The first approach was to use the optimized splits and optimize cycle length options provided by Synchro. If that change was not significant, other options were explored, including increasing total splits to allow more green time for the deficient movements. It also meant extending protected movements to see the effects that it would have on the movement delay, v/c ratio, and total delay. Once these options were exhausted, the following changes investigated the turn types for each movement and their effectiveness. Some of the options include split, permitted, protected, permitted plus protected, and reserved.

Once satisfactory results were achieved, Synchro's optimized split and optimize cycle length options were utilized. This final step will have produced optimal results and minimize delays for all approaches. It is worth noting that the delays shown are representative of rush hour volumes. As a result, it is not unusual to see LOS D or E during these times.

Geometric Design Improvements

A geometric change aims to provide a safe and efficient roadway that addresses the conflicting needs and concerns of all users of the intersection. Behind every intersection geometric design, there are basic elements that must be considered. These are human factors (driver habits, pedestrian characteristics, and bicyclists), traffic considerations (crash history, traffic movements, and turning movements), physical elements (sight distance, environmental factors, and traffic control), and economic factors (improvement costs, energy consumption, and right-of-way impacts) [9]. In order to create a well-designed intersection that can provide a safe and sufficient traffic flow for commuters, these factors must be taken into consideration. A series of geometric additions/changes were implemented into the intersections through software simulation on Synchro to determine any improvements. Some of the geometric improvements considered were the addition of left-turning lanes, auxiliary lanes with

right/left-turning movements to increase capacity and reduce crashes, roundabouts, pedestrian, and bike lanes as needed, and a lane addition.

2.2.3 Safety Countermeasures

Minor Improvements

Minor changes to a road facility may significantly improve its level of service and reduce the safety risks at a minimal cost. A minor improvement scheme on the existing intersection stays consistent with the general characteristics of the road network with the least amount of disruption. Some minor improvements include increasing the size of the regulatory signs, adding additional lights for visibility during night times, and optimizing pedestrian countdown signals (PCS). It is important to note that minor improvements for each intersections are different, depending on the intersection's characteristics. These minor improvements will follow the Ontario Traffic Manual (OTM) Books, the TAC manual, and other guidelines to follow the transportation regulations in Ontario.

Crash Modification Factors

To evaluate the safety conditions of the problematic intersections, improvements will be made based on the Crash Modification Factor (CMF) of the selected countermeasures. The CMF provides a quantitative estimate of the safety effect of the countermeasure and is used to identify countermeasures to reduce the number of collisions that will occur. The CMF values collected will be obtained from the Transportation Association of Canada manual and various studies provided by Crash Modification Factors Clearinghouse, where only the studies with a minimum of four stars are considered. Studies with a four or five-star rating indicate the studies' high level of accuracy, precision, and general applicability.

2.3 Additional Factors

Some of the factors that were considered while determining the countermeasures for each intersection were the cost, the environmental impacts, and the social impacts of the project. The cost considered was associated with any changes that were made per intersection. The environmental factors considered were the impact of each change on the air quality in the surrounding. Air quality was measured by the emissions report obtained from the SYNCHRO, which consists of carbon monoxide (CO), nitrogen oxides (NOx), and volatile organic compounds (VOC) emissions at each intersection. Some social factors that were considered included land acquisition, fuel consumption, and any noise or disruption caused by the changes. It was determined that land acquisition would be required to make any geometric design change to the intersection, which would consist of contacting the owner of the land under consideration. For this project, the design changes suggested were made assuming

that the required land has been acquired. While certain factors such as greenery, aesthetics, and disruption are not quantifiable, the suggestions made in this project were made after careful consideration of these factors as well.

3 Analysis and Results

The analysis and results for improving Intersection 1 (Dieppe Road/North Service Road/Ramp and Niagara Street) are presented in this section for illustration. For more details on evaluating the three intersections, the reader is referred to [8].

3.1 *Level of Service Improvements*

3.1.1 Signal Timing Optimization

In this analysis, the traffic data and signal timing plan were provided by Niagara Region to get a baseline understanding of how the intersection was performing. From this data, the peak PM hour was used to find the critical state of the intersection when it was experiencing its highest volumes. After inputting the different movement volumes and existing timing plan into Synchro 11, the resulting LOS of service calculated according to the Highway Capacity Manual Sixth Edition was LOS C with a 29.7-s control delay. For this study, these volumes were projected ten years into the future using a growth rate of 2% annually. This projection would ultimately lead to an increase of 22% in volume for each movement. The labelled diagrams of each movement and its associated volume before and after are shown in Figs. 2 and 3.

Considering the “do-nothing” approach, if this intersection was left as is with a 2% annual growth rate over ten years, it would have a LOS D with a 51.2-s delay according to HCM 6th. LOS D at PM peak is not unheard of and is still acceptable; however, this study looked to optimize conditions and improve conditions. From the HCM 6th report generated in Synchro 11, the failing traffic lanes traffic were identified as the through and through/right turn lanes in the southwest leg of the intersection, as shown in Figs. 4 and 5.

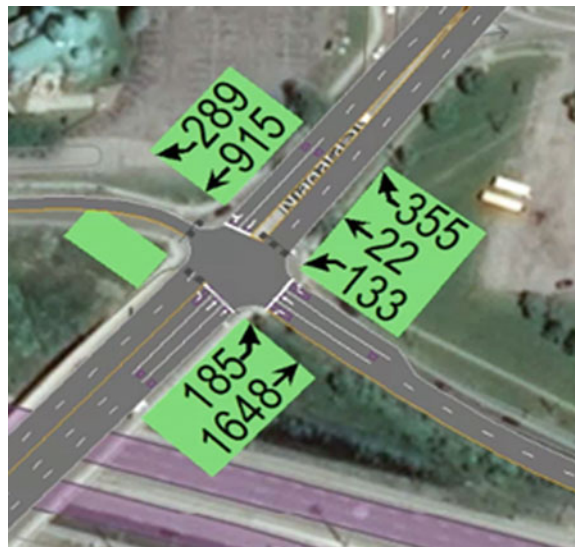
This approach had a LOS of F with an 82.9-s delay, significantly higher than the other two approaches, 43.0 s going northwest and 32.6 s going northeast. The v/c ratio for the through and through right were 1.05 and 1.07, respectively. This indicates the volume travelling in the lane is greater than the lane capacity and thus is failing. Based on these findings, this is where the adjustments began.

The solution found was a simple increase in green time for these movements. Although this solution was simple, there was some nuance in balancing the amount of green time added such that the other delays would not increase. In this case, adding 8.5 s of green time to the southwest and northeast splits proved to remedy the issue.

Fig. 2 Current traffic volume



Fig. 3 Projected traffic volume



After implementing this revised plan and increasing the volumes, the LOS of the intersection is D with a 38.7-s delay. The v/c ratio for the two critical lanes was reduced to 0.93 and 0.94, meaning although it is functioning at the upper end of its capacity, it is still within capacity. This solution improves each movement's LOS by distributing the delay between the northwest and southwest approaches, as shown in Figs. 6 and 7.

Fig. 4 Projected level of service



Fig. 5 Projected delay



Comparing these results to if no action was taken, the difference is a delay of 51.2 s (do-nothing) against 38.7 s (revised timing plan). Therefore, applying this change yields a 24.4% decrease in delay times for this intersection.

With the positive results found through just a signal timing change, this intersection does not warrant any geometric changes at this point. However, in the next 5–10 years, additional lanes or reconfiguration may be required as many of them are operating near capacity. This methodology of identifying critical spots, adjusting timing and turn, and implementing geometric changes will be carried through to the other two intersections that are in progress.

Fig. 6 Improved level of service



Fig. 7 Improved delay



3.2 Safety Countermeasures

3.2.1 Minor Improvements

The minor improvement at the intersection of Dieppe Road/North Service Road/Ramp and Niagara Street mainly consist of regulatory signs, which are essential in terms of the intersection's safety.

Locations of ONE-WAY signs should be placed near the right corner or the far-left corner of an intersection and facing the traffic perpendicular to a one-way street [6]. The current location of the ONE-WAY sign in the intersection of Dieppe Road/North Service Road/Ramp and Niagara Street needs to be changed. Currently, the sign is not visible to most drivers as they turn onto the one-way street into opposing traffic, as evident in the collision data. Thus, an additional ONE-WAY sign will be located along the westbound corner of Dieppe Road and Niagara Street.

A speed limit sign (maximum speed in km/h) will be added along Niagara Street within 150 m of the intersection lights. This is because no speed limits signs exist on any of the approaches, meaning drivers may unknowingly be above or below the posted limit. Adding a speed limit sign can reduce the number of accidents and increase safety precautions.

Another factor that will be considered at this intersection is a supervised school crossing by school cross guards. Since there is a school near the intersection, the children must travel safely alongside the high traffic volume. These school crossing guards will allow a right-of-way for school children as vehicles are yielded and will be located at Niagara Street and Facer Street corner during particular hours to provide children indicators of when to cross and deliver safety measures [5].

3.2.2 Crash Modification Factors

The Collision Details Report of Dieppe Road/North Service Road/Ramp and Niagara Street entails many angle collisions compared to the other impact types. To combat this issue and the overall safety of the intersection with the consideration of pedestrians and cyclists, red-light cameras and larger diameter signal displays will be installed, and signal operations will be optimized. Table 1 will present the effectiveness of each countermeasure previously listed:

Table 1 Dieppe road/north service road/ramp and Niagara street—safety countermeasures [3, 7, 10]

| Countermeasure | Targeted crash type | CMF ^a | CRF (%) ^b |
|--------------------------------|----------------------------|------------------|----------------------|
| Red-light camera | All angle | 0.916 0.623 | 8.4 37.7 |
| Larger diameter signal display | All angle | 0.8–0.9 ≥0.5 | 10–20 ≤50 |
| Optimizing signal operations | All pedestrian and cyclist | 0.92 0.63 | 8 37 |

^a CMF ≥ 1 indicates an increase in collisions; CMF ≤ 1 indicates a decrease in collisions

^b Crash Reduction Factor (CRF) = percentage of decrease in collisions

Table 2 Dieppe road/north service road/ramp and Niagara street—total gas emissions

| Emissions (g/h) | Current | Projected | Solution |
|---------------------------------|---------|-----------|----------|
| Carbon monoxide (CO) | 4258 | 6938 | 5792 |
| Nitrogen oxide (NOx) | 828 | 1350 | 1126 |
| Volatile organic compound (VOC) | 1078 | 1608 | 1342 |
| Total | 6164 | 9896 | 8260 |

Table 3 Dieppe road/north service road/ramp and Niagara street—fuel used

| | Current | Projected | Solution |
|-----------------|---------|-----------|----------|
| Fuel used (gal) | 35 | 60 | 51 |

3.3 Additional Factors

3.3.1 Environmental Impacts

Table 2 demonstrates the total emissions of carbon monoxide, nitrogen oxides, and volatile organic compounds observed currently, after the 22% projection, and after the proposed solution at Dieppe Road/North Service Road/Ramp and Niagara Street. It can be observed that the emission levels for the proposed solution are lower than the projected emissions, which was the ideal goal.

3.3.2 Social Impacts

The suggested changes at Dieppe Road/North Service Road/Ramp and Niagara Street are a combination of signal timing optimization and minor improvements. Implementing new signage and optimizing signal timings have minimal social impacts as the construction disruption would be temporary. In addition, it is also essential to consider that the signal timing optimization will result in a positive impact, in the form of reduced delay and emissions, thus improving the air quality. Moreover, it is also important to note that the changes will increase the overall functionality of the intersection, which is the goal for a positive social impact. Finally, as shown in Table 3, the fuel used at each intersection is reduced with the proposed solution as well.

3.3.3 Cost

Additional costs considered in implementing the minimal changes and optimizing the signal timings are listed in Table 4.

Table 4 Dieppe road/north service road/ramp and Niagara street—estimated cost (Canadian Automobile Association 2017; City of Toronto 2021) [11]

| Estimated cost | Information | |
|----------------|-----------------------|-----------|
| One-time cost | ONE-WAY sign | ×2 \$154 |
| | Speed limit sign | \$39 |
| | Signal timing change | \$7000 |
| | Red-light camera | \$150,000 |
| | Total cost | \$157,193 |
| Annual cost | School crossing guard | \$6370 |
| | Total cost | \$6370 |

4 Other Intersections

4.1 Geneva Street and Westchester Avenue

After applying the same evaluation process to Geneva Street and Westchester Avenue, it became apparent that the southbound left-turn lane on Geneva Street would become damaging to the intersection with its projected traffic volumes. A proposed right-turn storage lane will be added to the westbound lane to lower the volume to capacity ratio and improve traffic flow. This right-turn lane will reduce intersection delay by approximately 24%. However, the southbound left lane will remain problematic. Although the project’s scope for the team ended here, two possible solutions were proposed to combat the issue of the left-turn lane. Out of the two solutions, constructing an elevated ramp exclusively for left-turns was deemed the safest and with the most favorable long-term effects. Therefore, the team completed this particular intersection with optimization to the signal timings and minor improvements to the intersection that includes; adding MERGE sign, reducing the speed limit, and implementing bike lanes.

4.2 Carlton Street/North Service Road and Geneva Street

Finally, after analyzing Carlton Street/North Service Road and Geneva Street, the solution proposed is a geometric change in addition to signal timing optimization and minor improvements. In order to resolve the intersection level of service and delay time, a 150 m right-turn storage lane is proposed for the eastbound and westbound approaches. The addition of these storage lanes decreases the delay timing by 36%, allowing the intersection to operate smoothly even with a 22% increase in volume. Furthermore, including minor improvements such as auxiliary signal heads, red-light cameras, and necessary speed limit signs will allow this particular intersection to function at an acceptable level of service for the next ten years.

Table 5 Sensitivity analysis

| Intersection | Delay (seconds) | |
|--|-----------------|--------------|
| | 22% increase | 25% increase |
| Dieppe road/north service road/ramp and Niagara street | 38.7 | 41.8 |
| Geneva street and Westchester Avenue | 90.8 | 96.0 |
| Carlton street/north service road AND Geneva street | 79.4 | 85.8 |

5 Sensitivity Analysis

In this project, the design period for analysis was ten years. However, only the current traffic counts were available. As a result, traffic volume forecasts were created by assuming a growth factor of 2% per year for all movements. The 2% would be compounded over the 10-year design period, leading to a 22% increase in the current traffic volumes. This assumption was most significant as there is bound to be uncertainty associated with the forecasts. Therefore, a sensitivity analysis was performed to determine the impact of this uncertainty on the analysis results. The impact was investigated by increasing the current intersection volumes by 25% to determine the effect of such traffic growth on intersection delay. The results are shown in Table 5. As noted, the intersection delay did not substantially change with the increase in traffic volumes by 25%, compared to that of 22%.

6 Conclusions

To improve the traffic network in the Niagara Region, ten signalized intersections were evaluated. Based on this evaluation, the three most problematic intersections were selected for improvements. The alternative improvements to the intersections included the “do-nothing” alternative, level of service improvements, and safety countermeasures. In this study, minor improvements and signal timing optimization are recommended to improve the first and second intersections (Dieppe Road/North Service Road/Ramp and Niagara Street and Geneva Street and Westchester Avenue). In addition, the second intersection requires a significant geometric improvement beyond the scope of this study. The third intersection, Carlton Street/North Service Road and Geneva Street, includes minor changes, signal timing optimization, and geometric design changes.

The lessons learned from this study indicate that the nature of improvements to the signalized intersections in the Niagara Region depends on the intersection’s geometric and traffic characteristics and the location of the intersection within the network. For example, the left-turn movement at the second intersection involved mostly traffic that aims to enter the freeway. The LOS for this movement was F and could not be improved using traditional traffic management strategies. A separate ramp should be explored in the future to improve the level of service for this

movement. Moreover, sensitivity analysis is essential to assess the reliability of the proposed solution regarding the uncertainty associated with traffic volume forecasting. The solutions outlined in this study and the findings from this analysis should help improve the traffic network in the Niagara Region.

Acknowledgements The authors thank Farokh Iaq Kaker, Ryerson University for her assistance with this capstone project. This capstone project has been proposed by the Transportation Planning Department, Regional Municipality of Niagara.

References

1. City of Toronto (2017) Traffic signal installation. City of Toronto. <https://www.toronto.ca/services-payments/streets-parking-transportation/traffic-management/traffic-signals-street-signs/traffic-signals-in-toronto/traffic-signal-installation/>
2. De Souza A, Brennand CR, Yokoyama RS, Donato EA, Madeira E, Villas LA (2017) Traffic management systems: a classification, review, challenges, and future perspectives. *Int J Distrib Sens Netw* 13(4)
3. Izadpanah P, Thukral S, Zarei H, Hadayeghi A (2015) Safety evaluation of red light camera and intersection speed camera programs in Alberta. Presented at the 94th annual meeting of the Transportation Research Board, Paper No. 15-4704, Washington, D.C
4. Niagara Region (2016) How we grow: Niagara 2041. <https://www.niagararegion.ca/2041/pdf/mcr-pic3-boards.pdf>
5. Ontario Traffic Manual (2016) Pedestrian crossing treatments. Book 15
6. Ontario Traffic Manual (2000) Regulatory signs. Book 5
7. Retting R, Chapline J, Williams A (2002) Changes in crash risk following re-timing of traffic signal change intervals. In: *Accident analysis and prevention*, vol 34, no 2. Pergamon Press, Oxford, NY, pp 215–220
8. Tariq R, Nguyen M, Gobin N, Puthoor J, Yip M (2021) Miovision smart city intersection adoption across Niagara region. Transportation Capstone report. Department of Civil Engineering, Ryerson University, Toronto, Ontario
9. Taylor GJ (2015) Intersection geometric design. CED Engineering. <https://www.cedengineering.com/userfiles/Intersection%20Geometric%20Design-R1.pdf>
10. Transportation Association of Canada (2004) Canadian guide to in-service road safety reviews: a book in the Canadian road safety engineering handbook
11. Uline Canada (2021) Parking and traffic signs. Uline, Milton, Ontario. https://www.uline.ca/BL_8255/Parking-and-Traffic-Signs

Exploring the Use of Artificial Neural Networks for Scour Prediction



M. Marrocco, P. Williams, R. Balachandar, and R. Barron

1 Introduction

Scour around bridge piers has been a topic of interest in hydraulic engineering practice for many years. The scour process can be defined as the erosion or movement of bed material due to flowing water, and can be the result of general scour, local scour, contraction scour, or any combination of these. The present paper focuses on local scour (i.e. scour caused directly by the presence of an abutment and/or bridge pier). Scour around bridge piers can affect the stability of the structure and can ultimately cause failure. It has been established that scour and erosion are the leading cause of bridge failure in North America, with approximately 60% of bridge failures being attributed to scour and scour-related complications [21, 26, 29].

Currently, foundation design with respect to bridge pier scour is based on approved code-specified methods. These methods are typically empirical equations, developed by curve-fitting to experimental data, used to calculate the depth at which foundations should be placed to avoid a loss of lateral support due to scour. These equations tend to be limited to certain flow and geometric conditions and extrapolating to field situations should be done with caution. Recent research shows that there are several deficiencies with current prediction equations, including unnecessarily high estimates [30]. Challenges arise in developing an accurate equation with curve-fitting due to the complexity of the scour mechanism, which is further discussed in this paper. Additionally, some factors influencing scour depth, such as inter-particle behaviour in a given sediment, are difficult to quantify. Scale effects further increase the challenges of precisely estimating scour depth in the field by extrapolating laboratory based experimental results.

M. Marrocco (✉) · P. Williams · R. Balachandar · R. Barron
University of Windsor, Windsor, Canada
e-mail: marrocc1@uwindsor.ca

© Canadian Society for Civil Engineering 2022
S. Walbridge et al. (eds.), *Proceedings of the Canadian Society of Civil Engineering Annual Conference 2021*, Lecture Notes in Civil Engineering 250,
https://doi.org/10.1007/978-981-19-1065-4_13

In recent years, utilization of machine learning methods has become increasingly popular for analysis of the scour problem. A machine learning methodology that has been widely used for this application is artificial neural networks (ANN). This paper provides a review of research that has been conducted in relation to the use of ANNs in the prediction of scour depth. Various ANN models have been applied to the scour problem in an attempt to improve the accuracy of scour depth predictions. The efficacy of such models is discussed herein, and their results compared to the performance of accepted empirical formulae.

2 The Scour Process

Local scour can be defined as the erosion induced by the presence of a bridge pier or abutment. The scouring mechanism is complex and consists of downflow on the upstream surface of the pier, a horseshoe vortex around the pier, wake vortices, trailing vortices, or any combination of these. The scour process is initiated at the pier sides, when the pressure around the pier decreases in the downstream direction, as a result of flow acceleration around the sides of the pier. A ring of scoured material is then created as the scour increases in the upstream direction, until the upstream face of the pier is reached [12]. The downward pressure gradient that is created at the pier face causes an increased downward flow velocity, which induces the scouring action. The horseshoe vortex is created when this downflow curls up and around itself, getting trapped in the scour ring and causing a rapid removal of sediment [12]. The bed material continues to erode until an equilibrium scour depth (d_{se}) is reached.

The equilibrium state of local scour differs under clear-water and live-bed conditions. In clear-water conditions, the bed material is at rest and equilibrium state is reached at the point in time when bed material is no longer being removed from the scour hole, as the velocity of the circulating flow in the hole is no longer capable of doing so [6]. The equilibrium condition for clear-water scour is reached when the critical shear stress of the bed material at the bottom of the scour hole is equal to the shear stress caused by the horseshoe vortex [10]. For live-bed conditions (e.g. when bed material is transported by the approach flow), the equilibrium condition can be defined as the scour depth at which the rate of sediment transport out of the scour hole is equal to the rate of sediment transport into the hole [6].

2.1 Parameters Influencing Scour

There are many parameters that have an affect on the equilibrium scour depth around a bridge pier. Most of these parameters can be classified into three general groups: flow properties, bed sediment characteristics, and pier geometry [20]. Flow properties include density (ρ), dynamic viscosity (μ), temperature, flow depth (h), energy slope,

near-bed shear stress (τ), angle of incidence (θ) and mean flow velocity (U). Bed sediment characteristics include sediment density (ρ_s), median sediment size/diameter (d_{50}), uniformity of particle size distribution (σ_g), cohesiveness, shape factor (k), angle of repose (φ), fall velocity and critical velocity of the bed material (U_c). Pier geometry includes pier diameter (D), shape (α), surface condition, pier orientation (β) and debris accumulation. Since scour is a temporal process, time is another parameter that should be considered, but does not fit into the three aforementioned categories. A schematic of local scour around a bridge pier is shown in Fig. 1.

Typically, only those parameters which have the greatest influence on d_{se} are retained for analysis, i.e.,

$$d_{se} = f(\rho, \mu, U, U_c, h, \rho_s, d_{50}, \sigma_g, g, D, \alpha, \beta) \tag{1}$$

where g is the gravitational acceleration. Through dimensional analysis, the number of parameters can be further reduced. Typically, the equilibrium scour depth is normalized by either the flow depth or pier diameter. Using the pier diameter, the dimensionless form of the equation for equilibrium scour depth is

$$\frac{d_{se}}{D} = f\left(\frac{U}{U_c}, \frac{U}{\sqrt{gh}}, \frac{h}{D}, \frac{D}{d_{50}}, \frac{\rho U D}{\mu}\right) \tag{2}$$

In general, the input parameters utilized in developing the ANN models have been based on governing parameters previously defined in the literature. The blockage ratio, defined as the ratio of pier diameter (D) to channel width (b), is often not considered as a governing parameter since early research has suggested that blockage effects are negligible when the blockage ratio is less than 10% [6]. None of the ANNs in the studies reviewed considered the blockage ratio in the development of the prediction models. However, recent research has indicated that the blockage ratio has a discernable influence on the scour geometry, even when it is below 10% [8, 13, 31, 32]. The inclusion of blockage ratio may improve the performance of ANNs in predicting equilibrium scour depth.

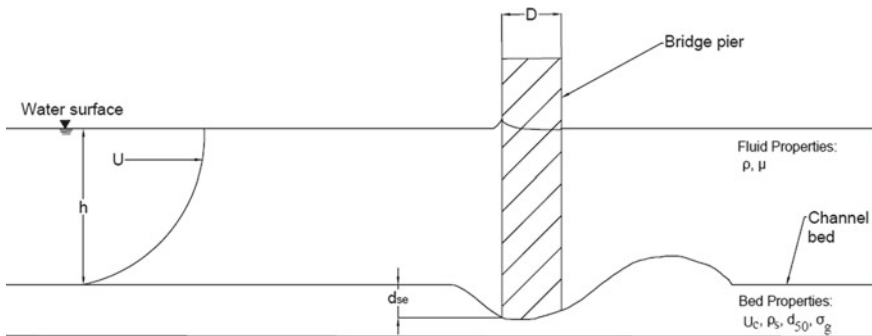


Fig. 1 Scour around a circular bridge pier

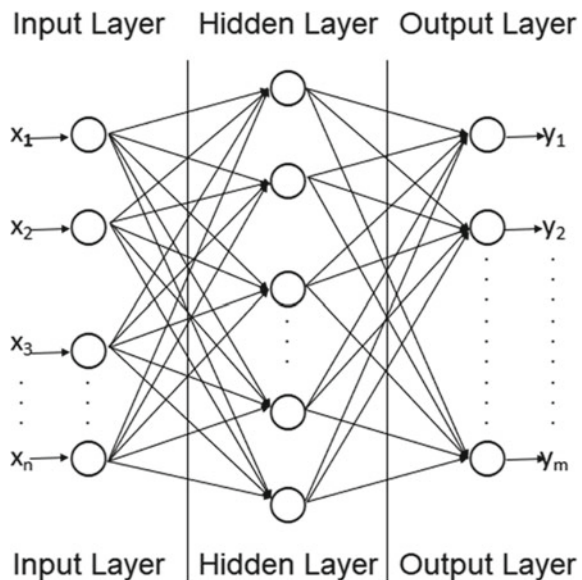
3 Artificial Neural Networks

There are many different tools in artificial intelligence (AI) and machine learning that can be applied to predict scour. These tools include, but are not limited to, ANNs, multivariate adaptive regression splines (MARS), group method of data handling (GMDH) and neuro-fuzzy models. Each of these tools attempts to find relations between selected input variables and a target output variable. This paper focuses on ANNs and their application to the scour process.

ANNs are mathematical models that are devised from an analogy to brain cells and biological neural networks. As an AI tool, ANNs can be applied to various fields and are capable of solving intricate problems such as optimization, simulation, estimation, prediction, as well as model complex problems. Three common types of ANNs are feed-forward (FF) neural networks, radial basis function (RBF) neural networks, and general regression (GR) neural networks [11]. Generally, ANN models consist of an input layer, a hidden layer or layers, and an output layer. Nodes, also called neurons, make up each of these layers, with the number of nodes varying, depending on the development of the model. Both the number of nodes in the hidden layer(s) and number of hidden layers used are typically selected through trial and error. The type of ANN most commonly applied to the scour problem is the feed-forward model.

The general schematic of an ANN with one hidden layer is illustrated in Fig. 2. Input signals are received and summed (perhaps weighted) by the nodes, the sum is then passed through an activation function and an output is generated. The output signals are then sent out to other nodes in the network. The resulting output variables

Fig. 2 General schematic of an ANN



are produced by merging the connection weights with each input node after passing through an activation function. Activation functions can be used to introduce non-linearity to the model. An example of a common activation function used for FF-ANNs is the sigmoid transfer function [Eq. 3]. On the other hand, RBF-ANNs utilize a Gaussian transfer function [Eq. 4]:

$$f(x) = \frac{1}{1 + \exp(-x)} \quad (3)$$

$$f_i(x) = \exp\left(-\frac{\|x - c_i\|^2}{2\sigma_i^2}\right) \quad (4)$$

where x is the known input vector, c_i is the centre of the radial basis function for node i , and σ_i is the width of the Gaussian function which indicates the selectivity of the neuron (i.e. controls the smoothness properties of the radial basis functions).

GR-ANNs are a variation of RBF-ANNs which do not require an iterative training process [11]. These types of ANNs consist of four layers: an input layer, pattern layer, summation layer and an output layer. The input layer is connected to the pattern layer where each neuron presents a training pattern. Each neuron in the pattern layer is then connected to two neurons in the summation layer, one which computes the sum of weighted responses and one which computes unweighted outputs of the pattern neurons. The summation and output layers then work together to normalize the output set. The hidden layers of a GR-ANN contain activation functions as in the other ANNs discussed above.

Two data sets are used in the development of a neural network, a training data set and a validating data set. Typically, the training data set is larger than the validation set, but this is not necessary. The training data set is first used by the ANN to train the network. Next, the validation data set is used to assess the performance of the network. The ANNs are trained using learning algorithms which are selected by the model developer. The performance of ANNs is typically measured using statistical quantities, such as mean absolute error (MAE) and coefficients of determination (R^2). Equations 5 and 6 show how the MAE and R^2 values are calculated, respectively.

$$\text{MAE}(\%) = \frac{\sum_{j=1}^n |y_j - y_{tj}|}{n} \times 100 \quad (5)$$

$$R^2 = 1 - \frac{\sum_{j=1}^n (y_j - y_{tj})^2}{\sum_{j=1}^n (y_j - \bar{y}_j)^2} \quad (6)$$

Here y_j is the output, y_{tj} is the target value, \bar{y}_j is the average of outputs and n is the total number of events considered.

Since the use of ANNs is a data-driven approach to solving problems, there are advantages to its use in the application of the scour problem over curve-fitting and linear regression. One main advantage is that no relationship between the input and

output variables is assumed. Furthermore, the underlying physics of the problem does not need to be known. These aspects are favourable for the scour problem because of the complexity of the scour mechanism—despite the considerable body of work in this area over the past several decades, the scour process is not fully understood.

4 Review of Applications of ANN to the Scour Problem

Many different approaches to development of ANNs for predicting equilibrium scour depth have been reported in the literature. The development of the models varies by the types of ANNs and learning algorithms applied. Most of the studies reviewed took multiple approaches in developing a model so that their performances could be compared. Some of the studies also applied different AI tools such as adaptive neuro-fuzzy inference systems (ANFIS) and took nonlinear regression (NLR) or multiple linear regression (MLR) approaches to compare their performances to that of the ANNs. However, only the ANN models are reviewed in this paper. Also, different approaches have been taken in the literature to reduce the data set, resulting in different input variables for each prediction model.

4.1 Influence of Training Data Sets

As previously mentioned, the majority of current methods of estimating scour depth were developed from curve-fitting experimental data and extending empirical equations for field prediction should be done with caution. Of the literature reviewed, many of the ANN models were developed and tested using data obtained from laboratory experiments. As such, the accuracy of these models for field use is also unknown. However, it is expected that the predicted scour values will lack the desired accuracy for use in practice, as ANNs produce optimal values when applied to data in the range from which it was trained and field data and laboratory data rarely occur in the same range.

Table 1 outlines MAE values from studies which developed models from both laboratory and field data sets. The MAE values from the models developed from field data are significantly higher than those developed from laboratory data. Of the

Table 1 MAE values for models developed from laboratory data and field data

| Source | MAE value [%] | |
|---------------------------|---------------|-------|
| | Laboratory | Field |
| Choi and Cheong [7] | 14.6 | 66.3 |
| Toth and Brandimarte [28] | 2.0 | 48.0 |
| Shamshirband et al. [24] | 2.9 | 52.0 |

literature reviewed, it was also discovered that, in general, R^2 values from field-developed models were lower than those from models developed from laboratory data. The difference in error values is most likely due to the larger variation and paucity of field data compared to laboratory data. Additionally, field data tends to be noisier than laboratory data, which can have an effect on the training of the ANNs and consequently the accuracy of their results. As indicated above, improved accuracy in ANN models developed from field data is required for use in practice.

In a study conducted by Toth and Brandimarte [28], the validation data was divided into seven subsets with increasing specificity. The first set contained all data, which was then separated into field and laboratory conditions, then separated once more into live-bed and clear-water conditions. The results of the study showed a strong improvement of the performance of the models as the specialization of the data increased. This phenomenon should be considered when comparing the performance of various models, a model trained on a larger data set, where data was obtained from multiple sources with varying conditions, may produce less accurate results than a model trained on a smaller dataset with fewer variations.

4.2 Learning Algorithms

ANNs require selection of a learning algorithm to train the model. Review of the literature found that the Levenberg–Marquardt (LM) algorithm, back propagation algorithm (BP), particle swarm optimization algorithm (PSO) and Firefly (FA) algorithm have been applied to train ANNs for the scour problem. The BP and LM algorithms were more commonly used than the PSO and FA algorithms, as they are implemented in the MATLAB neural network toolbox. Figure 3 is based on validation data and shows the predicted versus measured or actual scour depth values from the different learning algorithms applied to laboratory scour data. The 45° line indicates a perfect match between predicted and observed values. The other two solid lines show the 20% error bounds. As can be seen in the plots, each of the learning algorithms generate relatively accurate scour predictions, with most of the data falling in the 20% error range.

The coefficients of determination were averaged from the reviewed literature to determine the average R^2 value for each learning algorithm. The PSO algorithm yielded the highest R^2 value of 0.936, while the BP algorithm yielded the least with an R^2 value of 0.907. The LM and FA algorithms yielded R^2 values of 0.908 and 0.918, respectively. These values suggest that the PSO algorithm is most optimal for training ANNs for scour prediction, however the values do not vary significantly. These values agree with the study completed by Dang et al. [9], who compared the use of the LM algorithm to the PSO and FA algorithms and found that the model trained by the PSO algorithm was the most accurate. However, since the BP and LM algorithms have been applied to more models than the PSO algorithm, and the PSO algorithm is relatively new to the scour problem, there is more data for the applicability and accuracy of BP and LM algorithms. Also, it is important to note

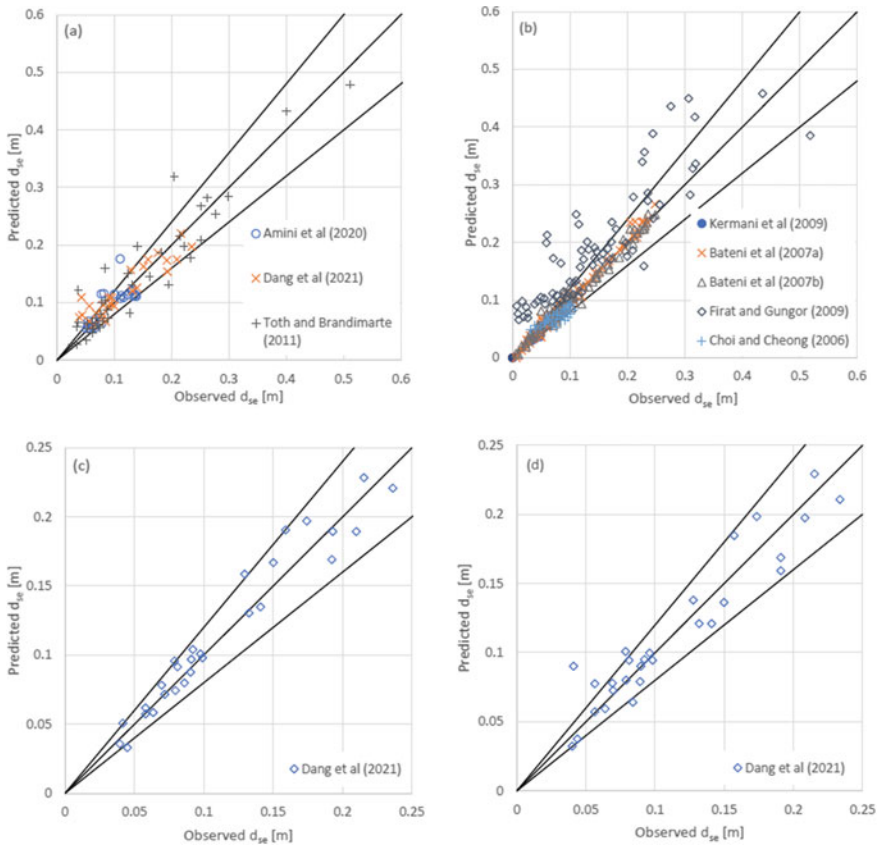


Fig. 3 Predicted versus observed equilibrium scour for **a** LM algorithm, **b** BP algorithm, **c** PSO algorithm, and **d** FA algorithm

that each of the literature models used their own dataset, making direct comparison of error values difficult.

4.3 Type of ANN

Figure 4 displays the plots of the predicted versus observed equilibrium scour depths for different ANN types developed from laboratory data, based on validation data. These plots can be compared to those in Fig. 3, which show the predicted versus observed scour depth plots for FF-ANNs. As can be seen from these plots, the GR-ANNs and RBF-ANNs have a higher percentage of data points outside of the 20% error range. When the R^2 values for the different types of ANNs are compared, the FF-ANNs yields the highest value of 0.924, while the GR-ANN and RBF-ANN yield

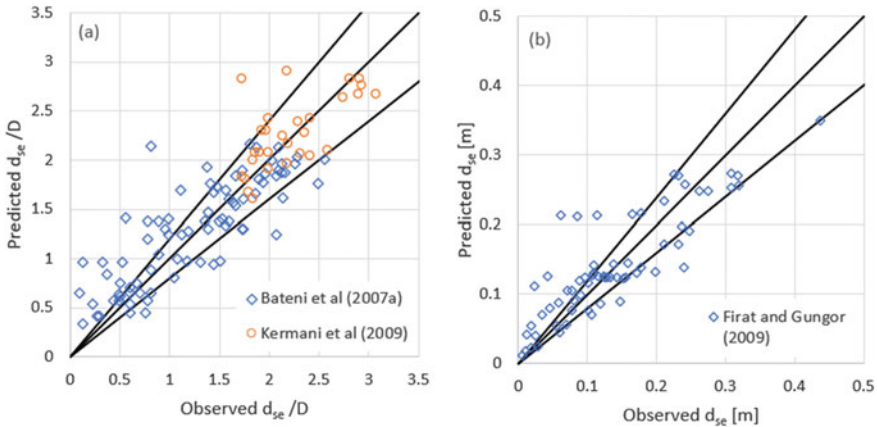


Fig. 4 Predicted versus observed equilibrium scour for **a** an RBF-ANN and **b** an GR-ANN

significantly lower values of 0.781 and 0.788, respectively. These results suggest that the FF-ANN is the most optimal type of ANN to be applied to the scour problem. Studies completed by Kermani et al. [16] and Bateni et al. [2] support these results, as the results of their studies suggest that FF-ANNs provide more accurate results than RBF-ANNs. However, the study completed by Firat and Gungor [11] contradicts these results, as it was concluded that the GR-ANNs provided more accurate results than the FF-ANN. As displayed in Fig. 3b, a significant number of data falls outside of the 20% error range for the FF-ANN developed by Firat and Gungor [11], which is not the case for the other models. This suggests that the model presented by Firat and Gungor [11] may not be an accurate representation of the performance of FF-ANNs in the application of scour. Since no other study that was reviewed developed a GR-ANN, there is no further data from the current study supporting or contradicting these results.

In a recent study, conducted by Pal [22], deep neural networks (DNN) were evaluated in the application of predicting scour around a bridge pier. DNNs can be defined as a neural network with “two or more hidden layers having a large number of nodes and using sophisticated mathematical modelling” [22]. The results of the DNN were evaluated against the predictions of an FF-ANN trained with a BP algorithm. The results of this study indicate that the predictions from the DNN were more accurate than the FF-ANN, with coefficients of determination of 0.925 and 0.878, respectively.

4.4 Comparison to Existing Formulae

The performance of the ANNs were also compared to various accepted empirical scour prediction methods. Results from the literature [1, 3, 11, 16, 18, 22, 24, 27]

Table 2 Values of average R^2 for different prediction models

| Model | R^2 |
|--------------------------------|-------|
| ANN | 0.883 |
| Breusers et al. [5] | 0.203 |
| Richardson et al. [23], HEC-18 | 0.518 |
| Johnson [14] | 0.729 |
| Laursen and Toch [17] | 0.654 |
| Melville and Chiew [19] | 0.515 |
| Shen [25] | 0.486 |
| Williams et al. [31] | 0.718 |

was reviewed and the average R^2 values were computed for those formulae which have been evaluated by more than one study. An equation proposed by Williams et al. [31] was also evaluated. A comparison of the average error values from these studies can be found in Table 2. In all of the literature reviewed, the ANN models produced more accurate results than the empirical methods. This solidifies the proposal that the application of ANNs in scour prediction should be further investigated for the use of foundation design.

A summary of the literature reviewed for this study can be found in Table 3. The table includes information on the type(s) of data used for development, the type(s) of ANNs used, the learning algorithm(s) (LA) used, as well as any other models that were evaluated in comparison. The optimal model from each study, along with the coefficient of determination (R^2) from that model, is also presented.

5 Conclusions and Recommendations

The current methods for predicting the equilibrium scour depth for bridge pier foundation design have various shortcomings. Literature was reviewed to explore the application of ANNs in the scour problem. Statistical analysis found that learning models developed from laboratory data were more accurate than those developed from field data. This is logical since the laboratory experiments were conducted under well controlled conditions and the data tended to show less variation. However, models developed from laboratory data may not be applicable in field conditions. Further analysis of different learning algorithms found that models trained using the PSO algorithm yielded the most optimal results. Comparison of different types of ANNs resulted in some discrepancies. Overall, the FF-ANNs seemed to produce the most accurate results. Further investigation into GR-ANNs is recommended to confirm the results presented by Firat and Gungor [11]. Pal [22] also introduced DNNs to the scour problem, which produced more accurate results than the FF-ANN. Since the PSO algorithm and DNNs are relatively new to the application of scour prediction, further investigation into their applicability and accuracy is recommended. Statistical

Table 3 Summary of literature reviewed

| Source | Number of data | | ANN type | | | LA ^a | Other models | Compared to formulae | Optimal model | R ² |
|---------------------------|----------------|-------|----------|-----|----|-----------------|---------------------|----------------------|---------------|----------------|
| | Lab | Field | FF | RBF | GR | | | | | |
| Shamshirband et al. [24] | 552 | 540 | ✓ | | | PSO | - | ✓ | ANN | 0.897 |
| Choi and Cheong [7] | 97 | 212 | ✓ | | | BP | - | ✓ | ANN | - |
| Lee et al. [18] | - | 380 | ✓ | | | BP | - | ✓ | ANN | 0.914 |
| Toth and Brandimarte [28] | 331 | 215 | ✓ | | | LM | - | ✓ | ANN | - |
| Bonakdari and Ebtahaj [4] | - | 467 | ✓ | | | BP | ANFIS, NLR | | ANFIS | - |
| Pal [22] | - | 232 | ✓ | | | BP | DNN | ✓ | DNN | 0.925 |
| Amini et al. [1] | b | - | ✓ | | | LM | - | ✓ | ANN | 0.950 |
| Dang et al. [9] | 192 | - | ✓ | | | LM, PSO, FA | - | | ANN (PSO) | 0.974 |
| Siahkali et al. [27] | 246 | - | ✓ | | | BP | GMDH, MARS, Kriging | ✓ | Kriging | 0.990 |
| Kaya [15] | - | 380 | ✓ | | | LM | - | | | 0.858 |
| Firat and Gungor [11] | 165 | - | ✓ | | ✓ | BP | MLR | ✓ | GR-ANN | 0.866 |
| Batani et al. [3] | 263 | - | ✓ | | | BP | - | ✓ | ANN | 0.979 |
| Batani et al. [2] | 263 | - | ✓ | ✓ | | BP | ANFIS | ✓ | FF-ANN | 0.988 |
| Kermani et al. [16] | 124 | - | ✓ | ✓ | | BP | ANFIS | ✓ | FF-ANN | 0.933 |

^a Learning algorithm(s)

^b Number of data points not provided

analysis also found that ANN models are capable of producing more accurate scour depth predictions than current accepted empirical methods. Future research into the use of ANNs to predict scour around bridge piers should include blockage ratio as a governing parameter, as studies have shown that blockage effects influence scour geometry.

References

1. Amini A, Hamidi S, Shirzadi A, Behmanesh J, Akib S (2020) Efficiency of artificial neural networks in determining scour depth at composite bridge piers. *Int J River Basin Manag* 1–7
2. Bateni S, Borghei S, Jeng D (2007) Neural network and neuro-fuzzy assessments for scour depth around bridge piers. *Eng Appl Artif Intell* 20(3):401–414
3. Bateni S, Jeng D, Melville B (2007) Bayesian neural networks for prediction of equilibrium and time-dependent scour depth around bridge piers. *Adv Eng Softw* 38(2):102–111
4. Bonakdari H, Ebtehaj I (2017) Scour depth prediction around bridge piers using neuro-fuzzy and neural network approaches. *Int J Civ Environ Eng* 11(6):835–839
5. Breusers H, Nicollet G, Shen H (1977) Local scour around cylindrical piers. *J Hydraul Res* 15(3):211–252
6. Chiew Y (1984) Local scour at bridge piers. Ph.D. thesis, Auckland University, Auckland, New Zealand
7. Choi S, Cheong S (2006) Prediction of local scour around bridge piers using artificial neural networks. *J Am Water Resour Assoc* 42(2):487–494
8. D'Alessandro C (2013) Effect of blockage on cylindrical bridge pier local scour. M.A.Sc. thesis, University of Windsor, Windsor, Canada
9. Dang N, Anh D, Dang T (2021) ANN optimized by PSO and firefly algorithms for predicting scour depths around bridge piers. *Eng Comput* 37(1):293–303
10. Deng L, Cai C (2010) Bridge scour: prediction, modeling, monitoring, and countermeasures—review. *Pract Period Struct Des Constr* 15(2):125–134
11. Firat M, Gungor M (2009) Generalized regression neural networks and feed forward neural networks for prediction of scour depth around bridge piers. *Adv Eng Softw* 40(8):731–737
12. Guo J (2012) Pier scour in clear water for sediment mixtures. *J Hydraul Res* 50(1):18–27
13. Hodi B (2009) Effect of blockage and densimetric Froude number on circular bridge pier local scour. M.A.Sc. thesis, University of Windsor, Windsor, Canada
14. Johnson P (1992) Reliability-based pier scour engineering. *J Hydraul Eng* 118(10):1334–1358
15. Kaya A (2010) Artificial neural network study of observed pattern of scour depth around bridge piers. *Comput Geotech* 37(3):413–418
16. Kermani M, Beheshti A, Ashtiani B, Yazdi S (2009) Estimation of current-induced scour depth around pile groups using neural network and adaptive neuro-fuzzy inference system. *Appl Soft Comput* 9(2):746–755
17. Laursen E, Toch A (1956) Scour around bridge piers and abutments. Iowa Highway Research Board, Ames IA
18. Lee T, Jeng D, Zhang G, Hong J (2007) Neural network modeling for estimation of scour depth around bridge piers. *J Hydrodyn* 19(3):378–386
19. Melville B, Chiew Y (1999) Time scale for local scour at bridge piers. *J Hydraul Eng* 125(1):59–65
20. Melville B, Coleman S (2000) Bridge scour. Water Resources Publications, Highlands Ranch, CO, USA
21. Miroff N (2007) Collapse spotlights weaknesses in U.S. infrastructure, August 3, 2007. <http://www.washingtonpost.com/wp-dyn/content/article/2007/08/02/AR2007080202410.html>

22. Pal M (2019) Deep neural network based pier scour modeling. *ISH J Hydraul Eng* 28 (sup1):80–85
23. Richardson E, Harrison L, Richardson J, Davis S (1993) Evaluating scour at bridges (No. HEC 18), 2nd edn. Administration (FHWA), Washington, D.C.
24. Shamshirband S, Mosavi A, Rabczuk T (2020) Particle swarm optimization model to predict scour depth around a bridge pier. *Front Struct Civ Eng* 14(4):855–866
25. Shen H (1971) Scour near piers. In: *River mechanics*, vol II, chap 23. Fort Collins, CO, USA
26. Shirhole A, Holt R (1991) Planning for a comprehensive bridge safety program. *Transp Res Rec* 1290:39–50
27. Siahkali M, Ghaderi A, Bahrpeyma A, Rashki M, Hamzehkolaei N (2021) Estimating pier scour depth: comparison of empirical formulations with ANNs, GMDH, MARS, and kriging. *J Artif Intell Data Min* 9:109–128
28. Toth E, Brandimarte L (2011) Prediction of local scour depth at bridge piers under clear-water and live-bed conditions: comparison of literature formulae and artificial neural networks. *J Hydroinf* 13(4):812–824
29. Wardhana K, Hadipriono F (2003) Analysis of recent bridge failures in the United States. *J Perform Constr Facil* 17(3):144–150
30. Williams P, Balachandar R, Bolisetti T (2013) Evaluation of local bridge pier scour depth estimation methods. In: *24th Canadian congress of applied mechanics*, Saskatoon, SK, Canada, vol 1, pp 2–6
31. Williams P, Bolisetti T, Balachandar R (2018) Blockage correction for pier scour experiments. *Can J Civ Eng* 45(5):413–417
32. Williams P (2019) The role of approach flow and blockage on local scour around circular cylinders with and without countermeasures. Ph.D. thesis, University of Windsor, Windsor, Canada

Valuing Demand Response on Electric Vehicles Charging



H. Yang and Y. H. Kim

1 Introduction

Due to the concerns of the greenhouse effect caused by fossil fuels, generating electricity from renewable energy becomes an ideal alternative to reduce emissions to the environment. Moreover, renewable energy can diversify electricity facilities' energy supply and reducing dependence on conventional fossil fuels. Currently, renewable energy sources provide about 18.9% of Canada's total primary energy supply [17]. Wind energy, one of the fastest-growing sources of electricity in Canada [17], has become a new contributor to decrease the power systems' marginal cost. In 2019, Canada's installed capacity of wind generation had reached 13,413 MW, and more than 300 wind farms are operating across Canada [7].

However, the renewable energy supply fluctuations raise a challenge for electric facilities to match real-time supply and demand. Generating electricity by wind turbines is not continuously available during the whole day, and greatly affected by the weather condition. When the wind speed is not sufficient to rotate the shaft, the renewable energy supply will be low, and utility operators must assign electricity generated from conventional fossil energy at a higher cost to meet the demand. On the other hand, the excess supply of wind energy leads to a decrease in the wholesale electricity price due to the merit-order effect [11].

Thus, electricity facilities have developed many strategies for the supply fluctuations of varying timescales, e.g., energy storage, demand response, and bilateral contract. Electric vehicles (EVs) connected with an electric grid can adopt these

H. Yang · Y. H. Kim (✉)

Department of Civil and Environmental Engineering, University of Windsor, Ontario N9B 3P4, Canada

e-mail: kim523@uwindsor.ca

H. Yang

e-mail: yang112u@uwindsor.ca

© Canadian Society for Civil Engineering 2022

S. Walbridge et al. (eds.), *Proceedings of the Canadian Society of Civil Engineering*

Annual Conference 2021, Lecture Notes in Civil Engineering 250,

https://doi.org/10.1007/978-981-19-1065-4_14

strategies to flat the supply fluctuations; nowadays, the number of electric vehicle sales in Canada keeps steady growth. Many studies have investigated the integration of EVs into power systems without increasing the grid stress. Borba et al. proposed to maximize the integration of wind energy supply in power grid through scheduling plug-in hybrid electric vehicles (PHEVs) charging times [6]. Bellekom et al.'s study showed that implementing EVs into power systems with load management can decrease the excess in wind electricity production [3]. Wang et al. analyzed the interactions among PHEVs, wind power, and demand response, concluding that the operating cost of the electricity system can be significantly reduced [23].

While these studies focus on the potential impact of EVs participating in DR programs, few studies discuss what are proper incentives for EV users shifting their behaviours of charging. According to Luft [14], the reward cause more considerable improvement for habit development relative to punishment. Thus, a DR program that customers are rewarded with appropriate incentives will have a greater performance in shaving the excess wind energy at the supply peak. Conversely, if the cost of the DR program for renewable energy is over than its benefit, electricity utilities have no motivation to introduce or expand the DR program to EV users. Therefore, quantifying the value of DR can better encourage more EV users involved in the DR program, and help the utility operators manage the power system more efficiently.

Different DR programs can be classified into two main categories: incentive-based programs and price-based programs [1]. Incentive-based programs usually offered to large industrial or commercial customers, provide incentives to targeted customers that adjust their power demand responding to the demand requests. Price-based programs vary the electricity price over time so that the end-use customers can change their electric usage at times of low unit prices.

In this study, we propose offering additional incentives through a put option to EVs users that participating in RTP, which can achieve valley filling at the renewable energy supply peak. In addition, we value three different price-based DR programs on EVs charging utilizing the historical data of HOEP. Results show the proposed DR program can significantly decrease EVs' overall cost and mitigate the price volatility.

Section 2 introduces the electricity market and DR program currently existing in Ontario, Canada. Then, Sect. 3 discusses the price-based DR programs and outlines the optimal EV charging strategies under different DR programs. Section 4 presents the detail of the proposed DR program, while Sect. 5 provides a numerical study on the Ontario electricity market. Lastly, Sect. 6 concludes with the finding.

2 Electricity Market

Similar to other commodities, electricity is bought and sold in the electricity market. Participants in the electricity market, including operators, producers, and wholesale consumers, help make pricing competitive. The electricity market cost-effectively aligns the electricity production with the need for electricity, which varies throughout the day. To achieve the real-time balance, the electricity system operators forecast

the expected demand for the days ahead, then producers can submit offers to supply and wholesale consumers bids on the consumption based on the forecast. The lowest cost to produce electricity offer will be accepted first to fulfill the forecast demand; hence, the renewable energy sources, which marginal costs are close to zero, are more frequently to meet demand than conventional fossil fuels. However, the operators have to address the excess/lack of electricity supplement in the real-time market. When the supply drops below the demand, flexible forms of power generation (more expensive start-up cost than non-flexible ones) are brought online to meet demand, causing the wholesale price to rise considerably; on the other hand, the electricity price will decrease when the power supply is higher than demand, and negative prices might occur under this circumstance. The decreasing price represents that it is hardly possible to reduce the supply from a technical perspective or involves relatively high shutdown costs for the producers [20]. Along with increased renewable energy source shares, the mismatch between supply and demand will occur more frequently. To address the balance concern, electricity utilities have offered a wide range of DR programs to the consumers that provide them an opportunity to reduce or shift their electricity usage during peak periods in response to the financial incentives [22].

In Ontario electricity market, both incentive-based programs and price-based programs are being used to actively engage end-users in adjusting their consumption in response to prices and system needs. The Independent Electricity System Operator's (IESO) capacity auction is one of the incentive-based programs, which allows DR program participants to submit bids including clearing prices and quantities, receive payments after satisfying their capacity reduction obligations. Also, the IESO, similar to other power operators, such as the California Independent System Operator (CAISO), provides ancillary services to help ensure the reliable operation of the power system easing the renewables integration [2, 15]. Price-based programs, such as time-of-use pricing (TOU) or real-time pricing (RTP), also play an important role in the operation of the electric grid to meet Ontario's reliability needs in a cost-effective way. Under TOU pricing, consumers (residents or small businesses) pay higher prices during the on-peak period, while the price is cheaper at mid-peak and cheapest during the off-peak. Mid-sized and large businesses are charge under real-time pricing (RTP), which is called Hourly Ontario Energy Price (HOEP). The HOEP is the average of the twelve market clearing prices set in each hour, changing hourly based on the real-time supply and demand availability [15]. In the next section, we will discuss the price-based programs in detail.

3 Price-Based DR Programs

Both electricity utilities and customers can receive benefits from DR programs. From the electricity utility perspectives, DR programs have been considered as a cost-effective resource to reduce the power system peak and defer the need for increasing the generating capacity; thus, the valuation of DR program can be estimated from cost-saving aspects: avoided generation capacity cost, transmission cost, ancillary

services costs, energy cost and so on [10]. As customers participate in DR programs, especially EV users, the major benefit is the significant savings on the electricity bills. Therefore, the intrinsic value for EV users changing their charging periods can be directly measured by the charging cost-saving comparison. Before evaluating these DR programs, it needs to be identified that EV users are less likely to participate the incentive-based programs because (1) there is some discrepancy between individual charging behaviour of EVs, (2) an individual EV user's load is small to reach the minimum load curtailment requirement of incentive-based programs; (3) EV user might have resistance to inconvenience resulting from unscheduled shifting loads and/or privacy issues [24] therefore, only price-based programs will be discussed in this study.

3.1 Time of Use Pricing

In Ontario, residents or small business consumers pay TOU prices facing on-peak (11 a.m. to 5 p.m. on weekdays), mid-peak (7 a.m. to 11 a.m. and 5 p.m. to 7 p.m. on weekdays), and off-peak (7 p.m. to the second day 7 a.m. on weekdays) demand charges that reflect the cost of generating electricity at different times of the day.¹ These prices are set by the Ontario Energy Board (OEB), which can fluctuate during the summer and wintertime. Since the majority of Ontario residents are currently enrolled in TOU pricing, TOU pricing can be treated as the baseline when comparing with other DR programs. Under TOU pricing, the best strategy for the EVs to minimize the electricity bill is to shift the entire charging process to off-peak hours.

$$\min Z(T) = y_{off} \cdot T \cdot PW \quad (1)$$

where $Z(T)$ is the total cost of charging, in cent/kW, y_{off} is the hourly off-peak price, in cent/kWh, T is the total charging time, in hour, and PW represents the EVs charging power, in unit of kW.

As the OEB only regulates the price plan bi-annually, the main drawback of TOU pricing is unable to capture the hourly variation in the demand and supply; thus, it only provides limited support for renewable energy integration [21]. In other words, TOU pricing has a minor impact on shaving the surplus wind energy supply.

¹ We ignore the current special case: Starting from January 1, 2020, the Government of Ontario is currently providing a temporary COVID-19 Recovery Rate: a flat price of 8.5 cents per kWh, 24 h a day, 7 days a week.

3.2 Real-Time Pricing

Under the RTP, large customers who participated in the market are charged the real-time electricity price on an hourly basis (in form of HOEP in Ontario). As HOEP is determined hourly by IESO, it reflects more accurately the cost of producing electricity based on supply and demand than TOU pricing due to a shorter adjusting period. Although the target client of RTP pricing is large businesses, there is still a possibility to provide individual customers electricity at wholesale price when the number of participating is large enough. Boisvert et al.'s study indicated that customers, who are getting involved in the New York wholesale market, have contributed to reducing market prices and emergency demand response program (EDRP) events [4]. For those electricity consumers in Ontario paying the wholesale price, the cost of purchasing electricity during the supply peak is significantly decreased as the wholesale price can be zero or even negative, e.g., the price at 7–8 a.m. on April 13, 2019, is 0 cent per kWh, and it drops below zero at the same day, –1.32 cents per kWh between 10 and 11 a.m. By subscribing to the RTP program, the real-time price can be a signal for consumers shifting their energy demand to meet the supply peak, which can contribute to integrate renewable energy sources into the power system. Currently, customers must rely on historical or forecast electricity prices to schedule their usages [21]. Also, to avoid the dramatic high energy price, the EVs users must keep monitoring the hourly-ahead price through certain smart technology when charging the EVs.

$$\min Z(T) = \sum_{i=1}^n x_n \cdot y_n \quad (2)$$

where

$$\begin{aligned} x_1 + x_2 + \cdots x_{n-1} + x_n &= 1 \\ x_1 &= x_2 = \cdots x_{n-1} \end{aligned}$$

For the charging period, T , it is charged at equal fractions from $t = 1$ to $t = n - 1$ hours, and x_n is the remaining fractions of the total EV batteries be fully charged after $n - 1$ hours, y_n is the real-time electricity rate (HOEP) at hour n , in cent/kWh.

4 Real-Time Pricing with Option

The electricity producers can hedge their financial risk of production through trading electricity options in the market [18]. An option is a contract that allows the contract holder to have the right to buy/sell a specified amount of asset at the maturity date at a fixed price called the strike price, while a forward contract required the holder

to buy/sell the asset regardless of their willingness. In such a DR program, it is incentivized with options to EVs users in addition to the real-time electricity rate. The producers purchase a put option 24 h in advance based on the day-ahead forecasting prices that allow them to sell a certain amount of electricity at the strike price during the delivery period, as illustrated in Figs. 1 and 2. Since renewable energy resources have great uncertainty and intermittency, significant deviations might exist between day-ahead scheduling and real-time delivery of electricity. At the time of maturity (1 h before physical delivery), the demand peak occurs causing the price inflation, then the producers choose to sell the electricity without exercising the option; meanwhile, the consumers receive the price increment signal so that they can decrease their energy demand to reduce the bill. On the other hand, it is expected that the producers exercise the option during the supply peak period. In that case, the consumers will follow the

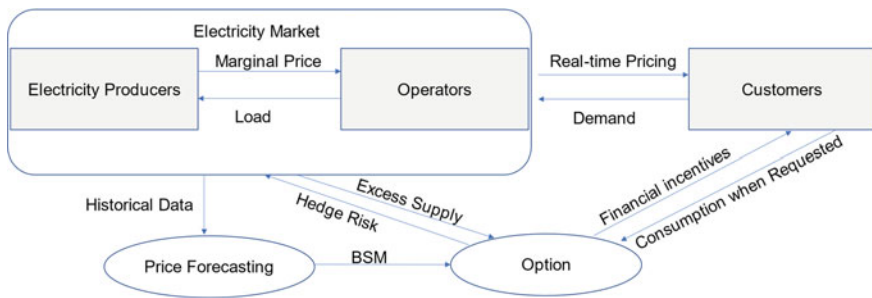


Fig. 1 Structure of proposed DR program

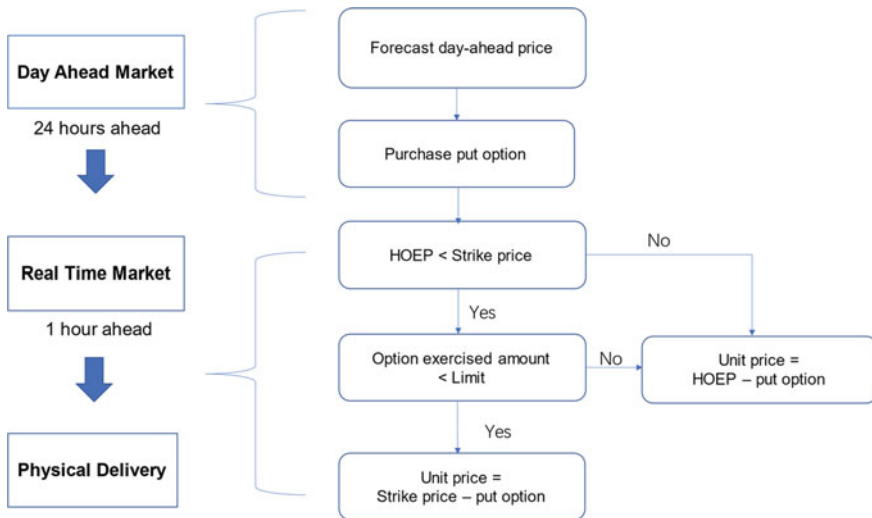


Fig. 2 Timeline for pricing the electricity price in the proposed DR

dispatching request to consume electricity at the supply peak, receive the benefit from the relatively low electricity pricing, and the option fee payment from the producers. Thus, the electricity utilities can hedge the price and production-availability risks with a properly constructed put option combined with RTP, while the customers can get appropriate incentives from the option payment P_n and lower risk involving the real-time market.

$$\min Z(T) = \sum_{i=1}^n (x_n \cdot z_n - P_n) \quad (3)$$

$$z_n = \begin{cases} \max(y_n, K) & ex < E_{tot} \\ y_n & ex \geq E_{tot} \end{cases} \quad (4)$$

where

$$x_1 + x_2 + \dots + x_{n-1} + x_n = 1$$

$$x_1 = x_2 = \dots = x_{n-1}$$

Similar to RTP, the EV is charged at equal fractions until $t = n - 1$ hour, K is the fixed electricity rate (strike price) through the put option, and z_n represents the actual electricity rate paid by EV customers at hour n . If the amount of put options already exercised, ex , is less than the upper put option exercise limit amount, E_{tot} , and the real-time electricity rate y_n is below the strike price K , the producers exercise the option of selling the electricity at the strike price K ; oppositely, if the real-time electricity rate y_n is above the strike price K , or the amount of put options already exercised, ex , is greater than the exercise limit amount, E_{tot} , the put option will be expired, and the real-time electricity rate becomes the electricity hourly wholesale price y_n . Considering the capacity of the EVs' batteries, the put option can not be exercised more than once per day. The value of the put option follows the Black-Scholes-Merton (BSM) formula, given by:

$$P_n = K e^{-rt_m} N(d_2) - S_t N(d_1) \quad (5)$$

where

$$d_1 = \frac{In \frac{S_t}{K} + \left(r + \frac{\sigma^2}{2}\right)t_m}{\sigma \sqrt{t_m}}$$

$$d_2 = d_1 - \sigma \sqrt{t_m}$$

r is the risk-free interest rate, t_m is the maturity date, in terms of years, σ is the implied volatility of the electricity, $N(\cdot)$ represents the normal distribution function.

5 Numerical Study

Different from other commodities, electricity cannot be stored efficiently like conventional fossil fuels, which need to balance production and consumption all the time. Kluge [12] indicated that the inelasticity of demand and supply makes electricity prices very sensitive to extreme events. When the penetration of renewable energy in the grid increases, the electricity price will be more volatile due to the uncertainty and intermittency of renewable energy resources. Therefore, it is essential to construct a price forecasting model for the electricity market so that the option incentives can be accurately valued, and the electricity facilities can incorporate them into the DR strategies. There are two different ways to estimate the future price in the market. One is to determine the day-ahead market price through matching bidding offers at the supply and demand equilibrium price, which is used by the electricity producers. The second approach is based on historical data with statistical techniques. We will follow the second approach to forecast the day-ahead electricity price since the statistical techniques allow us to be clear about the impact of exogenous factors [25], such as total consumption.

The generalized autoregressive conditional heteroskedasticity (GARCH) model has shown a great predictive performance when the data exhibits time-varying volatility and volatility clustering. Many studies have explored applying the GARCH model to predict electricity prices. Garcia et al. [9] proposed a GARCH methodology to forecast hourly prices in the deregulated electricity markets of Spain and California, and average forecast errors are around 9%. Liu and Shi's research [13] concluded that autoregressive moving average (ARMA) models with GARCH processes are in general an effective tool for modelling and forecasting the mean and volatility of electricity prices. Thus, this paper forecasts the day-ahead electricity price based on the AR-GARCH model.

5.1 Market Data

The data [15], which the analysis presented, covers the 5-year period from January 1st, 2015 to December 31st, 2019. It consists of 43,824 observations of the Hourly Ontario Energy Price (HOEP), which represents the wholesale price of electricity in the real-time market. Normally, a data point that exceeds three times the standard deviation of the sample is considered as an outlier representing significantly different from other observations. The outliers in the data, which can be detected through Table 1, should be replaced with the mean of the respective week before feeding into the models. Moreover, the data are divided into two parts: one part (January 1st, 2015 to December 31st, 2018, 35,064 observations) is used for model parameter estimation, and the other part (January 1st, 2019 to December 31st, 2019, 8760 observations) is used for forecasting and simulation.

Table 1 Descriptive statistics for HOEP 2015–2019 (Unit: cent/kWh)

| Year | Observations | Mean | Max | Min | Median | Std. dev. | Kurtosis | Skewness |
|------|--------------|-------|---------|--------|--------|-----------|----------|----------|
| 2015 | 8760 | 21.24 | 584.56 | -22.42 | 17.44 | 27.10 | 118.0 | 7.78 |
| 2016 | 8784 | 14.89 | 1619.60 | -9.82 | 10.38 | 32.16 | 1088.7 | 26.0 |
| 2017 | 8760 | 13.73 | 847.51 | -67.08 | 7.36 | 24.73 | 200.3 | 9.17 |
| 2018 | 8760 | 22.44 | 365.64 | -4.39 | 17.05 | 22.13 | 28.4 | 3.10 |
| 2019 | 8760 | 16.72 | 1028.52 | -59.29 | 14.36 | 23.55 | 716.6 | 18.8 |

5.2 Assumption and Setup

In the numerical simulation, the consumer price index (CPI) effect on purchasing electricity is neglected in the numerical simulation, i.e., the purchasing power of one dollar in 2019 is the same as one dollar in 2015. No friction exists when purchasing electricity in the market, i.e., there is no global adjustment fee or grid congestion fee. Also, there is no black swan (unpredictable) event that significantly affects the consumption of electricity, for example, health pandemic. For EVs charging under RTP, the current technology can not ensure finding a daily optimal charging time to minimize the cost among the charging day; therefore, it is assumed that EVs will be charged based on an optimal charging time of the same day in the previous year. Also, for EVs charging under RTP with option, the put options are exercised at the first hour period among the charging day that the real-time electricity rate y_n is below the strike price K ; moreover, the remaining portion will be charged at the daily optimal charging time in the previous year. According to Plötz et al. [19], we assume that the daily commute distance of EVs is 35 km/day that requires charging 1.62 h of charging every day, and the charging power is 3.6 kW, the annual limit of the option exercised, E_{tot} , equals 360.

5.3 Price Modelling and Forecast

Firstly, the data examined in the Engle's test [8] confirms that ARCH effects are presented. The null hypothesis of no ARCH effects in the residuals can be rejected with a highly significant test statistic of 6846 (<0.0001). In other words, an AR (m)-GARCH (p, q) model can be used to fit the data, which includes an autoregressive process of order m , and GARCH error components. The GARCH error components include a GARCH polynomial with a degree of p and an ARCH polynomial with a degree of q . The AR process can be defined as:

$$y_t = \mu + \sum_{i=1}^m \phi_i y_{t-i} + \xi_t \quad (6)$$

Table 2 The parameter estimates for AR (1)-GARCH (1,1) model

| Variable | Estimate value | Standard error |
|------------|----------------|----------------|
| μ | 0.2695 | 0.0278 |
| ϕ_1 | 0.9642 | 0.0014 |
| ω | 11.004 | 0.3451 |
| α_1 | 0.6706 | 0.0220 |
| β_1 | 0.3295 | 0.0070 |

and the GARCH error components, first proposed by Bellerose [5], can be represented as:

$$\xi_t = \sqrt{h_t} z_t \quad (7)$$

$$h_t = \omega + \sum_{i=1}^p \alpha_i h_{t-i} + \sum_{j=1}^q \beta_j \xi_{t-j}^2 \quad (8)$$

where y_t is the electricity price at time t , h_t denotes its conditional variance, z_t stands for $N(0,1)$, ω is the long-run variance, ξ is the error term, and α , β are the additional coefficients to set up a stationary model, which fulfill the requirement $\alpha + \beta < 1$ and $\alpha > 0$, $\beta > 0$ (Table 2).

After feeding in the processed market data, the parameters of the AR (1)-GARCH (1,1) model can be estimated through the least-squares estimation (for AR order) and maximum likelihood estimation (for GARCH order) [16]. It can provide a future insight into the hourly price and the conditional variance of electricity. Based on the day-ahead forecast price and predicted volatility, the value of the option can be established followed the BSM formula.

5.4 Result and Discussion

Based on Sect. 5.2, we evaluate three price-based DR programs discussed through the annual cost of charging for one EV, and the result is shown in Fig. 3. As Eq. (1) indicated, customers under TOU are charged at fixed electricity rate regardless of demand/supply peak; meanwhile, the daily cost of RTP and RTP with option keep fluctuating. From Fig. 3, in March and April 2019, the daily cost of RTP is significantly higher than RTP with option, which represents that the demand peak occurs causing RTP increases. In such a scenario, the electricity operators exercise the put option so that EVs users can shift their load from high demand to low demand period. However, the situation is reversed in May and June that EVs customers under RTP with option pay much more than under RTP. The main reason that contributed to the situation is that the excess wind energy cause price deflation. At renewable energy supply peak, electricity utilities exercise the put option to lock in their profits when

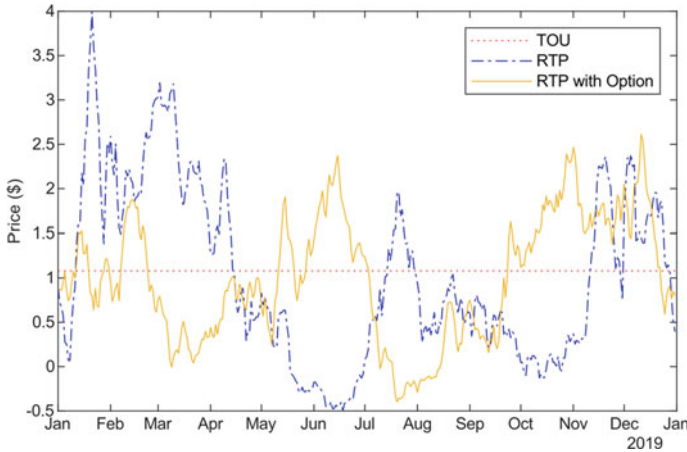


Fig. 3 Daily cost of EV charging

the real-time price is lower than the strike price. Therefore, the daily cost of EV users under RTP with option increases when the demand is low. Through purchasing the put option, electricity utilities avoid the potential losses due to the excess renewable energy supply and keep most of the benefits from high market-clearing price. In September, the RTP with option will be slightly cheaper than RTP due to the put option payment to customers, which implies that the wind energy supply is neither surplus nor insufficient.

Table 3 reveals that EVs customers receiving the option incentives (except RTP with Option (90)) have lower average daily cost than involving in TOU or RTP. Moreover, customers under RTP with option program can avoid high price volatility compared to RTP. As the options exercised limit increases, there is less financial risk to be confronted. Less financial risk cannot guarantee a lower cost because the real-time electricity fluctuates. Still, it is recommended that EV customers should

Table 3 Daily cost of EV charging for DR programs

| DR program | Average daily cost (\$) | Standard error (\$) |
|-----------------------------------|-------------------------|---------------------|
| TOU | 1.080 | 0 |
| RTP | 1.034 | 1.319 |
| RTP with option (90) ^a | 1.054 | 1.188 |
| RTP with option (180) | 0.965 | 1.027 |
| RTP with option (360) | 1.005 | 0.915 |

^a RTP with Option (.) represents the limit of options exercised annually

involve in RTP with Option (360), as it reduces the exposure to price inflation risk, and costs less than RTP and TOU. In addition, the option payment, P_n , as a direct reward that varies hourly, can reinforce their acceptance to DR program and attract more users subscribing to DR programs.

The numerical study shows the feasibility of valuing DR through the annual electricity cost to EVs users. RTP combined with option not only provides an appropriate incentive for customers responding to demand requests but also reduces the grid stress caused by the excess renewable energy supply. With consumers' participation, fewer fossil backup power systems are required allowing electricity utilities to optimize grid operations and resources dynamically.

6 Conclusion and Future Work

This paper proposes that a financial option can be used as an additional incentive to EV customers involving in RTP. We estimate the value of DR based on the average daily cost of EVs charging. Exploiting the historical data of HOEP, an AR (1)-GARCH (1,1) model is calibrated to predict the financial risks for electricity producers in the market. The electricity retailers will pay EVs users through the put option to hedge the potential losses at the supply peak. Through a comparison of three different DR programs, it can be validated that the RTP with option will deliver the most beneficial outcome to all those involved, improving the environment by increasing renewable energy usage, and significantly reducing electricity costs and price volatility for consumers. In future work, some improvements could be made by introducing more option trading strategies to cover both the supply and demand peak. In this study, only the supply peak is covered by the put option, while the demand peak relies on the real-time price signal. Thus, future work can be expected to increase a much greater renewable energy share in the electricity market.

References

1. Albida MH, El-Saadany EF (2008) A summary of demand response in electricity markets. *Electric power Syst Res* 78(11):1989–1996
2. Balaraman K (2020) Wind plants can provide grid services similar to gas, hydro, easing renewables integration. *Caiso*. Accessed 22 Feb 2021. <https://www.utilitydive.com/news/wind-plants-can-provide-grid-services-similar-to-gas-hydro-easing-renewab/574070/>
3. Bellekom S, Benders R, Pelgröm S, Moll H (2012) Electric cars and wind energy: two problems, one solution? A study to combine wind energy and electric cars in 2020 in the Netherlands. *Energy* 45(1):859–866
4. Boisvert RN, Cappers PA, Neenan B (2002) The benefits of customer participation in wholesale electricity markets. *Electr J* 15(3):41–51
5. Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. *J Econometrics* 31(3):307–327

6. Borba BSM, Szklo A, Schaeffer R (2012) Plug-in hybrid electric vehicles as a way to maximize the integration of variable renewable energy in power systems: the case of wind generation in northeastern Brazil. *Energy* 37(1):469–481
7. CanWEA, Canadian Wind Energy Association (2020) Installed capacity. Accessed 22 Feb 2021. <https://canwea.ca/wind-energy/installed-capacity/>
8. Engle RF (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: J Econometric Soc* 987–1007
9. Garcia RC, Contreras J, Van Akkeren M, Garcia JBC (2005) A GARCH forecasting model to predict day-ahead electricity prices. *IEEE Trans Power Syst* 20(2):867–874
10. Hledik R, Faruqui A (2015) Valuing demand response: International best practices, case studies, and applications. The Brattle Group, Inc.
11. Ketterer JC (2014) The impact of wind power generation on the electricity price in Germany. *Energy Econ* 44:270–280
12. Kluge, T. (2006). Pricing swing options and other electricity derivatives. *University of Oxford*
13. Liu H, Shi J (2013) Applying ARMA-GARCH approaches to forecasting short-term electricity prices. *Energy Econ* 37:152–166
14. Luft J (1994) Bonus and penalty incentives contract choice by employees. *J Account Econ* 18(2):181–206
15. Independent Electricity System Operator (IESO) (2020) Data directory. Accessed 22 Feb 2021. <https://www.ieso.ca/en/Power-Data/Data-Directory>
16. MathWorks, Inc. (2019) Econometric math toolbox. Massachusetts, United State. Accessed 22 Feb 2021. <https://www.mathworks.com/help/Econ/>
17. Natural Resources Canada (2017) About renewable energy. 13 Dec 2017. Accessed 22 Feb 2021. <https://www.nrcan.gc.ca/our-natural-resources/energy-sources-distribution/renewable-energy/about-renewable-energy/7295>
18. Pineda S, Conejo AJ (2013) Using electricity options to hedge against financial risks of power producers. *J Modern Power Syst Clean Energy* 1(2):101–109
19. Plötz P, Jakobsson N, Sprei F (2017) On the distribution of individual daily driving distances. *Transport Res Part B Methodol* 101:213–227
20. Sewalt M, De Jong C (2003) Negative prices in electricity markets. *Commodities Now* 7:74–77
21. Steen D, Le T, Bertling L (2012) Price-based demand-side management for reducing peak demand in electrical distribution systems—with examples from Gothenburg. In: Nordic conference on electricity distribution system management and development
22. US Department of Energy (2020) Demand response. Office of Electricity. Accessed 22 Feb 2021. <https://www.energy.gov/oe/activities/technology-development/grid-modernization-and-smart-grid/demand-response>
23. Wang J, Liu C, Dan T, Zhou Y, Kim J, Vyas A (2011) Impact of plug-in hybrid electric vehicles on power systems with demand response and wind power. *Energy Policy* 39(7):4016–4021
24. Wang Z, Paranjape R, Chen Z, Zeng K (2019) Layered stochastic approach for residential demand response based on real-time pricing and incentive mechanism. *IET Gener Transm Distrib* 14(3):423–431
25. Weron R (2014) Electricity Price Forecasting: A review of the state-of-the-art with a look into the future. *Int J Forecast* 30(4):1030–1081

Advancing the Practice of Soil Bioengineering in Alberta—The Bow River Bioengineering Demonstration and Education Project



M. Gallant, N. Posada, J. Slaney, and P. Raymond

1 Introduction

Along the Bow River in Calgary, approximately 26% of the river banks have been hardened using traditional riprap revetments that significantly impact the rivers environmental, aesthetic, and recreational opportunities [7]. The City of Calgary developed the Riparian Action Program (RAP) in 2016 to implement a comprehensive and coordinated approach to protect riparian areas in Calgary [2]. One of the key actions of the RAP is to implement soil bioengineering techniques due to their increased ecological benefits compared to hard engineering practices. The City of Calgary in partnership with The Province of Alberta developed the Bioengineering Demonstration and Education Project (BDEP) as a way of demonstrating and encouraging the use of alternative softer forms of bank protection including bioengineering. The project had several goals beyond erosion protection including:

- **Creating Room for the River**—The banks of the river were pulled back to reduce upstream water levels and create different overbank flood levels to benefit different riparian ecosystems;
- **Wildlife Corridor**—As wildlife preferentially use river valleys for migration, a wildlife corridor was created within the project to help guide animals (e.g., deer, coyote, porcupine) under the adjacent bridge rather than over it;

M. Gallant (✉)
Kerr Wood Leidal Associates Ltd., Calgary, Canada
e-mail: mgallant@kwl.ca

N. Posada · J. Slaney
The City of Calgary (Water Resources, Calgary, Canada)

P. Raymond
Terra Erosion Control Ltd., Nelson, Canada

- Fish Habitat—Fish habitat features were integrated into the design itself rather than post design which greatly improved the overall fish habitat value of the project; and,
- Education—Extensive interpretative signage has been placed at the site demarking the various soil bioengineering techniques. Student groups are viewing the site from the water on organized rafting guided tours as part of the RiverWatch program. In addition, an informative website has been created (Calgary.ca/BDEP) where design and monitoring documents are available for download.

2 Description of the Project Site

BDEP is located on the right bank of the Bow River in the historic community of Inglewood in Calgary, Alberta. The site is approximately 700 m long and featured several challenging bank conditions including:

- severely eroded riverbanks at 1:1 or steeper slopes;
- a wildlife passage pinch point between Pearce Estates Park and Inglewood Bird Sanctuary at Cushing Bridge and 17 AVE SE Bus Rapid Transit bridge crossings;
- a riverbank toe adjacent to one of the deepest scour holes on the Bow River in Calgary;
- unhealthy riparian areas due to severe bank erosion, historic bank protection, and human use; and,
- a section of existing riprap bank protection and two groyne structures that were constructed because of 40 m of bank erosion from the 2013 flood.

3 Bioengineering Techniques Used

A focus of the BDEP was to demonstrate that soil bioengineering is a feasible alternative to conventional riprap and to focus on strict application of the design guidelines developed for bioengineering in Calgary [1]. Fourteen different soil bioengineering techniques were constructed at 10 treatment areas with a variety of hydraulic, geotechnical and environmental conditions. The list of techniques used at BDEP is provided in Table 1. Of the techniques used, half were new to Calgary, including the vegetated timber crib wall with fish shelters, the box fascine, and the hedge brush layers illustrated in Fig. 1.

4 Monitoring Methods

Monitoring at the site is being conducted according to the BDEP Bioengineering Efficacy Monitoring Plan (BEMP) [6], where monitoring occurs in years 2019, 2020,

Table 1 Bioengineering techniques used at BDEP

| Bioengineering techniques | Length of bank treated (m) | Shear stress resistance (N/m ²) (initial—grown) |
|--|----------------------------|---|
| Vegetated riprap with rooted long live cuttings ^a | 75 | > 350 ^b |
| Vegetated timber crib wall with ^a and without fish shelters | 120 | 600 ^d |
| Vegetated soil wraps with brush layers | 145 | 100–300 ^c |
| Brush mattress with and without a rock toe | 102 | 50 to >300 ^b |
| Box fascine ^a | 135 | 250 ^b |
| Brush layer with contour fascine | 82 | 100–300 ^c |
| Contour fascine | 20 | 60–150 ^c |
| Hedge brush layers ^a | 25 | 100–300 ^c |
| Live staking | 25 | 100–150 ^b |
| Three techniques to vegetate an existing riprap revetment ^a | 215 | >350 ^b |
| Conventional riprap d50 = 500 mm | — | 400 ^b |

^a Bioengineering techniques that are new to Calgary Sources ^b Lachat [11], ^c Fischenich [5], ^d Evette et al. [4]

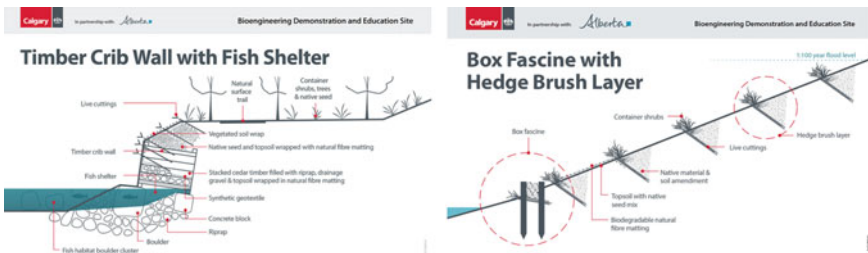


Fig. 1 Illustrations of BDEP bioengineering techniques developed for the education program

2021, 2023, and 2028 (year 1, year 2, year 3, year 5, and year 10 post-construction). Post-construction monitoring at the BDEP consists of the following:

- (1) **Fish and Fish Habitat**—monitoring methods include underwater photography and snorkel surveys to assess fish use of the site, visual redd surveys and kick sampling to assess fish spawning in the reach, electrofishing and minnow traps to assess fish abundance and diversity, water quality monitoring, and fish habitat assessment;
- (2) **Wildlife**—monitoring methods include breeding bird surveys, and wildlife cameras at the wildlife corridor to confirm use;

- (3) **Riparian Health**—monitored according to the *Large River Riparian Health Methodology* developed by Cows and Fish [3], and,
- (4) **Bioengineering Structural Integrity**—monitoring methods are based on detailed structural (i.e., inspections of erosion, condition of construction materials, evaluation of river flow and shear stress) and vegetation survival assessments that were developed as part of the City of Calgary Riparian Monitoring Program [8].

5 Results to Date

Most bioengineering techniques used at the site are flourishing. In particular, the hedge brush layer technique (Fig. 1), which involves adding a container plant species that does not form adventitious roots in a conventional brush layer, has shown to perform very well. Results by monitoring component are provided below [9, 10].

- (1) **Fish and Fish Habitat:** In 2019, 48 fish from 10 species and in 2020, 120 fish from 9 species were observed to be using the project area for migration, foraging, overwintering, rearing, and spawning purposes. Species including brown trout, rainbow trout, burbot, northern pike, and mountain whitefish were observed in the fish shelters, boulder clusters, and surrounding habitats during winter, spring and summer assessments.
- (2) **Wildlife:** In 2019 over 120 individual birds from 31 species, and in 2020 50 individual birds from 16 different species were observed during breeding bird surveys. The wildlife cameras captured 212 individuals in 2019 and 317 individuals in 2020 using the wildlife corridor with the most common species being white-tailed deer, coyote, and white-tailed jackrabbit.
- (3) **Riparian Health:** There has been a marked improvement in riparian health at all three BDEP sites since the baseline assessment in 2016. Vegetation ratings are substantially higher for Sites 1, 2 and 4, ranging from 2 to 2.5 times higher than the vegetation rating for a theoretical conventional riprap design site. Also, overall ratings for Sites 1, 2 and 4 range from 34 to 54% higher than the overall rating for a theoretical conventional riprap design site.
- (4) **Bioengineering Structural Integrity:** Flows in the Bow River from 2018 to 2020 were at or lower than the 2-year peak flow. In general, the physical condition of the site structures, including fish habitat structures appear to be stable, with no signs of major erosion, scour, or displacement. Vegetation survival results are good with overall survival (all sites included) of 80% in 2019.

6 Lessons Learned/Recommendations for Future Projects

Based on preliminary analysis, using soil bioengineering techniques at the BDEP resulted in an estimated cost savings of about \$1 M over a conventional riprap

approach. The site has achieved its goal of demonstrating that soil bioengineering is a more cost effective, environmentally friendly, and aesthetic alternative to conventional riprap where shear stresses allow. It is hoped that the education component of the project will encourage local practitioners to use these techniques in the future.

Additional technical lessons learned from the project are listed below.

- Hedge brush layers and rooted long live cuttings are successful new techniques used in the BDEP that should be incorporated in other projects. By adding rooted plant species in the brush layer to form a hedge brush layer, biodiversity is improved, and fruiting shrubs for wildlife food and nitrogen fixing species are included. At the BDEP, 5-gallon size chokecherry, river alder, red osier dogwood and balsam poplar were placed at one plant per metre in combination with regulator live cuttings.
- By using rooted long live cuttings in one site, the contractor was able to construct the site in August which is outside of the traditional construction window for soil bioengineering projects and was then able to meet a tight construction schedule.
- 2019 and 2020 monitoring results of the three techniques used to vegetate existing riprap revealed that covering the riprap with planting material resulted in greater vegetation establishment success versus the void-fill methods. However, hot and dry conditions for vegetation growth in existing riprap have limited the overall establishment of woody vegetation at these sites. Recommendations based on observations of these techniques are that topsoil void-fill appears to provide better growth than a well graded granular material but is more easily washed out; placing live cuttings in void spaces prior to placing void-fill is preferred; and, using a telebelt or similar method allows placement of void-fill material deep into the riprap voids on a site with limited construction equipment access.
- The BDEP highlights the importance of a reliable contractor with skilled personnel to implement a strong maintenance program for weeding and irrigation installation and scheduling. While vegetation has established throughout the site, at some locations seeded herbaceous vegetation and invasive weeds are competing with the planted woody vegetation for growing space, nutrients, moisture and light. This is causing stress on the woody vegetation and is restricting growth. Irrigation practices have led to good establishment in most areas but survival in the vegetated timber crib wall is below expectations due to challenging watering conditions.

Monitoring of the project will continue to 2028, where it is expected that the collected data will result in more refined recommendations to improve soil bioengineering practices in Calgary.

References

1. AMEC (2012) Design guidelines for erosion and flood control projects for streambank and riparian stability restoration. Technical Report prepared by AMEC Environment & Infrastructure for The City of Calgary. Available online at: Calgary.ca/BDEP

2. City of Calgary (2017) Riparian action program. Available online at: <https://www.calgary.ca/uep/water/watersheds-and-rivers/riparian-areas.html>
3. Cows and Fish (2018) Alberta Lotic Wetland Health Assessment for Large River Systems (Survey) User Manual (Technical Report prepared by Alberta Riparian Habitat Management Society [Cows and Fish])
4. Evette A, Jaymond D, Recking A, Piton G, Rauch HP, Frossard PA (2018) The limits of mechanical resistance in bioengineering for riverbank protection. International Association for Life-Cycle Civil Engineering, Ghent, Belgium
5. Fischenich C (2001) Stability thresholds for stream restoration materials—EMRRP Technical Notes Collection (ERDC TN EMRRP-SR-29). USACE Research and Development Center, Vicksburg, MS
6. Hemmera (2018) Bioengineering demonstration and education project bioengineering efficacy monitoring plan—final report. Technical Report prepared for Alberta Environment and Parks by Hemmera Envirochem Inc
7. KCBL (2017) The City of Calgary morphology and fish habitat study—main report. Technical Report prepared by Klohn Crippen Berger Ltd. for The City of Calgary
8. KWL (2018) The City of Calgary riparian monitoring program monitoring plan. Technical Report prepared for The City of Calgary by Kerr Wood Leidal Associates Ltd
9. KWL (2020) Bioengineering demonstration and education project—2019 monitoring report. Technical Report prepared for The City of Calgary by Kerr Wood Leidal Associates Ltd
10. KWL (2021) Bioengineering demonstration and education project—2020 monitoring report. Technical Report in preparation for The City of Calgary by Kerr Wood Leidal Associates Ltd
11. Lachat B (1999) Guide de protection des berges de cours d'eau en techniques vegetales, Ministère de l'Aménagement du territoire et de l'Environnement, France

Case Study of Flood Profile Hydraulic Modeling Uncertainty



Nathan A. Valsangkar and David G. McLean

1 Introduction

Flood maps are an important component of floodplain management and emergency planning. Flood maps can be used to communicate a variety of information, including inundation extents, hazards (such as flow depths and velocities), and consequences/damages. A primary flood mapping input is the design flood profile, defined as the predicted water levels corresponding to a design return period [5]. Standard engineering practice is to estimate design flood profiles using hydraulic models.

Flood profiles estimated by hydraulic modeling are subject to several sources of uncertainty [4, 5, 14, 15]. The model itself has structural and numerical uncertainties, as well as uncertainties in its parameterization. External to the model, there are informational uncertainties in boundary and initial conditions, as well as in observed data used for model calibration [14]. Natural variation, such as geomorphic processes and climate change, can modify flood profiles within the life of the maps and present additional uncertainty.

The conventional approach to addressing uncertainties is the addition of a fixed vertical freeboard to design flood profiles. In British Columbia, there is no provincial standard for freeboard. Historical flood mapping studies have used fixed freeboards of either 0.3 m or 0.6 m when flood profiles were estimated using peak instantaneous discharge or average daily discharge, respectively [1]. The use of standard fixed freeboards can be arbitrary and has been questioned by several authors [5, 16, 20]. They advocate for using a probabilistic approach to identify and evaluate the sources of uncertainty and their impacts on design flood profiles. To the authors' knowledge, probabilistic approaches are seldom applied to flood mapping projects in British Columbia.

N. A. Valsangkar (✉) · D. G. McLean
Northwest Hydraulic Consultants Ltd, Nanaimo, Canada
e-mail: nvalsangkar@nhcweb.com

The purpose of this study was to complete an uncertainty analysis of flood profile hydraulic modeling for a reach of the Cowichan River, Vancouver Island. In 2020, Northwest Hydraulic Consultants Ltd. (NHC) completed floodplain and erosion hazard mapping for the reach on behalf of the Cowichan Valley Regional District (CVRD). The outcomes of the analysis are compared to historical freeboard standards and are used as a defensible basis for design freeboard selection.

2 Study Area

The Cowichan River has its headwaters at Hooper Mountain, near the western end of Lake Cowichan, BC. From Lake Cowichan, the river flows east 46 km to tidewater near Duncan. The study area is an approximately 14 km long reach of the river near Riverbottom Road (Fig. 1). Within this reach, the channel has an average gradient of approximately 0.4% and a typical bank-full top width between 40 and 100 m. The river has an irregular wandering planform with frequent irregular bars, vegetated islands, and in-channel wood debris. It is often confined on one or both sides by steep terraces of glacial and glaciofluvial materials. Historically, the river has been subject to bank erosion and intermittent, rapid channel shifting and avulsions [17].

Peak flow events in the region typically occur in November to March when Pacific cyclonic depressions pass over the Strait of Georgia. Orographic effects result in high rainfall rates along the mountains of Vancouver Island. Severe flooding on the Cowichan is often generated by rain-on-snow events [17]. Land use within the floodplain generally consists of rural residential properties and forested land. There are no dikes in this reach, though many riverfront properties include erosion mitigation works such as riprap revetments.

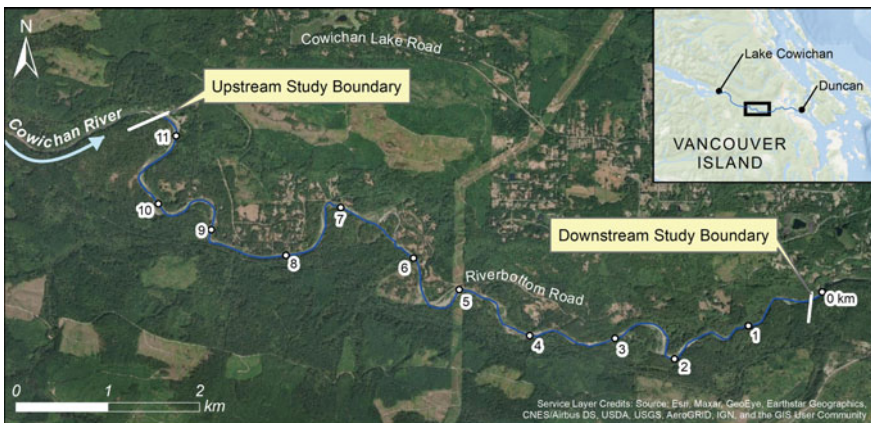


Fig. 1 Overview of study area

3 Hydraulic Model Development

3.1 Model Selection

The US Hydrologic Engineering Center (HEC) HEC-RAS 5.0.7 one-dimensional hydraulic model was selected for use in the study. The model was run for steady gradually varied flow, wherein water surface profiles are computed between stream cross sections using the standard step procedure. Flow resistance is computed using the Manning equation. A detailed description of HEC-RAS can be found in the HEC-RAS River Analysis System Hydraulic Reference Manual [6].

3.2 Model Geometry

The model geometry was developed using single-beam sonar bathymetry data collected in winter 2020, and topographic LiDAR data collected in summer 2019. The geometry includes 45 channel and floodplain cross sections, with an average spacing of approximately 250 m between cross sections. Ineffective flow areas were delineated based on qualitative analysis of the topographic LiDAR data, as well as known floodplain areas that are subject to ponding during overbank flow.

3.3 Boundary Conditions

The model's downstream boundary condition was set to normal depth at a bed slope of 0.4%. The downstream boundary is located approximately 250 m downstream of the study limit. The CVRD adopted the 200-year return period peak instantaneous discharge (Q_{200i}) as the design hydrologic boundary condition for regulatory floodplain map development. The design flood must also consider potential climate change impacts [1]. The following summarizes the approach used to estimate the design flood magnitude.

The study reach is located between two Water Survey of Canada (WSC) hydro-metric gauges: 08HA002 Cowichan River near Lake Cowichan (upstream of study reach) and 08HA011 Cowichan River near Duncan (downstream of study reach). The gauges' upstream drainage areas are 594 km² and 826 km², respectively; the drainage area for the study reach is 735 km². A flood frequency analysis was carried out to estimate the gauges' Q_{200i} magnitudes. The estimates were 344 m³/s and 747 m³/s, respectively.

The Modified Index Flood method (Eq. 1) was used to scale the incremental Q_{200i} flow gain between 08HA002 and 08HA011 to the incremental watershed area between 08HA002 and the study reach. The incremental Q_{200i} flow gain estimate was 277 m³/s. The total Q_{200i} estimate for the study reach is the sum of the 08HA002

flood magnitude and the scaled incremental flow gain to the study reach, equal to $622 \text{ m}^3/\text{s}$.

An overview climate change assessment was carried out in accordance with EGBC guidelines [1, 10]. CVRD climate change projections [8] for the Cowichan River watershed indicate increasing trends in the magnitude of extreme precipitation. The Mann–Kendall test was applied to peak flow records for the Cowichan River WSC gauges to evaluate whether temporal trends exist. Statistically significant increasing trends ($\alpha = 0.05$) were found in the Duncan gauge (08HA011). Per EGBC guidelines [18], a 20% increase on design discharges is appropriate when trend analyses indicate increasing flood magnitudes over time. The Q_{200} design flood magnitude is $746 \text{ m}^3/\text{s}$.

$$Q_2 = Q_1 \cdot (A_1/A_2)^n \quad (1)$$

Equation 1 relates the discharges (Q) of two watersheds based on their respective areas (A) and a scaling exponent (n). For British Columbia, Eaton et al. [9] reported an average n value of 0.75.

3.4 Model Calibration and Validation

Model calibration was carried out using a continuous water surface profile surveyed during a flood flow of approximately $79 \text{ m}^3/\text{s}$. Calibration was completed by varying channel and overbank Manning's n values; calibrated average channel and overbank Manning's n values were 0.048 and 0.11, respectively. The model was validated using additional data from high water mark surveys. The surveys were completed during the February 1, 2020 flood event, which had a return period between 30 and 50 years. The surveyed points include five water surface elevations and three inferred high-water marks, such as rafted wood debris. Surveys were completed at several flood stages; the approach described in Sect. 3.3 estimated a range of flows during the survey of approximately $278\text{--}464 \text{ m}^3/\text{s}$. Table 1 summarizes the calibration and validation results.

Note that the water surface profile used for calibration is subject to measurement uncertainties on the order of $\pm 0.07 \text{ m}$. The high-water marks used for validation have higher measurement uncertainties, particularly the inferred high-water marks. The total uncertainty is not quantifiable but may be on the order of ± 0.1 to 0.3 m .

Table 1 Summary of model calibration and validation

| Model run | Number of measurements | Mean error (m) | Maximum absolute error (m) | RMSD (m) |
|-------------|------------------------|----------------|----------------------------|----------|
| Calibration | 45 | 0.00 | 0.17 | 0.09 |
| Validation | 8 | 0.05 | 0.37 | 0.19 |

The flow estimates used as hydrologic boundary conditions for validation are also uncertain (see Sect. 4.1). Root mean squared difference (RMSD) values (Eq. 2) for calibration and validation are considered comparable to the inherent uncertainties of the calibration and validation datasets.

For the validation dataset, the model results had a non-zero mean error of + 0.05 m. However, a single-tailed t-test ($\alpha = 0.05$) indicates that the mean error is not statistically greater than zero. The authors acknowledge that Manning’s n can be systematically different at low and high flows (see Sturm [19]). This can introduce bias in design flood profiles if calibration is conducted at very low flows. The impact of systematic bias on flood profile uncertainty has not been assessed.

$$RMSD = \sqrt{\sum_{i=1}^n (e_i)^2 / n} \tag{2}$$

Equation 2 (see Willmott and Matsuura [21]) defines the root mean squared difference (RMSD), where e_i is the difference between pairwise matched observations (whether modeled or observed), and n is the number of pairwise matched observations.

3.5 Design Flood Profiles

The calibrated and validated model, along with the Q200_i design flood magnitude, was used to estimate the 200-year return period design flood profile for the study reach (Fig. 2).

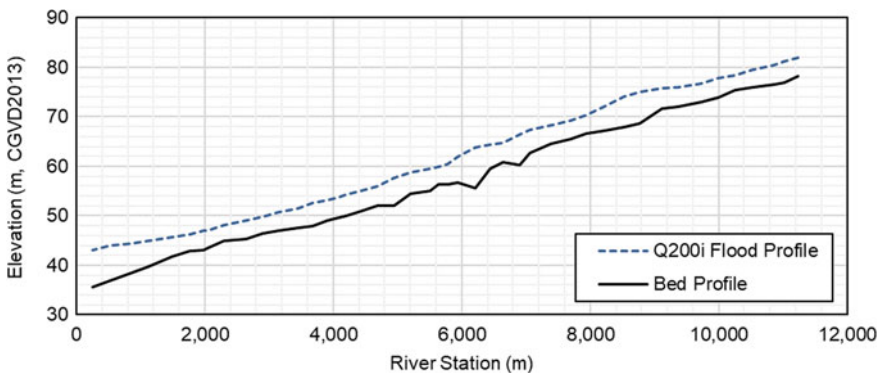


Fig. 2 Design flood profile for study reach, corresponding to Q200_i

4 Uncertainty Analysis

4.1 Uncertainty Quantification

A complete uncertainty analysis would involve the rigorous identification of all sources of uncertainty and their joint probabilistic impacts on design flood profiles [14]. Such an analysis would be computationally intense and likely impracticable. The uncertainty analysis instead focuses on two major contributors to uncertainty: hydrologic boundary conditions and Manning's n [4, 5].

Uncertainty quantification for the Q200_i hydrologic boundary focused on flood frequency analysis uncertainty and uncertainty in the scaling exponent used in Eq. 1. The uncertainty associated with climate change is beyond the scope of the study and has not been quantified. Flood frequency analysis uncertainties for the two WSCS gauges were estimated by percentile bootstrapping [7]. Uncertainty in the scaling exponent was quantified using the regression uncertainties reported by Eaton et al. [9]. The overall uncertainty was estimated using Monte Carlo methods, whereby each simulation applied the areal scaling methodology (see Sect. 3.3) to randomly sampled flow and scaling exponent values. The mean Q200_i estimate was 746 m³/s, with a 95% confidence interval of 668–834 m³/s over 25,000 Monte Carlo simulations.

Manning's n is an empirical parameter describing the flow resistance of the river channel and overbank floodplains [2]. Uncertainty quantification for Manning's n values focused on the linkages between model parameterization and calibration/validation error. The sensitivity of the calibration/validation flood profiles was evaluated by globally varying the model's base Manning's n values (Fig. 3). For calibration flows, Manning's n variations of ± 12 and $\pm 24\%$ produced mean flood profile changes equivalent to one and two times the calibration RMSD, respectively. For validation flows, the same Manning's n variations produced mean flood profile

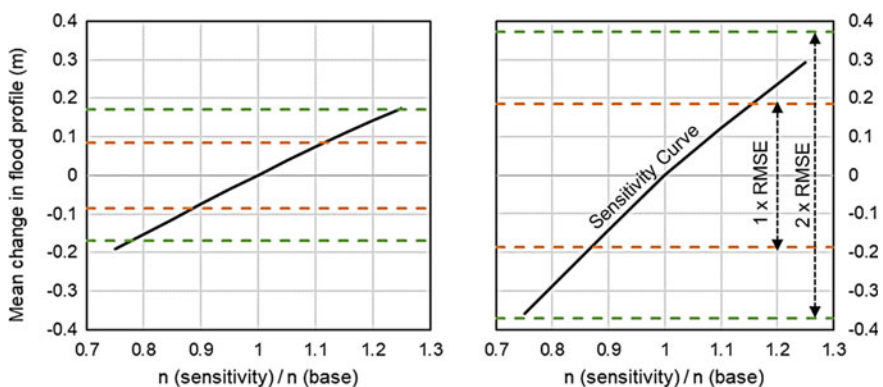


Fig. 3 Sensitivity of flood profiles to global changes in Manning's n for calibration (left figure) and validation (right figure) conditions. The orange and green lines, respectively, represent one and two-times the RMSD of calibration (left figure) and validation (right figure)

changes of somewhat less than one and two times the validation RMSD. Note that the validation’s hydrologic boundary conditions and surveyed high water marks are inherently more uncertain than those used for calibration (Sect. 3.4). The results indicate that global variations in Manning’s n of ± 12 and $\pm 24\%$ can approximate the statistical properties of calibration and validation uncertainty. We quantified the Manning’s n uncertainty using a normal distribution with mean of zero and a standard deviation of 12%.

4.2 Uncertainty Propagation Analysis

The joint probabilistic impacts of hydrologic boundary and Manning’s n uncertainty were evaluated using Monte Carlo uncertainty analysis. The analysis approach is adapted from the [20], described as follows and as illustrated in Fig. 4:

1. The uncertainty in Q_{200i} values was characterized by an empirical probability density function (PDF) derived from the Monte Carlo simulation ensemble described in Sect. 4.1.
2. The hydraulic model was used to develop a family of stage-discharge curves for variable Manning’s n values at each model cross section. The curves define the uncertainty in flood elevations for various discharge values.
3. For each Monte Carlo simulation a Q_{200i} value was randomly sampled from the Q_{200i} PDF and a Manning’s n value was randomly sampled from the normal distribution defined in Sect. 4.1.

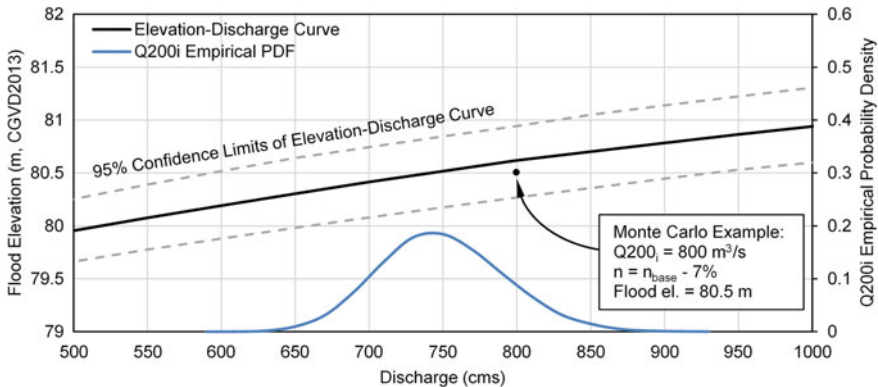


Fig. 4 Example of Q_{200i} PDF and stage-discharge curve uncertainties for a model cross section located at the upstream limit of the study reach. The Monte Carlo simulation example shows the modeled flood elevation corresponding to randomly sampled Q_{200i} and n values. This Monte Carlo simulation method was repeated 10,000 times to generate an ensemble of modeled flood elevations

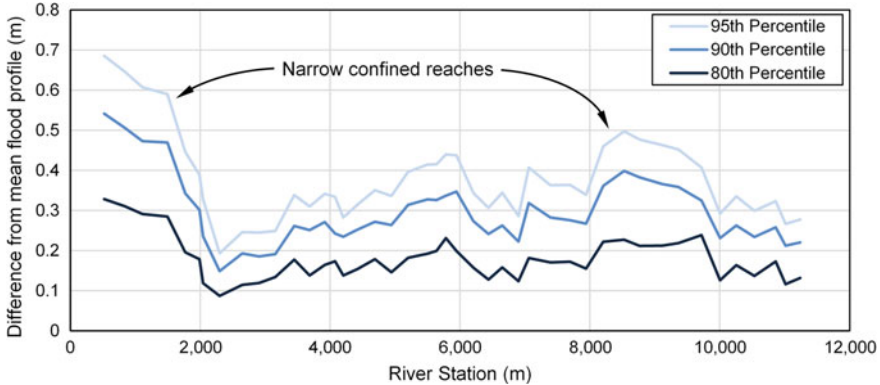


Fig. 5 Monte Carlo uncertainty propagation results, differences from the mean design flood profile at the 80th, 90th, and 95th percentiles after 10,000 simulations

4. The stage-discharge curves were used to estimate the flood elevation at each model cross section corresponding to the sampled Q_{200_i} and Manning's n values.
5. Steps 2 and 4 were repeated to generate a Monte Carlo simulation ensemble of flood elevations at each model cross section. Approximately 2500 simulations were required to converge the ensemble mean flood elevations to within 5 mm.

Note that this approach considers the global uncertainty on Manning's n and does not vary Manning's n independently for each cross section. HEC-RAS's standard step backwater procedure uses the hydraulic geometry both upstream and downstream of a given cross section when calculating flood profiles. By calibrating and validating the model to field data, the Manning's n uncertainties at each cross section cannot be considered independent.

The Monte Carlo simulations produced an ensemble of 10,000 design flood profiles. 80th percentile differences from the mean flood profile ranged from 0.09 to 0.34 m, with a RMSD of 0.19 m. 95th percentile differences ranged from 0.19 to 0.71 m, with RMSD of 0.40 m (Fig. 5). Uncertainties were highest in narrow, confined sub-reaches such as canyons and lowest in sub-reaches with broad, low-gradient floodplains.

4.3 Impacts of Geomorphic Change

Geomorphic change can impact hydraulics and flood elevations [11]. In BC it is not typical practice to prepare floodplain maps that consider potential future channel configurations [1]. The exception is alluvial fan hazard maps, where significant geomorphic changes can be expected over time and during the design event. Floodplain maps may include a recommended period of applicability, after which the

maps should be reviewed and updated to reflect changes to the design flood, channel geometry, diking, or floodplain use.

Air photo records indicate a history of lateral channel migration and avulsions (major channel-shifting events) within the study reach [17]. A 1993 bathymetric survey is available from historical floodplain mapping completed by the Province [3]. A comparison of the 1993 and 2020 bathymetries indicates bed level changes ranging from -1.5 to 2.0 m along the study reach. Four avulsions occurred between 1993 and 2020.

The potential for geomorphic change over the life of the maps can be considered an additional source of flood profile uncertainty. This was investigated by comparing the Province’s 1993 hydraulic model to the 2020 model. The Province’s model (HEC-2 format) was converted to HEC-RAS and updated to match the 2020 model boundary conditions. The model has 24 cross sections, compared to 45 used in the 2020 model. The Province’s Manning’s n values were originally calibrated to an incorrect hydrological boundary condition and are not considered reliable (NHC 2020). For simplicity, the Manning’s n values in both the 1993 and 2020 models were updated to match the mean calibrated values from the 2020 model (0.048 and 0.11 for the channel and overbanks, respectively.) This removes a component of the uncertainty associated with comparing the two models’ flood profiles.

Seventeen of the 1993 model cross sections were located within 50 m of the 2020 model cross sections. The 2020 model’s flood elevations at these locations were estimated by flood profile interpolation and compared to the 1993 model’s flood elevations (Fig. 6). The flood elevation differences were normally distributed (Shapiro–Wilk test, $\alpha = 0.05$) with mean and standard deviations of 0.04 m and 0.36 m, respectively. There does not appear to be a reach-scale trend of aggradation or degradation. The results indicate local variations in flood elevations, typically on the order of ± 0.25 to 0.5 m, due to morphological changes extending over scales of several channel widths.

Note that the differences are not solely attributable to geomorphic changes that occurred from 1993 to 2020. The two models’ flood profiles are subject to uncertainties from cross section spacing and numerical errors; it was necessary to complete

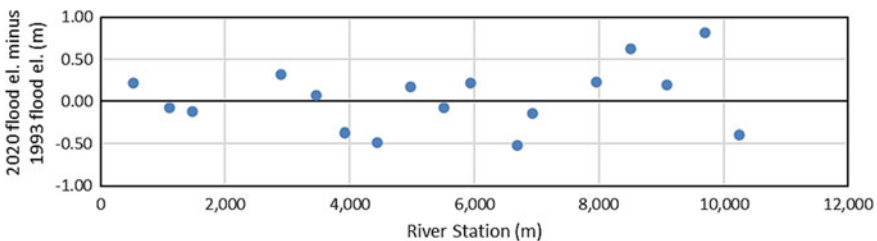


Fig. 6 Differences in design flood elevations between the 2020 hydraulic model and the updated 1993 hydraulic model. Both models use average Manning’s n values of 0.048 and 0.11 for the channel and floodplains, respectively

flood profile interpolation in the 2020 model; and the 1993 model used lower-resolution topographic data to define the overbank floodplain geometry. However, both models' cross section spacings are within the backwater threshold suggested by Samuels equation (see Samuels [18]). Some uncertainty from cross section spacing and flood profile interpolation is present, but it is likely smaller than other uncertainty sources. The uncertainties associated with overbank floodplain geometry and numerical errors have not been quantified.

4.4 Influence of Channel Confinement

The uncertainty propagation results (Sect. 4.2) indicate that the degree of channel confinement influences flood profile uncertainty with respect to hydrologic boundary conditions and Manning's n . This effect can be explored theoretically through the form of the Manning equation (Eq. 3). Partial derivatives of flow depth with respect to discharge and Manning's n ($\partial H/\partial Q$ and $\partial H/\partial n$) yield gradients that are inversely related to channel width. If bed slope is held constant, halving the channel width increases the gradients of the partial derivatives by approximately 52%. This implies that changes, or uncertainties, in discharge and Manning's n will produce greater changes in flood profiles for a narrow, confined channel than for a wide, unconstrained channel.

$$H = Q^{3/5} W^{-3/5} S^{-3/10} n^{3/5} \quad (3)$$

Equation 3 presents the wide rectangular channel approximation of the Manning equation for normal depth [19], where H is flow depth (m), Q is discharge (m^3/s), W is channel width (m), S is bed slope (m/m), and n is the Manning roughness coefficient.

It is well documented that constraining rivers through dike construction and floodplain encroachment can increase flood elevations [10, 20]. Klijn et al. [12] argue that channel confinement also has implications for the robustness of river systems from a flood risk management perspective. In BC, the design of flood defense works must consider the potential for future climate change within the design life of the works [10]. The impact of climate change on river flood magnitudes is highly uncertain and may not be quantifiable [13]. Flood defenses that are highly sensitive to discharge uncertainties may have an increased probability of failure over time [12]. Probabilistic design methods can mitigate uncertainty [20], however, freeboards that consider significant climate change uncertainty may be impracticably large.

We further investigated the impact of channel confinement by including hypothetical streambank dikes in the calibrated/validated 2020 model. The dikes were located on both channel banks to simulate a "traditional" dike configuration rather than a setback dike configuration. The Monte Carlo uncertainty propagation approach described in Sect. 4.2 was applied to the diked model. As expected, adding dikes increased the mean design flood profile, with a RMSD of 0.44 m from the base 2020

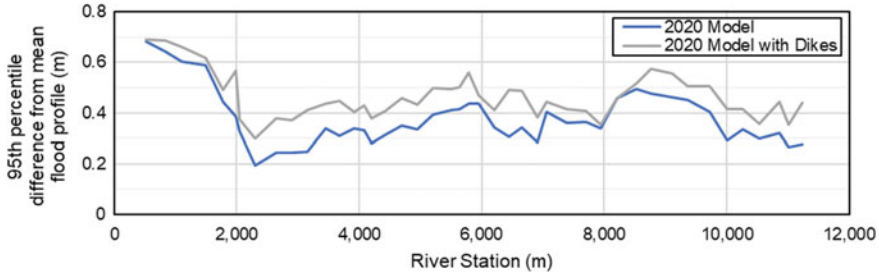


Fig. 7 Monte Carlo uncertainty propagation results for the 2020 calibrated/validated model and the 2020 model with dikes. Differences from the mean design flood profile are at the 95th percentile after 10,000 simulations

model. Uncertainty in the design flood profile also increased, as demonstrated by the two models' 95th percentile differences from the mean flood profile (Fig. 7). The diked model increased the 95th percentile difference from the mean flood profile by up to 0.18 m, with a RMSD 0.10 m. In areas that currently have wide, unconfined floodplains, adding hypothetical streambank dikes resulted in the greatest increases in flood elevations and flood profile uncertainties. An example is river stations 2500–3000 (Fig. 7).

5 Discussion

Hydraulic models are powerful tools for assessing flood hazards and designing flood mitigation measures. However, communicating realistic estimates of flood profile uncertainties to stakeholder and regulatory agencies should be recognized as a necessary part of any flood investigation [15]. The commonly used approach of adding a standard freeboard to design flood elevations does not account for local site conditions. Freeboard can often be misinterpreted by non-technical stakeholders as a very conservative additional “safety factor”. On some other projects in BC, stakeholders and planners have confused model calibration statistics with the likely accuracy of the overall design flood level. The results of this study demonstrate that good model calibration does not necessarily provide sufficient insight on the overall uncertainty of flood profile estimates.

This study demonstrated that relatively simple methods can be used to represent uncertainties in some of the key components that govern flood levels, including hydrologic boundary conditions, channel roughness, and changes in channel/floodplain topography due to geomorphic change. The results show that the uncertainty in predicted water levels is greater in narrow, confined sections of the river than in wide sections. This result reinforces the findings of [12] who showed that flood profile model sensitivity testing could be used to quantify the resiliency of flood defenses. Ultimately, the authors recommended a 0.6 m freeboard on the mean design flood

profile for use in the 2020 floodplain mapping study. This considered the uncertainties stemming from hydrologic boundary conditions and channel roughness, as well as the fact that other uncertainties were not quantified in the analysis.

Accounting for geomorphic change over the duration of flood map usage is difficult to assess given the limited data. Based on surveys from 1993 and 2020 (27-year span), the results showed no overall trend of aggradation or degradation in the reach (the mean change in flood elevations was 0 m). However, local flood elevation variations of ± 0.25 to ± 0.5 m occurred due to morphological changes that extended over length scales of several channel widths. Note that these variations are transient in nature because geomorphic change is ongoing. A location with increased flood elevations may experience lowered flood elevations 10 or 20 years from now, and vice versa. As such, it is challenging to directly incorporate geomorphic change into design freeboard levels. For the 2020 floodplain mapping study, the authors recommended that the maps be reviewed ten years from publication, after any significant avulsion, or after a flood with a recurrence interval greater than approximately 25 years.

Acknowledgements This study was supported by the Cowichan Valley Regional District (CVRD) as part of the Riverbottom Road Flood and Erosion Hazard Mapping Project. We acknowledge the guidance and direction provided by Jeff Moore, Kate Miller, and Hamid Hatami. Other NHC staff whose work contributed to the flood profile uncertainty analysis include Faye Hirshfield (hydrology), Wil Hilsen (geomorphology), Ben Humphreys (GIS), and Peter Kirillo (geomatics).

References

1. APEGBC (2017) Flood mapping in BC. APEGBC Professional Practice Guidelines V1.0. Engineers and Geoscientists BC, Burnaby, BC, Canada
2. Arcement GJ, Schneider VR (1989) Guide for selecting manning's roughness coefficients for natural channels and floodplains. United States Geological Survey, Denver, CO, USA
3. BC Ministry of Environment, Lands, and Parks (1997) A design brief on the floodplain mapping project. Cowichan and Koksilah Rivers near Duncan and the Cowichan River (Riverbottom Road Area). Water Inventory Section, Victoria, BC, Canada
4. Bessar MA, Matte P, Anctil F (2020) Uncertainty Analysis of a 1D river hydraulic model with adaptive calibration. *Water* 12(561)
5. Brandimarte L, Di Baldassarre G (2012) Uncertainty in design flood profiles derived by hydraulic modeling. *Hydrol Res* 43(6):753–761
6. Brunner GW (2016) HEC-RAS river analysis system hydraulic reference manual. US Army Corps of Engineers Hydrologic Engineering Center (HEC), Davis, California, USA
7. Chernick MR, LaBudde RA (2011) An introduction to bootstrap methods with applications to R. Wiley, Hoboken
8. Cowichan Valley Regional District (2017) Climate projections for the Cowichan Valley Regional District. Cowichan Valley Regional District, Duncan, BC, Canada. <https://www.cvrld.ca/DocumentCenter/View/81884/Climate-Projections-Report>
9. Eaton B, Church M, Ham D (2002) Scaling and regionalization of flood flows in British Columbia, Canada. *Hydrol Proces* 16(16):3245–3263
10. EGBC (2018) Legislated flood assessments in a changing climate in BC. Professional Practice Guidelines V2.1. Engineers and Geoscientists BC, Burnaby, BC, Canada

11. Hou J et al (2016) Effects of morphological change on fluvial flood patterns evaluated by a hydro-geomorphological model. *Procedia Engineering* 154:441–449
12. Klijn F, Asselman N, Mosselman E (2018) Robust river systems: on assessing the sensitivity of embanked rivers to discharge uncertainties, exemplified for the Netherlands' main rivers. *J Flood Risk Manage* 12(19)
13. Kundzewicz ZW et al (2014) Flood risk and climate change: global and regional perspectives. *Hydrol Sci J* 59(1):1–28
14. Loucks DP, van Beek E (2017) Water resource systems planning and management—an introduction to methods, models, and applications. Deltares and UNESCO-IHE, Delft, Netherlands
15. Mosselman E (2018) Modelling in Applied Hydraulics: More Accurate in Decision-Making than in Science? *SymHydro 2017*. Springer, Nice, France 13:741–749
16. National Research Council (2000) Risk Analysis and Uncertainty in Flood Damage Reduction Studies. The National Academic Press, Washington, DC
17. Northwest Hydraulic Consultants Ltd. 2020. *Cowichan River – Riverbottom Road Area Flood and Erosion Hazard Mapping. Final Report – Revision 1*. Cowichan Valley Regional District, Duncan, BC, Canada. <http://cvrd.ca/3195/Flooding>
18. Samuels PG (1989) Backwater lengths in rivers. *Proc Inst Civ Eng* 87(4):571–582
19. Sturm T (2009) *Open Channel Hydraulics*, 2nd edn. McGraw-Hill, New York, NY, USA
20. U.S. Army Corps of Engineers. 1996. *Risk-Based Analysis for Flood Damage Reduction Studies. EM 1110–2–1619*. Washington, DC.
21. Willmott CJ, Matsuura K (2005) Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance. *Climate Res* 30:79–82

Surrogate Model Development for Bioretention Cell Simulation-Optimisation Applications



R. Khalid, E. Snieder, and U. T. Khan

1 Introduction

With urbanization, rural and undeveloped lands are continually converted to urban areas. Impervious areas such as parking lots, buildings and streets replace the native vegetation, consequently disrupting the natural water balance of a watershed [8]. The amount of stormwater runoff and peak flow rate is increased as it no longer infiltrates into the native soil, contributing to a greater risk of pluvial flooding in urban cities. Additionally, urbanization is considered to be the biggest contributor to non-point source pollution, resulting in deteriorated stormwater quality due to increased nutrients, bacteria and toxic metals concentration [1, 3, 11, 25]. Further amplifying the risk of pluvial flooding is the increased frequency and intensity of extreme rainfall events due to climate change [16]. According to Gaur et al. [14] 40–60% of Canadian cities are projected to encounter increases in flooding frequency due to climate change, with 100-year storm events predicted to occur every 15 years in Toronto. One potential method to mitigate these risks is through the use of Low Impact Development (LID), a common approach that aims to mimic the pre-development hydrological conditions of the developed landscape through increased infiltration and evapotranspiration [10, 12, 28].

R. Khalid · E. Snieder · U. T. Khan (✉)
Department of Civil Engineering, York University, 4700 Keele Street, Toronto, ON M3J 1P3,
Canada
e-mail: usman.khan@lassonde.yorku.ca

© Canadian Society for Civil Engineering 2022
S. Walbridge et al. (eds.), *Proceedings of the Canadian Society of Civil Engineering
Annual Conference 2021*, Lecture Notes in Civil Engineering 250,
https://doi.org/10.1007/978-981-19-1065-4_17

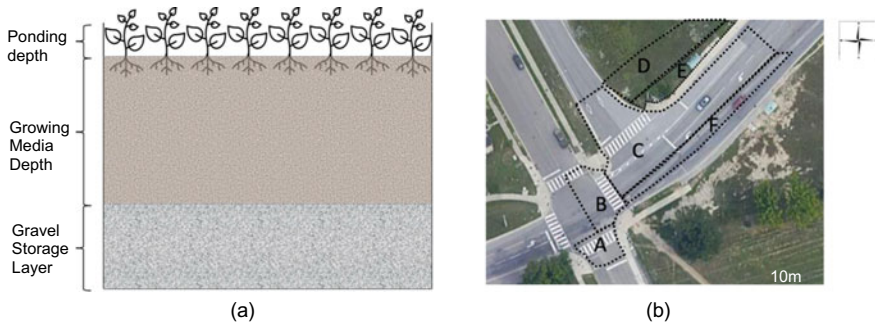


Fig. 1 **a** The main vertical layers of a bioretention cell. **b** Delineated subcatchments of the study site. The BRC is denoted by subcatchment F

1.1 Bioretention Cells

Bioretention cells (BRC), a type of LID, are designed to infiltrate, filtrate and temporarily store runoff. Their general profile is as follows; a top vegetation layer, a soil growing media layer, a gravel storage layer and possibly an underdrain, as shown in Fig. 1 [6]. BRCs have been extensively researched, both in field and laboratory settings, with proven performance in reducing runoff volume, peak flow rate and improving stormwater quality [1, 2, 18, 19, 28].

However, reported results have varying efficiencies ranging from 66 to 96% in peak flow reductions, 62–93.5% in runoff reduction [17, 18] and 48–99% in Total Nitrogen (TN) change of mass [17, 19]. These differences highlight the sensitivity of BRC performance to different design parameters (such as varying growing media depth, composition of growing media and gravel storage layer depth), land-use, inflow pollutant concentration, and site specific hydrological conditions. Additionally, existing design guidelines offer a range of values for design parameters [2], which can make it difficult to arrive at an optimal site specific design.

BRC modelling is typically performed using the Stormwater Management Model (SWMM) software, which is a physics-based model. Physics based models utilize existing principles of physics to model a system [29], and are often computationally expensive to run requiring a significant allocation of time for models with a large number of simulations. Surrogate models (SMs) are one alternative to physics-based processes, offering a number of advantages to modelling and designing BRCs, which are discussed in detail below.

1.2 Surrogate Models for LID Modelling

Surrogate models otherwise known as meta-models, are simplified computationally inexpensive numerical approximations of the original model [20]. SMs are

constructed by first generating outputs using a set of known values from the original model (such as SWMM) and subsequently training the SM to predict the known output. SMs have been utilized extensively in the fields of hydrology and environmental modeling, including LID applications. For example, Brunetti et al. [5] evaluated the efficacy of using SMs in place of physics based models by using a Kriging model in place of Hydrus-2D for LID numerical analysis. In a study by Raei et al. [26], an Artificial Neural Network (ANN) was used in place of SWMM to evaluate LID design in an multi-objective optimization framework. Mohammadiun et al. [24] also utilized ANNs as a surrogate to SWMM in their multi-objective optimisation framework aimed at designing Urban Stormwater Drainage Systems. Their study favourably highlighted the large saving in computational time when using surrogate instead of the original models. Therefore, existing literature demonstrates the effectiveness and need for SMs in LID research and modelling. However, there are limitations associated with the use of SMs including: the stochastic nature of ANNs, meaning the same results cannot be replicated twice, creating uncertainty with regards to their predictions [30]. Additionally, data driven models such as ANNs only produce reliable predictions for input values which lie within the training sample's range. Further, uncertainties associated with selecting the number of input candidates, selecting the best input candidates for optimal performance and specifying the number of training samples per input remain. The focus of this research will be on this last component.

BRCs are commonly designed using synthetic design storms to evaluate their efficacy in reducing runoff and improving water quality. However, their design can be a computationally intensive task as the given range of design values in existing guidelines calls for a trial-and-error approach. As such, evaluating BRC design in this iterative process using traditional models such as SWMM is time consuming and may limit designers from further refining their designs due to the added computational complexity. Therefore, we propose using SMs in place of physical-based models to perform BRC designs.

1.3 Objectives

In this research, the main objective is to develop a robust, efficient and accurate SM that is capable of simulating SWMM results for runoff quantity. The surrogate is deemed a reliable substitute for SWMM if the quantified difference in model results is minimal. ANN's are selected in this research as the basis of the SMs, since they are widely used in rainfall-runoff simulations [9]. ANNs can be also be considered universal approximators, capable of simulating complex non-linear relationships [9], making them an ideal candidate for SMs.

Secondly, the optimal number of training samples required to achieve high ANN accuracy is investigated. ANN accuracy is highly dependent on the quantity of training samples provided at the beginning of model synthesis. Too few training

samples may hinder the ANN's ability to infer the system's underlying relationships accurately, which may cause the ANN to under or overpredict result values. Conversely, providing too many training samples may negatively impact the ANN's ability to generalize, resulting in overfitting, which renders the ANN unable to accurately predict results for input variables not yet encountered [9]. To investigate this, the number of training samples are varied from 50 to 200 for various rainfall events to determine the impact the number of training samples has on ANN performance. Further, the optimal number of samples required, beyond which including additional training samples would not significantly increase model performance, is determined.

Finally, the performance of a universal SM versus an event-based model is compared. A universal model promotes ease of use amongst designers and policy-makers as it facilitates design, owing to the fact that the only necessary inputs become the design parameters of the BRC and rainfall depth associated with a certain event. Developing a universal SM would further strengthen the case for using SMs as it can be shown that they maintain high performance and the ability to generalize under various conflicting rainfall depth inputs.

This research uses a BRC located near York University, in Toronto Canada, as the case study to develop and test the BRC SM. The following sections describe the study site, the development of the SM, synthesis of selected rainfall events, sampling methods and performance metrics chosen to evaluate SM performance.

2 Methods

All the calculations conducted in this study were performed using MATLAB 2020a. The SWMM model was developed using EPA SWMM version 5.1.015.

2.1 Study Site

The proposed BRC will be built at the intersection of Evelyn Wiggins Drive and Murray Ross Parkway, situated at York University, shown in Fig. 1b as subcatchment F. This location mainly receives runoff from the residential area south of York university as well as from the University itself. Therefore, the runoff is originating from a residential and mixed land-use area. The area draining into the BRC was quantified to be approximately 2000m². According to the design guidelines, the recommended ratio of upstream impervious area to BRC pervious area are 5:1 to 15:1 (Credit Valley Conservation (CVC) and Toronto and Region Conservation Authority (TRCA) [7]. Accordingly, the maximum area of the BRC was estimated to be 400m². Additionally, the design guidelines recommend a growing media depth of 1–1.25 m and a gravel storage layer depth of 300 mm, which were implemented in the BRC.

Table 1 Characteristics of subcatchments draining into the BRC

| Subcatchment ID | Subcatchment | Width (m) | Area (m ²) | Slope (%) |
|-----------------|--------------------|-----------|------------------------|-----------|
| A | Collection point 1 | 13 | 229 | 1.5 |
| B | Collection point 2 | 15 | 183 | 1.5 |
| C | Roadway | 83 | 1000 | 1 |
| D | Grass hill | 47 | 461 | 2 |
| E | Grass slope | 47 | 221 | 10 |
| F | BRC | 3 | Variable | N/A |

2.2 SWMM Model

To develop the hydrological model in SWMM, the BRC subcatchment characteristics were established using ESRI’s ArcGIS Pro. A 1 m resolution Digital Terrain Model (DTM) was obtained from the Government of Canada [15] and used to determine the slope, flow direction and flow accumulation. The BRC watershed was then delineated to accurately calculate and map out the area draining into the BRC. Subsequently, the SWMM model was developed to reflect the various subcatchments draining into the BRC, as outlined in Table 1 and shown in Fig. 1b.

Initial BRC design was built following the design guideline recommendations with a ponding depth of 300 mm, a growing media depth of 1000 mm and a gravel storage layer depth of 300 mm with an underdrain offset 100 mm from the bottom. Calibration of the SWMM model was not possible as the BRC has not been built yet. However, model parameters were selected to represent the characteristics of the surrounding area as closely as possible. It is important to note however, calibration of the model does not affect the objective of this paper since interest lies in investigating the SM’s ability to replicate any physical process.

2.3 Surrogate Model Synthesis

An ANN consists of a set of interconnected layers, namely the input, hidden and output layers, in a feedforward network. ANNs commonly consists of 3 layers, where the information in the first input layer is moved through to the nodes in the hidden layer and subsequently into the output later, as illustrated in Fig. 2. Each connection from the inputs to the hidden nodes has an associated multiplicative weight, w_k , which is iteratively adjusted at each training step to minimize the difference between the predicted model output and observed output. Biases, b_k , are also adjusted in a similar manner. This method of training is known as error backpropagation, which is the most common training algorithm across ANNs [9]. The number of hidden nodes is generally determined by a trial-and-error approach terminated based on a user specified performance level. Too few hidden neurons may hinder the ANNs ability

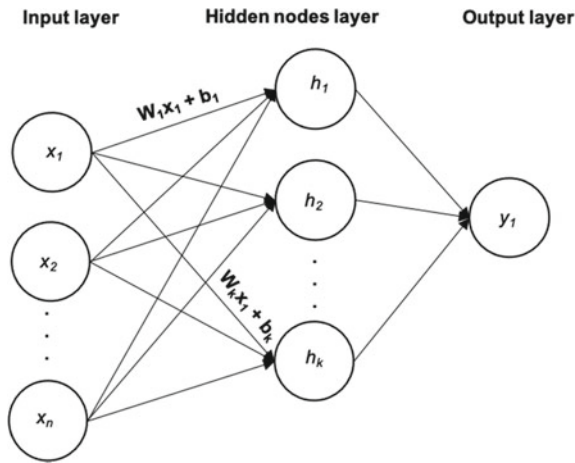


Fig. 2 Overview of ANN structure with: input nodes (x_n), hidden nodes (h_k), output node (y), weights (w_k) and biases (b_k)

to fully form connections that accurately describe the system. Conversely, too many hidden neurons can result in an ANN that requires excessive training computational effort [9].

The ANN in this study consists of 3 layers, with 6 inputs, 10 hidden nodes and one output to predict the runoff generated by the BRC. BRC area, growing media depth, storage layer depth, rainfall depth, rainfall average intensity and peak intensity are used as the input variables. Levenberg–Marquardt backpropagation is used as the training algorithm. It is often the fastest and more efficient training algorithm and thus commonly used in ANN training [27].

Commonly, the input dataset is subdivided into three subsets: training, validation and testing. The training data is used to adjust the network weights and biases, the validation subset is used as an early stopping technique to avoid overfitting, and the testing subset is used to quantify model performance. In this research, the training, validation and testing data split was 60–20–20% respectively.

In this research, an ensemble of 100 ANN's is created by randomly initializing the weights and biases 100 times and retraining the ANN, a procedure known as multi-start. An ensemble helps in capturing any uncertainties associated with using a single model, as well as the ensemble averaged output has less variability than what may be present in a single ANN [4].

2.4 Data Preparation

Selecting well-distributed training samples is critical to ANN development as it will facilitate the ANN's ability to interpolate or extrapolate values not yet encountered in future evaluations.

2.4.1 Rainfall Scenarios

To facilitate BRC design under various storm frequencies and durations, 18 different storms were selected for this study, 6 return periods and 3 durations. The return periods selected were 2, 5, 10, 25, 50 and 100-year storm events with durations of 6, 12 and 24 h each. These storms were selected as they are commonly used durations and return periods for designing hydrological infrastructure. The total rainfall depth was used to develop design storms following the Atmospheric Environmental Services (AES) Southern Ontario (SO) distributions. Utilizing AES-SO storms is advantageous over other distributions due to the fact that they were created based on Canadian data [21]. Total rainfall depths were obtained from the historical data for the Toronto North York Station (ID: 615S001) provided by the University of Western's climate change tool. Historical data fitted to the Gumbel distribution was selected, to facilitate in comparison with the historical values provided by the Ministry of Transportation. The historical information for the rain station is retrieved from Environment Canada's North York monitoring station.

2.4.2 Design of Experiments

One of the key components in the development of an SM is the selection of the initial sampling points. A well-distributed initial design set ensures the entire design space is explored [5, 13], and is well-represented. A number of sampling methods exist such as Random Sampling, Stratified Sampling and Latin Hypercube Sampling [23]. In this study, LHS has been used as the sampling method. LHS is an extension of stratified sampling first proposed by McKay et al. [23]. The range of each input X_i , $i = 1 \dots N$, is divided into N strata of equal marginal probability $1/N$, with only one sample taken from each stratum [23].

In this study, the number of input samples per rainfall event was selected to be 50, 100 and 200. MATLAB's built in *lhsdesign* function [22] was used to produce the stratified range for BRC area, growing media depth and storage layer depth. Initial values for BRC area ranged from 1 to 460 m² based on the recommended impervious to pervious ratio, the growing media depth values ranged from 1 to 1000 mm and the gravel storage depth values ranged from 200 to 300 mm, in accordance with the guidelines [7]. These initial minimum values were chosen to ensure the SM is exposed to the full spectrum of possible design parameter values. The gravel layer begins at 200 mm to allow for users to place an underdrain at least 100 mm away

from the bottom of the gravel storage. The inputs for all the rainfall events were then combined into one matrix to develop a universal SM. The total number of inputs per model was thus 900, 1800 and 3600. These inputs were first evaluated in SWMM to calculate the runoff generated in the study area. Subsequently, the inputs and output were grouped together and used to train the ANN.

2.5 Performance Metrics

To accurately gauge model performance, a number of performance metrics are used in this study: the Nash–Sutcliffe Efficiency (NSE), Continuous Ranked Probability Score (CRPS) and Percent Capture (PC). Given the nature of an ANN ensemble, its performance cannot be calculated using metrics such as NSE, therefore, given its deterministic nature, the NSE will be calculated using the averaged ensemble results for the test dataset per ANN model in the ensemble and finally averaged across all models. The CRPS is a probabilistic measure used for ensemble verification, taking into account both calibration (statistical compatibility between the ensemble forecasts and observed values) and sharpness (spread of a distribution) [4]. PC measures how frequently the observed value is found within the bounds of ensemble prediction, which can be indicative of whether a model is over or under predicting.

A value of 1 for NSE indicates perfect agreement between the observed and predicted values [27], which is desirable though difficult to achieve. Models with an NSE value greater than 0.75 are considered to have very good performance. A CRPS value of 0 is considered a perfect score, therefore, the aim is to minimise the CRPS to 0 as it has no upper bounds. Finally, a 100% percent capture indicates that the observed value always falls within the ensemble ranges.

3 Results and Discussion

Firstly, the use of an SM in place of SWMM for BRC design was evaluated. A number of BRC design configurations, obtained from the LHS stratified sampling, were simulated in SWMM to determine the amount of runoff leaving the BRC. These values were then used to train and evaluate BRC performance. The predicted SM runoff values, shown in Fig. 3, for the three models with 50, 100 and 200 inputs for the 10-year 24-h storm, are plotted against observed runoff values for multiple BRC area configurations, obtained from SWMM. The modelled storm is based on the Canadian Atmospheric Environment Service (AES) storm distribution. Area values go up to 400 m² in accordance with the impervious to pervious ratios dictated in the design guidelines. Given that the SM is an ensemble consisting of 100 ANNs, there are 100 different predictions for runoff values under each design configuration. These results are plotted in the light blue, depicting the ensemble spread. Averaged

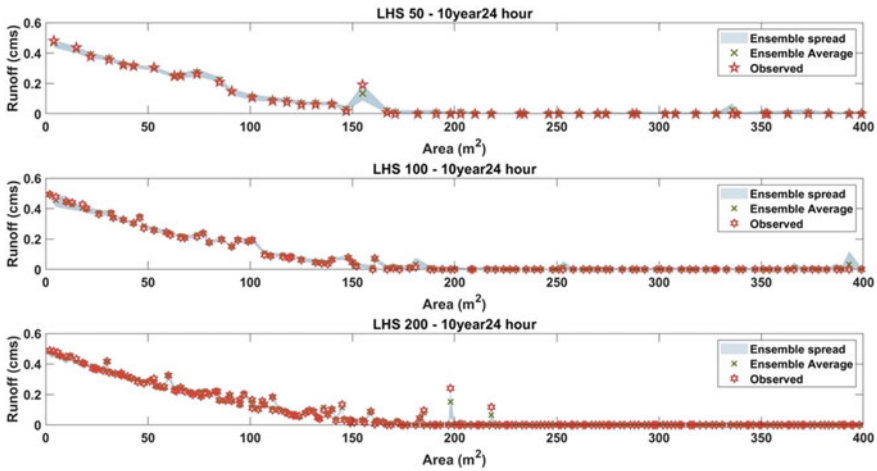


Fig. 3 Comparison plots of the models with different input size for the 10 year 24 h storm event. Top: SM with 50 inputs. Middle: SM with 100 inputs. Bottom: SM with 200 inputs

ensemble results are also plotted for comparison. The observed values are indicated by the red stars and the ensemble average is indicated by the green crosses.

3.1 Surrogate Model Performance in Place of SWMM

As is evident from the plots in Fig. 3, the averaged ensemble results, which represent the range of ANN runoff predictions, and observed values, i.e. actual runoff values obtained from SWMM, are very close with the exception of a few outliers. This indicates that averaging ensemble results can yield highly accurate results. Model performance in the outliers can be improved by including more high flow design configurations in the training dataset.

A more quantitative analysis of model performance using the three performance metrics: NSE, CRPS and PC is given in Table 2. As previously mentioned, the NSE was evaluated on the test dataset, which is completely independent from the training and validation dataset, providing a more accurate depiction of model performance. The CRPS and PC were calculated using the entire dataset as they are aimed at

Table 2 Performance metric results for the three models

| Model | NSE (unitless) | CRPS (cm) | PC (%) |
|-------------|----------------|-----------|--------|
| LHS 50 | 0.97 | 0.004 | 98.0 |
| LHS 100 | 0.98 | 0.003 | 95.0 |
| LHS 200 | 0.99 | 0.002 | 91.5 |
| Event based | 0.83 | 0.003 | 99.0 |

evaluating total ensemble performance. Based on the NSE results, the SM is able to simulate SWMM results with accuracy ranging from 97 to 99%. The very small CRPS results indicate that the ensemble range was mostly centered around the observed values with very minimal variance. Finally, the extremely high values for PC show that the observed values were usually within the ensemble range, indicating no bias within the SM. Both qualitative and quantitative results point towards the efficacy of using SMs in place of SWMM.

3.2 Model Performance Under Increasing Training Samples

SM performance was evaluated for 50 inputs, followed by 100 and 200. The observed values are plotted against the predicted ensemble range in Fig. 3. The figure indicates ensemble spread decreases as the number of inputs increase, which is expected as the more training samples the ANN has, the more precise the predicted results will be. Conversely, ensemble spread does not impact the accuracy of the averaged ensemble results.

While the figure comparison pointed to the LHS 200 model as the best model in terms of ensemble spread, the performance metrics are indicative of the opposite. All the models were extremely close in the performance metrics with no clear best model. The LHS 100 model is slightly better performing in terms of NSE and CRPS than LHS 50, however is lower for PC than LHS 50.

Prior to the development of an SM, all the inputs used for SM training must first be evaluated using the traditional model. This means that for each SM model, SWMM had to be run for 900, 1800 and 3600 times respectively to obtain the runoff values, which were then used as the target output for the SM. Thus, using less inputs to train the SM results in significant savings in computational time and effort as the original model is run for fewer evaluations. As such, given that the differences in performance between the three models is negligible, using 50 instances of each model input is recommended as the performance remains high without the added computational effort.

3.3 Universal and Event Based Surrogate Model

A comparison to determine whether an event based SM or universal SM is more accurate is performed. The universal model included all 18 storm events, and thus for the LHS 50 scenario, will include 900 training samples, whereas the event based surrogate will contain only the inputs and target output relevant to the modelled storm, in this case the 10-year, 24-h storm. The event based surrogate is also run using the multi-start method, therefore it is an ensemble with 100 ANNs. This facilitates comparison between the two models. The same inputs used to train the universal surrogate are

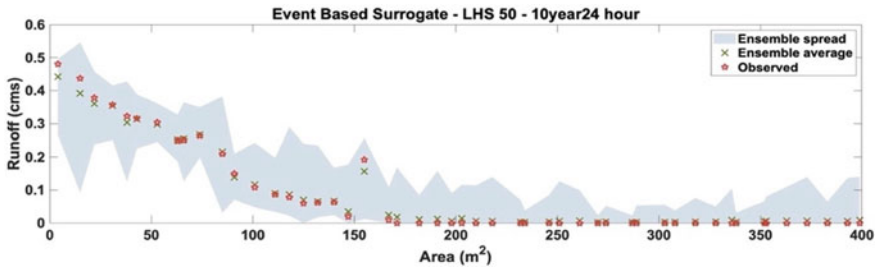


Fig. 4 Comparison plots of the universal and event based surrogate model. Top: Event based surrogate with LHS 50 sampling. Bottom: Universal surrogate with LHS 50 sampling

also used to train the event based surrogate. Based on previous performance results, the event based model was trained using 50 training samples.

A comparison of ensemble spread and averaged ensemble results plotted against observed values for both models is presented in Fig. 4 and the top plot in Fig. 3. Interestingly, the ensemble spread for the event based surrogate is much larger than that of the universal surrogate. This is due to the fact that the universal surrogate had 900 training samples and is thus able to more accurately discern which rainfall event is being modelled and its respective runoff value for each BRC design configuration. Additionally, as is clear from Fig. 4 due to the larger ensemble spread, the averaged runoff predictions are less accurate than those of the universal surrogate. Referring to Table 2, the event-based surrogate has a poorer NSE performance than the universal surrogate, and similar performances in CRPS and PC. These results indicate that a universal surrogate is more accurate than an event based surrogate. This is an important finding as universal surrogates are also more efficient for users, allowing users the opportunity to model BRC design under many different rainfall events in one simple click of a button.

By specifying a required runoff value and design storm, the corresponding area, growing media and storage layer depth can be determined instantaneously using the universal SM. This can aid policy-makers and designers with BRC design in a more efficient and expedient manner as many possible BRC design configurations can be obtained without the need to rely on slower physical-based models such as SWMM. This is demonstrated in Fig. 5 and the top plot of Fig. 3, where the user can specify a certain runoff value and find the corresponding BRC area, growing media depth and gravel storage layer depth. For example, for a runoff value of 0.2 cms, the approximate corresponding BRC area from Fig. 3 is 80 m² with a growing media depth of approximately 40 mm and a gravel storage depth of 250 mm, found in Fig. 5.

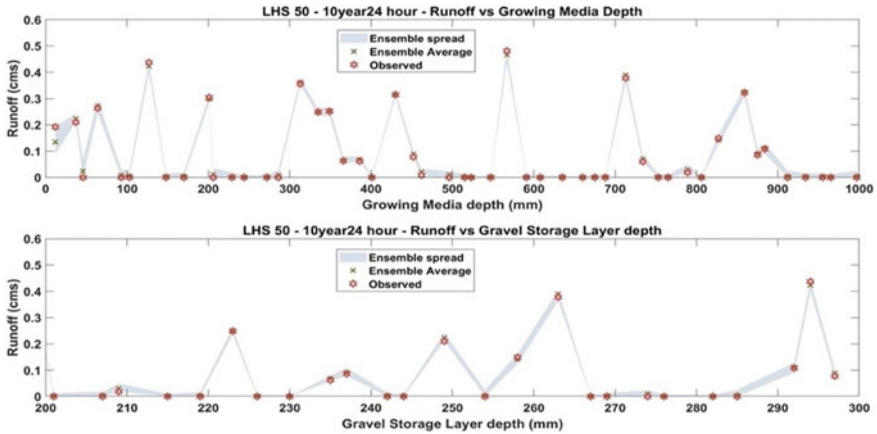


Fig. 5 A comparison plot of BRC design parameter values for given runoff values

4 Conclusions and Recommendations

In this research, a surrogate model was developed to be used for BRC design in place of SWMM. Using 50 training samples proved to be almost as accurate as using 100 and 200 samples per rainfall event. Therefore, it is recommended that surrogate models are trained using 50 samples as it much more efficient and less computationally intensive than using SWMM. Additionally, results showed that using an event based surrogate model is not as accurate as using a universal surrogate. This is an important finding as it allows users to easily switch between rainfall events, making the BRC design process more efficient and streamlined.

Universal model performance may be improved by ensuring the model is trained on more high flow scenarios to eliminate the outliers found. Future research should focus on determining the lowest possible amount of training samples without sacrificing the achieved accuracy of an LHS 50 model.

References

1. Ahiablame LM, Engel BA, Chaubey I (2012) Effectiveness of low impact development practices: literature review and suggestions for future research. *Water Air Soil Pollut* 223(7):4253–4273. <https://doi.org/10.1007/s11270-012-1189-2>
2. Bacys M, Khan UT, Sharma J, Bentzen TR (2019) hydrological efficacy of ontario's bioretention cell design recommendations: a case study from North York, Ontario. *J Water Manage Model*. <https://doi.org/10.14796/jwmm.c468>
3. Bedan ES, Clausen JC (2009) Stormwater runoff quality and quantity from traditional and low impact development watersheds. *JAWRA J Am Water Resour Assoc* 45(4):998–1008. <https://doi.org/10.1111/j.1752-1688.2009.00342.x>

4. Boucher MA, Perreault L, Anctil F (2009) Tools for the assessment of hydro logical ensemble forecasts obtained by neural networks. *J Hydroinfr* 11(3–4):297–307. <https://doi.org/10.2166/hydro.2009.037>
5. Brunetti G, Šimůnek J, Turco M, Piro P (2017) On the use of surrogate-based modeling for the numerical analysis of low impact development techniques. *J Hydrol* 548(May):263–277. <https://doi.org/10.1016/j.jhydrol.2017.03.013>
6. City of Toronto (2017) Toronto green streets technical guidelines
7. Credit Valley Conservation (CVC), and Toronto and Region Conservation Authority (TRCA) (2010) Low impact development stormwater management planning and design guide. https://cvc.ca/wp-content/uploads/2014/04/LID-SWM-Guide-v1.0_2010_1_no-appendices.pdf
8. Davis AP (2005) Green engineering principles promote low-impact development. *Environ Sci Technol*. <https://doi.org/10.1021/es053327e>
9. Dawson CW, Wilby RL (2001) hydrological modelling using artificial neural networks. *Progress Phys Geogr Earth Environ* 25(1):80–108. <https://doi.org/10.1177/030913330102500104>
10. Dietz ME (2007) Low impact development practices: a review of current research and recommendations for future directions. *Water Air Soil Pollution*. <https://doi.org/10.1007/s11270-007-9484-z>
11. Dietz ME, Clausen JC (2008) stormwater runoff and export changes with development in a traditional and low impact subdivision. *J Environ Manage* 87(4):560–566. <https://doi.org/10.1016/j.jenvman.2007.03.026>
12. Eckart K, McPhee Z, Bolisetti T (2017) Performance and implementation of low impact development—a review. *Sci Total Environ*. <https://doi.org/10.1016/j.scitotenv.2017.06.254>
13. Garud SS, Karimi IA, Kraft M (2017) Design of computer experiments: a review. *Computers Chem Eng*. <https://doi.org/10.1016/j.compchemeng.2017.05.010>
14. Gaur A, Gaur A, Yamazaki D, Simonovic SP (2019) Flooding related consequences of climate change on Canadian cities and flow regulation infrastructure. *Water* 11(1):63. <https://doi.org/10.3390/w11010063>
15. Government of Canada (2019) High Resolution digital elevation model (HRDEM)—CanElevation Series—Open Government Portal. 2019. <https://open.canada.ca/data/en/dataset/957782bf-847c-4644-a757-e383c0057995>
16. IPCC (2013) Summary for policymakers. In: *Climate Change 2013: The physical science basis*. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge
17. Jia H, Wang X, Ti C, Zhai T, Field R, Tafuri AN, Cai H, Yu SL (2015) Field Monitoring of a LID-BMP treatment train system in China. *Environ Monit Assess* 187(6):373. <https://doi.org/10.1007/s10661-015-4595-2>
18. Khan UT, Valeo C, Chu A, van Duin B (2012) Bioretention cell efficacy in cold climates: part 1—hydrologic performance. *Can J Civ Eng* 39(11):1210–1221. <https://doi.org/10.1139/I2012-110>
19. Khan UT, Valeo C, Chu A, van Duin B (2012) Bioretention cell efficacy in cold climates: part 2—water quality performance. *Can J Civ Eng* 39(11):1222–1233. <https://doi.org/10.1139/I2012-111>
20. Khu S-T, Savic D, Liu Y, Madsen H (2004) A fast evolutionary-based meta-modelling approach for the calibration of a rainfall-runoff model. In: *Transactions of the 2nd Biennial Meeting of the International Environmental Modelling and Software Society (IEMSS)*. Manno, Switzerland
21. Marsalek J, Watt WE (1984) Design storms for urban drainage design. *Can J Civ Eng* 11(3):574–584. <https://doi.org/10.1139/I84-075>
22. MathWorks Inc. (2021) Latin hypercube sample—MATLAB Lhsdesign—MATLAB & Simulink. Mathworks Simulink. <https://www.mathworks.com/help/stats/lhsdesign.html>
23. McKay MD, Beckman RJ, Conover WJ (1979) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21(2):239. <https://doi.org/10.2307/1268522>

24. Mohammadiun S, Yazdi J, Salehi Neyshabouri SAA, Sadiq R (2018) Development of a stochastic framework to design/rehabilitate urban stormwater drainage systems based on a resilient approach. *Urban Water J* 15(2):167–176. <https://doi.org/10.1080/1573062X.2018.1424218>
25. Qin HP, Zhuo XL, Fu G (2013) The effects of low impact development on urban flooding under different rainfall characteristics. *J Environ Manage* 129:577–585. <https://doi.org/10.1016/j.jenvman.2013.08.026>
26. Raei E, Alizadeh MR, Nikoo MR, Adamowski J (2019) Multi-objective decision-making for green infrastructure planning (LID-BMPs) in urban storm water management under uncertainty. *J Hydrol* 579:124091. <https://doi.org/10.1016/j.jhydrol.2019.124091>
27. Snieder E, Shakir R, Khan UT (2020) A comprehensive comparison of four input variable selection methods for artificial neural network flow forecasting models. *J Hydrol* 583:124299. <https://doi.org/10.1016/j.jhydrol.2019.124299>
28. Sohn W, Kim JH, Li MH, Brown R (2019) The influence of climate on the effectiveness of low impact development: a systematic review. *J Environ Manage*. <https://doi.org/10.1016/j.jenvman.2018.11.041>
29. Solomatine DP, Ostfeld A (2008) Data-driven modelling: some past experiences and new approaches. *J Hydroinform* 10:3–22. <https://doi.org/10.2166/hydro.2008.015>
30. Zhang W, Li J, Chen Y, Li Y (2019) A Surrogate-Based Optimization design and uncertainty analysis for urban flood mitigation. *Water Resour Manage* 33(12):4201–4214. <https://doi.org/10.1007/s11269-019-02355-z>

Segregation of Particles in Multi Size and Density Beds by Circular Wall Jets



H. Hernandez, A. Mostaani, and A. H. Azimi

1 Introduction

Prediction the geometrical characteristics of scour and deposition such as depth, width, and length downstream of hydraulic structures such as, spillways, gates, weirs, and jets is of utmost importance from the engineering point of view [4, 5, 15, 16]. The local erosion of erodible beds considerably affects the stability of hydraulic structures and bridge piers. Moreover, understanding the scour process is important in many environmental and industrial applications [17]. However, the sediment segregation process in mineral and coal mining operations by a 3D wall jet has not been thoroughly analyzed to date. The bed materials in natural streams can be transported and separated using more environmental friendly methods to separate in mining operations that will help environmental conservation.

Erosion induced by different types of jets has been studied for a variety of hydraulic conditions [6, 8, 20, 21]. Dimensional analysis has been used to predict the scour characteristic lengths downstream of submerged turbulent wall jets at the equilibrium stage [2, 6, 11]. However, less attempts have been made to investigate the scouring caused by a circular wall jet in confined channels. Some empirical equations have been developed to predict the scour dimensions using by employing dimensional analysis and most empirical models correlated scour dimensions with densimetric Froude number. The densimetric Froude number describes the relationship between the tractive and resistive forces on a single particle and is defined

H. Hernandez · A. Mostaani · A. H. Azimi (✉)

Department of Civil Engineering, Lakehead University, Thunder Bay, ON, Canada

e-mail: azimi@lakeheadu.ca

H. Hernandez

e-mail: halexan1@lakeheadu.ca

A. Mostaani

e-mail: amostaan@lakeheadu.ca

© Canadian Society for Civil Engineering 2022

S. Walbridge et al. (eds.), *Proceedings of the Canadian Society of Civil Engineering*

Annual Conference 2021, Lecture Notes in Civil Engineering 250,

https://doi.org/10.1007/978-981-19-1065-4_18

as $F_{De} = u_o/(gD_e\Delta\rho/\rho)^{1/2}$, where u_o is the mean jet velocity at the nozzle, D_e is the effective diameter of the bed particles, g is the acceleration due to gravity, ρ is the density of water, and $\Delta\rho$ is the density difference between the bed material and water. Recent studies have shown that other non-dimensional parameters such as the relationship between the particle diameter and the nozzle diameter, tailwater depth, outlet size, relative downstream channel width should be used in combination with the densimetric Froude number to properly predict the scour dimensions [7, 18, 23, 24].

As the jet advances over erodible beds, diffusion takes place rapidly and the thickness of the jet increases along the direction of the jet. The centerline velocity of the jet reduces due to intermixing of the jet with the ambient [9]. The jet diffusion is affected by the tailwater and hence the erosion process takes place. The deeply submerged wall jets allow almost unbounded diffusion in the streamwise, vertical, and transverse directions downstream of the jet. Lim [14] reported a distorted scour hole as a result of a very narrow channel and a strong recirculating flow at both sides of the scour hole. As the tailwater decreases, the jet diffusion increases which results in inducing more turbulence and higher backflow. The backflow effect alters the shape of scours and the scour hole no longer conserves the shape that was observed in deeply submerged conditions. It was also observed that at low tailwater depth, the upward jet expansion is limited, and the water surface absorbs a portion of the energy produced by the jet. The strong backflow at both sides of the nozzle exit is influenced by the channel width and it causes additional scour near the vicinity of the nozzle outlet.

The time evolution of local scour induced by submerged wall jets can significantly be affected by the bed particle size and the interlocking between particles [3, 9, 10, 13, 19–21]. In multi-size beds, the lighter and smaller particles are lifted from the bed and transported away by the mean flow of the jet. On the other hand, the larger and heavier particles remain in place due to higher resistive force. Therefore, over the time sediment segregation takes place in beds with wide sediment size distribution. The mixed particle beds promote the formation of a layer of coarse particles at the bed surface, known as armor coat, with a few sheltered small grains. The armor layer progressively increases the effective critical bed shear stress and reduces the expansion rate of scour hole over time [22]. Aderibigde and Rajaratnam [2] reported that the armor layer reduces the scour depth by 50%, and a linear correlation was developed between the scour depth and the densimetric Froude number F_{D95} , based on a particle size that 95% of the particles are smaller, D_{95} , to predict the maximum scour depth.

This study presents the results of erosion in loose beds caused by jets with a circular nozzle. The erosion experiments were tested in uniform, and multi-density and size beds with different tailwater depths and in clear water conditions. The main objective of the present study was to investigate the effect of tailwater depth and densimetric Froude number on the scour characteristics of multi-density and uniform loose beds. Such bed mixtures with different sizes and densities are naturally formed in some rivers and tailing ponds resulted from mining operations. The results of this study can be used for protection of local scour around bridge piers. The prediction

of the characteristic dimensions of scour hole will help to avoid possible failure in design and operation of hydraulic structures. Moreover, the full understanding of scour phenomenon could help mining operations to separate heavy particles from a multi-density mixture during extraction process. The controlled application of a circular wall jet helps the separation of heavy particles from sand that can be used for the optimization of mineral separation during preliminary extraction processes.

2 Experimental Setup

Erosion over multi-size and density beds were analyzed through a series of experiments conducted in the Multiphase Flow Research Laboratory (MFRL) at Lakehead University. Figure 1 shows the sketch of the experimental setup and parameters. A Glass-walled tank of 2.0-m-long, 0.46-m-wide, and 0.48-m-high was employed. A 1.0-m-long sediment bed section was used to study scour formation and the bed section was leveled with the invert of the nozzle outlet to form a wall jet with no offset. The sediment section depth was 0.14 m with the same width as the channel width ($B = 0.46$ m). Circular nozzles with a diameter of $d_o = 0.0127$ and 0.0254 m were used for wall jets and different flow rates ranging between $Q = 22.2$ and 32.1 L/min were tested. A sharp-crested weir with a variable height was placed downstream of the sediment bed section to control the water level, y_t . Three tailwater depths were used in this study with the tailwater depths of $y_t = 0.0508, 0.1016,$ and 0.1524 m.

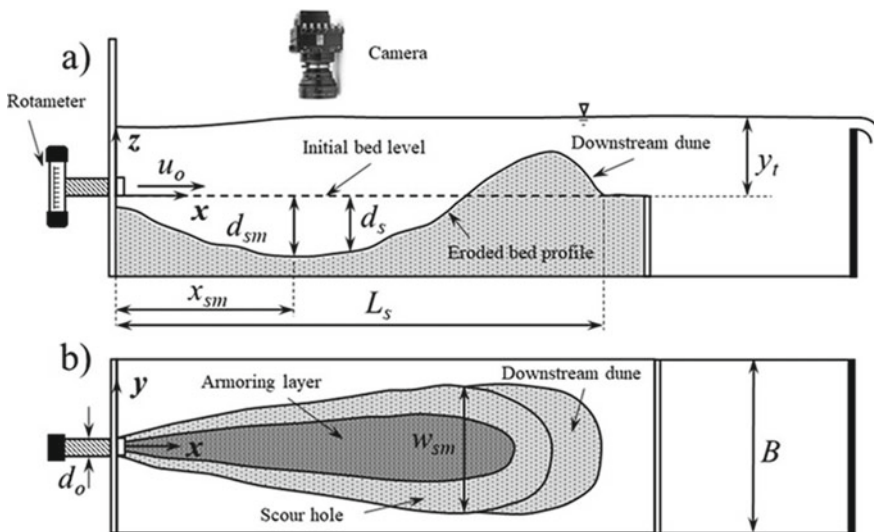


Fig. 1 Schematic of the experimental setup and coordinate system for the erosion of uniform and multi-size and density sediment beds downstream of a circular jet: **a** side view; **b** plan view

A uniform sand bed with a $D_{50} = 0.38$ mm was used as a benchmark to study the armoring effect, scour protection of the mixed bed downstream of 3D wall jets, and the material segregation of minerals present on the mixture bed. The mixture bed with a $D_{50} = 0.29$ mm consists of four particles as Lead balls, Lead particles, Magnetite particles, and Sand particles with different densities and sizes. A submersible pump was used to generate flow with different discharges and it was placed far from the tank to avoid transmitting any vibrations from the pump to the test section. The jet discharge, Q , was controlled by a valve, and flow discharge was measured using an accurate rotameter (FCH-C-PA, BioTech e.k, Vilshofen, Germany). The water temperature was kept constant to $20\text{ }^{\circ}\text{C} \pm 1\text{ }^{\circ}\text{C}$ for all experiments.

Great care was taken to monitor the local motion of particles to identify the steady-state flow condition and termination of each test. When the steady-state flow condition was achieved, the scour geometries were constant for 60 min, and the test was terminated. The overall experimental time for each test was approximately 8 h. A high-resolution camera (Photron Fastcam, 1024-PCI 100 KC, Tokyo, Japan) was placed above the flume to capture the images of the scour evolution with time. The LED strip lights (Zorsen LED 5050, Aseen, China) were placed around the flume to improve the quality of images and to set high contrast threshold between each type of particles. The scour depth, d_s , was measured in 0.1 m distance intervals using a point gauge with an accuracy of ± 0.1 mm. The maximum width, w_{sm} , and length, L_s , were extracted using the image processing toolbox developed in MATLAB (MathWorks, 2018. version 9.7.0, Natick, USA) as well as the boundaries of different particles (see Fig. 1). Details of the experimental parameters are stated in [12].

3 Results and Discussion

Figure 2 shows the temporal evolution of the scour hole generated by a circular wall jet over a uniform bed. The erosion starts almost immediately once the jet contacts the bed as the flow shear stress exceeds the critical shear stress of the bed particles. At the early stage, the jet starts to produce a narrow hole and mainly undermines the bed. During the first hour of experiments, noticeable growth was observed in the characteristic dimensions of the scour. The scour hole formed a symmetric elliptical shape [6, 8, 20, 21] along with a uniform ridge with a high sediment deposition after the ridge crest. After the first hour from the onset of experiment, the scour hole was almost developed and started to grow mainly in width and depth as well as accumulating more sediment behind the ridge crest. Once the scour width became comparable with the tank width (i.e., $t = 240$ min), the growth rate of scour hole significantly reduced. For partially and deeply submerged wall jets, no ripples were observed and the jet produced more defined scour holes and less erosion near the jet exit since the jet had a higher capacity to erode the bed due to the lower velocity diffusion.

Figure 3 depicts the erosion process of multi-size and density beds induced by a circular wall jet. Similar to uniform sand beds (see Fig. 2), the erosion started almost

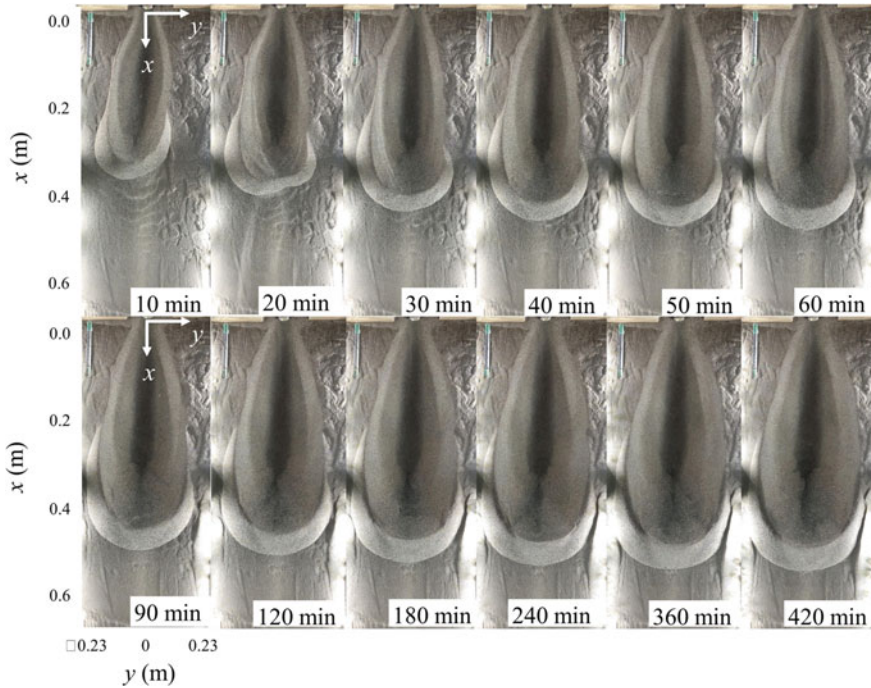


Fig. 2 Top view snap-shot images on the temporal evolution of scour hole over uniform sand bed by a circular wall jet with $d_o = 25.4$ mm, flow rate of $Q = 28.5$ L/min, and submergence ratio of $y_i/d_o = 4.0$

immediately in multi-size beds and lighter bed materials began to separate from the bed. The top view of scour profiles showed an irregular semi-elliptical scour shape that varied accordingly with the discharge [1, 19]. Once the erosion begins, the jet energy dissipates over the bed and the flow momentum separates and congregates particles of the same size and density in the scour hole and forms the armor layer on the surface of the scour hole. The armor layer is composed of Lead balls, Lead, and Magnetite particles. Due to the size and density of the Lead balls, they were always located in the center region of the scour hole, and the Lead and Magnetite particles surrounded the Lead balls to form different rings (see Fig. 3).

The first ring is composed of the Lead balls (dark gray), the second ring is composed of the Lead and Magnetite particles (light gray), and the third ring is composed of sand particles. The ring area formed by the Lead and Magnetite particles increased with decreasing tailwater depth by approximately 6% while the sand area did not change significantly with tailwater depth variations. The exerted jet subjected to the multi-size bed produces a narrower scour hole with two-sequence equilibrium scour depths and an irregular front dune. The first scour depth which is the maximum scour depth (i.e., d_{sm}) is located near the jet nozzle while the subsequent scour depth is shifted to the downstream side of the scour hole near the ridge.

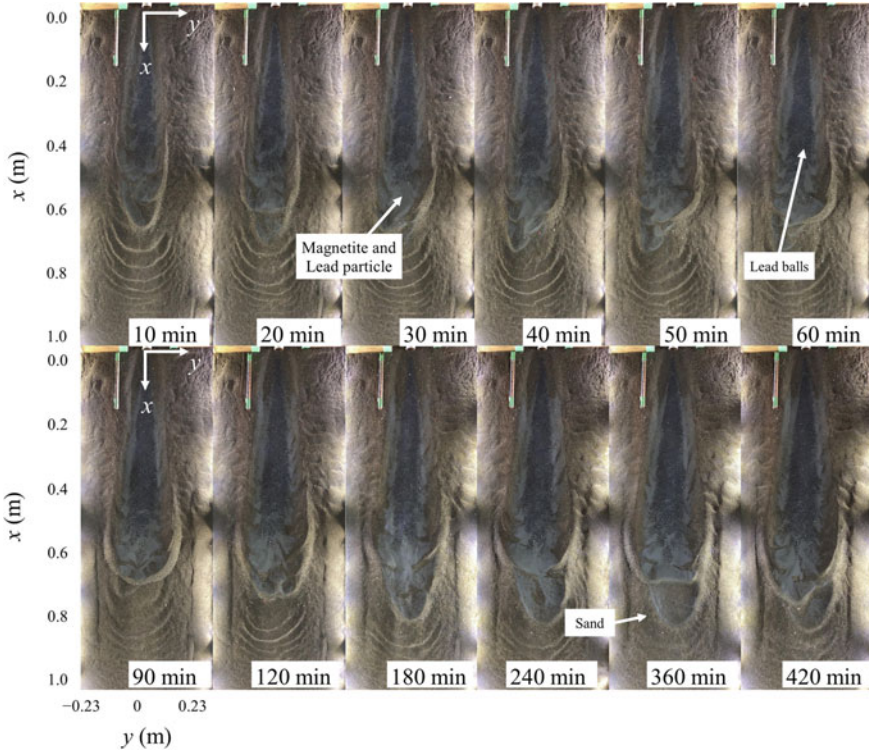


Fig. 3 Top view snap-shot images on the temporal evolution of scour hole over a multi-density and size bed by a circular wall jet with $d_o = 25.4$ mm, flow rate of $Q = 28.5$ L/min, and submergence ratio of $y_i/d_o = 4.0$

The limited lateral jet expansion and the restriction by tailwater depth also affected the ridge shape showing a flattened and irregular ridge. The scour hole in multi-size and density beds was almost developed at $t = 30$ min from the onset of the test, and the growth rate of scour was slower than the uniform bed due to formation of armor layer. At $t = 120$ min, the scour hole characteristic lengths did not considerably change, and the jet solely modified the shape of the particle rings.

Figure 4 shows the correlations of the maximum scour depth with densimetric Froude number in the equilibrium state (i.e., $t = 420$ min) to develop a model for prediction of the equilibrium scour depth at the centerline of scour. As was reported by Aderibigde and Rajaratnam [2], the lighter bed materials moved out from the scour hole, and the coarse materials (i.e., Lead balls) are mainly deposited at the scour maximum depth. Hence, the determination of the effective diameter that better represents the maximum scour hole was necessary. Figure 4a shows the correlation between the normalized maximum scour depth and densimetric Froude number based on the median grain size D_{50} , and the grain size for which 85% of materials are finer D_{85} , respectively. The scatter in Fig. 4a shows that both F_{D50} and F_{D85}

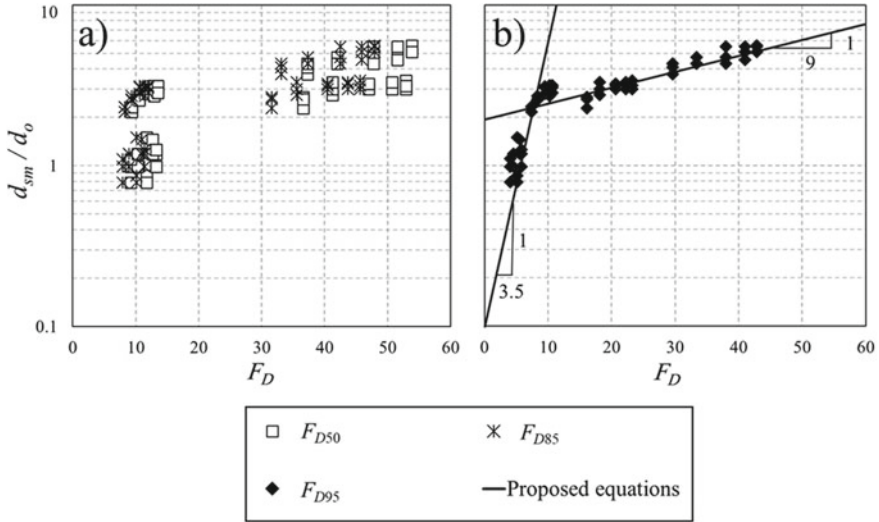


Fig. 4 Performance of the proposed model for prediction of the equilibrium scour depth downstream of a circular jet for both uniform and multi-size and density particle beds

are not suitable for prediction of scour depth. Figure 4b shows the correlation of the normalized maximum scours depth with densimetric Froude number F_{D95} . The densimetric Froude number calculated is based on the grain size for which 95% of materials are finer (D_{95}). As can be seen in Fig. 4b, for lower densimetric Froude numbers (i.e., $F_{D95} \leq 12$) the rate of increase is higher for the normalized maximum scour depth than the rate of increase for higher densimetric Froude numbers (i.e., $F_{D95} > 12$). Figure 4b shows that most of the data are close together in a well-defined band showing a strong correlation between d_{sm}/d_o and F_{D95} which can be expressed as:

$$d_{sm}/d_o = 0.28F_{D95} \quad \text{for } F_{D95} \leq 12 \tag{1}$$

$$d_{sm}/d_o = 0.28F_{D95} \quad \text{for } F_{D95} > 12 \tag{2}$$

In Eqs. 1 and 2, the coefficient of determinations are $R^2 = 0.978$ and 0.920 , respectively. It is reasonable to use a coarser particle diameter that representative of the bed mixture after the erosion due to the coarser particles remained on the eroded bed forming a coarse layer and the finer particles on the top layer of the eroded bed were washed away by turbulent resuspension mechanism.

Figure 5 shows the variation of maximum scours width, w_{sm} , normalized with the nozzle size, d_o , with densimetric Froude number F_D . Similar to Fig. 4, the choice of effective diameter better described the maximum width of the scour. The use of the mean diameter, D_{50} , showed the best relationship between the normalized maximum width with the nozzle diameter w_{sm}/d_o and the densimetric Froude number F_{D50} .

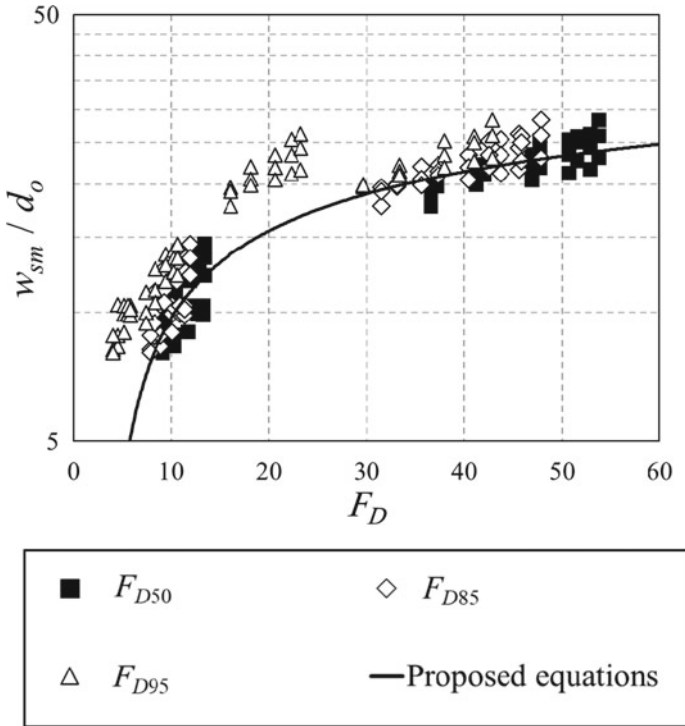


Fig. 5 Performance of the proposed equation for prediction of maximum scour width in equilibrium stage downstream of a circular jet for both uniform sand and multi-size and density particle beds and different submergence ratios

The proposed equation of the maximum scour width can be described as:

$$w_{sm}/d_o = 8.45 \ln(F_{D50}) - 9.77 \tag{3}$$

In Eq. 3 the coefficient of determinations is $R^2 = 0.934$. The different variations of normalized maximum width with densimetric Froude number proposed F_{D85} and F_{D95} showed bigger scatter making undesired to be use for prediction of the characteristic dimension of the scour hole.

As it is depicted in Fig. 6, the variation of the maximum length of the scour hole, L_s , for both multi-size and density bed and uniform beds can be analyzed by plotting L_s/d_o as a function of densimetric Froude number. The process to find the effective diameter that best describes the relationship between these two variables was the same as in Fig. 4.

It was found that the diameter that shows the lower scatter is the densimetric Froude number calculated with the particles mean diameter D_{50} and the proposed prediction equation of maximum scour length can be described as:

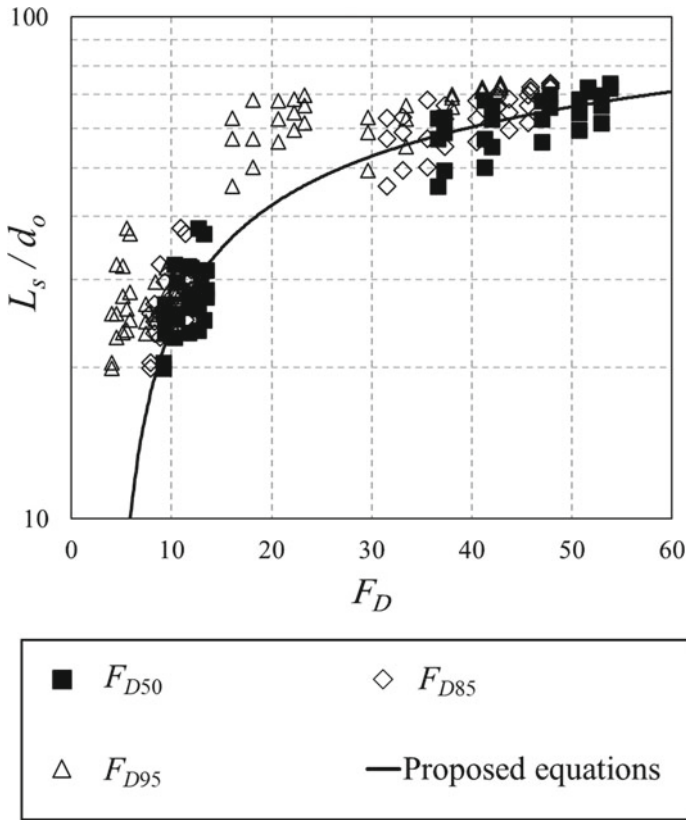


Fig. 6 Performance of the proposed equation for prediction of the scour length in equilibrium stage downstream of a circular jet for both uniform and multi-size and density particle beds and different submergence ratios

$$L_s/d_o = 23.52 \ln(F_{D50}) - 30.0 \tag{4}$$

with a coefficient of determination of $R^2 = 0.920$. The scour hole is better described with the median particle size (i.e., D_{50}) and density of particles rather than large and heavy particles since finer particles are deposited on the sides of the scour so the growth of the scour width and length are better described with the median size particles.

4 Conclusions

In this study, the temporal evolution of uniform and multi-size and density particle beds and their equilibrium scour dimensions induced by circular jet in a confined

channel were studied. The temporal evolution of the multi-size bed scour hole showed that an armor layer formed over time which limited the scour dimensions. It was found that the erosion of multi-size and density particle beds reached the equilibrium stage faster than the uniform bed due to the armoring effect. It was observed that the scour formation in uniform sand beds were symmetric elliptical shape while in multi-density beds, the scour hole formed an irregular asymmetric pit. In the multi-size and density beds, the armor layer consisted of Lead balls, Lead, and Magnetite particles which were separated by the jet's momentum and congregated over certain areas on the scour hole. The area occupied by the Lead and Magnetite particles increased with decreasing tailwater level by approximately 6% while the effect of tailwater on the magnitude of the sand area was negligible.

The 3D wall jet produced a narrower scour hole in the multi-size bed with two-sequence scour depths which the first one was the maximum scour depth (i.e., d_{sm}). The effect of different parameters in predicting the scour dimensions in the equilibrium stage were evaluated and it was found that densimetric Froude number, which represents the interaction between drag and friction forces on a particle, was the dominant parameter. The scour dimensions showed that densimetric Froude number based on median grain size D_{50} (i.e., F_{D50}) was suitable for predicting the maximum scour width and length while, densimetric Froude number based on the grain size for which 95% of materials are finer (i.e., F_{D95}) reasonably predicted the maximum scour depth since F_{D95} represents the effect of armor layer over the erodible bed. Accordingly, empirical prediction equations based on nonlinear regression analysis were proposed in order to predict the scour hole maximum depth, width, and length in the equilibrium stage.

References

1. Ade F, Rajaratnam N (1998) Generalized study of erosion by circular horizontal turbulent jets. *J Hydraul Res* 36(4):613–636
2. Aderibigbe O, Rajaratnam N (1998) Effect of sediment gradation on erosion by plane turbulent wall jets. *J Hydraul Eng* 124(10):1034–1042
3. Ali KHM, Lim SY (1986) Local scour caused by submerged wall jets. *Proc Inst Civ Eng* 81(4):607–645
4. Azimi AH, Zhu DZ, Rajaratnam N (2014) An experimental study of sand deposition from slurry wall jets. *J Eng Mech* 140(2):296–314
5. Azimi AH, Qian Y, Zhu DZ, Rajaratnam N (2015) An experimental study of circular sand–water wall jets. *Int J Multiph Flow* 74:34–44
6. Balachandar R, Kells JA, Thiessen RJ (2000) The effect of tailwater depth on the dynamics of local scour. *Can J Civ Eng* 27(1):138–150
7. Bashiri H, Sharifi E, Singh V (2018) Prediction of local scour depth downstream of sluice gates using harmony search algorithm and artificial neural networks. *J Irrig Drain Eng* 144(5):06018002
8. Blaisdell FW, Anderson CL (1988) A comprehensive generalized study of scour at cantilevered pipe outlets. *J Hydraul Res* 26(4):357–376
9. Chatterjee SS, Ghosh SN, Chatterjee M (1994) Local scour due to submerged horizontal jet. *J Hydraul Eng* 120(8):973–992

10. Dey S, Sarkar A (2006) Scour downstream of an apron due to submerged horizontal jets. *J Hydraul Eng* 132(3):246–257
11. Dey S, Sarkar A (2007) Effect of upward seepage on scour and flow downstream of an apron due to submerged jets. *J Hydraul Eng* 133(1):59
12. Hernandez H (2021) Experimental study of local scour formation on multi-density beds by a circular wall jet. M.Sc. Thesis, Department of Civil Engineering, Lakehead University, 118p
13. Kells JA, Balachandar R, Hagel KP (2001) Effect of grain size on local channel scour below a sluice gate. *Can J Civ Eng* 28(3):440–451
14. Lim SY (1995) Scour below unsubmerged full-flowing culvert outlets. *Proc Inst Civil Eng Water Maritime Energy* 112(2):136–149
15. Manzouri M, Azimi AH (2019) A study of mound formation by discharging sand particles through oblique pipes in stagnant water. *Int J Sedim Res* 34(6):564–576
16. Manzouri M, Azimi AH (2019b) Effects on oily sand jet evolution from impact momentum and channelization of particles through an immiscible interface. *Int J Multiphase Flow* 121:103124
17. Martino RG, Ciani FG, Paterson A, Piva MF (2019) Experimental study on the scour due to a water jet subjected to lateral confinement. *Eur J Mech B/Fluids* 75:219–227
18. Melville BW, Lim SY (2014) Scour caused by 2D horizontal jets. *J Hydraul Eng* 140(2):149–155
19. Rajaratnam N, Berry B (1977) Erosion by circular turbulent wall jets. *J Hydraul Res* 15(3):277–289
20. Rajaratnam N (1981) Erosion by plane turbulent jets. *J Hydraul Res* 19(4):339–358
21. Rajaratnam N, Macdougall RK (1983) Erosion by plane wall jets with minimum tailwater. *J Hydraul Eng* 109(7):1061–1064
22. Raudkivi AJ (1990) *Loose boundary hydraulics*, 3rd edn. Pergamon Press Inc., Tarrytown
23. Yu G, Tan SK (2007) Estimation of boundary shear stress distribution in open channels using flownet. *J Hydraul Res* 45(4):486–496
24. Zhao P, Yu G, Zhang M (2019) Local scour on noncohesive beds by a submerged horizontal circular wall jet. *J Hydraul Eng* 145(9):06019012

A Multiple Linear Regression (MLR) Model for the Application of Electrical Vehicles in the United States



L. Navarro and B. Bathaei

1 Introduction

Around 20% of the world's petroleum is consumed by the U.S. [9]. While 70% of the consumed petroleum is for transportation systems [30]. In addition, 28% of greenhouse gases (GHGs) emissions of the U.S. are from the transportation sector [31]. Due to an increase in vehicle ownership and demand, some policies are required to control the potential impacts of transportation on GHGs emissions [25]. The major contributions of GHGs emission mitigation are reducing air pollution, improving human health and the natural ecosystem, and conserving finite resources of fossil fuels [1, 12, 20, 22]. These contributions greatly pave the path toward achieving a more sustainable urban development. Urban sustainability requires different sustainable elements, which GHGs emission mitigation is one of those elements, to ensure a sustainable environment [4, 5]. Based on the Clean, Low-Emission, Affordable New Transportation Efficiency Act (CLEAN-TEA) introduced in March 2009, determining GHGs mitigation strategies and policies from U.S. local governments of the transportation sector are required [17]. These strategies are classified into three groups including reducing (reducing GHG emissions), avoiding (avoiding unnecessary energy consumption and promote other modes of transportation), and replacing (replacing fossil fuels with low-emission alternative fuels) [20, 21]. This study focuses on replacing the strategy of EVs. The main contributions of EVs are reducing vehicle operating costs, reducing GHG emissions, air pollution, and the dependency on fossil fuel [7]. The current research aims to explore potential factors that can be attributed to the penetration of Hybrid/Electric Vehicles in the U.S. In this regard, several Multiple Linear Regression (MLR) models are applied to find the significant

L. Navarro (✉) · B. Bathaei
The University of Texas Rio Grande Valley, Edinburg, TX, USA
e-mail: linda.navarro01@utrgv.edu

factors that impact the use of different types of hybrid vehicles compared to conventional vehicles. The types of hybrid vehicles assessed are Plug-in Hybrid Vehicles (PHEV), electric vehicles (EV), hybrid vehicles (HEV). This section is followed by a literature review, which is then succeeded by the data and variables description. Research methodology, results, and model validation are presented in subsequent sections. Finally, the last section provides the conclusion of the study.

2 Litreture Review

Despite the environmental benefits [2, 11, 19] and less financially related benefits [13, 27] of electric drive vehicles (including EVs, PHEVs, and hybrid-electric vehicles), adaptation and penetration of electric vehicles (EVs) in the market are one of the main concerns. Some scholars believe that some factors such as financial, battery-related concerns, and potential inconvenience of recharging are the main obstacles for penetration of them in the market [8, 15, 29]. For example, it is found out that three factors of (i) significant savings on monthly fuel costs, (ii) at-home recharge facilities, and (iii) a tax rebate of \$7000 (86.0%, 83.1%, and 82.3%, respectively) are the ones with highest weights amongst all eleven factors which contribute to the penetration of EVs in the market (Based on the survey between 1000 participants) [23]. In another research, probabilistic choice models estimate demand for electric cars. In this regard, separate utility functions for 51 subjects are made based on the survey data. Vehicle attributes such as price, operating cost, and range are preferred. The results showed that speed and range are the top value attributes among the consumers [8].

Since Plug-in Electrical Vehicles (PEVs) were newly introduced in the market, it is difficult to predict new adoptions of them. Although there is extensive emerged literature on exploring the most influential factors on PEVs' demand, given an early market in PEVs, the near future of their market is not clear (Carley et al. 2013). Several studies have been conducted on PEVs but lack important PEV buyers' information. PEV buyers are discouraged to purchase these types of vehicles because the electricity price is not the same as the gas price. For instance, a multiple regression analysis was conducted to assess potential factors that influence buyers in the counties of the state of California when buying PEVs. The results from this study illustrated that the factors that influence buyers to purchase PEV vehicles in the state of California are charging stations availability and retail gas price [21]. Moreover, based on two-analysis (Binary Factor Analysis and Principal Component Analysis) on consumer purchase motivations, interest in technology was a strong factor for consumers to purchase a hybrid vehicle. However, the overall results show that consumers consider financial benefits and affective aspects such as comfort, self-expression, manufacture's reputation, and style when purchasing a hybrid vehicle [27].

A conducted research to identify potential barriers for ordinary consumers to adopt plug-in electric cars concludes that an EV driving experience is not as good as driving

an internal combustion engine (ICE) vehicle. BEV users are also concerned about the sufficiency of stored energy for journey completion, and access to recharging points on the route. Furthermore, drivers are confident that EVs require further technological and infrastructural developments and significant price decreases. Most of them are unwilling to commit to purchasing EVs until sufficient advancements of EVs [13]. Therefore, based on the contribution of the previous studies, consumers prefer personal mobility needs over environmental benefits where eventually affect rapid technological and infrastructural developments. Moreover, socio-demographic characteristics, infrastructure variables, and vehicle-related factors seem to be the most important ones that influence the penetration and adaptation of EVs. Table 1 summarizes some of the most recent studies on this topic.

2.1 Financial and Economic Impacts of EVs

Electricity prices, battery costs, and market rules are the problems facing purchasing EVs. Despite the fact, some scholars find the financial and economic impacts of EVs are profitable, others investigate that V2G is not profitable to all users with current market revenue and policy rules. For example, they investigate the impacts of operating costs and market rules by developing a model of a centralized Vehicle to Grid (V2G) system and apply it to the 2015 wholesale electricity market in Texas (Houston Hub) for selling energy and capacity. They assess three potential market rule scenarios, characterized by different ways that EV owners are compensated. In the first scenario, electric vehicles are paid based on a fixed retail market price; in the second, they are paid a time-varying retail market price; in the third, the virtual power plant shares 50% of its total reward with the participating vehicles. Their results show that these vehicles with lower per-unit output-battery cost can lose more money because of extensive battery over-use and insufficient reward at current market prices. The authors suggest improving the rewards of EVs “policymakers could invest in reducing battery costs or could spur the introduction of newer market products such as a peak-shaving product or a renewable consuming/flexibility product” [6]. The contribution of their research would be an improvement on the efficiency of V2G systems in addressing climate change and make this system more economically sustainable.

2.2 Environmental Benefits and Energy Impacts of EVs

Another interesting topic in the field of EVs is their energy impacts them. Some scholars conduct a comparison of the life cycle of GHG emissions between coal-to-liquid (CTL) fuels, coal-to electricity for plug-in hybrid electric vehicles (PHEVs), coal-to-hydrogen for (FCVs), and petroleum-based fuels. Results demonstrate that PHEVs could reduce vehicle life cycle GHG emissions to about one-half when

Table 1 Overview of studies on electric vehicle adaptation behavior

| Study, country | Mode | EV types | Explanatory variables | Key findings | Limitations/future work |
|----------------|-----------|-----------|---|--|--|
| [21], U.S.A | MLRA | PEV, CV | Socioeconomic, demographic, and travel-related characteristics | It can be applied to evaluate changes in vehicle fleet composition and the levels of emissions in response to transportation policies | Further research is needed to integrate other indicators such as psychographics and their contribution to PEV adoption behavior |
| [20], U.S.A | AHP/MOVES | – | Air pollution, traffic congestion, investment cost, and natural environment | Its applications include ranking transportation strategies on mitigating CO ₂ emissions, prioritizing budgets | Future research is needed to integrate other GHGs and their contribution to both climate change and air pollution, as well as to explore multiple scenarios and integrate an appropriate cost analysis for evaluation of each strategy and their feasibility |
| [19], U.S.A | Lifecycle | PHEV, FCV | CTL fuels, coal-based electricity fuel, and coal-based hydrogen fuel | PHEVs could reduce vehicle life cycle GHG emissions by up to about one-half when coal is used to generate the electricity used by the vehicles | – |

(continued)

Table 1 (continued)

| Study, country | Mode | EV types | Explanatory variables | Key findings | Limitations/future work |
|---------------------|---|----------|---|---|---|
| [6], U.S., TX | Centralized (V2G) system | EVs | Market price, time-varying retail prices, and VPP shares 50% of its total reward with the participating EVs | These vehicles with lower per unit output—battery cost could lose more money because of extensive battery over-use and insufficient reward at current market prices | The DP-UC method used in this paper is applicable to any VPP-EV—market scenario that uses day ahead and/or real time prices. However, the results are only applicable to certain markets that have nodal prices for energy and ancillary services. Moreover, the assumptions impact the results |
| [27], UK | Binary factor analysis and principal component analysis | HEV | Financial benefits, interesting in technology, and affective aspects | The financial transport policies are an important factor in consumer hybrid purchase motivations, and social norms and consumers' willingness influence the purchase decision | This study is retrospective: the questionnaire was distributed to customers who had bought the Prius in 2007 and 2008. It is possible, therefore, that responses might have been different had the questions been posed at the time of purchase |
| Krupa et al. (2014) | JMPO Pro 10.0.0 | PHEV | Fuel costs saving, recharge facilities, and tax rebate | Financial and battery-related concerns remain major obstacles to widespread PHEV penetration | The data are readily available to other researchers and stakeholders who need to ground truth their assumptions and decisions with data |

(continued)

Table 1 (continued)

| Study, country | Mode | EV types | Explanatory variables | Key findings | Limitations/future work |
|----------------|--|-----------|---|---|--|
| [26], U.S.A | MATSim, with SAEV Modeling | SAEV | Charing station location, and charge time | Response time improvement through decreasing charge times/increasing vehicle range is its limitation | Future work will be analyzing the balance of charge times and fleet size with desired response times |
| [29] | – | PHEV, V2G | Socio-technical barriers | The literature review addresses “socio-technical” obstacles as important as V2G transition | Work to improve the technical performance of hardware must be coupled with attempts to overcome economic, behavioral, cultural, and infra-structural obstacles. These latter types of barriers do not fit neatly into the traditional R&D categories and remain deeply embedded in the social and institutional fabric |
| [2], U.S.A | Survey, Data Collection, Spreadsheet Model for Calculation | PHEV | Recharge patterns and electricity demand | Deferring all recharging to off-peak hours could eliminate all additions to daytime electricity demand from PHEVs, although less electricity is used, and less gasoline displaced | Radically changing travel behavior could invalidate the use of data on present-day real travel |

(continued)

Table 1 (continued)

| Study, country | Mode | EV types | Explanatory variables | Key findings | Limitations/future work |
|----------------|--------------------|----------------|---|---|---|
| [28], Canada | Nested Logit | HEV | Fuel price/availability, acceleration, pollution, taxes, fees, HOV ^a lane, socio-economic factors, and travel behavior | Incentives including parking fees and access to HOV lane do not affect preferences towards HEVs | Limited access and penetration to Internet are considered the main sources of bias in the collected samples |
| [16], Canada | Multinomial Probit | HEV, PHEV, BEV | Cost, performance, charging factors, warranty, vehicle attribute, cash, free parking, free toll road use, access to HOV ^a lane, and Socio-economic factors | PHEVs and BEVs are most attractive to households that are younger and highly educated. Those that care about fuel economy and reduced, or zero emissions show much higher probabilities of selecting HEV, PHEV, and BEV options | Future work should treat the different segments of the market as distinct as most relationships in our discrete choice model varied over vehicle types |
| [32], Germany | Multinomial Probit | HEV, PHEV, BEV | Purchase price, power, fuel costs, emissions, age, body type, horsepower, driving range, mileage, and vehicle to work | Younger and environmentally aware persons have positive intention towards HEVs, PHEVs, and BEVs | The practical obstacles for the broad application of flexible multinomial Probit models are when a high number of individual characteristics is included as explanatory variables |
| [24], U.S | Nested Logit | PHEV, BEV | Fuel-saving, range anxiety, willingness to pay for workplace, and public charging | Three types of recharge enhancement increase sales of PHEVs and BEVs | Future technology progress will have a larger impact if a better recharge infrastructure is in place |

(continued)

Table 1 (continued)

| Study, country | Mode | EV types | Explanatory variables | Key findings | Limitations/future work |
|----------------|--------------------|----------|---|---|--|
| [10], Germany | Multinomial Probit | HEV, BEV | Purchase price, fuel costs, power, CO2 emissions, and fuel availability | Increasing fuel availability could increase greater than threefold increase in HEV and BEV market penetration | A problem with experimental market shares of stated-preference studies, as opposed to observed shares in real markets of revealed-preference data, is that stated choices are a response to the experimental attribute variation. Thus, little can be said about the competitiveness of the different alternatives |

^a *HOV* High occupancy vehicle, *CTL* Coal-to-liquid, *HEV* Hybrid electric vehicle, *PHEV* Plug-in hybrid electric vehicle, *BEV* Battery electric vehicle, *V2G* Vehicle-to-grid, *SAEV* Autonomous electric vehicle, *PEV* PHEV or BEV, *CV* Conventional vehicle, *FCV* Fuel cell vehicle, *MLRA* Multiple logistic regression analysis

carbon is combined with coal to generate electricity for the vehicle [19]. In another research, authors estimate the electricity and gasoline use under three recharging scenarios (immediate and unconstrained, universal workplace access, and off-peak only) [2]. They conclude although PHEV vehicles can eliminate the use of gasoline relative to conventional vehicles, recharging them in peak hours can cause peak electricity demand. So, deferring all recharging to off-peak hours could eliminate all additions to daytime electricity demand from PHEVs, although less electricity is used, and less gasoline displaced [2].

PHEVs can be used as a transition technology to FCVs as fuel technology progressively develops. The transportation sector is considered as one of the largest sectors to produce CO₂ emissions and GHGs due to anthropogenic activities resulting in recent climate change. The U.S. Environmental Protection Agency (EPA) estimates that about 28% of total the U.S. GHG emissions came from the transportation sector (EPA 2008). Therefore, PHEVs are determined to help with the reduction of GHG emissions in the mainstream transportation sector [19]. They could reduce vehicles' life cycle GHG emissions to about one-half when carbon is combined with coal to generate electricity for the them. Annual CO₂ emissions in Texas are substantially higher than any other states in the U.S. Within the state of Texas, both large urban areas such as Dallas and smaller cities such as Lubbock have relatively higher per capita emissions, with large contributions from transportation sector [14]. Some other scholars evaluate and prioritize strategies that can reduce CO₂ emissions. The results after performing the analytical hierarchy process (AHP) method show that the reduction strategies have the highest score of 40% following 36% of avoiding strategies and 24% of replaced strategies [20]. This indicates that the reduction strategies have the highest impact on CO₂ mitigation and the lowest energy demand.

2.3 Social-Technical Factors Affecting EVs

Another crucial factor affecting the use of EVs are socio-technical ones. "Perhaps because they are often harder to identify, more difficult to overcome" [29]. Therefore, socio-technical barriers facing the use of PHEVs and V2G systems are identified and categorized as below:

- Battery technology and high costs compared to conventional,
- Impediments relating to customer acceptance due to high initial cost,
- The historical aversion to new technologies [29].

Hearty resistance from stakeholders to the existing infrastructure may be significant. Automobile manufacturers, oil companies, and repair businesses have sunk billions of dollars into supply and production infrastructure for conventional vehicles [29]. So, the authors concluded to improve the technical performance of hardware must be coupled with attempts to overcome economic, behavioral, cultural, and infrastructural obstacles [29]. Table 1 represents an overview of studies on electric

vehicle adaptation behavior. The next section provides methodology and a thorough description of all datasets, which are incorporated into this study.

The literature review identifies two main gaps in the existing disaggregate U.S. studies hybrid/electric vehicles adoption. First, none of these studies investigate potential factors that can be attributed to the penetration of PHEVs, EVs, and HEVs compared to CVs. Second, to our knowledge, none of these studies use MLR model to estimate the significant factors that impact the use of several types of Hybrid/Electric vehicles compared to conventional ones.

3 Methodology

This section focuses on the description of the development of the models for this study. Several multiple linear regression models were performed to determine significant factors that would impact the use of different types of hybrid vehicles compared to conventional vehicles. The types of hybrid vehicles assessed in this study are plug-in hybrids vehicles (PHEV), electric vehicles (EV), hybrid vehicles (HEV). The conventional vehicles (CV) evaluated in the study encompass only vehicles with fuel type as gas. The National Household Travel Survey website is used as the source to extract sample data of the US in relationship with hybrid and conventional vehicles. R Studio is the software used to conduct the statistical analysis for the sample data. The sample data consisted of quantifying the vehicle mile travel (VMT) with the relationship of each state and the factors estimated to impact the use of different types of vehicles. Some of the quantified factors include fuel expenditures between \$0 to \$1000 (FE1000), fuel expenditures between \$0 to \$1200 (FE1200), household income of \$50k and above (HouseIn50), household income of \$75k and above (HouseIn75), metropolitan statistical area (MSA), drivers with children (LifeCWC) or not (LifeCnC), vehicle model either Ford (Mford), Toyota (Mtoyota), Tesla (Mtesla), Nissan (Mleaf), travel day either a weekend or a weekday, and worker count either people work (Worker13) or not (Worker0). Since the types of vehicles assessed have different characteristics, there are different ranges for the factors impacting the use of hybrid and conventional vehicles. For fuel expenditures, FE1000 is the range used for the types of hybrid vehicles sample data and FE1200 for the CV vehicles sample data. For household income, HouseIn50 is used for PHEV, EV, and CV while HouseIn75 is used by the only HEV. Moreover, for vehicle models, Mford and Mtoyota are used for PHEV, HEV, and CV while Mtesla and Mleaf are used for EV (see Fig. 1). The sample data is limited to have values for every state in the US. Therefore, the sample data is adjusted concerning the values available for every state for each model resulting in different sample sizes. Four multiple linear regression models are performed and analyzed using the goodness of fit measure, r squared. Linear regression assumptions are identified by plotting the linear regression model. For every model four graphs are plotted: residual vs fitted, normal q-q, scale location, and residuals vs leverage graphs in R studio. These graphs contribute to the study significantly since it determines assumptions such as linearity and homoscedasticity.

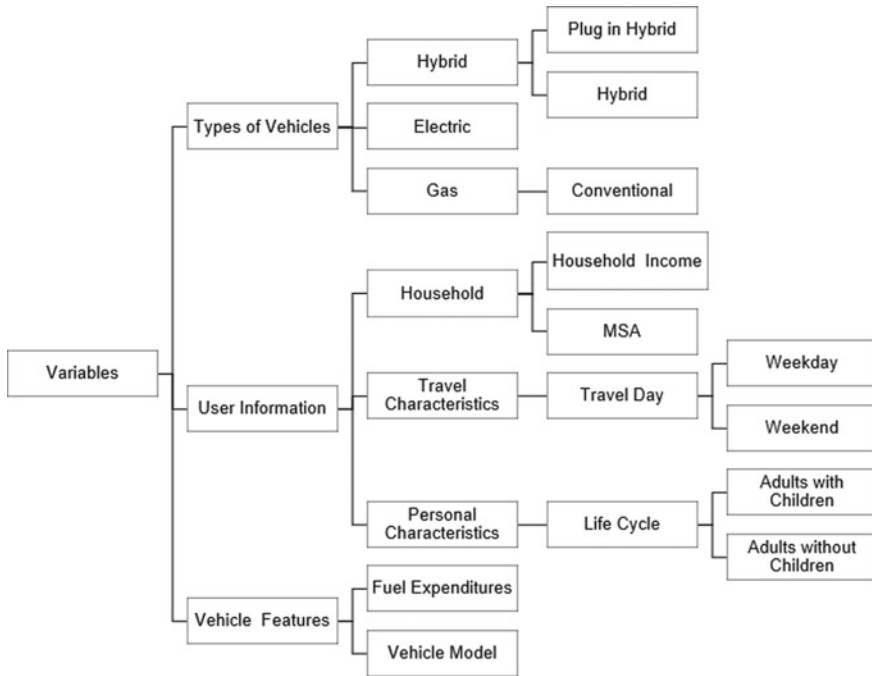


Fig. 1 Data distribution for the study

Also, *p*-values are assessed to determine if the values are statistically significant to the model. The following sections show the results from each model.

4 Sample Data

Four datasets are used to conduct this study. The first dataset consists of 474 observations for the PHEV vehicles model. The second dataset consists of 545 observations for the EV model. The third dataset consists of 4982 observations for the HEV models. The fourth dataset consists of 238, 320 observations for the CV. The values in each dataset are changed in terms of natural logs to have better outcomes in the model linear regression models (see Fig. 2). Figure 2 is only showing the states with relevant information to the study.

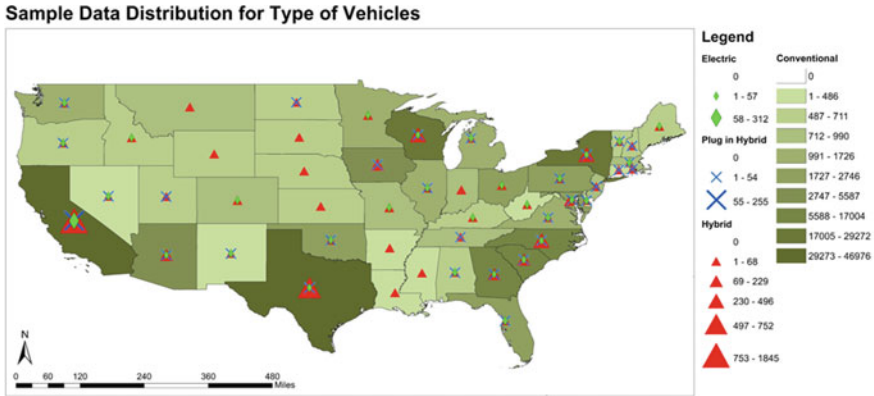


Fig. 2 Sample data distribution for type of vehicles

5 Model Results and Discussion

The results of the linear regression models for the hybrid and conventional vehicles are demonstrated in Table 2. The next following paragraphs describe the results for each one.

The regression output for PHEV shows that the FE1000 and MSA are significant which suggest that drivers that used that type of vehicle reside with urban areas and spend more than \$1000 dills on fuel. On the other hand, LifeCnoC, Mford, Mtoyota, Weekday, HouseIn75, LifeCWC, Worker13, Weekend, and Worker0 are not statistically significant because based on this study. For the independent variables

Table 2 Linear regression results for type of hybrid and conventional vehicles

| Explanatory variables | PHEV | | EV | | HEV | | CV | |
|---------------------------|--------|----------|--------|----------|-------|----------|-------|----------|
| | Coef | P-values | Coef | P-values | Coef | P-values | Coef | P-values |
| FE1000, F1200(CV) | -0.07 | 0.01 | 0.74 | 3.13e-05 | -0.18 | 2.94e-05 | 0.01 | 0.69 |
| HouseIn50, HouseIn75(HEV) | 0.33 | 0.03 | -0.031 | 0.62 | 0.08 | 0.03 | 0.00 | 0.40 |
| MSA | -0.03 | 0.40 | -0.26 | 0.11 | 0.043 | 0.42 | 0.06 | 0.03 |
| LifeCnoC | -0.02 | 0.43 | -0.00 | 0.99 | -0.01 | 0.43 | -0.00 | 0.96 |
| LifeCwC | 0.03 | 0.19 | -0.06 | 0.270 | -0.00 | 0.73 | -0.00 | 0.85 |
| Mford, Mtesla(EV) | 0.02 | 0.48 | -0.04 | 0.29 | -0.01 | 0.02 | -0.00 | 0.00 |
| Mtoyota, Mleaf(EV) | -0.03 | 0.21 | -0.27 | 0.47 | -0.01 | 0.87 | -0.02 | 0.00 |
| Weekday | 0.02 | 0.7 | 0.02 | 0.71 | 0.14 | 0.00 | 0.57 | 0.00 |
| Weekend | 0.05 | 0.05 | 0.01 | 0.76 | 0.04 | 0.00 | 0.25 | 0.00 |
| Worker0 | -0.006 | 0.78 | 0.03 | 0.5 | -0.02 | 0.07 | 0.04 | 0.00 |
| Worker13 | 0.04 | 0.38 | 0.29 | 0.04 | 0.23 | 1.92e-05 | 0.09 | 0.00 |
| R-Squared | 0.59 | | 0.87 | | 0.81 | | 0.99 | |

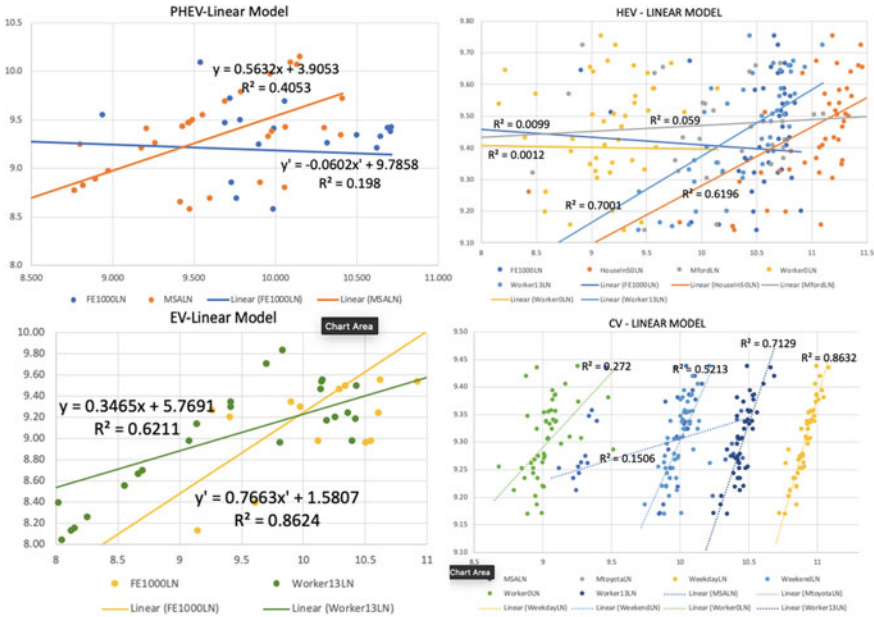


Fig. 3 Linear model for PHEV, HEV, EV, and CV

with p -values of less than 0.05, one can conclude that sample data provide enough evidence to reject the null hypothesis for the entire population. Changes in the independent variables are associated with changes in the responses at the population level. These variables are statistically significant and probably a worthwhile addition to the regression model. From the results above, we would consider removing the variables with p -values of more than 0.05, because keeping them in the model can reduce the model’s precision (see Fig. 3).

The regression model for HEV shows Fuel Expenditures, Household Income, Ford Model, Zero Worker Count, and Worker Count as statistically significant suggesting that the household income plays an important role for the users that go to work and not have a job. Since the adjusted R-square is not in between 0.90 and 1, the model for hybrid vehicles is only acceptable. Therefore, further investigation is required to be conducted to assess the variables that truly impact the model (see Fig. 3).

The regression output for EV shows that fuel expenditures and worker count 1–3 per household shows that drivers that go to work tend to use more EV. On the other hand, the rest of the independent variables are not statistically significant because their p -values greater than the usual significance level of 0.05. For the independent variables with p -values of less than 0.05, one can conclude that sample data provided enough evidence to reject the null hypothesis for the entire population. Changes in the independent variables are associated with changes in the responses at the population level. These variables are statistically significant and probably a worthwhile addition

to the regression model. As mentioned before, the variables with p -values of more than 0.05 are removed from the regression model (see Fig. 3).

Once the regression model for CV sample data is conducted, then it is evaluated by the goodness of fit measure. The adjusted R-squared is determined to be 0.99 that concludes to have the best fit of the model. The factors that impact the model are the following: MSA, vehicle models for Ford and Toyota, travel days for the weekdays and weekends, and worker count. These values are determined to be statistically significant with p -values below 0.05 (see Fig. 3).

6 Conclusion

The objective of this study is to evaluate the parameters which have impacts on the adaptation and penetration of hybrid/electric vehicles. One of the drawbacks of conventional vehicles is the high assumption of fossil fuel which has some negative impacts on the environment (emission of GHGs), society (health issues), and economics (energy demand). Changing the transportation mode with a sustainable fuel source such as electricity can reduce the usage of gasoline. So, the conducted study aimed to seek affecting attributes that contribute to the penetration of hybrid/electric vehicles. The types of hybrid vehicles assessed are PHEV, EV, HEV in the U.S. In this regard, the multiple linear regression model is applied as the methodology to determine the significant factors that impact the use of two types of hybrid vehicles compared to CV. The National Household Travel Survey website is used to gather relevant data and R Studio is the software for doing the statistical analysis of sample data. The VMT is considered as the dependent variable in Linear Regression Model and other factors such as fuel expenditures between \$0 to \$1000, fuel expenditures between \$0 to \$1200 household income of \$50k and above, household income of \$75 k and above, metropolitan statistical area (MSA), life cycle, vehicle model, travel day, and worker count are applied as independent variables (explanatory variables). Four linear regression models are performed, and the results show that the variables that have statistical significance with the type of Hybrid vehicles are Fuel expenditures and household income. The hybrid vehicles model presents more significant factors such as Ford models, vehicles driven on the weekdays and weekends, and 1–3 workers per household. The factors that impact the use of conventional vehicles compared to hybrid vehicles are MSA, Model Toyota vehicles, vehicles driven on the weekdays, weekends, zero workers, and one to three workers per household. Furthermore, people that have plug-in vehicles, electric vehicles, and hybrid vehicles tend to have more fuel expenditures and higher household income than conventional vehicles. Therefore, it is determined that “drivers with and without children” are not significant among the models.

References

1. Arce GLAF, Carvalho Jr JA, Nascimento LFC (2014) A time series sequestration and storage model of atmospheric carbon dioxide. *Ecol Model* 272:59–67
2. Axsen J, Kurani KS (2010) Anticipating plug-in hybrid vehicle energy impacts in California: constructing consumer-informed recharge profiles. *Transp Res Part D: Transp Environ* 15(4):212–219
3. Andryscio N, Gurney KR, Beneš B, Corbin K (2008) Visual exploration of the vulcan CO₂ data. *IEEE Computer Graph Appl* 29(1):6–11
4. Bathaei B (2018) Achieving sustainable city by the concept of Persian Garden. In: international conference—architecture technology and the city workshop questions, Romania, 2018. *Acta Technica Napocensis, Civil Engineering & Architecture, Special Issue*. <http://constructii.uctluj.ro/ActaCivilEng>
5. Bathaei B (2016) Change is of the essence, regenerating of brown fields (Landscape Revitalization of Tehran's Brick Kilns). In: 2nd international conference on architecture, structure and civil engineering (ICASCE'16), UK, London
6. Bhandari V, Sun K, Homans F (2018) The profitability of vehicle to grid for system participants—a case study from the Electricity Reliability Council of Texas. *Energy* 153:278–286
7. Bradley TH, Quinn CW (2010) Analysis of plug-in hybrid electric vehicle utility factors. *J Power Sources* 195(16):5399–5408
8. Calfee JE (1985) Estimating the demand for electric automobiles using fully disaggregated probabilistic choice analysis. *Transport Res Part B: Methodol* 19(4):287–301
9. Davis SC, Williams SE, Boundy RG (2018) Transportation energy data book, Edition 36.2 of ORNL-5198. Oak Ridge National Laboratory, Oak Ridge, TN, USA (2018)
10. Daziano RA, Achtnicht M (2014) Forecasting adoption of ultra-low-emission vehicles using Bayes estimates of a multinomial probit model and the GHK simulator. *Transp Sci* 48(4):671–683
11. EPRI, NRDC (2007) Environmental assessment of plug-in hybrid electric vehicles. Volume 1: Nationwide greenhouse gas emissions. Electric Power Research Institute (EPRI), Final Report
12. Ghommem M, Hajj MR, Puri IK (2012) Influence of natural and anthropogenic carbon dioxide sequestration on global warming. *Ecol Model* 235:1–7
13. Graham-Rowe E, Gardner B, Abraham C, Skippon S, Dittmar H, Hutchins R, Stannard J (2012) Mainstream consumers driving plug-in battery-electric and plug-in hybrid electric cars: a qualitative analysis of responses and evaluations. *Transport Res Part A: Policy Practice* 46(1):140–153
14. Gurney KR, Mendoza DL, Zhou Y, Fischer ML, Miller CC, Geethakumar S, de la Rue du Can S (2009) High resolution fossil fuel combustion CO₂ emission fluxes for the United States. *Environ Sci Technol* 43(14):5535–5541 (2009)
15. Hidrue MK, Parsons GR, Kempton W, Gardner MP (2011) Willingness to pay for electric vehicles and their attributes. *Resour Energy Econ* 33(3):686–705
16. Higgins CD, Mohamed M, Ferguson MR (2017) Size matters: How vehicle body type affects consumer preferences for electric vehicles. *Transport Res Part A Policy Practice* 100:182–201
17. Houk J (2010) Greenhouse gas emissions analysis of regional transportation plans with EPA's MOVES Model. In: Proceedings of 19th annual international emission inventory conference. Emissions inventories-informing emerging issues
18. Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990–2006 (2008) In the Federal Register/FIND. *Federal Inform. News Dispatch* 73(46):12413
19. Jaramillo P, Samaras C, Wakeley H, Meisterling K (2009) Greenhouse gas implications of using coal for transportation: life cycle assessment of coal-to-liquids, plug-in hybrids, and hydrogen pathways. *Energy Policy* 37(7):2689–2695
20. Javid RJ, Nejat A, Hayhoe K (2014) Selection of CO₂ mitigation strategies for road transportation in the United States using a multi-criteria approach. *Renew Sustain Energy Rev* 38:960–972

21. Javid RJ, Nejat A (2017) A comprehensive model of regional electric vehicle adoption and penetration. *Transp Policy* 54:30–42
22. Khatiwala S, Toste Tanhua S, Fletcher M, Gerber M, Doney SC, Graven HD, Gruber N et al (2013) Global ocean storage of anthropogenic carbon. *Biogeosciences* 10(4):2169–2191
23. Krupa JS, Rizzo DM, Eppstein MJ, Brad Lanute D, Gaalema DE, Lakkaraju K, Warrender CE (2014) Analysis of a consumer survey on plug-in hybrid electric vehicles. *Transport Res Part A: Policy Practice* 64:14–31
24. Lin Z, Greene DL (2011) Promoting the market for plug-in hybrid and battery electric vehicles: role of recharge availability. *Transp Res Rec* 2252(1):49–56
25. Liu Y, Tremblay J-M, Cirillo C (2014) An integrated model for discrete and continuous decisions with application to vehicle ownership, type and usage choices. *Transport Res Part A Policy Practice* 69:315–328
26. Loeb B, Kockelman KM, Liu J (2018) Shared autonomous electric vehicle (SAEV) operations across the Austin, Texas network with charging infrastructure decisions. *Transport Res Part C Emerg Technol* 89:222–233
27. Ozaki R, Sevastyanova K (2011) Going hybrid: an analysis of consumer purchase motivations. *Energy Policy* 39(5):2217–2227
28. Potoglou D, Kanaroglou PS (2007) Household demand and willingness to pay for clean vehicles. *Transp Res Part D: Transp Environ* 12(4):264–274
29. Sovacool BK, Hirsh RF (2009) Beyond batteries: An examination of the benefits and barriers to plug-in hybrid electric vehicles (PHEVs) and a vehicle-to-grid (V2G) transition. *Energy Policy* 37(3):1095–1103
30. U.S. Department of Energy (US DOE), Energy Information Administration (2021) Use of energy in the United States explained: How the United States uses energy. Accessed 04 Jan 2021. <https://www.eia.gov/energyexplained/use-of-energy/>
31. U.S. Environmental Protection Agency (US EPA) (2021) Fast facts: U.S. transportation sector greenhouse gas emissions 1990–2016. Accessed 04 Jan 2021. <https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=P100USI5.pdf>. EPA-420-F-18-013
32. Ziegler A (2012) Individual characteristics and stated preferences for alternative energy sources and propulsion technologies in vehicles: a discrete choice analysis for Germany. *Transport Res Part A: Policy Practice* 46(8):1372–1385

Travel Time Reliability with the Presence of Connected and Autonomous Vehicles



F. Hosseinzadeh, M. Ansari Esfeh, and L. Kattan

1 Introduction

1.1 Background and Motivation

The transportation system is designed to serve its users in normal conditions. Extreme events such as hurricanes, heavy rains, tornados, and heavy snowfalls cause system performance degradation. The evaluation of the transportation system performance under perturbation is critical in order to respond quickly to network disruptions. Such responses include issuing evacuation warnings before disruption occurrence, rescue operations, and recovery activities during and after perturbation [1]. Reliability, vulnerability, and resilience are various concepts that describe network performance. The reliability of a transportation system is the probability of a network to remain satisfactory under disruption. With their seamless connectivity potentials, CAVs and their enabling infrastructure create a connected cyber physical network capable of providing timely and accurate travel information at the network level; thereby improving travel time reliability (TTR) [26, 30, 31]. While Wang et al. [30, 31] and Mittal et al. [26] studied travel time reliability in the presence of CVs, their work focuses on the cases with 100% market penetration rates of CVs. This paper takes one step in this direction and examines the road network TTR under various CAV penetration levels to evaluate the system performance at different disruption phases. Two methods are used to evaluate the TTR value. First, the Monte-Carlo reliability method is utilized to generate random travel times based on a suggested mean and variance. Probability density function (PDF) and exceedance probability curves are then extracted for various CAV penetration scenarios and various disruption phases. The failure probability is obtained from the value of exceedance probability

F. Hosseinzadeh (✉) · M. Ansari Esfeh · L. Kattan
University of Calgary, Calgary, Canada
e-mail: fatemeh.hosseinzadeh@ucalgary.ca

© Canadian Society for Civil Engineering 2022
S. Walbridge et al. (eds.), *Proceedings of the Canadian Society of Civil Engineering Annual Conference 2021*, Lecture Notes in Civil Engineering 250,
https://doi.org/10.1007/978-981-19-1065-4_20

239

curves at the travel time threshold. Second, the first-order reliability method (FORM) is used to compute failure probability. Travel time reliability is calculated by using failure probability as will be explained in Sect. 2. The paper aims to conduct travel time reliability analyses for a road network subject to disruptions and under different market penetration rates of CAVs. The analysis was performed on the Nguyen-Dupius network that its resilience was evaluated by [2].

1.2 Review of the Previous Studies

Several studies evaluated TTR to evaluate transportation system performance. Asakura and Kashiwadani [5] proposed the concept of TTR considering the travel demand of road networks. Iida [19] defined TTR as the probability of travelers successfully reaching their destination within a specified time. Liu et al. [22] studied the TTR of stochastic networks under a user equilibrium (UE) assumption. Wang et al. [29] proposed a bi-objective model for TTR analysis based on the UE and punishing UE models. In addition, Pei and Gai [27] used the Monte Carlo simulation method to analyze the TTR of a road network. Other evaluation indices were also proposed to further analyze the TTR [7, 10, 11, 28]. Recently, Jose and Ram [20] reviewed the existing researches on network TTR, user based TTR, variability, and travel behaviour with particular reference to airport access.

CAVs are expected to positively affect transportation system performance. Through their vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication capabilities, CAVs can collect and dissipate vital traffic information in real time. Yang et al. [32] reported that the average travel time, total travel time, and average delay of vehicles in a mixed traffic environment were lower than those in a traditional traffic environment. Bento et al. [9] and Genders and Razavi [14] found that CVs can improve safety, mobility and reduce the environmental impact in work zones. Also, Goodall et al. [15] showed that using CVs improves travel time prediction and traffic management. Arnaout and Bowling [4] proposed an agent-based microsimulation to examine the impact of vehicle communication on traffic performance of a freeway with an on-ramp to induce a disturbance in the traffic network. Also, Kattan et al. [21] developed a virtual V2V communication platform for information dissemination in case of incidents. The results show overall improved network performance and decrease in average travel time.

Hannoun [17] and Bahaaldin et al. [6] studied the CAVs' impacts on network performance during natural disasters such as earthquakes and flooding. Hannoun [17] developed a framework based on time-dependent hyper-star algorithm to optimally route vehicles in a flooding scenario. Other studies estimated the level of service (LOS) in a mixed-traffic environment with different penetration rates of CAVs [13, 30, 31, 34].

More recently, Ahmed et al. [2] developed a model of surface transportation system resilience in a mixed-traffic environment consisting of human-driven vehicles and CAVs. Their model considered the sensitivity of the network resilience at varying

CAV penetration rates, namely 0, 25, 50, 75, and 100% CAV. The authors presented different formulations for estimating disruption impact factor (DIF), capacity of links, link flow, link travel time and resilience parameter. The authors considered five phases: a preparation phase (P1), immediately before disruption impact phase (P2), a disruption impact phase (P3), immediately after disruption impact phase (P4), and a recovery phase (P5) to analyze the performance of transportation system in smaller timescales.

1.3 Definition

Various representative evaluation concepts exist to explain network performance, namely vulnerability, resilience, and reliability. They are shown schematically in Fig. 1 and are defined below:

Vulnerability: vulnerability is defined as the susceptibility of a network to perturbations. The amount of reduction of functionality under disruption evaluates the vulnerability. Husdal [18] defined vulnerability as the propensity of a transportation network to become inoperable under disruption, which is an opposite concept to reliability. Generally, Gu et al. [16] classified the vulnerability as accessibility vulnerability, connectivity vulnerability, and capacity vulnerability.

Resilience: resilience is defined as the ability of a network to resist, adapt to, and recover from perturbations. In general, two components, the ability to resist the effects of a perturbation, and the ability to recover from a perturbation are inherently accounted in the concept of resilience. Resilience is a comprehensive evaluation concept used in evaluating system performance under perturbations [16].

Reliability: reliability is defined as the probability of network performance to remain satisfactory under perturbation [16]. The probability that the travel time on a given path is shorter than a given threshold can be estimated by applying the limit state

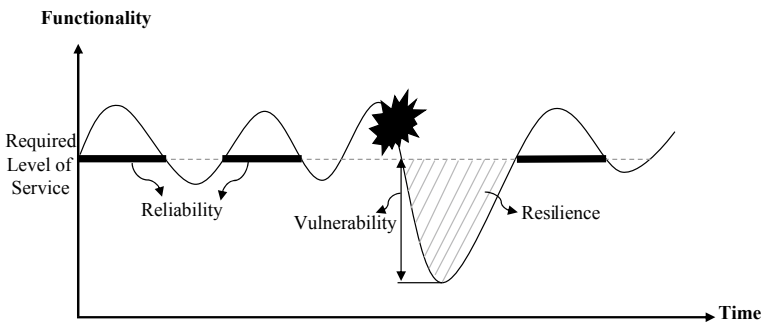


Fig. 1 Schematic view of evaluation concepts including vulnerability, resilience, and reliability [16]

Table 1 Performance indicators used in this paper

| Performance indicator | Abbreviation | Definition |
|--|--------------|---|
| Complementary cumulative distribution function | CCDF | The probability that travel time will be more than t for the time domain. It is also called the exceedance probabilities. it calculates from 1-CDF |
| Travel time reliability | TTR | The probability that the travel time on a given path is shorter than a given threshold. It is the CDF's value at threshold |
| Failure probability | P_f | The probability that the travel time on a given path is greater than a given threshold. It is the value of CCDF or exceedance probabilities for a travel time threshold |

function (LSF) [24]. LSF returns a negative value under system failure conditions and a positive value when the system is satisfied.

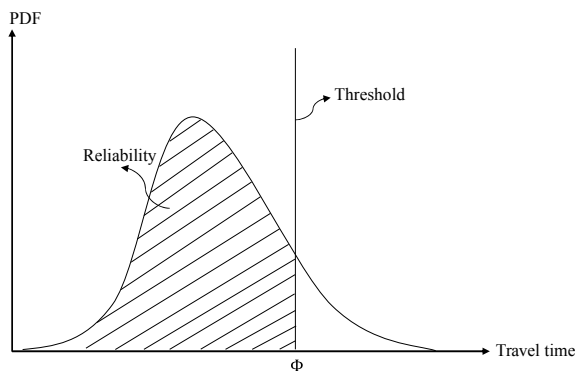
Table 1 shows Definitions of the performance indicators used in this paper.

2 Methodology

Unpredictable perturbations in a given transportation network lead to uncertainties in both travel demand and supply, which in turn influence the network reliability [16]. With the increasing demand for travel, transportation system reliability becomes an integral part of the planning, management, and operation of transportation networks. Transportation system reliability has three aspects: connectivity reliability, capacity reliability, and travel time reliability.

TTR is defined as the probability of a traveler successfully completing a trip and reaching a given destination within a specified time [8, 30, 31]. As shown in Fig. 2, the travel time reliability is obtained from Eq. 1 where t_i is the total travel time for

Fig. 2 Definition of travel time reliability



i th path and Φ is the travel time threshold related to the desired service level [30, 31]. TTR can be alternatively calculated based on the probability of failure using Eq. 2. Failure in Eq. 2 refers to the event that the trip in path i cannot be completed within a desired time Φ .

$$TTR(t) = Pr\{t_i \leq \Phi\} \tag{1}$$

$$TTR(t) = 1 - Pr\{failure\} = 1 - Pr\{(\Phi - t_i) \leq 0\} \tag{2}$$

Various methods can be used to calculate travel time reliability. As shown in Fig. 3, two methods are used, the Monte Carlo simulation and first-order reliability method (FORM).

Monte Carlo simulation [25] is a sampling method that generates random numbers based on a suggested mean and variance. Monte Carlo simulation can generate travel time distribution, which makes it possible to calculate the probability of failure and the probability that travel time between a given origin and destination remains satisfactory. Despite its simplicity, sampling methods such as Monte Carlo can be time-consuming because the generation of random numbers and the evaluation of an equation could be processor-intensive procedures.

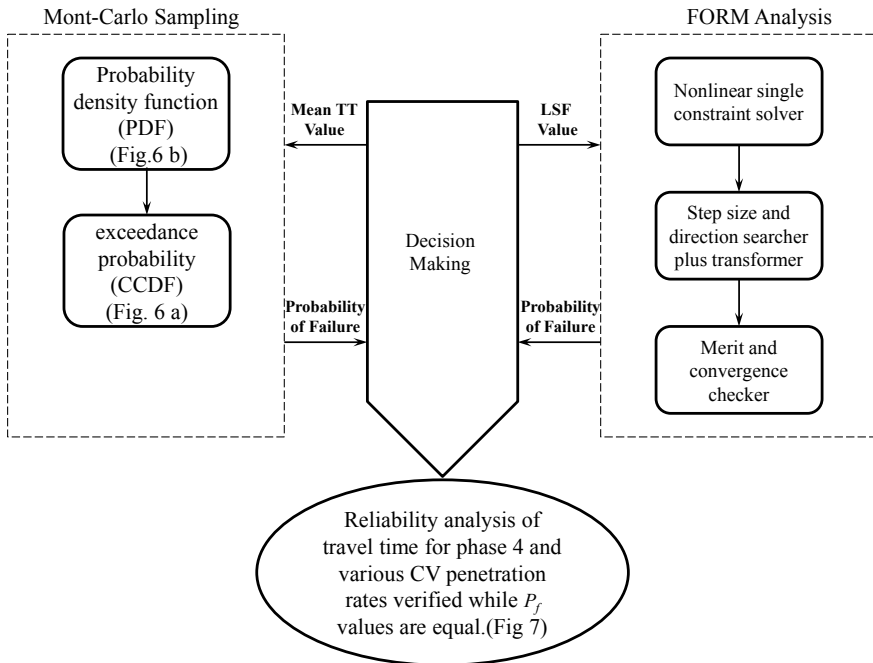


Fig. 3 The flowchart to evaluate travel time reliability for various CAV penetration rates immediately after disruption

Thus, in addition to Monte Carlo simulation, FORM is also used to evaluate travel time reliability as one of the most reliable computational methods [12]. In this method, first order Taylor series approximation of performance function and derivative-based iteration search algorithm is used at the design point [33]. Risk tools (Rt) [23], a computer program to carry out reliability and optimization analysis with interconnected probabilistic models, is used in this paper. RT implements a FORM analyzer with different subtotal analyses including nonlinear single constraint solver, step size and direction searcher, transformer, and convergence checker. The detail of FORM analysis, such as step size and convergence checker, for the case study will be presented in the following section.

2.1 The Study Case Network

This paper estimates and compares the travel time reliability of a network under disruption, with consideration of different CAV penetration rates (0, 25, 50, 75, and 100%). The analysis is conducted on the Nguyen-Dupuis network depicted in Fig. 4. The analysis mainly focuses on the travel time between node 10 (origin) and node 8 (destination) at various disruption phases. The network data are obtained from [1, 2] with the assumption of a 25 mph speed limit and one lane in each direction for collectors and 45 mph speed limit and two lanes in each direction for arterials. To model CAVs, Ahmed et al. [2] considered their effect on headway, capacity and travel time.

The saturation headway of mixed traffic (i.e. the constant headway achieved once a stable moving queue is established in mix traffic environment) was measured using the following equation:

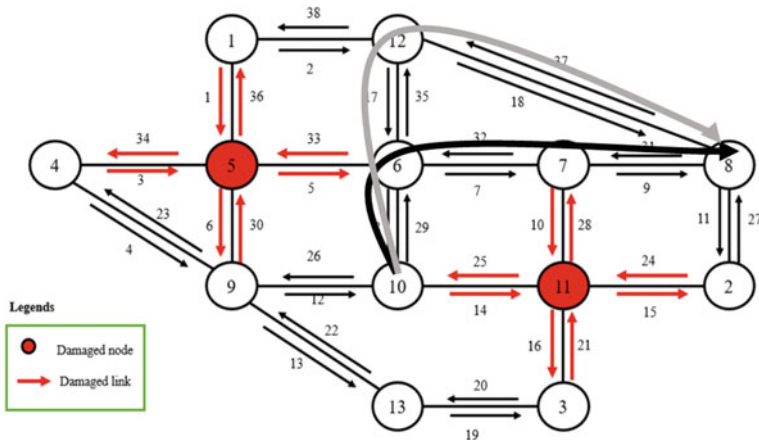


Fig. 4 Damaged nodes and links of the network in phase P4 [2]

Table 2 The value of headways for various CAV penetration rates

| CAV% | \bar{h}_{AA} (s) | \bar{h}_{AM} (s) | \bar{h}_{MA} (s) | \bar{h}_{MM} (s) |
|---------|--------------------|--------------------|--------------------|--------------------|
| 0–25% | 0.45 | 1.8 | 1.8 | 1.8 |
| 26–50% | 0.40 | 1.6 | 1.8 | 1.8 |
| 51–75% | 0.35 | 1.4 | 1.8 | 1.8 |
| 76–100% | 0.30 | 1.2 | 1.8 | 1.8 |

$$\bar{h}_{mix}(p) = p^2\bar{h}_{AA} + p(1-p)(\bar{h}_{AM} + \bar{h}_{MA}) + (1-p)^2\bar{h}_{MM} \quad (3)$$

where, p is the CAV penetration rate, \bar{h}_{AA} is the average headway for a CAV following a CAV, \bar{h}_{AM} is the average headway for a CAV following a non CAV, \bar{h}_{MA} is the average headway for a non-CAV following a CAV and \bar{h}_{MM} is the average headway for a non-CAV following a non-CAV which are different for various CAV penetration scenarios. Table 2 summarizes the value of these headways [1].

The reduction of headway in a mixed traffic increases the roadway capacity and decreases link travel time:

$$\begin{aligned} \overline{TT}_{l,mix}^{t_i} &= \left(1 + \frac{f_l^{t_i}}{C_l^{t_i}}\right) \times FTT_l^{t_i} \\ &= \left(1 + \frac{\sum_{ln} \delta_{ln} \times p_{ln,l}^{t_i} \times f_{ln,l}^{peak}}{N_{ln,l} \times \delta_{ln} \times \left(1 - \frac{6}{100} \times N_{ln,l}\right) \times f_{ln-w} \times \left(\frac{3600}{\bar{h}_{mix}}\right)}\right) \times FTT_l^{t_i} \end{aligned} \quad (4)$$

where $FTT_l^{t_i}$ is the free flow travel time on link l during time t_i , $N_{ln,l}$ is the number of lanes in link l , $C_l^{t_i}$ is the capacity of link l , $\bar{C}_{ln,l}^{t_i}$ is the average capacity per lane in link l during time t_i , f_{ln-w} is the capacity adjustment factor for lane width, $f_{ln,l}^{t_i}$ is the flow of lane ln in link l during time t_i , $f_{ln,l}^{peak}$ is the peak flow of lane ln in link l , $p_{ln,l}^{t_i}$ is the hourly flow rate as a percentage of peak flow, $f_l^{t_i}$ is the flow of link l during time t_i , δ_{ln} is a factor that represents lane's functionality as affected by disruption, and δ_{ln} is 1 if a lane is still operational, otherwise 0 [2].

As a case study, Nguyen-Dupuis network shown in Fig. 4 is used for the analysis. It is assumed that the acceptable level of service requires travel times to be less than 35 min for origin–destination (OD) pair 10 to 8. Based on the definition of reliability in Eqs. 1 and 2, travel time reliability $TTTR(P_i)$ and failure probability $P_f(P_i)$ for i th path are obtained from Eqs. 5 and 6 where P_i shows i th path, $TTT(P_i)$ is the total travel time for i th path, and LSF is the limit state function which is the difference between threshold time and total travel time:

$$TTTR(P_i) = Pr\{TTT(P_i) \leq 35\} = 1 - P_f(P_i) \quad (5)$$

$$P_f(P_i) = Pr\{(35 - TTT(P_i)) \leq 0\}, LSF = 35 - TTT(P_i) \quad (6)$$

As in Ahmed et al. [2], 5 phases of network disruption are considered. For example, for a tornado disruption, P1 refers to before nothing happens phase, P2 represents the raining situation before tornados, P3 is the time a tornado takes place, P4 is the time immediately after the tornado occurs, and P5 is when the network is completely recovered. Disruption makes some routes unavailable due to the failed nodes/links and increases the travel time for other paths. The nodes and links that are disrupted are shown in red in Fig. 4.

Table 3 shows the mean of travel time data for each link between nodes 10 and 8 presented for various CAV penetration rates. This data is collected from the network resilience analysis under disruption, during different disaster phases, for various CAV penetration rates [1].

Consistent with the assumptions in [1], we assumed the traffic demand reduction around 10% to 30%, speed reduction by 3–4 mph, and capacity reduction by 15–30% during adverse weather conditions. It is assumed that the traffic average headway increased by 3% in P2 and by 10% in P4. In recovery phase P5, it is assumed that traffic will get back to typical headway. For traffic flow, the availability of alternate routes and the route preference were analyzed considering the impact of the failed nodes and links. A reduction in traffic demand as of 15% was considered to comprehend the travelers' disinterest in making trips, harsh weather conditions, congestion, and partial small-scale damage on the links.

To evaluate the TTR value, first, the Monte-Carlo reliability method is utilized to generate random travel times based on mean travel time of Table 3 and coefficient of variation (COV) which assumed 0.2. The PDF can be extracted for generated random travel times for various CAV penetration scenarios and various disruption phases. The exceedance probability curves are then obtained using the cumulative distribution function (CDF) for each time based on the definitions in Table 1. The value of exceedance probability curves for travel time threshold gives the failure probability for various CAV penetration scenarios and various disruption phases. The travel time reliability can be calculated from Eq. 5, which is the difference between 1 and failure probability.

As a separate exercise, FORM analysis is followed to compute failure probability. FORM analysis is carried out to investigate the probability that the OD path's travel time exceeds the threshold which is 35 min. At first, the LSF, which is defined in Eq. 6, is introduced to the software. Then, the distribution of variables is defined by its mean value and COV. For travel time, a lognormal distribution with a mean travel time of Table 3 and COV of 0.2 is defined for each CAV penetration rate and at various disruption phases. The first order Taylor series approximation of LSF and iteration search algorithm is used to find failure probability. The maximum number of iterations is considered 100. The Armijo step size with an initial step size of 1 is used [3]. The result converges when the difference between the two iteration results is less than 0.001. Since the LSF is linear, the failure probabilities should be close to the sampling analysis.

Table 3 Mean values of the travel time (min) in Nguyen-Dupuis network

| Link | CAV 0% | | | | CAV 25% | | | | CAV 50% | | | | CAV 75% | | | | CAV 100% | | | |
|------|--------|------|------|------|---------|------|------|------|---------|------|------|------|---------|------|------|------|----------|------|-----|------|
| | (P1) | P2 | P3 | P4) | (P1) | P2 | P3 | P4) | (P1) | P2 | P3 | P4) | (P1) | P2 | P3 | P4) | (P1) | P2 | P3 | P4) |
| (7) | 9.2 | 10 | 9.3 | 8.6 | 9 | 9.8 | 9 | 8.4 | 8.3 | 9 | 8.3 | 7.8 | 7.1 | 7.8 | 7.1 | 6.8 | 5.7 | 6.3 | 5.7 | 5.6 |
| (9) | 6.4 | 7 | 9.3 | 6.1 | 6.3 | 6.9 | 9 | 6.1 | 6.1 | 6.7 | 8.3 | 5.9 | 5.7 | 6.3 | 7.1 | 5.6 | 5.2 | 5.8 | 5.7 | 5.2 |
| (29) | 13.4 | 14.9 | 31.2 | 13.3 | 13.4 | 14.8 | 30.3 | 13.3 | 13.3 | 14.8 | 27.2 | 13.2 | 13.2 | 14.7 | 22.2 | 13.2 | 13.1 | 14.6 | 16 | 13.1 |

Source Ahmed [1]

This approach is summarized in the flowchart of Fig. 3 for various CAV penetration rates and Phase P4. As is explained before, for other phases, we only used FORM method because of its low computation load.

3 Results and Discussion

10^6 probabilistic samples are generated using the Monte Carlo sampling method. Figure 5 shows the plot of the exceedance probability or the probability of failure with respect to travel time for different disruption phases, i.e., from pre-disruption phase, P1, to post-disruption phase, P5. Figure 5a, b show the results for the CAV penetration rate of 25% and 100% respectively. As shown in Fig. 5a, b, in phases P2 and P4, the performance of the path travel time is negatively affected by the disruptive events. However, with the CAV penetration of 100% the exceedance probabilities of total travel time, especially in disruption phase P4 significantly decrease.

Phase P4, which starts immediately after the disruption is more critical than P1, P2, and P5, which are phases before the disruption or after the recovery. So, sampling reliability analysis is reported for only this critical phase in Fig. 6.

In Fig. 6, two different diagrams in sampling reliability analysis for disruption phase 4 are illustrated alongside the LOS limit or 35 min threshold, namely, the exceedance probability and the resulting probability density function of the total travel time under various scenarios for CAV penetration rates, i.e., 0, 25, 50, 75, and 100%. Figure 6 shows that increasing the CAV penetration rate reduces both the total travel time and its variance. The failure probability of the total travel time from the threshold of 35 min is obtained from the exceedance probability diagram.

FORM analysis is performed to compare failure probabilities from this method with the failure probabilities resulting from sampling analysis. The LSF in FORM is the same as the LSF in sampling analysis which is shown in Eq. 6. Figure 7 shows the failure probabilities resulting from sampling and FORM analyses for different CAV penetration scenarios during the critical disruption phase P4. As shown, the

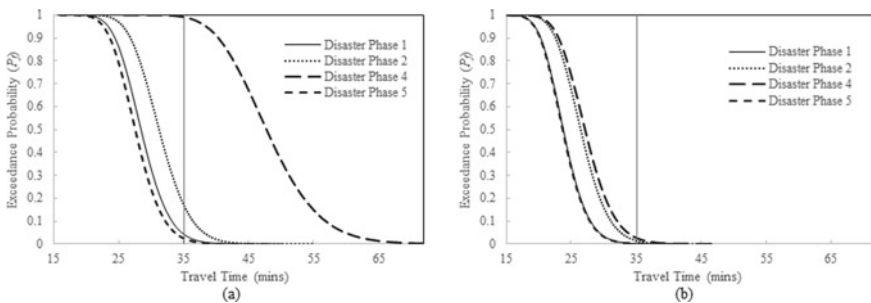


Fig. 5 Exceedance probability of travel time for OD pair 10 to 8 for different disruption phases with CAV penetration of **a** 25% and **b** 100%

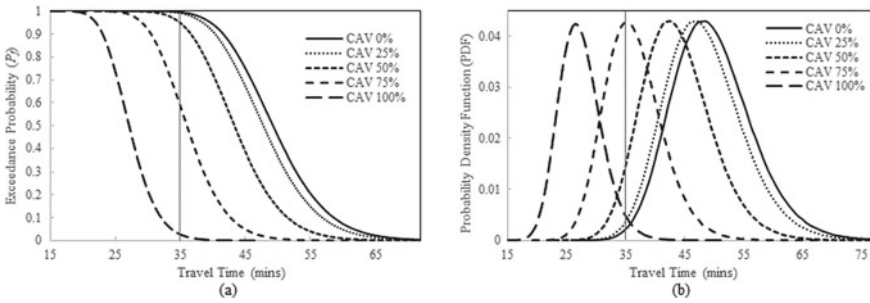


Fig. 6 a Exceedance probability and b probability density function for the OD 10 to 8 path travel time in disruption phase P4 with different CAV penetration alongside the LOS limit

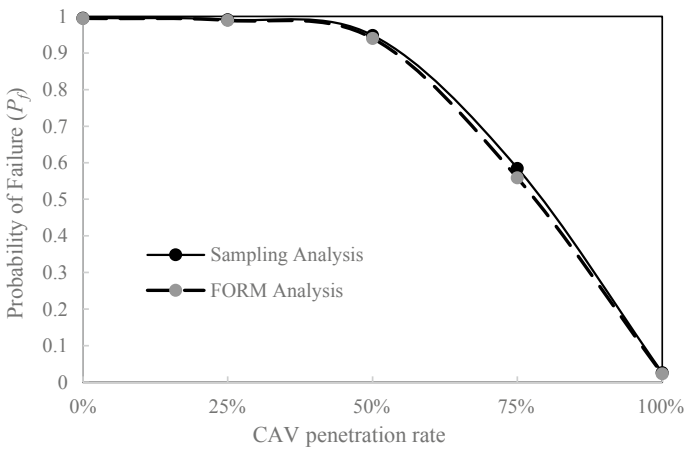


Fig. 7 Outputs of different reliability methods in disruption phase P4

probability of failure obtained from both analyses are very close to each other for the most critical phase P4. The lower computation time of FORM, which is ten times faster than that of the sampling analysis, makes it a better choice for the reliability analysis. The lower computation time in FORM is due to the fact that FORM doesn't need to draw the distribution function of the total travel time to calculate the probability of failure.

It can be concluded that the less critical phases P1, P2, and P5 which are less sensitive to CAV penetration rates, are expected to have similar findings.

Again, using FORM analysis, the reliability of total travel time in different disruption phases and different CAV penetration scenarios are obtained and are presented in Fig. 8. The results show that as the percentage of the market penetration rate of CAV increases, the travel time reliability is improved, especially in the case of P4. TTR follows a gradually decreasing trend in disruption phase P2 compared with phase P1. The reduction of TTR is 13.3% for a non-CAV penetration rate, suggesting

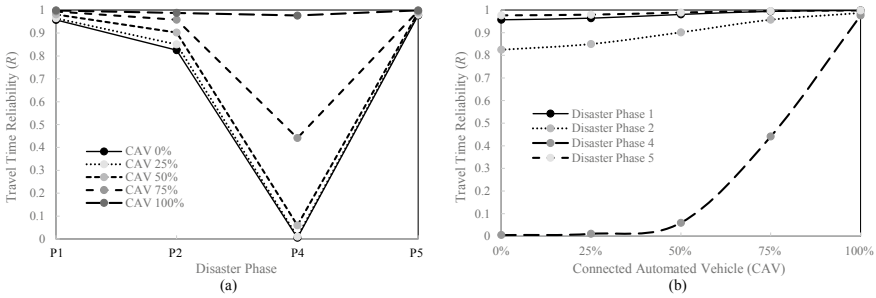


Fig. 8 OD 10 to 8 path travel reliability variation for **a** different CAV penetration and **b** different disruption phases

that 13.3% of the pre-disruption trips are affected by the disruption phase P2. These results also suggest that with higher CAV penetration rates, the disruption impacts could be mitigated. Specifically, the reduction of TTR for CAV penetration rate of 100% restricts the drop in the TTR to 1.1%.

In Fig. 8, it is observed that the TTR value of phase P4 depends heavily on CAV penetration rates because this phase is the critical situation immediately after the disaster and the presence of CAVs can help control the system. TTR in disruption phase P4 increases by 97.1% under 100% CAV penetration rate compared to a traffic environment without CAVs. In the post-disruption phase, P5, recovery actions are implemented to improve the functionality of the transportation system to the pre-disruption level. The maximum improvement of TTR in P5 compared with P4 (97.2%) occurs at a network with a non-CAV penetration rate. Therefore, more recovery time and an increase in the renovation budget are required for a non-CAV penetration rate network. In a network with a high CAV penetration rate, there is information exchange and connection between vehicles; thus, the low reduction of TTR at P4 happens. So, the high CAV penetration rate results in less needed recovery time and budget in P5 compared with P4.

4 Conclusion

This paper examines how travel time reliability as a performance measure of a transportation network can be evaluated in a mixed traffic environment. Analyses show that the performance of a transportation system is affected immediately before and after the disruption. However, CAVs are shown to reduce the mean travel time and thus improve the travel time reliability. Decreasing the travel time reliability immediately before and after disruption is due to some pre-disruption events such as rain before tornados and some post-disruption impacts such as failure of nodes and links. Increasing the CAV penetration rate is shown to significantly decreases the exceedance probabilities of total travel time in disruption phase P4. It also means

that the failure probability of travel time from a specific threshold decreased with increasing the CAV penetration rate and the related information exchange between connected and autonomous vehicles. So, with increasing the CAV penetration rate in networks, they will have higher travel time reliability before and after disruption than in the past.

One of the limitations of this research is the lack of access to real network-wide travel time data under various CAV penetration rates. Other performance metrics such as flow and capacity reliability can be calculated for a mixed traffic environment as a suggestion for future work. The approach presented in this paper can be a supportive tool for managers to determine unreliable paths in a transportation network. Therefore, more resources can be allocated to reinforce the operational structure and paths traffic conditions.

References

1. Ahmed S (2018) Resilience modeling of surface transportation system in mixed traffic environment. Graduate Theses, Dissertations, and Problem Reports. 3695
2. Ahmed S, Dey K, Fries R (2019) Evaluation of transportation system resilience in the presence of connected and automated vehicles. *Transport Res Record: J Transport Res Board* 2673(9):562–574
3. Armijo L (1966) Minimization of function having Lipschitz continuous first partial derivatives. *Pacific J Math* 16:1–13
4. Arnaout G, Bowling S (2011) Towards reducing traffic congestion using cooperative adaptive cruise control on a freeway with a ramp. *J Ind Eng Manage* 4(4):699–717
5. Asakura Y, Kashiwadani M (1991) Road network reliability caused by daily fluctuation of traffic flow. In: the 19th PTRC Summer annual meeting, University of Sussex, Brighton, UK
6. Bahaaldin K, Fries R, Bhavsar P, Das P (2017) A case study on the impacts of connected vehicle technology on no-notice evacuation clearance time. *J Adv Transp* 1964:1–9
7. Bell M, Cassir C, Iida Y (1999) A sensitivity-based approach to network reliability assessment. In: the 14th international symposium on transportation and traffic theory
8. Bell MGH, Lida Y (1997) Transportation network analysis
9. Bento LC, Parafita R, Rakha HA, Nunes UJ (2019) A study of the environmental impacts of intelligent automated vehicle control at intersections via V2V and V2I communications. *J Intell Transport Syst* 1–19
10. Chang JS (2010) Assessing travel time reliability in transport appraisal. *J Transp Geogr* 18(3):419–425
11. Chen K (2008) Models and algorithms for travel time reliability assessment of urban road networks based-on moving source data. Jiaotong University, Beijing, China
12. Cizelj L, Mavko B, Riesch-Oppermann H (1994) Application of first and second order reliability methods in the safety assessment of cracked steam generator tubing. *Nuclear Eng Design* 147
13. Fountoulakis M, Bekiaris-Liberis N, Roncoli C, Papamichail I, Papageorgiou M (2017) Highway traffic state estimation with mixed connected and conventional vehicles: microscopic simulation-based testing. *Transport Res Part C: Emerg Technol* 78:13–33
14. Genders W, Razavi S (2015) Impact of connected vehicle on work zone network safety through dynamic route guidance. *J Comput Civil Eng* 30(2)
15. Goodall NJ, Smith BL, Park B (2016) Microscopic estimation of freeway vehicle positions from the behavior of connected vehicles. *J Intell Transport Syst* 20(1):45–54
16. Gu Y, Fu X, Liu Z, Xu X, Chen A (2020) Performance of transportation network under perturbations: reliability, vulnerability, and resilience. *Transport Res Part E: Logistics Transport Rev* 133

17. Hannoun GJ (2017) Framework for better routing assistance for road users exposed to flooding in a connected vehicle environment. Dissertation, Virginia Tech
18. Husdal J (2005) The vulnerability of road networks in a cost-benefit perspective. In: The 84th Transportation Research Board annual meeting, Washington, DC
19. Iida Y (1999) Basic concepts and future directions of road network reliability analysis. *J Adv Transp* 33(2):125–134
20. Jose A, Ram S (2020) Travel time reliability to airport: review and assessment. *Transport Res Procedia* 48:2771–2783
21. Kattan L, Moussavi M, Far B (2010) Evaluating the potential benefits of vehicle to vehicle communication (V2V) under incident conditions in the PARAMICS model. In: The 13th International IEEE conference on intelligent transportation, Madeira, Portugal
22. Liu HX, Bu L, Pu Y (2004) Travel time reliability on stochastic road network. *Chin Civil Eng J* 37(8):102–105
23. Mahsuli M, Haukaas T (2013) Computer program for multimodel reliability and optimization analysis. *J Comput Civ Eng* 27(1):87–98
24. Melchers RE, Beck AT (2018) *Structural Reliability analysis and prediction*. Wiley, Hoboken
25. Mitra C, Singh J (1995) Monte Carlo simulation for reliability analysis of emergency and standby power systems. In: IEEE industry applications conference thirtieth IAS annual meeting
26. Mittal A, Mahmassani HS, Talebpour A (2017) Network flow relations and travel time reliability in a connected environment. *Transp Res Rec* 2622(1):24–37
27. Pei YL, Gai CY (2005) Study on operation reliability of highway network. *J Highway Transport Res Develop* 22(5):119–123
28. Tu H, Li H, Lint H, Zuylen H (2012) Modeling travel time reliability of freeways using risk assessment techniques. *Transport Res Part A: Policy Practice* 46(10):1528–1540
29. Wang JYT, Ehrgott M, Chen A (2014) A bi-objective user equilibrium model of travel time reliability in a road network. *Transport Res Part B: Methodol* 66:4–15
30. Wang J, Wang C, Lv J, Zhang Z, Li C (2017) Modeling travel time reliability of road network considering connected vehicle guidance characteristics indexes. *J Adv Transp* 2017(2):1–9
31. Wang R, Li Y, Work DB (2017) Comparing traffic state estimators for mixed human and automated traffic flows. *Transport Res Part C: Emerg Technol* 78:95–110
32. Yang F, Yun MP, Yang XG (2012) Single lane microscopic traffic flow model based on multi-agent in CVIS circumstance. *J Tongji Univ (Nat Sci)* 40(8):1189–1196
33. Zhao YG, Ono T (2001) Moment methods for structural reliability. *Struct Saf* 23:47–75
34. Zhu F, Ukkusuri SV (2017) An optimal estimation approach for the calibration of the car-following behavior of connected vehicles in a mixed traffic environment. *IEEE Trans Intell Transp Syst* 18(2):282–291

The Role of Microplastics' Size and Density on Their Vertical Turbulent Mixing and Transport



Arefeh Shamskhany and Shooka Karimpour

1 Introduction

Plastic input to the aquatic environment is expected to grow by an order of magnitude over the next decade [8]. Rivers are important transport pathways of aquatic plastics and are one of the major sources of transporting land-based plastic to oceans [11]. Plastic particles are categorized into two main groups based on their size: macroplastics, particles sized larger than 5 mm, and Microplastics (MPs), sized smaller than 5 mm. Both macroplastics and MPs are present in the aquatic environment for countless years, turning into smaller pieces through the fragmentation process [3]. Therefore, the presence of MP particles grows exponentially over the years. Due to polymers density, most MPs are expected to be present in surface layers [13]. However, the majority of MPs are missing from the epipelagic zone [17, 18]. This large missing portion of the MPs implies that MPs are transported to deeper parts of the aquatic environment. In-depth currents induced by thermal gradient, wind, or topographic changes, shape the transport of MPs and carry these small particles from uppermost layers to lower layers or resuspend settling particles [9].

Elimination of the aquatic MPs is impossible, and the presence of MPs in aquatic environments has detrimental effects on aquatic ecosystem health. For instance, aquatic species ingest MPs, which exposes them to the verge of annihilation [16]. Thus, knowing MP particle trajectories and final destinations will help us identify their risk to the aquatic ecosystem, which makes future risk mitigation possible. Recognizing the fate of these particles and their transport in the aquatic environment

A. Shamskhany (✉) · S. Karimpour
Department of Civil Engineering, Lassonde School of Engineering, York University, Toronto,
Canada
e-mail: shams019@yorku.ca

S. Karimpour
e-mail: shooka.karimpour@lassonde.yorku.ca

is a significant environmental challenge in recent years. MPs possess a wide range of properties, and their in-depth transport differs widely from other contaminants and sediments. The most important MPs' physical properties that affect their transport in aquatic environments are density, size, and shape. In contrast with sediment particles, the physical characteristics of MPs are very diverse. For instance, all silt particles have spherical shapes and the same range of sizes. However, MPs with the same density or the same size ranges are present in various shapes, affecting their dispersal, transport, and spread in aquatic systems [15]. The main reason for this difference is the drag force, which depends on the particle shape. Recent studies have conducted limited laboratory experiments with specific controlled conditions to establish equations for the settling velocity of irregular-shaped MPs (e.g., [10]). However, a natural water body does not meet the laboratory conditions of these experiments. Therefore, more studies are required to find a comprehensive equation for calculating the irregular MPs' drag force and settling velocity.

Due to the complexity of the process, a comprehensive field-based study to evaluate MPs' trajectory is not feasible. In natural aquatic ecosystems, a combination of hydrodynamic characteristics of ambient flow and physical properties of MPs, such as shape, density, and size, affect their transport and temporary and final destination of these particles. Jalón-Rojas et al. [7] are among the few recent studies that developed a 3D numerical package, TrackMPD, to track MP particles in the marine environment to investigate their disposal and spread. This model considers physical characteristics (including dimensions, shape, and density), physical processes (including biofouling and degradation), and hydrodynamic interactions (including advection, turbulent diffusion, windage, sinking, beaching, and washing off). In this study, the coupled impacts of flow hydrodynamics and properties of MPs are analyzed together. However, understanding MPs' trajectories and their transport require meticulous consideration and isolation of these characteristics. Computational hydrodynamic modeling provides an efficient tool, helping us better understand and quantify MP transport. This research numerically investigates the impact of selected MP properties on their vertical in-depth transport, deposition, and accumulation. Different numerical experiments have been conducted to evaluate the effectiveness of MPs' selected properties in their transport and final fate. We have used OpenFoam, an open-source computational fluidic tool, and have modified the Lagrangian transport module to accommodate various MP properties. Our preliminary results highlight the importance of MP size and density and ambient mixing on MP transport.

2 Methodology

2.1 Numerical Modeling and Governing Equations

This study aims to investigate the effect of the size and density of MPs on their transport. To isolate the effect of other key parameters, we assumed a spherical shape for

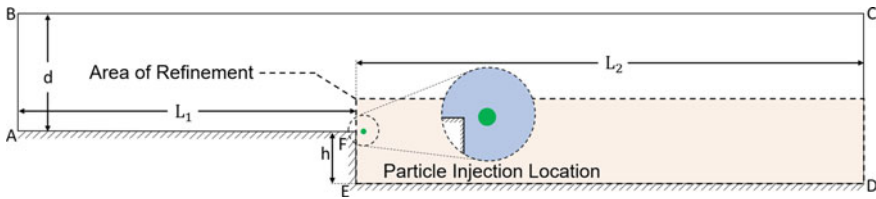


Fig. 1 Geometry of the BFS; $L_1 = 20$, $L_2 = 30$, $d = 7$, and $h = 3$ m; Injection point = (20.1, 3.0)

all particles in the present simulations. The first reason for this selection is that calculating the drag force and settling velocity for irregular MPs shapes is a challenging open-ended question that still requires further research [4, 10]. In addition, spherule particles or beads are reported among abundant aquatic MP shapes [2, 15]. In order to induce in-depth mixing, flow over a 2D transverse averaged Backward-Facing Step (BFS) is modeled. We used OpenFoam, an open-source computational fluidic tool with different pre-developed Eulerian and Lagrangian solvers for single- and multi-phase flow simulations. Here, we modified one of the pre-existing single-phase flow solvers to contain MP properties. The current solver is a hybrid Lagrangian–Eulerian computational model, where Lagrangian MP particle tracking is paired with an Eulerian hydrodynamic model. The simulation’s geometry is demonstrated in Fig. 1, where d is the inlet water depth, h is the step height, and L_1 and L_2 are lengths before and after the step, respectively. Boundary conditions used in this study are uniform flow with an inlet velocity of U_0 at the upstream inlet boundary (A-B), zero-gradient at the outlet and free surface boundaries (C-D and B-C, respectively), and no-slip at the lower wall boundary (A-F-E-D). The particle is injected into the ambient flow at $t = t_q$, when the turbulent flow is fully developed and reaches the quasi-steady state. The location of MP injection is illustrated in Fig. 1.

Lagrangian models are classified into four groups based on the volume fraction of the solid phase, which dictates the interaction between the particle and the ambient flow, or the coupling scheme [6]. In the current simulation, the mixture is dilute and dispersed. Therefore, here a one-way coupling scheme is applied to describe the interaction between particles and the ambient flow. In a one-way coupling system, only the ambient flow affects the behavior and trajectories of the particles, and the effect of particles on the surrounding flow is negligible. In this research, a single particle is injected into the fully developed turbulent flow, and consideration for particle interactions is not included. The first step is to solve for the hydrodynamic of the ambient flow. For this purpose, a set of conservation of mass and momentum equations, as demonstrated in Eqs. 1 and 2, are solved.

$$\nabla \cdot (\rho U) = 0 \tag{1}$$

$$\frac{\partial U}{\partial t} + \nabla \cdot (UU) = -\frac{1}{\rho_f} \nabla P + \nabla \cdot (\nu \nabla U) + \frac{1}{\rho_f} \nabla \tau^{SGS} \tag{2}$$

where ρ_f is the density of the ambient flow, and here it is assumed as 1020 kg/m^3 (the density of the seawater), P is the pressure, ν is the kinematic viscosity of the fluid, g , is the gravitational acceleration, U is the velocity of the flow, and τ^{SGS} is the turbulent stress. In order to investigate the effect of the turbulent ambient flow, MPs transport in a turbulent flow over a BFS is simulated using the Large Eddy Simulation (LES) model. In the LES approach, the effects of the eddies smaller than the grid-scale (sub-grid scale) are modeled using Smagorinsky sub-grid scale viscosity [14]. The position of each particle needs to be updated at the end of each time step to find the trajectories of MPs. Therefore, the MP displacement in two directions is calculated using Eq. 3:

$$\frac{dx_p}{dt} = U_p \quad (3)$$

In Eq. 3, U_p is the particle's velocity, dt is the Lagrangian time-step, and x_p is the particle's new position. Here, we assumed two different time steps for Eulerian and Lagrangian solutions. Since the simulation is one-way coupling, the Lagrangian time-step should be equal to or smaller than the Eulerian time-step. Newton's second law is used for each particle as a free body to calculate the particle's velocity (Eq. 4).

$$m_p \frac{dU_p}{dt} = F_D + F_G + F_B \quad (4)$$

where, m_p is the particle's mass, and F_D , F_G , and F_B are drag, gravitational, and buoyancy force components. This paper focuses on perfectly spherical MPs, known as microbeads. Thus, the active vertical force components, including drag force, are calculated based on Eqs. 5 to 7.

$$F_D = \frac{1}{2} \rho_f C_D A_p |U - U_p| (U - U_p) \quad (5)$$

$$F_G = \rho_p g V_p \quad (6)$$

$$F_B = -\rho_f g V_p \quad (7)$$

where A_p is the projected area, ρ_p is the density of the particle, V_p is the volume of the particle, and C_D is the drag coefficient calculated using the Schiller-Naumann (1935) method for spherules:

$$C_D = \begin{cases} \frac{24}{\text{Re}_p} \left(1 + 0.15 \text{Re}_p^{0.687} \right), & \text{Re}_p \leq 1000 \\ 0.44, & \text{Re}_p > 1000 \end{cases} \quad (8)$$

where Re_p is the particle Reynolds number, and is calculated based on Eq. 9:

$$\text{Re}_p = \frac{D_p |U - U_p|}{\nu} \quad (9)$$

where D_p is the particle diameter. An explicit scheme is used to solve Eq. 4, and the particle's velocity is calculated based on the previous time-step.

2.2 *MPs' Physical Properties*

In the present research, two different densities and two size classes are selected based on several field observations of aquatic MP abundance. [13] carried out a comprehensive literature review on the prevalence of different polymer types in aquatic environments, demonstrating that the most abundant polymer type is Polyethylene. This polymer is manufactured for a variety of purposes and thus includes a wide range of densities. Here we used two polymer types of Polyethylene (PE) and Polyethylene Terephthalate (PET) as two density classes (Table 1). In a similar approach, we have looked at MP size distribution in deep-sea sediments and surface layers. The size distribution of floating plastic debris in the open ocean has demonstrated a peak of distribution, both in terms of weight and count, at around 2 mm [5]. In addition, in a recent field study, based on samples collected from surface layers to deep sediments, the majority of MPs possess a size range of 0.5–2 mm [12]. Accordingly, two size classes of 2 and 0.5 mm are selected for the current study. In addition to MPs' physical characteristics, hydrodynamic properties of the ambient flow also have a significant effect on MPs' trajectories and final fate. Thus, we also investigated the role of turbulent coherent structures for four different inlet velocities of $U_o = 0.04, 0.1, 0.2,$ and 0.4 m/s. For each inlet velocity, four cases with different sizes and densities are simulated. Table 1 represents the selected physical properties of executed cases for each inlet velocity.

Table 1 Particle properties in executed cases

| Case name | Polymer type | Density (kg/m ³) | Size (mm) |
|-----------|--------------|------------------------------|-----------|
| PET-L | PET | 1410 | 2 |
| PET-S | PET | 1410 | 0.5 |
| PE-L | PE | 940 | 2 |
| PE-S | PE | 940 | 0.5 |

3 Results and Discussion

3.1 Contours

In our simulations, flow over the BFS reaches a quasi-steady state before the injection of MP particle. This is illustrated in Figs. 1 and 2. The velocity gradient at the corner of the BFS induces the formation of a shear layer. This generates continuous vortex shedding. Some of the vortices are advected with the mainstream, and others are entrained behind the BFS. Therefore, MP particles are exposed to instantaneous variations in the flow domain, affecting their transport and dispersal. Figure 2 demonstrates results for four executed cases (Table 1), each for two inlet velocities of $U_0 = 0.1$ and 0.2 m/s. This figure shows the time lapse of MP movement up to 30 s after

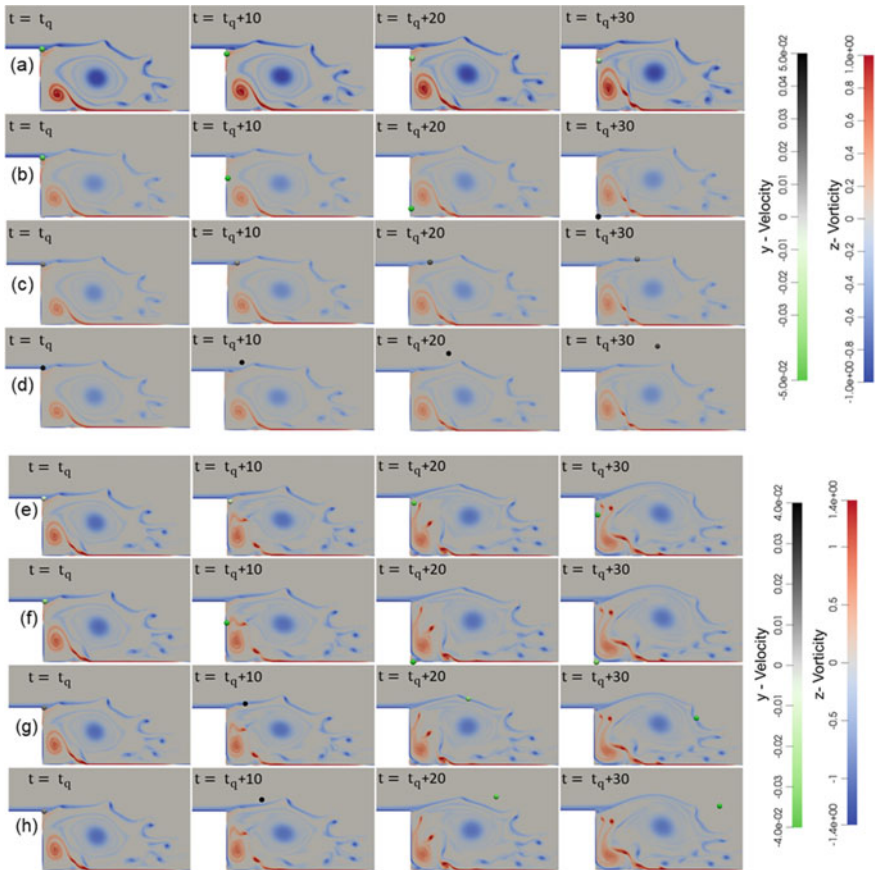


Fig. 2 MP trajectories behind the BFS; top four rows: PET-S (a), PET-L (b), PE-S (c), and PE-L (d), at $U_0 = 0.1$ m/s; bottom four rows: PET-S (e), PET-L (f), PE-S (g), and PE-L (h), at $U_0 = 0.2$ m/s

the injection. The coherent turbulent structure of the ambient flow is visualized using the vorticity contour and particle's color shows its vertical velocity. These contours are focused on the recirculation zone immediately downstream of the step. For both inlet velocities of $U_0 = 0.1$ and 0.2 m/s, heavy PET particles in both sizes settle with minimal involvement with the ambient flow. This is illustrated in Fig. 2a–f. While the small PE particles, Figs. 2c, g, follow the ambient flow closely. Whereas, as shown in Figs. 2d, h, pathways of larger PE particles tend to recede from the vorticity profile and float in both inflow velocities.

3.2 Particle Froude Number

Based on the vertical active forces' equilibrium, we defined criteria to investigate the effect of MPs' physical properties on their vertical transport. When a particle is injected into the ambient flow, three forces are acting upon the particle in the vertical y -direction. Two force terms, buoyancy in the positive and weight in the negative y -direction, only depend on the physical characteristics of the particle and the ambient fluid. The third force term is the drag force, which also depends on the particle's relative velocity with respect to the ambient flow and acts opposite to the relative motion of particles. In the present simulations, for buoyant PE particles that tend to move up, the drag would act downward. In contrast, for PET particles that tend to sink, drag exerts a resistance in the upward direction. Figure 3 demonstrates active vertical forces on PE and PET microbeads right after the injection, where the particle's velocity is zero.

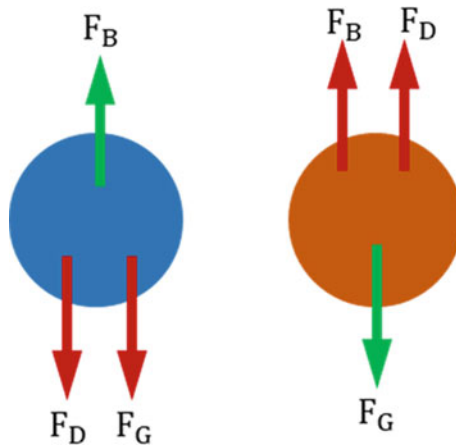


Fig. 3 Free body diagram for PE (blue) and PET (orange) particles for initial particle velocity of $U_p = 0$. F_D is the drag force, F_B is the buoyancy, and F_G is the self-weight of the particle

For a microbead substituting Eqs. 5–7, and assuming that the initial velocity of the particles right after the injection ($t = t_q$) is zero, vertical forces are:

$$\sum F_y = \mp \frac{1}{2} \rho_f C_D A_p U^2 - \rho_p g V_p + \rho_f g V_p \quad (10)$$

Buoyancy and gravitational force components are constant for a specific particle; however, the drag force varies based on the ambient flow velocity at the location where the particle is introduced. In order to investigate the effect of turbulent intensity, as a simplifying assumption, the inlet velocity is used for drag calculation at the injection time. Assuming a state of equilibrium, we rearranged Eq. 10 and put the drag force in the Left Hand-Side (LHS), and the self-weight and buoyance in the Right Hand-Side (RHS), which gives us Eq. 11.

$$\frac{U^2}{\Delta g \frac{V_p}{A_p}} = \frac{2}{C_D} \quad (11)$$

where Δg is the reduced gravity and equals to $\left(1 - \frac{\rho_p}{\rho_f}\right)g$. The dimensionless particle Froude number is defined based on the incoming flow velocity as [1]:

$$Fr_p = \frac{U_o}{\sqrt{\Delta g \frac{V_p}{A_p}}} \quad (12)$$

In Eq. 11, as long as $LHS \leq RHS$, the particle tends to move naturally, and the in-depth y -velocity due to the coherent structure behind the BFS is not strong enough to deflect MP from its natural trajectory. However, once the ambient flow in-depth y -velocity reaches the threshold value, $LHS > RHS$, the particle tends to move vertically with eddies. In other words, PE particles that are naturally buoyant, sink, and PET particles that naturally sink, rise up. For all cases in Table 2, the particle Froude number, Fr_p is examined at $t = t_q$. In Fig. 4, the particle Froude number, Fr_p is reported with respect to $\sqrt{\frac{2}{C_D}}$. It is expected that particles in which Fr_p is comparable to or smaller than $\sqrt{\frac{2}{C_D}}$ tend to move naturally. Figure 4 demonstrates that in two cases of PET-L at inlet velocities of 0.04 and 0.1 m/s and one case of PE-L at inlet velocity of 0.04 m/s, the particle Froude number is less than the threshold value. This means, in these cases, the particle moves rather passively and the microbead is not engaged with the turbulent structures. Instead, the particle would follow its natural trajectory due to its buoyance and weight. The trajectory of these three passive cases, lying above the dashed line in Fig. 4, are presented in Fig. 5. In Fig. 5, the MP particle's position is nondimensionalized using the step height. For the PET-L case, in which the particle is heavier than the ambient flow, for both inlet velocities of 0.04 and 0.1 m/s, particles sink immediately after the injection, and by 30 s after the injection, they settled on the bed. Also, in the PE-L case, in which the particle

Table 2 Executed cases analytical properties

| U_0 (m/s) | Case name | Fr_p | τ (s) | ξ_{max} (1/s) | St |
|-------------|-----------|---------|------------|-------------------|--------|
| 0.04 | PET-L | 0.5656 | 0.3133 | 0.5 | 0.1567 |
| | PET-S | 1.1312 | 0.0196 | 0.5 | 0.0098 |
| | PE-L | 1.2489 | 0.2089 | 0.5 | 0.1044 |
| | PE-S | 2.4977 | 0.0131 | 0.5 | 0.0065 |
| 0.1 | PET-L | 1.4140 | 0.3133 | 1.0 | 0.3133 |
| | PET-S | 2.8281 | 0.0196 | 1.0 | 0.0196 |
| | PE-L | 3.1221 | 0.2089 | 1.0 | 0.2089 |
| | PE-S | 6.2443 | 0.0131 | 1.0 | 0.0131 |
| 0.2 | PET-L | 2.8281 | 0.3133 | 1.4 | 0.4387 |
| | PET-S | 5.6562 | 0.0196 | 1.4 | 0.0274 |
| | PE-L | 6.2443 | 0.2089 | 1.4 | 0.2924 |
| | PE-S | 12.4885 | 0.0131 | 1.4 | 0.0183 |
| 0.4 | PET-L | 5.6562 | 0.3133 | 2.0 | 0.6267 |
| | PET-S | 11.3124 | 0.0196 | 2.0 | 0.0392 |
| | PE-L | 12.4885 | 0.2089 | 2.0 | 0.4178 |
| | PE-S | 24.9771 | 0.0131 | 2.0 | 0.0261 |

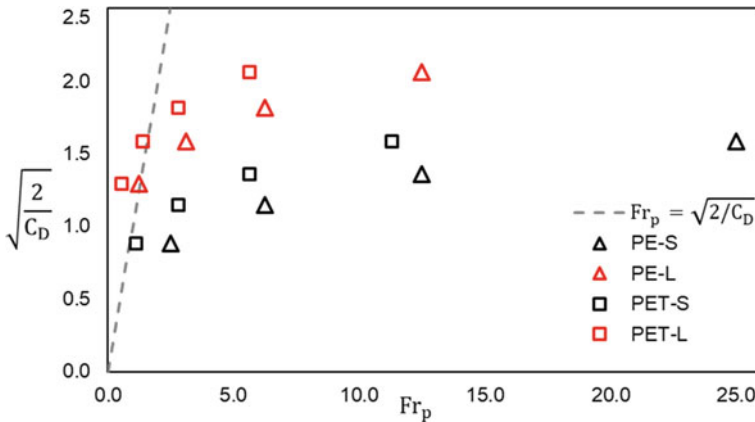


Fig. 4 Particle Froude number, Fr_p versus drag coefficient at $t = t_q$; cases above the dashed line did not reach the threshold Fr_p , and tend to move naturally; cases below the dashed line are transported by the ambient flow circulations

is lighter than the ambient flow, the particle starts to rise up immediately after the injection. In the latter case, the particle is carried in the streamwise direction due to the x-component of the flow velocity. This is while at $Fr_p > 10$, for inlet velocity of $U_0 = 0.4$ m/s, all particles, regardless of the size and density, are heavily entrained with the ambient flow (Fig. 6(b)).

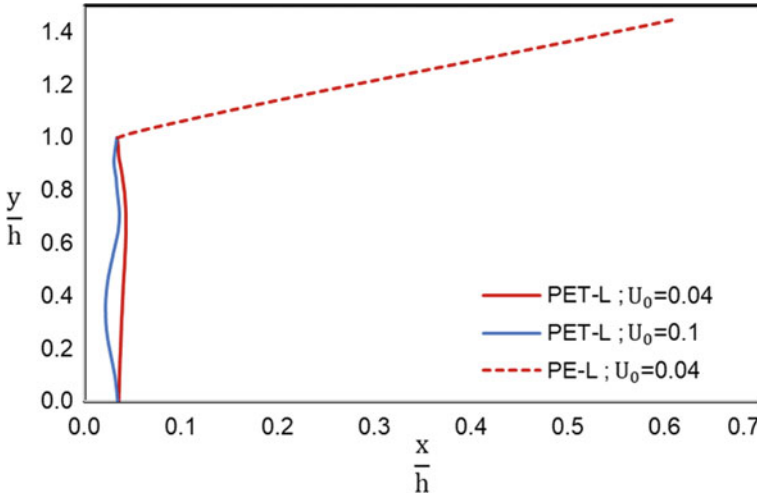


Fig. 5 MP trajectories for three passive cases which $Fr_p < \sqrt{\frac{2}{C_D}}$; simulation duration = $t_q + 30$ s

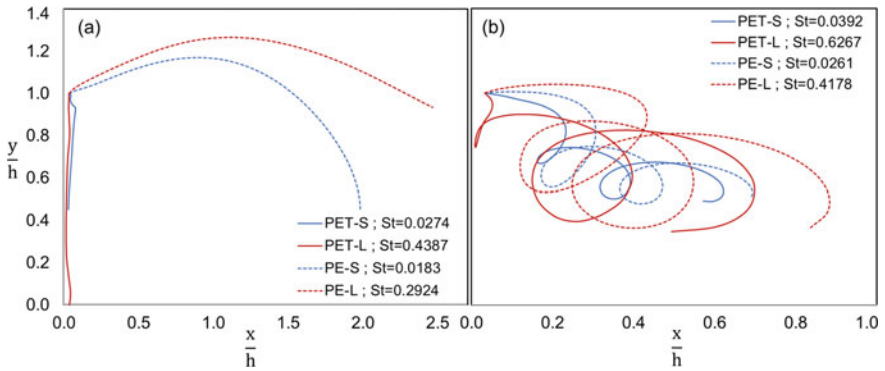


Fig. 6 MP trajectories for inlet velocity of **a** $U_0 = 0.2$, and **b** 0.4 m/s; simulation duration = $t_q + 30$ s

3.3 Relaxation Time and Stokes Number

In a turbulent regime, particle transfer and entrainment are often characterized by particle relaxation time. The particle relaxation time used here is estimated based on the difference between the particle density and the fluid density. This parameter quantifies the time-scale required for MP particles to respond to the changes in the ambient flow. This relaxation time, τ , is defined based on Eq. 13:

$$\tau = \frac{|\rho_p - \rho_f| D_p^2}{18\nu} \tag{13}$$

In addition to the physical characteristics of the particle and the flow, hydrodynamic characteristics of the flow also affect the trajectory of a particle. i.e., for two MP particles with the same physical properties, trajectories would not be the same if the hydrodynamic characteristics of the surrounding flow are different. Therefore, to anticipate the involvement of MPs with the turbulent ambient flow, we need a criterion that depends on the physical characteristics of MP particles as well as the hydrodynamics of the ambient flow. In the current simulation, MP particles are transported in vortices larger than sub-grid scale, with coherent structures. To signify the involvement of MPs with larger eddies, we have defined a Stokes number, St , based on the product of the relaxation time and the vorticity time-scale:

$$St = \tau \xi_{\max} \quad (14)$$

where St is the Stokes number, and ξ_{\max} is the maximum field vorticity. Table 2 presents relaxation time, τ , maximum field vorticity, ξ_{\max} , and Stokes number, St for all sixteen cases of this research.

At low velocities of $U_0 = 0.04$ and 0.1 m/s, the particle Froude number is small. Therefore, the ambient flow does not impact the movement of naturally buoyant or sinking plastics. This is despite observing a very low Stokes number, for instance, $St = 0.0065$ and 0.0131 for PE-S at inlet velocities of $U_0 = 0.04$ and 0.1 m/s, respectively. Figure 6a shows a similar trend in all cases at $U_0 = 0.2$ m/s. At higher velocities, particle Froude number increases, and hence MPs can potentially entrain in the ambient flow's turbulent structure. Therefore, as the particle Froude number and Stokes number increase, the particle involvement with the coherent structure enhances. For larger particle sizes, both relaxation time and the Stokes number are higher than those of smaller particles. A higher particle Froude number ensures engagement of the particle with the ambient flow. However, a higher particle Stokes number translates to a greater lag between the ambient flow and the particle. For two sizes of microbeads with the same polymer type and the same ambient flow characteristics, larger particles' trajectories are closer to the natural movement of these particles than smaller MPs. Figure 6a, b shows the trajectories of four cases at $U_0 = 0.2$ and 0.4 m/s. At $U_0 = 0.4$ m/s, both PET-L and PE-L cases have the highest Stokes number. Therefore, in these cases, the pattern of movement greatly deviates from the ambient flow. In Fig. 6b, the upward movement of the PE-L case, which is a buoyant particle, is more dominant compared to its downward movement. Similarly, the sinking of the PET-L case, which is a dense particle, is dominant in the direction of gravity compared to its upward movements. On the other hand, based on Fig. 6b, in PET-S and PE-S cases with comparable Stokes numbers, particle trajectories are similar. Thus, for the smaller Stokes number, the effect of density is less important, and despite the significant density difference, at a higher inlet velocity, both PET-S and PE-S particles follow close trajectories. At $U_0 = 0.2$ m/s, all cases have a smaller particle Froude number and Stokes number than the corresponding cases at $U_0 = 0.4$ m/s.

4 Conclusion

Aquatic MP transport and fate are dominated by many factors, including particle size and density. Ambient flow characteristics, including in-depth and streamwise turbulent motions, also affect the dispersal of MPs. We have examined the role of MP size and density based on results obtained from sixteen cases of numerical simulations. We have identified two dimensionless numbers, particle Froude number and Stokes number, that describe the entrainment and vertical transport of MP particles. At smaller particle Froude numbers, MPs are passive to the ambient flow, and their self-weight and buoyance mainly dominate their trajectory. In contrast, MP particles with larger Froude numbers are entrained in the in-depth recirculation behind the BFS. On the other hand, the Stokes number quantifies the lag between the MP particles and the ambient flow. Due to their low Stokes number, small MP particles introduced to a higher inlet velocity follow the flow path closely. In such conditions, the effect of density is negligible, and all particles, regardless of their density, follow a similar trajectory in the turbulent structure.

Our simulations demonstrate that aquatic MP dispersal depends on their selected properties. A further detailed investigation into the role of particle Stokes and Froude number and numerical validation will be presented in a future publication.

References

1. Aguirre-Pe J, Olivero ML, Moncada AT (2003) Particle densimetric Froude number for estimating sediment transport. *J Hydraulic Eng* 129(6):428–437
2. Ballent A, Corcoran PL, Madden O, Helm PA, Longstaffe FJ (2016) Sources and sinks of microplastics in Canadian Lake Ontario nearshore, tributary and beach sediments. *Mar Pollut Bull* 110(1):383–395. <https://doi.org/10.1016/j.marpolbul.2016.06.037>
3. Barnes DKA, Galgani F, Thompson RC, Barlaz M (2009) Accumulation and fragmentation of plastic debris in global environments. *Philos Trans R Soc B: Biol Sci* 364(1526):1985–1998
4. Chubarenko I, Bagaev A, Zobkov M, Esiukova E (2016) On some physical and dynamical properties of microplastic particles in marine environment. *Mar Pollut Bull* 108(1–2):105–112. <https://doi.org/10.1016/j.marpolbul.2016.04.048>
5. Cózar A, Echevarría F, González-Gordillo JJ, Irigoien X, Úbeda B, Hernández-León S et al (2014) Plastic debris in the open ocean. *Proc Nat Acad Sci* 111(28):10239–10244
6. Elghobashi S (1994) On predicting particle-laden turbulent flows. *Appl Sci Res* 52(4):309–329
7. Jalón-Rojas I, Wang XH, Fredj E (2019) A 3D numerical model to track marine plastic debris (TrackMPD): sensitivity of microplastic trajectories and fates to particle dynamical properties and physical processes. *Mar Pollut Bull* 141:256–272
8. Jambeck JR, Geyer R, Wilcox C, Siegler TR, Perryman M, Andrady A, Narayan R, Law KL (2015) Plastic waste inputs from land into the ocean. *Science* 347(6223):768–771
9. Kane IA, Clare MA (2019) Dispersion, accumulation, and the ultimate fate of microplastics in deep-marine environments: a review and future directions. *Front Earth Sci* 7. <https://doi.org/10.3389/feart.2019.00080>
10. Khatmullina L, Isachenko I (2017) Settling velocity of microplastic particles of regular shapes. *Mar Pollut Bull* 114(2):871–880
11. Lebreton LCM, Van Der Zwet J, Damsteeg J-W, Slat B, Andrady A, Reisser J (2017) River plastic emissions to the world's oceans. *Nat Commun* 8:15611

12. Sagawa N, Kawaai K, Hinata H (2018) Abundance and size of microplastics in a coastal sea: comparison among bottom sediment, beach sediment, and surface water. *Marine Pollution Bull* 133:532–542. <https://doi.org/10.1016/j.marpolbul.2018.05.036>
13. Schwarz AE, Ligthart TN, Boukris E, van Harmelen T (2019) Sources, transport, and accumulation of different types of plastic litter in aquatic environments: a review study. *Mar Pollut Bull* 143(March):92–100. <https://doi.org/10.1016/j.marpolbul.2019.04.029>
14. Smagorinsky J (1963) General circulation experiments with the primitive equations: I. The basic experiment. *Mon Weather Rev* 91(3):99–164
15. Song YK, Hong SH, Jang M, Kang JH, Kwon OY, Han GM, Shim WJ (2014) Large accumulation of micro-sized synthetic polymer particles in the sea surface microlayer. *Environ Sci Technol* 48(16):9014–9021. <https://doi.org/10.1021/es501757s>
16. Taylor ML, Gwinnett C, Robinson LF, Woodall LC (2016) Plastic microfibre ingestion by deep-sea organisms. *Sci Rep* 6(1):1–9
17. Ter Halle A, Ladirat L, Gendre X, Goudounèche D, Pusineri C, Routaboult C, Tenailleau C, Duployer B, Perez E (2016) Understanding the fragmentation pattern of marine plastic debris. *Environ Sci Technol* 50(11):5668–5675
18. Van Sebille E, England MH, Froyland G (2012) Origin, dynamics and evolution of ocean garbage patches from observed surface drifters. *Environ Res Lett* 7(4):44040

Establishing Workflow Processes for Bridge Inspections Using RPAS



C. Schmelzle and M. Paulsen

1 Introduction

Building upon work completed in 2018, RPAS continue to be used to supplement bridge inspection projects. RPAS are an effective tool for complex bridge structures when compared to typical inspection methodologies, such as Under Bridge Inspection Vehicles (UBIV's) or Rope Access. They provide benefits in numerous areas, as detailed below in Table 1.

1.1 Case Studies

In recent years, wide use of RPAS has led to the development of workflows that are consistent with a range of structure types and challenging site-specific conditions. A selection of these projects was chosen as case-studies for this paper which capture the range of conditions, as detailed below in Table 2.

These projects varied significantly in terms of geographical area, client type and RPAS system used. Each inspection required a review of the site conditions to determine the most suitable approach and discussions with the asset owner to ensure all necessary documentation was prepared. Full photo coverage of the structures was obtained to set a clear baseline on condition and be able to compare images of the same location or defect, inspection over inspection. All structures inspected above were major bridges over waterbodies, railyards and mountain canyons. RPAS

C. Schmelzle (✉) · M. Paulsen
Associated Engineering, Edmonton, AB, Canada
e-mail: schmelzlec@ae.ca

M. Paulsen
e-mail: paulsenm@ae.ca

Table 1 RPAS benefits

| Criteria | Benefit |
|-----------------------|--|
| Access | Provides full access to bridge underside and piers/bearings |
| Traffic accommodation | Removes need to close lanes and place barricades |
| Safety | Reduces time spent by inspector requiring specialized access |
| Cost | Limits labour and large equipment requirements |
| Load restrictions | Usable regardless of load restrictions on the bridge |

provide the greatest benefit on large and complex structures, rather than standard smaller bridges with easier access.

2 Workflow Development

As asset owners incorporate RPAS into inspection programs, there is an increased need to develop a structured workflow process. An established methodology will allow for consistent collection of bridge inspection data. This workflow is guided by the following points;

- Establishing the expectations and capabilities of the RPAS
- Determining the qualifications and requirements for relevant sub-consultants
- Developing a clear and coherent site-specific inspection plan
- Defining specifications/criteria for the final deliverable.

These ideas will be used to develop the workflow. The process will be broken down by the various phases of a bridge inspection project using RPAS (Fig. 1).

2.1 *Assessment of RPAS and Role in Bridge Inspection Programs*

RPAS to supplement visual inspections is still relatively new practice, and as such assessing if the technology is suitable for the work is a necessary first step. Before proceeding with RPAS inspections the site(s) should be reviewed and a high-level assessment performed to determine if RPAS are suitable. Some questions that should be investigated include;




- Can the site be easily accessed by foot?
- What is the level of information/detail required?

Table 2 RPAS bridge inspection case studies

| Bridge | Location | Year | Photo | Lessons learned/notes |
|---------------|------------------|------------|---|--|
| Maple avenue | Medicine Hat, AB | 2018 |  | <ul style="list-style-type: none"> • Flat bank allowed for good access for takeoff/landing • System used had short battery life, resulting in longer inspection time |
| Camrose drive | Camrose, AB | 2018, 2021 |  | <ul style="list-style-type: none"> • Bridge was located along pedestrian trail, requiring traffic accommodation • System used was not able to closely reach the pier caps and bearings |


(continued)

Table 2 (continued)

| Bridge | Location | Year | Photo | Lessons learned/notes |
|------------------------|----------------------|------------------|---|--|
| Athabasca River Bridge | Fort McMurray, AB | 2019, 2020, 2021 |  | <ul style="list-style-type: none"> • System lost GPS signal, requiring a manual flight • Length of bridge made maintaining visual contact difficult • Presence of pipelines and steel girders made flights more hazardous |
| 4th Ave Viaduct | Moose Jaw, SK | 2020 |  | <ul style="list-style-type: none"> • Flight over active CP rail required special planning • Operators had to work around pedestrians |
| Battleford bridges | North Battleford, SK | 2020 |  | <ul style="list-style-type: none"> • The arches blocked signal and video feed, removing the ability to fly from the deck • Timed inspection around low water levels to provide best access from banks |

(continued)

Table 2 (continued)

| Bridge | Location | Year | Photo | Lessons learned/notes |
|-----------------|-----------|------|---|---|
| Johnston Canyon | Banff, AB | 2020 |  | <ul style="list-style-type: none"> • Very tight areas and small catwalks made access challenging • New system was chosen to satisfy site conditions |

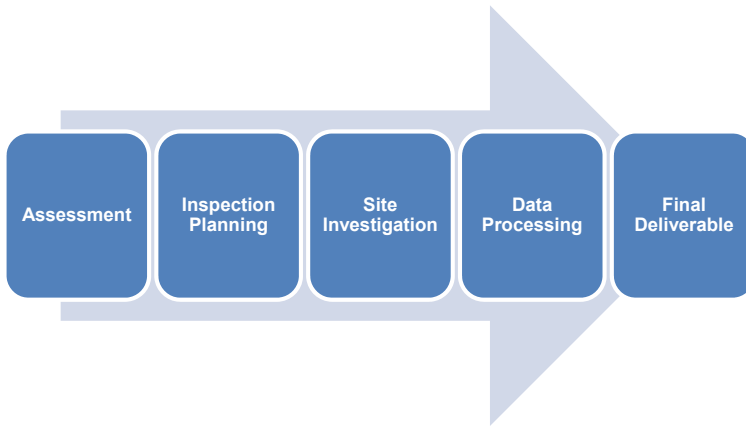


Fig. 1 Bridge inspection project phases

- Are there operators available that can complete the work?

RPAS provide the most significant benefit for large/complex bridges that cannot be easily or cost-effectively accessed. The inspections often require a high level of detail, as these structures are likely critical to the client's overall roadway or trail network. There may also be special permits needed based on location of the structure and general airspace in the area. For the Johnston Canyon work (Table 2), while drone use is generally prohibited in Banff National Park, special permission and permits were granted by the Project Team to allow the flights to occur.

Once established as a viable option, extent of data gathering and uses need to be determined: for example, developing a 3D model of the structure through photogrammetry, or only referencing plans with documented photos. It is also worth considering the role of RPAS inspection data in the overall bridge maintenance program. As RPAS allow for full photo coverage at each inspection, deterioration can be mapped and tracked through photos.

2.2 Current Industry Practice in Use of RPAS

In preparing this paper, a review was conducted of current bridge inspection practices, with a focus on RPAS. RPAS is a common tool in the construction industry for a wide range of tasks, including survey and construction management. In addition to this, regulations around the use of RPAS have significantly loosened in the past five years, resulting in a high number of companies forming and individuals becoming pilots. In an effort to grow, companies may take on scope that they are not directly familiar with, with the idea of 'learning on the fly'. It is essential that prospective operators can speak to direct experience inspecting bridges with RPAS and have the

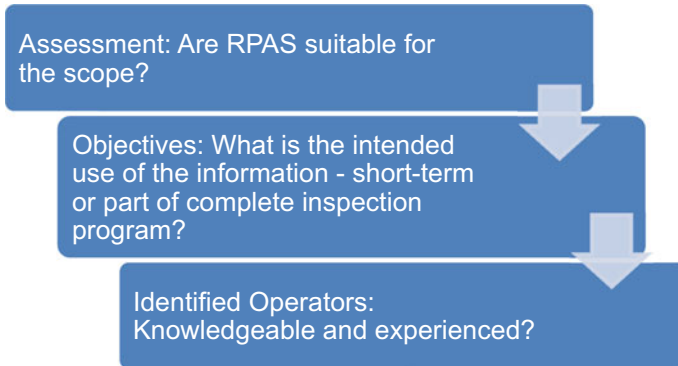


Fig. 2 Assessment phase breakdown

required certifications based on their jurisdiction. In Canada, there are two levels of RPAS Certification; Basic and Advanced. Any pilot flying commercially for the purpose of inspecting a bridge should have their Advanced certificate, as it allows them to fly in controlled areas and above the public. Once the pilot has been vetted, they should be shown the site and structure type to be inspected to ensure that their equipment will function properly. The Athabasca River Bridge (Table 2) had large deep steel girders, which affected the communication between the drone and the controller. The operators had to return to site with a different system to complete the flights. Similarly, for the Battleford Bridges (Table 2), large steel truss members adjacent to the deck blocked the signal and prevented the operators from flying from the deck.

Going through this process will help to understand the requirements for the data and suitability of RPAS for the inspection (Fig. 2).

2.3 Inspection Planning

Inspection planning involves the inspection team and RPAS operator team reviewing the site, confirming the scope and defining the details before heading to site. All team members involved with the inspection in any capacity should participate in these discussions.

The focus should be the flight plan. The flight plan should run through all aspects of the RPAS flight, as detailed below in Table 3.

If possible, a site visit should be conducted with the inspection team prior to the actual inspection and a short test flight carried out. This will allow everyone to become familiar with the site and potentially identify previously unknown hazards. This proved to be especially useful for the Athabasca River Bridge, as access to the banks was restricted in certain areas.

Table 3 Flight plan aspects

| Aspect | Description |
|------------------------|--|
| Takeoff and landing | Establish where the RPAS will take off from and land. Multiple points may be required based on bridge length and complexity. It is important that the operators maintain visual line of sight with the RPAS at all times |
| Flight path | Define the path of the RPAS along the structure. This will ensure that full coverage of the bridge is caught and allow any hazards with the bridge structure itself to be identified |
| Battery life | Confirm the battery life for the RPAS system and how it relates to how long/far the unit can fly. Ensure that replacement batteries are available and be aware of cold weather which may result in reduced battery life |
| Coverage requirements | Explain the extent of photo/video coverage that is needed as part of the scope. Ensure all bridge terminology is described and that the operators are fully aware of the bridge elements they need to capture. Providing and reviewing an elevation drawing is recommended |
| GPS connectivity | Review the RPAS system that will be used for GPS connectivity requirements. Based on the bridge type and nearby utilities, the units may have difficulty maintaining GPS connection, resulting in the operators flying the bridge manually |
| Photo properties | Based on the intended usage of the photos, review the camera being used and if it will capture high quality photos (minimum 12 megapixels). Discuss how the operator will account for poor lighting and glare |
| Safety hazards | Identify all safety hazards in how it relates to the RPAS unit, operators and public. The operators are trained pilots and should not be bothered while flying. The Advanced certification in Canada allows for pilots to fly over and in the proximity of pedestrians, but it is always recommended to avoid the public if possible |
| Weather | Review the weather forecasts, as extreme cold and high winds will significantly impact the ability to carry out the flights. The RPAS system specifications should be reviewed for weather restrictions |
| Required documentation | Review all safety, insurance and certification documentation with the client and have a copy on site if needed |

2.4 Site Investigation

Once the inspection team meets on site, the flight plan should be reviewed, and all aspects confirmed before starting the inspection. The weather should also be double-checked to ensure that the flight can still move ahead. Once the plan is reviewed and no issues are identified, the operators can move to the first takeoff point and commence the flight. As RPAS typically act as a supplement to the standard bridge inspection process, while the operators are flying, the inspection team can still inspect all aspects of the bridge that are within hands reach.

While flying, the operators should not be disturbed. Asset owners are often interested in this unique type of work and stop by the site to visit and discuss. All discussions with the operators should happen once the RPAS unit has safely landed. It is recommended that the inspection team remains on site until the flights have been completed. This ensures that they are available to answer any questions as it pertains to the required coverage. This approach worked well on the Maple Avenue Bridge, Camrose Drive Bridge and 4th Ave Viaduct (Table 2). Since these bridges are in pedestrian corridors, people would often stop to watch the flight and want to ask questions. The inspector was able to explain to them what was happening and remind them to not bother the operators.

2.5 Data Processing

The data processing phase involves the operators reviewing the photos for clarity/color and assigning the naming convention for all photos taken. All photos should be geotagged, allowing for a visual representation of where each photo was taken. This phase of the project can take a significant amount of time, ranging from 2 to 6 weeks. The time will vary depending on the number of photos taken. It is recommended to stay in constant communication with the operators as they process the data. Receiving a sample set of one span or one pier may allow gaps in the coverage to be identified. Depending on the location of the project, RPAS require a relatively low effort for mobilization, and it may be possible to revisit the sites if certain areas were missed or more data is needed.

It is important that operators properly label each photo and describe the element shown. This is critical for when one photo captures multiples of the same element. For example, when taking photos of cables for a stay-cable bridge, multiple cables will be visible in one photo, thus making it difficult to distinguish exactly which numbered cable is being focused on. A naming convention should be described and agreed upon with the team to allow for an efficient review and proper identification of defects.

The photos are then passed on to the inspection team and undergo a thorough review for defects. The need for high quality photos is reinforced at this stage, as the clarity and zoom capabilities are essential to identifying defects. This is shown in Fig. 3, with the original photo on the left and the zoom-in of the bearing on the right.

The inspector must take time to review all photos for potential defects. If at this stage a significant defect is found that requires further investigation or a hands-on inspection, the typical inspection methodologies discussed in Sect. 1 can be used. The benefit is that resources have not been spent implementing these methodologies throughout the full length of the structure, but only the specific area for which it was required.

This phase of the project will also depend on the final deliverable. If the deliverable is only the photos themselves, this phase will be limited to the above. Options exist for further processing, including photogrammetry. Typical deliverables have been



Fig. 3 Recommended zoom capabilities

standard bridge inspection reports supplemented with the additional photo coverage provided by the RPAS with plans to use photogrammetry software in the future.

2.6 Final Deliverable

This phase is largely dependent on the assessment phase and discussions had with the asset owner. The inspection team will compile all the information from the inspection, including the RPAS data and their own hands-on data. This information will then help supplement the recommendations for future investigations or repairs.

2.7 Workflow Benefits

The benefits to breaking down the RPAS inspection process in the above manner are outlined below;

- Ensuring that RPAS are suitable based on the site-specific constraints of the selected bridge
- Establishing open and clear expectations between the inspector and the RPAS operator
- Identifying all known hazards or potential safety issues prior to conducting the site inspection
- Confirming that the operator has the proper equipment for flying the selected bridge
- Capturing consistent data and images that can be directly compared year-over-year
- Ensuring the final deliverable meets the specific needs of the client
- Providing a clear process that outlines the requirements for others to effectively use RPAS for their bridge inspections.

3 Technical Initiatives and Future Developments

The above workflow assumes a standard approach to bridge inspections using RPAS and a final deliverable from the operators of a folder with all the photos or video. This section will lay out some new developments in the industry that are currently in use, and those expected in the coming years.

3.1 *Photogrammetry*

The process of photogrammetry is creating a 3D model of a structure by stitching together photos. This process is already widely used in the inspection industry but is becoming more prevalent and affordable for asset owners. In terms of the workflow, using photogrammetry would involve uploading the photo coverage to a third-party software company that will process the data and output a 3D model. This model can be accessed on any web browser and allows the user to navigate the areas of the structure that were captured with the RPAS. As the photos are geotagged, the user can see where the different photos were taken that make up the model and click to see the specific photo instantly.

The models can be fully scaled, allowing for measurements to be taken directly on the model. This is especially useful for confirming the area of defects and for establishing quantities for repairs. If multiple years of inspection photos are present, the user can scroll through the years and visually see the progression of deterioration for a certain element.

An inspection that is slated to use photogrammetry will require more photos than a typical inspection, as the photos require 70% to 80% overlap to create the model. Advancements are ongoing to allow the software to create the model from video instead of using the photos. Figure 4 shows an example of a bridge model made from photogrammetry.

3.2 *Deck Cracking*

In discussions with software companies, the ability to perform a detailed crack survey of bridges using RPAS is currently being developed. This will be possible even as the bridge is in use by the public. Deck replacements and repairs are a common maintenance item for almost all bridges. Being able to quickly perform a crack survey and compare the deterioration over time will help to understand the durability of certain deck types and aid owners in life-cycle costing decisions.

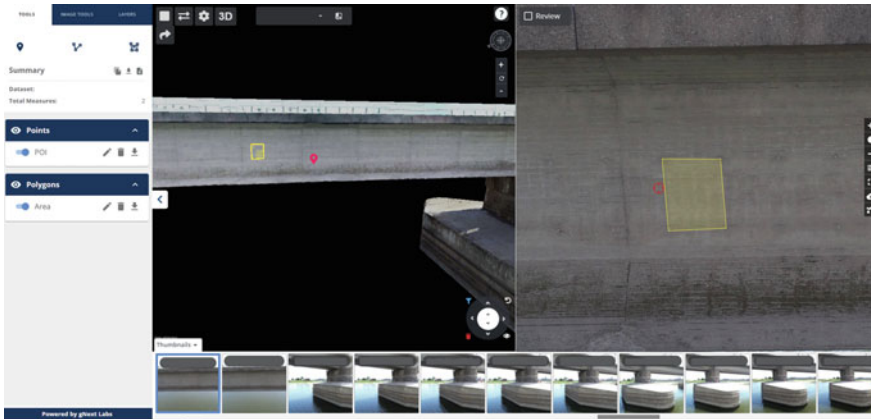


Fig. 4 3D bridge model made from photogrammetry

3.3 Payloads

RPAS systems are becoming highly sophisticated and capable of carrying various payloads to aid with different inspection types. Some units can be installed with infrared cameras, which have a range of uses. In addition, some units are outfitted with NDT tools, such as hammers for concrete testing.

3.4 Autonomous Flights

Preprogramming of flight paths and autonomous flights has potential to help reduce costs and risk for owners and operators. Flight planning software exists that can help operators efficiently plan out their flights around structures and reduce the risk of having the RPAS unit lose connectivity and potentially crash.

4 Conclusions

The benefits that RPAS provide to a standard bridge inspection have been well documented, as seen in the case studies presented. They offer a cost-effective approach to obtaining detailed inspection data with minimal safety hazards. As we expect the use of RPAS to increase moving forward, a structured workflow was needed to evaluate the effectiveness of this tool on new bridge inspection projects. The workflow was broken down based on the typical phases of a bridge inspection project (Assessment, Inspection Planning, Site Investigation, Data Processing and Final Deliverable). Essential tasks were identified for each stage and general notes were included

that should be considered. As well, technological developments in the RPAS and inspection software industries will add more value to our clients and inform how they manage their bridge inventory in the long-term.

Acknowledgements The authors would like to acknowledge the input from Skyreel Aerial Imaging Inc. who were the UAV operators for the case studies presented and gNext Labs who provided the photogrammetry information and model screenshot.

Numerical Investigation of Turbulent Structures and Air Entrainment in Positive Surge Waves



Z. Li and S. Karimpour

1 Introduction

Surge waves form due to sudden changes in flow conditions. Positive surge waves commonly occur in the natural system and hydraulic conveyance structures. Closure of sluice gates in water management and hydropower plants induce an abrupt change in velocity leading to the formation of a pressure wave. Tsunami waves and tidal bores also exhibit the characteristics of flow discontinuity with respect to pressure and velocity profiles. Such phenomena exhibit high turbulence, which contributes to air entrainment, sediment gathering, and induces contaminant and debris transport [14]. Turbulent structures located under the front of the tidal bore may lead to bed erosion [5] and the adjacent community's flooding damage [22]. Bed erosion results in river channel expansion, adverse effect on water safety and habitat environment, and jeopardizes bridges over the river [7]. Like tidal bores, the dam-break waves cause sediment movement and channel bed erosion and can cause swift changes in the bed formation across the channel downstream. Consequently, the flow will be altered, and the estimation of factors such as peak water depth and time to reach the residence area for alert purposes will become difficult [24]. The related studies of dam-break waves date back to [19], who applied equations of motion to the dam-break case. Reference [8] presented laboratory studies for dam-break waves over various roughness of flume beds and verified its outcomes with the theoretical solutions. Although the experimental studies of dam-break waves are getting more attention, the overall number of experimental studies is still limited (e.g., [6, 18, 20]). In recent years, several studies have been carried out to investigate the air entrainment and turbulent

Z. Li (✉) · S. Karimpour
York University, Toronto, Canada
e-mail: zli29@my.yorku.ca

S. Karimpour
e-mail: shooka.karimpour@lassonde.yorku.ca

patterns of moving and stationary surge waves experimentally (e.g., [11, 12]). Most of the existing experimental studies are based on single point measurements and the overall turbulent characteristics cannot be obtained. Reference [15] are among the first to numerically simulate surge waves, where they have presented the turbulent behaviour within the waves.

In the 2000s, the air entrainment nature of the dam-break waves, also known as “white water”, has been focused on. Several experimental studies of dam-break waves were conducted over the stepped channel and using conductivity probes have determined the air entrainment over the vertical axis and bubble dimensions (e.g., [4, 2]). As for the numerical simulation, the widely used volume of fluid (VOF) method for interface capture was proposed by Reference [9]. For aeration near the interface due to the turbulence, Hirt and Souders [10] firstly introduced for FLOW-3D® to calculate the air fraction based on the comparison between “stabilizing forces” and “destabilizing forces” near the interface. More recently, [23] have developed a calibration approach to determine the ideal magnitude for adjustable parameters in the same model. Lubin and Glockner [16] also demonstrated the numerical work of aeration in plunging waves. Advanced numerical methods allow us to address mixing and air entrainment in dam-break waves and surge waves. Furthermore, this allows a detailed description of complex features of flow [10]. Based on the literature review, there are several numerical and experimental approaches to study the turbulent structure and the air entrainment properties of the surge waves individually. However, few studies have connected two aspects at different Froude numbers numerically. Therefore, within the positive surge waves, the research project studies the interconnection between the turbulence and aeration characteristics numerically and improves compliance with the experimental data available.

2 Methodology

2.1 Governing Equations and Numerical Solutions

In order to resolve the turbulent motion at the surge front, Large Eddy Simulation (LES) is employed. The governing equations for LES are obtained by filtering the Navier–Stokes equation:

$$\rho \frac{\partial \bar{u}_i}{\partial t} + \rho \frac{\partial \bar{u}_i \bar{u}_j}{\partial x_j} = - \frac{\partial \bar{p}}{\partial x_i} + \frac{\partial}{\partial x_j} \left(\mu \left(\frac{\partial \bar{u}_i}{\partial x_j} + \frac{\partial \bar{u}_j}{\partial x_i} \right) \right) + \frac{\partial \tau_{ij}^{SGS}}{\partial x_j} \quad (1)$$

where \bar{u}_i and \bar{p} are the filtered velocity and pressure, respectively. Equation 1 is presented in the suffix notation and i or $j = 1$ in this notation corresponds to the x -direction, i or $j = 2$ corresponds to the y -direction, and i or $j = 3$ to the z -direction. The sub-grid scale (SGS) turbulent shear stress is expressed as:

$$\tau_{ij}^{\text{SGS}} = \rho \overline{u_i u_j} - \rho \overline{u_i} \overline{u_j} \quad (2)$$

The SGS turbulence model defines this filtered turbulent shear stress:

$$\tau_{ij}^{\text{SGS}} = -2\mu_t S_{ij} \quad (3)$$

where S_{ij} is the strain rate and defined as:

$$S_{ij} = \frac{1}{2} \left(\frac{\partial \overline{u_i}}{\partial x_j} + \frac{\partial \overline{u_j}}{\partial x_i} \right) \quad (4)$$

The eddy viscosity of the SGS motion is constructed based on the Smagorinsky-Lilly model [21]:

$$\mu_t = \rho (C_s \Delta)^2 \sqrt{2 S_{ij} S_{ij}} \quad (5)$$

where Δ is the filter size of the sub-grid scale model, represented as the size of the mesh, and C_s is the Smagorinsky constant.

The air–water interface model is based on the volume of fluid (VOF) method for identifying the boundary between water and air. In this method, the interface can be determined by introducing the water volume fraction, α_w [1]:

$$\frac{\partial \alpha_w}{\partial t} + \nabla \cdot (\alpha_w \mathbf{u}) + \nabla \cdot [\mathbf{u}_c \alpha_w (1 - \alpha_w)] = 0 \quad (6)$$

In Eq. 6 \mathbf{u}_c represents the “relative velocity” for water and air. $\alpha_w = 1$ demonstrates the cells filled with water, and $\alpha_w = 0$ for the cells filled with air only. The interface, surge front, and in-depth where air entrainment is expected, the water volume fraction is $0 < \alpha_w < 1$.

2.2 Implementation in OpenFOAM

The computational domain is shown in Fig. 1. The gate is located at $x = 0$. The computational domain is extended from $x = -L_{xu}$ upstream to $x = L_{xd}$ downstream of the gate. With the removal of the gate, a surge wave propagates downstream with a celerity of c and depth of d_2 . The initial water depth before the gate is d_0 and the downstream water depth is d_1 . The domain is surrounded by rigid walls on three sides with the upper face open to the atmosphere. The boundary conditions for velocity and pressure are coupled as *no-slip* and *fixed flux pressure*, respectively. Surge Froude number is defined based on the unperturbed water depth and velocity downstream of the surge wave:

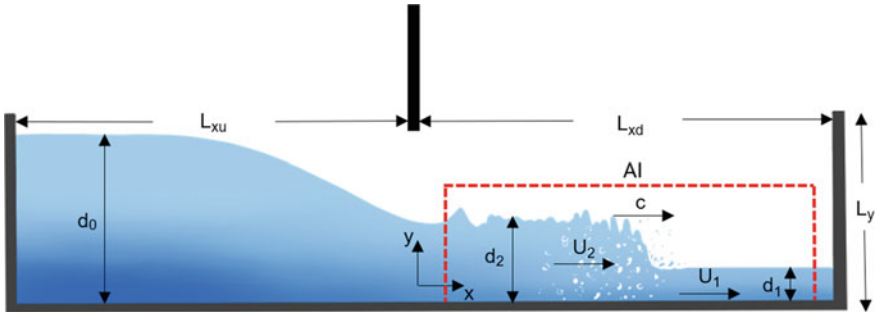


Fig. 1 Sketch of the moving surge wave in the computational domain

$$Fr_1 = \frac{c + U_1}{\sqrt{gd_1}} \tag{7}$$

Here, the initial velocity $U_1 = 0$. References [13] and [25] have reported undular waves at surge Froude numbers up to $Fr_1 \approx 1.5$. This paper, therefore, covers surge Froude numbers beyond this range as it aims to investigate the turbulent properties across the surge breaking front. The initial water depths, d_0 and d_1 , and the domain dimensions, L_{xu} and L_{xd} , are determined based on the Method of Characteristics (MOC). According to the MOC, the d_0 and d_1 ratio dominates the expecting Fr_1 [2]:

$$Fr_1 = \frac{c}{\sqrt{gd_1}} = \frac{0.63545 + 0.3286\left(\frac{d_1}{d_0}\right)^{0.65167}}{0.00251 + \left(\frac{d_1}{d_0}\right)^{0.65167}} \tag{8}$$

Using the theoretical celerity of positive surge (c), obtained from Eq. 8, and the celerity of negative surge ($c_0 = \sqrt{gd_0}$), the length of the computational domain before and after the gate, L_{xu} and L_{xd} , are selected. The uniform square mesh of the size of $\Delta x = \Delta y$ is implemented. Figure 1 illustrates the Area of Interest (AI), where local mesh refinement of the size of $dx = dy$ is implemented. This provided increased resolution around the surge front and behind the surge. Table 1 summarizes the flow conditions for 3 surge Froude numbers $Fr_1 = 1.60, 2.13, \text{ and } 2.49$.

Table 1 Computational domain and flow conditions of the simulated cases

| Reference IDs | Δx (m) | dx (m) | L_{xu} (m) | L_{xd} (m) | L_y (m) | d_1 (m) | d_0 (m) | c (m/s) | c_0 (m/s) | d_2 (m) | Fr_1 |
|---------------|----------------|----------|--------------|--------------|-----------|-----------|-----------|-----------|-------------|-----------|--------|
| 1 | 0.01 | 0.005 | 25 | 25 | 1.2 | 0.34 | 1 | 2.922 | 3.132 | 0.624 | 1.60 |
| 2 | 0.01 | 0.005 | 25 | 24 | 1.2 | 0.20 | 1 | 2.984 | 3.132 | 0.513 | 2.13 |
| 3 | 0.01 | 0.005 | 25 | 25 | 1.2 | 0.15 | 1 | 3.021 | 3.132 | 0.461 | 2.49 |

3 Results and Discussion

3.1 Mixing and Vorticity Profile

The simulations are conducted for 3 surge Froude numbers. VOF provides one velocity for water and air combined. To distinguish between water and air in Fig. 2, the product of vorticity, ζ , and water volume fraction, α_w , is plotted as $\zeta_w = \alpha_w \zeta$. Figure 2a, b and c illustrate the ζ_w profiles at 3 time steps for $Fr_1 = 1.60$, 2.13, and 2.49, respectively. As seen in Fig. 2a, the mixing at $Fr_1 = 1.60$ over the length of the surge is confined to the surface. Behind the surge, the advected vortices induce mixing in depth. However, this in-depth mixing behind the surge is mainly confined to $y > d_1$. For $Fr_1 = 2.49$, on the other hand, the mixing in the surge front induces more in-depth mixing. The surge wave is often characterized by the recirculating flow at the air–water interface and a shear layer that is induced at the surge toe [3]. The deeper reach of vortices behind the surge can be explained by the extension of the mixing cone in the shear layer. At lower surge Froude numbers, the velocity gradient is smaller, which leads to a weaker shear layer. This, in turn, results in limited mixing at the surge front.

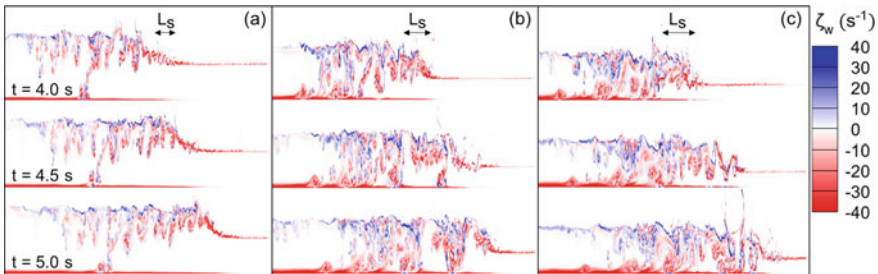


Fig. 2 Contour lines of the ζ_w at $t = 4.0, 4.5,$ and 5.0 s for **a** $Fr_1 = 1.60$; **b** $Fr_1 = 2.13$; and **c** $Fr_1 = 2.49$, with surge length L_s (m) labelled, as estimated in Fig. 5

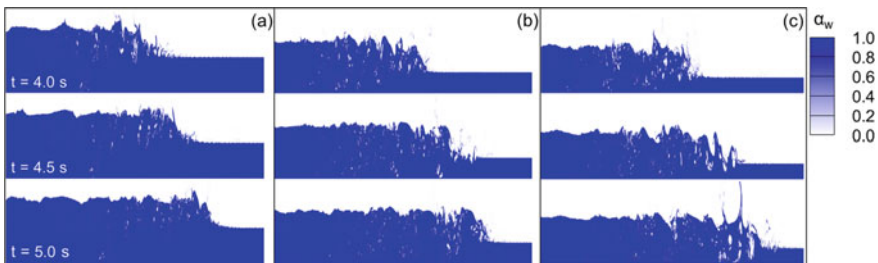


Fig. 3 Contour lines of the α_w at $t = 4.0, 4.5,$ and 5.0 s for **a** $Fr_1 = 1.60$; **b** $Fr_1 = 2.13$; and **c** $Fr_1 = 2.49$

The corresponding air concentration profiles for Fig. 2 are given in Fig. 3. In Fig. 3a, the air entrainment is confined to behind the surge at all three instances. Furthermore, the size of the air pockets (shown in white) for $Fr_1 = 1.60$ is small. On the other hand, as shown in Fig. 3c, air entrainment induced by the coherent structures across the surge front reaches the full depth. The air pockets also entail larger envelopes, visible by large white patches in this figure. This is similar to observations made by Reference [17]. For stationary hydraulic jumps, with Froude numbers ranging from 5.1 to 8.3, they reported that bubble size increases with the Froude number.

3.2 Water Depth Profiles and Perturbation

In order to investigate the water depth and phase change with time, the profiles of surge waves have to be translated. All the obtained profiles at different times, t , are shifted by a length of $L_{lag} = ct$, where c is the theoretical surge wave celerity. This space lag has translated the moving surge wave at different instances to a standing wave, where the approximate positions of the surge front at different instances coincide. Water depth profiles are selected at $\alpha_w = 0.5$.

The average space lag profiles are presented in Fig. 4(a1) and (b1) for $Fr_1 = 1.60$ and 2.49, respectively. The average profile for each Froude number is obtained using the instantaneous profiles at about 60 instances after a fully developed turbulent surge wave was observed. Similarly, the water depth perturbations are estimated based on average and instantaneous profiles. These perturbations are plotted in Fig. 4(a2) and (b2). Water depth perturbation envelopes, marking the minimum and maximum

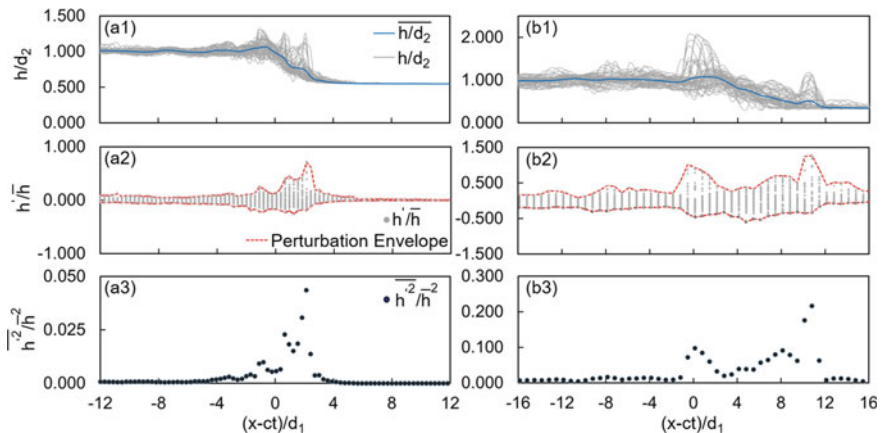


Fig. 4 Plots of normalized parameters (1) h/d_2 , (2) h'/\bar{h} , and (3) h'^2/\bar{h}^2 for (a) $Fr_1 = 1.60$ and (b) $Fr_1 = 2.49$, where h is the instantaneous water depth, \bar{h} is the average water depth, and $h' = h - \bar{h}$ is the water depth perturbation

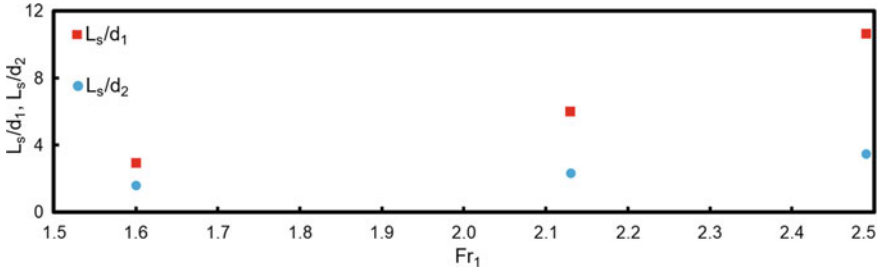


Fig. 5 Plots of L_s/d_1 and L_s/d_2 with $Fr_1 = 1.60, 2.13, 2.49$

levels of water depth perturbations, are also plotted in these figures. In both surge Froude numbers one distinct peak can be identified close to the surge toe. Across the length of the surge, the normal water depth perturbation, h'/\bar{h} , is generally lower. However, the perturbation peaks again around the surge heel. This is clear, specially at higher surge Froude number of $Fr_1 = 2.49$. Figure 4(a3) and (b3) demonstrate the magnitude of water depth perturbation, where the primary peak at the toe and the secondary peak at the heel are visible.

Based on the observation from water depth and magnitude of water depth perturbation plots in Figs. 4(a3) and (b3), surge length, L_s , at different Fr_1 can be estimated. The surge lengths plotted in Fig. 5 are the distance between two peaks observed in the magnitude of water depth perturbation. The estimated L_s/d_1 and L_s/d_2 are plotted against Fr_1 in Fig. 5. As shown in this figure, both normalized surge lengths, L_s/d_1 and L_s/d_2 , increase with surge Froude number, Fr_1 . Along with our observations from Figs. 2 and 3, this indicated that both the mixing length and strength of the recirculating flow near the air–water interface change with the surge Froude number.

3.3 Air Concentration Near Surge Toe and Heel

For air entrainment profiles, we have also applied the space-lag approach. All the profiles are translated with a distance of $L_{lag} = ct$. Fig. 6 illustrates 10, 50, and 90 percentiles of air concentration in depth, as well as the average air concentration profile. Figure 6a and c demonstrate such profiles close to heel and toe, respectively. The depth between the 10 and 90 percentiles in the profile closer to the toe is smaller than the analogous depth for a profile closer to the heel. This trend demonstrates lower air entrainment depth closer to the toe, which is attributed to this profile’s vicinity to the tip of the mixing cone.

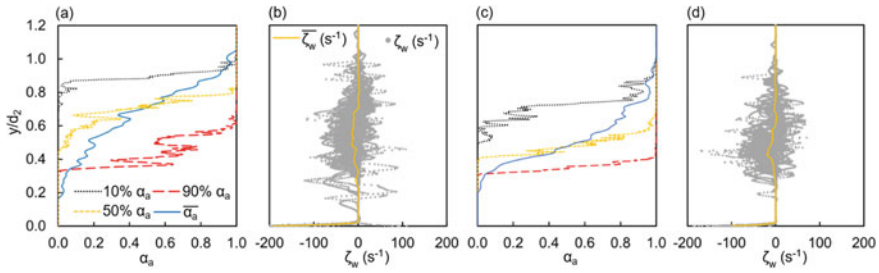


Fig. 6 Plots of y direction distribution of air volume fraction, α_a and ζ_w for $Fr_1 = 2.13$ at **a, b** $x = 12.4$ m (close to surge heel) and **c, d** 12.8 m (close to surge toe) at $t = 4$ s

4 Conclusion

Positive surge waves generated from a sudden opening of the sluice gate are simulated in this study. Various breaking surge Froude numbers in the range of $Fr_1 = 1.60$ – 2.49 are studied. We have employed the combination of Large Eddy Simulation and Volume of Fluid to capture the detailed structures of the turbulent flow and patterns of air entrainment behind the surge wave. At low Froude numbers, the mixing induced by vortices is limited to the vicinity of the air–water interface. These vortices reach a greater depth at higher Froude numbers, leading to higher levels of air entrainment both near the surface and in-depth. Average water depth profiles, perturbation envelope and magnitude, are also generated by shifting and averaging the instantaneous water depth profiles. Two perturbation peaks are observed in the vicinity of the surge heel and surge toe. These peaks are used to define the surge length, which then is plotted against the surge Froude number. Furthermore, for $Fr_1 = 2.13$, a smaller aeration depth is observed at the surge toe compared to the heel. Our numerical simulations shed light on the detailed structures of a transient surge wave. Validation with existing experimental studies and further detailed analysis will be reported in a future publication.

References

1. Almeland SK (2018) Implementation of an air-entrainment Model in InterFoam. Technical Report, Proceedings of CFD with OpenSource Software, Chalmers University of Technology, Gothenburg, Sweden. https://doi.org/10.17196/OS_CFD#YEAR_2018
2. Chanson H (2004) The hydraulics of open channel flow: an introduction, 2nd edn. Elsevier, Burlington, MA, USA
3. Chanson H, Lubin P, Glockner S (2012) Unsteady turbulence in a shock: physical and numerical modelling in tidal bores and hydraulic jumps. Turbulence: theory, types and simulations. Nova Science Publishers, Hauppauge, NY, USA, pp 113–148
4. Chanson H (2003) Two-phase flow characteristics of an unsteady dam break wave flow. In: 30th IAHR congress, Thessaloniki, Greece, vol C2, pp 237–244

5. Chanson H, Tan Y (2018) Particle dispersion under tidal bores: application to sediments and fish eggs. In: 7th international conference on multiphase flow, Tampa, FL, USA, paper No 12.7.3
6. Chen YH, Simon DB (1979) An experimental study of hydraulic and geomorphic changes in an alluvial channel induced by failure of a dam. *Water Resour Res* 19(5):1183–1188
7. Department of Environment and Resource Management of Queensland (2009) What causes streambed erosion? Government Publications, Queensland, Australia
8. Dressler RF (1954) Comparison of theories and experiments for the hydraulic dam-break wave. *Int Assoc Sci Hydrology* 3(38):319–328
9. Hirt CW, Nichols BD (1981) Volume of fluid (VOF) method for the dynamics of free boundaries. *J Comput Phys* 39(1):201–225
10. Hirt CW, Souders DT (2004) Modeling entrainment of air at turbulent free surfaces. In: World water and environmental resources congress, Salt Lake City, Utah, USA
11. Koch C, Chanson H (2009) Turbulence measurements in positive surges and bores. *J Hydraul Res* 47(1):29–40
12. Leng X, Chanson H (2016) Coupling between free-surface fluctuations, velocity fluctuations and turbulent reynolds stresses during the upstream propagation of positive surges, bores and compression waves. *Environ Fluid Mech* 16(4):695–719
13. Leng X, Chanson H (2017) Upstream propagation of surges and bores: free-surface observations. *Coast Eng J* 59(01):1750003
14. Li Y, Chanson H (2018) Sediment motion beneath surges and bores. In: 7th IAHR international symposium on hydraulic structures, Aachen, Germany
15. Lubin P, Glockner S, Chanson H (2010) Large eddy simulation of turbulence generated by a weak breaking tidal bore. *Environ Fluid Mech* 10(5):587–602
16. Lubin P, Glockner S (2015) Numerical simulations of three-dimensional plunging breaking waves: generation and evolution of aerated vortex filaments. *J Fluid Mech* 767:364–393
17. Murzyn F, Chanson H (2007) Free surface, bubbly flow and turbulence measurements in hydraulic jumps. Hydraulic Model Reports, University of Queensland, Brisbane, Australia, CH63/07
18. Nakagawa H, Nakamura S, Ichihashi K (1969) Generation and development of a hydraulic bore due to the breaking of a dam. *Bull Disaster Prev Res Inst* 19(2):1–17
19. Pohle FV (1952) Motion of water due to breaking of a dam and related problems. In: Symposium on gravity waves, NBS Circ, 521
20. Schmidgall T, Strange JN (1961) Floods resulting from suddenly breached dams. Conditions of high resistance. Hydraulic model investigation. Miscellaneous Paper, Army Engineer Waterways Experiment Station, Vicksburg, MS, USA
21. Smagorinsky J (1963) General circulation experiments with the primitive equations: I the basic experiment. *Mon Weather Rev* 91(3):99–164
22. Sulaiman A (2017) Environmental effect of tidal bore propagation in Kampar River. In: MATEC web of conferences, EDP Science, vol 103, p 01015
23. Valero D, García-Bartual R (2016) Calibration of an air entrainment model for CFD spillway applications. *Adv Hydroinformatics* 571–582
24. Zech Y, Soares-Frazão S, Spinewine B, Le Grelle N (2008) Dam-break induced sediment movement: experimental approaches and numerical modelling. *J Hydraul Res* 46(2):176–190
25. Zheng F, Li Y, Xuan G, Li Z, Zhu L (2018) Characteristics of positive surges in a rectangular channel. *Water* 10(10):1473

A Review of the Literature on Design and Performance of Multi Lane Roundabouts in Canada: The Case for Turbo Roundabouts



Juniper Scott and Lovegrove Gordon

1 Introduction

Modern Canadian transportation is supported by a diverse road network; when these roads cross, intersections are needed. Single lane roundabouts (SLRs) have grown to become a standard intersection choice over the last two decades however, due to capacity their application is limited to small volume crossings. Roundabouts are preferred over traditional intersections (traffic lights or stop sign crossings) due to reducing collision and injury frequency over 35% and 75% respectively [3], and greatly increasing traffic flow. Multi lane roundabouts (MLRs) have been implemented for roughly two decades in North America however, their success does not include the sweeping safety benefits of SLRs. As time is lost within transportation systems due to ‘clogging’ of intersections and traffic backup, the need for solutions that offer better safety and traffic flow capacity exist.

Turbo roundabouts (TRs) are a relatively new intersection concept implemented in the Netherlands nearly twenty years ago; these MLRs include raised lane dividers, spiral circulating lanes, often shelter islands for pedestrians, and sometimes raised pedestrian crosswalks. While these roundabouts improve safety over existing MLR’s [20], and capacity over SLRs, the relative impact of specific TR geometry and placement of appurtenances is still being researched. Approach angles influence speeds of entering or exiting vehicles [14], and consequently capacity and safety are functions of these angles. Crossing shelters assist pedestrians and cyclists by shortening road crossing distance [14], while raised crossings lower approaching vehicle speeds

J. Scott (✉) · L. Gordon

School of Engineering, University of British Columbia, 1137 Alumni Avenue, Kelowna, BC V1V1V7, Canada

e-mail: scott.h.juniper@gmail.com

L. Gordon

e-mail: gord.lovegrove@ubc.ca

© Canadian Society for Civil Engineering 2022

S. Walbridge et al. (eds.), *Proceedings of the Canadian Society of Civil Engineering*

Annual Conference 2021, Lecture Notes in Civil Engineering 250,

https://doi.org/10.1007/978-981-19-1065-4_24

without disrupting two-way traffic flow. With rise of active transportation in municipal communities [1], inclusion of appurtenances in design guidelines cannot be overemphasized.

With Turbo Roundabouts only seemingly being actively applied in Europe, and somewhat addressed in the U.S. via the FWHA, there appears to be a knowledge gap for Canada to promote implementation; despite some MLRs that mimic turbo roundabout design, there are no ‘true’ TRs existing in Canada. Due to harsh winters and large snowfall, parts of Canada might create environmental challenges for snow removal. Additionally, there are various studies that somewhat conflict with the safety benefits purported. Finally, Appurtenances that make our intersections functional and desirable parts of the build environment seem to lack definition in existing design aids and research. This paper reviews the background of circular intersections, and illustrates most recent credible research outlining ideal Turbo Roundabout design guidelines in a Canadian context. Existing MLR projects are reviewed to give an idea of the lessons learned and provide insight towards ideal design guidelines. Appurtenance traffic calming and pedestrian aids are reviewed for insight into expected benefits and perceived best practices.

2 Background

This background serves to summarize the adoption of circular intersections from historical to modern transportation system improvements within a North American context.

2.1 *Initial Circular Intersections*

The first circular intersection utilized in North America was commissioned in 1905 in New York City—aptly named the Columbus Circle (Transportation Association of Canada [7]). A few decades followed where traffic circles and rotaries were used as an alternative road improvement. These ‘initial circular intersections’ (ICI) were different than ‘modern roundabouts’ (MR) due to their geometric design, entry/exit requirements, right of way procedures, circulating speed, and pedestrian allowances; particularly dangerous was the requirement for circulating vehicles to yield to entering vehicles, a large internal diameter, and high circulating vehicle speeds. Due to the various differences, ICI’s led to many high-speed collisions and congestion and were eventually phased out as a viable design improvement. During the 1960s the British varied rules of the road to require entering traffic to yield towards circulating traffic on all circular intersections. Around this same time, designs were being implemented using smaller internal diameters and slower circulating speeds. Due to these changes, safety and capacity of ICI’s improved drastically and led to the implementation and design of what are now called Modern Roundabouts.

2.1.1 Rotaries

Rotaries were installed within Canada prior to the 1960s and included large diameter central islands (sometimes larger than 100 m). These transportation improvements were designed to maintain high speeds while moving throughout the circle. Lane changes would be required to enter the circulating ring, contributing to large size. Implementation of these improvements was limited to Eastern Canada and United States. Rotaries are otherwise known as traffic circles in Western Canada [6].

2.1.2 Urban Traffic Circles

Traffic circles are circular intersections in urban settings. Traffic circles were often built around historical monuments or items of significant cultural value. Due to high volume pedestrian crossings at many of these urban sites, traffic signals were often used to control multiple modes of transportation (pedestrian, cycling, automotive, etc.) through the circles.

2.1.3 Neighbourhood Traffic Circles

Neighborhood traffic circles are residential improvements, often constructed within existing four-way intersections. These improvements may include yield signs at the entrance, however, are often minimalistic in design and include little more than a central island that forces one way circulating traffic.

2.2 Modern Roundabouts

Canada's first MR emerged in the 1990s. Since implementation, single lane MR's have been extended to interchanges around highways, intersections on highways, and have been given preference in policy as a first choice for highway intersections in British Columbia to support climate targets [2]. As mentioned, modern roundabouts utilize yielding to circulating traffic. Due to this characteristic, it's a commonly understood roundabouts hold advantages over intersections (especially signalized) in low volume situations [10].

2.2.1 Mini Roundabout

The smallest of modern roundabouts is the mini roundabout. The ICD for this type of roundabout is typically 13.5–27 m [12]. The central island is often traversable, and for four legs the maximum daily capacity is estimated at approximately 15,000 vehicles.

2.2.2 Single Lane Roundabout

The single lane roundabout typically has an inscribed circular diameter (ICD) of 27 m to 54 m, and as the name suggests, only supports movement of one lane of traffic. Depending on level of traffic and pedestrian / cyclist requirements, crossing signals may be used at entrance and exit of these roundabouts. Single lane roundabouts have demonstrated desirable safety metrics over traditional intersection design for severity and frequency of injury [5]; injury rate and collision frequency for SLRs are typically reduced by 73% and 51% respectively [16], in the United States.

2.2.3 Mixed Multi Lane Roundabout

Mixed Multi Lane Roundabouts are, by definition, multi lane roundabouts that allow driver lane changing. Mixed lane roundabouts were implemented to increase capacity over traditional roundabouts, however driver behavior (specifically the proclivity of drivers to change lanes multiple times within the roundabout), has led to a reduction in the efficiency of this intersection; specifically, it has been suggested that the reduced traffic safety of lane changing leads to more accidents, and in turn lowers the efficiency/capacity/robustness of this intersection [17].

2.2.4 Turbo Roundabout

The turbo roundabout was first designed by Dr. Lambertus Fortuijn in 1996 while he was a senior lecturer at Delft University; the design was hypothesized to solve challenges towards multilane roundabout safety performance in the Netherlands. Key features of the TR include raised lane dividers separating inner and outer traffic lanes, spiral circulating roadway from inside to outside, and divergent entry lanes which restrict driver exit choices. The raised lane dividers eliminate lane weaving and reduce conflict points; the spiral roadway ensures that lane restricted vehicles may still access their desired exit point; the divergent entry lanes ensure any lane changes happen well before entering the TR. A depiction of the turbo roundabout and its' key features may be found in Fig. 1. It should be noted that typical turbo roundabout design give allowances for appurtenances that encourage more comfortable road passage and crossing for cyclists and pedestrians; these appurtenances include separated cycle track, crossing refuge, and sometimes raised crossings that encourage auto drivers to reduce speed.

3 Literature Review

The main literature used for this review comes from research papers and manuals produced in Europe; one of the main predecessors of these sources, while also being

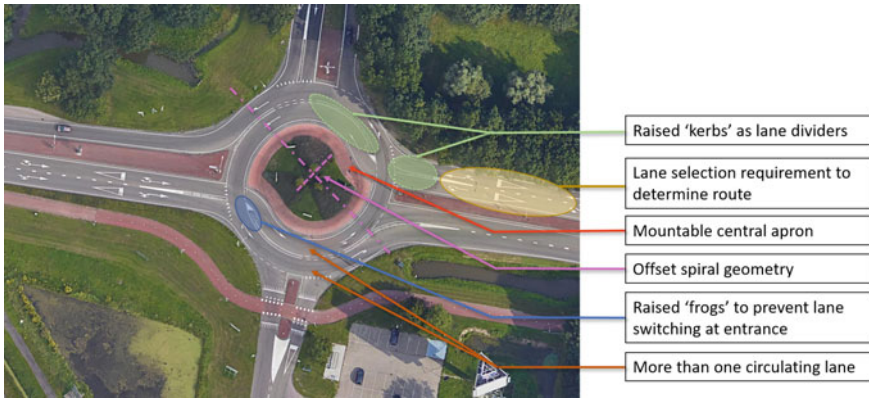


Fig. 1 Turbo roundabout key features. *Source* Google Maps; authors' labels

the most well defined, continues to be the Dutch Roundabout Design Manual. As a principle design source, much of the research created by others either supports or questions the assertions of the manual. This literature review serves to analyze the roundabout design manual by comparing it primarily with peer reviewed papers. For the purposes of analysis, this review will be limited to unsignalized or self regulating turbo and mixed multi lane roundabouts.

The first four subsections of this literature review will discuss features of a turbo roundabout that distinguish it from other similar intersection improvements. The final section will discuss how turbo roundabouts operate, their limitations, and opportunities within a transportation network.

3.1 *Spiral Lane Geometry*

There are seven types of lane geometry prescribed within the manual: basic, egg, knee, spiral, rotor, stretched knee and star [14]. The FHWA defines a turbo roundabout as having no more than two circulating lanes; and prescribed only the basic, egg, knee, spiral, and rotor geometries. The additional roundabout types in the Dutch catalogue (stretched knee and star) allowed for much higher capacity with additional lanes, which is something not recommended by FHWA [15].

Performance was mainly linked to safety statistics and not speed or volume for Dutch applications. The parameter found to be most closely correlated towards safety/speed was the radii of the inner curve and inner lane (R1); the ideal radii was found to be 12 m. The inside spiral R1 may be seen on the below figure and found highlighted in the below table. The below figure (turbo block detail) illustrates typical design geometry; four radii are used in laying out any two lane turbo roundabout geometry, and up to six radii are used if a three lane turbo roundabout is desirable. The radii are offset from center along a translational axis approximately

half a lane distance. Using a smaller radii roundabout (as found in the turbo roundabout) ensures that vehicles will maintain a reasonable speed throughout the circulating lane; other research has confirmed that lower vehicle speeds and drastically reduce the potential for serious injuries, however in 2012 some research lended credit to the roundabout design manual by suggesting 40 km/h as the optimal design speed for turbo roundabouts [8] (Fig. 2).

Czech researchers examining implementation of turbo roundabouts found the optimal lane width for turbo roundabouts was between 4.5 and 5.8 m in width [19]. This width coincides with widths pronounced in the Dutch Roundabout Design Manual, however this varied lane width allowed some flexibility to ensure that lanes were not too wide to encourage excessive speed, nor too narrow to encourage oversize vehicle lane encroachment.

Perhaps one of the most easily understood and well documented arguments for use of a turbo roundabout is from a conflict assessment approach; utilizing turbo roundabouts in place of multi lane roundabouts drastically reduces the amount of potential conflict points within an intersection design, as shown in Fig. 3.

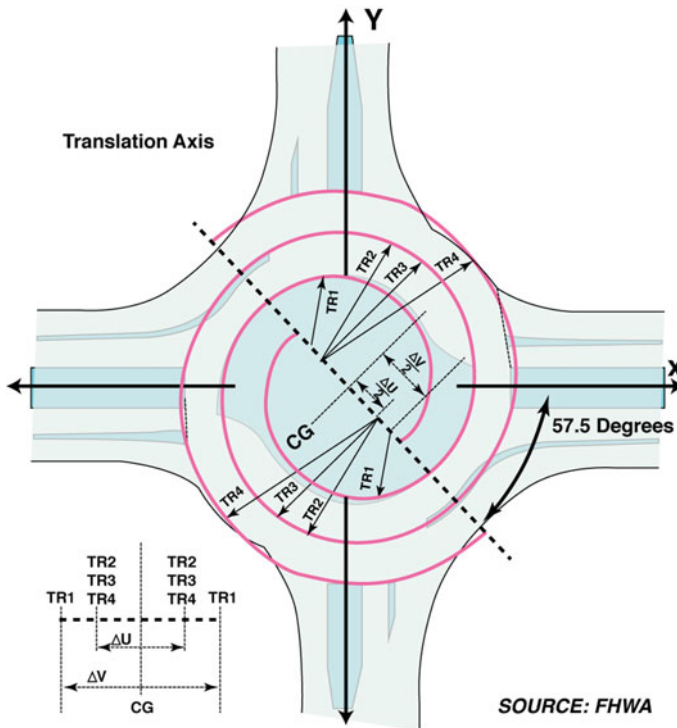


Fig. 2 Turbo block detail. Adapted from [13]

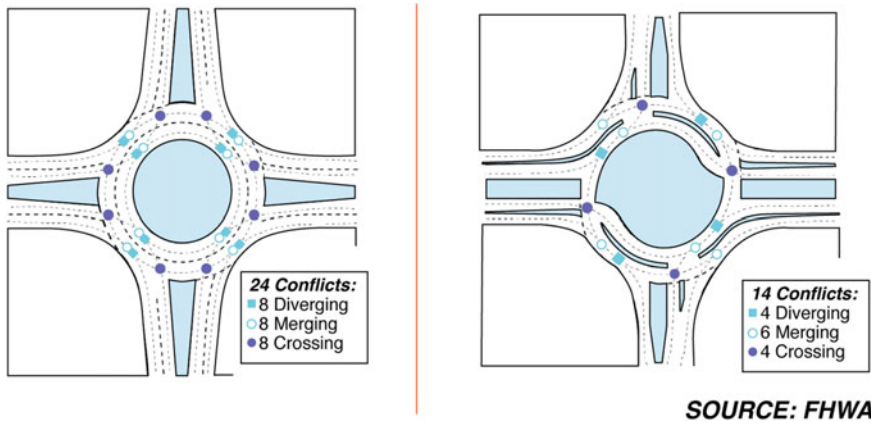


Fig. 3 Comparison of multi lane and turbo roundabout conflict points. Adapted from [13]

3.2 Physical Lane Dividers

Lane dividers are a necessary component of turbo roundabouts to prevent lane weaving, reduce speed of circulating vehicles, and reduce fear of being cut off while travelling in other lanes [14]. A large ‘Frog’ is placed near the start of the inner travel lane to allow higher visibility of the lane divider, and encourage entering the appropriate lane. Reflectors are advised and may be placed either on top of the lane divider, or on slopes. The original lane dividers had a ‘soft’ raised curb of approximately 1 inch, and then a sloped raise of another 2 more inches to the total height of barrier; the soft curb allowed vehicles to pass over if required however was a significant deterrent. To allow snowplowing the overall height of the divider was unchanged, but the soft curbs were removed, as shown in the below diagram. The overall width of divider in both diagrams is 1 foot wide, and a foundation or footing was suggested as a concrete structure embedded in the roadway (Fig. 4).

In 2015, a Polish research paper investigated the effect of raised lane dividers on reduction of certain types of collisions, as compared to multilane roundabouts with no raised lane dividers. At that time Poland had constructed both turbo roundabouts with and without physical lane dividers; the type and severity of collisions that occurred at the turbo roundabout with a physical lane divider were strikingly different; the collisions resulted in approximately 20% fewer serious injuries, and less than half the amount of vehicle side impacts when physical dividers were employed. The author concluded there was little difference between turbo roundabouts without a lane divider and a standard multi lane roundabout from a safety perspective [11], and suggested that multilane roundabouts without raised lane dividers are undesirable (Fig. 5).

Safety benefits of Turbo roundabouts with a raised lane divider cannot be overstated. Another group of Polish researchers analyzed collision data of nine multilane roundabouts (five of which were turbo roundabouts) over four years to prescribe

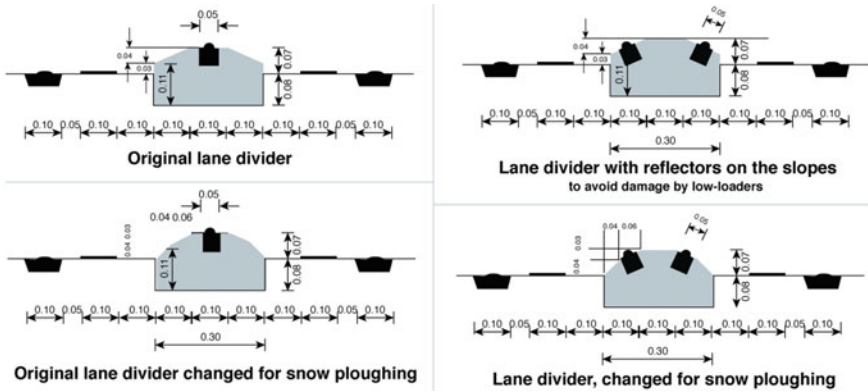
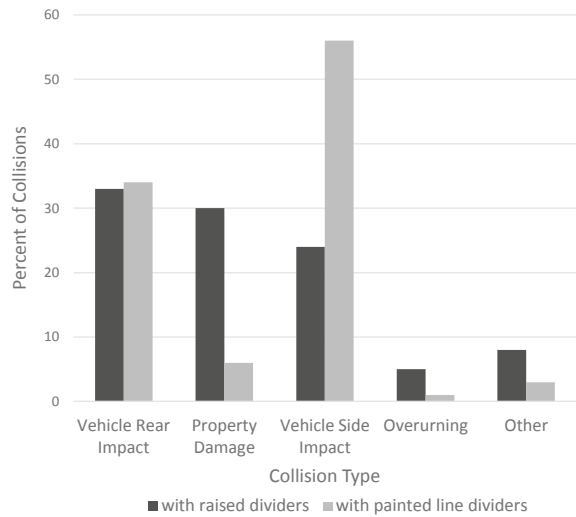


Fig. 4 Physical lane dividers. Adapted from [14]

Fig. 5 Collisions type comparisons between turbo roundabouts with and without raised dividers [11]



predictive models of safety performance factor. The data and modelling suggested a 90 percent confidence that with given traffic volumes, a raised lane divider would result in up to 60 percent less collisions at a turbo roundabout site [9], most definitely in US applications where higher approach speeds are common. According to this team’s research, the likelihood of drivers to follow a swept path was highly influenced by the presence of lane dividers; up to 40% of drivers using multilane and turbo roundabouts violated the selected lanes and changed lanes within the circulating lane. Having lane dividers reduces the potential conflict points significantly for circulating vehicles, and ensures that smaller radii geometry is followed, which induces lower speed and risk towards vehicles within the turbo roundabout.

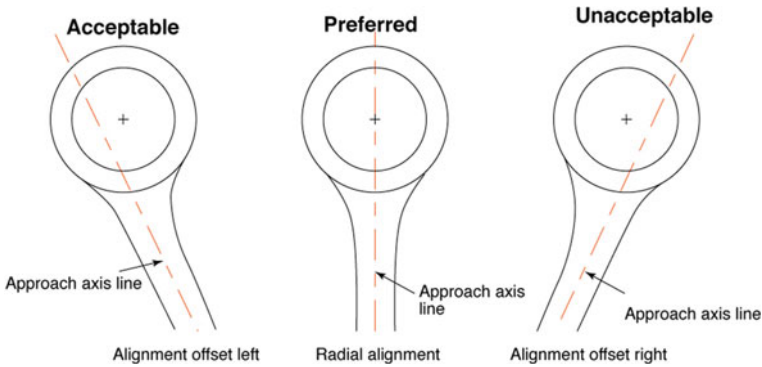


Fig. 6 Approach axis line [14]

Further, Italian research demonstrated that the physical lane dividers guarantee a reduced running speed throughout the circulating lane, by enforcing compliance to the smaller radii geometry [8]. This research suggested that physical lane dividers should be used in all circumstances when considering turbo roundabouts in an urban context.

3.3 Approach Lane Vectors

Alignment of the approach lane is a critical determinant of incoming speed. It is desirable to have drivers reduce speed when entering roundabouts for safety reasons. Due to this reason it's allowable to have offsets to the left of the center of the roundabout, but not to the right. Having an offset to the left is only advised if there are no cyclists and few pedestrians crossing the roadway, as higher vehicle exiting speed would be expected [14]. Ensuring vehicles approach roundabouts nearly perpendicular encourages reduction of speed due to the radius of curvature to enter the circulating lane; without a smaller radius curve to enter the lane, reduced speed (and therefore safety benefits) cannot be guaranteed (Fig. 6).

3.4 Pedestrian Crossings

Brilon summarized lessons learned with all types of roundabouts in Germany in 2011 at the TRB roundabout conference; particularly useful were the findings about effective placement of appurtenances. It was suggested that the only significant risk at compact roundabouts was with their connection to cyclists; Brilon suggested that cycle track crossings of entry and exit lanes should be placed at least 5 m from the circulating road, and might only be necessary when traffic carried at the roundabout

exceeds 15,000 vehicles per day. While statistical information wasn't presented to support some of the lessons learned, it was suggested that cyclists suffer from poor visibility when their crossing distance is too close to the circulating lane [4]. Conclusions of this paper suggested that signalized intersection is still a well accepted solution for traffic amounts above 40,000 vehicles per day.

Brilon's findings were supported in a British case trial study where different cycle track and crossing schemes were tested to analyze how roundabouts affected the safety of cyclists and pedestrians and their proclivity towards active transportation in a network where roundabouts were constructed. Overall, the cyclists and pedestrians were supportive of the roundabout improvements towards an active transportation, and the designs most successful included segregation of all three modes, and separation from the circulating traffic ring [22].

It should be noted that pedestrian crossing locations and inclusion of appurtenances likely has significant effects on traffic capacity in urban context, if at grade crossings are included. Italian researchers estimating capacity of turbo roundabouts in urban context had contentious findings for capacity that did not agree with previous research and may warrant further investigation [8].

3.5 Placement Within Network and Capacity

Researchers in Spain analyzed the safety and capacity of Turbo Roundabouts using gap acceptance theory along with other complex capacity formulation techniques. Without getting into too much detail about this research, the findings indicated that turbo roundabouts had limited application in high capacity circumstances. The researchers concluded that previous authors used much too simplistic approaches for comparison of turbo roundabouts to mixed multi lane roundabouts (specifically for lane allocation and saturation); the circumstances where turbo roundabouts perform best are for locations where turns in the minor traffic flow direction are above sixty percent [18]. This point was further examined and agreed upon that in order for turbo roundabout capacity to exceed multi lane roundabouts, more than sixty percent of vehicles must be turning right [21].

Further research by Silva and others analyzed performance of turbo roundabouts compared to multilane roundabouts with microsimulation and real case study in corridor applications. The corridor analyzed had three roundabouts spaced at 400 and 470 m on a two lane road; the research was aimed at analyzing capacity of these turbo roundabouts at saturated or near saturated conditions. The findings by this research suggest capacity performance degrades rapidly as turbo roundabouts exceed capacity [17], as opposed to conventional multilane roundabouts; it appears the multilane roundabouts were less susceptible to clogging at saturated conditions.

4 Discussion

One of the main benefits recognized from numerous sources within the literature review was the safety benefits of turbo roundabouts (and roundabouts in general). Added safety via reduced frequency and severity of collisions has been attributed to radial curvature imposing a reduction of speed to under 40 km/h. Low radius curves require drivers to slow down to comfortably move throughout the circulating lanes. As shown in prior research, vehicles will take the fastest path through multi lane roundabouts where lane compliance is not enforced via raised lane dividers, making illegal movements commonplace throughout multilane roundabouts—thus undermining their purpose of improving road safety. Case studies have shown collision frequency reduction of at least 60% when converting a priority intersection to a turbo roundabout [19], however this amount varies based on country of implementation (expected speeds and rules of road vary).

One item differing between research papers was an agreed upon capacity formula, especially in urban context where pedestrian crossings are expected at grade. The method of calculating capacity with only vehicles wasn't necessarily agreed upon, and certainly the location of appurtenances and their affect on intersection capacity is not very well understood. It was however recognized that pedestrian and cyclist crossings should be kept back from these intersections; the appropriate distance from the circulating lane to place pedestrian crossings was not found in this literature review.

5 Conclusion

Turbo roundabouts seem to operate consistently, but not exclusively, at high capacity in locations with three intersecting legs. These roundabouts have the potential to replace many signalized intersections in three and four leg scenarios in urban applications however the likelihood of success is dependant on accurate modelling and understanding of how pedestrian flows affect the safety and function of the intersection itself. It seems reasonable that further research into modelling of the different types of these roundabouts in urban circumstances would be beneficial, so an analytical tool may be developed to adequately assess different locations for suitability of turbo roundabouts as a first choice.

One of the underlying program tenets for the Dutch roundabout design manual (referenced frequently in this paper) is a vision zero goal. The vision zero goal for transportation, as the name implies, is a goal of having zero fatal collisions; the goal places the onus on adequate design to reduce risks for road users. North American design manuals as recent as “Guidelines for the Planning and Design of Roundabouts” [12] seem to encourage multilane roundabout designs that focus on vehicles maintaining their directed path when entering roundabouts, with no physical lane controls. Research suggests up to 40% of road users ignore lane markings when

using multi lane roundabouts for the fastest path; it seems redundant to design multi lane roundabouts for circumstances which aren't, in fact, observed or realistic. One conclusion is that opportunity for driver error isn't necessarily viewed as a designer's responsibility, and perhaps that is a fault of the road culture of North America. If safety is to be encouraged from a design perspective, it may be worth reviewing the program foundations supporting it.

Finally, it is apparent from the research reviewed that multi lane roundabouts are not desirable improvements from a safety perspective without raised lane dividers.

Acknowledgements This literature review was made possible through the financial support of the University of British Columbia, and the oversight of the UBC School of Engineering.

References

1. Assunção-Denis M-E, Tomalty R (2019) Increasing cycling for transportation in Canadian communities: understanding what works. *Transp Res Part A: Policy Pract* 123:288–304. <https://doi.org/10.1016/j.tra.2018.11.010>
2. BCMoTI (2019) Supplement to TAC geometric design guide. BCMoTi, Kamloops, BC 2019. <https://www2.gov.bc.ca/assets/gov/driving-and-transportation/transportation-infrastructure/engineering-standards-and-guidelines/highway-design-and-survey/tac/tac-2019-supplement/bctac2019-chapter-0700.pdf>
3. Board, transportation research, and engineering national academies of sciences and medicine (2010) Roundabouts: an informational guide second edition. The National Academies Press, Washington, DC. <https://doi.org/10.17226/22914>
4. Brilon W (2011) Studies on roundabouts in germany: lessons learned. In: TRB roundabout conference, 15
5. Brown M (1995) The design of roundabouts. State-of-the-art review/transport research laboratory, department of transport. HMSO, London. file://catalog.hathitrust.org/Record/101655424
6. Canadian Institute of Transportation Engineers, Technical Liason Committee (2013) Roundabouts in Canada: a primer for decision makers. Minnow Lake. [http://www.minnowlake.ca/Roundabouts_in_Canada_2013\(1\).pdf](http://www.minnowlake.ca/Roundabouts_in_Canada_2013(1).pdf)
7. Chartier G (Paradigm Transportation Solutions Ltd.), Eichenbaum T (Hatch Infrastructure), Jacobson M (Watt Consulting Group), MacDonald M (Harbourside Transportation Consultants), Ridling C (Mott MacDonald), Rodegerdts L (Kitt, M. (Watt C. G. (2017). Canadian Roundabout Design Guide. Transportation Association of Canada. <https://www.tac-atc.ca/en/publications/ptm-crdg-e>
8. Corriere F, Guerrieri M (2012) Performance analysis of basic turbo-roundabouts in urban context. *Procedia Soc Behav Sci* 53: 622–32. <https://doi.org/10.1016/j.sbspro.2012.09.912>
9. Kieć M, Ambros J, Bąk R, Gogoliń O (2019) Evaluation of safety effect of turbo-roundabout lane dividers using floating car data and video observation. *Accid Anal Prev* 125:302–10. <https://doi.org/10.1016/j.aap.2018.05.009>
10. Lakouari N, Ez-Zahraouy H, Benyoussef A (2014) Traffic flow behavior at a single lane roundabout as compared to traffic circle. *Phys Lett A* 378(43):3169–76. <https://doi.org/10.1016/j.physleta.2014.09.001>
11. Macioszek E (2015) The road safety at turbo roundabouts in Poland. *Arch Transp* 33(1). <https://doi.org/10.5604/08669546.1160927>
12. MassDOT, and Kittleston & Associates Inc. (2020) Guidelines for the planning and design of roundabouts. Edited by Septe 2020. MassDOT, Boston. <https://www.mass.gov/doc/massdot-guidelines-for-the-planning-and-design-of-roundabouts/download>

13. Office of Safety FHWA (2019) Turbo roundabouts: informational primer. Washington, DC. <https://safety.fhwa.dot.gov/intersection/roundabouts/fhwasa20019.pdf>
14. Overkamp DP, van der Wijk W, Haskoning R (2009) Roundabouts—application and design a practical manual
15. Porter R, Gooch J, Peach K, Chestnutt C, Moore B, Broeren P, Tigelaar J (2019) Advancing turbo roundabouts in the United States: synthesis report. Washington, DC
16. Robinson BW, Rodegerdts LA, Scarborough W, Kittelson W, Troutbeck RJ, Brilon W, Bondzio L et al (2000) Roundabouts: an informational guide. Fed Highw Adm. file://catalog.hathitrust.org/Record/005529942
17. Silva AB, Mariano P, Silva JP (2015) Performance assessment of turbo-roundabouts in corridors. *Transp Res Procedia* 10:124–33. <https://doi.org/10.1016/j.trpro.2015.09.062>
18. Silva AB, Vasconcelos L, Santos S (2014) Moving from conventional roundabouts to turbo-roundabouts. *Procedia Soc Behav Sci* 111: 137–46. <https://doi.org/10.1016/j.sbspro.2014.01.046>
19. Skvain V, Petrů J, Krivda V (2017) Turbo—roundabouts and their basic evaluation at realized constructions in Czech Republic. *Procedia Eng* 190(December):283–290. <https://doi.org/10.1016/j.proeng.2017.05.339>
20. Tollazzi T, Renčelj M (2014) Comparative analyse of the two new alternative types of roundabouts—turbo and flower roundabout. *Baltic J Road Bridge Eng* 9(3):164–170. <https://doi.org/10.3846/bjrbe.2014.21>
21. Vasconcelos L, Silva AB, Seco ÁM, Fernandes P, Coelho MC (2014) Turboroundabouts: multicriterion assessment of intersection capacity, safety, and emissions. *Transp Res Rec* 2402(1):28–37. <https://doi.org/10.3141/2402-04>
22. Yor I, Helman S, Vermaat P (2015) Off street trials of a dutch style roundabout. London. https://www.cycling-embassy.org.uk/sites/cycling-embassy.org.uk/files/documents/ppr751_dutch_roundabout_safety_v1.pdf

Smartphone: A Source for Transit Service Planning and Management Using Wifi Sensor Data



Muhammad Arslan Asim, Lina Kattan, and S. C. Wirasinghe

1 Introduction

Complete information on passenger activity in a transit system is important for transit planning, operation, demand management and service improvement. Estimated or actual ridership information is a basic requirement for both the design of new transit services and the modification of existing transit services. Fleet size, schedule or headways and duration of service are designed based on passenger demand along with other variables. Passenger demand and other variables like passenger arrival rate, in-vehicle and at-station crowding levels, and dwell time are periodically or continuously monitored. This allows transit services to meet seasonal and daily fluctuations in demand to provide a certain level of service to transit customers.

A transit passenger's activity tracking data can include information such as the time a passenger entered into a transit station, the path taken to access a platform, waiting location and time, the car or door of a transit vehicle used to board, the stations passed through to reach the destination, the time of alighting, the path taken to exit the station or access a connecting route, total distance travelled and total time spent in the transit network [8, 12].

Transit service operators use various types of data collection systems in tandem to monitor the above transit passenger activity information. These data collection

M. A. Asim (✉)

The City of Calgary, University of Calgary, Calgary, AB, Canada
e-mail: muhammad.asim@calgary.ca

L. Kattan · S. C. Wirasinghe

The University of Calgary, Calgary, AB, Canada
e-mail: lkattan@ucalgary.ca

S. C. Wirasinghe

e-mail: chan.wirasinghe@ucalgary.ca

© Canadian Society for Civil Engineering 2022

S. Walbridge et al. (eds.), *Proceedings of the Canadian Society of Civil Engineering Annual Conference 2021*, Lecture Notes in Civil Engineering 250,
https://doi.org/10.1007/978-981-19-1065-4_25

systems are Closed Captioning Television (CCTV) cameras, manual counts, Automated Passenger Counts (APC), turn gates, automatic vehicle locations and Automated Fare Collection (AFC) systems that include smart cards and mobile ticketing. None of these systems can gather complete transit passenger activity on its own; each of these systems gathers one or more elements of transit passenger activity data.

Turn gates and manual counts are the traditional methods used for estimating transit ridership at stations and on transit vehicles. Although these systems are still used by many transit agencies, they are resource-intensive and time-consuming [9, 14]. APC systems are considered as an industry standard and are widely used on transit vehicles for collecting continuous ridership (boarding, alighting and bus loads) data [9]. AFC systems are being continually adopted by transit agencies around the world and can be used to gather ridership information, although this is not the primary purpose of these systems. All these systems are expensive, require extensive capital investments [9] and maintenance budgets, yet offer limited data.

Advancements in information and communication technology and big data analytics offer the opportunity to estimate transportation travel data more accurately and cost-effectively. The increasing number of passengers that carry smartphones coupled with the wide availability of free WiFi at transit facilities have made it possible to gather information on transit passenger activity using WiFi sensing devices. WiFi sensing devices can identify the unique Media Access Control (MAC) address of every smartphone that comes within its specific detection range, without associating this address to any personal information like the device owner's name or cell number [1].

Additionally, other useful data can be reported, including the date and time, RSSI (Received Signal Strength Indicators), duration of contact and location coordinates of the device. The data collected through the WiFi sensing technology has the potential to automatically and continually infer important information on passengers' activity with relatively low investment and maintenance cost. WiFi sensors can be installed at strategic locations in transit facilities to detect transit passengers' smartphone data. This data can then be processed to gather ridership information, inter and intra stations origin-destination, wait time and crowding level. This data can also be used to estimate key states of a transit system (dwell time, travel time, on-time performance etc.) [12] and possibly to make real-time decisions in case of emergencies.

The objective of this paper is to demonstrate that WiFi sensors are an efficient, accurate and cost-effective technology that transit agencies can use to collect passenger activity data. Data collected from WiFi sensors can be noisy and can capture the MAC addresses of both passengers and non-passengers. Advanced filtering techniques are thus needed to remove noisy measurements. This paper also describes a pilot project being initiated at the City of Calgary, AB by deploying commercially available WiFi sensing devices installed at seven LRT stations to continually gather passengers' data for six months.

In the following sections of this paper, the background and relevant literature are first described. WiFi sensors and their functionality (how they collect the cell phone information, sensors coverage etc.) are then presented, along with a discussion of how this data can be stored, processed/parsed and analyzed to determine/estimate various

parameters related to transit passengers' activity and transit operational analysis. Finally, the preparedness of Calgary Transit is described followed by conclusions, recommendations and suggestions for future work.

2 Literature Review

Advancements in the field of information and communication technology, wireless connectivity in public transportation service and increases in the use of mobile devices can significantly impact travellers' behaviour. Mobile devices allow travellers to spend their travel time doing useful activities during what could otherwise be wasted time. Dong et al. [3] conducted a study in 2012 in California to examine the impact of providing free WiFi onboard transit vehicles. The study found that survey participants would make up to 8.6% more trips if free WiFi was provided on train, with young and new riders making the highest percentage preferring free WiFi. Market penetration of cell phones (have WiFi and/or Bluetooth) is more than 80% worldwide [13]. As of July 2020, the cell phone penetration rate in Canada was 79% [13].

The smartphone data collected through WiFi has the potential to quantify transit parameters such as ridership (boarding, alighting and bus loads), passenger arrival and departure rates at stations, passengers' origins destinations, routes, wait times travel times and frequency of transit use [4]. Transit agencies can also use the same WiFi dataset to monitor crowding levels at transit facilities. In case of transit vehicle overloads, transit service disruptions and other emergencies they can then deploy appropriate mitigation strategies such as, dispatching additional transit vehicles based on real-time demand [5].

Smartphones can be detected by various mediums, like cellular towers, Bluetooth and WiFi devices. Cellular mode is almost always turned on in smartphones, making them easily detectable by cellular towers. However, the data collected from cellular towers are not accessible due to low coverage from different private operators. WiFi and Bluetooth have good local coverage. WiFi detection rate is significantly higher than Bluetooth due to higher connection speed, range and level of access to the internet (i.e. smartphone users almost always keep the WiFi mode turned-on) [1, 8, 9].

WiFi sensors are commonly used by retail businesses to gather information from customers' smartphones and produce statistics such as customers visit and physical and temporal loyalty patterns. This information is useful for business monitoring and improvement [11]. Some airport authorities have installed WiFi sensors to track passengers' time spent in various activities (check-in, security, waiting in lobby, boarding) from the terminal entrance until aircraft boarding. These data sets are being used for monitoring and for devising new strategies for increased operational efficiencies in airport terminals [7, 12].

Transit agencies have realized the importance of internet availability for transit customers and have adopted the trend of providing complimentary WiFi at transit facilities and on vehicles. Complimentary WiFi has become one of the parameters

for measuring transit system's attractiveness (Level of service). However, the use of smartphone data for monitoring transit passengers' activities is still under research and has not been deployed in the field yet. Ongoing research has tested the feasibility and effectiveness of the WiFi technology for monitoring transit passengers' activity. Transit agencies that conducted pilot studies to infer WiFi-based transit passengers' information are mainly in the US and include New York City, Charlottesville, Virginia and Santa Catarina, Brazil.

Paradedda et al. [8] developed a model to estimate bus ridership and origin-destination using smartphone data collected by WiFi sensors, at three stops of a local bus route in Santa Catarina, Brazil. They validated this estimation with the counts collected by a video camera and found that the WiFi sensors were 86% successful at detecting smartphones in the first 40 s. They also noted that the estimated passenger counts were always less than the actual passenger counts at the aggregate level. This could be because the WiFi mode may have been turned off for some devices. The number of device detections and their consecutive detection, signal strength (RSSI), the elapsed time between the origin and destination stops are used as bases to identify the customers making a trip between that origin-destination pair.

EI-Tawab et al. [4] tested a low-cost WiFi data-based methodology for bus service improvements by tracking passenger smartphones at two stops of a local bus route on a university campus in Virginia, USA. Their main objective was to determine the number of passengers and their waiting time. The data collected consisted of timestamps, MAC addresses, and RSSI. Using the minimum and maximum waiting times for a specific device in a bus station, they estimated the duration of the passengers' waiting time at the bus stops. The arrival of the bus was used to filter the smartphones which were present at the bus stop. They did not provide data validation results.

Pu et al. [9] used the combination of WiFi and Bluetooth data on three local bus routes in Seattle, USA to estimate transit ridership flow and origin-destination pairs in real-time. Data collection devices were installed on busses. The collected data was transferred from the vehicles to a remote management center for analysis. The researchers proposed a process to broadcast determined passenger load and transit operational status to the transit customers in real-time. To distinguish passengers from non-passengers a fuzzy clustering mean algorithm was developed. The experimental results indicated that their proposed algorithm can highly improve the estimation accuracy. They found that mean values of the detection times, detection duration, average and maximum RSSI of passengers' MAC addresses were much larger than those of non-passenger's MAC address.

Ryu et al. [10] conducted a small-scale proof-of-concept using WiFi sensing technology to determine transit origin-destination and passenger wait time at bus stops of a local route along a corridor in Charlottesville, Virginia, US. They validated their system using video data and found 91% accuracy in OD estimation and 7 s error in travelers' wait time. They used WiFi signal strength (RSSI) to identify individuals along the bus route.

Chon et al. [2] studied the feasibility of using WiFi sensors for tracking pedestrian mobility activities at a large scale. They recruited 25 Korean university students to act as mobile WiFi sensors by holding a smartphone while on the university campus. The

smartphones were installed with special applications to detect smartphones within its WiFi range. They used consecutive detection of the same device within a certain period and signal strength of a certain threshold to filter and detect pedestrians around the mobile WiFi sensors. These pedestrian detections were used to monitor crowding levels at various places at different times of day. Their detection technique was 91% accurate.

Shlayan et al. [12] proposed the use of Bluetooth and WiFi to collect transit passengers' cell phone data for estimation of time-dependent origin-destination within a transit facility and wait time at Metropolitan Transit Authority (MTA) bus and subway stations, New York City, USA. They placed more than one sensor at a station to track movement patterns within the station. They found that sensor placement needs to be site-specific for achieving accurate results.

WiFi sensing technology has the potential to collect complete transit passengers' activity information. Previous research has focused on collecting one or two variables related to transit passengers' activity at bus stops and short bus routes. A few researchers have applied WiFi sensing to transit systems at a larger level, like LRT system to automatically collect passenger activity data.

3 WiFi Sensors and Functionality

When the WiFi mode is turned on in a smartphone, the device frequently emits packets of information to search for an available internet connection. WiFi sensing devices, usually referred to as *Access Points (APs)*, detect these packets of information, record the identity of the device and then record the device as connected to their network [8]. Every smartphone has a unique MAC identification when it tries to search for an AP. Therefore, every smartphone can be uniquely identified within a WiFi detection zone using MAC identification. The detection range of WiFi sensors varies between brands but is usually around 200ft. Other than MAC address, the date, time and receiver signal strength of a smartphone are also recorded by an AP. APs continuously search for packets of information coming from different smartphones. If the same device is detected again the same set of information for that particular smartphone is recorded again.

RSSI is the minimum signal strength (after all possible losses) at which data from a source can be received at a specific rate. RSSI scale can be from 0, being best to -120 dbm, being worst. Every smartphone sensed by an AP has a different RSSI value. The RSSI of a smartphone device varies depending on the distance from an AP, phone brand and model [8]. Additionally, obstructions such as human bodies, windows and walls, and high voltage and high frequency devices can cause interference and weaken the RSSI strength. According to previous studies RSSI of more than -80 dbm is required for a smartphone to be detected by an AP and transmit information in transit facilities. More advanced modems/routers can be used to improve signal strength (RSSI) for an area.

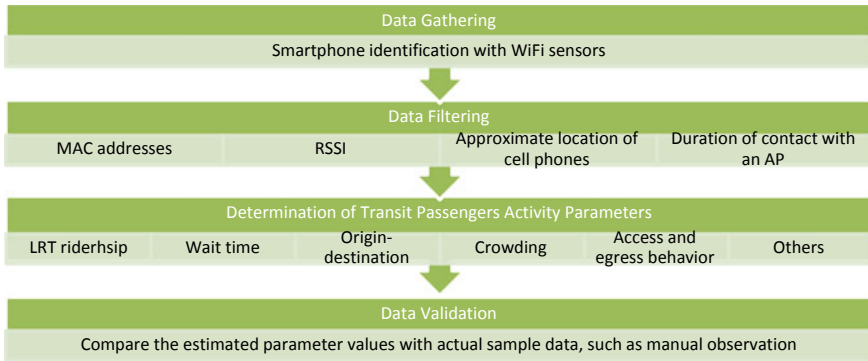


Fig. 1 Steps involved in the determination of transit passengers’ activity parameters using their smartphone IDs

Discovery time is the time that a smartphone takes to be detected by an AP. Discovery time depends on the technology of a smartphone. The newest smartphones connect with an AP virtually instantly as they enter a WiFi zone of an AP. Older smartphones take up to 10 seconds to be detected with the screen turned on and up to a minute with the screen turned off. Based on data from 11 mobile WiFi sensors, [2] found that 82% of the smartphones scan APs every minute and 90% scan every 130 seconds [2]. Three or more WiFi sensors are required to find the exact location of a smartphone with a facility [11]. The position of APs is also important in defining the sensing range. WiFi sensors’ antennae are positioned in a way to limit the sensing range to the transit facility [1]. This makes it easier to distinguish transit passengers’ devices from other devices.

APs usually have built-in storage, but this storage is not normally enough to store large amount of data. Therefore, a database server should be set up with which APs connect wirelessly through WiFi to send information in real-time for storage. In the database information of each smartphone is recorded by time and location. Average data push rates of APs can be adjusted. These rates may range from few seconds to few minutes. The size of each packet of information that an AP transmits to the database server depends on the traffic in the AP’s detection zone. Figure 1 shows steps involved in the determination of passengers’ activity information. The following sections describe these steps in more detail.

3.1 Installation of WiFi Sensors, Data Collection and Testing

Depending on the specific objectives, WiFi sensors are installed at strategic locations in transit facilities such as entrances to stations and platforms. The layout of a transit facility and platform can make it challenging to completely cover an area of interest with WiFi sensors. Transit stations can be underground with cover-controlled access

points, or outdoors, with uncontrolled/open access. Platforms can also be located in the middle of or parallel to rail tracks/running ways, or tracks/running ways can be located in between two platforms. If detection devices are placed too close to each other then they can both detect the same device simultaneously. Therefore, some criterion needs to be defined for placement of sensors to avoid detection duplication of the same smartphone, based on RSSI values [12].

Passengers may prefer to wait either inside the station building, in a shelter located at the platform or in open area, depending on prevailing weather conditions. This makes it challenging to differentiate passengers who are waiting at a central platform between the different train routes they are waiting for. After the placement of WiFi sensors, testing must be conducted with signal strength measuring devices to note RSSI value at various locations within a transit facility to calibrate the RSSI values with respect to various locations in the transit facility. Volunteers holding smartphones of various brands and models and of known MAC addresses can walk through the transit facility to confirm that the data is efficiently collected and stores by the WiFi sensors and database server.

3.2 Data Filtration

Differentiating transit passengers from non-passengers is a challenge as the WiFi sensing devices detect all the smartphones within range. The devices that can be detected include fixed WiFi devices [9], vehicles and nearby pedestrians within the sensing range. Fortunately, the number of smartphone detections (i.e., the number of consecutive transmissions a particular smartphone is detected at a given location), detection duration, RSSI, date and time of the detected smartphones can be used in combination to differentiate transit passengers from non-passengers. Fusing data from other sensors such as APC or automatic fare collection can also improve the filtration process.

3.3 Estimation

Once the data is filtered and transit passengers' smartphone IDs have been separated, analyses are conducted to determine passenger activity (ridership, wait time, travel time, crowding levels, etc.) and system operations (arrival/departure timing, dwell time, travel time between stations, etc.) parameters. Smartphone MAC addresses, dates and times are used in this process. Commercial software packages such as Power BI are available to perform the analysis in real-time and show the Key Performance Indicator (KPI) analytics.

3.4 Data Validation

Field data from a reliable source can be used to check the accuracy of the passenger activity data as estimated from the smartphone sensors. Usually, manual surveys or observations from surveillance cameras are used to validate smartphone-based transit passengers' activity related data [4, 6, 8, 9]. Various methods, such as Kalman filtering and neural networks can be used to validate smartphone data. These methods can also be used to predict passenger activity based on observed historical patterns.

4 Potential of Smartphone Data Collected Through WiFi

This section describes how various parameters and transit performance indicators can be determined using transit passenger smartphone data collected through WiFi sensors.

4.1 Ridership and Wait Time at Stations/Stops

Transit ridership information primarily refers to the number of passengers boarding and alighting, origin and destination demand, and transit vehicle loading. This information is important in route planning, transit service design and demand management. Turn gates and manual counts are still used by many transit agencies to estimate ridership at stations and on vehicles. Other automated passenger measurement includes APCs, camera detection and axle load sensors [6, 12]. APC systems are considered an industry standard and are widely used on transit vehicles for collecting continuous ridership data. Transit agencies also use AFC systems to gather ridership information and WiFi technology can be used to efficiently collect continuous ridership and passenger wait time data at transit stations through passengers' smartphone ID. WiFi sensors are installed at strategic locations on platforms to sense smartphones when passengers arrive at the station. The smartphone data need to be filtered to remove the noisy detections of non-passenger WiFi signals. The data need to be cleaned further by filtering out the passengers who were waiting at the same platform but for different transit routes. For example, when two vehicles arrive at the platform simultaneously, it is challenging to differentiate passengers boarding to and alighting from a specific vehicle. Different strategies are required to filter the data route wise at a specific time when more than one transit vehicles arrive at a platform from opposite directions. Based on the arrival pattern of transit vehicles it may not be possible in some cases to filter out the boarding, alighting and waiting passengers at a platform and passenger load on a transit vehicle, unless the ridership information from upstream stations is known.

As WiFi sensors continually scan for smartphones, the data can be manipulated to determine the ridership and passenger wait times. Smartphone ID data from all transit stations can be used to determine a passenger's origin and destination by MAC address matching analysis. The WiFi sensor data should be compared to actual ridership data from an authentic source, such as an APC system or CCTV cameras, to evaluate the accuracy of ridership and passenger wait time.

Using this technique to determine ridership could allow transit agencies to determine metrics such as arrival loads, departed loads, boarding, alighting, origin and destination matrices, Max Load Points (MLP are the locations on a transit route where passenger loads are maximized), aggregate and disaggregate passenger wait times and the temporal distribution for all routes and directions of route serving a station.

4.2 Demand at Transit Stations

Scheduling and dispatch are essential steps of transit route planning. These steps require passenger demand estimates at each station along a route. In fixed-route fixed scheduling, high passenger demand at transit stations along a route may cause overcrowding. To manage overcrowding, additional transit vehicles can be dispatched to carry the leftover passengers in case reliable passenger arrival data is not available. Fixed routes with dynamic schedules require accurate passenger arrival patterns at stations along a route to optimize vehicle scheduling and dispatching. Transit agencies can thus mitigate crowding and excessive passenger wait time if real-time or estimated passenger demand information is available.

Currently, transit agencies use vehicle onboard APCs, manual observations and/or CCTV cameras at stations to gather crowding information. Data from WiFi sensors can be used to estimate and predict passenger demand in real-time at all stations of a route at a lower cost than current systems. This data could allow transit agencies to optimize the scheduling and dispatching of main routes as well as those of feeder bus routes. WiFi sensors installed at access points of stations can continuously detect passengers arriving at a station by timestamps. The MAC addresses, dates and detection time data for every passenger's smartphone at every station can be analysed to determine real-time passenger demand. However, data filtering is required, as mentioned above, to distinguish transit users from non-transit users. Also, the demand data collected over time can be further analysed for predicting future daily and hourly transit passenger arrival rates (passenger/unit time).

4.3 Customers Access and Egress at Stations

Most major transit stations have more than one passenger entrance/exit points. Some of these entrance/exit points are well used and others are underused. Data

on the number of passengers entering and exiting through all station entry/exits could provide insights for future station design and modification of existing stations. Currently, entrance/exit points at transit are mainly manually monitored by CCTV cameras from control centers. WiFi technology can continuously record passenger counts in and out of a station through entrance/exit points. Transit agencies can use this information to identify which entrance/exit points are most frequently used by time of day and when and what is peak use. WiFi sensors can be installed at all entrance/exit points as well as on the station's platform. As a passenger enters a station through an entrance, the passenger's smartphone is detected by the WiFi sensor installed at that entrance. After few seconds or minutes, depending on the walking time from the entrance to the platform, the same smartphone will be detected by the WiFi sensor installed at the platform. The RSSI value of detections will increase as the passenger approaches the platform.

Similarly, a passenger alighting from a transit vehicle will be detected by WiFi sensors installed at the platform. After few seconds or minutes the passenger will be detected by WiFi sensors installed at the station exit point. As the passenger leaves the station, the RSSI value detected by the WiFi sensor installed at the exit will fade out until it is no longer detectable. By the MAC address matching process, transit agencies can determine the number of passengers entering and leaving a station. Total transit passengers' usage of a specific entrance/exit point at a station can thus be monitored using this technique. As in the other applications, the data needs to be filtered out to distinguish transit users from non-transit users.

4.4 Station Crowding

Transit agencies can use crowding information at platforms to improve dwell times, manage passengers' evacuations during emergencies and handle transit service disruptions. Crowding information at platforms can help improve safety during the COVID-19 pandemic. Passenger boarding and alighting time can be reduced by spreading passengers along the platform before transit vehicle arrival so, they can board less congested train cars. A luminous band along the platform (equal to the length of the train) can be provided to show occupancy of the arriving train cars. Transit apps are a lesser costly solution in this regard, as these apps can show the occupancy of arriving train cars in real-time, allowing passengers to choose a less busy train car [5].

Currently, APC systems installed on every transit vehicle are used to track passenger loads at stations and transmit this information to the downstream station, as the transit vehicle leaves the upstream station. Crowding levels at platforms are monitored through CCTV cameras.

WiFi sensors can be installed at transit stations to determine passenger loads on transit vehicles and platforms. An arrangement of WiFi sensors at the platform is required to record passenger crowd data as was described above for the ridership and

passengers waiting time. Deliverables of this technique will include real-time and historical transit vehicle loads and platform load patterns.

4.5 Transit Operational Monitoring and Management

Transit agencies use precise transit arrival and departure information to calculate KPIs such as on-time performance and dwell time. These KPIs are then used as inputs for scheduling and resource optimization. Transit agencies also use these KPIs to identify areas for service improvement such as travel time reliability improvement by introducing transit priority measures, decreasing station dwell time by improving boarding processes, and other service improvements.

Currently, check-in and check-out sensors are installed on train tracks at the ends of platforms. These record the arrival and departure times of every train. APCs can also provide this information whenever the doors of a vehicle are opened or closed at the station. WiFi sensor technology can be used to gather continuous transit vehicle arrival and departure timing at stations at a lower cost than current methods.

During revenue service hours, transit vehicles are full of passengers holding smartphones. As a transit vehicle arrives, many smartphones would be detected simultaneously, indicating the arrival of the vehicle. As the transit vehicle departs from the platform the RSSI values of the smartphone on the transit vehicle will decrease until the smartphones are no longer detected. The reported time and date of these events can determine transit vehicle arrival and departure times.

It could be challenging to detect transit vehicles during off-peak periods when transit vehicles may be empty, as no smartphones would be detected by a vehicle's arrival. One potential solution to this problem could be to install WiFi sensors at all stations along a route to identify this type of unusual event. Alternatively, a WiFi-enabled device with a known MAC address could be deployed in every transit vehicle. This system would make it easy to detect transit vehicles the time the transit vehicle arrives and departs. The majority of new transit vehicles are installed with GPS modems, which are cellular and WiFi-enabled. The MAC addresses of these GPS modems can thus be used to detect transit vehicles at stations. The cost of the WiFi sensors-based solution would be still less than standard APC systems.

Transit vehicles' arrival and departure timings data can be used to calculate dwell time at stations. Schedule adherence can be determined by comparing the recorded arrival with the scheduled times. Travel time and average vehicle speed between two stations can be calculated based on the recorded arrival and departures at both the stations.

5 Calgary Transit's Pilot

The City of Calgary is the third largest municipality in Canada (after Toronto and Montreal) and the largest city in the province of Alberta. Calgary Transit is the public transit service provider in Calgary. Calgary Transit's rapid transit system consists of BRT (MAX) and LRT services. Two LRT lines (the Red and Blue lines shown in Fig. 2) are the backbone of Calgary's transit system and carry over 200,000 customers daily. Calgary Transit has been providing free WiFi to its customers at all LRT stations for many years. Recently, Calgary Transit, Shaw and the University of Calgary have partnered to conduct a pilot for collecting passenger activity data by tracking smartphones through WiFi sensing technology. Seven LRT stations (out of a total of 45) were selected for the pilot. A Cisco Meraki MR42 WiFi sensor have been installed at each station (Fig. 2). The pilot aims to test the accuracy of the WiFi sensing technology and to infer parameters including station ridership, line ridership at MLPs, wait times, origin–destination matrices, train travel times between stations and dwell times at the stations.

The WiFi sensors installed at the stations will not store the smartphone data that they will capture. Therefore, a database server is being set up to host the data remotely in real-time, continuously for six months. Appropriate data filtering techniques will be applied to separate the LRT passengers' smartphones out of all the smartphones detected by the WiFi sensors. The filtered data will then be analysed to infer passenger activity and LRT operational parameters. Ridership data will be collected manually and with APC devices during the study period to validate the accuracy of the parameters inferred from the smartphone data. Finally, analytics of various key performance indicators will be created to monitor the passenger activity and LRT operations in real-time.

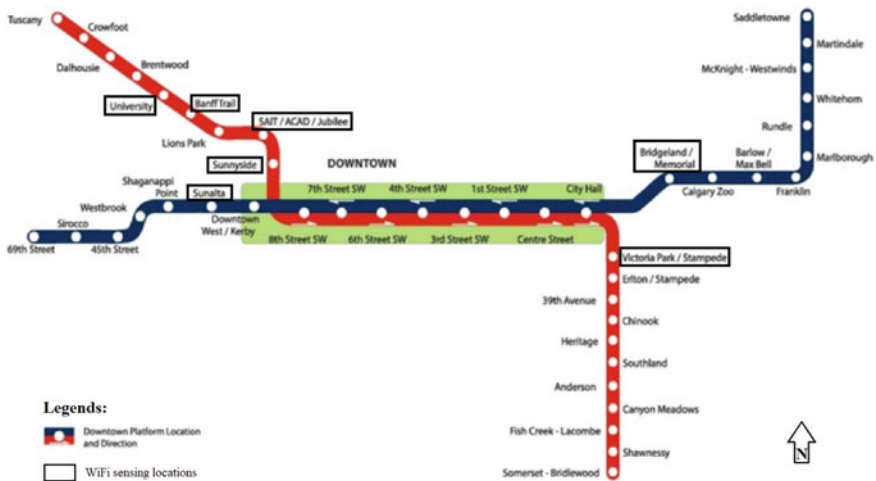


Fig. 2 Location of the WiFi sensors at selected stations on the LRT network in Calgary, AB, Canada

6 Conclusions

Transit agencies use various types of data collection systems that are expensive to install and maintain to collect passenger activity and system performance data. This paper has explored WiFi sensing technology based on relevant literature and recent research findings. WiFi technology has the potential to collect information on transit passenger activity and transit system performance efficiently and with low investment. One system consisting of WiFi sensors and database servers can be set up to gather all the necessary information for which multiple devices are currently being used. Yet, the application of WiFi for monitoring purposes in the transit industry worldwide is still rare. With the expected increase in data transfer speed (5G) and wider bandwidth, WiFi-based monitoring will become even more accessible and inexpensive for transit agencies. Data privacy will still be an issue; transit agencies need to assure the security of passengers' information.

7 Recommendations

Pilot studies must be conducted by installing WiFi sensors at one or more stations as a proof-of-concept before transit agencies can undertake network-wide deployments of WiFi sensing technology. Multiple WiFi sensors must be installed at platforms or station access points, as at least three sensors are required to triangulate a location, thus improving the accuracy of the location coordinates of a detected smartphone. While one WiFi sensor may suffice for a small bus stop, effective placement of multiple WiFi sensors is needed to achieve high accuracy and minimize filtering requirements.

Manual data or observation from video data is needed to validate passenger and transit vehicle activity information determined from smartphone data collected through WiFi sensors. In addition, data from other industry proven and in-use sources such as, Automated Passengers Counts (APC), which have an accuracy of 98%, if available, must also be used as a base to confirm the accuracy of the passenger information inferred from the WiFi smart phone data set. Transit agencies, IT consultants and legal agencies should cooperate to establish standards around smartphone data exchange in a secure way with transit agencies through WiFi sensors for transit improvement.

8 Further Work

The use of WiFi sensing technology for inferring transit passenger activity and system performance is an emerging research topic. Much research is required to further

explore the potential of this technology to generate more accurate data, without compromising the privacy of smartphone users.

To improve privacy, smartphone manufacturers have started implementing MAC address randomization. This implies that a new address can be generated for every time a WiFi-enabled device sends to all channels for a connection. Due to this security protocol, it might be no longer possible to identify every element of transit passenger's activity [15]. However, the smartphone data through WiFi still offers a promising opportunity to determine ridership and identify other localised information at the same station (e.g., ridership counts, wait time, platform crowding level, etc.). Such limitations of the smartphone data collected through WiFi can be further explored and appropriate solutions can be developed to gather transit passenger activity information at a disaggregate level.

Acknowledgements The authors would like to acknowledge Calgary Transit for their support and interest in this research, Shaw for providing the WiFi sensors and their technical services, the University of Calgary's IT team for creating database servers for this pilot study. Also, thanks to Sam Pollock, graduate student at the University of Calgary, for his help proofreading this paper.

References

1. Abedi N, Bhaskar A, Chung E, Miska M (2015) Assessment of antenna characteristic effects on pedestrian and cyclists travel-time estimation based on Bluetooth and WiFi MAC addresses. *Transp Res Part C: Emerg Technol* 60:124–141. <https://doi.org/10.1016/j.trc.2015.08.010>
2. Chon Y, Kim S, Lee S, Kim D, Kim Y, Cha H (2014) Sensing WiFi packets in the air: practicality and implications in urban mobility monitoring. In: *UbiComp 2014—proceedings of the 2014 acm international joint conference on pervasive and ubiquitous computing*, pp 189–200. <https://doi.org/10.1145/2632048.2636066>
3. Dong Z, Mokhtarian PL, Circella G, Allison JR (2014) The estimation of changes in rail ridership through an onboard survey: did free Wi-Fi make a difference to Amtrak's Capitol Corridor service? *Transportation*. <https://doi.org/10.1007/s11116-014-9532-7>
4. El-Tawab S, Arai I, Salman A, Brian Park B (2020) A framework for transit monitoring system using IoT technology: two case studies. In: *2020 IEEE international conference on pervasive computing and communications workshops, PerCom Workshops 2020*, pp 1–6. <https://doi.org/10.1109/PerComWorkshops48775.2020.9156130>
5. Hänseler FS, van den Heuvel JPA, Cats O, Daamen W, Hoogendoorn SP (2020) A passenger-pedestrian model to assess platform and train usage from automated data. *Transp Res Part A: Policy Pract* 132(January):948–968. <https://doi.org/10.1016/j.tra.2019.12.032>
6. Kim H, Kwon S, Wu SK, Sohn K (2014) Why do passengers choose a specific car of a metro train during the morning peak hours? *Transp Res Part A: Policy Pract* 61:249–258. <https://doi.org/10.1016/j.tra.2014.02.015>
7. Min H, Zhou F, Jui S, Wang T, Chen X (2004) White paper white paper. *Distribution* (January):1–23
8. Paradedda DB, Kraus Junior W, Carlson RC (2019) Bus passenger counts using Wi-Fi signals: some cautionary findings. *Transportes* 27(3):115–130. <https://doi.org/10.14295/transportes.v27i3.2039>
9. Pu Z, Zhu M, Li W, Cui Z, Guo X, Wang Y (2021) Monitoring public transit ridership flow by passively sensing Wi-Fi and bluetooth mobile devices. *IEEE Internet Things J* 8(1):474–486. <https://doi.org/10.1109/JIOT.2020.3007373>

10. Ryu S, Park BB, El-Tawab S (2020) WiFi sensing system for monitoring public transportation ridership: a case study. *KSCE J Civ Eng* 24(10):3092–3104. <https://doi.org/10.1007/s12205-020-0316-7>
11. Shaw (2020) How to use shaw smart WiFi analytics
12. Shlayan N, Kurkcu A, Ozbay K (2016) Exploring pedestrian bluetooth and WiFi detection at public transportation terminals. In: *IEEE Conference on intelligent transportation systems, proceedings, ITSC.*; 229–234. <https://doi.org/10.1109/ITSC.2016.7795559>
13. Statista Research Department (2019) Global mobile phone internet user penetration-2019
14. Statistics (2020) Mobile internet user penetration rate in selected countries in 2020| Statistic. Available from <https://www.statista.com/statistics/239114/global-mobile-internet-penetration/>
15. Vanhoef M, Matte C, Cunche M, Cardoso LS, Piessens F (2016) Why MAC address randomization is not enough: an analysis of Wi-Fi network discovery mechanisms. In: *ASIA CCS 2016—proceedings of the 11th ACM Asia conference on computer and communications security*, pp 413–424. <https://doi.org/10.1145/2897845.2897883>.

Assessment of Adaptation Solutions to Floods with PCSWMM and a Multicriteria Analysis for a Very Small Watershed



Audrey Coulombe, Jean-Luc Martel, Annie Poulin, Mathias Glaus, Geneviève Audet, and Steve Girard

1 Introduction

In southern Quebec, the spring freshet generates flooding water levels in a large number of rivers [1, 2], affecting around 80% of riparian municipalities [3]. In summer and fall, heavy rainfall caused by convective or tropical storms can also cause their share of inconvenience [4], in particular over small rural watersheds. These pluvial flooding are also expected to be impacted by climate change through an intensification of extreme rainfall events [5]. Small cities located in very small ungauged watersheds are particularly ill-equipped to handle this situation.

Therefore, the main objective of this study is to develop and test a methodology to be used by municipalities located in very small ungauged watersheds in order to assess flood adaptation scenarios while considering climate change. A case study was conducted in the Municipality of Saint-Isidore (hereafter Saint-Isidore) located in Quebec, Canada (Fig. 1) at the head of the Saint-Régis River watershed. The aim is to apply the developed methodology to classify various potential adaptation

A. Coulombe (✉) · J.-L. Martel · A. Poulin · M. Glaus
École de Technologie Supérieure, Construction Engineering Department, Montreal, Canada
e-mail: audrey.coulombe.1@ens.etsmtl.ca

A. Coulombe · J.-L. Martel · A. Poulin
Hydrology, Climate, and Climate Change (HC3) Research Laboratory, Montreal, Canada

M. Glaus
Station Expérimentale des Procédés Pilotes en Environnement (STEPPE) Research Laboratory,
Montreal, Canada

G. Audet
Société de Conservation et d'aménagement des Bassins Versants de la Zone Châteauguay (OBV
SCABRIC), Montreal, Canada

S. Girard
Municipality of Saint-Isidore—Public Works, Montreal, Canada

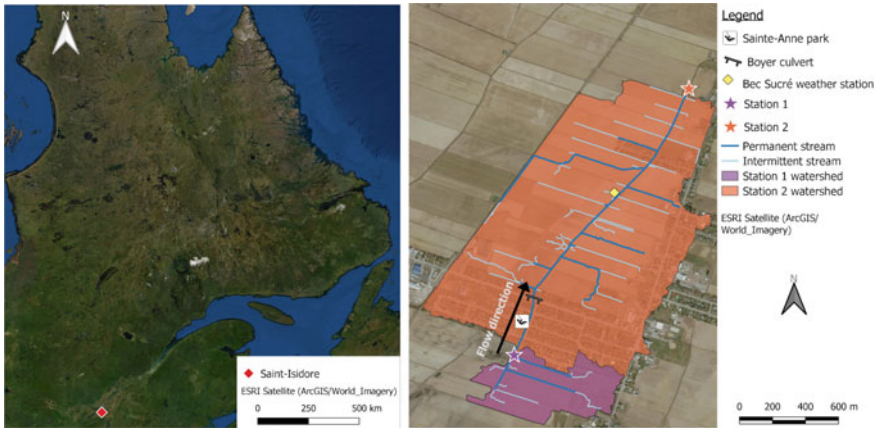


Fig. 1 Localization of the study site

solutions against flooding of fluvial origin caused mainly by the high water level of the river following summer intense rainfall. This initiative has been spearheaded by the watershed organization of the OBV SCABRIC (Châteauguay River conservation and development society), aiming to better understand flooding and the solutions that can be put forward while integrating the aspect of climate change as well as decision support for analysis.

2 Study Site

The majority of the municipality's watershed is made up of agricultural lands (75%), with the remaining being occupied by residential areas (25%) [6]. Several agricultural plots oriented perpendicular to the stream are drained using tile drainage directly into the Saint-Régis River to evacuate water from fields more efficiently to improve crop productivity. A large portion of the Saint-Isidore sewer network is pseudo-separative, meaning that foundation drains and roof gutters are connected directly to the sanitary sewer rather than the storm sewer. The storm sewer network collects runoff water through sumps and channels it directly into the Saint-Régis River.

The Boyer culvert serves as the outlet of the Saint-Régis River in the municipality's residential area (see Fig. 1). During intense summer convective storms, the network upstream of the culvert can become completely overloaded, causing local flooding in the city's area near the stream. These flood events are amplified with the water level remaining high near Sainte-Anne Park due to the culvert's raised elevation preventing proper drainage following rainfall events, also creating a safety issue for children.

3 Methodology

This section covers the three main steps to the proposed methodology to evaluate and compare various potential adaptation solutions against flooding: (1) the data collection, (2) the modelling component (hydrological, hydraulic and climate change) and (3) the decision support component (multi-criteria analysis).

3.1 Data Collection

To build a hydrological and hydraulic model over an ungauged watershed, data collection is the first step that must be accomplished. With respect to this case study, two water-level stations (Stations 1 and 2; Fig. 1) composed of a pressure sensor were installed early September 2019 allowing the water level to be deduced by using an additional barometric pressure sensor. Unfortunately, following the flood preventive de-icing of the Saint-Régis River, Station 2 sensor was torn apart on March 13, 2020.

A land survey activity made it possible to generate cross sections and the longitudinal profile of the Saint-Régis River and to obtain the position and rim elevations of the storm manholes and culverts. In order to properly represent the storm sewer system, the invert elevation of the majority of the accessible storm manholes was measured and completed using as-built plans or by interpolation if needed.

The delimitation of the watershed is required to determine the surface area which is drained towards the water-level stations. To this end, a digital terrain model (DTM) produced using a LiDAR survey in 2017 with an altimetric and planimetric precision of ± 15 cm was used. The ArcGIS software was used to process the DTM and to make it hydrologically logical. During these manipulations, the culverts (invisible to the DTM) are artificially burned in the DTM allowing a complete and realistic layout of the hydrographic network. Stations 1 and 2 drainage areas were found to be 0.26 km^2 and 2.08 km^2 , respectively. The delineation of the watershed then makes it possible to define the search for additional information to supply data to the hydrological and hydraulic model, such as the site pedology and detailed land use.

3.2 Hydrological and Hydraulic Modelling

The purpose of this study's hydrological and hydraulic modelling is to better understand all the processes generating runoff and its path over the territory leading to flooding problems for which a solution is to be identified. Considering the rural and urban component of the territory under study, the hydraulic model must be able to simulate both the flow processes for the rural drainage network (tile drainage, ditches and the Saint-Régis river channel) as well as the storm sewer network. In addition, since the land use of the study site is diversified and the flooding issues are located

in different places in the watershed, semi-distributed modelling (whose parameters vary depending on the territory) is necessary. Finally, the watershed being of a very small area ($\sim 2 \text{ km}^2$), the hydrological and hydraulic processes must be resolved at a sub-daily time step to avoid missing peak flows.

The hydrological and hydraulic modelling tool used to take these different characteristics into account is the PCSWMM software developed by Computational Hydraulic Int. (CHI) incorporating directly into its calculation engine the latest version of the Storm Water Management Model (SWMM) developed by the United States Environmental Protection Agency (EPA). The following subsections present the steps to be taken in order to build and calibrate the model.

3.2.1 Building of the Model and Calibration

The PCSWMM model was built using the data collected at Sect. 3.1. The storm sewer network and the Saint-Régis River channel and banks were modelled using the surveyed data and the as-built plans. The rural drainage ditches were also modelled and extracted from the DTM in ArcGIS. The Green-Ampt infiltration scheme was selected and its parameters were adjusted according to the study site known pedology. Groundwater flow was added to the model and also parametrized according to the site's pedology. All rugosity values (Manning coefficient), impervious percentages, depths of depression storage, infiltration scheme parameters were initially adjusted based on typical values [7, 8].

The calibration of the PCSWMM model is conducted by further adjusting parameters such as the rugosity from the sub-watersheds impervious and pervious areas and the Saint-Régis River channel and banks. The sub-watersheds' conceptual lengths and widths and impervious percentages were also modified to improve the hydrograph timing. The calibration and validation results will be discussed in Sect. 4.1.

3.2.2 Potential Adaptation Solutions

Various potential adaptation solutions were identified together with Saint-Isidore that could potentially be implemented. These are presented in Fig. 2 and can be summarized as follows:

- *Lowering of Boyer culvert (C)*: this culvert is currently raised above the river bed, creating an accumulation of water upstream within the residential area. Lowering of the culvert would allow for a proper drainage, providing more storage capacity during subsequent rainfall events.
- *Reprofiling of the Saint-Régis River (DR and UR)*: the river bed would be reprofiled using an average slope from its source to the outlet (Station 2) to remove the non-uniform accumulation of deposits, providing more conveyance. Downstream (LR)

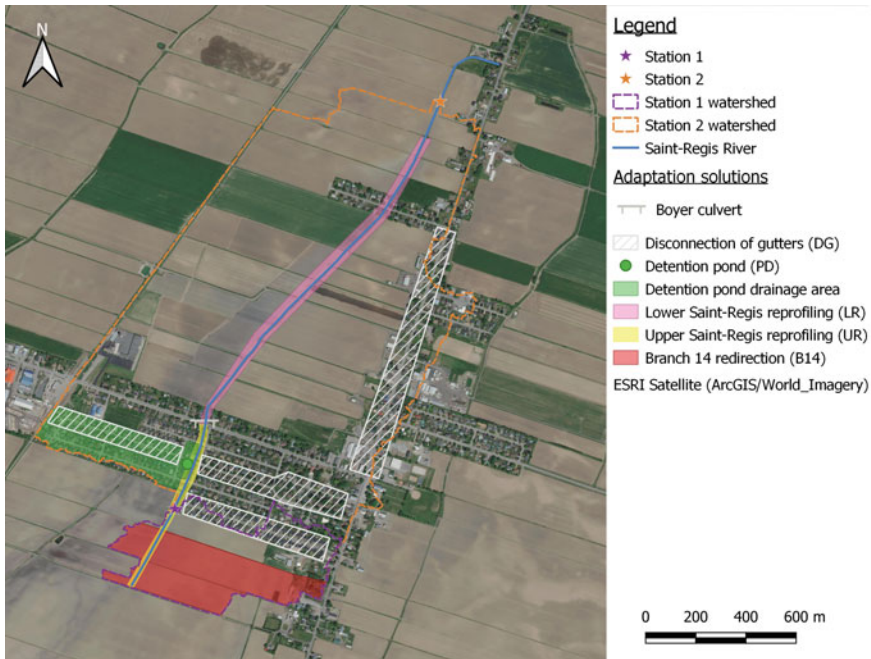


Fig. 2 Localization of potential adaptation solutions within the territory of Saint-Isidore

and upstream (UR) reprofiling from the residential area (near the Boyer culvert) are considered as two distinct measures.

- *Detention pond (DP)*: construction of a dry detention pond in the Sainte-Anne Park, storing runoff volumes from approximately an 8000 m² drainage area and gradually releasing it in the river. The storm sewer network from the targeted areas would be redirected towards the detention pond.
- *Branch 14 redirection (B14)*: historically, a small portion of the Station 1 watershed was diverted towards the municipality. The Branch 14 could be restored to its previous configuration, but would likely require environmental administrative procedures to do so.
- *Disconnection of gutters (DG)*: even though a portion of the municipality’s sewer network is pseudo-separative, multiple gutters were found to be connected to the storm sewer network. These gutters would be disconnected from the network and redirected towards permeable surfaces or in rain barrels.

The calibrated PCSWMM model is used to evaluate individual adaptation solutions and combinations of those thereof. A 3-h Chicago synthetic storm with a 25-year return period was selected as the design event under which all scenarios are evaluated and compared. This storm was generated using the intensity-duration-frequency (IDF) curves from Environment and Climate Change Canada’s Ste-Clothilde weather station closest to the study site. In order to take the impact of climate change into

account, an 18% correction factor was applied to the storm intensities following the Quebec Ministry of Sustainable Development, Environment and Fight Against Climate Change recommendations [9].

3.3 Multi-criteria Analysis for Decision Support

The comparative analysis of the various adaptation solutions is based on socio-economic, technical and environmental criteria, which can be determined using a multi-criteria analysis. This analysis makes it possible to rank the measures and requires the identification as well as the weighting of the criteria considered. The choice of criteria and their units is based on the literature as well as the specificities of the proposed technical solutions. The attribution of the value to each adaptation solution for each selected criterion is based on the results of the modelling component and on knowledge from the literature or expert judgments. The decision support process is based on the PROMETHEE method [10] for which the method as well as the methodological approach are explained by [11]. The study’s selection of criteria is shown in Table 1 and are divided into three categories according to the fulfillment of three conditions: (1) the list of criteria relevant to the study must be exhaustive, (2) consistency in the role of a criterion at its local level and when it is immersed in its category, and (3) the non-redundancy of the criteria [12].

In the PROMETHEE methods, the adaptation solutions $A = (a_1, a_2, \dots, a_n)$ are noted for each criterion $C = (c_1, c_2, \dots, c_n)$. Then, the difference $d_c(a_i, a_j)$ is computed for all pairwise comparison of adaptation solutions according to each criterion c : $d_c(a_i, a_j) = c_c(a_i) - c_c(a_j)$. Afterwards, these differences are translated in a preference degree P ranging from 0 to 1 based on a preference function. A score of 1 indicating a strong preference for an adaptation solution against another for one criterion comparison, and a score of 0 means no preference. In this study, two preference functions are employed:

$$\text{Linear } P(a_i, a_j) = \begin{cases} 0 & \text{if } |d_c(a_i, a_j)| \leq q \\ \frac{|d_c(a_i, a_j) - q|}{p - q} & \text{if } q < |d_c(a_i, a_j)| \leq p \\ 1 & \text{if } |d_c(a_i, a_j)| > p \end{cases} \quad (1)$$

$$\text{Usual } P(a_i, a_j) = \begin{cases} 0 & \text{if } d_c(a_i, a_j) \leq 0 \\ 1 & \text{if } d_c(a_i, a_j) > 0 \end{cases} \quad (2)$$

Then, these preference degrees are paired with the weight’s criterion given by the authorities. Due to COVID-19, synthetic weights were defined instead until public consultation can be conducted.

$$\pi(a_i, a_j) = \sum_{c=1}^n P_c(a_i, a_j)w_c \quad \text{and} \quad \pi(a_j, a_i) = \sum_{c=1}^n P_c(a_j, a_i)w_c \quad (3)$$

Table 1 Description and units of the selected criteria

| Context | Criteria | Description of the criterion | Units |
|---------------|--------------------------------|---|--------------------------------|
| Socioeconomic | Administrative complexity | Solution requires the mobilization of resources from the municipality to achieve the solution | \$ |
| | Cost of investment | Cost required to design and build the solution | \$/year |
| | Operation and maintenance cost | Cost required for the operation of the solution and its routine maintenance | \$/year |
| | Repair cost | Cost to be expected to bring the solution up to standard | 1-5 |
| Technical | Ease of implementation | Ease with which a solution can be implemented | 1-5 |
| | Ease of maintenance | Simplicity with which it is possible to keep the solution running smoothly | 1-5 |
| | Adaptable | Ease of upgrading the solution | 1-5 |
| | Water flow | Diminish or maintain the water flow at the watershed's outlet | m ³ s ⁻¹ |
| | Water level | Peak water height at Sainte-Anne park | m ³ s ⁻¹ |
| | Proven solution | Solution commonly used occasionally, or in development | 1-5 |
| Environment | Water quality | The solution minimize the contribution of contaminants to waterways | 1-5 |
| | Faunic habitat quality | The solution optimize or improve the quality of habitats for aquatic and terrestrial fauna | 1-5 |
| | Erosion potential | The solution reduces sediment inputs downstream | 1-5 |
| | Infiltration potential | The solution promotes the recharge of the water table | 1-5 |

Thus, the leaving (positive) flow (Φ^+) and the entering (negative) flow (Φ^-) are deducted. The former means how this adaptation solution outranks all the other adaptation solutions according to one criterion, in contrast to the entering flow, which means how the adaptation solution is outranked by all the other adaptation solutions.

$$\Phi^+(a_i) = \frac{1}{n-1} \sum_{x \in A} \pi(a_i, x) \quad \text{and} \quad \Phi^-(a_i) = \frac{1}{n-1} \sum_{x \in A} \pi(x, a_i) \quad (4)$$

where n is the number of adaptation solutions considered in the analysis. The net flow (Φ) is the difference of the leaving flow (Φ^+) and the entering flow (Φ^-). The flow score is computed as the difference between the positive flow and the negative flow: $\Phi(a_i) = \Phi^+(a_i) - \Phi^-(a_i)$.

4 Results and Discussion

This section presents the results obtained for the modelling components (hydrological, hydraulic and climate change), followed by the multi-criteria analysis for decision support. First, the calibration and validation of the PCSWMM model are presented. Then, the potential adaptation solutions are simulated in current and future climates to determine their strengths, weaknesses and possible synergies between them. Finally, the data obtained from PCSWMM modelling as well as other data from practice and expert judgments are fed into a multi-criteria analysis in order to obtain a ranking of the adaptation solutions.

4.1 PCSWMM Model Calibration

The objective of the PCSWMM model is to reproduce the water dynamics of the Saint-Isidore watershed. A few rainfall events were selected to calibrate and validate the PCSWMM model using the water level gauging (between September 12, 2019, and September 11, 2020). Consecutive rainfall events that took place between October 15, 2019, and November 3, 2019, were selected to calibrate the model, providing diversity in the intensity and duration of those events [13]. According to Moriasi et al. [14], several indicators can be used to jointly assess the performance of a hydrological model as satisfactory or not. The model is deemed acceptable when the following conditions are met: 1) the Nash-Sutcliffe coefficient (NSE) > 0.6 , 2) the coefficient of determination (R^2) > 0.6 and 3) the percentage bias (P_{BIAS}) $\pm 15\%$.

Due to the limited number of events with enough rainfall to generate above average water levels, two sequences of two rainfall events (occurring in summer and in fall) were selected for the validation, whereas the remaining ones served for the calibration. Similar values were obtained for both calibration and validation events. The two validation events between October 25 and November 4 are presented in Fig. 3 and the performance indicators for both fall and summer event sequences are shown in Table 2.

Validation results shown in Table 2 indicate that satisfactory performance is obtained for all indicators at Station 2 during the fall event, and Station 1 for the summer events. For the fall events, Station 1 performance is poorer (as seen in Fig. 3) and is, in fact, systematically lower than Station 2 performance. This can possibly be explained by the presence of the Boyer culvert and the historical diversion (see Sect. 3.2.2) that may affect the runoff and routing processes in that particular area.

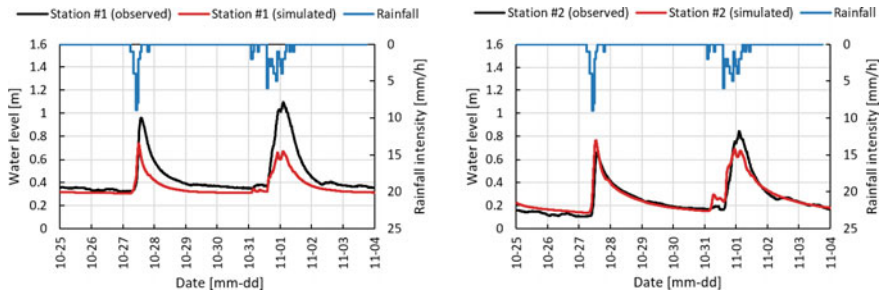


Fig. 3 Validation at Station 1 (left) and Station 2 (right) from October 25 to November 4, 2019

Table 2 Performance indicators for both validation events

| Events dates | Station # | NSE | R ² | PBIAS (%) |
|--------------------------|-----------|------|----------------|-----------|
| 2019-10-25 to 2019-11-04 | Station 1 | 0.46 | 0.90 | 19.0 |
| | Station 2 | 0.90 | 0.90 | -2.5 |
| 2020-08-02 to 2020-08-09 | Station 1 | 0.75 | 0.91 | 10.0 |

Note that Station 2 data is not available beyond March 13, 2020, due to equipment damages and malfunctioning from that date on. Overall, the calibration was considered satisfactory to conduct the case study and evaluate the performance of the various adaptation solutions. However, a larger monitoring time would have provided more rainfall events for both calibration and validation periods, leading to more robust results for the analysis.

4.2 Evaluating the Different Adaptation Solutions

4.2.1 Baseline Scenario

As seen in Fig. 3 for Station 1, the restriction caused by the Boyer culvert is preventing the water level to decrease below approximately 0.4 m following any significant rainfall event. Indeed, the municipality has mentioned that this is a recurrent problem where the water level remains high for a long period of time, further contributing to potential flooding in the residential areas. Figure 4 shows the maximum water levels at a cross-section upstream of the Boyer culvert reached during the 3-h Chicago storm with a 25-year return period under present climate conditions for the six individual solutions described in Sect. 3.2. A 25 mm rainfall event was used in the model 3 days before the event to better demonstrate the problem of drainage in the residential area.

Results shown in Fig. 4 suggest that the lower Saint-Régis reprofiling and the lowering of the Boyer culvert are needed together (LR+C) to allow the upstream section to properly drain itself and lower the water level following significant rainfall

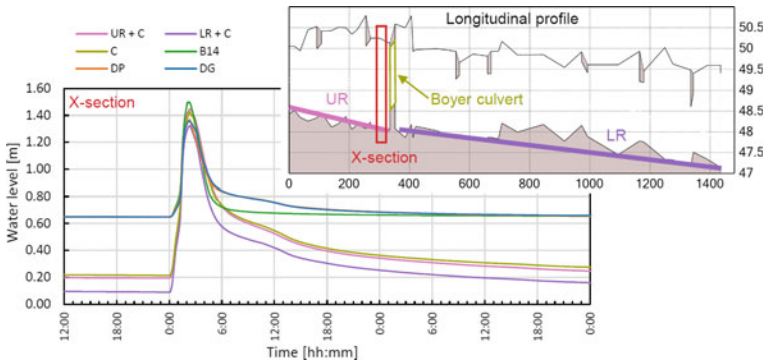


Fig. 4 Cross section (X-section) and longitudinal profile of maximum water levels for each adaptation solution alone for a 3-h 25-year return period Chicago synthetic storm following a 25 mm rainfall event

events (similar behaviours were obtained for more frequent events). This can be explained by the culvert's raised elevation with respect to the river bed as well as the important accumulation of sediments downstream of the culvert preventing its lowering alone from achieving the desired result. Thus, the combination of both lower Saint-Régis reprofiling and the lowering of the Boyer culvert (LR+C) are considered as the minimum solution and will serve as the baseline scenario hereafter.

4.2.2 Comparing All Solutions Together in Future Climate

Combining all individual adaptation solutions shown in Fig. 2 with the baseline scenario (LR+C), the PCSWMM simulations were run once again, and the results are shown in Fig. 5 in both the current and future climate conditions (i.e. with the 18% majoration factor). Furthermore, the maximum potential gain is shown with the black curve, combining all solutions together. All possible combinations of two or more measures were also tested for the multicriteria analysis, but are not shown in Fig. 5.

The results suggest that each individual solution provides improvement over the baseline scenario in both the current and future climate in terms of flow at the outlet of the Saint-Régis River segment studied. The only exception is the Branch 14 redirection (B14) which results in a lower total volume, but an increased peak flow. This could be explained due to the smaller storage area within the river bed in the residential area, leading to higher water levels and flows. The combination of the remaining solutions provided the best improvements in terms of peak flow reduction. While the detention pond (DP) and the disconnection of gutters (DG) both provide the largest gain in terms of peak flow reduction, the remaining options are intertwined together making it difficult to distinguish them. The multi-criteria analysis will allow to obtain a ranking of each individual and combination of adaptation solutions to determine the

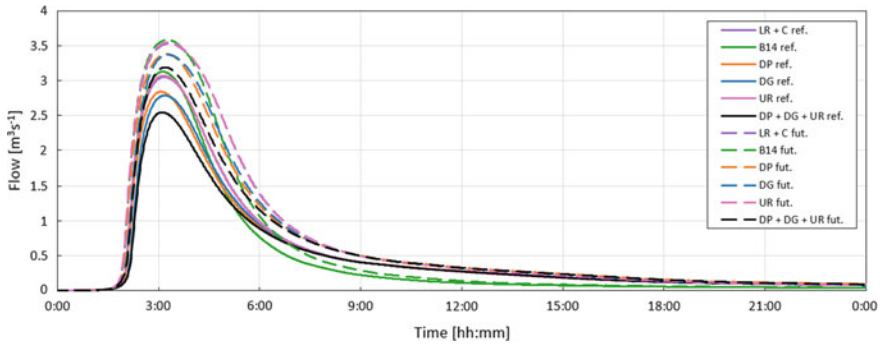


Fig. 5 Hydrographs of the flows at the outlet of the Saint-Régis River (Station 2), for the portion under study, for the baseline (LR+C), the four individual adaptation solutions combined with the LR+C baseline, and the best combination for peak flow reduction for a 3-h Chicago storm with a 25-year return period under the future climate

best scenario to be selected according to socioeconomic, technical and environmental factors.

4.2.3 Multi-criteria Analysis

The analysis of the robustness of the multi-criteria analysis was carried out by constructing various scenarios with which a specific distribution of the weighting of the criteria is associated. The total weight of 100% is equally distributed among the three major categories of criteria (socioeconomic, technical and environmental) for the base scenario, while 75% is assigned to the major category for the other three scenarios. Table 3 presents the parameters as modeled in PROMETHEE, with q and p respectively defining the indifference and preference thresholds in the case of a linear function. Better performances are obtained for smaller values over all sub-criteria.

All 16 adaptation solutions combinations were included in the multicriteria analysis. To assess the performance of the solutions against each other and against the criteria, leaving, entering and net flows are computed, allowing to rank the solutions. Figure 6 presents the net (Φ), leaving (Φ^+) and entering (Φ^-) flows as well as the classification of the best five adaptation solutions obtained with PROMETHEE.

Results indicate that the disconnection of gutters (DG) is found in the first position for all scenarios, except for the environmental one where it is in second position. The base scenario shows that both DG and the detention pond (DP) share the first three ranks. With respect to the Branch 14 redirection (B14), it reaches second and third positions in the socioeconomic scenario, but only makes it at the fifth position at best for the other scenarios. It is possible that the cost for investment, operation and maintenance does not allow DP to rank among the best in the socioeconomic scenario. This analysis shows that the ranking is sensitive to the weights given to each scenario, ultimately depending on the preferences of decision-makers.

Table 3 The sub-criteria and their properties for the multicriteria analysis. The units of q and p are the same as their respective sub-criterion's units

| Sub-criteria | Weights (%) | | | Technical | | Environmental | Type | Function | q | p |
|----------------|--------------------------------|---------------|-----------|---------------|-------|---------------|--------------|----------|--------|---------|
| | Base case | Socioeconomic | Technical | Environmental | | | | | | |
| Socio-economic | Administrative complexity | 8.33 | 18.75 | 3.13 | 3.13 | 3.13 | Quantitative | Linear | n/a | n/a |
| | Cost of investment | 8.33 | 18.75 | 3.13 | 3.13 | 3.13 | Quantitative | Linear | 82 867 | 199 472 |
| | Operation and maintenance cost | 8.33 | 18.75 | 3.13 | 3.13 | 3.13 | Quantitative | Linear | 1 980 | 4 124 |
| | Repair cost | 8.33 | 18.75 | 3.13 | 3.13 | 3.13 | Quantitative | Linear | 5 325 | 13 282 |
| Technical | Ease of implementation | 5.56 | 2.08 | 12.50 | 2.08 | 2.08 | Qualitative | Usual | n/a | n/a |
| | Ease of maintenance | 5.56 | 2.08 | 12.50 | 2.08 | 2.08 | Qualitative | Usual | n/a | n/a |
| | Adaptable | 5.56 | 2.08 | 12.50 | 2.08 | 2.08 | Qualitative | Usual | n/a | n/a |
| | Water flow | 5.56 | 2.08 | 12.50 | 2.08 | 2.08 | Quantitative | Linear | 0.09 | 0.14 |
| Environm | Water level | 5.56 | 2.08 | 12.50 | 2.08 | 2.08 | Quantitative | Linear | 0.21 | 0.35 |
| | Proven solution | 5.56 | 2.08 | 12.50 | 2.08 | 2.08 | Qualitative | Usual | n/a | n/a |
| | Water quality | 8.33 | 3.13 | 3.13 | 18.75 | 18.75 | Qualitative | Usual | n/a | n/a |
| | Faunic habitat quality | 8.33 | 3.13 | 3.13 | 18.75 | 18.75 | Qualitative | Usual | n/a | n/a |
| | Erosion potential | 8.33 | 3.13 | 3.13 | 18.75 | 18.75 | Qualitative | Usual | n/a | n/a |
| | Infiltration potential | 8.33 | 3.13 | 3.13 | 18.75 | 18.75 | Qualitative | Usual | n/a | n/a |
| | Total | 100 | 100 | 100 | 100 | 100 | | | | |

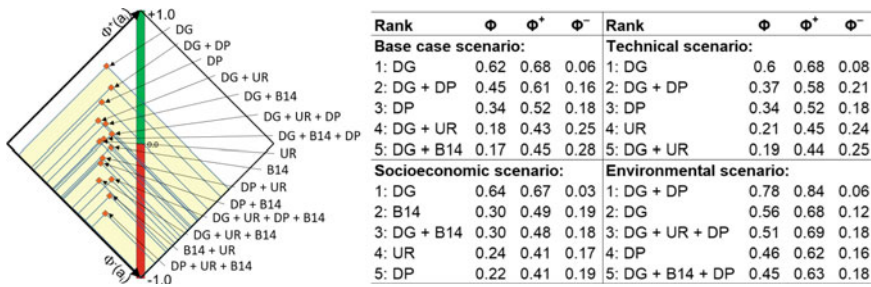


Fig. 6 Results from the multi-criteria analysis. The figure on the left is a visual representation of the base case scenario results, where the leaving flow (Φ^+) and entering (Φ^-) are added together to evaluate the net flow (Φ) corresponding to the red/green bar in the middle. A higher solution on the diamond is preferred. The flows and final ranks for all scenarios are shown in the table on the right

5 Conclusion

A case study was conducted during which the developed methodology was applied to serve as a benchmark for future studies involving very small, ungauged watersheds with the aim of helping small municipalities identify adaptation solutions to flooding. The proposed methodology includes three main steps: field data collection; hydrological and hydraulic modelling under present and future climate conditions; and multi-criteria analysis. Field data collection involved installing two water-level stations and recorded data over a one-year period were used to calibrate an event-based PCSWMM model. Then different possible adaptation scenarios were analysed and compared using the PROMETHEE method.

Overall, the results obtained from the hydraulic and hydrological simulations, as well as the multicriteria analysis indicate that a combination of adaptation solutions is likely the best option for Saint-Isidore. First, a minimal intervention consisting of downstream reprofiling (LR) and the lowering of the Boyer culvert (C) are necessary to provide proper drainage in the residential area following a significant rainfall event. In terms of resilience against climate change, the addition of gutter disconnection (DG), a dry detention pond (DP), and upstream reprofiling (UR) would lead to the largest peak flow reduction. However, while the DG scenario is an ideal solution based on the multi-criteria analysis conducted, the addition of the DP and/or UR solutions would ultimately rely on the decision-makers scenario selection (i.e. socioeconomic, technical and environmental) as well as their respective weighting. It should be noted that the multi-criteria analysis performed in this case study allows the mitigation measures to be further investigated in a design phase and could serve as a preliminary study.

We believe that this methodology could be applied to other cities located in very small ungauged watersheds to select optimal solutions to increase their resilience against climate change. The degree of complexity and tools required to conduct

such an analysis will ultimately depend on the desired objectives as well as the municipality's capabilities and resources.

References

1. ILCRRS (2019) The causes and impacts of past floods in the Lake Champlain-Richelieu River basin—historical information on flooding. Technical report, 108p
2. MAMH (Ministère des Affaires municipales et de l'Habitation) (2020) Des solutions durables pour mieux protéger nos milieux de vie—plan de protection du territoire face aux inondations. Gouvernement du Québec, Québec, QC, Canada
3. Bélanger D, Gosselin P, Poitras P (2006) Changements climatiques au Québec méridional: perceptions des gestionnaires municipaux et de la santé publique. Gouvernement du Québec, Québec, QC Canada
4. AlQasimi E, Mahdi T-F (2019) Flooding of the Saguenay region in 1996: Part 1—modeling River Ha! Ha! Flooding. *Nat Hazards* 96:1–15
5. Martel J-L, Mailhot A, Brissette F (2020) Global and regional projected changes in 100-yr subdaily, daily, and multiday precipitation extremes estimated from three large ensembles of climate simulations. *J Clim* 33(3):1089–1103
6. Audet G et al (2019) Vers l'adaptation aux changements hydroclimatiques—Portrait du bassin versant de la rivière Saint-Régis, Montérégie, Rés-Alliance de la Saint-Régis, version finale, SCABRIC: Sainte-Martine, 36 p. + 1 annexe
7. Rawls W, Brakensiek D, Miller N (1983) Green-Ampt infiltration parameters from soils data. *J Hydraul Eng* 109(1):62–70
8. Te Chow V, Maidment DR, Mays LW (2010) Applied hydrology. McGraw-Hill Education, New York, NY, USA
9. MDDELCC (Ministère du Développement durable et de la Lutte contre les changements climatiques) (2017) Manuel de calcul et de conception des ouvrages municipaux de gestion des eaux pluviales. Gouvernement du Québec, Québec, QC, Canada
10. Mareschal B (2011) Promethee & GAIA software. <http://www.promethee-gaia.net/software.html>. Accessed 21 Mar 2021
11. Yan C, Rousse D, Glaus M (2019) Multi-criteria decision analysis ranking alternative heating systems for remote communities in Nunavik. *J Clean Prod* 208(20):1488–1497
12. Roy B, Bouyssou D (1993) Aide Multicritère à la Décision: Méthodes et Cas. Economica, Paris, France
13. Reynolds JE, Halldin S, Seibert J, Xu CY, Grabcs T (2019) Robustness of flood-model calibration using single and multiple events. *Hydrol Sci J* 65(5):842–853
14. Moriasi DN, Arnold JG, Van Liew MW, Bingner RL, Harmel RD, Veith TL (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans ASABE* 50(3):885–900

Hydrail One: A Standard Gauge Low Power Fuel Cell/battery Hybrid Hydrail Vehicle



M. Hegazi and G. Lovegrove

1 Introduction

In Canada, Class 1 railroads moved 413 billion tonne-kilometers in 2015, transporting more than \$C 280 billion worth of goods in support of the Canadian economy [1]. However, this rail also impacted the environment, as over 2.15 billion liters of fuel were burned by Class 1 railroads in Canada. Approximately 3.2% of this fuel were consumed by switcher locomotives [2]. While 3.2% is a small portion of the overall 2.15-billion-liter total, yard-switching operations contribute disproportionately higher levels of emissions (>6%) as compared to mainline freight or passenger railway operations. This is due to factors inherent to internal combustion engines that result in a relationship between the engine's efficiency and the dynamics of the power demand (duty cycle). It is reported that yard switching operations have almost twice as much NO_x and particulate matter emissions as mainline operations, more than twice the hydrocarbons emissions, and similar levels of carbon monoxide (CO) and Sulfur Oxide [2]. This leads us to conclude that a net reduction of approximately 6–7% of railway sector air pollution and greenhouse gases (GHG) can be achieved through implementing cleaner, zero-emission propulsion technology for yard switcher rail locomotives.

Transportation electrification is the area of research that aims to rectify the impact of the emissions. Most of this research is dedicated to road vehicles given their proliferation. However, recently there has been growing interest and research in zero emission railway vehicles as well. The pursuit of optimal powertrain designs through power source hybridization requires the use of locomotive driving cycles

M. Hegazi · G. Lovegrove (✉)
UBC School of Engineering, 1137 Alumni Avenue, Kelowna, BC V1Y1C9, Canada
e-mail: gord.lovegrove@ubc.ca

M. Hegazi
e-mail: mohamed.hegazi@ubc.ca

for power source sizing, [3–6] and energy management [4, 5]. In North America, two projects have aimed to build a prototype zero-emission switcher locomotive. Both occurred in the post-2008 period during a spike in NA oil prices. The first to be completed and tested in September 2009 was the NS 999, a battery all-electric locomotive by Norfolk Southern. The NS 999 was successful in performing all the required switching operations; however, the powertrain experienced routine failure of non-battery related electrical components in a testing period that lasted three months. Norfolk Southern concluded that Lithium-ion battery technology was the only battery chemistry suitable for switching operations. At USD\$1 million, NS also concluded that the cost was not economically justifiable [7]. The second successful project was led by Dr. Arnold Miller's Vehicle Projects Inc., which collaborated with a consortium of industry and government agencies including the US Army Corps of Engineers and BNSF. This consortium built the world's largest fuel cell land vehicle and tested it in 2010 [8]. The BNSF fuel cell/battery hybrid switcher was successful in performing all the required switching operations but proved to not be user friendly, which required significant crew training. Its lack of user-friendliness, coupled with subsiding oil prices, reduced the economic justification for its ever going into revenue service. It has now been dismantled to use for spare parts.

Hydrogen fuel cells convert chemical energy to electrical energy by combining hydrogen and oxygen in a process that generates electric current and water as the by-product in an emission free process. Performance of existing FC technology warrants their use in various industries. Previously, FC technology suffered from very low efficiencies; however, their current efficiency ranges from 40 to 50%. These new advancements motivated R&D into the usage of FC tech as alternatives to combustion engines. Additionally, the very high density of hydrogen makes FC tech an economically wise option in the long term. They have been successfully employed in the automotive sector despite the limitations that exist on their dynamic performance. This limitation is often resolved through the usage of hybrid powertrains with a more power dense energy source. Batteries and ultracapacitors have been proposed as potential secondary sources.

This paper details the work done to build a prototype low power standard gauge railway vehicle. The goal of this research is to prototype a hybrid hydrogen fuel cell/battery powertrain onboard a railway vehicle in order to develop the know how required for such a project and to study the impact of the propulsion dynamics on the different power sources.

2 Methodology

The Fairmont A4-D is a rear wheel driven motor car with a 2.3 L Ford engine at the front of the vehicle, a manual gear box, and a differential. The gear box is a 4 speed manual transmission with the highest gear ratio being an 8 for driving in reverse. The differential distributes the power from the engine and transmission to the wheels at an additional ratio of 3. This arrangement means that the engine speed

is reduced by a factor of 3–24 depending on the gears engaged. The reason for that being the typically poor torque output of combustion engines. The location of these components can be observed in Figs. 1 and 2 showing the before and after removal photos.

The weight of the speeder is distributed on the frame presented in Fig. 3. The chassis has eight main parts connected together with 0.75" steel grade hex head bolts;

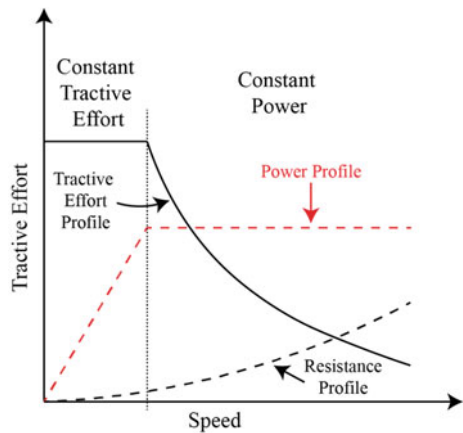
Fig. 1 The fairmont A4-D prior to any modifications



Fig. 2 The fairmont A4-D after the engine and transmission were removed



Fig. 3 Vehicle tractive effort, translational resistance and power as a function of vehicle's velocity



two frames on each side, a front and back bumper, and two connecting frames in the middle. The chassis is completely made of steel and is estimated to weigh 350 kg making it approximately 40–50% of the entire vehicle mass. The chassis is assumed to be capable of supporting the expected mass given that the retrofit is estimated to weigh lower than the original vehicle. The rear connecting frame has mounting brackets for the differential welded onto it. Similarly, the front connecting frame had supporting beams for the combustion engine. Minimal alterations are planned for these two frames since they will be repurposed to support two battery housing boxes, and a traction package. The speeder had a simplistic tread brake mechanically applied hand brake system. A system of levers connected the hand brake to the brake shoe which when applied by the driver would bring the shoe in contact with the wheel. The biggest drawback of this mechanism is that it was mechanically actuated relying on muscle power. In addition, the poor state of the mechanism including the brake shoes due to lack of maintenance (rust, etc....) made its reuse unsafe. The entire mechanism was abandoned.

The tractive effort of a railway vehicle is the force required to propel that vehicle and overcome any resistances to motion. A vehicle moving from rest is capable of a certain fixed starting tractive effort that is independent of power and is a function of the locomotive's mass and wheel/rail adhesion coefficient. Figure 3 presents the relationship between the forward translational force (tractive effort), retardation force (resistance), and supply power. The traction package (motor and gear combination) must be carefully selected so that the required tractive effort is generated, and the vehicle manages to move from rest. As the figure shows, the required tractive effort declines as the vehicle gains velocity making the starting tractive effort the highest required in any trip. This section will outline the method by which the traction package was sized and selected.

The starting tractive effort is calculated by the following equation:

$$TE_{start} = \mu Wn$$

where TE_{start} is the starting tractive effort in Newtons, μ is the adhesion coefficient, W is the estimated vehicle weight in Newtons, and n is the ratio of powered axles. The vehicle's weight (W) is the product of its mass and gravitational acceleration. A 20% adhesion was assumed, and a 50% ratio of powered axles since only half the axles were powered (the rear axle as demonstrated in Fig. 8).

$$TE_{start} = 0.2 \times 1169 \times 9.8 \times 0.5 \sim 1146 \text{ N}$$

Tractive effort is a translational force which needs to be translated to rotational torque at the wheel, differential, speed reducer, and motor output.

$$\tau_{wheel} = TE_{start} \times r_{wheel}$$

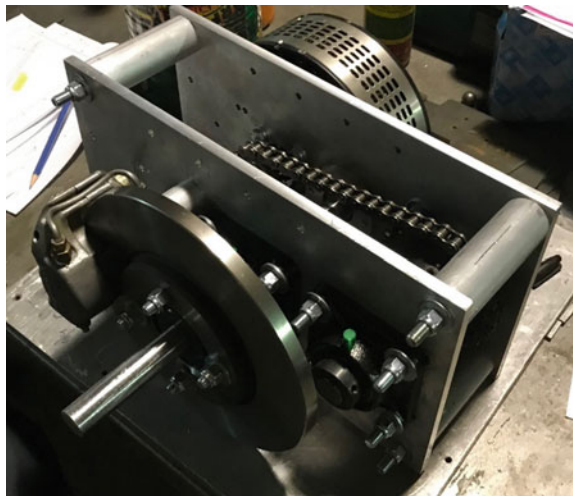
where τ_{wheel} is the starting torque at the wheels in Newton-meters, and r_{wheel} is the wheel radius in meters.

Given a wheel radius of 0.23 m, the combined torque at the rear wheels was 263.6 Nm. A differential gear ratio of 3, and assuming ideal gears, results in approximately 88 Nm required torque output of the traction package (motor and speed reducer).

A decision was made to use a permanent magnet direct current (PMDC) traction motor. The justification for this design choice was the high-power density, relatively constant torque, and simplicity of control. Being a permanent magnet motor allowed for high torque levels and compact size, and as a DC motor it eliminated the need for an inverter. This design choice ensured that the entire electrical system was DC which greatly simplified the electrical design. The motor was selected to meet the required torque output, overall size, and budget constraints. The D135RAG variant of the LEM-200 design by Lemco Motors was the motor of choice. It is rated for approximately 40 Nm of torque at approximately 83 Nm of peak torque which it could provide for a few seconds. It had a diameter of 21 cm, and a thickness of 14 cm excluding the shaft making it a very compact design.

A chain drive speed reducer was custom built to meet the rated torque of 88 Nm, and perhaps surpass it allowing for a higher mass to be moved. A gear ratio of 5 allowed for enough torque multiplication so that the traction motor could operate at below its rated torque at the estimated mass. A starting motor torque of approximately 18 Nm could be achieved at rated mass, and 36 Nm at double the mass. The custom-built reducer allowed for easy gear ratio modifications if higher speeds were required. This traction package will be directly connected to the rear differential eliminating the need for a drive shaft. This is a design improvement as it reduces weight and cost and increases system energy efficiency. Figure 4 shows the assembled traction package.

Fig. 4 Assembled traction package showing traction motor, chain drive speed reducer, brake disk and caliper



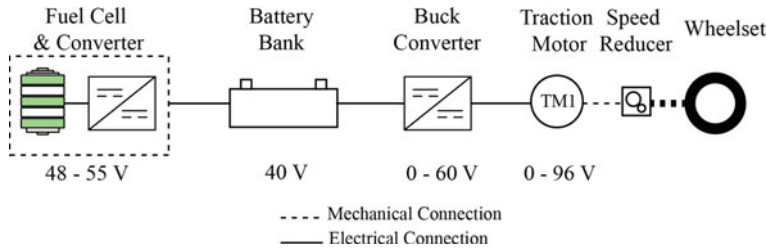


Fig. 5 The powertrain architecture of hydrail one

The powertrain of the retrofitted vehicle relies on a fuel cell (FC) prime mover supplying average power, and a battery bank supplying instantaneous power. As illustrated in Fig. 5, a 40-V battery bank will supply the traction motors through a motor controller (buck converter). The controller's efficiency is typically 98% delivering power at a reduced voltage and increased current. This reduces the average current supplied by the battery bank which enhances the life cycle of the batteries as well as reduce conductor size. The fuel cell system's voltage ranges from 48 to 55 V. Being at a higher voltage than the battery bank ensures correct current flow. The traction motor is rated at approximately 4000 RPM at 96 V. The motor controller is rated at 60 V peak output which ensures that the entire electrical system is considered low voltage. These design choices restrict the maximum velocity of the vehicle to just below 6 mph given the previously discussed gear ratios.

The power demand of the vehicle is the product of its tractive effort and velocity. This is translated to voltage and current at the power supply side after factoring the various efficiencies. The FC system is rated at 1.7 kW, the battery bank at 12 kW (40 V—300 Ah at 1C discharge), and the traction motor at 16 kW. Assuming a power demand equal to the rated power of the FC, a 7-h continuous operation time per tank is expected. However, if the power demand of the traction motor is higher than the rated power of the FC system, the vehicle will operate in charge depletion mode. The vehicle is built to operate in battery only mode if required, which at 12 kW draw will only last 1 h. However, the motor is expected to draw only 1–3 kW of power at maximum velocity (6 mph) allowing for 4–12 h of battery only operation. The peak power demand of the traction motor is approximately 34.3 kW which can be supplied by the battery bank at a 3C discharge rate, a recommended maximum by the manufacturer. These peak powers are typically rated for a 30 s duration before components start to fail. The design ensures that those levels are never reached as the conductors, circuit breaker, disconnect switch, fuse and diode are not rated for such high power (current).

The fuel cell system of choice for this project was the 1.7 kW FCgen H2PM system by Ballard Power. Designed as a backup power for critical power applications such as telecommunications sites, this system is relatively simple in comparison with other automotive grade systems. The simplicity stems from the reliance on ambient air for stack cooling unlike the dedicated cooling systems of other FCs. The system is also safer as it operates on lower pressure hydrogen gas regulating standard 200 bar(g)

tanks to 5 bar(g) as the hydrogen enters the valve block. The valve block can connect to 3 hydrogen cylinders, each one can provide rated power for approximately 7 h. The only replaceable item within the FC is the panel filter which is designed for lifetime operation. Diagnostics can be performed by monitoring the data communicated by the FC system over CAN and TCP/IP networks. The advertised system efficiency of the FC is 42% with a maximum airflow of 600 m³/h at maximum load and maximum ambient temperature.

A high-power density battery with superior temperature performance including cold discharging was needed for the purposes of this project. A Lithium Iron Phosphate chemistry rated at 3.3 V/300Ah was the system of choice. At approximately 1 kWh per a 10 kg cell, the energy density of these cells was around 100 Wh/kg. This density is lower than other lithium-ion chemistries and designs. Specifically, it is half that of the cylindrical NMC cells often used for electric vehicle (EV) applications such as the ones manufactured by Tesla. The improved temperate performance, cycle life, and reduced overall number of cells which translates to a simpler battery management system and wiring had a heavier weight in deciding which technology to adopt. These cells are rated for 5000–7000 cycles at 80%–70% depth of discharge values respectively.

The powertrain of the retrofitted vehicle relies on a fuel cell (FC) prime mover supplying average power, and a battery bank supplying instantaneous power. As illustrated in Fig. 5, a 40-V battery bank will supply the traction motors through a motor controller (buck converter). The controller's efficiency is typically 98% delivering power at a reduced voltage and increased current. This reduces the average current supplied by the battery bank which enhances the life cycle of the batteries as well as reduce conductor size. The fuel cell system's voltage ranges from 48 to 55 V. Being at a higher voltage than the battery bank ensures correct current flow. The traction motor is rated at approximately 4000 RPM at 96 V. The motor controller is rated at 60 V peak output which ensures that the entire electrical system is considered low voltage. These design choices restrict the maximum velocity of the vehicle to just below 6 mph given the previously discussed gear ratios.

The power demand of the vehicle is the product of its tractive effort and velocity. This is translated to voltage and current at the power supply side after factoring the various efficiencies. The FC system is rated at 1.7 kW, the battery bank at 12 kW (40 V—300 Ah at 1C discharge), and the traction motor at 16 kW. Assuming a power demand equal to the rated power of the FC, a 7-h continuous operation time per tank is expected. However, if the power demand of the traction motor is higher than the rated power of the FC system, the vehicle will operate in charge depletion mode. The vehicle is built to operate in battery only mode if required, which at 12 kW draw will only last 1 h. However, the motor is expected to draw only 1–3 kW of power at maximum velocity (6 mph) allowing for 4–12 h of battery only operation. The peak power demand of the traction motor is approximately 34.3 kW which can be supplied by the battery bank at a 3C discharge rate, a recommended maximum by the manufacturer. These peak powers are typically rated for a 30 s duration before components start to fail. The design ensures that those levels are never reached as

the conductors, circuit breaker, disconnect switch, fuse and diode are not rated for such high power (current).

3 Results

Figure 6 shows the as-built 3D CAD drawing of the vehicle while Fig. 7 shows the actual finished vehicle. We note the presence of the fuel cell at the front of the vehicle to improve the airflow into the fuel cell as the vehicle is driving forward. The exhaust vent of the fuel cell is channeled to the rear of the vehicle passing through the traction motor and gearbox to enhance heat distribution and ventilation of the moving parts. The batteries are hidden in two boxes beneath the floor of the vehicle. Each box contains 6 battery cells which evenly distributes the mass of the vehicle as well as maintain a very low center of gravity for stability. Behind the two seats sits a hydrogen cylinder rack capable of carrying three hydrogen tanks and secured by a combination of gravity and industrial straps.

At the time of writing this manuscript, the vehicle had not yet undergone on track testing. Off-track testing was conducted to confirm that the powertrain system was functional. The no-load load-side current was between 15 and 20 amperes. The source side current was not significantly higher at slightly above 20 amperes as presented in Fig. 8. Sudden spikes in source side currents coinciding with sudden drops in load side currents are a result of applying the handbrake as a test of the overall system response to sudden frictional braking.

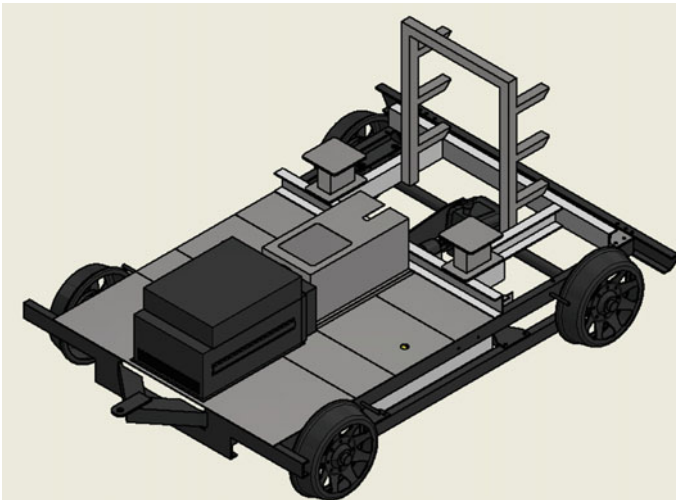


Fig. 6 3D CAD drawing of the retrofitted vehicle as-built before being tested



Fig. 7 Result of the retrofitted hydrail speeder

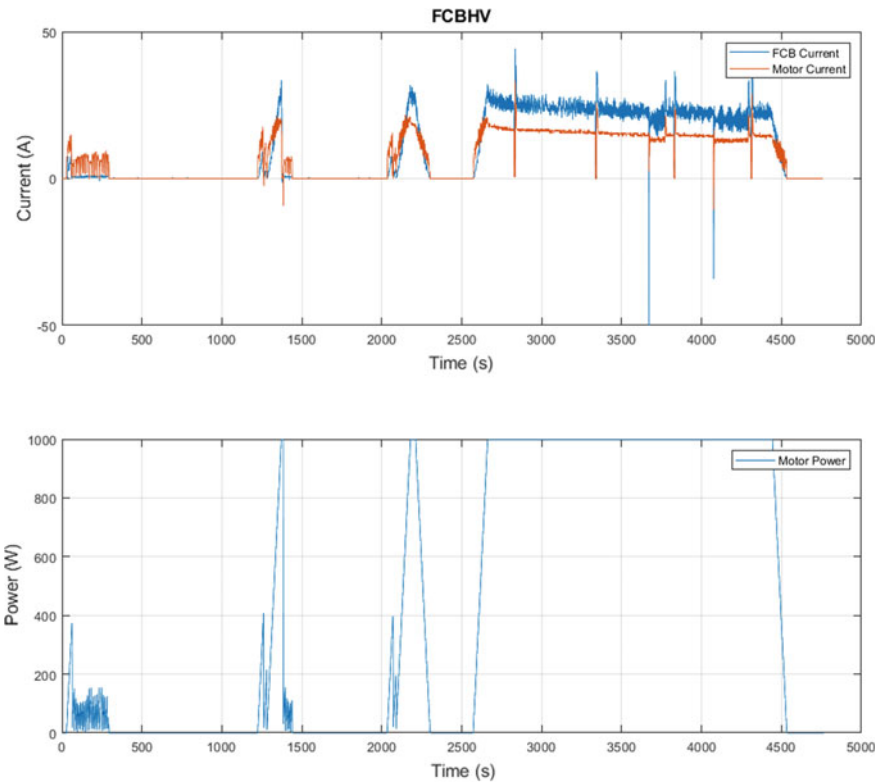


Fig. 8 Off-track test results in full hybrid mode showing the combined current from the fuel cell and battery bank in blue and motor current in red in the top figure. Bottom figure shows the control command signals for propulsion

4 Conclusions

A regiment for retrofitting a low power standard gauge railway speeder was presented. Details about the different components were revealed along with the corresponding justification. The rationale behind the selected powertrain configuration was discussed and the result of the retrofit was presented. Preliminary stationary (off-track) test result was presented showing that the powertrain was functional, and that the electrical system responded to command propulsion signals. On-track testing remains to be completed.

References

1. Railway Association of Canada (2016) Rail Trends
2. Railway Association of Canada (2015) Locomotive emissions monitoring program
3. Jaafar A, Sareni B, Roboam X (2013) A systemic approach integrating driving cycles for the design of hybrid locomotives. *IEEE Trans Veh Technol* 62:3541–3550
4. Jaafar A, Akli CR, Sareni B, Roboam X, Jeunesse A (2009) Sizing and energy management of a hybrid locomotive based on flywheel and accumulators. *IEEE Trans Veh Technol* 58:3947–3958
5. Kebriaei M, Sandidzadeh MA, Asaei B, Mirabadi A (2017) Component sizing and intelligent energy management of a heavy hybrid electric vehicle based on a real drive cycle. *Proc Inst Mech Eng Part F J Rail Rapid Transit* 231:122–132
6. Sarma U, Ganguly S (2018) Determination of the component sizing for the PEM fuel cell-battery hybrid energy system for locomotive application using particle swarm optimization. *J Energy Storage* 19:247–259
7. Office of research, development and technology (2017) Hybrid locomotive for energy savings and reduced emissions. Federal Railroad Administration, Washington, DC
8. Miller A, Hess K, Barnes D, Erickson T (2007) System design of a large fuel cell hybrid locomotive. *J Power Sources* 173(2):935–942

Suppression of Vortex Shedding from Bridge Pier Using Attached Splitter Plate



S. Samuel Li and S. Kheshtgar Darvazeh

1 Introduction

There have been numerous cases of bridge pier scour causing bridge failures. The resulting social and economic impacts are huge, as well-documented. Extensive research has been undertaken to understand pier scour processes and find effective ways to control scour. Previous researchers have performed laboratory experiments and numerical simulations of turbulent flow and bed scour around piers. Practical experimental results are empirical formulae of the maximum scour depth, as summarised in Brandimarte et al. [4]. As the flow approaches a pier, horseshoe vortices form in front of the pier as a result of the stagnating pressure gradient [19]. The vertical component of the flow removes bed sediments [11], in which the bed shear stress is a key variable. Horseshoe vortices were observed to form closer to a submerged pier than to a surface piercing one, and the bed shear stress increased with increasing level of submergence [20]. The scour depth is influenced by pier diameter, sediment size and flow depth [10] as well as by debris carried by a natural stream [9]. The horseshoe vortices have complicated distributions of velocity and turbulence. The region of the maximum turbulence intensity and maximum Reynolds shear stress was observed upstream of the main vortex, in the place of the largest turbulent eddy, and vortices evolved from a small vortex upstream of the pier and developed into multiple vortices at equilibrium [13]. In the wake of a cylinder, the number of vortices and flow patterns depend on the Reynolds number [7], and there is a strong turbulence [12].

Previously, some numerical studies deal with flow around a single pier using Direct Numerical Simulation [8], Large Eddy Simulation [24] or Reynolds-averaged

S. S. Li (✉) · S. Kheshtgar Darvazeh

Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, QC H3G 1M8, Canada

e-mail: sam.li@concordia.ca

© Canadian Society for Civil Engineering 2022

S. Walbridge et al. (eds.), *Proceedings of the Canadian Society of Civil Engineering*

Annual Conference 2021, Lecture Notes in Civil Engineering 250,

https://doi.org/10.1007/978-981-19-1065-4_28

simulation [2, 14, 15, 21, 25], and flow around multiple piers [5]. Other studies consider two-dimensional (2-D) flow [1, 3, 8] and three-dimensional flow [6, 22] around a cylinder fitted with a splitter plate of different lengths. The previous studies have a common limitation that the Reynolds number is low or the vertical dimension is missing (2-D). This paper aims at predicting 3-D flow around a cylinder with a splitter plate at the upstream or downstream side, at high Reynolds number.

2 Computational Methods

The computational model domain was a 3-D open channel, where a pier with or without a splitter plate was present at the centreline (Fig. 1). The flow entered the channel at the inlet, passed around the pier, and exited the channel at the outlet. The pier had a diameter of 10 cm and a height of 50 cm. The channel had a length of 40 cm upstream of the pier and 110 cm downstream. The channel had a width of 60 cm between two sidewalls. The channel had a horizontal bottom. The model domain was discretised into mesh of finite volume. The mesh size was 1 cm, with refinement for near-wall regions (pier surface and channel-bottom). The refinement ensured that the dimensionless wall distance (y^+) was about unity. Finite volume solutions to 3-D unsteady Reynolds-averaged Navier-Stokes equations were computed. The shear stress transport k - ω model [17] was used for turbulence closure. Sensitivity runs of the computational model were carried to confirm that the numerical results are independent of mesh configuration and turbulence closure.

The model domains had three types of boundaries: (1) lateral open boundary (inlet and outlet); (2) non-slippery solid walls (pier surface and channel bottom); (3) slippery walls (sidewalls and top lid). Conditions imposed at the boundaries were as follows: At the inlet, a vertical profile of the horizontal streamwise velocity was prescribed according to the 1/7th power law, and the horizontal velocity increased to a maximum velocity of 1.8 m/s at the top lid; the lateral and vertical velocities were set to zero, meaning that the inlet velocity was normal to the boundary. At the pier surface and the channel bottom, the three components of velocity were zero. On the sidewalls and top lid, the shear stresses were set to zero. The inlet conditions matched those of [6] to allow for data comparison. A total of six runs of the 3-D CFD model were carried out under conditions listed in Table 1.

3 Results

3.1 Turbulence Kinetic Energy

Run 1 (Table 1) used a pure cylinder without a splitter plate, serving as the base case for comparison with other runs. For Run 1, an example of TKE distributions

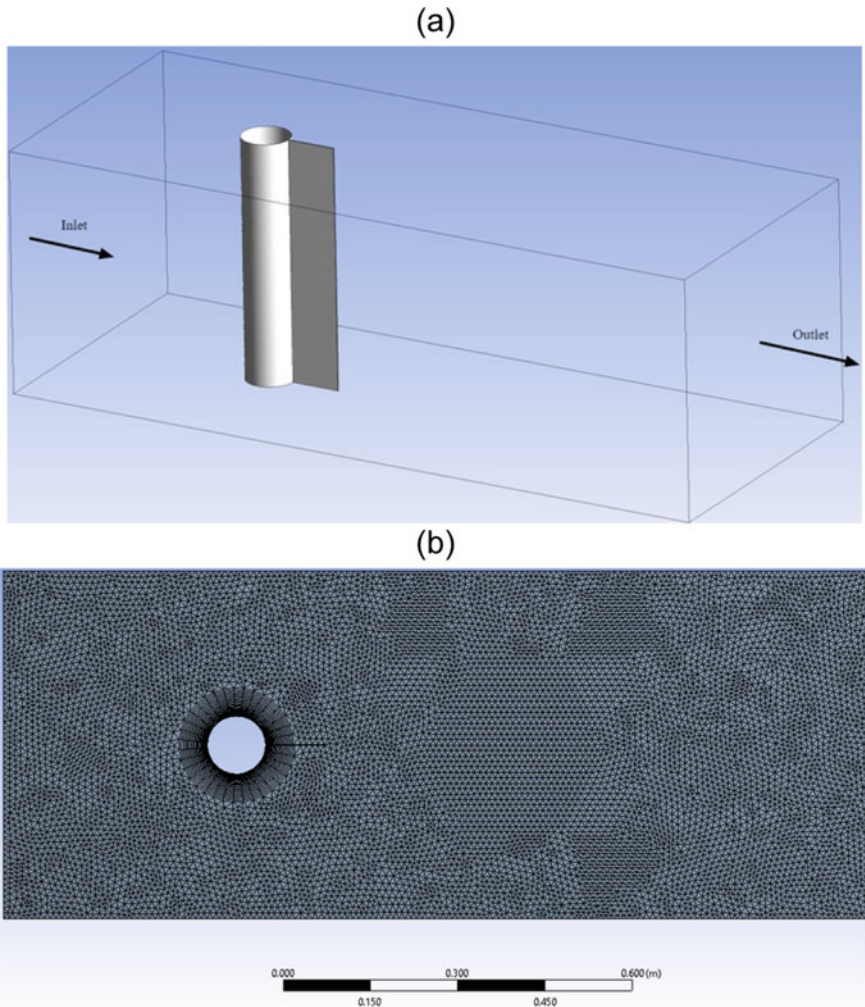


Fig. 1 a Computational model channel for the case where a splitter plate is attached to the pier at the downstream side; b plane view of the mesh

Table 1 Summary of conditions for numerical simulations. The time step was 0.0025 s. The number of nodes ranged from 848,957 to 857,614. The number of elements was between 3,978,663 and 4,002,085

| Run ID | Ratio of splitter length to pier diameter | Location of splitter plate |
|--------|---|----------------------------|
| 1 | 0 | Absent (pure cylinder) |
| 2 | 0.5 | Downstream |
| 3 | 0.6 | Downstream |
| 4 | 1 | Downstream |
| 5 | 0.5 | Upstream |
| 6 | 1 | Upstream |

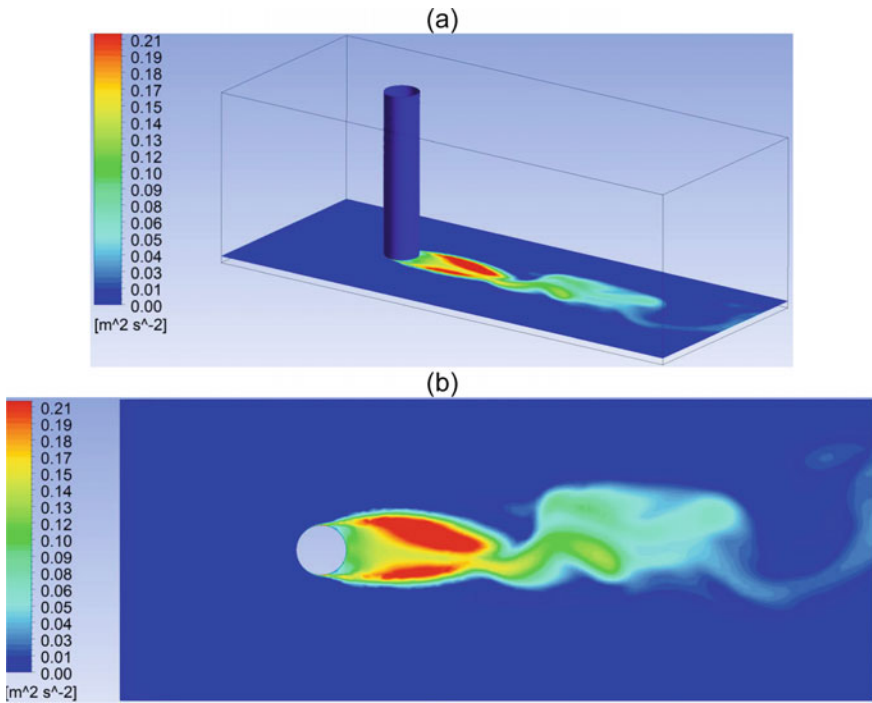


Fig. 2 Contours of TKE in the horizontal plane at a vertical distance of 2 cm from the channel-bottom for Run 1: **a** 3-D view; **b** plane view. The reynolds-averaged flow reached a state of quasi-equilibrium

is plotted in Fig. 2a, b for a horizontal plane near the channel-bottom. There was a region of high levels of TKE immediately downstream of the pier. Strong turbulence has the potential to scour an erodible bed.

In Fig. 3a–d, TKE distributions in the same plane as in Fig. 2 are plotted for Runs 2, 4, 5 and 6. A comparison of the TKE values in these figures with those in Fig. 2b shows that the splitter plate caused a decrease not only in the maximum TKE but also in the area of high TKE levels. The regions of predicted high TKE levels match sediment scour regions downstream of a pier as observed in existing experiments [9, 10]. In Fig. 3a, b, the splitter plate attached to the downstream side of the pier weakened turbulence from the base case (Fig. 2b), in particular in the wake. In Fig. 3c, the splitter plate attached to the upstream side was not as effective as the downstream splitter plate (Fig. 2a) in terms of TKE reduction. In Fig. 3d, the longer upstream splitter plate gave some improvement in terms of TKE reduction from the shorter upstream splitter plate (Fig. 3c).

The average of TKE values laterally across the pier width varied with longitudinal distance from the pier. The averaged TKE had a peak value of $0.142 \text{ m}^2/\text{s}^2$ for Run 1. The peak value was reduced by 14% to $0.12 \text{ m}^2/\text{s}^2$ for Run 5. For Runs 2, 4 and 6, the peak values were reduced by about 39% to $0.087 \text{ m}^2/\text{s}^2$. Run 4 gave a shift of

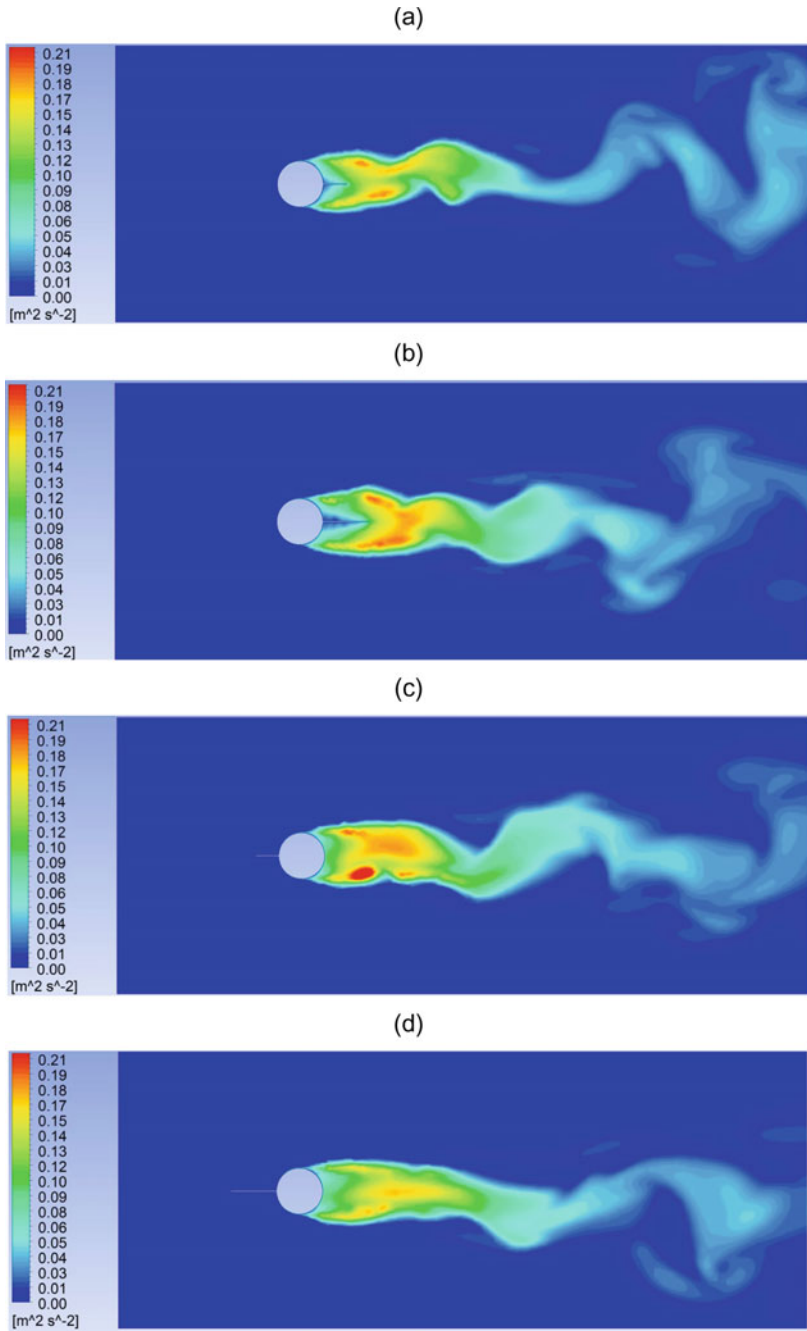


Fig. 3 Contours of TKE in the horizontal plane at 2 cm from the channel-bottom for: **a** Run 2; **b** Run 4; **c** Run 5; **d** Run 6. The Reynolds-averaged flow reached a state of quasi-equilibrium

the location of the peak value towards downstream and a gradual variation in TKE around the peak (Fig. 3b). Previously, [8, 18] demonstrated that at the ratio of the splitter length to the pier diameter equal to one, vortex formation shifted location towards downstream. Since the formation of vortices increases TKE, the results from this paper show some consistency with those of the previous studies.

3.2 Bottom Shear Stress

For the base case (Run1), contours of the bed shear stress at equilibrium are shown in Fig. 4. Just downstream of the pier, the flow had negative velocities pointing toward upstream, giving rise to negative values for the bed shear stress. The negative values were due to vortices in the wake region. Strong bed shear stresses in the longitudinal direction occurred on both sides of the upstream half of the pier. The locations are where horseshoe vortex forms around the base of the pier.

In Figs. 5a–d, contours of the bed shear stress at equilibrium are plotted for Runs 2, 4, 5 and 6. It has been checked that the equilibrium bed shear stress in the vicinity of the pier had little fluctuations in time. By comparing the bed shear stresses in these figures to those in Fig. 4, one notices three significant differences: (1) with either an upstream splitter plate or a downstream splitter, the strongest positive and negative shear stresses decreased from the base case; (2) the areas of the strong shear stresses shrank; (3) there was an overall reduction of bed shear stresses around the pier for cases with a splitter plate.

The average of bed shear stress values laterally across the pier varied with longitudinal distance from the pier. The average bed shear stress had a peak value of 11 N/m² for Run 1, which occurred immediately downstream of the pier. The peak values were reduced for Run 2, 4, 5 and 6. Run 4 gave the most significant reduction, with the peak value dropped by 33% down to 7.39 N/m² (Fig. 5b). For Run 4, the peak bed shear stress occurred immediately upstream of the pier, whereas for Runs 2, 5 and 6, the peak bed shear stresses appeared immediately downstream of the pier.

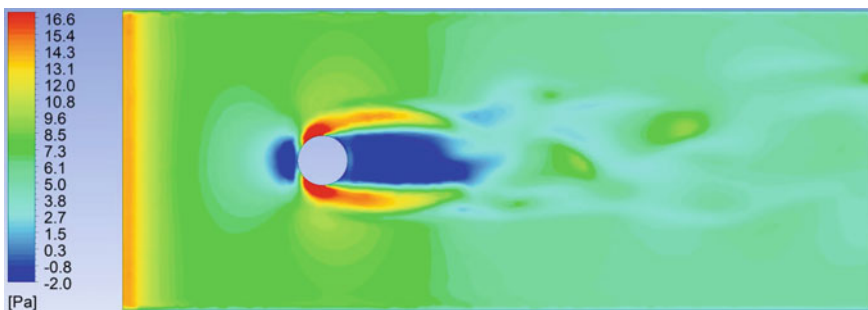


Fig. 4 Bottom shear stress for Run 1

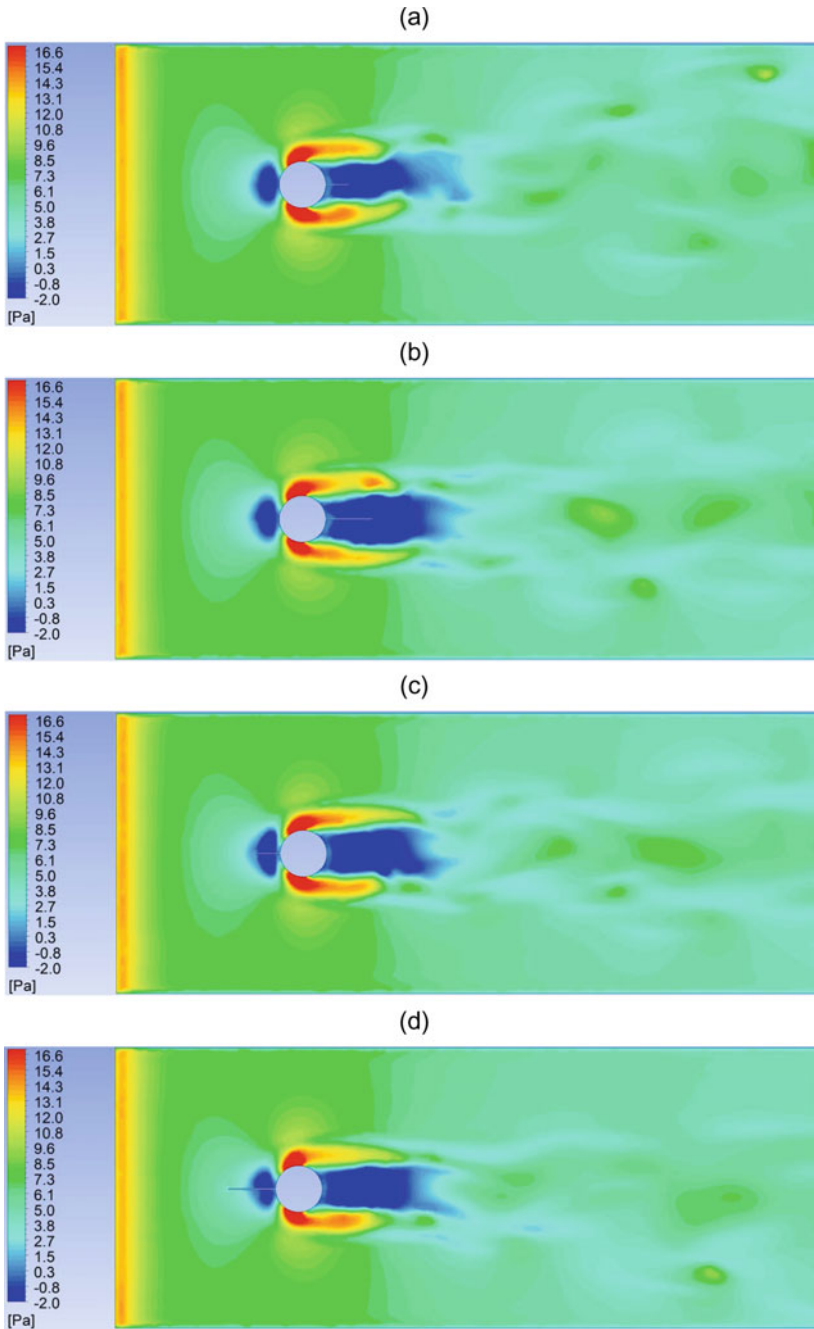


Fig. 5 Distributions of bed shear stress at equilibrium for: a Run 2; b Run 4; c Run 5; d Run 6

Runs 2, 5 and 6 gave approximately the same performance in terms of reduction in the peak bed shear stress. For all runs, the bed shear stresses decreased to a minimum value downstream of the peak value locations. In summary, Runs 2, 4 and 5 demonstrated the effectiveness of a splitter plate in achieving bed shear stress reductions. Further downstream, the bed shear stresses for all runs gradually increased to around 5 N/m^2 . The results for Run 4 (Fig. 5b) appeared to be the most desirable.

3.3 Cross-Channel Flow

Lateral flow in the cross-channel direction contributes to velocity shear and hence the production of turbulence. Also, lateral flow contributes to vibrations of bridge piers. For these reasons, it is desirable to suppress the cross-channel velocity as much as possible. The lateral velocity component in the centre plane is shown in Fig. 6 for

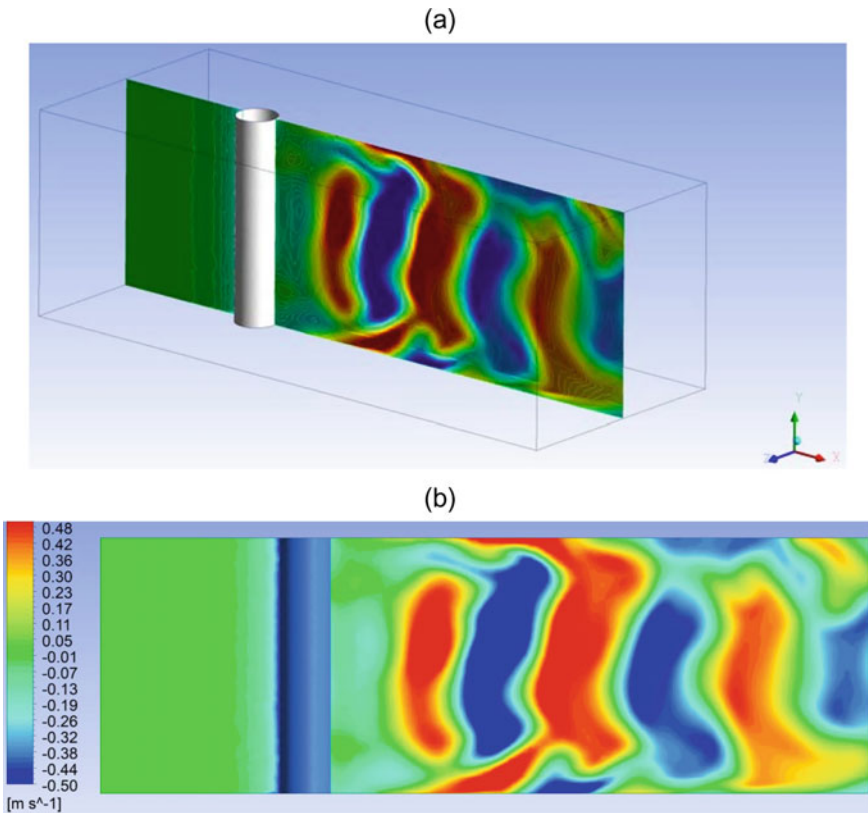


Fig. 6 Distribution of the cross-channel velocity component in the centre plane: **a** 3-D view; **b** elevation view

Run 1. The alternate patterns displayed large scale eddy motions and velocity shear in the interface between fluids flowing in the opposite directions. The shear extended almost over the entire water column. The strong lateral velocities were a few times of the inlet velocity entering the model domain.

The downstream splitter plate for Run 2 reduced the regions of strong lateral velocities (not shown). There was only a small region of high velocities near the top lid. No significant lateral motions were near the bottom or the pier. For Run 4, lateral flow downstream of the pier was suppressed (Fig. 7a), which was a substantial improvement from the results for Runs 1 and 2. Run 5 reduced the lateral velocities, compared to Run 1. A comparison of lateral velocities for Run 5 to those for Runs 2 and 4 showed that Run 5 was not as capable as Runs 2 and 4 in suppressing lateral velocities. For Run 5, strong lateral velocities appeared in some near-bottom regions. Run 6 gave better results (Fig. 7b) than Runs 2 and 5, in terms of suppressing lateral velocities. Between Runs 6 and 4, the performance was approximately the same. In conclusion, either an upstream or a downstream splitter plate with a length equal to the pier diameter (Runs 4 and 6) is highly capable of reducing vortices and lateral velocities downstream of the pier.

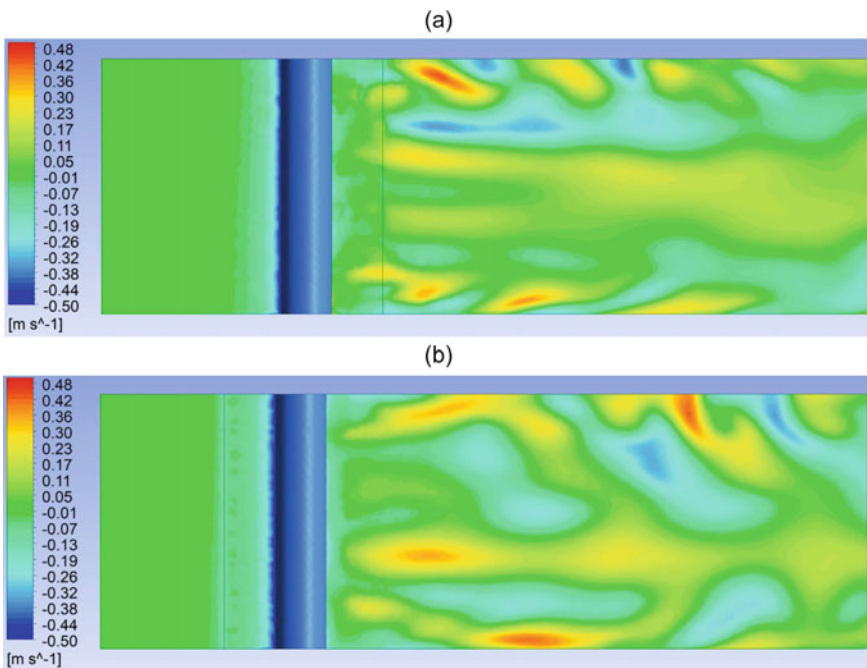


Fig. 7 Distributions of the cross-channel velocity component in the centre plane for: **a** Run 4; **b** Run 6

4 Discussion and Conclusions

The foregoing results of turbulence characteristics were computed using the shear stress transport k - ω model for turbulence closure. Computations of turbulent flow under identical hydraulic conditions were also performed using the standard k - ϵ model [16] and the k - ω model [23] for turbulence closure. Among the three turbulence models, the results extracted for a series of selected cells show consistency, leading to the conclusion that the foregoing results are independent of the choice of turbulence closure schemes. Note that the shear stress transport k - ω model is more sophisticated. Computations were also performed using a finer mesh (0.5 cm resolution) and a smaller time step (0.0025 s). It has been verified that the mesh of 1 cm resolution and time step of 0.0025 s (Table 1) produce results at a series of selected cells in consistency with those from the computations using the finer mesh and smaller time step.

This paper presents computational results of 3-D turbulent flow around a cylindrical pier. The computations cover both the case of a bare pier and cases of a pier fitted with a splitter plate on either the upstream or the downstream side. The key conclusions are that fitting a splitter plate on the downstream side (Run 4, see Table 1) effectively reduces the peak turbulence kinetic energy by 39%, shifts the location of the peak turbulence kinetic energy toward downstream (away from the pier), and reduces the peak bed shear stress by 33%. In the absence of a splitter plate (Run 1), spatially alternating high velocities in the cross-channel direction take place, which are stronger than the approach flow in the channel-centre plane. This is problematic. The presence of a splitter plate (Runs 2, 4, 5 and 6) produces an improvement. A splitter plate with a longitudinal length equal to the pier diameter (Runs 4 and 6) gives the best improvement of suppressing high lateral flow velocities.

The results are analysed to quantify the effect of a splitter plate on the drag and lift coefficients, eddy viscosity, eddy frequency, and vorticity. An upstream splitter plate (Runs 5 and 6) reduces the drag coefficient by 45%, being more effective than a downstream splitter plate. The computed drag coefficient for Runs 1 and 3 are in agreement with the results of Dai et al. [6]. A reduction of the drag coefficient means reduced drag forces exerted by the flow on piers. A downstream splitter plate (Runs 2 and 4) reduces the lift coefficient by about 80%. Lift forces would cause vibrations on piers and increase vortex shedding power. A downstream splitter plate with a longitudinal length equal to the pier diameter effectively reduces the strength and area of vorticity in the wake. It also has beneficial effect with respect to both drag and lift forces. Further study should perform physical model experiments to validate the numerical results reported in this paper and address practical challenges in using a splitter plate for mitigation of pier scour in the field.

Acknowledgements This study received financial support from NSERC through Discovery Grant held by S.S. Li.

References

1. Abdi R, Rezazadeh N, Abdi M (2017) Reduction of fluid forces and vortex shedding frequency of a circular cylinder using rigid splitter plates. *Eur J Comput Mech* 26(3):225–244. <https://doi.org/10.1080/17797179.2017.1306826>
2. Ali KHM, Karim O (2002) Simulation of flow around piers. *J Hydraul Res* 40(2):161–174. <https://doi.org/10.1080/00221680209499859>
3. An X, Song B, Tian W, Ma C (2019) Numerical research of flow past a circular cylinder with splitter plate at a subcritical reynolds number region. *J Shanghai Jiaotong Univ (Sci)* 24:113–121. <https://doi.org/10.1007/s12204-019-2045-y>
4. Brandimarte L, Paron P, Di Baldassarre G (2012) Bridge pier scour: a review of processes, measurements and estimates. *Environ Eng Manage J* 11(5):975–989
5. Da Silva BL, Luciano RD, Utzig J, Meier HF (2019) Analysis of flow behavior and fluid forces in large cylinder bundles by numerical simulations. *Int J Heat Fluid Flow* 75:209–226. <https://doi.org/10.1016/j.ijheatfluidflow.2019.01.006>
6. Dai S, Younis BA, Zhang H, Guo C (2018) Prediction of vortex shedding suppression from circular cylinders at high reynolds number using base splitter plates. *J Wind Eng Ind Aerodyn* 182:115–127. <https://doi.org/10.1016/j.jweia.2018.09.006>
7. Dargahi B (1989) The turbulent flow field around a circular cylinder. *Exp Fluids* 8:1–12
8. De Araujo LA, Schettini EBC, Silvestrini JH (2018) Direct numerical simulation of the flow around a cylinder with splitter plate: analysis for moderated reynolds numbers. *J Braz Soc Mech Sci Eng* 40:276. <https://doi.org/10.1007/s40430-018-1199-0>
9. Dias AJ, Fael CS, Nunez-Gonzalez F (2019) Effect of debris on the local scour at bridge piers. In: IOP conference series: materials science and engineering, vol 471, p 022024. <https://doi.org/10.1088/1757-899x/471/2/022024>
10. Ettema R, Kirkil G, Muste M (2006) Similitude of large-scale turbulence in experiments on local scour at cylinders. *J Hydraul Eng ASCE* 132(1):33–40. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2006\)132:1\(33\)](https://doi.org/10.1061/(ASCE)0733-9429(2006)132:1(33))
11. Graf WH (1998) *Fluvial hydraulics*. Wiley, Chichester, U.K.
12. Graf W, Istiarto I (2002) Flow pattern in the scour hole around a cylinder. *J Hydraul Res* 40(1):13–20. <https://doi.org/10.1080/00221680209499869>
13. Guan D, Chiew YM, Wei M, Hsieh SC (2019) Characterization of horseshoe vortex in a developing scour hole at a cylindrical bridge pier. *Int J Sedim Res* 34(2):118–124. <https://doi.org/10.1016/j.ijsrc.2018.07.001>
14. Huang W, Yang Q, Xiao H (2009) CFD Modeling of scale effects on turbulence flow and scour around bridge piers. *Comput Fluids* 38(5):1050–1058. <https://doi.org/10.1016/j.compfluid.2008.01.029>
15. Jia Y, Altinakar M, Guney MS (2018) Three-Dimensional numerical simulations of local scouring around bridge piers. *J Hydraul Res* 56(3):351–366. <https://doi.org/10.1080/00221686.2017.1356389>
16. Jones W, Launder B (1972) The prediction of laminarization with a two-equation model of turbulence. *Int J Heat Mass Transf* 15(2):301–314. [https://doi.org/10.1016/0017-9310\(72\)90076-2](https://doi.org/10.1016/0017-9310(72)90076-2)
17. Menter FR (1994) Two-Equation eddy-viscosity turbulence models for engineering applications. *AIAA J* 32(8):1598–1605. <https://doi.org/10.2514/3.12149>
18. Ozkan GM, Firat E, Akilli H (2017) Passive flow control in the near wake of a circular cylinder using attached permeable and inclined short plates. *Ocean Eng* 134(April):35–49. <https://doi.org/10.1016/j.oceaneng.2017.02.014>
19. Raudkivi AJ (1991) Scour at Bridge Piers. In: Breusers HNC, Raudkivi AJ (eds) *Scour: hydraulic structures design manual*, vol 2. Taylor and Francis, New York, NY, USA, pp 61–98
20. Sadeque MA, Rajaratnam N, Loewen MR (2008) Flow around cylinders in open channels. *J Eng Mech ASCE* 134(1):60–71. [https://doi.org/10.1061/\(ASCE\)0733-9399\(2008\)134:1\(60\)](https://doi.org/10.1061/(ASCE)0733-9399(2008)134:1(60))

21. Salaheldin TM, Imran J, Chaudhry MH (2004) Numerical modeling of three-dimensional flow field around circular piers. *J Hydraul Eng ASCE* 130(2):91–100. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2004\)130:2\(91\)](https://doi.org/10.1061/(ASCE)0733-9429(2004)130:2(91))
22. Vu H, Ahn J, Hwang J (2016) Numerical investigation of flow around circular cylinder with splitter plate. *KSCE J Civ Eng* 20(6):2559–2568. <https://doi.org/10.1007/s12205-015-0209-3>
23. Wilcox DC (1993) *Turbulence modelling for CDF*. DCW Industries Inc., La Canada, CA, USA
24. Yen CL, Lai JS, Chang WY (2001) Modeling of 3D flow and scouring around circular piers. *Proc National Sci Counc Repub China* 25(1):17–26
25. Zaid M, Yazdanfar Z, Chowdhury H, Alam F (2019) Numerical modeling of flow around a pier mounted in a flat and fixed bed. *Energy Procedia* 160(February):51–59. <https://doi.org/10.1016/j.egypro.2019.02.118>

A Comparative Study of Mean and Turbulent Flow Fields Around a Boulder Within Boulder Array and an Individual Boulder



R. Pierce, A. Golpira, and A. B. M. Baki

1 Introduction

The presence of large roughness elements (LRE), such as boulders, considerably varies surrounding mean and turbulent flow characteristics [1]. This variation may affect sediment transport dynamics and habitat complexity in the reaches containing in-stream boulders. Therefore, boulder placement, as a common method, has been applied to enhance degraded streams and create nature-like fish passages. In nature and practical application, in-stream boulders can be found individually or within an array. Flow in the vicinity of an individual boulder is different from the flow characteristics around a single boulder within a boulder array [2]. This difference can be associated with the flow retardation and backwater effects [1] due to a boulder array.

Several studies have investigated the flow characteristics around single isolated LREs and individual LREs within an array of LREs [1–4], and others). Reference [1] found that the velocity defect law was able to describe mean flow around a fully submerged boulder within a boulder array under an isolated-roughness flow regime. Reference [4] investigated mean and turbulent flow fields in the wake of a single cube. They reported changes in the spatial distribution of the Reynolds shear stress (RSS) and turbulent kinetic energy (TKE) by varying the relative depth of the cube. Reference [3] studied mean velocity and turbulence intensity profiles around a boulder within an array of boulders under a wake-interference regime over a smooth bed. They proposed relationships to describe streamwise turbulence intensities around the boulder. Reference [2] compared mean velocity and turbulence intensity profiles as well as bed shear stress around a fully submerged individual boulder with a boulder within an array. They found that for both conditions, boulders

R. Pierce (✉) · A. Golpira · A. B. M. Baki
Department of Civil and Environmental Engineering, Clarkson University, Potsdam, NY, USA
e-mail: piercera@clarkson.edu

have a local effect on the mean velocity profiles. They also reported a more significant decrease in the turbulence intensities due to the boulder array [2].

The objective of this experimental study is to compare mean and turbulent flow profiles in the vicinity of a single boulder, with those around a boulder within an array of boulders under an isolated-roughness flow regime over a rough bed. Stream-wise velocity profiles along with the turbulence intensity, TKE, and RSS profiles were investigated to perform the comparison. The velocity and turbulence intensity profiles also were compared with the logarithmic law and exponential decay functions, respectively. The results of this study can be used to investigate the implications of individual and boulder arrays for the surrounding sediment transport and aquatic habitat suitability.

2 Methodology

2.1 Experimental Setup

Experiments were conducted in the Ecohydraulics Flume, a 13-m-long, 0.96-m-wide rectangular water-recirculating flume located in the Water Resources Laboratory at Clarkson University. The flume bed consisted of a 5-cm layer of fine gravel, with $d_{10} = 3.8$ mm, $d_{50} = 6.1$ mm, and $d_{90} = 8.8$ mm, where d_p refers to the grain size for which $p\%$ of particles are finer. To prevent erosion, especially around the boulders, the gravel was glued to the bed. Experiments were carried out starting at $x = 6.82$ m, where a fully turbulent flow was developed. Here, x refers to the distance from the flume entrance. Also, all measurements were taken along $y = 30$ cm, where y indicates the distance from the right sidewall (standing toward downstream). For experiments with boulder placement, natural semi-spherical boulders with an equivalent diameter of $D = 12$ cm were used. Hereafter, the point at $x_0 = 6.94$ m is considered as the origin point, therefore, the location of the other points will be expressed relative to x_0 and D . For instance, $\times 1 = -1D$ refers to the station 12 cm upstream of the x_0 and $\times 2 = 1D$ indicates the station 12 cm downstream of the x_0 .

Three-dimensional velocity time series were measured by using a Vectrino Plus acoustic Doppler velocimeter with a sampling frequency of 100 Hz for a duration of 120 s. A carriage with a precise automatic positioning system carried the ADV to the desired position in longitudinal, transverse, and vertical directions. It should be noted that the Vectrino Plus obtains the velocity from a measuring cylindrical volume located 5 cm below the probe. Such an arrangement prevents near-surface measurements. A measurement volume height of 5.5 mm was used, except for the near-bed points at which the measurement volume height was set to 1.9 mm.

In order to measure velocity time series, three scenarios were designed: no-boulder scenario, S1, single-boulder scenario, S2, and boulder array scenario, S3. At all scenarios, a constant flow rate, $Q = 75$ L/s, was used. For S1, to obtain an undisturbed velocity time series, no boulders were placed on the flume bed. Measurements were

taken only at a single station, $\times 1 = -1D$ and 6 points over the bed. The vertical distance between points varied between 0.5 to 1.2 cm covering the relative depth, $z/H = 0.05-0.47$, where z is the vertical distance from the bed and H is the average flow depth. For S2, a single boulder was placed at x_0 , and velocity time series were obtained at four stations including boulder upstream at $\times 1 = -1D$, and boulder downstream at $\times 3 = 1.5D$, $\times 4 = 2D$, and $\times 5 = 3D$. Similar to S1, at each station, 6 points were measured over the flow depth covering $z/H = 0.04-0.46$. It is worth mentioning that a strong jump occurred immediately downstream of the boulder, preventing velocity profile measurements at $\times 2 = 1D$ due to insufficient depth. For S3, boulders were placed in an array throughout the flume with the longitudinal and transverse spacing of $6D$ and $3D$, respectively. This arrangement resulted in an areal density of 2%, which corresponds to an isolated-roughness flow regime. Under this flow regime, boulders' wakes are not expected to interfere with each other [5]. Measuring stations located at x_0 , at $\times 1 = -1D$, $\times 2 = 1D$, $\times 3 = 1.5D$, $\times 4 = 2D$ and $\times 5 = 3D$. To obtain velocity profiles at each station, measurements were conducted at 7 points over the flow depth corresponding to $z/H = 0.04-0.51$. Figure 1 shows the boulder arrangement for S3 and the location of the target boulder, around which measurements were taken. Figure 2 illustrates the described measuring stations around the target boulder. Table 1 provides a brief description of each scenario including flow depth at each station and the number of placed boulders for each scenario.

To ensure sufficient data quality, Ref. [6] phase-space threshold method, modified by Ref. [7], was implemented to remove spikes from velocity time series. Furthermore, signal correlation (COR) and signal-to-noise ratio (SNR) were used to filter

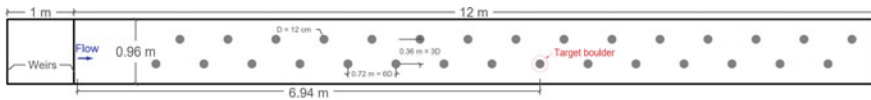


Fig. 1 Plan view of the Ecohydraulics Flume with the boulder arrangement for S3. The target boulder for the detailed measurements has been shown with a red circle

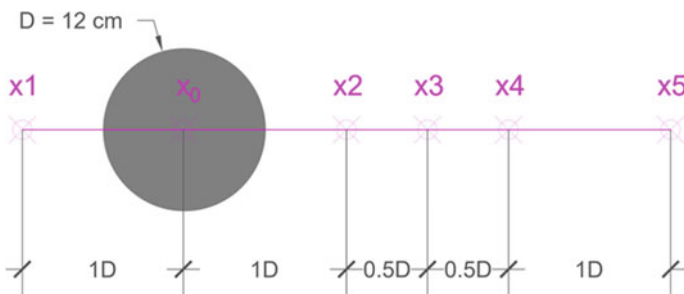


Fig. 2 Measuring stations around the target boulder for S2 and S3. Note that for S1 the only measuring station was $\times 1$ and for S2, station $\times 2$ was excluded from the other stations

Table 1 General description of the scenarios

| Scenario | Flow rate [L/s] | Flow depth [cm] | | | | | Number of placed boulders |
|--------------------------|-----------------|-----------------|------|------|------|------|---------------------------|
| | | × 1 | × 2 | × 3 | × 4 | × 5 | |
| S1 (Undisturbed flow) | 75 | 9.7 | 9.7 | 9.7 | 9.7 | 9.7 | 0 |
| S2 (Individual boulder) | | 11.9 | | 9.7 | 11.4 | 9.7 | 1 |
| S3 (Boulder array) | | 13.1 | 11.4 | 12.3 | 13.2 | 12.8 | 32 |

out low-quality data. Although $COR \leq 70\%$ has been generally used to eliminate poor data, in highly turbulent flow, like this study, lower thresholds can be used to prevent losing a large amount of data while still providing acceptable data quality [8]. In this study, a filtering criterion $COR \leq 55\text{--}65\%$ and $SNR \leq 15$ dB was used to filter out bad data. This resulted in keeping an average of around 69% of data for all scenarios.

2.2 Mean and Turbulent Flow Parameters

All the mean and turbulent profiles were normalized using the shear velocity, u_* , which was found from the following equation [9, 10]:

$$\tau = \rho u_*^2 \left(1 - \frac{z}{\delta}\right) \cong -\rho \overline{u'w'} \quad (1)$$

Here, $-\rho \overline{u'w'}$ represents the measured RSS on the vertical plane, and u' and w' represent the streamwise and vertical fluctuating velocities, respectively. Also, δ refers to the distance from the reference bed to the point of the maximum velocity.

The logarithmic law was used to evaluate streamwise velocity profiles in this study. The logarithmic law can be found from the following equation [9]:

$$\frac{u}{u_*} = \frac{1}{\kappa} \ln\left(\frac{z + z_0}{k}\right) + B_r \quad (2)$$

where u is the time-averaged streamwise velocity at a specific point; κ is the von Karman constant with a value of 0.40; and z_0 is the distance between the crest of the rough bed and the zero-velocity point, determined by using $z_0 = 0.20k$; $k = 0.26$ cm is the characteristic bed roughness height; and $B_r = 8.47 \pm 0.9$ is a constant given by Ref. [9] for flat rough beds.

The streamwise and vertical turbulence intensity profiles were compared with the proposed exponential decay functions by Ref. [10] (Eq. 3), Ref. [9] (Eq. 4), and Ref.

[3] (Eq. 5):

$$\frac{u'_{rms}}{u_*} = 2.3 \exp\left(-\frac{z}{H}\right), \quad \frac{w'_{rms}}{u_*} = 1.27 \exp\left(-\frac{z}{H}\right) \quad (3)$$

$$\frac{u'_{rms}}{u_*} = -2.04 * \exp\left(-0.97 \frac{z}{H}\right), \quad \frac{w'_{rms}}{u_*} = 1.14 \exp\left(-0.76 \frac{z}{H}\right) \quad (4)$$

$$\frac{u'_{rms}}{u_*} = \exp\left(-0.57 \frac{z}{H}\right) \quad (5)$$

where u'_{rms} and w'_{rms} refer to the root mean square of streamwise and vertical velocity fluctuations, respectively.

Finally, the TKE per unit mass, K , was found from the following equation [11]:

$$K = \frac{1}{2} \left(u'^2_{rms} + v'^2_{rms} + w'^2_{rms} \right) \quad (6)$$

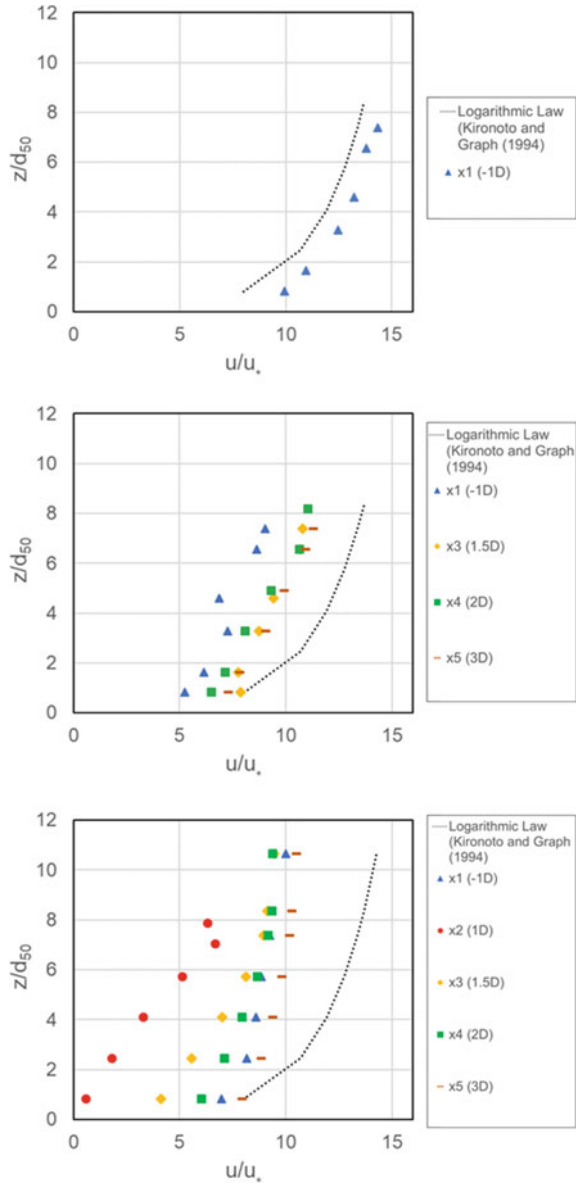
where v'_{rms} stands for the root mean square of transverse velocity fluctuations.

3 Results and Discussion

3.1 Velocity Profiles

Figure 3 shows the normalized streamwise velocity profiles at the stations around the boulder for each scenario. For S1, the shape of the measured profile was similar to the logarithmic law; however, values of the measured profile slightly deviated from the logarithmic law and shifted toward higher velocities. For S2 and S3, except for S3 at $\times 2$, the shape of velocity profiles can be described by the logarithmic law but the profiles were shifted toward lower velocities. For S3, i.e., in the presence of a boulder array, at $\times 2$ velocity significantly decreased over the whole measured depth. Near-zero velocities near the bed at $\times 2$ represented the reattachment point close to $x = 1D$. The reattachment length was in agreement with findings of [3] for a wake-interference regime and was lower than the reported length by Ref. [1] for a boulder under the isolated-roughness flow regime. Similar to Tsakiris' [2] findings, the velocity profile of $\times 3$ for S3 showed a parallel shift toward lower velocities in comparison with the same velocity profile for S2. It can be caused by the higher form drag due to the boulder array (Tsakiris 2014). At further downstream stations, i.e., $\times 4$ and $\times 5$, velocity profiles for both S2 and S3 showed a reasonable degree of similarity; however, velocities closer to the middle ($z/d_{50} > 6$) were generally lower for S3. At $\times 1$, the upstream stations, velocities for S3 were generally lower than S2 velocities.

Fig. 3 Normalized streamwise velocity profiles for all scenarios and stations



3.2 Turbulence Intensity Profiles

Figure 4 shows the normalized streamwise and vertical turbulence intensity profiles, u'_{rms}/u_* , and w'_{rms}/u_* , respectively, for all the scenarios around the boulder. The measured profiles were compared with the proposed exponential decay functions

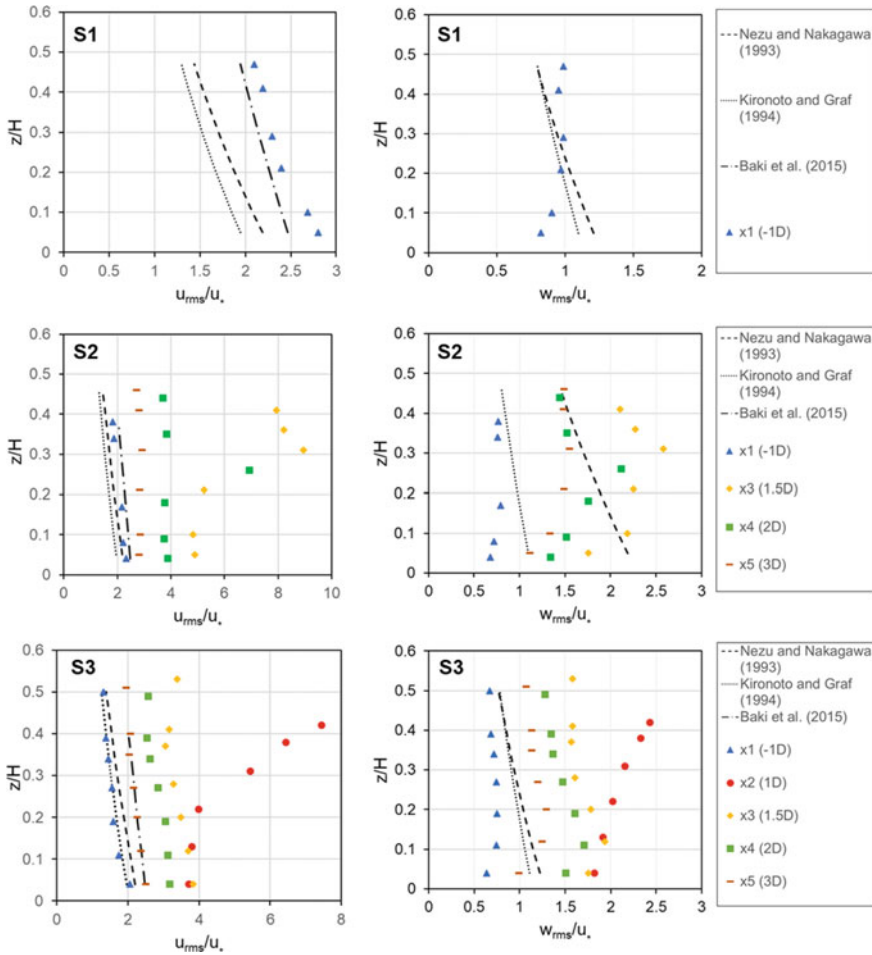


Fig. 4 Normalized streamwise (left column) and vertical (right column) turbulence intensity profiles for all scenarios and stations

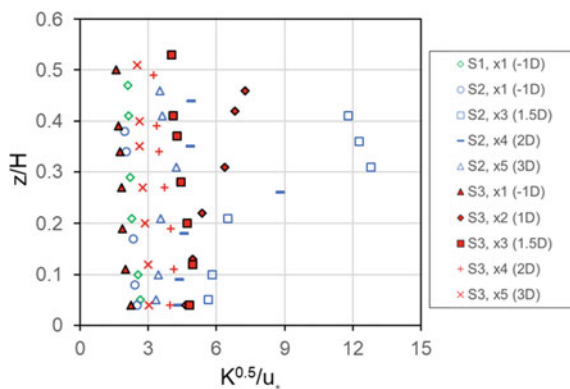
of [9, 10], and [3]. For S1, the measured u'_{rms}/u_* profile followed the shape of exponential decay functions, and the proposed profile by Ref. [3] described u'_{rms}/u_* values more precisely. For w'_{rms}/u_* , the measured profile of S1 showed lower values in the inner layer ($z/H < 0.2$) and did not follow the proposed profiles by Refs. [9, 10] in that region. The profiles of u'_{rms}/u_* for both S2 and S3 at $\times 1$ (the upstream station) followed the exponential decay functions. For S2, in the wake of the boulder at $\times 3$, the profile of u'_{rms}/u_* deviated from the exponential decay functions in the outer layer ($z/H > 0.2$). It can be attributed to the hydraulic jump immediately downstream of the boulder. As flow approached the further downstream stations, i.e., $\times 4$ and $\times 5$, u'_{rms}/u_* the turbulence dampened in the outer layer and profiles started to return to the shape of the exponential decay profiles; however, they showed

higher values relative to the common proposed functions. For S3, profiles of u'_{rms}/u_* at $\times 2$ highly deviated from the exponential decay function in the outer layer, due to high turbulence in this region. At $\times 3$ and $\times 4$, the profiles showed higher values than the commonly proposed profiles but could be described by the exponential decay functions. At $\times 5$, u'_{rms}/u_* could be perfectly described by Ref. [3] proposed function. The findings of this study were in disagreement with [2] results, in which all the profiles for both individual and array scenarios showed lower values compared to the [10] proposed profile. The results also indicated deviation of the streamwise turbulence intensity profiles from the [10] profile in the inner layer for both individual and array scenarios while in this study the profiles for both scenarios followed the shape of exponential decay functions, including [10] function, in the inner layer. Besides, the values of u'_{rms}/u_* for S3 were generally smaller relative to S2, implying the homogenizing effect of boulder arrays, as indicated by Tsakiris (2014). For S2, the profiles of w'_{rms}/u_* , deviated from the exponential decay functions in the inner layer, especially at downstream stations. For S3, the profile of w'_{rms}/u_* at $\times 2$ completely deviated the exponential decay functions and the values increased at higher z/H . For the other stations of S3, similar to S2, they deviated from the exponential decay functions in the inner layer; however, this deviation was less significant compared to the S2 scenario. In the wake of the boulder, the maximum w'_{rms}/u_* for S2 occurred around $z/H = 0.25$ while for S3 it occurred at lower z/H with a smaller magnitude (except for $\times 2$). Greater values of w'_{rms}/u_* for S2 can be attributed to the stronger upwelling region downstream of the boulder compared to the S3 [4].

3.3 Turbulent Kinetic Energy Profiles

Figure 5 shows the normalized TKE profiles, $K^{0.5}/u_*$, for all the scenarios and stations. It can be seen that for the undisturbed flow, S1, $K^{0.5}/u_*$ slightly decreased at higher z/H . At the upstream station of both S2 and S3, $K^{0.5}/u_*$ profiles followed the profile of undisturbed flow. For the individual boulder scenario, S2, $K^{0.5}/u_*$

Fig. 5 Normalized TKE profiles for all the scenarios and stations



values $\times 3$ and $\times 4$ in the outer layer highly increased then gradually decreased as z/H increased. As downstream distance increased, the $K^{0.5}/u_*$ profiles began to return to the shape of the undisturbed flow profile while still having a shift toward higher values. For the boulder array scenario, S3, except at $\times 2$, all the other $K^{0.5}/u_*$ profiles at downstream stations followed the undisturbed flow profile with a parallel shift toward higher values. This shift toward higher values decreased as downstream distance increased. For S3, immediately downstream of the boulder, at $\times 2$, $K^{0.5}/u_*$ consistently increased at a higher z/H showing a different profile shape compared to the disturbed flow in both inner and outer layer.

3.4 Reynolds Shear Stress Profiles

Figure 6 shows profiles of the normalized RSS, $-\overline{u'w'}/u_*^2$, for all the scenarios and stations. Except at $\times 3$ for S2, and $\times 2$ for S3, the rest of $-\overline{u'w'}/u_*^2$ profiles remained in a band between 0 and 2 and showed a similar profile shape; a slight general decrease in $-\overline{u'w'}/u_*^2$ by increasing z/H . At $\times 2$ for S3, the $-\overline{u'w'}/u_*^2$ values increased in the outer layer and at $\times 3$ for S2, $-\overline{u'w'}/u_*^2$ values consistently increased as z/H increased in both inner and outer layers. In the wake of the boulder, the shape of profiles for S3 was in agreement with the findings of [11] for an intermediately submerged boulder array under a wake-interference regime. In the near-bed region ($z/H < 0.1$) of the boulder's wake, the $-\overline{u'w'}/u_*^2$ values of S3 were smaller relative to the S2. This implied that the presence of a boulder array reduced the bed shear stress more significantly compared to the individual boulder scenario. Reference [2] also indicated that a collective array of boulders noticeably reduced bed shear

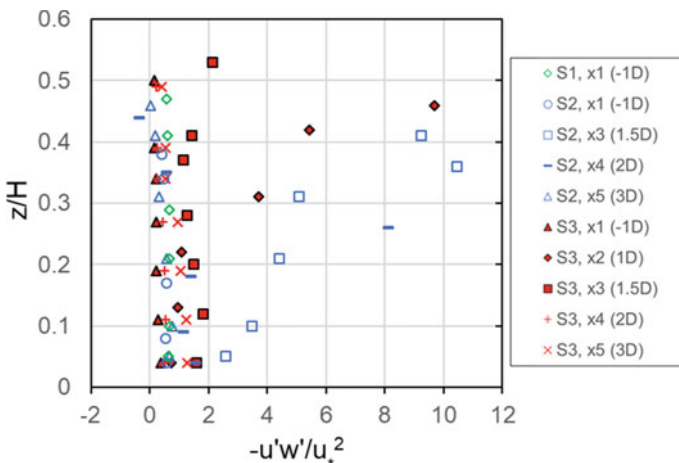


Fig. 6 Normalized RSS profiles for all scenarios and stations

stress and ignoring this reduction can lead to overestimation of bed shear stress and subsequently bedload transport.

4 Conclusions

Understanding the mean and turbulent flow characteristics due to the presence of LREs provides insightful information about the surrounding sediment transport and aquatic habitat. Effects of individual and boulder arrays on the surrounding flow field can be distinct. This study compared mean and turbulent flow profiles around an individual boulder (scenario S2) and a boulder within a boulder array under the isolated-roughness flow (scenario S3). The most important findings are summarized below:

1. Although the logarithmic law overestimated streamwise velocities for both S2 and S3, it was able to describe the shape of profiles around the boulder. The velocity profiles for S3 showed a parallel shift toward lower values compared to the S2.
2. Normalized streamwise turbulence intensity profiles followed the shape of exponential decay function in the inner layer for both S2 and S3 while in the wake of the boulder deviated from those functions in the outer layer up to $x/D = 2$. The common exponential decay functions generally underestimated the values of normalized streamwise turbulence intensity for both scenarios.
3. Normalized vertical turbulence intensity profiles deviated from the exponential decay function in the inner layer, with a higher deviation for S2, and could not describe properly the common exponential decay functions.
4. Normalized TKE profiles for S2 and S3 deviated from the profiles of an undisturbed flow (S1) in the boulder's wake up to $x/D = 2$ and 1.5, respectively. Also, they showed higher normalized TKE values compared to the S1 all over the depth.
5. The normalized RSS profiles showed a similar trend except for the closest downstream stations for both S2 and S3. Also, S3 resulted in generally lower normalized RSS values close to the bed in comparison with S2. It indicates that the effects of boulder arrays on sediment transport rate and local erosion and deposition should not be considered the same as an individual boulder for design purposes.

References

1. Papanicolaou AN, Kramer CM, Tsakiris AG, Stoesser T, Bomminayuni S, Chen Z (2012) Effects of a fully submerged boulder within a boulder array on the mean and turbulent flow fields: implications to bedload transport. *Acta Geophys* 60(6):1502–1546

2. Tsakiris AG, Papanicolaou AT, Hajimirzaie SM, Buchholz JH (2014) Influence of collective boulder array on the surrounding time-averaged and turbulent flow fields. *J Mt Sci* 11(6):1420–1428
3. Baki ABM, Zhang W, Zhu DZ, Rajaratnam N (2016) Flow structures in the vicinity of a submerged boulder within a boulder array. *J Hydraul Eng* 143(5):04016104
4. Lacey RJ, Rennie CD (2012) Laboratory investigation of turbulent flow structure around a bed-mounted cube at multiple flow stages. *J Hydraul Eng* 138(1):71–84
5. Morris HM Jr (1955) Flow in rough conduits. *Trans Am Soc Civ Eng* 120(1):373–398
6. Goring DG, Nikora VI (2002) Despiking acoustic Doppler velocimeter data. *J Hydraul Eng* 128(1):117–126
7. Wahl TL (2003) Discussion of “Despiking acoustic Doppler velocimeter data” by Derek G. Goring and Vladimir I. Nikora. *J Hydraul Eng* 129(6): 484–487
8. Wahl TL (2000) Analyzing ADV data using WinADV. In: Joint conference on water resource engineering and water resources planning and management. ASCE, Minneapolis, Minnesota, United States
9. Kironoto BA, Graf WH (1994) Turbulence characteristics in rough uniform open-channel flow. *Proc Inst Civ Eng-Water Maritime and Energy* 106(4):333–344
10. Nezu I, Nakagawa H (1993) Turbulence in open-channel flows. IAHR monograph series. AA Balkema, Rotterdam, Netherlands
11. Baki ABM, Zhu DZ, Rajaratnam N (2015) Turbulence characteristics in a rock-ramp-type fish pass. *J Hydraul Eng* 141(2):04014075

Predicting the Geometrical Characteristics of an Inclined Negatively-Buoyant Jet for Angles from 30° to 60° Using GMDH Neural Network



Hassan Alfaifi, Abdulmajid Mohammadian, and Hossein Bonakdari

1 Introduction

The wastewater produced from seawater desalination plants (brine water) is commonly disposed of back into the near coastal water body. This process is done with a submerged pipe creating a buoyant effluent jet. This effluent (brine) has a high concentration of salinity, which causes negative impacts on the marine ecosystem, particularly in the near-field zone of the discharge point [7, 29]. The discharged effluent is either a positively-buoyant jet (lighter than the receiving water) or a negatively-buoyant jet (denser than the receiving water), depending on the method used in the desalination plant. The dense jet (negatively-buoyant jet) has highly adverse effects on the marine environment. Commonly, desalination plants that use a Reverse Osmosis (RO) system use submerged discharges, which formed a negatively-buoyant jet, to ensure a high dilution of this effluent in order to minimize the adverse impacts on the marine environment.

Numerous previous research has focused on laboratory studies of the mixing of the negatively-buoyant jet either alone or in combination with modeling efforts ([1–3, 6, 9, 18, 19, 23, 26, 27, 31]; Jiang and Lee 2014; Papakonstantis and Papanicolaou 2011a, 2011b; Roberts and Daviero 1997). Less effort, however, has been given to numerical modeling and developing an analytical solution and to predict the behaviors and geometric characteristics of inclined dense jets [4, 16, 17, 20, 22], which still

H. Alfaifi (✉)

King Abdulaziz City for Science and Technology (KACST), National Center of Environmental Technology, Riyadh, Saudi Arabia

e-mail: halfa103@uottawa.ca; halfaifi@kacst.edu.sa

A. Mohammadian

Department of Civil Engineering, University of Ottawa, Ottawa, Canada

H. Bonakdari

Department of Soils and Agri-Food Engineering, Université Laval, Quebec City, Canada

© Canadian Society for Civil Engineering 2022

S. Walbridge et al. (eds.), *Proceedings of the Canadian Society of Civil Engineering*

Annual Conference 2021, Lecture Notes in Civil Engineering 250,

https://doi.org/10.1007/978-981-19-1065-4_30

requires further investigation. Therefore, more studies using numerical modeling as an efficient tool can be useful for predicting the behaviors of brine discharges into seawater.

Recently, efforts in the use of Artificial Neural Network (ANN) methods have been done in different fields of engineering science. One of these methods is the group method of data handling (GMDH). This method is used for solving nonlinear engineering problems and providing explicit equations for modeling the hydraulic phenomena. Results obtained from this method have shown accurate results compared to those of other methods [5, 8, 10–12, 21, 28, 32].

For example, [12] conducted a study to exam the ability of the GMDH model for predicting the geometric variables of stable channels (width, depth, and slope). The results showed that the GMDH model performed well. Moreover, they compared the result obtained from GMDH with those from previous theoretical equations (based on regression analysis), and this comparison showed the ability of GMDH model to predict the geometric variables of channels better than other theoretical equations. Alitaleshi and Daghbandan [5] conducted a study to evaluate the quality of treated water in a water treatment plant using GMDH-type neural network. The results showed the success of the model used (GMDH) for prediction as the results of the determination coefficient were higher than 0.97.

A recent study conducted by [4], where they used a GMDH model to predict the geometrical characteristics of inclined dense jets for angles 15° and 52° . This study was considered as the first use of this new approach in the field of mixing jet behavior. The empirical equations for estimating the geometrical characteristics of inclined buoyant jets have been established in this study. The GMDH results gave an accurate prediction for the parameters tested compared to the other analytical models.

Therefore, the aim of this study is to examine the ability of the GMDH model in predicting geometrical characteristics of inclined negatively-buoyant jets for angles ranging from 30° to 60° .

2 Theory

2.1 Dimensional Analysis of an Inclined Negatively-Buoyant Jet

Figure 1 illustrates an inclined negatively-buoyant dense jet being discharge into a stagnant water body with zero ambient velocity ($U_a = 0$). The jet (has a diameter D) is positioned at an angle (θ) to the horizontal plane. The states of the ambient water is an unstratified and has a constant density ρ_a . The height of the ambient water is deep enough to allow the jet to fully develop below the water surface. The initial velocity of the discharge jet is U_0 and it has a density ρ_0 ($\rho_a < \rho_0$).

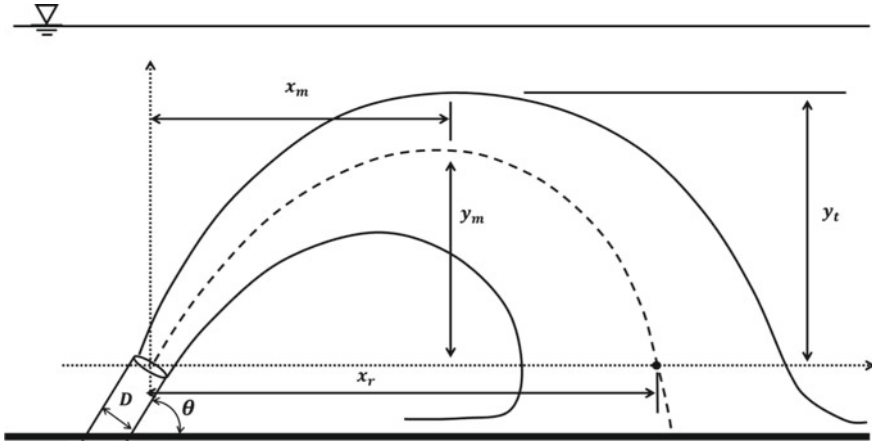


Fig. 1 Parameters of the inclined negatively-buoyant dense jet in stagnant ambient water [4]

The characteristics of the geometry of the mixing of the jet can be defined as: (1) the jet terminal rise height (y_t), (2) the horizontal distance from the point of jet maximum height of the centerline to the nozzle (x_m), (3) the jet centerline height (y_m), and (4) the jet return point (x_r), which is the point of the centerline of the same elevation to the tip of the nozzle (Fig. 1). The inclined dense jets can also be characterized by another parameters such as its discharge volume flux $Q_0 = U_0\pi D^2/4$, momentum flux $M_0 = U_0^2D^2\pi/4$, and buoyancy flux $B_0 = g'_0 U_0\pi D^2/4$. The g' is defined as $g' = g(\rho_0 - \rho_a)/\rho_a$, which is the effective gravitational acceleration, and g is the gravitational acceleration. Two length scales can be formed from these fluxes: the momentum length scale $L_M = M_0^{3/4}/B_0^{1/2}$, which determines the distance over which the buoyancy of the jet is less important than its momentum, and the source length scale $L_Q = Q_0/M_0^{1/2}$, which determines the distance over which the source discharge is important. These two length scales are employed to determine the mixing and geometric characteristics of a turbulent buoyant jet. The jet densimetric Froude number Fr_d , which is the ratio of inertia to buoyancy, can be derived for turbulent buoyant jets and is proportional to the ratio of length scales: $Fr_d = L_M/L_Q = U_0/(g'D)^{0.5}$, which means that the L_M and L_Q can be related to the densimetric Froude number and the jet diameter as:

$$L_M = Fr_d D \left(\frac{\pi}{4}\right)^{0.25} \tag{1}$$

$$L_Q = D \left(\frac{\pi}{4}\right)^{0.5} \tag{2}$$

For inclined dense jet, the M_0 will always be an important parameter where the jet is originally moving upwards, because the vertical component of M_0 and B_0 are not acting in the same direction. Therefore, any dependent variable will be a function of

M_0 , B_0 , Q_0 and θ as below [30]:

$$y_t = f(M_0, B_0, Q_0, \theta) \quad (3)$$

Based on the above, the terminal rise height of an inclined dense jet can be written in terms of the length scales as:

$$\frac{y_t}{L_M} = f\left(\frac{L_M}{L_Q}, \theta\right) \quad (4)$$

Instead of the length scales, y_t can also be alternatively expressed by using the jet's densimetric Froude number and nozzle diameter as:

$$\frac{y_t}{Fr_d D} = C_{y_t}(\theta) \quad (5)$$

Similarly, the other geometric parameters of the jet (i.e. x_r , x_m , and y_m) can be derived by using the coefficient values (C_{x_r} , C_{x_m} , and C_{y_m}), which can be established through experiments.

2.2 The Group Method of Data Handling (GMDH)

The GMDH is a computer program based algorithm, which was first proposed by [14]. This type of algorithm is used to model problems that involve data series for multi-input–single-output systems. The concept of the GMDH of selection and hybridization corresponds to the genetic algorithm. The main advantage of the GMDH, as opposed to traditional neural networks, is its use in establishing mathematical models for a given procedure. Like other types of neural networks, the GMDH can have one or more hidden layers. More information regarding GMDH processing is provided in a recently published article by the same authors in [4].

To use the GMDH method for prediction of the geometrical characteristics of a negatively-buoyant dense jet, the output variables were set to be the dimensionless of jet characteristics (i.e., x_m/D , y_m/D , x_r/D , and y_t/D) by using one output variable for each run. For the input variables, the Fr_d and θ parameters were used for all cases. The explicit equations for the GMDH were developed using the GMDH Shell software (GMDH [13]). The experimental data presented in previous studies for angles ranging from 30° to 60° are used in the GMDH processing. Obtaining a good prediction from the GMDH model is based on the preparation of the dataset, and therefore the observed dataset was randomly divided into two groups one for training and one for testing. For each case, the GMDH model was trained using 70% to 80% of the dataset, while the remaining data were used for testing the model. For the training dataset, the models developed were used to determine the dimensionless of the jet geometrics as a function of Fr_d and θ , with the optimal models being selected. After

that, the models selected were used to perform additional calculations for the testing dataset, and then their performance was evaluated. In this study, the layers were limited to a maximum of two layers in order to ensure the simplicity of the model.

3 Results and Discussion

In this section, the method of obtaining the GMDH model and its results are presented. The predicted model is compared with the actual data of angles obtained from previous studies. The comparison data includes angles 30°, 35°, 38°, 40°, 45°, 50°, 52°, 55°, 60°. For brevity, only the results of x_r/D are shown in this section. Figure 2 shows the results of the dimensionless geometrical parameters x_r/D of the actual data for training and testing plotted against Fr_d . The predicted model of the GMDH was added to the figures to show how it accurately fits the data. In order to evaluate the performance of the model, the coefficient of determination (R^2) was added to each figure to show the accuracy of the proposed model in predicting and fitting the actual data. From these Figures, it can be seen that the R^2 values show an excellent

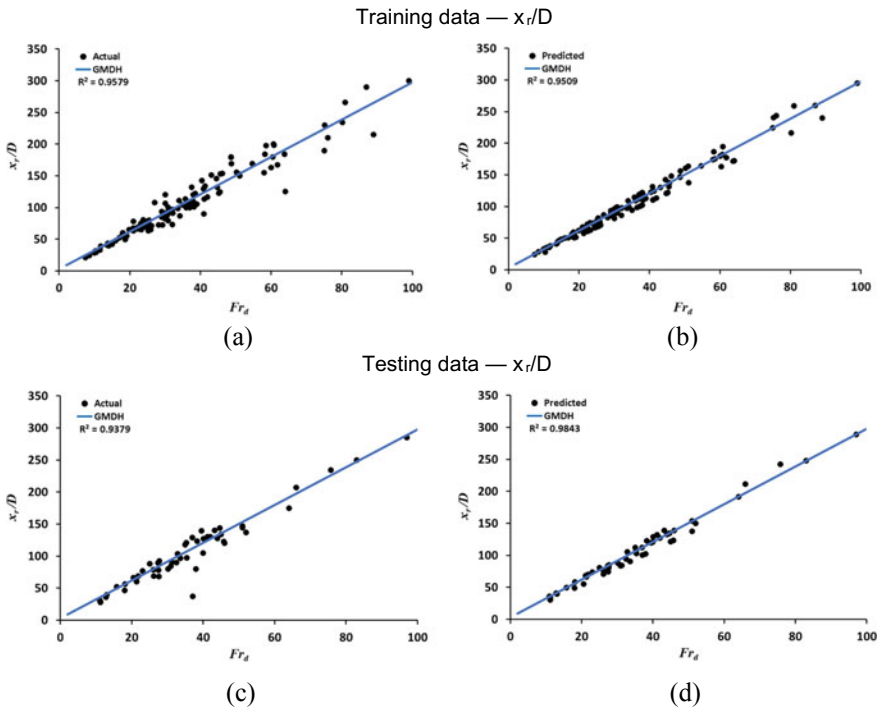


Fig. 2 The scatter plot with fitting lines of the actual data and GMDH predicted model for angles from 30° to 60°

prediction for the proposed model (GMDH), where 0.9579 and 0.9509 for training and 0.9379 and 0.9843 for testing. The testing results show a good prediction of the model. Figure 2a, b, c and b illustrate the comparison between the actual and predicted model for the training data, and the comparison between the actual and predicted model for the testing data, respectively.

Although there was a decrease in the value of the R^2 for the training model compared to the testing model, the accuracy remains high and acceptable. This indicates that all fitting lines pass through most of the data and the proposed model is highly accurate and satisfactory.

Figure 3 shows a detailed depiction of the performance of the GMDH method for x_r/D as well as how the model was trained and tested by the data index. As can be shown, the model was trained by 70% of the data while the rest was tested the model. Therefore, the key to influencing the change of training ratio is to obtain a simple model with high accuracy.

The equations extracted from the GMDH method for two main geometrical parameters (x_r/D and y_t/D) are presented in Table 1. These equations were used to predict the geometrical parameters in this study. As can be seen, all equations are functions of the Fr_d and the θ . Whereas, each model has limitations, these equations are valid for Fr_d ranging from 10 to 100. The equations are presented in dimensionless geometric

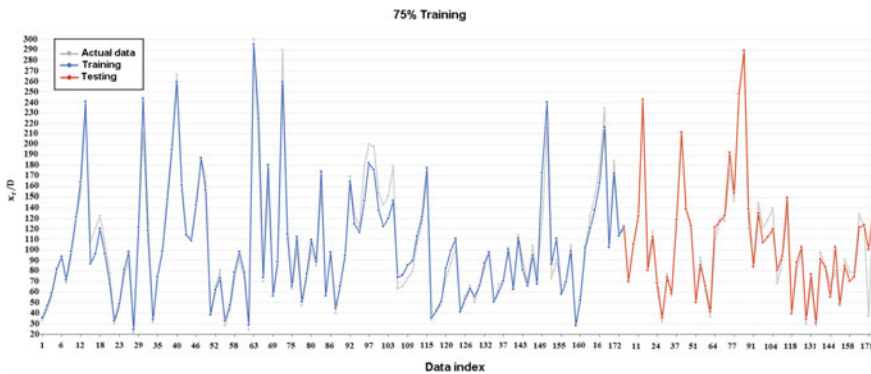


Fig. 3 The performance of GMDH method in the training and testing stage

Table 1 Nonlinear equations extracted from GMDH for x_r/D and y_t/D

| Geometrical parameter | GMDH proposed equations |
|-----------------------|--|
| x_r/D | $= -24.874 + 1.18264 \theta - 0.0132652 \theta^2 + 1.0016 N3 - 0.005735 N3^2$ $N3 = 1.85728 - 0.016678 Fr_d \theta^2 + 3.69098 Fr_d$ |
| y_t/D | $= 25.8315 + 0.0767749 Fr_d + 0.0254024 Fr_d N3 - 0.984546 N3 + 0.00825135 N3^2$ $N3 = -108.957 + 2.1931 \theta + 0.387391 \theta \theta^{1/3} - 0.0391048 \theta^2 + 5.14246 \theta^{1/3} + 5.89474 \theta^{1/3^2}$ |

Table 2 Postprocessed results of the GMDH method for training and testing

| Geometrical parameter | x_r/D | | y_t/D | |
|------------------------|-----------|------------|-----------|------------|
| | Model fit | Prediction | Model fit | Prediction |
| Number of Observations | 122 | 52 | 92 | 39 |
| Max. negative error | - 33.22 | - | - -17.99 | - |
| MAE | 8.23 | 7.87 | 4.85 | 5.54 |
| RMSE | 11.85 | 12.29 | 7.94 | 9.00 |
| R2 | 0.96 | 0.95 | 0.95 | 0.94 |

parameters and can be used for predicting the main geometrical parameters for the inclined dense jet for any angles located between 30° and 60°. Moreover, in order to ensure the simplicity of the model, the layers were limited to a maximum of two layers. In fact, we can increase model layers where the accuracy will be increased, but in return the model will be more complicated.

Moreover, Table 2 illustrates some postprocessed results of the GMDH method (statistical indices) to evaluate the performance of the GMDH model. These indices can clearly show the accuracy of the proposed model as it is commonly used in literature.

The three most important types of statistical indices were used: the coefficient of determination (R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE), which are defined as follows:

$$R^2 = 1 - \frac{\sum(\hat{y} - \bar{y})^2}{\sum(y - \bar{y})^2} \tag{6}$$

$$MAE = \frac{\sum_{i=1}^n |\hat{y} - y|}{n} \tag{7}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y} - y)^2}{n}} \tag{8}$$

where \hat{y} is the predicted values of the GMDH model, \bar{y} is the mean of the actual data, y is the actual data, and n is the number of actual data. The statistical results obtained for the GMDH model from these Equations are presented in Table 2. As known, the low values of these parameters (except R^2) indicate the high accuracy of the model. All geometrical results of R^2 are higher than 0.94, which means that the model proposed produced excellent predicting results.

4 Conclusions

Studying the mixing behavior of the brine water produced from desalination plants in the near-field zone has become a research priority to reduce the impact of this water on the marine environment. The discharged water can be formed as an inclined negatively-buoyant jet when its density is higher than the density of the receiving water. Therefore, predicting the geometrical characteristics of this jet is crucial important for improving the understanding of the mixing behavior of this type of discharging and its dilution. In this study, two main geometrical characteristics of the inclined negatively-buoyant jet for angles ranging from 30° to 60° was modeled and predicted using a type of artificial neural network (ANN) called group method of data handling (GMDH). Results of this model were evaluated statistically. The results show that this model is able to accurately predict the geometrical characteristics of the negatively-buoyant jet.

Acknowledgements Hassan Alfaifi would like to thank King Abdulaziz City for Science and Technology (KACST) for their financial support.

References

1. Abessi O, Roberts PJW (2014) Multiport diffusers for dense discharges. *J Hydraul Eng* 140(8):04014032
2. Abessi O, Roberts PJW (2015) Dense jet discharges in shallow water. *J Hydraul Eng* 142(1):04015033
3. Abessi O, Roberts PJW (2015) Effect of nozzle orientation on dense jets in stagnant environments. *J Hydraul Eng* 141(8):06015009
4. Alfaifi H, Mohammadian A, Bonakdari H (2020) Experimental investigation and model development of geometric characteristics of negatively buoyant jets inclined at 15° and 52° using GMDH method. *J Coastal Res* 36(3)
5. Alitaloshi F, Daghbandan D (2019) Using a multi-objective optimal design of GMDH type neural networks to evaluate the quality of treated water in a water treatment plant. *Desalin Water Treat* 139:123–132
6. Bashitialshaaer R, Larson M, Persson KM (2012) An experimental investigation on inclined negatively buoyant jets. *Water (Switzerland)* 4(3):720–738
7. Bombar G, Dölgen D, Necdet AM (2016) Environmental impacts and impact mitigation plans for desalination facilities. *Desalin Water Treat* 57(25):11528–11539
8. Bonakdari H, Mohammadian A (2019) Evolutionary prediction of an inclined dense jet in shallow water. *Desalin Water Treat* 151:1–16
9. Crowe AT (2013) Inclined negatively buoyant jets and boundary interaction. Ph.D. Thesis, University of Canterbury, Christchurch, New Zealand
10. Ebtehaj I, Bonakdari H, Khoshbin F, Bong CHJ, Ghani AA (2017) Development of group method of data handling based on genetic algorithm to predict incipient motion in rigid rectangular storm water channel. *Scientia Iranica* 24(3):1000–1009
11. Ebtehaj I, Bonakdari H, Zaji AH, Azimi H, Khoshbin F (2015) GMDH-type neural network approach for modeling the discharge coefficient of rectangular sharp-crested side weirs. *Eng Sci Technol Int J* 18(4):746–757

12. Gholami A, Bonakdari H, Ebtehaj I, Shaghghi S, Khoshbin F (2017) Developing an expert group method of data handling system for predicting the geometry of a stable channel with a gravel bed. *Earth Surf Proc Land* 42(10):1460–1471
13. Llc GMDH (2019) GMDH SHELL Ds. Version 3(8):9
14. Ivakhnenko AG (1971) Polynomial theory of complex systems. *IEEE Trans Syst Man Cybern* 4:364–378
15. Jiang B, Law AW-K, Lee JH-W (2014) Mixing of 30° and 45° inclined dense jets in shallow coastal waters. *J Hydraul Eng* 140(3):241–253
16. Jirka GH (2004) Integral model for turbulent buoyant jets in unbounded stratified flows. Part I: single round jet. *Environ Fluid Mech* 4(1):1–56
17. Kikkert GA, Davidson MJ, Nokes (2007) Inclined negatively buoyant discharges. *J Hydraul Eng* 133(5):545–54
18. Kikkert GA (2006) Buoyant jets with two and three-dimensional trajectories. Ph.D. Thesis, University of Canterbury, Christchurch, New Zealand
19. Lai CCK, Lee JHW (2012) Mixing of inclined dense jets in stationary ambient. *J Hydro Environ Res* 6(1):9–28
20. Lee JHW, Chu VH (2003) Turbulent jets and plumes: a Lagrangian approach. Kluwer Academic Publishers, The Netherlands, Dordrecht, The Netherlands
21. Mehri Y, Soltani J, Khashehchi M (2019) Predicting the coefficient of discharge for piano key side weirs Using GMDH and DGMDH techniques. *Flow Meas Instrum* 65:1–6
22. Oliver CJ, Davidson MJ, Nokes RI (2013) Removing the boundary influence on negatively buoyant jets. *Environ Fluid Mech* 13(6):625–648
23. Oliver CJ (2012) Near field mixing of negatively buoyant jets. Ph.D. Thesis, University of Canterbury, Christchurch, New Zealand
24. Papakonstantis IG, Christodoulou GC, Papanicolaou PN (2011) Inclined negatively buoyant jets 1: geometrical characteristics. *J Hydraul Res* 49(1):3–12
25. Papakonstantis IG, Christodoulou GC, Papanicolaou PN (2011) Inclined negatively buoyant jets 2: concentration measurements. *J Hydraul Res* 49(1):13–22
26. Papakonstantis IG, Tsatsara EI (2018) Trajectory characteristics of inclined turbulent dense jets. *Environ Process* 5(3):539–554
27. Papakonstantis IG, Tsatsara EI (2019) Mixing characteristics of inclined turbulent dense jets. *Environ Process* 1–17
28. Parsaie A, Azamathulla HM, Haghiabi AH (2018) Prediction of discharge coefficient of cylindrical weir-gate using GMDH-PSO. *ISH J Hydraul ineeing* 24(2):116–123
29. Petersen KL, Heck N, Reguero BG, Potts D, Hovagimian A, Paytan A (2019) Biological and physical effects of brine discharge from the carlsbad desalination plant and implications for future desalination plant constructions. *Water (Switzerland)* 11(2)
30. Roberts PJW, Ferrier A, Daviero G (1997) Mixing in inclined dense jets. *J Hydraul Eng* 123(8):693–699
31. Shao D, Law AWK (2010) Mixing and boundary interactions of 30° and 45° inclined dense jets. *Environ Fluid Mech* 10(5):521–553
32. Zaji AH, Bonakdari H (2019) Robustness lake water level prediction using the search heuristic-based artificial intelligence methods. *ISH J Hydraul Eng* 25(3):316–324

Influence of Rainfall on the Probability of Red-Light Running at Signalised Intersections



J. Oyaro and J. Ben-Edigbe

1 Introduction

In South Africa, approximately 1 million road accidents are reported per year. The majority of the accidents that occur are caused by human factors, with, on average, over 40 people a day being fatally injured and at least 20 being left permanently disabled. The number of fatal crashes was reduced by three per cent from 1438 in 2018/19 to 1390 in 2019/2020. It resulted in a 10% reduction in the number of fatalities from 1789 fatalities in 2018/19 to 1617 fatalities in the 2019/2020 festive season. South Africa has made strides in reducing road crash fatalities since its peak in 2006. However, the numbers remain high, thus, motivating the research study. Traffic signal timing is the technique used to distribute right-of-way at signalised intersections. The process allows traffic signal lights to emit intermittently, by which each intersection movement may be serviced without allowing conflicting movements to enter the intersection simultaneously. The amount of time required to display all phases for each directional movement of an intersection before returning to the starting point is sacrosanct to signalised intersection performance.

When the light is yellow, it is up to the driver's discretion because the yellow light is a warning to prepare for red. Some drivers accelerate to clear the intersection at the onset of yellow light, and others decelerate to stop at the stop-line. Approximately 20% of crashes at signalised intersections occur due to signal violation [23]. Measures to reduce aberrant driver behaviour and improve intersection safety have been researched and implemented. They include re-design of yellow time (especially lengthening yellow time), camera enforcement, and countdown timers. Despite these attempts, red-light running is still an evasive challenge. Drivers can be proactive or responsive; their behaviour determined by traffic signal compliance. They can

J. Oyaro · J. Ben-Edigbe (✉)

Sustainable Transportation Research Group, University of KwaZulu-Natal, Durban, South Africa
e-mail: ben-edigbe@ukzn.ac.za

choose to be normal, aggressive or conservative, especially in response to traffic lights. Drivers must make the right decisions; inappropriate decisions could cause serious safety concerns for road users. A dilemma zone is where drivers approaching a signalised intersection must decide to either proceed or stop at the onset of the yellow indication. In the dilemma zone, problems arise when different drivers caught up in these zones make complex decisions. These decisions could cause rear-end collisions or red-light running, all of which are unsafe actions. There is no clear agreement among researchers and practitioners about the definition and demarcation of dilemma zones. During the yellow interval, driver behaviour is the biggest contributor to red-light running (RLR), and signal timing design is considered one way to manage that driver behaviour to mitigate red light violations (Cohn et al.). While many drivers obey traffic signals, the possibility for violations exists due to driver distraction, indecision, or a deliberate decision to ignore the traffic signal. Rain causes accidents through a combination of several physical effects, including poor road surface grip and aquaplaning. Drivers will need to reduce speed during rainfall due to the loss of visibility caused by rainfall intensity. These changes contribute to intersection capacity loss [16]. The paper explored the impact of rainfall on the probability of red-light running at signalised intersections in Durban, South Africa. The key research question raised is: to what extent will rainfall intensity (light, moderate and heavy) influence red-light running.

2 Literature Review

This paper deals with red-light running at signalised intersections for which a satisfactory solution has yet to be prescribed. It analyses drivers' decision-making at the onset of yellow time interval at signalised intersections during rainfall. The South Africa manual for traffic signals [18] recommends setting the yellow time interval duration based on the probability of stopping. In South Africa, the controversial yellow time interval traffic law states that 'when a vehicle enters the intersection during the entire yellow interval, it is permitted to proceed and clear the intersection safety, however in a situation where the vehicle can neither enter nor be in the intersection on red, it must stop upon receiving the yellow interval' [18]. When a vehicle enters the intersection during the entire yellow interval suggests that the vehicle has cleared the approach link length and already inside the intersection; permission to proceed is inconsequential. However, where the vehicle can neither enter nor be in the intersection on red suggests that the vehicle is on the approach link length away from the intersection; it must stop upon receiving the entire yellow interval. The most severe yellow time interval problem observed at survey sites is red-light running (RLR), defined as attempting to cross the intersections during the red phases. Red-light running has been identified as one of the prominent factors involved in signalised intersection crashes [24]. When the lights are yellow, it is a warning to the drivers of an upcoming red light, and the red light means an end to the right of way. When the light turns yellow, drivers are faced with the decision to either

slow down and stop, continue at their speed and pass or accelerate to go through the intersection. A driver may decide to either maintain their speed or accelerate to clear the intersection. It leads to red-light running. Previous work has concluded that the dilemma zone contributes to red-light running behaviour, especially for high-speed intersections [23]. A dilemma zone is when a vehicle approaching an intersection stop-line at the onset of yellow can neither safely/ comfortably stop nor clear the intersection. Traffic accidents during the yellow time account for more than half of the traffic accidents at signalised intersections [22].

Studies have been carried out to understand driver behaviour at the onset of the yellow time and how it contributes to intersection safety, especially red-light running cases. Elmitiny et al. [7] their study found that the distance from the intersection at the start of yellow, operating speed and traffic flow are important factors that determine driver decision to either stop or cross the intersection. Acceleration and deceleration during yellow time have also been found to contribute to red-light running [1]. Further, Arash et al. [1], in their study based on simulation data, found that the period between 2 and 6 s after the onset of the yellow time was the most important to monitor to predict red-light running behaviour. The study also found time to an intersection at the onset of yellow required deceleration and distance to an intersection at a yellow onset to be among the most important factors that affect red-light running. Driver behaviour is also affected by several other factors: road infrastructure, vehicle type, and prevailing weather and traffic conditions. Speed is also a parameter that is affected by prevailing weather conditions. Perrin et al. [17] study conducted in Salt Lake City found that start-up delay increased by 5–23% (the highest snowy conditions). Li et al. [11] on the other hand, it concluded that compared to clear weather conditions, driver dilemma zone boundaries started farther away from the stop-line by approximately 0.2 s, driver perception reaction time (PRT) increased by about 0.11 s, and driver deceleration levels decreased by about 8%. The conclusion is that driver behaviour during the yellow interval affects red-light running, and that driver behaviour is affected by prevailing weather conditions.

This study sought to answer the question of how red-light running is affected by rainfall under different intensities. In Malaysia's research, stopping sight distance was reduced by rainfall [2–4]. The study concluded that irrespective of their intensities, rainfalls have a significant impact on sight distances. Mashros et al. [14] explored the impact of various rainfall conditions on traffic flow and speed in Malaysia and concluded that rainfall, irrespective of its intensities, impacts traffic flow and speed except for average volume. Some other factors affected by rainfall at signalised intersections include; saturation headway, saturation flow rate, and start-up lost time [22]. When it comes to perception reaction time, it was found to be affected by rainfall conditions [6]. A study to determine the weather effect on perception reaction time (PRT) at the onset of the yellow interval concluded that the driver's perception reaction time (PRT) increased as the time to intersection increased and was longer when the driver was travelling along an upgrade approach section. Overall, PRTs under wet conditions were found to be higher than those under dry conditions.

The South Africa manual for traffic signals [18] recommends an approach to signal timing where the yellow period is reduced and the all-red period correspondingly

increased while effectively retaining the inter-green period. Because, with a long yellow period, there is a tendency by drivers to abuse it, using it as an extension of the green period, and this can result in unsafe conditions. Equations 1 and 2 are the sight distance and yellow time interval equations, respectively [18].

$$\text{Sight Distance, } S = vt + \frac{v^2}{2a} \quad (1)$$

$$\text{Yellow time interval, } \tau_c = \left(t + \frac{v}{2a}\right) \quad (2)$$

where: v is approach speed; t denotes perception time (s); a denotes acceleration/deceleration (m/s^2).

In previous studies [8, 9, 12], drivers' behaviours were assessed using binary logistic regression or fuzzy logic. Gates et al. [8] recommend the logit regression model. When conducting a field investigation of RLR in Shanghai, China, Zhang et al. [24] used the random effects, logistic regression model. When modelling driver behaviour during the onset of yellow intervals at signalised intersections, Long et al. [13] used the agent-based behavioural model. Wang et al. [21] conducted an empirical study of modelling driver response during yellow signal intervals based on the probit model. The model was successful in predicting yellow light running occurrences. The fuzzy logistic model was used by Hurwitz et al. [10] to describe driver behaviour at high-speed signalised intersections. Moore and Hurwitz [15] used the fuzzy logic model when conducting a driving simulator study for dilemma zone identification to assist drivers when confronted with yellow signal lights. Gates et al. [9] used the logistic regression model proved to be a good fit probabilistic method of predicting red-light running. The regression model is of the form:

$$\text{Logit}(P) = \ln(P/1 - P) \quad (3)$$

Several factors influence drivers' behaviour when approaching intersections. These include the vehicle speed (v), Time to Intersection (τ), Distance to Intersection (d), among others. In this 'with/without' rainfall paper, travel time (τ) to stop line and approach speed (v) used as independent parameters. Travel time (τ) is a function of free-flow speed and saturation; however, the degree of saturation is set at zero because only lead vehicles approaching the signalised intersection considered in the study. In the logistic regression model, the dependent variable is the driver's choice behaviour. At the onset of a yellow time interval, $y = 1$ represents red-light running (RLR) and $y = 0$ means that the driver chooses the alternative by way of bringing the vehicle to a halt at the stop line. The logit regression model Eq. 4 can be rewritten as:

$$L_p = \ln\left(\frac{P}{1 - P}\right) = \beta_0 + \beta_1\tau + \beta_2v + \beta_3d \quad \text{for } 0 < P < 1 \quad (4)$$

The probability of a driver’s choice at the onset of a yellow signal can be calculated by;

$$\text{The probability of RLR estimated by } P(Y = 1) = \frac{e^{\beta_0 + \beta_1 \tau + \beta_2 v + \beta_3 d_3}}{1 + e^{\beta_0 + \beta_1 \tau + \beta_2 v + \beta_3 d_3}} \quad (5)$$

$$\text{The probability of stopping estimated by } P(Y = 0) = \frac{1}{1 + e^{\beta_0 + \beta_1 \tau + \beta_2 v + \beta_3 d_3}} \quad (6)$$

where

P represents the probability of a driver’s choice, d is stopping distance, β_0 represents the constant of the model and $\beta_1, \beta_2, \beta_3$ are model coefficients, τ is travel time to the stop line, $t_f \left(1 + \partial \left[\frac{T_v}{Q} \right]^\mu \right)$, T_v denotes traffic volume, Q is signalised intersection capacity, v is the vehicle’s approach speed at the onset of a yellow time interval.

3 Data Collection and Preparation

Data were collected at three selected four-legged intersections in the city of Durban, South Africa. Data were collected continuously for six weeks between October and December 2019 during weekdays. Selected signalised intersections (Umgeni road/Alpine Road and Sandile Thusi road/Stalwart Simelane Street) are in Durban, South Africa. At each signalised intersection, reference lines were marked with 5 m clearance to acquire accurate location and speed of the vehicles. The intersections were selected to meet the criteria; proximity to rain gauge station, be within Durban city and posted speed limit of 60 km/h (meaning similar yellow and all-red durations). Traffic was collected from video footage obtained with assistance from the eThek-wini Transport Authority and Durban Metro Police. Signalisation and geometric data were obtained from the sites using google earth (see Fig. 1) and image processing video cameras.

The field data collection was used for verification purposes. The rain gauge was at 0.20 mm/tip. Logged rainfall data were downloaded and exported to an excel sheet using a 5 min (0.083 h) time interval, then converted to intensity by dividing with 0.083 to have the intensity in mm/h. Rainfall intensities were classified as light rain (<2.5 mm/h), moderate rain (2.5 and 10 mm/h), heavy rain (10 and 50 mm/h) using World Meteorological Organisation classification system. Weighty rain with an intensity greater than 50 mm/h was not considered due to excessive aquaplaning and drag force effect on tyres. Typically at all sites, three-speed regimes (free flow, transition and posted speed limit) were in operation. As shown below in Fig. 2, each site was divided into three zones (free flow-FVZ, transition-TVZ, and posted speed-PVZ), and posted speed zone sub-divided into dilemma-DS and braking-BS sections.



Fig. 1 Google earth snapshot of the field intersection in Durban, South Africa

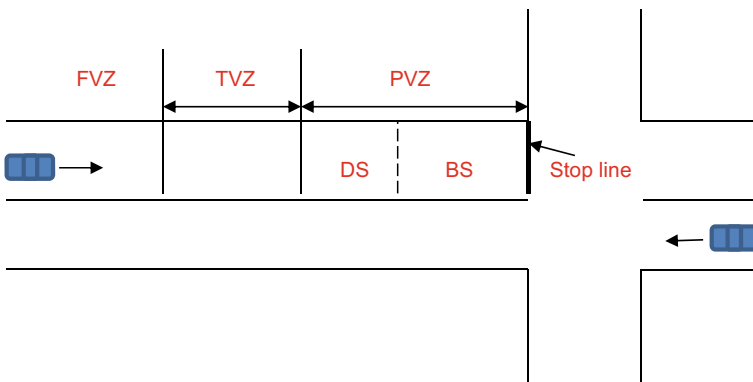


Fig. 2 Typical site sections

The free flow speed represents the average speed at the free flow zone, and the desired drivers speed in the absence of traffic control devices. At the transition speed zone, drivers may reduce or increase speed in this zone at the onset of yellow light. The posted speed limit is influenced by drivers decision and road safety measures in place (often speed cameras). Free flow speeds were collected during the green light phase and used to determine sight distance. The posted speeds were collected at the yellow time interval and used to estimate travel time. Note travel time equals free flow time where the lead vehicle speed in traffic flow is unimpeded, and traffic flow is subjected to an inconsequential degree of saturation. The generalised 85th percentile free-flow speed is 80 km/h (22.24 m/s), assuming an average reaction time of 2 s; the ensuing FVZ sight distance is approximately 112 m from the stop line. In contrast, the 50 m distance from PSZ to the stop line is a function of posted speed multiply

by yellow time interval. The dilemma section within PVZ is 20 m, and the braking section is 30 m.

The speed data were collected at two points; the first at a point upstream of the intersection, and the second speed was collected at the intersection stop-line. The two-speed parameters and the time difference were then used to determine the acceleration/deceleration. An extraction software was used to extract both the speed and time data from the obtained video cameras. Time was obtained by measuring how long after the yellow time, the vehicle passed the dilemma zone mark. For example, 0 s means the vehicle passed the zone at the start of yellow time, while 2 s means the vehicle passed the zone 2 s after yellow. The data collected at selected sites are speed, travel time, and determination of whether it was red or yellow light running.

4 Analysis and Discussion

The computed 85th percentile free-flow speed is 80 km/h during dry weather condition. Preliminary analysis reveals that rainfall caused average speed reduction from 80 km/h during dry weather to 50 km/h (37.5%) during light rain; to an average speed of 40 km/h (50%) during moderate rain and an average speed of 33 km/h (58.7%) during heavy rain Start-up Lost Times (SULT) for dry weather conditions are lower than start-up lost times for rainfall. Light rainfall has an average increase in start-up lost time of 7.47% (ranging from 0.55 to 19.17%), moderate rainfall has an average increase of 19.72% (ranging from 5 to 26%). In contrast, heavy rainfall has an average increase of 25% (ranging from 10 to 33%). These results are within the findings of previous studies [19, 20]. Headways are lower for dry weather than those of rainy conditions in all cases. Typical discharge headways and their statistical fitness are shown in Table 1.

However, in this paper, free-flow speed (v), travel time to stop line (t), and sight distance (d) is the independent parameters, and driver’s choice behaviour is the dependent variable. Note that free-flow speed samples were collected during the green phase, and sight distance is measured from the free flow speed zone to the stop line. Travel time to the stop line is measured as the time taken to cover PVZ 50 m length to the stop line. At the onset of yellow light, the dependent variable is represented by y , where $y = 0$ means the driver brings the vehicle to a halt at the stop-line and $y = 1$ denotes that the driver proceeded beyond the stop-line into the intersection at the onset of red light. In other to determine the probability of drivers stopping or RLR, a binary logistic regression model is used where the regression models are:

$$\text{The probability of RLR is estimated by } P(Y = 1) = \frac{e^{\beta_0 + \beta_1 v + \beta_2 \tau + \beta_3 d}}{1 + e^{\beta_0 + \beta_1 v + \beta_2 \tau + \beta_3 d}}$$

$$\text{The probability of stopping is estimated by } P(Y = 0) = \frac{1}{1 + e^{\beta_0 + \beta_1 v + \beta_2 \tau + \beta_3 d}}$$

Table 1 Typical discharge headways and their statistical fitness

| Site | Sample size | Average | Median | Standard error | Skewness | Kurtosis |
|---------------|-------------|---------|--------|----------------|----------|----------|
| <i>Site 1</i> | | | | | | |
| Dry | 1500 | 1.42 | 1.41 | 0.01 | 0.01 | -0.56 |
| Light | 400 | 1.56 | 1.54 | 0.02 | 0.36 | -0.16 |
| Moderate | 450 | 1.56 | 1.53 | 0.02 | 0.34 | -0.63 |
| Heavy | 400 | 1.68 | 1.63 | 0.02 | 0.35 | -0.33 |
| <i>Site 2</i> | | | | | | |
| Dry | 1300 | 1.53 | 1.49 | 0.02 | 0.50 | 0.16 |
| Light | 150 | 1.60 | 1.58 | 0.03 | 0.44 | -0.16 |
| Moderate | 200 | 1.62 | 1.58 | 0.03 | 0.66 | 0.34 |
| Heavy | 490 | 1.73 | 1.69 | 0.02 | 0.64 | 0.29 |
| <i>Site 3</i> | | | | | | |
| Dry | 1000 | 1.52 | 1.51 | 0.01 | 0.31 | 0.09 |
| Light | 280 | 1.68 | 1.60 | 0.02 | 0.61 | 0.25 |
| Moderate | 200 | 1.75 | 1.70 | 0.03 | 0.56 | 0.23 |
| Heavy | 470 | 1.92 | 1.92 | 0.02 | 0.66 | 1.28 |

4.1 Binary Logistic Model

Rainfall reduces vehicle speed. Speed decrease triggers travel time increase. Many studies show that rainfall affects speed; however, the probability of red-light running during adverse weather condition has yet to be studied. However, speed is constrained by rainfall intensity and, by extension, the probability of red-light running. A vehicle driving at the posted speed limit under dry conditions has a lower probability of running a red light than in rainy conditions, so achieving the posted speed limit is difficult. Stopping distance, vehicle approach speed and free-flow travel time used as independent variables, whilst driver behaviour is the dependent variable. Note that the stopping distance is not the same as the distance from the stop line. Distance from the stop line is a function of a yellow time interval and vehicle speed. For example, the control regression model for dry weather and statistical fitness is shown in Table 2.

From Table 2, the t' values all have an absolute value greater than 2.5; therefore, all the variables used in the regression equation are useful in predicting red-light running violation. The R^2 value is greater than 0.5, and the F value is greater than 4 suggesting the model equation did not happen by chance. The control regression model for dry weather;

$$Y = -8.71 + 0.6656t + 0.8487v - 0.1474d \tag{7}$$

Table 2 Summary of statistical fitness and coefficients of dry weather regression model

| | d | v | t | Constant |
|----------------|----------|----------|----------|----------|
| t' values | -7.94666 | 8.875503 | 3.661122 | -7.08302 |
| Coefficients | -0.14741 | 0.848691 | 0.665578 | -8.70664 |
| Std. error | 0.01855 | 0.095622 | 0.181796 | 1.22922 |
| R ² | 0.722499 | 0.252962 | N/A | N/A |
| F df | 68.56128 | 79.00 | N/A | N/A |
| Residuals | 13.16167 | 5.055196 | N/A | N/A |

Note Ad—approach distance, v—speed, t is travel time

4.2 Sensitivity Analysis

Assume distance to stop line, $d = 50$ m; yellow interval time, $t = 3$ s and posted speed limit,

$$v = 16.67 \text{ m/s (60km/h)}$$

where $Y = -25.1 + 2.264t + 2.0045v - 0.2208d$ from Table 1.

Probability of RLR, $P_{RLR}(Y = 1)$

$$= \frac{e^{\beta_0 + \beta_1 t + \beta_2 v + \beta_3 d_3}}{1 + e^{\beta_0 + \beta_1 t + \beta_2 v + \beta_3 d_3}} = \frac{e^{-8.71 + 0.6656t + 0.8487v - 0.1474d}}{1 + e^{-8.71 + 0.6656t + 0.8487v - 0.1474d}} = 14.9\%$$

Probability of stopping, $P_s(Y = 0)$

$$= \frac{1}{1 + e^{\beta_0 + \beta_1 t + \beta_2 v + \beta_3 d_3}} = \frac{1}{1 + e^{-8.71 + 0.6656t + 0.8487v - 0.1474d}} = 85.1\%.$$

The rainfall regression model for site 01 is:

$$Y = -25.1 + 2.264t + 2.0045v - 0.2208d \tag{8}$$

From Table 3, the t' values all have an absolute value greater than 2.5; therefore, all the variables used in the regression equation are useful in predicting red-light running violation. The R^2 value is greater than 0.5, and the F value is greater than 4 suggesting the model equation did not happen by chance.

Table 3 Summary of statistical model tests

| | d | v | t | Constant |
|----------------|----------|----------|----------|----------|
| t' values | -4.70929 | 4.909271 | 4.490781 | -4.80789 |
| Coefficients | -0.2208 | 2.004551 | 2.264139 | -25.1 |
| Std. Error | 0.046886 | 0.408319 | 0.504175 | 5.2 |
| R ² | 0.81512 | 0.228905 | #N/A | #N/A |
| F df | 33.80169 | 23 | #N/A | #N/A |
| Residuals | 5.313375 | 1.205143 | #N/A | #N/A |

Note d—stopping distance, v—speed and t—travel time

4.3 Sensitivity Analysis

Assume distance to stop line, $d = 50$ m; yellow interval time, $t = 3$ s and posted speed limit,

$$v = 16.67 \text{ m/s (60 km/h)}$$

where $Y = -25.1 + 2.264t + 2.0045v - 0.2208d$ from Table 1

$$\begin{aligned} &\text{Probability of RLR, } P_{RLR}(Y = 1) \\ &= \frac{e^{\beta_0 + \beta_1 t + \beta_2 v + \beta_3 d}}{1 + e^{\beta_0 + \beta_1 t + \beta_2 v + \beta_3 d}} = \frac{e^{-25.1 + 2.2641t + 2.0045v - 0.2208d}}{1 + e^{-25.1 + 2.2641t + 2.0045v - 0.2208d}} = 2\% \end{aligned}$$

$$\begin{aligned} &\text{Probability of stopping, } P_s(Y = 0) \\ &= \frac{1}{1 + e^{\beta_0 + \beta_1 t + \beta_2 v + \beta_3 d}} = \frac{1}{1 + e^{-25.1 + 2.2641t + 2.0045v - 0.2208d}} = 98\%. \end{aligned}$$

The summary for all sites is in Tables 2, 3 and 4. The probability of red-light running during light rainfall is 2%, moderate rainfall is 1%, and 0% during heavy rainfall. There is a significant reduction in the probability of red-light running during rainfall because of the speed reduction caused by rainfall (Tables 5 and 6).

During dry weather, the average probability of RLR is 14.9%, and the probability of stopping is 85.1% compared to the rainy condition where the average probability of RLR 2% for light rain, 1% for moderate rain and 0% heavy rain. The average probability of stopping is 98% for light rain, 99% for moderate rain 100% for heavy rain. The average time needed to safely cross the stop line at the onset of yellow time interval increased significantly from 3 s during dry weather to 3.73 s for light rain, 3.94 s for moderate rain and 5.03 s for heavy rain.

Table 4 Summary of probability RLR during light rainfall

| Site | Model equations | R ² | Y | e ^y | Probability (%) | |
|------|---|----------------|--------|----------------|-----------------|----------|
| | | | | | RLR | Stopping |
| Dry | Y = -8.71 - 0.1474d + 0.8487v + 0.6656t | 0.72 | 0.0646 | 0.1756 | 14.9 | 85.1 |
| 001 | Y = -25.1 - 0.2208d + 2.0045v + 2.2641t | 0.81 | -11.04 | 0.017118 | 2 | 98 |
| 002 | Y = -24.987 - 0.205d + 1.9261v + 2.292t | 0.70 | -10.25 | 0.023598 | 2 | 98 |
| 003 | Y = -27.466 - 0.2077d + 1.998v + 2.765t | 0.82 | -10.39 | 0.023266 | 2 | 98 |

Note RLR denotes red-light running, d—stopping distance, v—speed and t—travel time

Table 5 Summary of probability RLR during moderate rainfall

| Site | Model equations | R ² | Y | e ^y | Probability (%) | |
|------|---|----------------|--------|----------------|-----------------|----------|
| | | | | | RLR | Stopping |
| Dry | Y = -8.71 - 0.1474d + 0.8487v + 0.6656t | 0.72 | 0.0646 | 0.1756 | 14.9 | 85.1 |
| 01 | Y = -27.443 - 0.2204d + 2.0726v + 2.6402t | 0.81 | -11.02 | 0.018163 | 1 | 99 |
| 02 | Y = -32.329 - 0.2227d + 2.2236v + 3.4277t | 0.72 | -11.14 | 0.020587 | 2 | 99 |
| 03 | Y = -37.291 - 0.2967d + 2.7656v + 3.7274t | 0.71 | -14.83 | 0.005739 | 1 | 99 |

Note RLR denotes red-light running, d—stopping distance, v—speed and t—travel time

5 Conclusions

The study aims to assess the influence of rainfall on red-light running at a signalised intersection. Reduced visibility, reduced speed, increased stopping distance, reduced

Table 6 Summary of probability RLR during heavy rainfall

| Site | Model equations | R ² | Y | e ^y | Probability (%) | |
|------|--|----------------|--------|----------------|-----------------|----------|
| | | | | | RLR | Stopping |
| Dry | Y = -8.71 - 0.1474d + 0.8487v + 0.6656t | 0.72 | 0.0646 | 0.1756 | 14.9 | 85.1 |
| 01 | Y = -41.2951 - 0.3318d + 3.0716v + 4.14t | 0.78 | -16.59 | 0.003234 | 1 | 99 |
| 02 | Y = -42.0546 - 0.2753 + 2.789v + 4.65t | 0.79 | -13.77 | 0.009816 | 0 | 100 |
| 03 | Y = -57.0331 - 0.4608d + 4.2267v + 5.79t | 0.73 | -23.04 | 0.000422 | 0 | 100 |

Note RLR denotes red-light running, d—stopping distance, v—speed and t—travel time

traction, and extension deceleration was anticipated during rainfall. The key objectives are determining the probability of red-light running under dry, rainy conditions and comparing the outcomes. Modified passenger car equivalent values were used to convert observed traffic volumes to flows. Generally, rainfall caused an average speed reduction from 80 km/h during dry weather to 50 km/h (37.5%) during light rain; to an average speed of 40 km/h (50%) during moderate rain and an average speed of 33 km/h (58.7%) during heavy rain. The probability of red-light running diminished with rainfall irrespective of intensity. RLR’s probability during a light rain is 2%, moderate rain is 1%, and heavy rainfall is 0%. The average time needed to safely cross the stop line at the onset of yellow time interval increased significantly from 3 s during dry weather to 3.73 s for light rain, 3.94 s for moderate rain and 5.03 s for heavy rain. Thus, suggesting approach vehicles cannot enter the signalised intersection safely, hence must wait at the stop line for a green signal. The paper concluded that it is suicidal to attempt red-light running during rainfall, none recorded at all selected sites. In any case, it is recommended that future studies be conducted on the impact of rainfall on red-light running given variable yellow time interval.

Acknowledgements The University of KwaZulu-Natal, Durban, South Africa, sponsored this work as part of the Department of Civil Engineering PhD research projects. This paper’s contents reflect the authors’ views, responsible for the facts and accuracy of the data presented herein.

References

1. Arash J, Rakha H, Thomas AD (2016) Red-light running violation prediction using observational and simulator data. *Accid Anal Prev* 316–328
2. Ben-Edigbe J, Ferguson N (2005) The extent of capacity loss resulting from pavement distress. *Proc Inst Civ Eng Transp* 158(1):27–32
3. Ben-Edigbe J (2010) Assessment of speed-flow-density functions under adverse pavement condition. *Int J Sustain Dev Plan* 5(3):238–252
4. Ben-Edigbe J, Mashros N, Rahman R (2013) Extent of sight distance reductions caused by rainfall on single carriageway roads. *Int J Traffic Transp* 3(3):291–301
5. Cohn EG, Kakar S, Perkins C, Steinbach R, Edwards P (2020) Red light camera interventions for reducing traffic violations and traffic crashes. A systematic review. *Campbell Syst Rev* 16:2. <https://doi.org/10.1002/cl2.1091>
6. EL-Sharwaby I, Abdel-Salam G, Rakha H (2013) Evaluation of driver perception-reaction time under rainy or wet roadway conditions at onset of yellow indication. *J Transp Res Board* 18–24
7. Elmitiny N, Yan X, Radwan E, Russo C, Nashar D (2010) Classification analysis of driver's stop/go decision and red-light running violation. *Accid Anal Prev* 101–111
8. Gates TJ, Noyce DA, Laracuente L, Nordheim EV (2007) Analysis of driver behaviour in dilemma zones at signalized intersections. *Transp Res* 29–39
9. Gates TJ, McGee H, Moriarty K, Maria H (2012) A comprehensive evaluation of driver behaviour to establish parameters for the timing of yellow change and red clearance intervals. *Transp Res Record: J Transp Res Board* 2298(1):9–21
10. Hurwitz DS, Wang H, Knodler MA Jr, Ni D, Moore D (2012) Fuzzy sets to describe driver behaviour in the dilemma zone of high-speed signalised intersections. *Transp Res F: Traffic Psychol Behav* 15(2):132–143
11. Li H, Rakha H, EL-Shawaby I (2012) Designing yellow intervals for rainy and wet road conditions. *Int J Transp Sci Technol* 2(1):171–189
12. Rakha H, Amer A, El-Shawarby I (2008) Modelling driver behaviour within a signalised intersection approach decision dilemma zone. *Transp Res Rec* 2069(1):16–25
13. Long L, Liu Y, Han LD (2013) Impact of countdown timer on driving manoeuvres after the yellow onset at signalised intersections: an empirical study in Changsha, China. *Saf Sci* 55:8–16
14. Mashros N, Ben-Edigbe J, Hassan SA, Hassan NA (2014) Impact of rainfall condition on traffic flow and speed: a case study in Johor and Terengganu. *Jurnal Teknologi* 70:4
15. Moore D, Hurwitz S (2013) Fuzzy logic for improved dilemma zone identification. *Transp Res Rec* 2348(1):25–34
16. Oyaro J, Ben-Edigbe J (2020) Extent of capacity loss at signalised intersections caused by rainfall. *Open Transp J* 14:1. <https://doi.org/10.2174/1874447802014010214>
17. Perrin J, Martin P, Hansen B (2001) Modifying signal timing during inclemently weather. In: Annual meeting of Transportation Research Board—TRB, Washington, DC
18. SARTSM (2012) In: Transport NDO (ed) South African road traffic signs manual, vol 3: traffic signal design. Committee of Transport Officials (COTO), Pretoria, South Africa
19. Shao CQ, Rong J, Liu XM (2011) Study on the saturation flow rate and its influence factors at signalised intersections in China. *Procedia Soc Behav Sci* 504–514. <https://doi.org/10.1016/j.sbspro.2011.04.471>
20. Sun H, Yang J, Wang L, Li L, Wu B (2013) Saturation flow rate and start-up lost time of dual-left lanes at signalized intersection in rainy weather condition. In: International conference of transportation professionals, pp 270–279. <https://doi.org/10.1016/j.sbspro.2013.08.034>
21. Wang J, Dixon K, Li H, Ogle J (2005) Normal deceleration behaviour of passenger vehicles at stop sign-controlled intersections evaluated with in-vehicle global positioning system data. *Transp Res Rec: J Transp Res Board* 1937:120–127
22. Yang Z, Tian X, Wang W, Zhou X, Liang H (2014) Research on driver behaviour in yellow interval at signalized intersections. *Math Probl Eng*. <https://doi.org/10.1155/2014/518782>

23. Zhang L, Zhou K, Zhang W-B, Misener JA (2008) Empirical observations of red light running at arterial signalized intersection. University of California, Berkeley. <https://rosap.nrl.bts.gov/view/dot/20332>
24. Zhang Y, Yan X, Wu J, Dixit VV (2018) Red-light running crashes' classification, comparison and risk analysis based on General Estimates System (GES) crash database. *Int J Environ Res Public Health* 15(6):1290

Pressure Distribution in a Dam-Break Flow Simulated by a Mesh-Free Method



Tibing Xu, Jinlong Zuo, and Zhijian Huang

1 Introduction

With climate change, dam-break flows frequently occur, posing threats to the downstream communities. Numerical simulation can provide a useful and robust tool to figure out the threatened areas by the dam-break flow. However, the dam-break flow usually involves large deformation and severe fragmentation of free surface, which could be difficult to be captured by numerical schemes.

In recent years, the mesh-free method has become a robust numerical tool to simulate violent free-surface flows [2, 9]. In the method, the fluid is discretized into a set of particles or material points and they can move according to the velocity that they take. In the mesh-free method, there are two schemes to calculate the pressure field, one is to use the equation of state and the other is to solve the pressure Poisson equation. When using the equation of state, the incompressible fluid flow has to be treated as the weakly-compressible [5, 6, 8]. This scheme can cause numerical noise in the simulation which can be found in previous publications [7, 11]. This numerical noise can be eliminated by using the particle shifting technique [1] and the stabilization technique [11]. Xu and Jin [11] proposed a Laplacian model and a stabilization technique for the scheme to discretize the viscous term in the governing equations. They also applied this type of numerical scheme to simulate the broad-crested weir flow [12], fishway flow [10], and flow interaction with porous media [13].

In the other scheme to calculate the pressure field, the pressure Poisson equation is solved. In this scheme, a large matrix equation could be calculated in each time step as the semi-implicit scheme [3, 4, 14]. This scheme can be found in the Moving Particle Semi-implicit (MPS) method proposed by Koshizuka et al. [3] and applied

T. Xu (✉) · J. Zuo · Z. Huang
School of Civil and Environmental Engineering, Ningbo University, Ningbo, China
e-mail: xu320@uregina.ca; xutibing@nbu.edu.cn

into hydrodynamics modeling by other researchers [8, 10, 12]. Solving the matrix equation requires considerable time in the simulation and reduces the computing efficiency. The implicit scheme is also difficult to be implemented in parallel computing. To overcome the difficulty to obtain pressure field implicitly in the mesh-free method, Ye et al. [15] proposed an explicit scheme to solve the pressure Poisson equation by using the Laplacian model proposed by Xu and Jin [11]. In their simulation, the smooth pressure field can be achieved in simulating dam-break flow over dry/wet downstream conditions. The free surface and wave front movement are reproduced by the method agreeing with experimental measurements well.

In this study, the numerical scheme proposed by Ye et al. [15] is employed to simulate a dam-break flow. The simulated pressure distribution at different time steps and locations is investigated in the flow. The remained parts for the study are arranged as: the method is presented in Sect. 2, numerical results are illustrated in Sect. 3, and following by conclusions.

2 Mesh-Free Method

In this section, the governing equations and discretized scheme are introduced.

2.1 Governing Equations

The governing equations are expressed in the framework of the Lagrangian method as

$$\frac{1}{\rho} \frac{D\rho}{Dt} = \nabla \cdot \mathbf{u} = 0 \quad (1)$$

$$\frac{D\mathbf{u}}{Dt} = -\frac{1}{\rho} \nabla p + \mu \nabla^2 \mathbf{u} + \mathbf{g} \quad (2)$$

where ρ is the density, t is time, \mathbf{u} the velocity vector, p is the pressure, μ is the viscosity, and \mathbf{g} the gravity.

2.2 Mesh-Free Method

In the mesh-free method, the fluid is discretized into a set of particles or material points with the same distance Δl , that take the fluid properties and flow information such as velocity and pressure. These discretized particles are indexed as i and j . Here, i denotes the target particle in the calculation and j represents the neighboring

particles. To weigh the contribution from the neighboring particles, the weighting function is adopted. In this study, the weighting function is expressed as follows:

$$w_{ij} = \begin{cases} \left(1 - \frac{r_{ij}}{r_e}\right)^3 & r_{ij} < r_e \\ 0 & \text{else} \end{cases} \tag{3}$$

where r_{ij} the distance between particle i and particle j , and r_e is the interaction circle. The weighting function defines an interaction circle in which the neighboring particles j can interact with particle i . The interaction circle has a constant radius in the whole calculation as r_e . r_e is related to the discretized particle distance Δl . In the previous studies [11], $r_e = 4.0\Delta l$.

The particle number density is defined as the summation of the weighting function values as

$$\langle n \rangle_i = \sum_{j \neq i} w_{ij} \tag{4}$$

In the initial condition, when the particles are in the regular distribution with the same distance Δl , the particle number density is calculated as the initial particle number density n_0 .

According to the weighting function, the gradient model to discretize the pressure gradient term in the governing equations can be developed which is expressed as [3]:

$$\langle \nabla \varphi \rangle_i = \frac{D_m}{n_0} \sum_{j \neq i} \frac{\varphi_j - \varphi_i}{r_{ij}^2} \mathbf{r}_{ij} w_{ij} \tag{5}$$

where D_m is the dimension coefficient, φ is a scalar, and \mathbf{r}_{ij} is the position vector.

By taking the divergence of the gradient model, the Laplacian model can be derived [11]:

$$\langle \nabla^2 \varphi \rangle_i = \frac{D_m}{n_0} \sum_{j \neq i} \frac{\varphi_j - \varphi_i}{r_{ij}^2} G_{ij} \tag{6}$$

$$G_{ij} = (4D_m - 2)w_{ij} + r_{ij} \frac{\partial w_{ij}}{\partial r_{ij}} \tag{7}$$

The Laplacian model can be used to discretize the viscous term in the governing equations.

To solve the governing equations, the predictor-and-corrector time splitting scheme is adopted. That means that in the predictor, the viscous term and gravity force term are solved to obtain the intermediate flow information such as intermediate velocity, intermediate particle position, and intermediate particle number density.

The equations can be obtained in the predictor as:

$$\mathbf{u}_i^* = \mathbf{u}_i^k + \Delta t (\mu \langle \nabla^2 \mathbf{u}^k \rangle_i + \mathbf{g}) \quad (8)$$

$$\mathbf{r}_i^* = \mathbf{r}_i^k + \Delta t \mathbf{u}_i^* \quad (9)$$

where * denotes the intermediate time step, k is the previous time step, and Δt is the time step.

After getting the intermediate flow information, with the continuity equation, the following pressure Poisson equation can be obtained as

$$\langle \nabla^2 p \rangle_i^{k+1} = \frac{\rho}{\Delta t} \langle \nabla \cdot \mathbf{u} \rangle_i^* \quad (10)$$

By using a small time step, according to the Laplacian model and gradient model, an explicit relationship between the pressure field and the known flow information can be established as [15]:

$$p_i^{k+1} = \frac{D_m \sum_{j \neq i} \frac{p_j^k G_{ij}^*}{(r_{ij}^*)^2} - \frac{\rho}{\Delta t} \left(\frac{Dn}{Dt} \right)_i^*}{D_m \sum_{j \neq i} \frac{G_{ij}^*}{(r_{ij}^*)^2}} \quad (11)$$

In the corrector, the pressure gradient can be solved with the pressure field calculated in the predictor so that the velocity and particle position can be updated as:

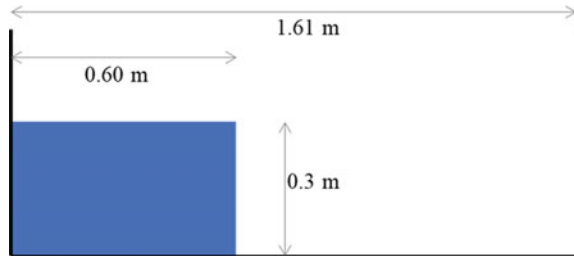
$$\mathbf{u}_i^{k+1} = \mathbf{u}_i^* + \Delta t \langle \nabla p \rangle_i^{k+1} \quad (12)$$

$$\mathbf{r}_i^{k+1} = \mathbf{r}_i^{k+1} + \Delta t \mathbf{u}_i^{k+1} \quad (13)$$

where $k + 1$ represents the next time step.

By using Eq. 11, the pressure field can be calculated explicitly, and the incompressible condition is satisfied. In this sense, the equation of state to calculate the pressure field is not necessary and the weakly-compressible assumption to model the incompressible fluid flow is avoided. This numerical scheme is used to simulate the dam-break flow over the dry and wet beds by Ye et al. [15]. The pressure field is smooth with less noise and the free-surface profiles and wave front movement are shown to agree with experimental measurements in their study.

Fig. 1 Diagram for the dam-break flow



3 Pressure Distribution in the Simulated Dam-Break Flow

In this section, the mesh-free method is used to simulate a dam-break flow and the pressure distribution in the simulation is analyzed.

3.1 Diagram for the Flow

The diagram for the dam-break flow is shown in Fig. 1. In a closed system, the water column is initially set up with a length of 0.6 m and a height of 0.3 m. The length of the system is 1.61 m. From $t = 0.0$ s, the water column collapses, propagating to the right side as the dam-break flow. After impacting the right wall, a dam-break surge can be formed.

The proposed mesh-free method is used to simulate this dam-break flow. In the simulation, the particle distance $\Delta l = 0.0025$ m is used to discretize the fluid column. The interaction radius $r_e = 4.0\Delta l$ and the time step satisfies the CFL criterion.

3.2 Flow Pattern

The flow patterns at different time steps for the dam-break flow in the simulation are shown in Fig. 2. It shows that the proposed numerical scheme can calculate a smooth pressure field. The water propagates to the right side from the initial condition and the water impacts against the solid wall, causing a reversed wave to the left side. The free surface during the reversed wave traveling backward is very fragmented in the simulation which can be observed in the figure at $t = 1.6$ s and 2.0 s. The mesh-free method can calculate the dam-break flow to reflect the flow characteristics.

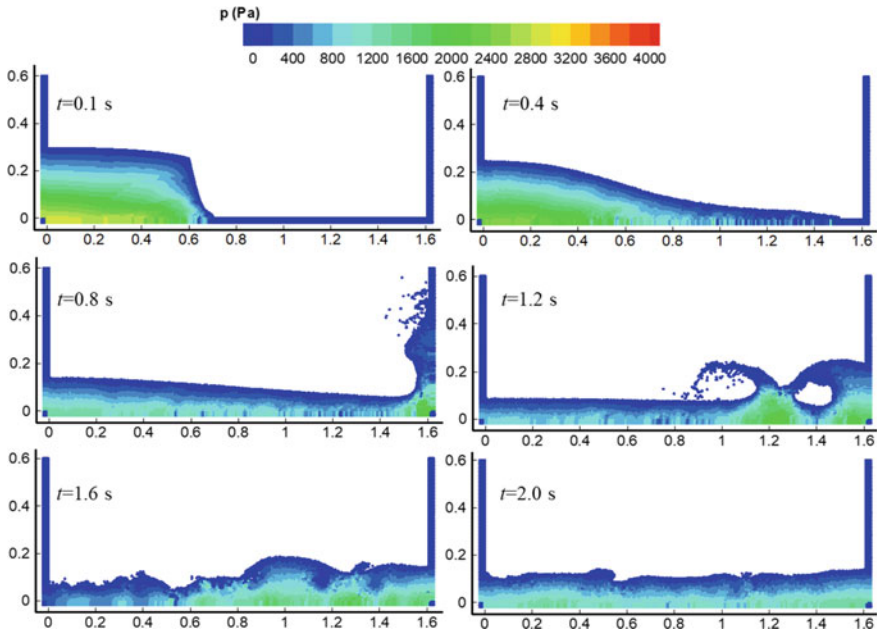


Fig. 2 Simulated flow patterns for the dam-break flow

3.3 Pressure Distribution

After examining the simulated flow patterns in the dam-break flow, the pressure distribution in the flow at different vertical sections is investigated, including the sections $x = 0.25$ m, 0.75 m, 1.0 m, and 1.25 m. The pressure is dimensionless as p/p_{max} and the distance at the vertical section is dimensionless as y/y_{max} . The pressure distribution is shown in Fig. 3. At the section $x = 0.25$ m, the pressure distribution is almost the linear trend at different time steps in the simulation. At $x = 0.75$ m, the pressure distribution is also linear except for the time step $t = 0.15$ s. This is because, at this time step, the wave just arrived at this section. At sections $x = 1.0$ m and 1.25 m, the pressure distribution approximates the linear profile. However, when at the time when the section is close to the wave front, the pressure deviates from the linear distribution.

4 Conclusions

In this study, an explicit mesh-free method is employed to simulate a dam-break flow. It is found that the numerical method can reflect the flow patterns with free surface large deformation and fragmentation. The simulated pressure field is very smooth

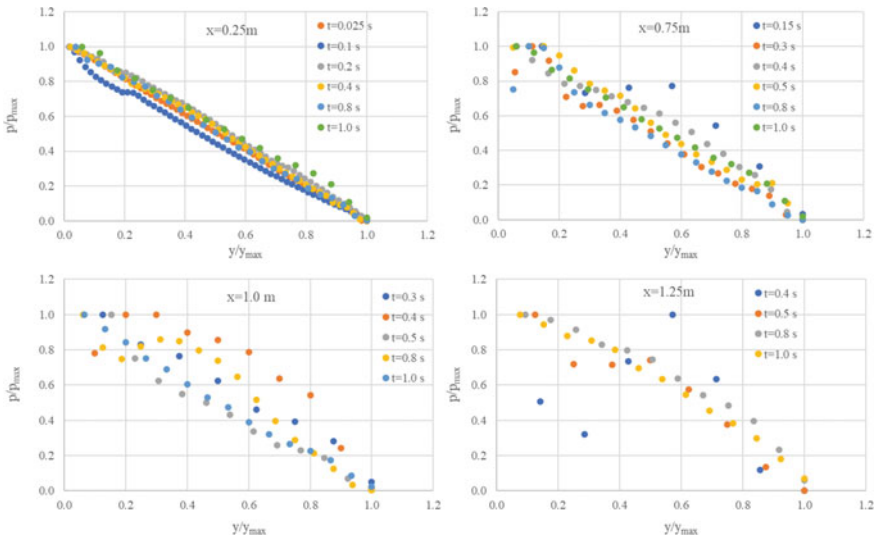


Fig. 3 Pressure distribution during the dam-break flow

with less noise which is different from the explicit calculation by using the equation of state. The pressure at different vertical sections in the dam-break is investigated and found that the pressure is almost the linear distribution at the sections. For sections close to the wave front in the flow, the pressure distribution deviates from the linear distribution.

References

1. Antuono M, Sun PN, Marrone S, Colagrossi A (2021) The δ -ALE-SPH model: an arbitrary Lagrangian-Eulerian framework for the δ -SPH model with particle shifting technique. *Comput Fluids* 216:104806
2. Gotoh H, Khayyer A (2018) On the state-of-the-art of particle methods for coastal and ocean engineering. *Coast Eng J* 60(1):79–103
3. Koshizuka S, Nobe A, Oka Y (1998) Numerical analysis of breaking waves using the moving particle semi-implicit method. *Int J Numer Methods Fluids* 26(7):751–769
4. Khayyer A, Gotoh H, Shimizu Y (2017) Comparative study on accuracy and conservation properties of two particle regularization schemes and proposal of an optimized particle shifting scheme in ISPH context. *J Comput Phys* 332:236–256
5. Monaghan JJ (1994) Simulating free surface flows with SPH. *J Comput Phys* 110(2):399–406
6. Monaghan JJ (2012) Smoothed particle hydrodynamics and its diverse applications. *Annu Rev Fluid Mech* 44:323–346
7. Oger G, Marrone SL, Touzé D, De Leffe M (2016) SPH accuracy improvement through the combination of a quasi-Lagrangian shifting transport velocity and consistent ALE formalisms. *J Comput Phys* 313:76–98
8. Shakibaeinia A, Jin YC (2010) A weakly compressible MPS method for modeling of open-boundary free-surface flow. *Int J Numer Methods Fluids* 63(10):1208–1232

9. Violeau D, Rogers BD (2016) Smoothed particle hydrodynamics (SPH) for free-surface flows: past, present and future. *J Hydraul Res* 54(1):1–26
10. Xu T, Jin YC (2014) Numerical investigation of flow in pool-and-weir fishways using a meshless particle method. *J Hydraul Res* 52(6):849–861
11. Xu T, Jin YC (2016) Improvements for accuracy and stability in a weakly-compressible particle method. *Comput Fluids* 137:1–14
12. Xu T, Jin YC (2017) Numerical study of the flow over broad-crested weirs by a mesh-free method. *J Irrig Drain Eng* 143(9):04017034
13. Xu T, Jin YC (2019) Modeling impact pressure on the surface of porous structure by macroscopic mesh-free method. *Ocean Eng* 182:1–13
14. Xu T, Jin YC (2019) Improvement of a projection-based particle method in free-surface flows by improved Laplacian model and stabilization techniques. *Comput Fluids* 191:104235
15. Ye Y, Xu T, Zhu DZ (2020) Numerical analysis of dam-break waves propagating over dry and wet beds by the mesh-free method. *Ocean Eng* 217:107969

Estimates of River Bathymetry from Satellite Images: A Case Study of the Nicolect River in Quebec



Shayan Salavitarbar and S. Samuel Li

1 Introduction

Morphology, sediment transport and flow condition are the main parameters that shape the form and behaviour of rivers. Consequently, knowledge of topography, flow condition and bed material properties is the main element in river studies. Nevertheless, in many cases, some parts of a river are inaccessible in remote locations. Moreover, the inadequacy of funding sources puts a limitation on a traditional investigation of river topography [9]. In comparison to some common field measurements for bathymetric or topographic data, satellite imagery can provide remote sensing approach to cover a large area of a river with a limited amount of field measurements [4].

Remote sensing bathymetry is based on the total amount of radioactive energy reflected from a water column. This means that once sunlight broadcasts through the water, water constituents and molecules absorb and scatter sun-light, then backscatters leaving energy are recorded by satellite sensors [5].

Three decades ago, Landsat multi-spectral satellite imagery and airborne photography were known as the only source for remote sensing of water bodies. Although airborne image can provide high-resolution data either spatially or spectrally but deployment issues and high costs of application put restrictions on them for general remote sensing purposes. On the other hand, the global coverage of Landsat-7 provides only a 30-m resolution. Later, IKONOS and QuickBird satellites in 1999 and 2004 provided an enhanced resolution to 4-m or better for remote sensing purposes [13]. It seems that a 4-m resolution is enough for ocean and coastal bathymetry

S. Salavitarbar (✉) · S. S. Li

Department of Building, Civil and Environmental Engineering, Concordia University, 1455 de Maisonneuve Blvd. W., Montreal, QC 3G 1M8, Canada

e-mail: shayan.salavitarbar@gmail.com

© Canadian Society for Civil Engineering 2022

S. Walbridge et al. (eds.), *Proceedings of the Canadian Society of Civil Engineering*

Annual Conference 2021, Lecture Notes in Civil Engineering 250,

https://doi.org/10.1007/978-981-19-1065-4_33

studies because of large scales. In river bathymetry studies, however, due to ecological importance of vertical structures and relatively small horizontal spatial scales, higher resolution of satellite images is needed [13]. In this paper, to increase the accuracy of river bathymetry, multi-spectra images of satellite World View3 with a 1.2-m resolution are selected.

The relationship between captured satellite images and bathymetry begins from the translation of upwelling spectral radiance into digital image data by describing a ratio-band algorithm that reveals a linear quantity related to water depth. Absolute depths obtained from this approach require an additional coefficient to calibrate the relationship between the band ratio and the water depth [10]. This coefficient is called attenuation coefficient, K_d . It is a critical parameter for the light intensity at depth. For many remote sensing studies, especially in the case of ocean studies, the standard method for estimating the attenuation coefficient is by comparing water depth with the spectral ratio between two water leaving wavelengths [1].

This paper represents an initial attempt to describe the bathymetry of a river through using multi-spectral high-resolution images and algorithms for the attenuation coefficient. This is different from previous studies, which rely on comparing measured depth to band ratio in order to obtain the attenuation coefficient. By adopting existing analytical algorithms for ocean application and testing them on rivers, this paper introduces a novel method for deriving river bathymetry from high-resolution satellite images without the necessity of using depth measurements.

2 Methodologies

2.1 Study Area

The Nicolet River (Fig. 1) is located within the Nicolet subbasin, in Quebec, Canada. This river is approximately 150 km long and is estimated to have a drainage area of 265 km². The nearby area is typically coniferous trees and agricultural grass. In the 1980s, biological diversity in fish population decreased in the Nicolet River and it was noticed by the Government of Québec. To protect the population of fish, La Corporation de gestion des rivières des Bois-Francs (CGRBF) undertook the task to create a fish habitat to restore their population. The outcome was the installation of a set of deflectors in V-shape in a reach upstream [12].

In this paper, the study area consists of an upstream part of the Nicolet River that flows between the Municipality of Norte-Dame-de-Ham and the Town of Victoriaville, which includes the deflectors. The deflector section is about 1.3 km long. The river width is around 35 m near the deflector; and the discharge varies from 0.6 to 30 m³/s [3].

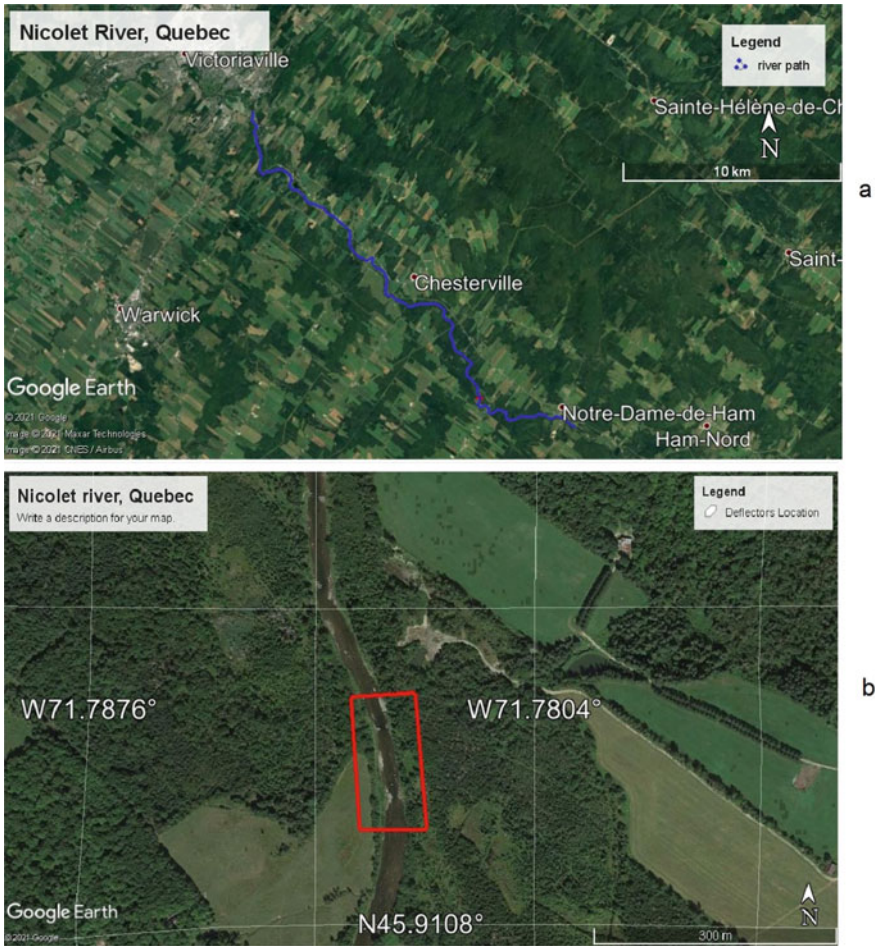


Fig. 1 A map showing the region of the Nicolet River: **a** site of study; **b** detailed location of the deflectors in the river channel (image downloaded from Google Earth on March 1, 2021)

2.2 *Determination of Channel Bathymetry Through Satellite Image*

This paper processed images of the WorldView-3 (WV3) high-resolution commercial imaging satellite. This satellite was launched on August 13, 2014, from Vandenberg Air Force Base. The worldView-3 sensor acquires 14-bit data in multi-spectral bands covering panchromatic, coastal, blue, green, yellow, red, red edge, NIR1, and NIR2. Images are commercially available at a 0.3-m resolution for panchromatic, and a 1.2-m for multispectral images. Images for the Nicolet River region were captured on 20 May 2016 with the condition of clear sky without any cloud coverage with Sun

Elevation (SE) of 62.30° and off-Nadir 26.50° . To increase the bathymetry mapping accuracy, geometric and radiometric corrections for high-resolution WV3 imagery should be applied to obtain reliable output results. This step of preprocessing was implemented by using ArcGIS to convert the raw data to practical data for river bathymetry.

2.2.1 Geometric Correction

Geometric correction is one of the essential steps of satellite image processing to match point coordinates to their real geographic positions. The implement of the Digital Elevation Model (DEM) images with actual coordinates of points to satellite images replaced image pixels in their correct locations. To validate that pixels were in the correct locations after geometric correction, the river path was created in Google Earth and attached to images before and after geometric correction. In Fig. 2, a comparison of the Nicolet River flow path between before- and after-correction is shown. The correction made a small shift in the channel centerline (flow path).

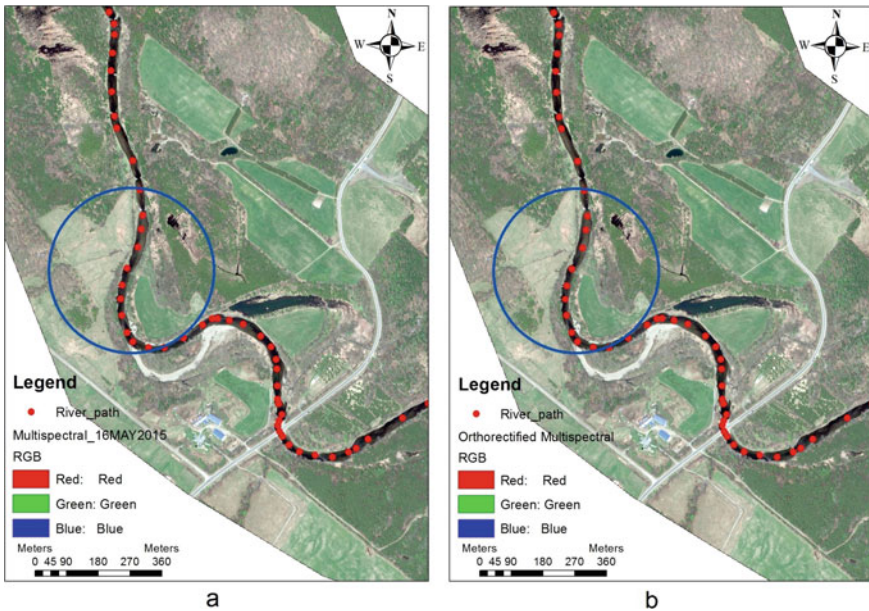


Fig. 2 A comparison between satellite images: **a** without geometric correction; **b** with geometric correction. Subtle differences can be discerned in the area marked by a blue circle

2.2.2 Conversion to Top-of-Atmosphere Reflectance

Uniform scene does not create a uniform image in terms of Digital Numbers (DN). To implement analysis technique for band ratio and Normalized Difference Water Index (NDWI), DN of images were converted to Top-of-Atmosphere Radiance, L, then to Top-of-Atmosphere Reflectance through Eqs. 1 and 2, respectively

$$L = \text{Gain} \times \text{DN} \times (\text{abscal factor/effective band width}) + \text{Offset} \tag{1}$$

$$\rho = \frac{LD^2\pi}{E_\lambda \text{Cos}\Theta_s} \tag{2}$$

where the Gain and Offset are the absolute radiometric calibration band dependent adjustment factors [11], D is defined as Earth-Sun distance; E_λ is known as Solar Exoatmospheric Irradiance; Θ_s is solar zenith angle, obtained from Eq. 3

$$\Theta_s = 90 - \text{sunElv} \tag{3}$$

2.3 Image Data and Processing

Before using any depth measurements, the river-channel boundary was specified through NDWI to discriminate wet pixels from other pixels. Then, for bathymetry detection, this paper followed the approach based on a ratio of two bands that has greater dissimilarities in terms of water absorption properties [9]. Therefore, red band (630–690 nm) and green band (510–580 nm) were extracted from multi-spectra image to obtain depth from the ratio given in Eq. 4

$$\text{Depth} = K_d \text{Ln } R(\lambda_r)/R(\lambda_g) \tag{4}$$

where $R(\lambda_r)$ and $R(\lambda_g)$ are the reflectance of red and green bands, respectively; K_d is the effective diffuse attenuation coefficient, which is defined as the difference between attenuation coefficients of red band and green band [4]

$$K_{de} = K_{dr} - K_{dg} \tag{5}$$

Unlike the common empirical algorithms based on the relationship between red-green ratio of water leaving reflectance, the strategy in this paper works on absorption coefficient, a, and backscattering coefficient, b_b , as indispensable properties to determination of K_d value. K_d can be calculated from semi-analytical models through using the wavelength of green band (555 nm) as a reference wavelength [7].

This paper took the same approach as in ocean application. In the case of ocean application, Lee et al. [6] indicated that the correlation of K_d with the absorption and backscattering coefficients can be obtained through Quasi-Analytical Algorithm (QAA):

$$K_d = m_0 + m_1(1 - m_2e^{-m_3a})b_b \tag{6}$$

In this equation, Lee et al. [7] introduced $m_0 = 1 + 0.005\Theta_s$, $m_1 = 4.18$, $m_2 = 0.52$, and $m_3 = 10.8$. Since the solar zenith angle in the captured satellite image of the Nicolet River was 27.7° , then m_0 was calculated as 1.1385.

This paper used $\lambda_0 = 555$ nm as the reference band. Dissolved and suspended constituents have no contribution in total absorption $a(\lambda)$ [6], leading to

$$a(\lambda) = \frac{[1 - u(\lambda)](b_{bw} + b_{bp})}{u(\lambda)} \tag{7}$$

where b_{pb} is the backscattering coefficient of suspended particles; b_{bw} is the backscattering coefficient of pure seawater. The sum of these backscattering coefficients provides b_b as the total backscattering coefficient. Here, $u(\lambda)$ is just a ratio of the backscattering coefficient to the sum of absorption and backscattering coefficients [6]

$$b_{bp} = b_{bp}(555) \left(\frac{555}{\lambda} \right)^Y \tag{8}$$

$$b_{bp}(555) = \frac{u(555)a(555)}{1 - u(555)} - b_{bw}(555) \tag{9}$$

where b_w for $\lambda = 555$ nm and $\lambda = 585$ nm in green and red band spectral were taken as 0.0014 and 0.0011, respectively [2], giving

$$u(\lambda) = \frac{-g_0 + [g_0^2 + 4g_1r_{rs}(\lambda)]^{0.5}}{2g_1} \tag{10}$$

This paper used $g_1 = 0.01247$ and $g_0 = 0.0895$. Values of Y and r_{rs} [6] were

$$r_{rs} = R_{rs}/(0.52 + 1.7R_{rs}) \tag{11}$$

$$Y = 2.2 \left\{ 1 - 1.2 \exp \left[-0.9 \frac{r_{rs}(440)}{r_{rs}(555)} \right] \right\} \tag{12}$$

Table 1 Calculated coefficients of backscattering (from Eqs. 8 and 9), absorption (Eq. 7), reflectance (Eq. 11) and attenuation coefficient (Eq. 6) for green and red bands

| Coefficient | Green band | Red band |
|--------------|------------|----------|
| b_w | 0.0014 | 0.0011 |
| b_{bp} | 0.020028 | 0.021236 |
| b_b | 0.021428 | 0.022336 |
| r_{rs} | 0.003214 | 0.003992 |
| $u(\lambda)$ | 0.0342790 | 0.042135 |
| $a(\lambda)$ | 0.2712839 | 0.209963 |
| K_d | 0.1187767 | 0.275570 |

where R_{rs} is the above-surface remote-sensing reflectance; r_{rs} is the below-surface remote-sensing reflectance. The calculated values for green band as reference band and red band are listed in Table 1.

3 Results and Discussion

Using the constants listed in Table 1, Eq. 5 produces an effective diffuse attenuation coefficient as $K_{de} = 0.156793 \text{ m}^{-1}$ for each pixel of the river. Thus, the determination of riverbed elevation for every river pixel is possible through the implementation of calculated K_d to Eq. 4. The reliability of the bathymetry model was tested using the 1.3 km section of the Nicolet River containing the deflectors. This selection allowed a comparison of the calculated elevations with measured elevations reported in [3]. In their study, the measured elevation was presented as DEM. In this study, the estimated bathymetry was not in the form of DEM. For this reason, the nearest compatible datum elevation station was identified, and the estimated elevation was transferred into DEM as shown in Fig. 3. For the deflector section, the estimated elevations from this study were compared with measured elevations from [3] in Fig. 4.

The model for estimating riverbed elevations (Fig. 4b) was validated in the following steps: (1) select a rectangular set of pixels including six pixels in the horizontal direction (image orientation) and eight pixels in the vertical direction. (2) Determine the area of the selected section from the resolution (1.2 m) of the captured image, the horizontal dimension being 7.2 m, and the vertical dimension being 9.6 m. The selected section included 48 pixels. Among them, 33 pixels showed DEM values between 96.5 and 97 m (Fig. 5), which are compatible with the measured data. The remaining 15 pixels showed a DEM value larger than 97 m or smaller than 96.5 m, meaning shallower and deeper water depths, respectively, in comparison to the measurements. The good comparison of 33 out of 48 data points amounts to a 69% accuracy of the model. It is important to note that the model requires no field measurement as input to predict riverbed elevations. This is a significant advantage.

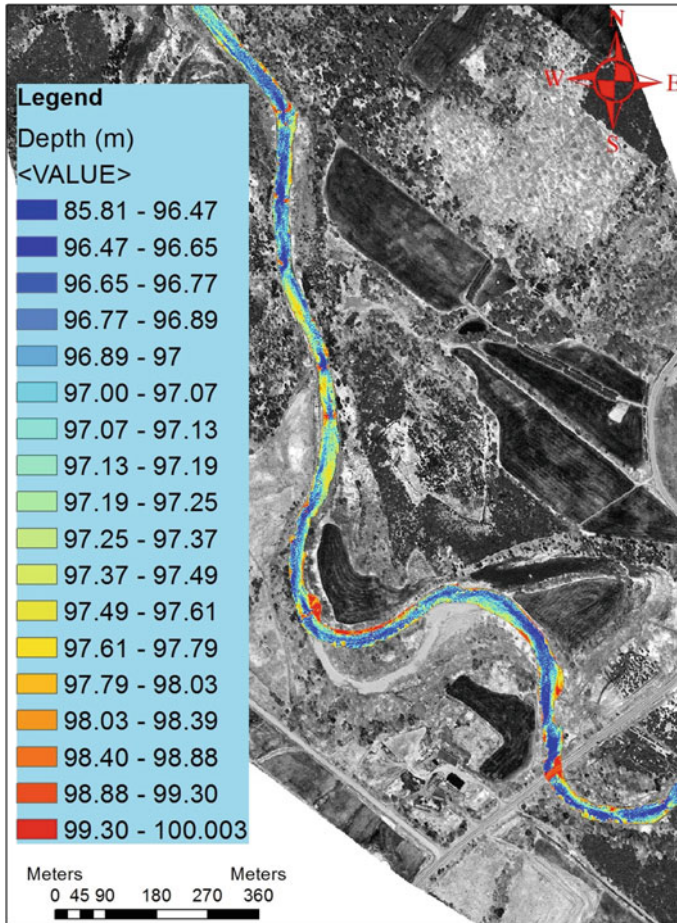


Fig. 3 Estimated elevation in form of DEM. The vertical datum is the tidal datum defined as the Mean Sea Level at five tidal gauges in the Pacific Ocean

The discrepancies are explained as follows. In Fig. 4a, the pool area between the two deflectors does not follow a regular shape. In modelled bathymetry (Fig. 4b), the pool area is a regular rectangle due to a limitation in the processor software for selecting pixels. Only a rectangular could be selected. As a result, there was a difference in shape between the selected area and the actual area. This limitation led to some data points outside the DEM range of 96.5–97 m. WV3 satellite imagery provides a high resolution of 1.2 m for each pixel. The measured DEM (Fig. 4a) provides approximately a 3-m resolution. This is 2.5 times less precise than the satellite image. It is understood that both field measurements and remote sensing data are subject to errors. The determination of errors is beyond the scope of this paper. For application to sediment transport studies, one would need very accurate

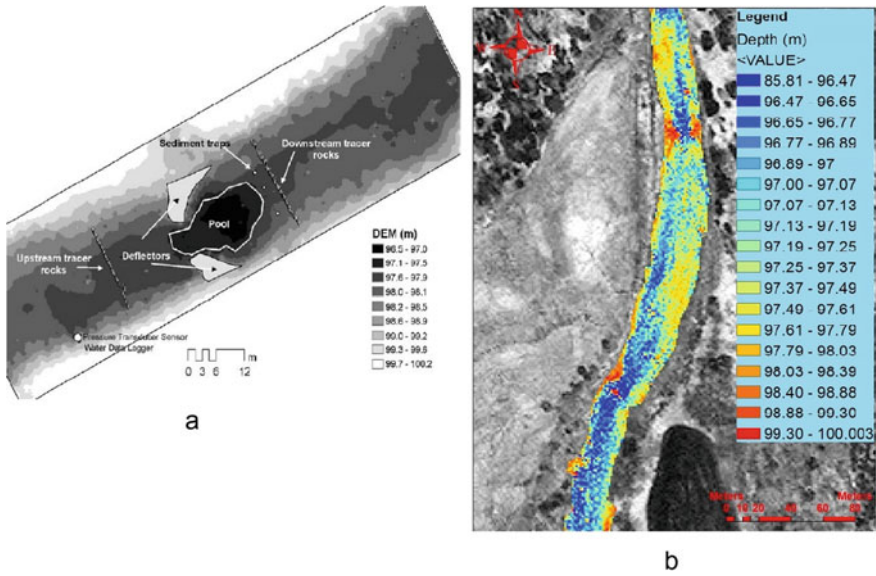


Fig. 4 a Measured elevation at the deflector location by [3], b estimated elevation from satellite images. The vertical datum is the same as in Fig. 3

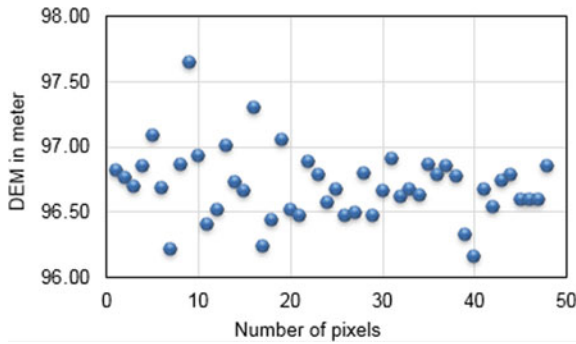


Fig. 5 Pixel value in digital elevation model

bathymetry as input. The model presented in this paper has the capacity to provide more details of elevations. This is especially useful for river flow regions where the depths vary rapidly. It is understood that the model uses flow depth to determine river bathymetry, which has dependency on fluvial condition.

4 Conclusion

One important aspect of river management is the preservation of the aquatic environment, which requires to identify locations of sediment erosion, deposition, and riverbank changes. To achieve preservation, it is essential to have accurate data of river bathymetry and its morphology. In complement to the traditional field methods for depth measurements, technology advancement offers a high quality of satellite images with multi-spectral bands within the visible spectrum as a new remote sensing approach in producing river bathymetry.

The conventional remote sensing approach compares measured depth to band ratio to define a coefficient and converts band ratio to the closet physical river depth. This paper demonstrates new methods without the need for depth measurement. The methods employ available analytical equations for the attenuation coefficient used in ocean application, along with radiometric correction and the use of logarithmic band ratio, for bathymetry estimates. This paper has demonstrated that for shallow rivers, the depth estimates are adequately close to field measurements.

In some image pixels of the flow, there are discrepancies in the estimated depths when compared to measured depths. One reason is that the selection of a comparison area by pixels is limited a rectangle, whereas the shape of the corresponding measurement area may not be exactly rectangular. Other plausible reasons are measurement errors and small shift in bathymetry mapping. The model for converting World View3 high-resolution multi-spectral images to flow depths works well, as demonstrated by a comparison for the pool location of the deflectors in the Nicolet River.

Acknowledgements This study received financial support from the Natural Sciences and Engineering Research Council of Canada through Discovery Grant held by S. Li.

References

1. Austin RW, Petzold TJ (1981) The determination of the diffuse attenuation coefficient of sea water using the coastal zone color scanner. Springer, New York, NY, USA
2. Buiteveld H, Hakvoort JHM, Donze M (1994) The optical properties of pure water. *Ocean Opt XII* 2258:174–183
3. Carré DM, Biron PM, Gaskin SJ (2007) Flow dynamics and bedload sediment transport around paired deflectors for fish habitat enhancement: a field survey in the Nicolet river. *Can J Civ Eng* 34(6):761–769
4. Harada S, Li SS (2018) Combining remote sensing with physical flow laws to estimate river channel geometry. *River Res Appl* 34(7):697–708
5. Jawak SD, Luis AJ (2015) Spectral information analysis for the semiautomatic derivation of shallow lake bathymetry using high-resolution multispectral imagery: a case study of Antarctic coastal oasis. *Aquat Procedia* 4:1331–1338
6. Lee ZP, Carder KL, Arnone RA (2002) Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters. *Appl Opt OSA* 41(27):5755–5772

7. Lee ZP, Darecki M, Carder KL, Davis CO, Stramski D, Rhea WJ (2005) Diffuse attenuation coefficient of downwelling irradiance: an evaluation of remote sensing methods. *J Geophys Res AGU* 110(C2):C02017:1–9. <https://doi.org/10.1029/2004JC002573>
8. Lee ZP, Du KP, Arnone R (2005) A model for the diffuse attenuation coefficient of downwelling irradiance. *J Geophys Res AGU* 110(C2):C02016:1–10. <https://doi.org/10.1029/2004JC002275>
9. Legleiter CJ, Overstreet BT (2012) Mapping gravel bed river bathymetry from space. *J Geophys Res AGU* 117(F4):1–24. <https://doi.org/10.1029/2012JF002539>
10. Legleiter CJ, Roberts DA, Marcus WA, Fonstad MA (2004) Passive optical remote sensing of river channel morphology and in-stream habitat: physical basis and feasibility. *Remote Sens Environ* 93(4):493–510
11. Kuester M (2016) Radiometric use of WorldView-3 imagery. DigitalGlobe, Longmont, Colorado, USA
12. Ng PTK (2005) Two-dimensional hydraulic-habitat modeling of a rehabilitated river. Department of Civil Engineering and Applied Mechanics, McGill University, Montréal, Quebec, Canada
13. Stumpf RP, Holderied K, Sinclair M (2003) Determination of water depth with high-resolution satellite imagery over variable bottom types. *Limnol Oceanogr Am Soc Limnol Oceanogr* 48(1):547–556

Numerical Simulation of Turbulent Offset Dense Jet Flow



Mostafa Bigdeli and Abdolmajid Mohammadian

1 Introduction

The number of in operation or being planned desalination plants has increased significantly in large coastal cities across the world given the increasing population growth, reduction in desalination costs, rapid advancements in desalination plant technology, and shortages of potable water [10]. The suspended particles and the dissolved minerals (in addition to salt) are removed by implementing the desalination process on seawater or saline water to obtain clean and potable water. On the other hand, the waste stream with higher concentrations, called brine, is returned (i.e., discharged back) to the coastal water bodies during the desalination process. Desalination can pose a risk of serious environmental crises given some of its environmental effects [1] such as destroying the surrounding ecosystem of the plant (i.e. natural habitat of marine species and fish cultures) [1, 4, 7, 16], brine and chemical disposal, high energy consumption, atmospheric pollutants emission, noise pollution, etc. [34]. As stated above, dense effluents are usually disposed of into coastal waters as submerged negatively buoyant jets [6]. The optimum brine submerged outfall, with a discharge angle of 60° to the horizontal seabed, has been reported widely in the previous experimental studies as obtaining the most appropriate mixing efficiency (i.e. maximum dilution) [28, 30, 31, 33, 42] since it can generate a curvilinear trajectory for the dense jet prior to the jet impaction with the seabed. However, given the cost and difficulty of inclined discharge construction, a horizontal discharge configuration has been suggested in engineering practices [35].

While a jet is discharged into a region over a wall in parallel to the jet exit axis, offset by a specified distance, turbulent offset jet flows are formed. The jet flow disposal may have a variable density that may be due to different fluid densities or

M. Bigdeli (✉) · A. Mohammadian
Department of Civil Engineering, University of Ottawa, Ottawa, ON K1N6N5, Canada
e-mail: mbigd068@uottawa.ca

© Canadian Society for Civil Engineering 2022
S. Walbridge et al. (eds.), *Proceedings of the Canadian Society of Civil Engineering Annual Conference 2021*, Lecture Notes in Civil Engineering 250,
https://doi.org/10.1007/978-981-19-1065-4_34

temperature variations [3]. Offset jets are considered to be a type of flow configuration prototype used in studying the physical features of complex flows that are employed in a large number of environmental and engineering applications, given their individual and complicated characteristics [3], and their application is relevant to energy dissipation tools used downstream of hydraulic structures. Since the offset jets are a combination of free, impingement, and wall jets, they are relatively more complex than other types of jet flows [3]. Many experimental studies can be found on two-dimensional offset jets [25] and three-dimensional wall jets [17, 18] in the literature.

Since the brine discharge is heavier than the surrounding water, the dense jet will finally move back onto the seabed. A density current is formed after dense jet impaction with the bed that disperses horizontally. Hence, it is vital to mix and disperse the concentrated brine discharge rapidly in order to decrease adverse effects on marine environments [6]. The prediction of dense jet mixing is indispensable for designing outfalls and assessing the environmental effects. Numerous studies have been conducted to determine the dense jet characteristics. Point-based conductivity probes were used [28, 34] to implement an experimental study to investigate the brine dilution discharged into a flowing stream. An experimental study of a vertical dense jet dilution was conducted in a motionless ambient [11] to simulate the brine disposal along the Gulf of Mexico coast. Another study [21] was carried out in the Gulf of Mexico to measure the bottom dilution and the height of the vertical brine jet in order to operate multi-port outfall.

Despite the numerous experimental studies that have been conducted in this field, very few numerical studies have been carried out for dense inclined and horizontal jets. A numerical study [37] investigated inclined negatively buoyant jets using CFX-5 as a three-dimensional model. The shear stress transport (SST) model was used for turbulence closure, which is based on a combination of the $k - \epsilon$ and the $k - \omega$ models. The results of the study indicated the slight underestimation of the terminal rise height and considerable return point in comparison with the experimental data [30]. Oliver et al. [27] studied inclined negatively buoyant jets by considering the geometrical and mixing characteristics using the standard $k - \epsilon$ turbulence model. Turbulent buoyant wall jets were studied [13, 39–41] numerically, using a finite volume model. Different turbulence models were employed so that the LRR model was introduced as the best turbulence model among all the models examined.

The present study provides the results of numerical simulation of a horizontal three-dimensional offset dense jet using an open-source computational fluid dynamics (CFD) code named open field operation and manipulation (OpenFOAM). In this study, numerical equations were solved using a finite volume method. Turbulence characteristics of the jets along with Reynolds shear stresses were studied, as well as mesh sensitivity. A highly accurate Large Eddy Simulation (LES) turbulence model, i.e., the wall-adapting local eddy-viscosity (WALE) model, was tested and the numerical results were compared with experimental measurements provided by Shao and Law [33].

2 Methodology, Formulation, and the Numerical Model

In this study, CFD as a useful tool that can solve a locally averaged Navier–Stokes equation [2, 19] was employed to simulate a horizontal three-dimensional offset dense jet. A free open-source software package, OpenFOAM, was adopted for the CFD simulations [38]. OpenFOAM is a C++ toolbox that is available for different operating system platforms and has been used in some of the previous hydraulic engineering studies [14, 23, 32]. In this study, twoLiquidMixingFoam was employed, which is a solver for mixing two miscible incompressible fluids, and solves the multi-phase problem using the volume of fluid method.

2.1 Governing Equations for Turbulent Flows

The three-dimensional RANS equations that were used in this study include continuity and momentum for incompressible flows as follows [8, 9]:

$$\nabla \cdot U = 0 \quad (1)$$

$$\frac{\partial \rho U}{\partial t} + \nabla \cdot (\rho U U) = \nabla \cdot \tau + \rho g - \nabla p \quad (2)$$

where U is the flow velocity vector, ρ is the fluid density, t is time, τ is the shear-rate (viscous stresses) tensor, g is the gravitational acceleration vector, and p is the pressure.

2.2 LES Turbulence Model

To simulate the transition from laminar to turbulent phase in wall-bounded boundary layer flow, LES has been employed in previous studies (e.g., [15]). The filtered Navier–Stokes equations based on the LES calculation can be defined as follows [36]:

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho \bar{u}_i}{\partial x_i} = 0 \quad (3)$$

$$\frac{\partial \rho \bar{u}_i}{\partial t} + \frac{\partial \rho \bar{u}_i \bar{u}_j}{\partial x_j} = -\frac{\partial P}{\partial x_i} + \frac{\partial}{\partial x_j} \left[\mu \left(\frac{\partial \bar{u}_i}{\partial x_j} + \frac{\partial \bar{u}_j}{\partial x_i} \right) - \frac{2}{3} \delta_{ij} \frac{\partial \bar{u}_l}{\partial x_l} \right] + \frac{\partial}{\partial x_j} \left(-\rho \overline{u'_i u'_j} \right) \quad (4)$$

Reynolds stress tensor, i.e., $\overline{u'_i u'_j}$, needs to be defined in turbulence models. It can be resolved based on several approaches. One such approach is the algebraic turbulence model, which states that the Reynolds stress tensor is a function of mean velocity gradient, turbulent length scale, and turbulent time scale as shown below [5]:

$$\overline{u'_i u'_j} = F\left(\frac{\partial \overline{u}_i}{\partial x_j}, l_t, t_t\right) \quad (5)$$

Thus, $-\rho \overline{u'_i u'_j}$ maybe approximated as follows [36]:

$$-\rho \overline{u'_i u'_j} = \mu_t \left(\frac{\partial \overline{u}_i}{\partial x_j} + \frac{\partial \overline{u}_j}{\partial x_i} \right) - \frac{2}{3} \left(\rho k + \mu_t \frac{\partial \overline{u}_k}{\partial x_k} \right) \delta_{ij} \quad (6)$$

where μ_t indicates the turbulent eddy viscosity.

In this study, turbulent offset jet flows were simulated by using one of the most precise LES turbulence models, i.e., the WALE model. The methods are described below.

2.2.1 WALE

The WALE model [26], as an algebraic model has been applied in previous studies for complex flow problems [12, 20, 22, 24, 29]. WALE formulation is appropriate for turbulent as well as for transitional wall-bounded flow. The form of its turbulent eddy viscosity is shown below [36]:

$$\mu_t = \rho \Delta_s^2 \frac{\left(S_{ij}^d S_{ij}^d \right)^{3/2}}{\left(S_{ij} S_{ij} \right)^{5/2} + \left(S_{ij}^d S_{ij}^d \right)^{5/4}} \quad (7)$$

where the model parameters are [36]:

$$\Delta_s = Cw(\Delta x \Delta y \Delta z)^{1/3} \quad (8)$$

$$S_{ij}^d = \frac{1}{2} (\overline{g_{ij}^2} + \overline{g_{ji}^2}) - \frac{1}{3} \delta_{ij} \overline{g_{kk}^2} \quad (9)$$

$$\overline{g_{ij}} = \frac{\partial \overline{u}_i}{\partial x_j} \quad (10)$$

The purpose of designing the WALE model is to generate the true wall asymptotic variation of the sub-grid scale viscosity so that damping functions are not needed,

and to introduce the WALE model as a suitable model for the unsteady turbulent flows that have complex turbulence structure [36].

2.3 Numerical Model Setup and Mesh Sensitivity

A rectangular flume with dimensions of 1000 mm long, 850 mm wide, and 58.48 mm high was modeled to conduct the numerical study considering the mentioned initial conditions, as shown in Fig. 1.

The 2D inlet (nozzle) was aligned horizontally with an area equal to the area of the experimental jet with a diameter of 12.95 mm [33]. A sensitivity analysis was carried out to investigate the influence of the mesh size, and accordingly, a mesh of fixed rectangular cells using Cartesian coordinates was employed as the computational domain. The computational domain was constructed with square cells of size $\Delta x = \Delta y = \Delta z = 13$ mm via a Cartesian cell-centered discretization method. Seven grid sizes of 5, 10, 11, 13, and 20 mm were analyzed to determine the optimum mesh size (Fig. 2). Since no significant differences were seen between the wave-fronts under grid sizes higher than 13 mm, a grid size of 13 mm was adopted in both directions to reduce the computational effort. The numbers of cells in the x, y, and z directions were 77, 5, and 65, respectively. The mesh system encompassed 25,025 cells in total for all turbulence models. All surfaces of the channel were assumed to be hydraulically smooth. The lower boundary and side-walls were set as *wall*, given no inflow except for the inlet, which was set as *patch*. The density of the effluent was fixed as 1017.7 kg/m^3 , while the surrounding water density was set at 997.8 kg/m^3 .

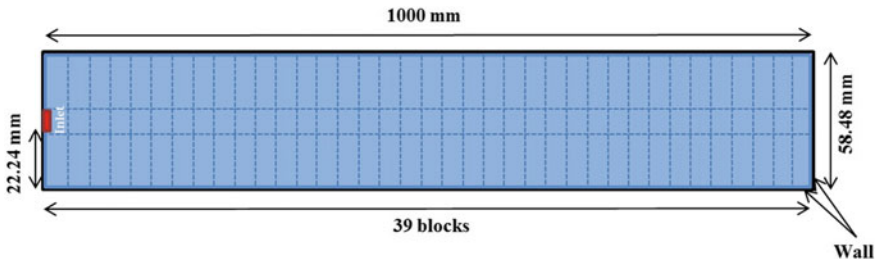


Fig. 1 2D scheme of numerical model setup and boundary conditions

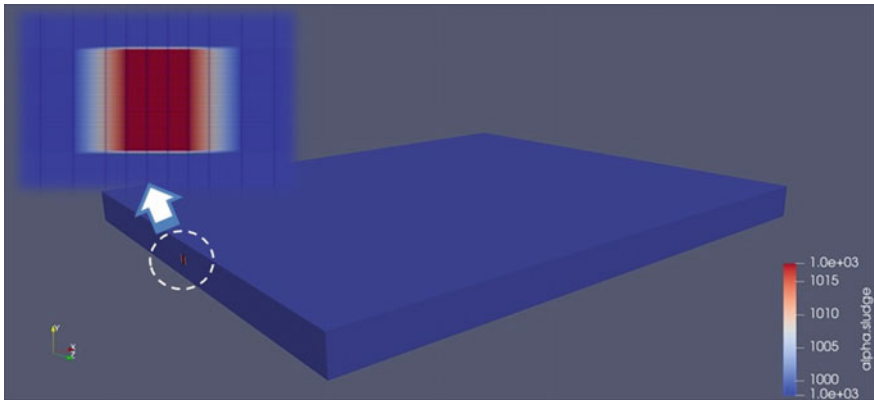


Fig. 2 Computational domain and the refined mesh system

3 Results and Discussion

3.1 Turbulence Characteristics

In the present study, the numerical evolution of the horizontal dense jet profile was extracted along a cross-section as seen in Fig. 3. In the regions closer to the nozzle tap, the concentration of the jet flow was larger due to larger momentum forces. The momentum forces decreased and buoyancy forces dominated the flow, as the jet went farther, and finally caused distortion, particularly for the lower half of the jet. In addition, there was a point after $x/D = 10$ where the concentration centerline started to shift below the jet velocity centerline. The concentration distribution fell into the general wall jet pattern close to the impact point. This indicates that the jet had a transition from plume to wall jet.

The contours of numerical Reynolds shear stress (i.e., $\overline{u'w'}$) are presented in Fig. 4. It can be seen that the contours with the higher value of $\overline{u'w'}$ were obtained from the WALE model mainly at $z/D = 3$. On the other hand, the lowest values were

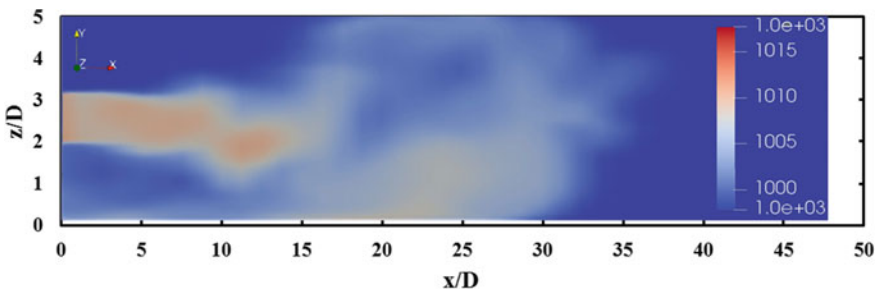


Fig. 3 Cross-sectional flow pattern evolution of horizontal dense jets

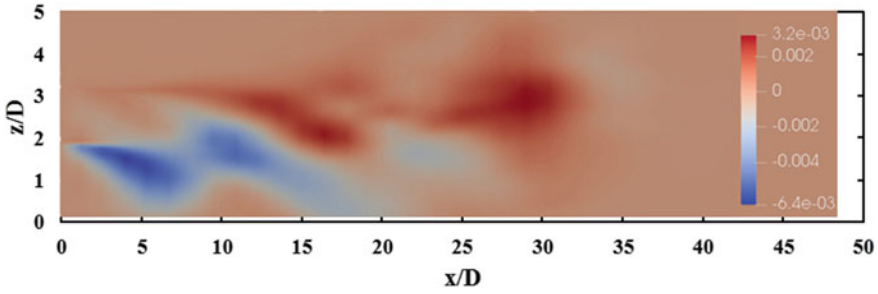


Fig. 4 Contours of Reynolds shear stress $\overline{u'w'}$

achieved at z/D lower than 2. In addition, the Reynolds shear stresses in the lower shear layer had slightly greater magnitude, regardless of their sign, than in the upper shear layer, which was consistent with the experimental results [33].

The variation of the numerical turbulence characteristics along the centerline in comparison with the experimental measurements was plotted in Fig. 5. The WALE turbulence model showed the best performance in the simulation of the maximum Reynolds shear stresses, especially before the impingement points where the fluctuations were weaker. The fluctuations of non-dimensionalized Reynolds shear stress (obtained via dividing $\overline{u'_m w'_m}$ by the square of the corresponding centerline maximum velocity, i.e. U_m) were also captured by the WALE model in good general agreement with the experimental results. It should be noted that the WALE model was unable to accurately exhibit the values of some peaks, especially at $x/D = 17.26$ and 17.62 .

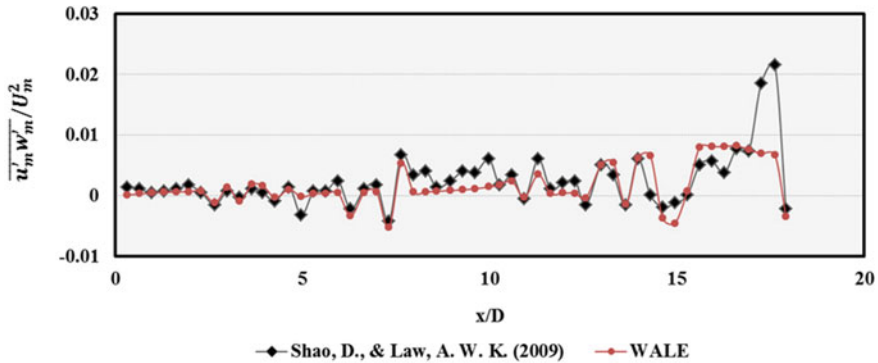


Fig. 5 Variation of turbulence characteristics $\overline{u'_m w'_m} / U_m^2$ versus x/D along the centerline

4 Conclusion

In this paper, one of the most accurate LES turbulence models, i.e., the WALE model, was employed to simulate the turbulence characteristics of horizontal three-dimensional offset dense jet flows. The horizontal dense jet flow profiles, as well as Reynolds shear stresses, were simulated in different areas either closer to or farther from the nozzle tap, and there were small differences between the numerical and experimental results.

Some concluding remarks for the current study are summarized below:

- The concentrations of the jet flows were larger in the regions closer to the effluent. The momentum forces decreased and buoyancy forces dominated the flow as the jet moved forward. It was numerically shown that the jet had a transition from plume to wall jet.
- The Reynolds shear stresses in the lower shear layer were slightly greater than in the upper shear layer, which was consistent with the experimental results.
- The WALE model showed the best performance in the simulation of the Reynolds shear stresses, especially before the impingement points, so that the numerical results were in good general agreement with the experimental measurements.

In order to reach a comprehensive understanding of the behavior of offset dense jet flows with higher levels of accuracy, it is suggested that further studies be conducted using various precise RANS, LES, and DES turbulence models along with different wall functions. More in-depth studies on assessing the physical features of complex flows are proposed for the future, given that their results can have wide applications in various environmental and engineering topics, particularly in energy dissipation and hydraulic structures.

References

1. Al Sajwani TbMA (1998) The desalination plants of Oman: past, present and future. *Desalination* 120:53–59
2. Anderson TB, Jackson R (1967) Fluid mechanical description of fluidized beds. *Equations of motion*. *Ind Eng Chem Fundam* 6:527–539
3. Assoudi A, Amamou A, Saïd NM, Bournot H (2020) Characteristics and analysis of a turbulent offset jet including the effect of density and offset height. *Int J Mech Sci* 174:105477
4. Einav R, Lokiec F (2003) Environmental aspects of a desalination plant in Ashkelon. *Desalination* 156:79–85
5. Gildeh HK (2013) Numerical modeling of thermal/saline discharges in coastal waters. University of Ottawa (Canada)
6. Gildeh HK, Mohammadian A, Nistor I, Qiblawey H (2015) Numerical modeling of 30° and 45° inclined dense turbulent jets in stationary ambient. *Environ Fluid Mech* 15:537–562
7. Hashim A, Hajjaj M (2005) Impact of desalination plants fluid effluents on the integrity of seawater, with the Arabian Gulf in perspective. *Desalination* 182:373–393
8. Holzmann T (2017) Mathematics, numerics, derivations and OpenFOAM (R), Holzmann CFD. Leoben, Austria. Retrieved from: www.holzmann-cfd.de

9. Imanian H, Mohammadian A (2019) Numerical simulation of flow over ogee crested spillways under high hydraulic head ratio. *Eng Appl Comput Fluid Mech* 13:983–1000
10. Intelligence GW (2004) Desalination markets 2005–2015. A global assessment and forecast. Media Analytics Ltd., Oxford, UK
11. James WP, Vergara I, Kim K (1983) Dilution of a dense vertical jet. *J Environ Eng* 109:1273–1283
12. Jee S, Joo J, Medic G (2016) Large-eddy simulation of a high-pressure turbine vane with inlet turbulence. Turbo expo: power for land, sea, and air. American Society of Mechanical Engineers, p V02DT44A019
13. Kheirkhah Gildeh H, Mohammadian A, Nistor I, Qiblawey H (2014) Numerical modeling of turbulent buoyant wall jets in stationary ambient water. *J Hydraul Eng* 140:04014012
14. Kheirkhah Gildeh H, Mohammadian A, Nistor I, Qiblawey H, Yan X (2016) CFD modeling and analysis of the behavior of 30 and 45 inclined dense jets—new numerical insights. *J Appl Water Eng Res* 4:112–127
15. Kim M, Lim J, Kim S, Jee S, Park D (2020) Assessment of the wall-adapting local eddy-viscosity model in transitional boundary layer. *Comput Methods Appl Mech Eng* 371:113287
16. Lattemann S, Höpner T (2008) Environmental impact and impact assessment of seawater desalination. *Desalination* 220:1–15
17. Launder B, Rodi W (1983) The turbulent wall jet measurements and modeling. *Annu Rev Fluid Mech* 15:429–459
18. Law AW-K, Herlina (2002) An experimental study on turbulent circular wall jets. *J Hydraul Eng* 128:161–174
19. Li X, Zhao J (2018) Dam-break of mixtures consisting of non-Newtonian liquids and granular particles. *Powder Technol* 338:493–505
20. Marty J (2014) Numerical investigations of separation-induced transition on high-lift low-pressure turbine using RANS and LES methods. *Proc Inst Mech Eng Part A: J Power Energy* 228:924–952
21. McLellan TN, Randall RE (1986) Measurement of brine jet height and dilution. *J Waterw Port Coast Ocean Eng* 112:200–216
22. Medic G, Zhang V, Wang G, Joo J, Sharma OP (2016) Prediction of transition and losses in compressor cascades using large-eddy simulation. *J Turbomach* 138
23. Mohsenabadi SE, Mohammadian M, Nistor I, Gildeh HK (2019) Un monde de barrages durables et sécuritaires (CFD modelling of near-field dam break flow, sustainable and safe dams around the world). In: Publications du symposium CIGB 2019, juin 9–14, 2019, Ottawa, Canada (Proceedings of the ICOLD 2019 symposium (ICOLD 2019), 9–14 June 2019, Ottawa, Canada). CRC Press, p 47
24. Morata EC, Gourdain N, Duchaine F, Gicquel L (2012) Effects of free-stream turbulence on high pressure turbine blade heat transfer predicted by structured and unstructured LES. *Int J Heat Mass Transf* 55:5754–5768
25. Nasr A, Lai J (1998) A turbulent plane offset jet with small offset ratio. *Exp Fluids* 24:47–57
26. Nicoud F, Ducros F (1999) Subgrid-scale stress modelling based on the square of the velocity gradient tensor. *Flow Turbul Combust* 62:183–200
27. Oliver C, Davidson M, Nokes R (2008) k- ϵ Predictions of the initial mixing of desalination discharges. *Environ Fluid Mech* 8:617–625
28. Pincince AB, List EJ (1973) Disposal of brine into an estuary. *J (Water Pollut Control Fed)* 2335–2344
29. Qin S, Koochesfahani M, Jaber F (2018) Large eddy simulations of unsteady flows over a stationary airfoil. *Comput Fluids* 161:155–170
30. Roberts PJ, Ferrier A, Daviero G (1997) Mixing in inclined dense jets. *J Hydraul Eng* 123:693–699
31. Roberts PJ, Toms G (1987) Inclined dense jets in flowing current. *J Hydraul Eng* 113:323–340
32. Shaheed R, Mohammadian A, Gildeh HK (2019) A comparison of standard k- ϵ and realizable k- ϵ turbulence models in curved and confluent channels. *Environ Fluid Mech* 19:543–568

33. Shao D, Law AW-K (2009) Turbulent mass and momentum transport of a circular offset dense jet. *J Turbul* N40
34. Shao D, Law AW-K (2010) Mixing and boundary interactions of 30 and 45 inclined dense jets. *Environ Fluid Mech* 10:521–553
35. Talavera JP, Ruiz JQ (2001) Identification of the mixing processes in brine discharges carried out in Barranco del Toro Beach, south of Gran Canaria (Canary Islands). *Desalination* 139:277–286
36. Tiwari P, Xia Z, Han X (2020) Comparison of VLES and LES turbulence modeling for swirling turbulent flow. *J Appl Fluid Mech* 13:1107–1116
37. Vafeiadou P, Papakonstantis I, Christodoulou G (2005) Numerical simulation of inclined negatively buoyant jets. In: *The 9th international conference on environmental science and technology*, September, pp 1–3
38. Weller HG, Tabor G, Jasak H, Fureby C (1998) A tensorial approach to computational continuum mechanics using object-oriented techniques. *Comput Phys* 12:620–631
39. Yan X, Ghodoosipour B, Mohammadian A (2020) Three-dimensional numerical study of multiple vertical buoyant jets in stationary ambient water. *J Hydraul Eng* 146:04020049
40. Yan X, Mohammadian A (2020) Evolutionary prediction of the trajectory of a rosette momentum jet group in flowing currents. *J Coastal Res* 36:1059–1067
41. Yan X, Mohammadian A, Chen X (2020) Numerical modeling of inclined plane jets in a linearly stratified environment. *Alex Eng J* 59:1857–1867
42. Zeitoun M, McIlhenny WF (1971) Conceptual designs of outfall systems for desalination plants. In: *Offshore technology conference*

Comparison of Three Minimization Theories for River Morphological Adjustments



Eman AlQasimi and Tew-Fik Mahdi

1 Introduction

Several minimization theories exist for the determination of river's bed and width adjustments. Three of these theories are compared in this paper using an experimental study by Cantelli et al. [4]. For the first theory, *minimization of total stream power* [7], if lower total stream power is the result of alteration of the channel widths, then channel adjustments are made in the lateral direction. Otherwise, the adjustments progress in the vertical direction. For the second theory, *minimization of energy slope* [6]. If the energy slope at a cross-section is greater than the weighted average energy slope of its adjacent sections, then the channel width at this section is reduced during deposition or the depth is increased during erosion. However, if the energy slope is smaller, the channel depth at this section is decreased during deposition or an increase of width occurs during erosion. Finally, for the third theory, *minimization of bed slope*. If the bed slope at a cross-section is greater than the weighted average bed slope of its adjacent sections, then the channel width at this section is reduced during deposition, or the depth is increased during erosion. Otherwise, the channel depth at this section is decreased during deposition or the width is increased during erosion.

The methodology to compare these three minimization theories is described in Sect. 2, while the results and discussion section are presented in Sect. 3, followed by the conclusion.

E. AlQasimi · T.-F. Mahdi (✉)

Department of Civil, Geological and Mining Engineering, Polytechnique Montreal, 2900 Boul. Édouard-Montpetit, University of Montréal Campus, 2500 Chemin de Polytechnique, Montreal, QC 3T 1J4, Canada

e-mail: tewfik.mahdi@polymtl.ca

© Canadian Society for Civil Engineering 2022

S. Walbridge et al. (eds.), *Proceedings of the Canadian Society of Civil Engineering*

Annual Conference 2021, Lecture Notes in Civil Engineering 250,

https://doi.org/10.1007/978-981-19-1065-4_35

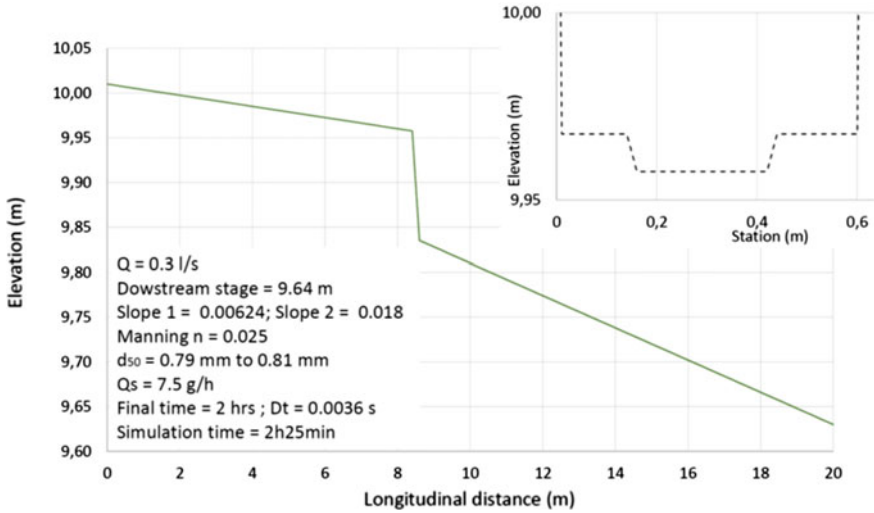


Fig. 1 Initial longitudinal profile of the experimental flume and typical cross-section

2 Methodology

Figure 1 summarizes the experimental case of Cantelli et al. [4]. The longitudinal profile has 2 different slopes and the material on the flume's bed is uniform (d_{50} between 0.79 and 0.81 mm). The upstream water and sediment discharges are 0.3 l/s and 7.5 g/h respectively.

The numerical modeling is done using the model UMHYSER-1D by AlQasimi and Mahdi [1–3] where the three minimization options are implemented. The model is run for each minimization option and the results are compared. UMHYSER-1D solves the Saint–Venant equations and a convection–diffusion equation for sediment mass conservation. For the present case, even if the bed slope smallness is locally violated, the Saint–Venant equations are used in this case as an extreme case to show the performances of the model.

3 Results and Discussion

Figure 2 shows simulated longitudinal profiles without minimization. The longitudinal profile evolution trend is well captured by UMHYSER-1D. However, at the slope break, the bed evolution is surely false. At the knick point, the Saint–Venant equations are not valid, but the numerical algorithm minimizes error growth which make UMHYSER-1D able to model a river reach with sudden thalweg changes. Overall, UMHYSER-1D is able to well reproduce the experimental results except at the slope break.

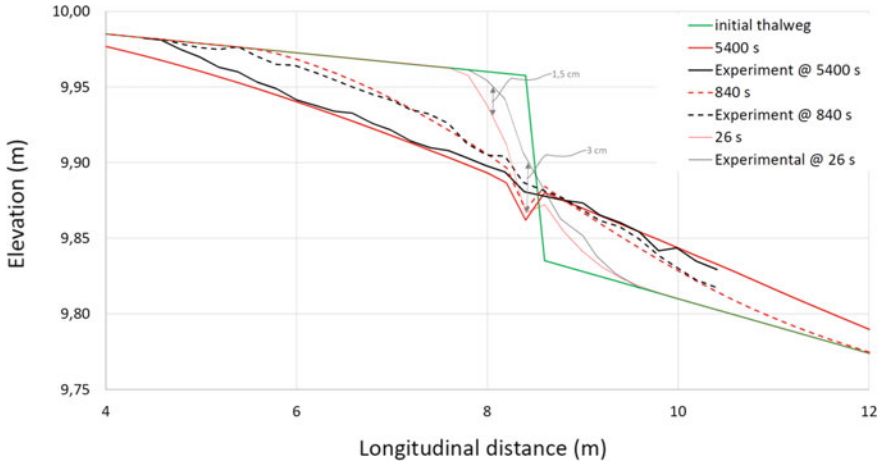


Fig. 2 Thalweg evolution without minimization options

Figure 3 shows simulated longitudinal profiles with three minimization options: minimization of bed slope, minimization of total stream power and minimization of energy slope. The best results are achieved by the minimum total stream power option while bed slope minimization has no effect and energy slope minimization doesn't significantly improve the results.

The minimum total stream power theory gives the best results because, mathematically speaking, this theory minimizes the following expression [5]:

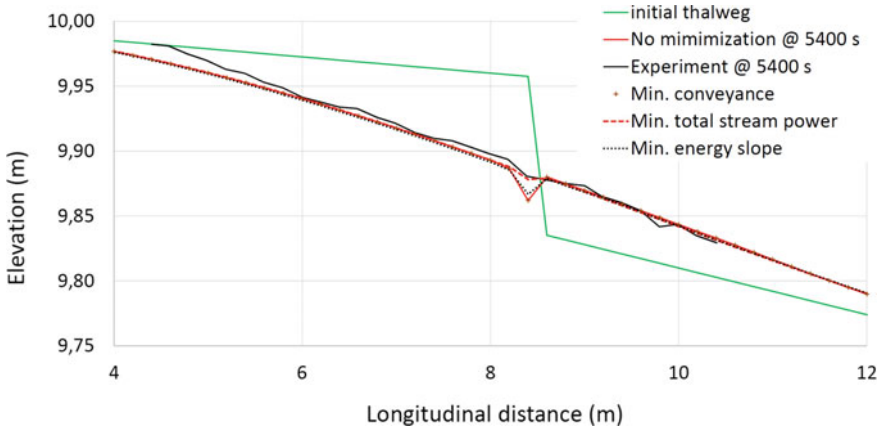


Fig. 3 Thalweg evolution using three minimization options

$$\phi_T = \sum_{i=1}^N \frac{1}{2} \gamma (Q_i S_i + Q_{i+1} S_{i+1}) \Delta x_i \quad (1)$$

where ϕ_T = total stream power, N = number of stations along the reach; γ = specific weight of water, Δx_i = distance between stations i and $i + 1$; and Q_i , S_i = discharge and slope at station i , respectively.

Equation (1) acts as an averaging process over the length of the channel (flume) involving the slopes at all the stations. As a result, the slopes at all the stations are milled providing the smooth simulated slope at the knick point.

4 Conclusions

To improve numerical modeling of rivers' longitudinal profile evolution, three minimization options, minimum of bed slope, minimum total stream power and minimum of energy slope, were compared in this study using an experimental study.

The bed slope minimization has no effect on thalweg evolution and energy slope minimization doesn't significantly improve the results. The minimum total stream power option improved the simulation results at the slope change only. It has to be noted that at the slope break, the total stream power minimization smooths the numerical solution even if, at this location, the Saint–Venant equations are not valid since the slope is not small anymore and the hydrostatic pressure distribution is violated. More investigations are needed to demonstrate the role of minimization theories to predict river morphology under unsteady flow conditions.

Acknowledgements This research was supported in part by a National Science and Engineering Research Council (NSERC) Discovery Grant, application No: RGPIN-2016-06413.

References

1. AlQasimi E, Mahdi T-F (2018) Unsteady model for the hydraulics of sediments in rivers. In: 26th annual conference of the computational fluid dynamics society of Canada, Winnipeg, MB, Canada, 10–12 June 2018
2. AlQasimi E, Mahdi T-F (2019) Flooding of the Saguenay region in 1996. Part 1: Modeling river Ha! Ha! flooding. *Nat Hazards* 96(1):1–15
3. AlQasimi E, Mahdi T-F (2020) A new one-dimensional numerical model for unsteady hydraulics of sediments in rivers. Special issue “Natural hazards and hydrological risks: climate change-water-sustainable society nexus”. *J Springer Nat Appl Sci* 2(9):1480
4. Cantelli et al (2004) Experiments on upstream-migrating erosional narrowing and widening of an incisional channel caused by dam removal. *Water Resour Res* 40
5. Chang HH (1982) Mathematical model for erodible channels. *J Hydraul Div ASCE* 108(HY5)

6. Chang HC (1988) *Fluvial processes in river engineering*. Wiley, New York, NY, USA
7. Yang CT (1972) Unit stream power and sediment transport. *ASCE J Hydraul Div* 98(HY 10):1805–1826

Investigating Event Selection for GA-Based SWMM Rainfall-Runoff Model Calibration



E. Snieder and U. T. Khan

1 Introduction

Physics-based models are commonly used to simulate dynamic hydrological processes in order to forecast future conditions. SWMM is a popular semi-distributed hydrological discrete-time simulation engine developed and freely distributed by the United States Environmental Protection Agency (US-EPA). SWMM is capable of simulating processes including surface runoff, groundwater, flow routing, infiltration, snowmelt, surface ponding, and water quality routing; many of these can be simulated using a variety of methods. As is typical for physics-based rainfall-runoff models, SWMM models are parameterised based mainly on measurable watershed characteristics. Many model parameters are highly uncertain, which is attributable to factors such as measurement error, non-stationary parameters, and unknown degrees of spatio-temporal heterogeneity [17]. In fact, the uncertainty of model parameters is among the most important uncertainty sources to consider for hydrological modelling [17, 24].

Typically, SWMM parameters are calibrated within their range of uncertainty in order to make predictions that more closely reflect observable values, which is quantified in the form of one or several performance criteria [14]. Automated calibration techniques have been demonstrated throughout the past two decades, nevertheless, many aspects, or the entire model calibration is still performed manually in an ad-hoc fashion [19].

Several studies have demonstrated the application of methods such as generalised likelihood uncertainty estimate (GLUE) and genetic algorithms (GA) for calibrating model parameters. However, model performance and generalisation remain heavily

E. Snieder (✉) · U. T. Khan

Department of Civil Engineering, York University, 4700 Keele Street, Toronto, ON M3J 1P3, Canada

e-mail: esnieder@yorku.ca

© Canadian Society for Civil Engineering 2022

S. Walbridge et al. (eds.), *Proceedings of the Canadian Society of Civil Engineering*

Annual Conference 2021, Lecture Notes in Civil Engineering 250,

https://doi.org/10.1007/978-981-19-1065-4_36

dependent on the selection of training data, which is typically done manually. A model calibrated to a particular event(s) is not guaranteed to maintain its accuracy on other events [3]. Therefore, selection of calibration and validation events is an important consideration, regardless of the calibration method. Multi-model approaches have emerged as a popular way to account for the dynamic variability of the rainfall-runoff relationship [3]. Such approaches can simultaneously consider multiple possible calibrated parameter sets. Considering the statements above, the objectives of this paper are to (1) investigate data-driven methods for identifying calibration events for SWMM calibration and (2) comparing different multi-model combination strategies.

2 Study Area

Fourteen Mile Creek (FMC) is in and upstream of the town of Oakville ON, Canada. FMC has a total area of 31 km², which is discretised into 61 subcatchments. The watershed is characterised as semi-urban, containing predominantly agricultural land-use upstream, while downstream consists of commercial and medium-density residential. The watercourses in the FMC model representation are highly detailed, as reflected by hundreds of irregular cross-sections used to represent the main creek. The FMC has historically experienced flooding and has been subject of several flood mitigation projects over the past decade [27]. The FMC SWMM model was obtained in an uncalibrated state, evidenced by the low precision and variability of parameter values. Cold weather conditions were removed from the dataset based on historic data obtained from Environment Canada [9], as cold weather processes (e.g., snow accumulation) are not parameterised in the SWMM model. To accommodate varying initial conditions for each event-based simulation, Hot Start Files (HSF) are produced for each hydrological event using a period of one week prior to the event start (Fig. 1).

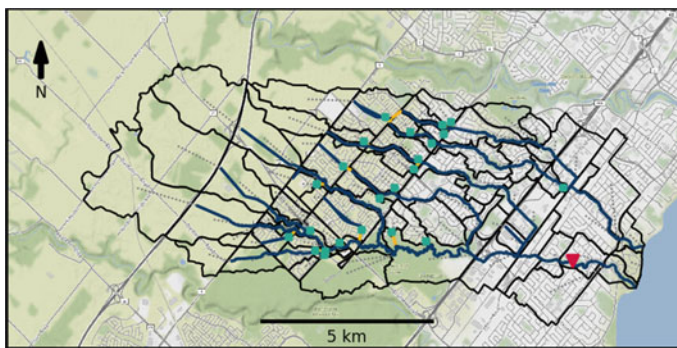


Fig. 1 FMC SWMM model

3 Methods

The following section provides a brief review of state-of-the-art methods for event selection, followed by a description of the methods used for model calibration, performance evaluation, and multi-model combination.

3.1 *Event Discretisation and Selection*

The selection of rainfall event(s) is an important consideration, as it will influence how the model performs on future events, or how well the model generalises. It is widely known that calibrated models will perform best on data that resembles the calibration data. Thus, event selection strategies tend to reflect the model application; for flood forecasting, a popular application, high flow rainfall-runoff events are selected for calibration. Several studies have used ad-hoc approaches for identifying rainfall events. Qin et al. [20] create synthetic events in 3 categories, with varying characteristics such as return period, duration, rainfall intensity, rainfall amount, and time-to-peak. In this study, grouping rainfall events is motivated by the unique rainfall-runoff relationship across different rainfall characteristics, which is confirmed by simulation results. Blackler and Guo [4] distinguish early spring, spring, summer, and intense summer rainfall events; this study noted that long, low intensity events achieved stronger calibration performance compared to short, intense events. Broekhuizen et al. [7] evaluate a multi-event calibration scheme in which events are selected from a set of discretised events based on one of rainfall, rainfall intensity, mean flow, total flow relative to total rainfall, or rainfall duration. In their calibration scheme, events that fall above and below a runoff threshold are used in a two-stage calibration technique, which is found to outperform a single stage multi-event approach.

Data-driven hydrological models, often considered to be counterparts to physics-based models such as SWMM, often employ sophisticated data splitting methods for identifying data for model calibration and validation. It is well-known that data used for calibration and validation should have similar statistical properties [5, 29]. Unsupervised clustering provides an objective way to identify groups of events, each with distinct characteristics. Clustering techniques have previously been used to classify individual observations, for the dual purpose of ensuring that each all flow regimes are equally represented (i.e., an equal number of low and high flows), and to ensure similar statistical properties between the calibration and validation data [1, 5, 26]. Toth [26] used an SOM to classify streamflow and rainfall data (with several lagged copies of each), evaluating the effect that varying the number of clusters has on overall model performance. The study also links the SOM-based classifications to physical hydrological phenomena (e.g., a given class may correspond to the rising limb of the hydrograph). Bowden et al. [5] compared GA-based and SOM-based data splitting methods, finding that they both outperformed traditional approach in which data is split arbitrarily without any consideration for the statistical properties

of each split. Similarly, Anctil and Lauzon [1] use an SOM to create data subsets with statistically similar properties.

However, because physics-based models require time-continuous data [11], methods that classify individual samples such as those mentioned above are unsuitable; instead, a clustering approach must be adapted to classify complete rainfall-runoff timeseries', rather than individual observations as has been done in previous works. Steffen and Gomes [25] demonstrate a framework for classifying flood severity using based on a set of indicators in order to better understand flood formation mechanics. In our study, a comparable timeseries clustering approach is employed, however, in our case for calibration and validation data selection.

As mentioned above, clustering is conducted using features derived from rainfall-runoff events that have first been discretised from continuous data, which are summarised in Table 2. For clustering, both rainfall and streamflow event statistics are used in order to achieve a subset of events with both diverse streamflow and forcing rainfall event characteristics [26]. The event statistics (called feature set) proposed in this study include total precipitation (P), peak intensity (i_p), mean intensity (i_μ), duration (D), date (date), peak flow (q_p), flow volume (Q), hydrograph centroid coordinates (q_{ct}) and (q_{cy}), and hyetograph centroid coordinates (q_{ct}) and (q_{cy}).

The feature set has a high dimensionality (11) relative to the number of events, or samples (25). Thus, before clustering, principal component analysis (PCA) is used for dimensionality reduction. This combination of dimensionality reduction and clustering has been widely used in literature to reduce the effects of the curse of dimensionality [10, 16, 25]. K-means clustering is used to cluster the 5 first principal components (PCs).

3.2 Genetic Algorithm-Based Optimisation

A variety of optimisation algorithms have been applied for SWMM calibration [19]. A recent comparison of three evolutionary optimisation algorithms found that the GA outperforms the other algorithms [15]. The GA is an evolutionary technique that searches for a global minimum by iteratively breeding new populations of parameter estimates. Each new population is created by crossing-over and mutating the best population of each previous generation. The initial population is generated by randomly sampling within the uncertainty range of each individual parameter. The uncertain parameters subject to calibration, and their degrees of uncertainty, are identified based on expert knowledge and are listed in Table 1. The uncertainty value indicates the uncertainty of each parameter relative to its initial value. The lower and upper constraints are identified based on physically realistic values for each parameter (irrespective of site conditions). For example, the upper constraint for saturated hydraulic conductivity (KSat) is 120 mm/h, which is largest value listed in the SWMM manual [13].

Table 1 GA calibration parameters and uncertainty ranges

| Class | Attribute | Units | Uncertainty | Lower constraint | Upper constraint |
|---------------|---------------|-------|-------------|------------------|------------------|
| Subcatchments | Width | m | ± 4.0 | 0 | 1,000,000 |
| Subcatchments | PercentSlope | % | ± 0.5 | 0 | 5 |
| Subcatchments | PercentImperv | % | ± 0.4 | 0 | 100 |
| Subareas | N_Imperv | – | ± 0.2 | 0 | 0.024 |
| Subareas | N_Perv | – | ± 1.0 | 0 | 0.8 |
| Subareas | S_Imperv | mm | ± 0.4 | 0 | 2.54 |
| Subareas | S_Perv | mm | ± 1.0 | 0 | 7.62 |
| Infiltration | Suction | mm | ± 1.0 | 0 | 320 |
| Infiltration | Ksat | mm/hr | ± 1.0 | 0 | 120 |
| Infiltration | IMD | – | ± 0.5 | 0 | 0.4 |

Table 2 Discretised event dates, statistics, and classification result

| Event start (dd-MM-yy HHmm) | Duration (h) | Precip. (mm) | Peak intensity (mm/h) | Mean flow rate (m ³ /s) | Peak flow rate (m ³ /s) | Class |
|-----------------------------------|--------------|--------------|-----------------------------|---------------------------------------|---------------------------------------|-------|
| 12-05-19 1505 | 39.5 | 13 | 2.8 | 1.2 | 1.8 | 2 |
| 25-05-19 0615 | 12.3 | 26.5 | 22.5 | 2.1 | 5.3 | 2 |
| 05-06-19 1130 | 9.8 | 23.6 | 21.7 | 1.9 | 3.3 | 2 |
| 10-06-19 0915 | 9.4 | 8.3 | 8.3 | 0.4 | 2.1 | 1 |
| 13-06-19 0400 | 12.6 | 10.4 | 14.9 | 0.4 | 1.8 | 1 |
| 20-06-19 0800 | 15.1 | 9.8 | 5.8 | 0.4 | 1.9 | 1 |
| 24-06-19 2220 | 3.3 | 11.5 | 10.8 | 0.6 | 2.5 | 1 |
| 29-07-19 2110 | 12.1 | 10.8 | 6.9 | 0.2 | 1.2 | 1 |
| 26-10-19 1945 | 13.7 | 33.7 | 11.9 | 1.8 | 5.6 | 2 |
| 30-10-19 1235 | 33.1 | 31.7 | 5.5 | 1.4 | 3.2 | 2 |
| 27-11-19 0715 | 5.3 | 6.1 | 6.5 | 0.3 | 0.6 | 1 |
| 29-12-19 1200 | 21.3 | 16.4 | 5.8 | 2.4 | 5.4 | 2 |

3.3 Performance Criteria

The Kling-Gupta Efficiency (KGE) is used as the objective function for the GA [12].

$$\text{KGE} = 1 - \sqrt{\left(\frac{\text{cov}(q, \hat{q})}{\sigma_q \sigma_{\hat{q}}} - 1\right)^2 + \left(\frac{\sigma_q}{\sigma_{\hat{q}}} - 1\right)^2 + \left(\frac{\mu_{\hat{q}}}{\mu_q} - 1\right)^2} \quad (1)$$

where q and \hat{q} are observed and predicted stage values, μ is the mean, $\text{cov}()$ is the covariance, and σ is the standard deviation.

$$x_i \leftarrow \text{argmax}(g(Q_i, f(X, P_i))) \quad (2)$$

where x_i is the parameter set obtained by the maximisation of the objective function (KGE), $g()$. The objective function quantifies the similarity between the observed flow Q for a given rainfall-runoff event i , and the predicted flow, $f()$. The predicted flow is calculated in SWMM and is a function of all possible model parameter sets X (bound by their uncertainty range) and the precipitation for the same event, P_i .

Two additional performance criteria are used for model evaluation: the mean volume error (MVE) and peak flow error (PFE), which are calculated as follows:

$$\text{MVE} = \frac{\sum(q - \hat{q})}{\sum q} \quad (3)$$

$$\text{PFE} = \frac{\max(q) - \max(\hat{q})}{\max(q)} \quad (4)$$

3.4 Multi-model Prediction

A multi-model approach is used in this study, that treats collections of SWM models, each calibrated to distinct hydrological events, as an ensemble; discrete predictions are obtained by combining the output of each ensemble member. Combining predictions from different parameter sets may be seen as counter-intuitive, especially for time-invariant parameters which are expected to have one true value, significant work in hydrology has demonstrated that there is no single optimum parameter sets; instead, there may be numerous parameter sets that are equally suitable for a given model [2, 3, 18]. Multi-model predictions are supported by pure literature on model ensembles, which asserts that predictions produced by ensembles are known to outperform single models; the success of ensemble is attributable to the degree of disagreement between the ensemble members, also known as ensemble diversity

[8]. Combining the predictions from multiple rainfall-runoff events has been demonstrated improving model generalisation in literature [3, 22]. Most commonly, multi-model approaches combined different types of rainfall-runoff models [28]. Several studies have demonstrated the effectiveness of combining homogeneous models, calibrated to different observations. Seibert and Beven [21] use a Monte Carlo simulation to randomly select observations for model calibration and obtain predictions by taking the weighted mean of the 100 best performing models. The multi-model approach is found to outperform the single model [21].

Our proposed ensemble consists of models that are each calibrated to different rainfall-runoff events; ensemble diversity, which can be measured as the covariance between the predictions of ensemble members, is achieved implicitly by including diverse training data in the form of rainfall events with distinct characteristics. To create an ensemble, n models are calibrated by using a GA so find a solution for Eq. (2). Predictions are computed for each individually calibrated model and combined, which is generally expressed by:

$$\widehat{Q}_i = h(f(x_1, P_i), f(x_2, P_i), \dots, f(x_n, P_i)) \quad (5)$$

where Q_i is the runoff corresponding to event i , $h()$ is the model combiner (e.g., equal model weighting, ANN, etc.), and $f(x_n, P_i)$ is a SWMM model with input parameters x_n and forcing precipitation for event i , P_i . There are a wide variety of model combination techniques; the following subsections outline two common methods: uniform weighting and stacking.

3.4.1 Weighted Combination

Weighted combination is among the most popular techniques for aggregating ensemble predictions [23]. Uniform weighting is the most common weighting scheme, but various techniques exist for determining non-uniform model weights. We evaluate both uniformly weighted and non-uniformly weighted schemes; for the latter, we utilise the predictions of each calibrated model on the concatenated calibration timeseries. The weights for the SWMM models are determined based on the following equation:

$$w \leftarrow \operatorname{argmax} \left(g \left(Q_i, \widehat{Q}_i \right) \right) \quad (6)$$

where w is the optimum weight vector and \widehat{Q}_i is the weighted combination of SWMM predictions:

$$\widehat{Q}_i = w_1(f(x_1, P_i)), w_2(f(x_2, P_i)), \dots, w_n(f(x_n, P_i)) \quad (7)$$

The weights are determined using the Nelder-Mead simplex direct search algorithm.

3.4.2 Stacking

Stacking is a generalisation technique in which a model is trained to combine other models (sometimes referred to as a meta-learner). Linear regression is the most common stacking model type [6]. However, more complex model types such as artificial neural networks are capable of mapping non-linear relationships between input data (SWMM predictions) and target data (observed flows), providing greater compensation for potential deficiencies of the SWMM models [23]. We include both a linear model and ANN in our comparison of combination methods. The configuration of the linear model is trivial; more consideration is necessary for the ANN parameterisation. State-of-the-art modelling practices are used to ensure the ANN is well-generalised. Specifically, stop-training and Bagging are employed; the Levenburg-Marquardt backpropagation is used for training. Grid-search optimisation is performed for the number of Bagged models and hidden layer architecture, which identified optimum sizes of 128 models and 8 hidden neurons.

A benefit of stacking is the ability to incorporate exogenous input variables such as precipitation forecasts and observed flow. In this research, we utilise aerial weighted rainfall for the FMC watershed and observed flow (lagged by 36 h).

3.5 Model Validation

In order to assess the ability for the model(s) to generalise, we assess model performance using events that are isolated from those used for predictions. For SM predictions, the validation performance for each event is calculated using the remaining 11 events. For the MM predictions, each possible combination is considered; for example, for 6 class 1 events (numbered 1–6) and 6 class 2 events (7–12), validating event #1 is performed using all 6 unique combinations of 5 events from each class ([2–11], [2–10,12], [2–9,11–12], [2–8,10–12], [2–7,9–12], and [2–6, 8–12]). High variability between each combination of selected events is indicative of high sensitivity to the event selection procedure.

4 Results and Discussion

The following sections outline the outcome of the unsupervised event clustering and presents a comparison of the single and multi-model methods outlined above.

4.1 Unsupervised Event Clustering

The outcome of the unsupervised event clustering is illustrated in Fig. 2, which shows the relationships between select indicator variables using scatterplots, and a summary is included in Table 2.

The final row and columns show the class values. The resulting classes can be described as large and small events, in terms of flow. There is clear stratification of classes for the total flow (Q), while the stratification is less defined for precipitation (P) and intensity (i_{μ}). Large events tend to occur in spring and fall, while small events occur throughout the year. Six events from each class are selected for model calibration and validation; models are calibrated to individual models, which are used together according to the methodologies described above.

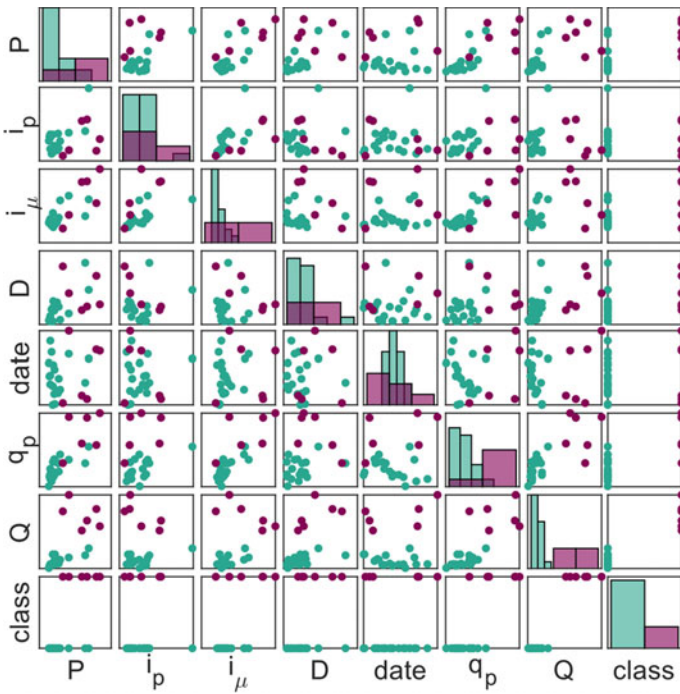


Fig. 2 Scatterplot matrix for clustering indicator variables; classes are distinguished by colour

4.2 Comparison of Single Model and Multi-model Predictions

As mentioned above, this research compares single model with several multi-model prediction techniques; the latter include uniformly weighted, weighted, multi-variable linear regression-based stacking, and artificial neural network-based stacking. The comparison is illustrated in Fig. 3 and summarised in Table 3.

The range of the boxes indicates the range in performance due to the selection of 5 events from each of classes 1 and 2. In Table 3, the mean performance (ranked) and variance are calculated for each event, then averaged across all events. The highest-ranking performance and lowest variance are indicated in bold. The uniformly weighted model predictions have the strongest KGE and MVE performance, whereas the STACK_ANNX model has the strongest PFE.

The variance in Table 3 corresponds to the variance associated with the specific event(s) used for the single or multi-model prediction; for example, for validating event 1, the single model performance is calculated for events 2–12; the variance

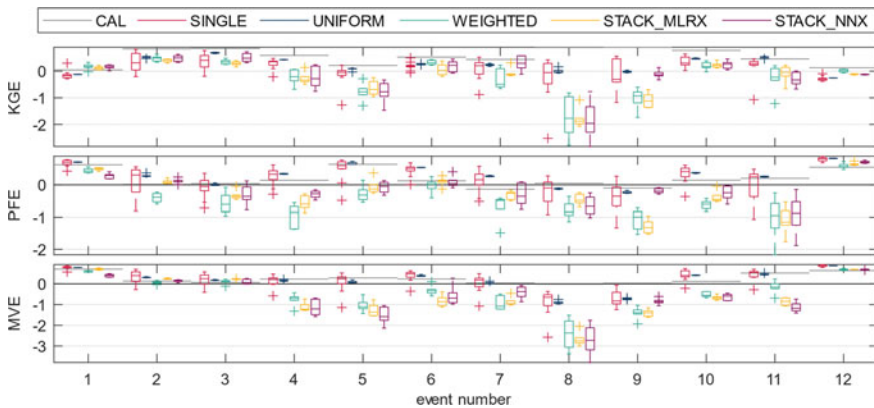


Fig. 3 Validation performance of SWMM configurations for single model (red), uniform weighting (blue), variable weighting (green), MLR-based stacking (yellow), and ANN-based stacking (purple). Calibration performance is indicated by a grey bar

Table 3 Ranked performance and variance for KGE, PFE, and MVE

| | KGE rank | KGE var | PFE rank | PFE var | MVE rank | MVE var |
|------------|--------------|--------------|--------------|--------------|--------------|--------------|
| CAL | 1.083 | – | 2.167 | – | 2.250 | – |
| SINGLE | 4.000 | 0.156 | 3.167 | 0.119 | 3.250 | 0.119 |
| UNIFORM | 2.583 | 0.001 | 4.000 | 0.001 | 3.250 | 0.001 |
| WEIGHTED | 4.167 | 0.116 | 4.667 | 0.101 | 3.250 | 0.091 |
| STACK_MLRX | 4.917 | 0.044 | 3.917 | 0.038 | 4.750 | 0.037 |
| STACK_ANNX | 4.250 | 0.094 | 3.083 | 0.072 | 4.250 | 0.105 |

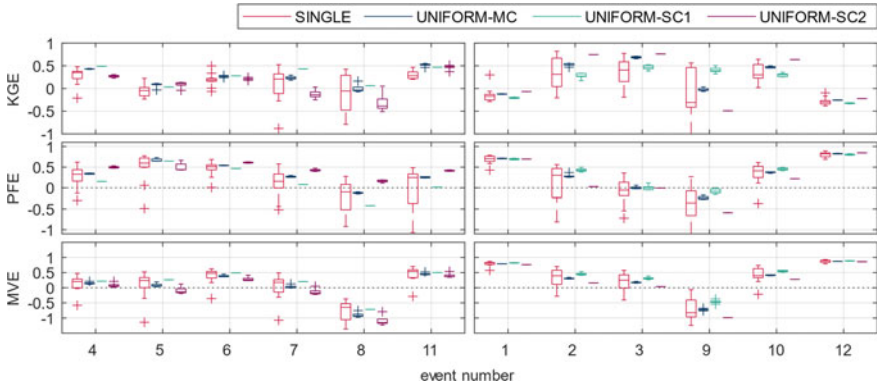


Fig. 4 Validation performance of SWMM configurations for single model (red), uniformly weighted multi-class (blue), class 1 (green), and class 2 (purple)

is calculated for the distribution of these 11 performance values. The uniformly weighted model has the lowest variance compared to all other methods, while the single model predictions typically have the highest, regardless of the performance metric. This means that model validation performance is highly sensitive to calibration event selection; combining predictions eliminates this variance.

Next, we compare the effects of three subsets of events for multi-model prediction: 5 events from each class (as presented above), using 5 class 1 events, and 5 class 2 events. Validation events belonging to class 1 and 2 are separated into left and right subplots, respectively. Generally, predictions made using models calibrated to a given class perform best on events of the same class; this is observed across different classes and performance criteria (Fig. 4).

These trends in performance can be used to guide event selection for calibration procedures. Balanced performance is achieved using a combination of diverse event classifications; if improved performance is desired on events of a given class, predictions can be made using events calibrated based on events in the same class.

5 Conclusion

This research outlines a novel approach, based on unsupervised clustering, for objectively selecting rainfall-runoff events for SWMM calibration. Once clustered, an event number of events from each class are selected for calibration, ensuring statistically diverse training data. Next, several multi-model prediction strategies, including two weighting schemes and two stacking methods. Overall, the uniformly weighted model combinations have the best performance and lowest sensitivity to specific event selection, across three criteria. Uniformly weighted predictions are compared for predictions made using only events from a single event class and predictions

made using an even number of predictions from each class. Predictions made using a single class tend to perform best on validation events with the same class, and worst on events from the other class; the multi-class performance typically lies between the two.

Recommendations for future work include a more robust comparison of multi-event calibration methods (e.g., using multi-objective optimisation) within the proposed framework for event selection. Furthermore, a detailed analysis of calibrated parameter values for models calibrated to distinct events will provide a greater understanding and physics-based support for parameter estimates.

Acknowledgements The authors would like to thank the National Sciences and Engineering Research Council of Canada (NSERC) and York University for providing funding for this research. Additionally, a special thank you to Civica Infrastructure for providing us with the FMC model and data files, and for their recommendations on model calibration parameter identification and uncertainty estimation.

References

1. Anctil F, Lauzon N (2004) Generalisation for neural networks through data sampling and training procedures, with applications to streamflow predictions. *Hydrol Earth Syst Sci* 8(5):940–958. <https://doi.org/10.5194/hess-8-940-2004>
2. Awol FS, Coulibaly P, Tolson BA (2018) Event-based model calibration approaches for selecting representative distributed parameters in semi-urban watersheds. *Adv Water Resour* 118:12–27. <https://doi.org/10.1016/j.advwatres.2018.05.013>
3. Beven K (2012) *Rainfall-runoff modelling: the primer*, 2nd edn. <https://doi.org/10.1002/9781119951001>
4. Blackler GE, Guo JCY (2014) Field test of paved area reduction factors using a storm water management model and water quality test site. *J Irrig Drain Eng* 140(4). [https://doi.org/10.1061/\(ASCE\)IR.1943-4774.0000680](https://doi.org/10.1061/(ASCE)IR.1943-4774.0000680)
5. Bowden GJ, Maier HR, Dandy GC (2002) Optimal division of data for neural network models in water resources applications. *Water Resour Res* 38(2):2-1–2-11. <https://doi.org/10.1029/2001WR000266>
6. Breiman L (1996) Stacked regressions. *Mach Learn* 24(1):49–64. <https://doi.org/10.1007/BF00117832>
7. Broekhuizen I, Leonhardt G, Marsalek J, Viklander M (2020) Event selection and two-stage approach for calibrating models of green urban drainage systems. *Hydrol Earth Syst Sci* 24(2):869–885. <https://doi.org/10.5194/hess-24-869-2020>
8. Brown G, Wyatt JL, Tiño P (2005) Managing diversity in regression ensembles. *J Mach Learn Res*
9. Environment Canada (2021) *Water level and flow—Environment Canada*. <https://wateroffice.ec.gc.ca/>
10. Frank M, Drikakis D, Charissis V (2020) Machine-learning methods for computational science and engineering. *Computation* 8(1):15. <https://doi.org/10.3390/computation8010015>
11. Guo D, Zheng F, Gupta H, Maier HR (2020) On the robustness of conceptual rainfall-runoff models to calibration and evaluation dataset splits selection: a large sample investigation. *Water Resour Res*. <https://doi.org/10.1029/2019WR026752>
12. Gupta HV, Kling H, Yilmaz KK, Martinez GF (2009) Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *J Hydrol* 377(1–2):80–91. <https://doi.org/10.1016/j.jhydrol.2009.08.003>

13. James W, Rossman LE, Robert W, James C (2010) User's guide to SWMM5, 13th edn. Computational Hydraulics International, Guelph. www.chiwater.com
14. Krebs G, Kokkonen T, Valtanen M, Koivusalo H, Setälä H (2013) A high resolution application of a stormwater management model (SWMM) using genetic parameter optimization. *Urban Water J* 10(6):394–410. <https://doi.org/10.1080/1573062X.2012.739631>
15. Kumar S, Kaushal DR, Gosain AK (2019) Evaluation of evolutionary algorithms for the optimization of storm water drainage network for an urbanized area. *Acta Geophys* 67(1):149–165. <https://doi.org/10.1007/s11600-018-00240-8>
16. Li S, Ma K, Jin Z, Zhu Y (2016) A new flood forecasting model based on SVM and boosting learning algorithms. In: 2016 IEEE congress on evolutionary computation (CEC). IEEE, pp 1343–1348. <https://doi.org/10.1109/CEC.2016.7743944>
17. Li Z, Shao Q, Xu Z, Cai X (2010) Analysis of parameter uncertainty in semi-distributed hydrological models using bootstrap method: a case study of SWAT model applied to Yingluoxia watershed in Northwest China. *J Hydrol* 385(1–4):76–83. <https://doi.org/10.1016/j.jhydrol.2010.01.025>
18. Mediero L, Garrote L, Martín-Carrasco FJ (2011) Probabilistic calibration of a distributed hydrological model for flood forecasting. *Hydrol Sci J* 56(7):1129–1149. <https://doi.org/10.1080/02626667.2011.610322>
19. Niazi M, Nietch C, Maghrebi M, Jackson N, Bennett BR, Tryby M, Massoudieh A (2017) Storm water management model: performance review and gap analysis. *J Sustain Water Built Environ*. <https://doi.org/10.1061/JSWBAY.0000817>
20. Qin H-P, Li Z-x, Fu G (2013) The effects of low impact development on urban flooding under different rainfall characteristics. *J Environ Manag* 129:577–585. <https://doi.org/10.1016/j.jenvman.2013.08.026>
21. Seibert J, Beven KJ (2009) Gauging the ungauged basin: how many discharge measurements are needed? *Hydrol Earth Syst Sci* 13(6):883–892. <https://doi.org/10.5194/hess-13-883-2009>
22. Shamseldin AY, O'Connor KM, Liang GC (1997) Methods for combining the outputs of different rainfall-runoff models. *J Hydrol* 197(1–4):203–229. [https://doi.org/10.1016/S0022-1694\(96\)03259-3](https://doi.org/10.1016/S0022-1694(96)03259-3)
23. Sharkey AJC (1996) On combining artificial neural nets. *Connect Sci* 8(3–4):299–314. <https://doi.org/10.1080/095400996116785>
24. Shin M-J, Choi Y (2018) Sensitivity analysis to investigate the reliability of the grid-based rainfall-runoff model. *Water* 10(12):1839. <https://doi.org/10.3390/w10121839>
25. Steffen PC, Gomes J (2018) Clustering of historical floods observed on Iguaçú River, in União Da Vitória, Paraná. *RBRH* 23(23). <https://doi.org/10.1590/2318-0331.231820170107>
26. Toth E (2009) Combined use of SOM-classification and feed-forward networks for multinet-work streamflow forecasting. *Assembly* 11:11962–11962. www.hydrol-earth-syst-sci.net/13/1555/2009/
27. Town of Oakville, AMEC (2014) Public Information Centre # 2 Flood Mitigation Opportunities Study Fourteen Mile Creek & McCraney Creek Systems Municipal Class Environmental Assessment. Town of Oakville. <https://www.oakville.ca/assets/general-environment/14McPIC2boards.pdf>
28. Zahmatkesh Z, Karamouz M, Nazif S (2015) Uncertainty based modeling of rainfall-runoff: combined differential evolution adaptive metropolis (DREAM) and K-means clustering. *Adv Water Resour* 83:405–420. <https://doi.org/10.1016/j.advwatres.2015.06.012>
29. Zheng F, Maier HR, Wu W, Dandy GC, Gupta HV, Zhang T (2018) On lack of robustness in hydrological model development due to absence of guidelines for selecting calibration and evaluation data: demonstration for data-driven models. *Water Resour Res* 54(2):1013–1030. <https://doi.org/10.1002/2017WR021470>

How Can a State-of-the-Art Prediction Technique for Random Parameter Models Improve the Construction Work Environment in Ontarian Highways?



Seyedata Nahidi and Susan Tighe

1 Introduction

Throughout the last couple of decades, road safety and collision analysis have been advanced due to developing innovative techniques. More conventional methods, where data analysis and statistical modelling are the foundation of it, are also forged ahead with the introduction of “Random Parameter Models” or “Mixed Models”. In 1961, Clark introduced the concept of these models. He suggested that these models could be beneficial in terms of future event predictions, while there is uncertainty about collected data and unobserved heterogeneity issue [3]. These models let the coefficients change across the observations following a given distribution [11]. Since 1980, these models, which were originally used for econometric analysis, were started to being popular in the transportation engineering field. Anastasopoulos and Mannering [2] published an article where they have used these models for the first time for analyzing the accident frequency assessment with count-data models. Later on, these models were used in numerous studies by numerous researchers in the transportation field to investigate the impact of unobserved factors in their studies [1, 5, 7]. Apart from the advantages that these models offer, there are two main concerns about them: (1) It is complicated to decide which one of the coefficients to choose for future predictions based on the characteristic of the events (whether the unobserved heterogeneity is expandable to all scenarios), and (2) They are data dependant, and generally they are not spatially or temporally transferable. This study focuses on these concerns and tries to develop an efficient technique to overcome these issues.

S. Nahidi (✉)
University of Waterloo, Waterloo, Canada
e-mail: snahidi@uwaterloo.ca

S. Tighe
McMaster University, Hamilton, Canada
e-mail: sltighe@uwaterloo.ca

2 Methodological Framework

In 2019, authors published a paper where they investigated the factors that significantly have an impact on injury-severity of the collisions in the work zones located in high-volume highways. They concluded that the random parameter ordered logit model has the highest performance among other developed models. This model is considered the foundation of this study [8]. Fixed parameter ordered logit model could be formulated as Eq. 1:

$$z = \beta X + \varepsilon \quad (1)$$

where z is the ordinal ranking estimator, β is a matrix of estimated fixed coefficients for independent variables, X is a vector matrix of independent variables' values, and ε is the random disturbance [11]. Comparing the value of the z with estimated thresholds can directly demonstrate the estimated category of the introduced event to the model. This process could be demonstrated in Eq. 2:

$$\begin{aligned} y &= 1 \quad \text{if } z \leq \mu_0 \\ y &= 2 \quad \text{if } \mu_0 < z \leq \mu_1 \\ &\dots \\ y &= k \quad \text{if } z \geq \mu_{k-1} \end{aligned} \quad (2)$$

where y is the allocated hierarchal code of the categories of the dependent variable (1 = PDO, 2 = Minor Injury, 3 = Major Injury, and 4 = Fatality), μ_k is the decision-making borderline and threshold.

Unlike the fixed-parameter models, in random parameter models, β is not a fixed coefficient, and it changes across the observation to account for the unobserved heterogeneity. Thus, depending on the distribution, the probability function, and the nature of the unobserved factors β can vary as Eq. 3:

$$\beta_i = \beta + \Delta z_i + \Gamma v_i \quad (3)$$

where β is the mean of the estimated random parameters. Δ is the coefficient matrix, z_i is a set of observed variables which do not alter by time and enter the averages of the random parameters, Γ is a lower triangular matrix which produces the covariance matrix of the random parameters. Finally, v_i is the unobservable latent stochastic term in the i th observation in β_i [8].

In previous studies, the prediction technique for future events was only limited to taking the average of all of the generated coefficients of the randomly varying factor. There are some issues associated with this method. Since there was an attempt to capture the unobserved heterogeneity impact with this method, using average value will not be a good representative for such a purpose. Thus, the effect of the

unobserved factors will not be equal in all cases. This fact will be neglected while using the average value and give them the identical weight of unobserved heterogeneity impact on all of the scenarios. An innovative methodology is recommended to avoid such mispredictions. This method uses the “similarity level” by comparing the possible future events with the historical database. Therefore, the following steps are recommended to be followed while predicting using the random parameter models:

Step 1: Extract the relevant information for future events: This information could be gathered by checking the design details, road segment characteristics, setups, and demographics of the neighbourhood and surrounding areas, guidelines, and road condition (Fig. 1).

Step 2: Comparison of events: This step compares the detailed information derived from the previous step with the existing database in which models were developed based on them (Fig. 2).

Step 3: Selection of the events from historical data: This step selects the scenarios which satisfy the “similarity level” condition. The sensitivity of the similarity level could be changed upon the requirements of the project, and the availability of sufficient variation in the dataset. For this study, the “similarity level” of 75% was considered. Thus, the coefficients from the scenarios with more than the assigned trigger were chosen and used for further analysis. The main reason for selecting 75% similarity level was to have at least 10% of the estimated coefficients (approximately 25 coefficients based on training dataset size) for prediction.

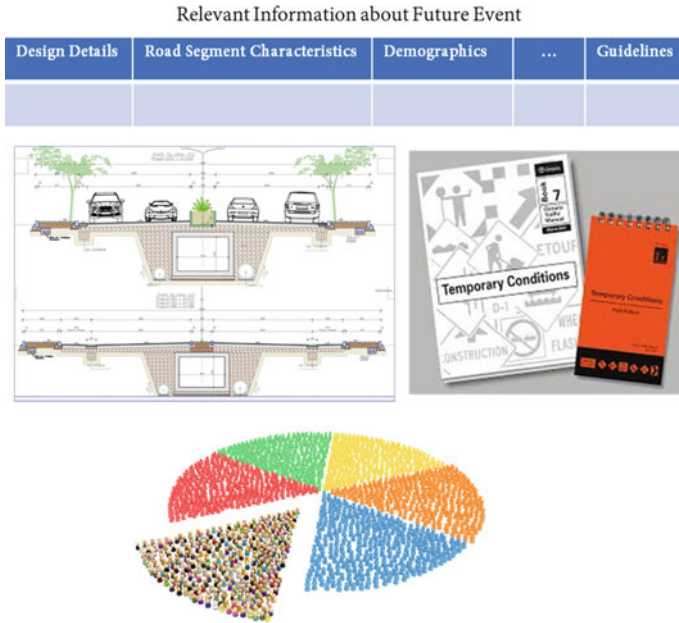


Fig. 1 Step 1: gathering the required relevant information about future events

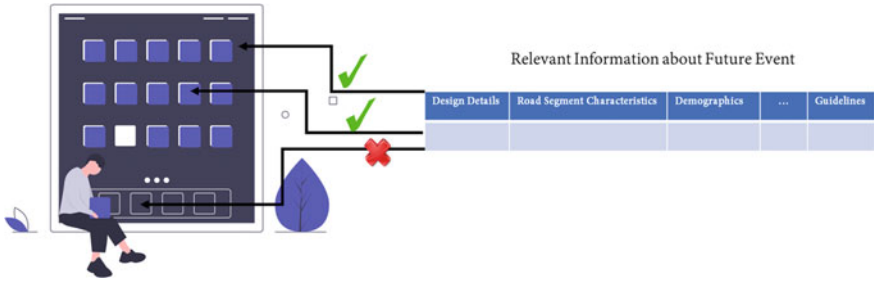


Fig. 2 Step 2: comparison of events, and step 3: selection of the events from the historical database

Step 4: Unifying the scenarios: The average of the chosen scenarios’ coefficients was taken and used for prediction purpose to capture the effect of the selected plots from step 3. Figure 3 depicts the summary of steps 3 and 4.

Lastly, the comparison between the recommended method and the conventional method was performed to check the accuracy and performance of the developed strategy. To achieve this goal, the stratified tenfold cross-validation method was utilized. This method provides advantages such as minimizing the over-fitting issue, estimating the prediction performance of the models, and assurance of consistency



Fig. 3 Demonstration of step 3 and step 4

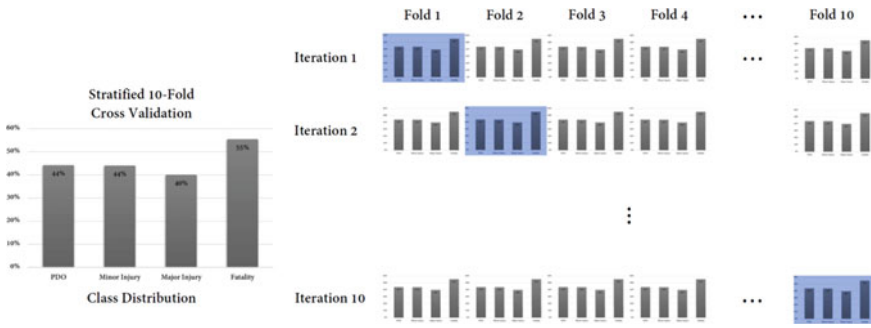


Fig. 4 Stratified tenfold cross-validation methodology for constructing folds [9]

of each fold with the database in terms of characteristic (specific to stratified tenfold cross-validation) [10]. Figure 4 schematically shows the whole procedure.

The overall strategy to achieve the required folds are shown in Fig. 4. It is necessary to follow specific steps, which summarized in Fig. 5.

After setting up the folds, the next step is to deploy an iterative model development process. This step can provide a detailed perspective of the performance and accuracy of the developed models.

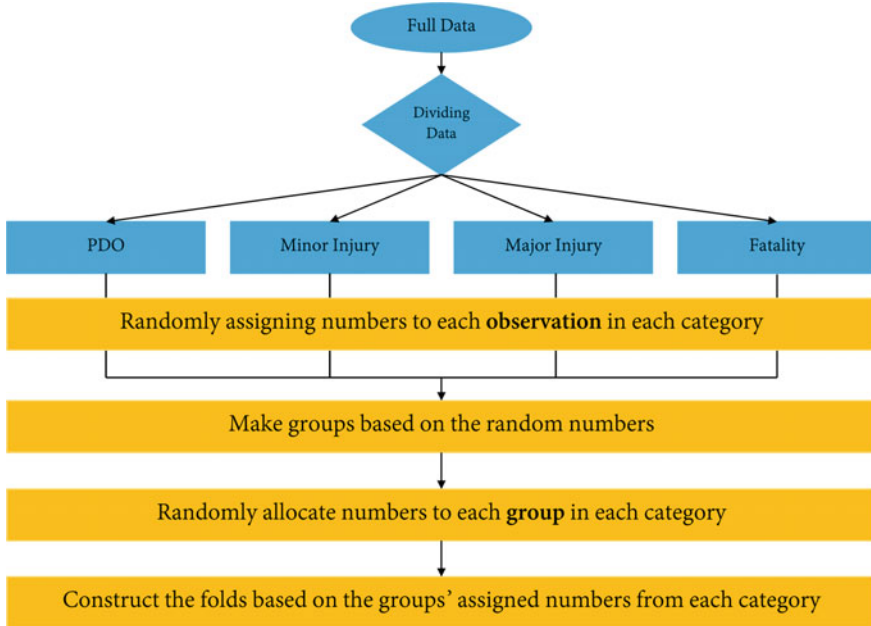


Fig. 5 Overall strategy to construct folds for stratified tenfold cross-validation [9]

3 Data

The dataset summarizes the work zone collisions with various injury-severity levels in the US interstate highways of New York, Pennsylvania, Michigan, and Illinois between 2013 and 2016. This dataset consists of more than 250 data points with multiple layers of information such as driver-related information, vehicle-related information, pre-crash information, and other information. This data was driven from the National Highway Traffic Safety Administration (NHTSA) database. (Please see [8] for more descriptive details of the dataset.)

4 Discussion and Results

According to the Capital Region Intersection Safety Partnership [6], apart from legal and societal damages of work-related fatalities, each lethal collision can cost approximately 228,000\$. Unfortunately, the current statistics demonstrated that costs associated with work environment fatalities increased by 10%; this includes the fatalities related to the workers and staff who lost their lives in traffic accidents. To enhance the prediction performance of unpleasant incidents on highways and work zones, identifying the factors that can significantly increase the risk of fatal collisions is necessary. The previous study by the authors Nahidi and Tighe [8] investigated these factors, and this study focuses on predicting future events using those statistically significant factors.

After dividing the dataset into ten homogenous folds, nine folds were used for the stratified cross-validation process, and one fold was held as the test data set. Later on, a set of models including ten random parameters ordered logit models were developed, and their performances were recorded. Table 1 summarizes the prediction performance of each iteration.

Table 1 demonstrates the summary of the actual injury-severity level versus the predicted injury severity level using the “similarity level” concept. The range of training fold dataset size (90% of data) is altering from 229 to 232, and this range for the test dataset (10% of data) is between 24 and 27. As shown in Table 1, PDO has the lowest accuracy level (approximately 68%). However, the lack of accuracy for this specific category could not be considered “risky”. Since models predicted future events in higher classes, and lead to more conservative designs and decisions. Furthermore, after careful investigation of possible reasons for these mispredictions, it was identified that in more than 80% of the cases with misprediction, the model predictions were only slightly different from the estimated thresholds shown in Eq. 2. Small changes around the threshold values and similarity of several minor injuries and PDO collisions can lead to this issue. It is believed that the development of “Random Threshold Random Parameter Ordered Logit models” can minimize this issue [4]. Table 2 represents the accuracy of predictions for each injury-severity category for each fold. In Table 2, two prediction interpretation strategies were

Table 1 Summary of stratified tenfold cross-validation using “similarity level”

| Fold 1 | Training dataset | | | | | Test dataset | | | | | Fold 2 | | | | | Training dataset | | | | | Test dataset | | | | | | | | | | | |
|-------------|------------------|-------|-------|----|------------|--------------|-------|-------|---|------------|-------------|-------|-------|----|------------|------------------|-------|-------|---|------------|--------------|-------------|-------|----|------------|-----|-------|-------|---|------------|--------------|----|
| | PDO | Mi-In | Ma-In | F | Tot (Act.) | PDO | Mi-In | Ma-In | F | Tot (Act.) | PDO | Mi-In | Ma-In | F | Tot (Act.) | PDO | Mi-In | Ma-In | F | Tot (Act.) | PDO | Mi-In | Ma-In | F | Tot (Act.) | PDO | Mi-In | Ma-In | F | Tot (Act.) | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | Test dataset | |
| PDO | 68 | 1 | 0 | 0 | 69 | 5 | 2 | 1 | 0 | 8 | PDO | 69 | 0 | 0 | 0 | 69 | 5 | 1 | 2 | 0 | 8 | PDO | 69 | 0 | 0 | 0 | 69 | 5 | 1 | 2 | 0 | 8 |
| Mi-In | 0 | 50 | 2 | 0 | 52 | 0 | 6 | 0 | 0 | 6 | Mi-In | 0 | 52 | 1 | 0 | 53 | 0 | 3 | 2 | 0 | 5 | Mi-In | 0 | 52 | 1 | 0 | 53 | 0 | 3 | 2 | 0 | 5 |
| Ma-In | 0 | 1 | 33 | 0 | 34 | 0 | 1 | 3 | 0 | 4 | Ma-In | 0 | 0 | 33 | 1 | 34 | 0 | 0 | 3 | 1 | 4 | Ma-In | 0 | 0 | 33 | 1 | 34 | 0 | 0 | 3 | 1 | 4 |
| F | 0 | 0 | 2 | 72 | 74 | 0 | 1 | 1 | 7 | 9 | F | 0 | 0 | 1 | 74 | 75 | 0 | 3 | 0 | 5 | 8 | F | 0 | 0 | 1 | 74 | 75 | 0 | 3 | 0 | 5 | 8 |
| Tot. (Pre.) | 68 | 52 | 37 | 72 | 229 | 5 | 10 | 5 | 7 | 27 | Tot. (Pre.) | 69 | 52 | 35 | 75 | 231 | 5 | 7 | 6 | 25 | | Tot. (Pre.) | 69 | 52 | 35 | 75 | 231 | 5 | 7 | 6 | 25 | |
| Fold 3 | Training dataset | | | | | Test dataset | | | | | Fold 4 | | | | | Training dataset | | | | | Test dataset | | | | | | | | | | | |
| PDO | 69 | 0 | 0 | 0 | 69 | 5 | 2 | 1 | 0 | 8 | PDO | 68 | 1 | 0 | 0 | 69 | 5 | 2 | 1 | 0 | 8 | PDO | 68 | 1 | 0 | 0 | 69 | 5 | 2 | 1 | 0 | 8 |
| Mi-In | 0 | 52 | 0 | 0 | 52 | 0 | 5 | 1 | 0 | 6 | Mi-In | 0 | 52 | 1 | 0 | 53 | 1 | 3 | 0 | 1 | 5 | Mi-In | 0 | 52 | 1 | 0 | 53 | 1 | 3 | 0 | 1 | 5 |
| Ma-In | 0 | 0 | 32 | 2 | 34 | 0 | 1 | 3 | 0 | 4 | Ma-In | 0 | 0 | 33 | 2 | 35 | 0 | 1 | 2 | 0 | 3 | Ma-In | 0 | 0 | 33 | 2 | 35 | 0 | 1 | 2 | 0 | 3 |
| F | 0 | 0 | 2 | 73 | 75 | 1 | 0 | 1 | 6 | 8 | F | 0 | 0 | 1 | 74 | 75 | 0 | 1 | 1 | 6 | 8 | F | 0 | 0 | 1 | 74 | 75 | 0 | 1 | 1 | 6 | 8 |
| Tot. (Pre.) | 69 | 52 | 34 | 75 | 230 | 6 | 8 | 6 | 6 | 26 | Tot. (Pre.) | 68 | 53 | 35 | 76 | 232 | 6 | 7 | 4 | 7 | 24 | Tot. (Pre.) | 68 | 53 | 35 | 76 | 232 | 6 | 7 | 4 | 7 | 24 |
| Fold 5 | Training dataset | | | | | Test dataset | | | | | Fold 6 | | | | | Training dataset | | | | | Test dataset | | | | | | | | | | | |
| PDO | 69 | 1 | 0 | 0 | 70 | 5 | 1 | 1 | 0 | 7 | PDO | 69 | 0 | 0 | 0 | 69 | 6 | 1 | 1 | 0 | 8 | PDO | 69 | 0 | 0 | 0 | 69 | 6 | 1 | 1 | 0 | 8 |
| Mi-In | 0 | 52 | 0 | 0 | 52 | 1 | 4 | 0 | 1 | 6 | Mi-In | 0 | 50 | 2 | 0 | 52 | 1 | 3 | 2 | 0 | 6 | Mi-In | 0 | 50 | 2 | 0 | 52 | 1 | 3 | 2 | 0 | 6 |

(continued)

Table 1 (continued)

| | | | | | | | | | | | | | | | | | | | | | |
|------------------|-------|-------|----|------------|-----|-------|-------|---|------------|-------------|--------------|-------|----|------------|-----|-------|-------|---|------------|----|----|
| Ma-In | 0 | 0 | 32 | 2 | 34 | 0 | 1 | 2 | 1 | 4 | Ma-In | 0 | 0 | 32 | 2 | 34 | 0 | 1 | 3 | 0 | 4 |
| F | 0 | 0 | 1 | 73 | 74 | 0 | 1 | 2 | 6 | 9 | F | 0 | 0 | 3 | 72 | 75 | 0 | 0 | 1 | 7 | 8 |
| Tot. (Pre.) | 69 | 53 | 33 | 75 | 230 | 6 | 7 | 5 | 8 | 26 | Tot. (Pre.) | 69 | 50 | 37 | 74 | 230 | 7 | 5 | 7 | 7 | 26 |
| Fold 7 | | | | | | | | | | | | | | | | | | | | | |
| Training dataset | | | | | | | | | | | Test dataset | | | | | | | | | | |
| PDO | Mi-In | Ma-In | F | Tot (Act.) | PDO | Mi-In | Ma-In | F | Tot (Act.) | PDO | Mi-In | Ma-In | F | Tot (Act.) | PDO | Mi-In | Ma-In | F | Tot (Act.) | | |
| 69 | 0 | 0 | 0 | 69 | 6 | 2 | 0 | 0 | 8 | 69 | 1 | 0 | 0 | 70 | 5 | 2 | 0 | 0 | 7 | | |
| Mi-In | 0 | 52 | 0 | 52 | 0 | 4 | 1 | 1 | 6 | Mi-In | 0 | 52 | 0 | 52 | 0 | 4 | 2 | 0 | 6 | | |
| Ma-In | 0 | 0 | 31 | 34 | 0 | 0 | 3 | 1 | 4 | Ma-In | 0 | 0 | 33 | 2 | 35 | 0 | 2 | 1 | 3 | | |
| F | 0 | 0 | 1 | 73 | 74 | 1 | 0 | 1 | 9 | F | 0 | 0 | 1 | 74 | 75 | 0 | 1 | 1 | 6 | 8 | |
| Tot. (Pre.) | 69 | 52 | 32 | 76 | 229 | 7 | 6 | 5 | 27 | Tot. (Pre.) | 69 | 53 | 34 | 76 | 232 | 5 | 7 | 5 | 7 | 24 | |
| Fold 9 | | | | | | | | | | | | | | | | | | | | | |
| Training dataset | | | | | | | | | | | Test dataset | | | | | | | | | | |
| PDO | Mi-In | Ma-In | F | Tot (Act.) | PDO | Mi-In | Ma-In | F | Tot (Act.) | PDO | Mi-In | Ma-In | F | Tot (Act.) | PDO | Mi-In | Ma-In | F | Tot (Act.) | | |
| 69 | 0 | 0 | 0 | 69 | 5 | 3 | 0 | 0 | 8 | 70 | 0 | 0 | 0 | 70 | 5 | 1 | 1 | 0 | 7 | | |
| Mi-In | 0 | 51 | 1 | 52 | 0 | 4 | 2 | 0 | 6 | Mi-In | 0 | 51 | 1 | 52 | 0 | 5 | 1 | 0 | 6 | | |
| Ma-In | 0 | 0 | 32 | 34 | 0 | 1 | 2 | 1 | 4 | Ma-In | 0 | 0 | 32 | 2 | 34 | 0 | 0 | 4 | 4 | | |
| F | 0 | 0 | 1 | 74 | 75 | 0 | 1 | 1 | 6 | F | 0 | 0 | 1 | 74 | 75 | 1 | 0 | 1 | 6 | 8 | |

(continued)

Table 1 (continued)

| | | | | | | | | | | | | | | | | | | | | | |
|-------------|----|----|----|----|-----|---|---|---|---|----|-------------|----|----|----|----|-----|---|---|---|---|----|
| Tot. (Pre.) | 69 | 51 | 34 | 76 | 230 | 5 | 9 | 5 | 7 | 26 | Tot. (Pre.) | 70 | 51 | 34 | 76 | 231 | 6 | 6 | 7 | 6 | 25 |
|-------------|----|----|----|----|-----|---|---|---|---|----|-------------|----|----|----|----|-----|---|---|---|---|----|

PDO Property damage only, *Mi-In* Minor injury, *Ma-In* Major injury, *F* Fatality

Table 2 Prediction accuracy of random parameter ordered probability models—category specific prediction versus conservative prediction

| Fold 1 | | Training dataset | | Test dataset | | Fold 2 | | Training dataset | | Test dataset | |
|--------|----------------------------------|------------------------------|----------------------------------|------------------------------|----------------------------------|------------------------------|--|----------------------------------|------------------------------|----------------------------------|------------------------------|
| | Category specific prediction (%) | Conservative predictions (%) | Category specific prediction (%) | Conservative predictions (%) | Category specific prediction (%) | Conservative predictions (%) | | Category specific prediction (%) | Conservative predictions (%) | Category specific prediction (%) | Conservative predictions (%) |
| PDO | 98.55 | 100.00 | 62.50 | 100.00 | PDO | 100.00 | | 100.00 | 100.00 | 62.50 | 100.00 |
| Mi-In | 96.15 | 100.00 | 100.00 | 100.00 | Mi-In | 98.11 | | 98.11 | 100.00 | 60.00 | 100.00 |
| Ma-In | 97.06 | 97.06 | 75.00 | 75.00 | Ma-In | 97.06 | | 97.06 | 100.00 | 75.00 | 100.00 |
| F | 97.30 | 97.30 | 77.78 | 77.78 | F | 98.67 | | 98.67 | 98.67 | 62.50 | 62.50 |
| Fold 3 | | Training dataset | | Training dataset | | Fold 4 | | Training dataset | | Test dataset | |
| | Category specific prediction (%) | Conservative predictions (%) | Category specific prediction (%) | Conservative predictions (%) | Category specific prediction (%) | Conservative predictions (%) | | Category specific prediction (%) | Conservative predictions (%) | Category specific prediction (%) | Conservative predictions (%) |
| PDO | 100.00 | 100.00 | 62.50 | 100.00 | PDO | 98.55 | | 98.55 | 100.00 | 62.50 | 100.00 |
| Mi-In | 100.00 | 100.00 | 83.33 | 100.00 | Mi-In | 98.11 | | 98.11 | 100.00 | 60.00 | 80.00 |
| Ma-In | 94.12 | 100.00 | 75.00 | 75.00 | Ma-In | 94.29 | | 94.29 | 100.00 | 66.67 | 66.67 |
| F | 97.33 | 97.33 | 75.00 | 75.00 | F | 98.67 | | 98.67 | 98.67 | 75.00 | 75.00 |
| Fold 5 | | Training dataset | | Training dataset | | Fold 6 | | Training dataset | | Test Dataset | |
| | Category specific prediction (%) | Conservative predictions (%) | Category specific prediction (%) | Conservative predictions (%) | Category specific prediction (%) | Conservative predictions (%) | | Category specific prediction (%) | Conservative predictions (%) | Category specific prediction (%) | Conservative predictions (%) |
| PDO | 98.57 | 100.00 | 71.43 | 100.00 | PDO | 100.00 | | 100.00 | 100.00 | 75.00 | 100.00 |
| Mi-In | 100.00 | 100.00 | 66.67 | 83.33 | Mi-In | 96.15 | | 96.15 | 100.00 | 50.00 | 83.33 |
| Ma-In | 94.12 | 100.00 | 50.00 | 75.00 | Ma-In | 94.12 | | 94.12 | 100.00 | 75.00 | 75.00 |
| F | 98.65 | 98.65 | 66.67 | 66.67 | F | 96.00 | | 96.00 | 96.00 | 87.50 | 87.50 |

(continued)

Table 2 (continued)

| Fold 1 | Training dataset | | Test dataset | | Fold 2 | Training dataset | | Test dataset | |
|--------|----------------------------------|------------------------------|----------------------------------|------------------------------|---------|----------------------------------|------------------------------|----------------------------------|------------------------------|
| | Category specific prediction (%) | Conservative predictions (%) | Category specific prediction (%) | Conservative predictions (%) | | Category specific prediction (%) | Conservative predictions (%) | Category specific prediction (%) | Conservative predictions (%) |
| Fold 7 | Training dataset | | Training dataset | | Fold 8 | Training dataset | | Test Dataset | |
| | Category specific prediction (%) | Conservative predictions (%) | Category specific prediction (%) | Conservative predictions (%) | | Category specific prediction (%) | Conservative predictions (%) | Category specific prediction (%) | Conservative predictions (%) |
| PDO | 100.00 | 100.00 | 75.00 | 100.00 | PDO | 98.57 | 100.00 | 71.43 | 100.00 |
| Mi-In | 100.00 | 100.00 | 66.67 | 100.00 | Mi-In | 100.00 | 100.00 | 66.67 | 100.00 |
| Ma-In | 91.18 | 100.00 | 75.00 | 100.00 | Ma-In | 94.29 | 100.00 | 66.67 | 100.00 |
| F | 98.65 | 98.65 | 77.78 | 77.78 | F | 98.67 | 98.67 | 75.00 | 75.00 |
| Fold 9 | Training dataset | | Training dataset | | Fold 10 | Training dataset | | Test Dataset | |
| | Category specific prediction (%) | Conservative predictions (%) | Category specific prediction (%) | Conservative predictions (%) | | Category specific prediction (%) | Conservative predictions (%) | Category specific prediction (%) | Conservative predictions (%) |
| PDO | 100.00 | 100.00 | 62.50 | 100.00 | PDO | 100.00 | 100.00 | 71.43 | 100.00 |
| Mi-In | 98.08 | 100.00 | 66.67 | 100.00 | Mi-In | 98.08 | 100.00 | 83.33 | 100.00 |
| Ma-In | 94.12 | 100.00 | 50.00 | 75.00 | Ma-In | 94.12 | 100.00 | 100.00 | 100.00 |
| F | 98.67 | 98.67 | 75.00 | 75.00 | F | 98.67 | 98.67 | 75.00 | 75.00 |

Table 3 Average performance of all 10-folds

| | Average of all folds | | | |
|-------|----------------------------------|-----------------------------|----------------------------------|-----------------------------|
| | Training dataset | | Test dataset | |
| | Category specific prediction (%) | Conservative prediction (%) | Category specific prediction (%) | Conservative prediction (%) |
| PDO | 99.42 | 100.00 | 67.68 | 100.00 |
| Mi-In | 98.47 | 100.00 | 70.33 | 94.67 |
| Ma-In | 94.45 | 99.71 | 70.83 | 84.17 |
| F | 98.13 | 98.13 | 74.72 | 74.72 |

considered: Category-specific predictions, and conservative predictions. More conservative predictions result in consideration of additional safety countermeasures and avoid unpleasant events on a larger scale.

For having a better perspective toward the average performance of the models, the average performance of all folds is presented in Table 3.

Table 3 demonstrates that as expected, models can predict the training dataset events with more than 95% accuracy for all injury-severity categories under both category-specific and conservative prediction strategies. This performance dropped after introducing the test datasets to the developed models for each fold. The category-specific predictions for the test dataset have an overall accuracy of 70%; however, this value for conservative prediction is 88.5%. Adapting the “similarity level” strategy as well as utilizing the conservative prediction technique enhanced the prediction performance of the developed models. Table 4 recapitulates the performance of conventional prediction techniques.

Comparing Tables 3 and 4 shows that the overall prediction performance has improved by at least 27% in category-specific predictions; for conservative prediction this value (apart from PDO collisions) 5%. It is assumed that introducing the similarity level concept combined with the random thresholds technique could improve the accuracy and performance of the predictions even more than what has achieved [4].

Table 4 Prediction performance of model without similarity level and tenfold cross-validation techniques

| | Conventional prediction results (without tenfold cross-validation and similarity level) | |
|-------|---|-----------------------------|
| | Category specific prediction (%) | Conservative prediction (%) |
| PDO | 40.26 | 100.00 |
| Mi-In | 31.03 | 68.97 |
| Ma-In | 15.79 | 73.68 |
| F | 43.37 | 69.74 |

5 Conclusion

This study examined and combined various techniques to improve the prediction performance of “Random Parameter Ordered Logit Models” for future events and tackle one of the challenges of these models, which was their data dependency. The recommended method incorporates the stratified tenfold cross-validation technique and innovative similarity level variable. The combination of these techniques meaningfully enhanced the prediction performance of these models by at least 27% in category-specific predictions and 5% in conservative forecasts. The findings of this study could help designers and project managers involved in highway construction, maintenance, and rehabilitation activities, to have safer planning toward the possible hazards. This study is also in line with Vision Zero objectives which aim to minimize and optimistically eliminate the fatalities on Canadian roads.

References

1. Anastasopoulos PC, Fountas G, Sarwar MT, Karlaftis MG, Sadek AW (2017) Transport habits of travelers using new energy type modes: a random parameters hazard-based approach of travel distance. *Transp Res Part C: Emerg Technol* 77:516–528. <https://doi.org/10.1016/j.trc.2017.01.017>
2. Anastasopoulos C, Mannering FL (2009) A note on modeling vehicle accident frequencies with random-parameters count models. *Accid Anal Prev* 41:153–159. <https://doi.org/10.1016/j.aap.2008.10.005>
3. Clark C (1961) The greatest of a finite set of random variables. *Oper Res* 9(2):145–162. <https://doi.org/10.1063/1.3061193>
4. Fountas G, Anastasopoulos PC (2017) A random thresholds random parameters hierarchical ordered probit analysis of highway accident injury-severities. *Anal Methods Acc Res*. <https://doi.org/10.1016/j.amar.2017.03.002>
5. Lee J, Mannering F (2002) Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. *Accid Anal Prev* 34(2):149–161. [https://doi.org/10.1016/S0001-4575\(01\)00009-4](https://doi.org/10.1016/S0001-4575(01)00009-4)
6. Leur P de, De P (2018) Collision cost study update FINAL report
7. Nahidi S, Fountas G, Sarvani SP, Sarwar MT, Anastasopoulos PC (2017) Project discrepancies in roadway construction and preservation: a statistical analysis of public-private partnership contract types in the US. *Front Built Environ* 3. <https://doi.org/10.3389/fbuil.2017.00015>
8. Nahidi S, Tighe SL (2019) Comparative assessment of various random parameter ordered models: a comprehensive evaluation of work zone collision injury-severities. In: 2019 CSCE transportation specialty conference. Canadian Society of Civil Engineers, Montreal
9. Nahidi S (2019) Work zone safety assessment and throughput analysis for high volume highways using random parameter models. University of Waterloo, Waterloo. <https://uwspace.uwaterloo.ca/handle/10012/15318>
10. Refaeilzadeh P, Tang L, Liu H (2009) Cross-validation. In: Liu L, Tamer Özsu M (eds) *Encyclopedia of database systems*. Springer, Boston, MA, pp 532–538. https://doi.org/10.1007/978-0-387-39940-9_565
11. Washington S, Karlaftis M, Mannering F, Anastasopoulos P (2020) Statistical and econometric methods for transportation data analysis. *Stat Econom Methods Transp Data Anal* <https://doi.org/10.1201/9780429244018>

Quality of Life Outcomes of the Smarter Growth Neighborhood Design Principles: Case Study City of Kelowna



Abdul Rahman Masoud, Ahmed Idris, and Gordon Lovegrove

1 Introduction

The built environment influences many aspects of human behaviour and our quality of life (QoL). For instance, neighbourhood design determines how children engage with public places and their level of access to green spaces and their level of freedom and independent mobility [30]. These factors, in turn, affect children's cognitive, emotional, social, and bodily development [14]. As for adults, it has been well established in the literature that mental health problems, including social isolation, are connected to the built environment [16]. In addition, neighbourhood design shapes our travel behaviour and traffic pattern, which influences air quality and noise pollution.

No standard published definition for QoL was found in the literature, due in large part to its historically abstract references. However, many researchers argue that it is a multidimensional construct of interacting objective and subjective dimensions [4], with [22] arguing QoL can be classified into four categories: (1) objective, (2) subjective, (3) combination of objective and subjective, and (4) domain specific. QoL in its simplest form can be defined as "how well we are doing, and how well things are going for us" [10]. The question then becomes how to assess QoL, and thus the success of a neighborhood's design.

Given that QoL covers a wide array of life domains, several indicators and indices have been proposed to measure it. For instance, [27] proposed a list of 22 indicators

A. R. Masoud (✉) · G. Lovegrove
School of Engineering, University of British Columbia, 1137 Alumni Avenue, Kelowna, BC
V1V1V7, Canada
e-mail: amasoud@mail.ubc.ca

A. Idris
Construction and Building Engineering Department, Arab Academy for Science, Technology and
Maritime Transport, Abu Kir, P.O. Box 1029, Alexandria, Egypt

based on an extensive literature review of human values and wellbeing in regard to sustainable development. The list includes indicators related to, but not limited to, social aspects, health, safety, security, privacy, freedom, environmental quality, spirituality, materialism, access to nature, education, and leisure. The indicators can be used to estimate the change in people's QoL by constructing a multi-attribute evaluation (MAE) matrix given predefined weights for each indicator. Similar to [9, 27] proposed a QoL model that incorporates objective and subjective dimensions, as well as a new dimension that captures personal autonomy of individuals' values and preferences across domains. In particular, their model accounts for (1) physical wellbeing, (2) material wellbeing, (3) social wellbeing, (4) development and activity, and (5) emotional wellbeing.

These previous measures come from the sociology literature, as do most QoL measures, which is not surprising since the concept 'QoL' evolved from research on social indicators [11]. Few researchers have explored developing urban and/or transportation QoL measures. For urban-specific measures, the Dutch have identified six sustainability principles related to improving QoL: (1) health, climate resistance, mobility, circularity (i.e. recycling and reusing), energy, and social economy. Serag El Din et al. [29] identified seven domains of QoL: (1) environmental, (2) physical, (3) mobility, (4) social, (5) psychological, (6) economical, and (7) political. In addition, [3] presented more compact urban QoL measures with four domains: (1) architectural and town-planning (i.e. peoples' perception of space and green areas), (2) socio relationship, (3) functionality (e.g. commercial, and transportation services), and (4) Contextual (i.e. psychological wellbeing).

As for transportation-related QoL, [19] identified seven QoL indicators based on an extensive literature review including (1) energy saving, (2) air quality, (3) health, (4) road safety, (5) congestion and travel time, (6) noise, and (7) equity. In addition, [22] proposed a framework for the interactions of transportation planning and QoL. In particular, their framework linked three components of the transportation system (vehicular traffic, mobility/accessibility, and the built environment) to four domains of QoL (physical, mental, economic, and social well-being).

Overall, the varied results of these numerous studies suggest that measuring the influence of land use and transportation systems on QoL is exceedingly complex. In addition, the wide variation of the proposed indicators reflects how there is no global agreement on the definition of quality life nor the importance of the factors that influence QoL. This paper will not attempt to develop the perfect neighborhood system design that 'maximizes' human QoL, as humans by nature are each unique, experience life differently, and thereby value the quality of their life experiences differently. Therefore, the objectives of this paper are to:

1. Identify a reliable suite of empirical tools that could be utilized by community planners and engineers to objectively evaluate key factors of a neighbourhood's design.
2. Utilize the identified suite of tools to evaluate the impact of the full-fledged SG principles on community QoL.

2 Literature Review

Several QoL evaluation tools have been previously identified and/or developed by researchers at the Sustainable Transport Safety lab at the University of British Columbia Okanagan [19, 25]. In particular, six tools and indicators are considered and discussed below, including: (1) I-Thrive, (2) air quality, (3) noise pollution, (4) walkability, (5) bikeability, and (6) Social Interactions.

First, The Interactive Sustainable Transport Safety/Healthy Development Index Valuation (I-THRIVE) is an excel-based tool that has been developed to help community planners and engineers in assessing the sustainability benefits of developing/retrofitting transportation and land use designs of a neighbourhood [23]. The tool consists of 12 elements, including (1) density, (2) service proximity, (3) land use mix, (4) street connectivity, (5) road network and sidewalk characteristics, (6) parking, (7) aesthetics and human Scale, (8) functionability, (9) predictability, (10) homogeneity, (11) forgivingness, and (12) state awareness. The first seven elements are related to the built environment evaluation and were mostly adopted from the Peel Healthy Development Index [7]. The remaining five elements are related to road safety evaluation and were adopted from the Sustainable Transport Safety (STS) principles. The STS principles were proposed in 1992 by the Dutch Road Safety Research Institute with two goals: (1) to prevent road crashes by decreasing the risk of human errors and, (2) recognizing that human error is bound to happen, to reduce the severity of any crashes that do occur to a level that the human body can tolerate [33].

Second, the transportation sector is a major contributor of other air pollutants such as particulate matter, carbon monoxide, nitrogen oxide, and volatile organic compounds [35]. Passenger vehicles account for a large sum of the total transportation related emissions in Canada including 4% of particulate matter, 21% of nitrogen oxide, and 51% of volatile organic compound emissions [8]. Numerous studies have been conducted to assess the impact of the built environment on transportation-related emissions [5, 6]. Similar to [28], this research utilizes emission factors (EF) to calculate total traffic emissions based on vehicle kilometres travelled (VKT). The GHG emission factors were extracted from BC Community Energy and Emissions Inventory [1].

Third, noise pollution is a major concern for many urban areas. Gandelman et al. [13] suggested that eliminating noise pollution significantly improves individuals' leisure and social life satisfaction. Traffic noise is a complex phenomenon that is influenced by several factors including traffic speed, volume and composition, road surface, meteorological factors, ground surface, tire type, etc., most of which are country specific [15]. As a result, several noise forecasting models have been developed around the world. This research will utilize the Ontario Road Noise Analysis Method for Environment and Transportation (ORNAMENT), illustrated in Eq. 1 [20], which accounts for traffic composition, volume, and speed. The model was adopted in this research for several reasons including its simplicity, being developed in a region close to the study area, and it accounts for the variables that studied in

this research.

$$L_{eq} = 42.3 + 10.2 \times \log(V_C + 6 \times V_T) - 13.9 \times \log(D) + 0.1S \quad (1)$$

where L_{eq} is the one hour equivalent sound level at a point of reception, V_C is the light-vehicle traffic volume per hour, V_T is the heavy-vehicle traffic volume per hour, D is the distance from the road to the observation point, and S is the average traffic speed (km/hr).

Forth, walkability has emerged as a new area of research that aims at identifying metrics of the level to which the built environment, including transportation ROW, is walking friendly. A number of approaches have been previously proposed to measure walkability [12, 31]. A common theme between most of the proposed walkability indices is that they try to capture three components of the built environment: (1) network connectivity, (2) residential density, and (3) land use diversity. Walkability in this research is measured using an index proposed by Frank et al. [12] for a QoL study. The index is a function of four components as follows: (1) net residential density, (2) retail floor area ratio, (3) Intersection density, and (4) Land use mix. The four variables were then normalized using z-score. Finally, the walkability index for each block group was calculated as the sum of the z-scores of the four components as shown in Eq. 2.

$$WI = (2 \times z - \text{intersection density}) + (z - \text{residential density}) \\ + (z - \text{retail FAR}) + (z - \text{land use entropy}) \quad (2)$$

Fifth, similar to walkability, bikeability refers to the extent to which the urban built environment is cycling friendly. However, the development of a bikeability index has received relatively lesser attention than walkability. Nevertheless, few researchers have proposed a bikeability index [34], of which the index developed by [34] was the most cited. This index measures bikeability based on metrics of five components: (1) cycling route density, (2) cycling route separation level, (3) connectivity of cycling friendly streets, (4) topography, and (5) destination density. The index value is computed by first creating a high-resolution surface of 10 m grid-cell raster in ArcGIS for each of the five components. Second, each surface was reclassified to deciles with a scale from 1 to 10, where 1 is least cycling friendly and 10 is the most cycling friendly. Finally, the bikeability index for each grid-cell is computed by summing the score of the five components.

Finally, the Canadian 24-h movement guidelines identify four key elements that are associated with notable health and cognitive benefits for children and youth: (1) accumulation of no less than 60 min per day of moderate to vigorous physical activity, (2) several hours of a variety of structured and unstructured light physical activities, (3) uninterrupted sleep of eight to 11 h, and (4) no more than 2 h per day of recreational screen time [32]. However, only less than fifth of Canadian children and youth adhere to these guidelines [18]. Researchers have not yet proposed a playability index for the built environment. Nevertheless, several studies in the literature have identified some

key elements of the built environment that are associated with children playability. In particular, playability is positively associated with close proximity to destinations [2], walkability [21], perception of traffic safety [26], and negatively associated with residential density [24]. Due to the lack of an index, playability in this research will be subjectively evaluated based on the aforementioned built environment elements.

3 Methodology

To demonstrate the proposed suite of tools, it was applied to three urban centres in the city of Kelowna: (1) Capri-Landmark, (2) South Pandosy, and (3) Rutland. These urban centres have been selected as a study area for two main reasons: (1) the city has mandated to accommodate 26% of its future growth in them and (2) the selected urban centres represent various levels of land use mix. All the three neighbourhoods were hypothetically retrofitted using the SMARTer Growth (SG) principles [17], as shown in Fig. 1. The retrofitting process involved a series of local road closures to preclude through traffic on local roads and direct it to the perimeter, without altering the existing non-motorized network. Moreover, some areas that resulted from road closures were used to provide green spaces in the neighborhood. These green spaces eventually supplemented the local road network to provide more ambient off-road pathways that allow convenient walking and cycling across the neighborhood. Finally, the distribution of population and employment within the neighbourhood was changed according to the SG Principles such as higher density and mixed land use located along the perimeter arterial and major collector corridors. Accordingly, two scenarios were examined for each of the urban centres: (1) the 2016 scenario which represents existing conditions (the year of 2016 was selected due to data availability) and (2) the 2040 scenario which is based on the city’s projected urban growth and the SG design principles.

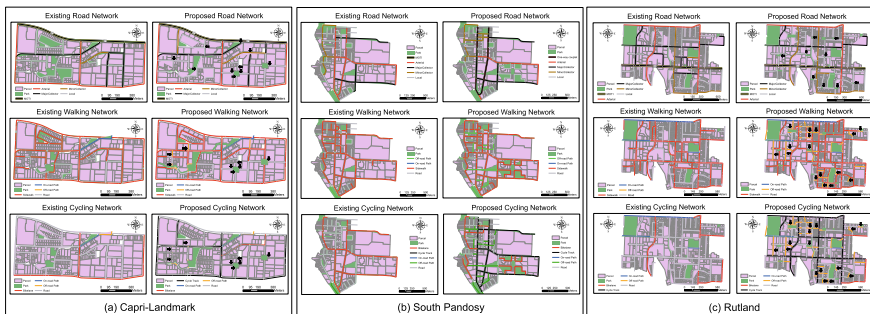


Fig. 1 Existing and SMARTer Growth design alternatives

4 Results and Discussion

This section presents the modelling results of evaluating the impact of the SG neighbourhood design principles on residents' QoL. As discussed earlier, Three urban centres in the city of Kelowna are examined in this research including: Capri-Landmark, South Pandosy, and Rutland. Two scenarios were examined for each of the urban centres, as follows: (1) 2016 transportation and land use systems (S1) and (2) 2040 projected transportation and land use systems utilizing SG design principles, given the city's projected urban growth (S3). The following subsections will report and discuss the results of each of the QoL evaluation tools for the study areas.

4.1 *i-THRIVE*

As shown in Fig. 2, the existing scenario (S1) for all three urban centres scored poorly on the i-THRIVE, with a total score of 33%, 38%, and 35% for Capri-Landmark, South Pandosy, and Rutland, respectively. One of the factors that contributed to the low i-THRIVE scores is the low residential densities and floor area ratio (FAR). The average dwelling unit density for Capri-landmark, South Pandosy, and Rutland are 23, 27, and 26 units/hectare, well below the minimum i-THRIVE requirement of 35 units/hectare. In addition, the non-residential floor area ratio (FAR) in the three urban centres are 0.62, 0.94, and 0.54, respectively. The low FAR values are attributed stand-alone parkade structure in the capri-landmark, and strip malls with vast parking lots in all three urban centres. In addition, outdoor public spaces comprises approximately 2, 4, and 7% of the total neighbourhood land, which means that only the Rutland neighbourhood meets the i-THRIVE requirement of dedicating at least 5% of the total community land as an outdoor public space. In addition, all three urban centres provide new services to the community and a mix of these housing types while only the South Pandosy and Rutland has a public school within 800 m of the community centre. On the other hand, none of the urban centres provide pedestrian use on the ground floor of all mixed-use buildings and 50% of multifamily residential buildings. Similarly, non of the urban centres met the minimum requirement of the predictability,

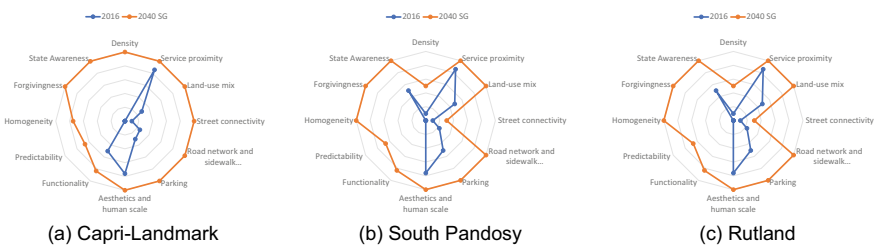


Fig. 2 i-THRIVE evaluation of the study area

homogeneity, forgivingness, and state awareness elements. Nevertheless, the three urban centres achieved high service proximity score of 85% due to the fact that more than 75% of the residential units are within 800 m of 20 neighbourhood services, three food markets, and a park. In addition, 100% of residents are within 800 m of a bus stop and 96, 84, and 88% are within 400 m of a bus stop.

On the other hand, the 2040 SG scenario achieved high i-THRIVE score of 96, 87, and 84% for the Capri-Landmark, South Pandosy, and Rutland, respectively. According to the city's OCP, The Capri-Landmark urban, South Pandosy, and Rutland centre are expected to grow by approximately 7800, 2000, and 3600 people by 2040. Given a forecasted average occupancy of 1.7 person per new household, it is expected that around 4500, 1200, and 2100 new units are needed to accommodate the growth in the the three neighbourhoods. Therefore, the residential unit densities for this scenario are assumed to be 84.5, 30.1, and 44, respectively. For the FAR, the intensified mixed-use zones in the arterial corridor along with the high residential and employment densities ensures that the SG scenario will have a FAR value greater than 2.5. In the retrofitted neighbourhoods, mixed, high residential and commercial land uses are located along calmed collectors and arterials. In addition, the SG design principles requires that each quadrant centre is within a maximum of 400 m to the nearest collectors and/or arterial. This configuration allows residents to walk or bike to services in all four directions of the neighbourhood within five-minutes. Similarly, the SG design principles facilitates a self-contained community with mixed land use that provides short commuting distances to work, school, or other services. In addition, SG provides at least 5% of the community land as outdoor public spaces since they provide the core of the AT off-road pathway network. The most significant i-THRIVE scores improvement was observed in the Road network and sidewalk characteristics and the STS elements. The SG design principles assigns higher priority to AT users through AT cut-throughs, and off-street pedestrians and cyclist paths, which separate AT users from motorists. However, in cases where AT users and motorists are sharing the same space on local roads, the posted speed is reduced to 15 km/h. In addition, wide sidewalks and cycle tracks are provided on arterial roads where the traffic speed volume is high. In addition, the SG principles includes many traffic calming measures, such as roundabout, three-way intersection, and raised cross walks on major collectors to reduce the severity of collisions. Finally, the three SG retrofits follow a strict hierarchy street network system in which each road fulfils a single function only. For instance, local roads in the SG retrofits only provide accessibility to final destinations within neighborhoods. On the other hand, minor and major collectors distribute traffic between local and arterial roads.

4.2 Air Quality

Traffic emissions in this research are assumed to be correlated with VKT only (i.e. speed, acceleration, and # of stops are not considered). The results below show the GHG forecast for the three urban centres, nevertheless, similar trends are expected

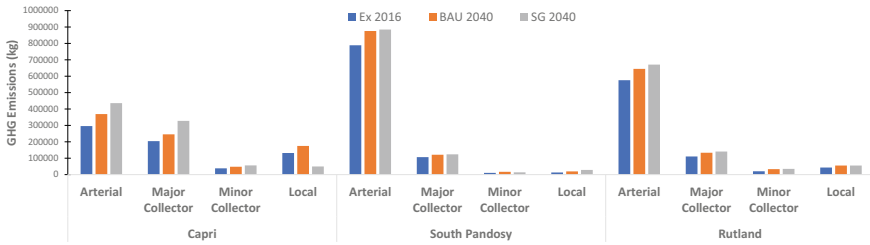


Fig. 3 Traffic emissions aggregated by road classification

for the other pollutants (e.g. particulate matter, carbon monoxide). Traffic emissions were estimated for each road link in the three urban centres for three scenarios: (1) 2016 scenario, 2040 business as usual (BAU), and 2040 SG. Surprisingly, the SG retrofit resulted in 1.82, 1.80, and 0.44% increase in GHG emissions compared to the BAU scenario, due to discontinuous local vehicular network. Figure 3 shows the estimated emissions, aggregated by road classification. The figure shows that the Capri-Landmark SG retrofit significantly reduced emissions on local roads while the emissions on arterials and collectors increased, compared to BAU. Similar trend but of reduced magnitude was observed in the Rutland urban centre. These results demonstrate that the well-defined SG’s street hierarchy did succeed in re-routing through traffic away from local roads to arterials and connectors. On the other hand, the results of the South Pandosy urban centre revealed an increase of emissions on local roads, compared to BAU. This counterintuitive result could be attributed to discrepancies in road classification between the BAU and SG scenarios. Some of the roads in South Pandosy were not classified according to the function they serve; these roads were reclassified in the SG scenario.

4.3 Noise Pollution

Similar to traffic emissions, noise pollution was estimated for each road link and then aggregated by road classification, as shown in Fig. 4. The results show that the SG retrofits have the lowest noise pollution on local roads. The low noise pollution in the SG retrofits are due to the low posted speed limit on local roads and the strict street hierarchy that preclude shortcutting through the residential core.

4.4 Walkability

A walkability surface was generated at the dissemination area level for each of the considered scenarios. The dissemination areas were grouped into deciles based on their walkability score, where the top deciles represent high walkability and the

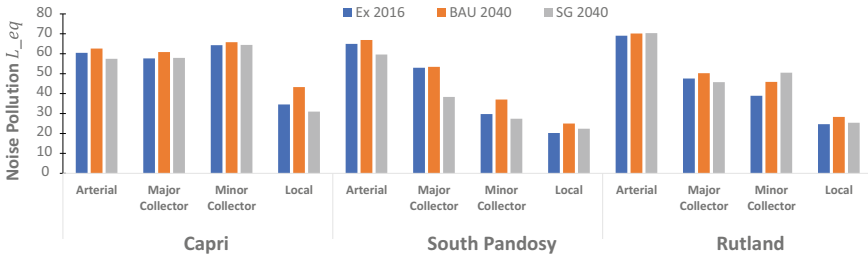


Fig. 4 Noise pollution aggregated by road classification

bottom deciles represent low walkability. Figure 5 shows Kelowna’s walkability surface for the 2016 and 2040 SG scenarios, respectively. For the 2016 scenario, the walkability index scores ranged from -4.49 to 10.77. The South Pandosy urban centre was classified in the highest walkability decile, while the Capri-Landmark and Rutland were classified in the second highest decile. Specifically, the capri-Landmark had high scores for residential density and FAR, moderately high score for intersection density, but scored moderately in land use mix. The South Pandosy urban centre scored high in terms of residential density, FAR, and Land use mix; however, it scored poorly in intersection density, which suggests that it would benefit greatly from improving its walking network via the SG principles. Finally, Rutland

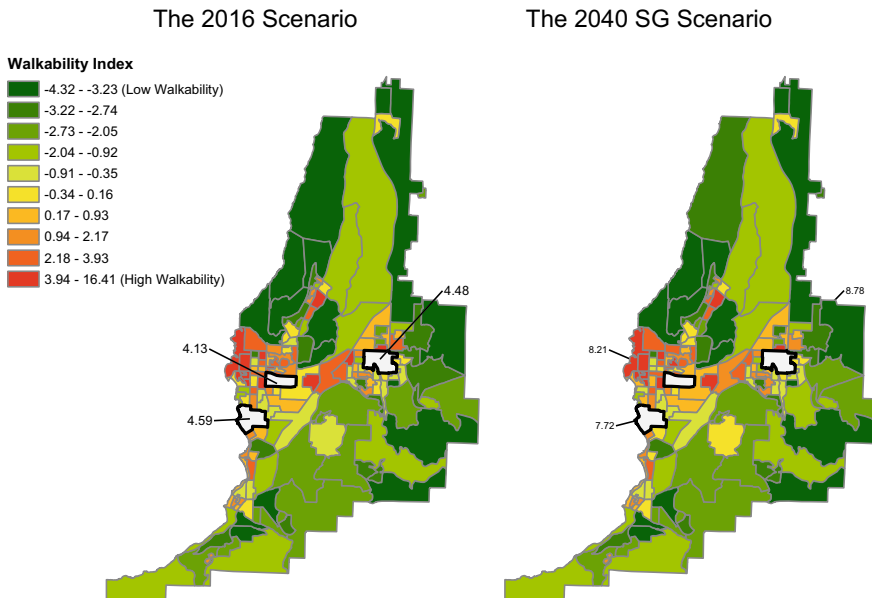


Fig. 5 The walkability Index for the city of Kelowna

had high scores for residential density and land use mix and moderately high scores for FAR and intersection density.

For the 2040 SG scenario's walkability index scores ranged from -4.32 to 16.41 . This scenario achieved a significant increase in walkability for the three urban centres. This increase can be attributed to the significant increase in population and intersection density. The later is attributed to the SG's highly connected walking grid network of an on-road and off-road paths that allow convenient walking and cycling across the neighbourhood in less than 5 min.

4.5 Bikeability

A bikeability surface of 10 m grid-cell raster was generated for each of the considered scenarios, as shown in Fig. 6. First, a raster data was generated for each of the bikeability components (cycling route density, cycling route separation level, connectivity of cycling friendly streets, topography, destination density). Second, each surface was reclassified into deciles and recoded from 1 to 10, with the 1st decile assigned a value of 1 and the 10th decile assigned a value of 10. Third, the five surfaces were combined to generate the bikeability score for each 10 m cell. Finally, the urban centre's bikeability score was generated by aggregating the cells values that are located within them. For the 2016 scenario, the three urban centres

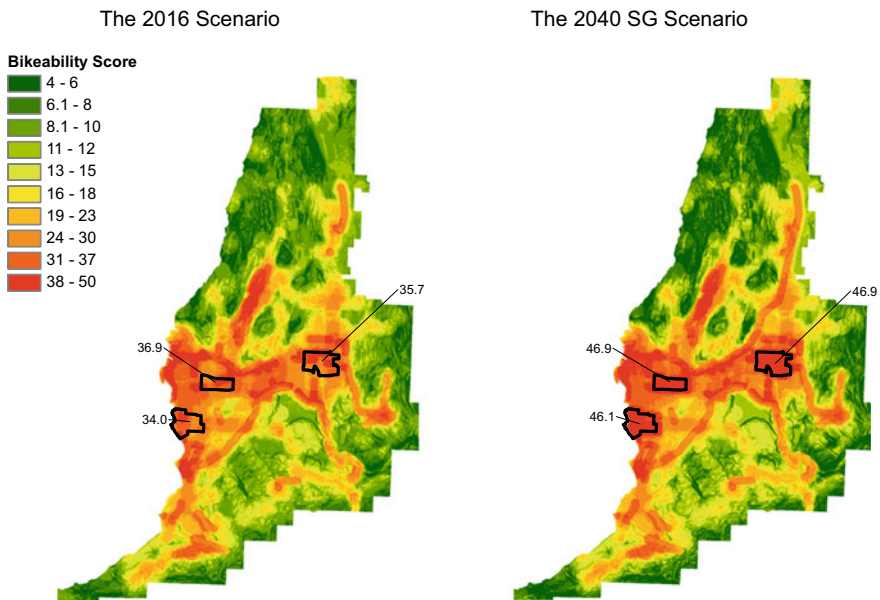


Fig. 6 The bikeability index for the city of Kelowna

scored well on topography, due to their flat slopes. In addition, they scored well on connectivity of cycling friendly streets and potential destination density for the 2016 scenario, which is expected due to their relatively high employment and retail area. On the other hand, the bikeability evaluation revealed a lack of dedicated cycling routes in the Capri-Landmark and Rutland urban centres as well as a lack of the availability of separated bike infrastructure in all three urban centres.

Similar to walkability, the 2040 SG scenario achieved the most significant increase in bikeability for the three urban centres. This increase could be attributed to two factors. First, the SG's continuous grid network of off-road paths and cycle tracks (physically separated) to provide full walk/bike connectivity. Second, the intensified mixed land uses, commercial, and services along collector the arterial corridors in a SG neighbourhood.

4.6 *Playability*

As discussed in Sect. 2, playability in this research will be evaluated subjectively based on the following elements: (1) close proximity to destinations, (2) walkability, (3) traffic safety, and (4) residential density. First, numerous research has found that playability is positively associated with having accessible child-friendly destinations such as schools, parks and play spaces. For the 2016 scenario, all three urban centres score high on this element as 100% of the population live within a 400 m buffer to a park and around 80, 100, and 88% are within an 800 m radius buffer of a school. Similarly, a high proximity to child-friendly destinations is expected for the SG scenario, especially for the later scenario as all residents are expected to live within one-minute walk of a park without crossing any streets. Second, playability was found to be positively associated with high walkable neighbourhoods. The walkability of each urban centre was discussed in Sect. 4.4 in which it was demonstrated that the 2040 SG scenario achieved the highest walkability score. Third, higher perception of traffic safety is associated with higher playabilities for girls. Previous research has predicted that the SG design would significantly improve road safety, with over 60% fewer crashes compared to the traditional grid and cul-de-sac pattern neighbourhood designs (sun and Lovegrove). In addition, the SG design controls the speed (15 km/h) and volume of traffic on local roads maintains low traffic volume on local roads even in its highest density scenario. Therefore, the 2040 SG scenario is expected to score higher on this element than the 2016 and 2040 OCP scenarios.

Finally, playability is positively associated with low residential density. This association could be explained by other features that usually accompany high density such as higher traffic volume and speed and lack of sufficient play and public spaces. The SG design by definition provides traffic calmed with self-enforcing infrastructure that ensures low vehicular speeds in the neighbourhood core. These traffic calming measures include but are not limited to 15 km/h maximum design speed, roundabouts/traffic circles, three-way intersections, discontinuous internal vehicle grid (culs-de-sac), and continuous internal walking and cycling network. All these

features are expected to mitigate the negative effects of high residential density in the 2040 SG scenario.

5 Conclusion

SMARTer Growth (SG) Neighbourhood Design is a novel sustainable community design strategy developed by researchers from the University of British Columbia (UBC) and the Canada Mortgage and Housing Corporation, that aim to promote more livable and sustainable communities. It combines the best characteristics of the conventional culs-de-sac and traditional grid neighbourhood patterns and features many design elements including compact higher density and mixed land uses, car-free cores, and central greens. This research aimed at evaluating the impact of the full-fledged SG principles on community of quality of life (QoL). Six tools were utilized in this research to give an indicator about community QoL including: (1) i-Thrive, (2) air quality, (3) noise pollution, (4) walkability, (5) bikeability, and (6) Social Interactions.

Three urban centres in Kelowna were selected as study areas: (1) Capri-Landmark, (2) South Pandosy, and (3) Rutland. All the three neighbourhoods were hypothetically retrofitted using the SG principles, a process that involved a series of local road closures and redistribution of population and employment, such as high density and mixed land uses are located on the perimeter arterial and major collector corridors. The results show that applying the SG principles in the three urban centres resulted in an increase in the i-THRIVE score, walkability, bikeability, and playability and a decrease in noise pollution. However, the SG scenario resulted in an increase in traffic air pollution compared to the business as usual scenario.

References

1. BC Ministry of Environment (2014) Kelowna city: 2010 community energy and emissions inventory, Monitoring and reporting on progress towards greenhouse gas emissions reduction targets
2. Bohn-Goldbaum EE, Phongsavan P, Merom D, Rogers K, Kamalesh V, Bauman AE (2013) Does playground improvement increase physical activity among children? A Quasi-experimental study of a natural experiment. *J Environ Public Health* 2013:1–9
3. Bonaiuto M, Fornara F, Bonnes M (2006) Perceived residential environment quality in middle- and low-extension Italian cities. *Revue Européenne De Psychologie Appliquée* 56(1):23–34
4. Bowling A, Gabriel Z, Dykes J, Dowding LM, Evans O, Fleissig A, Banister D, Sutton S (2003) Let's ask them: a national survey of definitions of quality of life and its enhancement among people aged 65 and over. *Int J Aging Hum Dev* 56(4):269–306
5. Derrible S, Saneinejad S, Sugar L, Kennedy C (2010) Macroscopic model of greenhouse gas emissions for municipalities. *Transp Res Rec* 2191(2191):174–181
6. Duncan D, Nadella V, Giroux S, Bowers A, Graham JD (2017) The road mileage user-fee: Level, intensity, and predictors of public support. *Transp Policy* 53:70–78

7. Dunn J, Creatore M, Peterson E, Weyman J, Glazier R (2009) Peel healthy development index, The Centre for Research on Inner City Health at St. Michael's Hospital, Toronto, ON
8. Environment and Climate Change Canada (2017) Air pollution from cars, trucks, vans and SUVs. Retrieved from <https://www.canada.ca/en/environment-climate-change/services/air-pollution/sources/transportation/cars-trucks-vans-suvs.html>
9. Felce D, Perry J (1995) Quality of life: its definition and measurement. *Res Dev Disabil* 16(1):51–74. [https://doi.org/10.1016/0891-4222\(94\)00028-8](https://doi.org/10.1016/0891-4222(94)00028-8)
10. Ferkany M (2012) The objectivity of wellbeing. *Pac Philos Q* 93(4):472–492
11. Ferriss AL (2004) The quality of life concept in sociology. *Am Sociol* 35(3):37–51
12. Frank LD, Sallis JF, Saelens BE, Leary L, Cain K, Conway TL, Hess PM (2009) The development of a walkability index: application to the neighborhood quality of life study. *Br J Sports Med* 44(13):924–933. <https://doi.org/10.1136/bjism.2009.058701>
13. Gandelman N, Gandelman N, Piani G, Piani G, Ferre Z, Ferre Z (2012) Neighborhood determinants of quality of life. *J Happiness Stud* 13(3):547–563
14. Garau C, Annunziata A (2020) Supporting children's independent activities in smart and playable public places. *Sustainability (Basel, Switzerland)* 12(8352):8352. <https://doi.org/10.3390/su12208352>
15. Garg N, Maji S (2014) A critical review of principal traffic noise models: strategies and implications. *Environ Impact Assess Rev* 46:68–81. <https://doi.org/10.1016/j.ear.2014.02.001>
16. Gehl J, Ebrary I (2011) *Life between buildings: using public space*. Island Press, Washington, D.C.
17. Grammenos F, Lovegrove GR (2015) *Remaking the city street grid: a model for urban and suburban development*. McFarland & Company Inc., Publishers, Jefferson, N.C.
18. Guerrero M, Livingstone F, Vanderloo L (2020) *ParticipACTION Report card on physical activity for children and youth*, ParticipACTION
19. Jamal EA (2019) *Sustainable community planning: the business case to address declining transport-related quality of life in the Kuwait urban area*. University of British Columbia
20. Khalil MA (2015) *Predictive models for free-field road traffic noise in Sharjah city*. University of Sharjah, United Arab Emirates
21. Kurka JM, Adams MA, Todd M, Colburn T, Sallis JF, Cain KL, Glanz K, Frank LD, Saelens BE (2015) Patterns of neighborhood environment attributes in relation to children's physical activity. *Health Place* 34:164–170. <https://doi.org/10.1016/j.healthplace.2015.05.006>
22. Lee RJ, Sener IN (2016) Transportation planning and quality of life: Where do they intersect? *Transp Policy* 48:146–155. <https://doi.org/10.1016/j.tranpol.2016.03.004>
23. Masoud AR (2015) Sustainable transport safety: a review of barriers to, and promising engineering tools for, promoting safer bicycling and walking in Canadian communities. <https://doi.org/10.14288/1.0166305>
24. Mitra R, Moore SA, Gillespie M, Faulkner G, Vanderloo LM, Chulak-Bozzer T, Rhodes RE, Brussoni M, Tremblay MS (2020) Healthy movement behaviours in children and youth during the COVID-19 pandemic: exploring the role of the neighbourhood environment. *Health Place* 65:102418
25. Ovi MFM (2020) *SMARTer growth neighborhood design manual: application to existing neighborhoods*. University of British Columbia
26. Page AS, Cooper AR, Griew P, Jago R (2010) Independent mobility, perceptions of the built environment and children's participation in play, active travel and structured exercise and sport: the PEACH project. *Int J Behav Nutr Phys Act* 7(1):17
27. Poortinga W, Steg L, Vlek C (2004) Values, environmental concern, and environmental behavior: a study into household energy use. *Environ Behav* 36(1):70–93
28. Rahman MN, Idris AO (2017) TRIBUTE: Trip-based urban transportation emissions model for municipalities. *Int J Sustain Transp* 11(7):540–552
29. Serag El Din H, Shalaby A, Farouh HE, Elariane SA (2013) Principles of urban quality of life for a neighborhood. *HBRC J* 9(1):86–92. <https://doi.org/10.1016/j.hbrj.2013.02.007>
30. Smith M, Amann R, Cavadino A, Raphael D, Kearns R, Mackett R, Mackay L, Carroll P, Forsyth E, Mavoa S, Zhao J, Ikeda E, Witten K (2019) Children's transport built environments: a mixed

- methods study of associations between perceived and objective measures and relationships with parent licence for independent mobility in Auckland, New Zealand. *Int J Environ Res Public Health* 16(8):1361. <https://doi.org/10.3390/ijerph16081361>
31. Stockton JC, Duke-Williams O, Stamatakis E, Mindell JS, Brunner EJ, Shelton NJ (2016) Development of a novel walkability index for London, United Kingdom: cross-sectional application to the Whitehall II Study. *BMC Public Health* 16(1):416. <https://doi.org/10.1186/s12889-016-3012-2>
 32. Tremblay MS, Carson V, Chaput J, Gorber SC, Dinh T, Duggan M, Faulkner G, Gray CE, Gruber R, Janson K, Janssen I, Katzmarzyk PT, Kho ME, Latimer-Cheung AE, LeBlanc C, Okely AD, Olds T, Pate RR, Phillips A, Poitras VJ, Rodenburg S, Sampson M, Saunders TJ, Stone JA, Stratton G, Weiss SK, Zehr L (2016) Canadian 24-Hour movement guidelines for children and youth: an integration of physical activity, sedentary behaviour, and sleep. *Appl Physiol Nutr Metab* 41(6):S311–S327. <https://doi.org/10.1139/apnm-2016-0151>
 33. Wegman F, Aarts L, Bax C (2008) Advancing sustainable safety. *Saf Sci* 46(2):323–343
 34. Winters M, Brauer M, Setton EM, Teschke K (2013) Mapping bikeability: a spatial tool to support sustainable travel. *Environ Plann B Plann Des* 40(5):865–883
 35. Xia T, Nitschke M, Zhang Y, Shah P, Crabb S, Hansen A (2015) Traffic-related air pollution and health co-benefits of alternative transport in Adelaide, South Australia. *Environ Int* 74:281–290. <https://doi.org/10.1016/j.envint.2014.10.004>

Reducing the Risk of Basement Flooding Through Building- and Lot-Scale Flood Mitigation Approaches: Performance of Foundation Drainage Systems



B. Kaur, A. Binns, D. Sandink, B. Gharabaghi, and E. McBean

1 Introduction

Extreme, short-duration rainfall events resulting in overwhelmed private- and public-side drainage and wastewater systems cause substantial insured and uninsured damage each year. With respect to property damage, one of the most significant events to date was the July 8, 2013 extreme-rainfall related urban flood event in the Greater Toronto Area. Roughly half of the \$1 billion insured losses experienced during this event were directly attributed to flooded residential buildings [7]. Cities across Canada are dealing with similar issues resulting in severe and repeated flooding of residential buildings from stormwater, sewer backup, and infiltration flooding.

Residential buildings can be flooded through various mechanisms, including sewer surcharge, infiltration flooding and overland flooding. Sewer surcharge occurs when the capacity of the sewer system is exceeded with the result that raw sewage is forced to back-up into residential basements. Infiltration flooding occurs due to a combination of groundwater and moisture in the soil surrounding the home foundation which may enter the basement through the connection between the foundation wall and slab or by overwhelming the flow through the weeping tile system into the sump pit. Overland flooding occurs as a result of water flowing on the ground surface entering the basement through windows, doors and other fixtures [2].

Factors including urbanization and land-use, watershed characteristics and hydrology, municipal infrastructure and stormwater management may influence the

B. Kaur · A. Binns (✉) · B. Gharabaghi · E. McBean
School of Engineering, University of Guelph, Guelph, ON, Canada
e-mail: binns@uoguelph.ca

B. Kaur
e-mail: baljeet@uoguelph.ca

D. Sandink
Institute for Catastrophic Loss Reduction, Toronto, ON, Canada

risk of flooding for a region. The presence of effective flood management mitigation technologies implemented at the building- and lot-scale (or private-side) can reduce the risk of occurrences of basement flooding for a particular home. Prominent approaches include backwater valves to reduce the risk of sewer surcharge into basements, low impact development strategies to reduce the volume of runoff entering municipal infrastructure, and foundation drainage (or weeping tile) systems to prevent water from entering the basement through infiltration. While all of these types of measures offer improved resiliency to water-related disasters, their success is dependent upon consistent installation procedures and public awareness to ensure continued maintenance and performance. For example, backwater valves have been documented to be affected by solids deposition in the valve that inhibit their performance and reliability [3, 6]. The effectiveness of foundation drainage systems can be affected by foundation drainage system material selection, backfilling and lot drainage practices, and sump pump failure [5].

There is a growing need for research into private-side flood mitigation approaches to ensure their successful implementation and maintenance and to develop a better understanding of their performance and suitability under complex, site-specific conditions. The goal of this research is to investigate the role of the various complex inter-related factors that affect the risk of basement flooding in order to identify opportunities for improved adoption and success of private-side flood mitigation approaches. This research focuses on foundation drainage systems through exploring the design of a novel laboratory foundation drainage system to evaluate the effects of various lot-scale factors on the performance of foundation drainage systems. This experimental approach will lead to the development of an improved understanding of the performance of foundation drainage systems and assist in better understanding the nature of flood vulnerability for individual homes in order to better appoint strategies to reduce this risk.

2 Foundation Drainage Systems

2.1 *Infiltration Flooding*

Infiltration flooding occurs when groundwater accumulates outside the basement wall or foundation and enters the basement through cracks or defects on the wall or the floor. It can be a critical issue for older homes where mortar settles, and cracks develop in the foundations. In addition to older homes, some newer homes in circumstances where there is poor lot drainage or where the foundation drainage system is not adequately maintained, may also have greater chance of developing problems related to infiltration flooding [8]. Ensuring that the perimeter of the building is well-drained (with a foundation drainage (or weeping tile) system) can assist in reducing the risk of infiltration flooding [5, 9].

The inter-relationship between various lot-level, watershed and municipal factors can influence the risk of flooding for a particular residential basement. Lot-level factors such as lot-grading, watershed factors such as the groundwater level and municipal factors such as the type of sewer system (combined or separate storm and sanitary sewers), and the possibility of sewer surcharge causing backflow into gravity-drained foundation drainage systems, can significantly affect the amount of water accumulating near the foundation wall of a basement. Therefore, integrated measures are required to reduce moisture-related problems in basements. This research seeks to evaluate the performance of a foundation drainage system and the effect of various environmental and lot-specific factors on the effectiveness of the system. An overview of the foundation drainage system is provided in the next section.

2.2 Residential Foundation Drainage System Overview

A typical residential foundation drainage system consists of a weeping tile or foundation drain pipe, which is a continuous and perforated pipe installed along the perimeter of the basement footing and covered with gravel before soil is backfilled around the foundation. This system is designed to drain excess groundwater away from the foundation wall and footing to help reduce the risk of infiltration flooding. The National Building Code of Canada (NBC) has provided a set of minimum requirements for the installation of a foundation drainage system [1].

According to the NBC (2015), a drain pipe should not be less than 4" in diameter and should be laid on undisturbed or well-compacted soil. Generally, a flexible PVC drain pipe is used for the foundation system (Horizon Engineering 2021). The pipe should be covered with a layer of at least 6" of gravel on the top and around the sides of the drain pipe. The size of the gravel should be such that not more than 10% of the material passes through a 4 mm sieve. Following the placement of the gravel pack, the foundation wall should be backfilled with clean and free-draining material. It should be free of any deleterious debris and boulders larger than 250 mm.

According to the NBC (2015), the water received by the foundation drain can be drained into a sewer, ditch or drywell. A sump pit can be installed in the basement where the water can be collected and then disposed of, either through gravity flow or by using an automatic sump pump. The foundation drainage system can also receive water from a window well which can increase the amount of flow needed to be carried in the drain pipe.

The effectiveness of the foundation drainage system and the reasons of failure of the system are closely linked with lot-specific practises; thus, it is important to analyse the effect of these practises on the efficiency of the system.

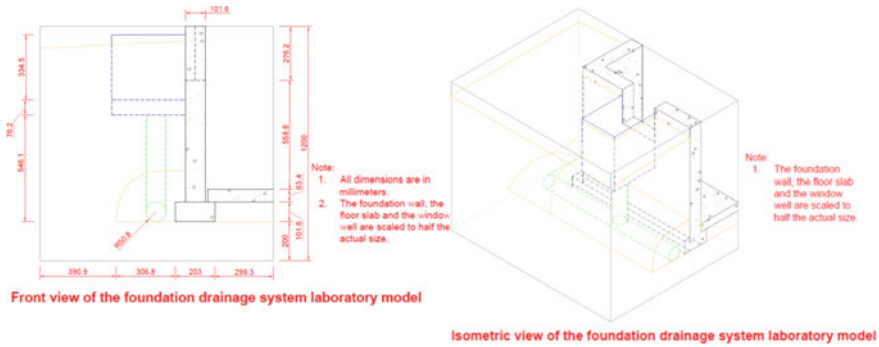


Fig. 1 Schematic of foundation drainage system model: front view (left); and isometric view (right)

3 Experimental Approach

3.1 Laboratory Model

In order to assess the performance of foundation drainage systems under varying lot-specific and hydrological conditions, a physical laboratory model of a residential foundation and drainage system is being designed and constructed. Laboratory models of private-side flood mitigation technologies have previously proven effective at developing a greater understanding of the performance of backwater valves through the ability to conduct controlled experimental testing to understand the influence of individual variables (see, e.g., [4]).

The model has been designed to simulate a residential foundation drainage system with a window well connection. The apparatus is shown in Fig. 1. The key components include: 1. A foundation wall, footing and floor slab to simulate a basement (scaled to half of the size of an actual basement); 2. A 4" PVC perforated and corrugated pipe laid around the footing; 3. A gravel pack around the drain pipe; 4. Soil representative of the backfill and native soil around the basement; 5. A window well connected to the drain pipe; 6. A rainfall and groundwater simulator; and 7. Collection tanks for sedimentation and recirculation. The model will be outfitted with additional measurement equipment and sensors, including flow measuring devices, pressure gauges and valves to control the flow.

3.2 Experimental Scenarios

The model will be used to evaluate the influence of a variety of factors that influence the risk of infiltration flooding and may affect the performance (and possible failure) of the foundation drainage system. The scenarios will be developed based

Table 1 Overview of experimental scenarios

| Rainfall intensity | Lot-grading | Groundwater level | Backfill material | % clogged |
|--|---|-------------------|--|-----------|
| Low (2-year return period storm) | Positive (sloped away from foundation) | Low | Silty-clay (fine texture) | 0 |
| Moderate (5/10-year return period storm) | Flat/no slope | Moderate | Loam or sandy-silt (medium to moderately-fine texture) | 50% |
| High (25-year return period storm) | Negative (sloped toward the foundation) | High | Coarse sand (coarse texture; free-draining material) | 75% |

on combinations of the following five factors: rainfall intensity, lot-grading, groundwater level, extent of clogging and type of backfill. Table 1 identifies the proposed conditions to be considered under these five factors.

Rainfall intensity ranging from a 2-year return period storm to a 25-year return period storm will be simulated to obtain an understanding of the role of the rate of rainfall impacting the lot. Consideration of storms of greater return periods may be considered in the future. Positive (sloped away from the foundation, as recommended), flat (no slope conditions) and negative (sloped toward the foundation) lot-grading conditions will be tested to assess the influence of lot-grading on the infiltration mechanisms and flow to the foundation drainage system. Various water table depths (i.e., groundwater level) will also be considered to gain an understanding of the sensitivity of infiltration flooding for homes in areas of high water table. The model is also constructed such that the backfill material may be replaced to consider the effect of particular soil conditions which result in various infiltration conditions and different potential for clogging of the foundation drainage system. The model will allow for the assessment of the effect of clogging of the foundation drainage system as fine material from the surrounding soil (e.g., clay and silt material) will clog the openings in the foundation drainage pipe with time, reducing the capacity of moisture from the surrounding soil to enter the drainage pipe. Thus, the time to clogging and the performance of the drainage system under varying clogging conditions will be evaluated using the model. The model can also consider the effect of settlement of the backfill material in the area adjacent to the foundation wall. This can produce a depression next to the foundation wall which can be a cause of water problems and basement flooding in older homes.

Measurements of the temporal response of the system to the various simulated conditions will be made to quantify the performance of the foundation drainage system. These measurements include assessment of the flow rate through the weeping tile system. The groundwater level change with time will also be monitored and measured to evaluate the localized groundwater response to precipitation events.

Clogging, and the build-up of fine material in the pipe, will be evaluated with the aid of video inspection of the foundation drainage pipe.

From these measurements, comparisons will be drawn between various scenarios and assist in determining how the performance of foundation drainage systems varies with time. Tests will also be conducted to assess the effectiveness of accessories like filter fabric on the lifespan of the weeping tile. Those lot-specific practises will be highlighted which contribute to increased clogging and thus failure of the foundation drainage system.

Results from the experimental research with the laboratory model will assist with calibration and validation for future numerical modeling of infiltration flooding of residential basements. This will enable assessment of longer-term effects and allow for consideration of more complex scenarios in order to produce greater insight into infiltration flooding mechanisms and the performance of foundation drainage systems.

4 Concluding Remarks

This research seeks to address the issue of infiltration flooding through the development of a novel experimental protocol which can be used to evaluate and test the performance of a residential foundation drainage system. These results will provide better understanding of the behaviour of a foundation system under various environmental and lot-specific scenarios and will highlight the reasons for the failure of the system. The results may also be used to determine the lifespan of a foundation pipe and maintenance intervals for the system. The knowledge attained by conducting these experiments will be extremely valuable for government officials, insurance companies and homeowners and can help in reducing the risk of infiltration flooding.

Acknowledgements This research was supported by the Institute for Catastrophic Loss Reduction (ICLR) and the Natural Sciences and Engineering Research Council of Canada (NSERC) through the Collaborative Research and Development grant program.

References

1. CCBFC, NRCC (2015) National building Code of Canada: 2015. National Research Council of Canada. Canadian commission on building and fire codes
2. CSA Group (2018) Z800-18: Basement flood risk reduction and protection guideline. Canadian Standards Association, Toronto
3. Dusolt S, Binns AD, McBean E, Gharabaghi B, Sandink D (2020) Characterization of backwater valves in sanitary sewer laterals and associated failures in a Canadian context. *Can J Civ Eng.* <https://doi.org/10.1139/cjce-2020-0026>

4. Dusolt S (2019) Evaluating the performance of backwater valves as a lot-level approach to reduce basement flooding in Canadian homes. PhD thesis. University of Guelph, Guelph, Canada
5. Horizon Engineering (2020) Builder guide to site & foundation drainage best practices for Part 9 buildings in British Columbia. Prepared for BC Housing. Horizon Engineering, Vancouver
6. Irwin S, Howlett C, Binns AD, Sandink D (2018) Mitigation of basement flooding due to sewer backup: overview and experimental investigation of backwater valve performance. *Nat Hazard Rev* 19(4):04018020
7. Robinson B, Sandink D (2021) Developing an efficient and cost-effective inflow and infiltration (I/I) reduction program. Institute for Catastrophic Loss Reduction/Standards Council of Canada, Toronto/Ottawa
8. Sandink D (2009) Handbook for reducing basement flooding. Institute for Catastrophic Loss Reduction, Toronto
9. Swinton MC, Kesik TJ (2008) Site grading and drainage to achieve high-performance basements. Institute for Research in Construction, National Research Council Canada

Predicting Navigability in the Lower Athabasca River System Through Numerical Modelling



S. Kashyap, E. Kerkhoven, Z. Islam, A. Petty, and S. Depoe

1 Introduction

The Athabasca River is the second largest river in Alberta. It has a length of about 1500 km from its headwaters at the Columbia Glacier to its outlet at Lake Athabasca, and drains an area of about 159,000 km², about 24% of Alberta's total land area. The Lower Athabasca River (LAR) extends over 200 km from Fort McMurray, and drains into the Peace-Athabasca Delta (PAD) (Fig. 1), recognized as an important habitat area for waterfowl and bison, and designated as one of the UNESCO World Heritage Sites. The LAR system, includes the main stem, its tributaries, and the waterways within the PAD, and has long been used for subsistence practice and transportation for Indigenous peoples, including the Athabasca Chipewyan First Nation (ACFN) and the Mikisew Cree First Nation (MCFN) [6]. The LAR system is a transportation corridor for accessing traditional lands, and for travelling between Fort Chipewyan and Fort McMurray [6]. Access is often needed to smaller side channels and adjoining tributaries for activities such as hunting, as animals such as moose tend to avoid banks facing the main stem LAR due to traffic and noise [6]. The PAD also contains many ancestral village settlements, and bountiful wildlife that are a critical source of subsistence [6].

Adequate flows, water levels and depths are required to ensure navigability of the River System. In the 1940s challenges to navigating the LAR were experienced, and a dredging channel was maintained up until 1996 [28]. Dredging was aimed in part at maintaining depth of 1.2 m [10, 15]. In recent years there have been additional

S. Kashyap (✉) · Z. Islam · A. Petty · S. Depoe
Alberta Environment and Parks, 3rd floor, South Petroleum Plaza, 9915-108 Street NW,
Edmonton, AB T5K 2G6, Canada
e-mail: shalini.kashyap@gov.ab.ca

E. Kerkhoven
Alberta Energy Regulator, 4999 98Ave NW, Edmonton, AB T6B2X3, Canada

© Canadian Society for Civil Engineering 2022
S. Walbridge et al. (eds.), *Proceedings of the Canadian Society of Civil Engineering Annual Conference 2021*, Lecture Notes in Civil Engineering 250,
https://doi.org/10.1007/978-981-19-1065-4_40

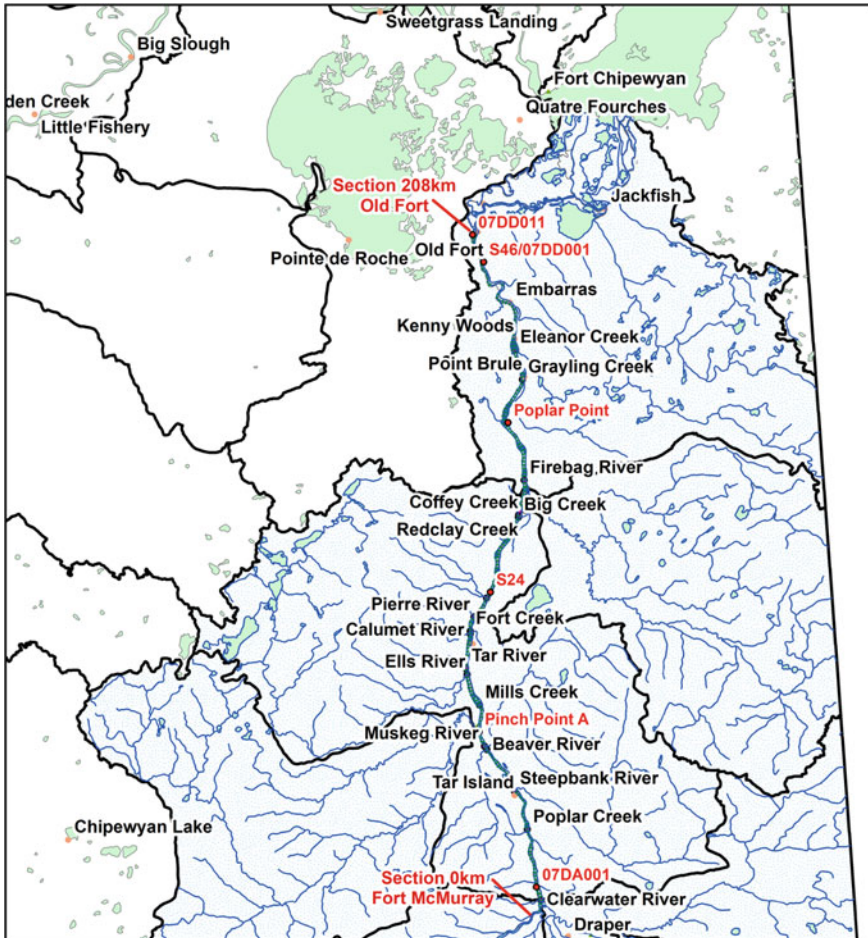


Fig. 1 Lower Athabasca River, with model domain extending from 0 km at Fort McMurray to 208 km at Old Fort. Gauging stations and pinch point locations (pinch point A u/s of Fort McKay and Poplar Point) are shown in red

concerns raised by the First Nations regarding possible impacts of mineable Oil Sands withdrawals and climate change on the water depths of LAR system [3, 8, 10].

While boats come in a variety of sizes and can navigate widely varying conditions [1], this study uses 1.2 m as the minimum navigational depth for the analysis. Candler et al. [6] identified a safe navigational depth of 1.2 m as the required depth for a fully loaded boat, after a successful hunt or for outfitting a trapping cabin, with an outboard motor. A depth of 1.2 m also corresponds to the historic dredging depth in portions of the Lower Athabasca River [10]. Other Indigenous communities and groups that use the Lower Athabasca River may have different minimum navigational depths.

In addition, a flow of approximately $1600 \text{ m}^3/\text{s}$ at 07DA001 was used for comparison when calculating the general probability (see Eq. (8)). This was cited in [6] as an initial threshold, subject to monitoring and refinement, for identifying where Treaty and aboriginal rights with regard to navigation, access and water level may be practiced fully along the LAR and adjoining tributaries [6]

2 Literature Review

Early work in numerical modelling of the LAR included the Northern River Basins Study (NRBS) [16], and the Mackenzie River Basin Hydraulic Model (MRBHM) [1, 18]. In these one dimensional (1D) river models, simplified rectangular cross-sections based on channel widths and bed elevations were used instead of actual river bathymetry. Other modelling attempts include a flow routing model called the Athabasca River Model [12–14, 22], although this model did not incorporate detailed bathymetry for the river, and the modelling focus was assessing water quality rather than navigation. Notable attempts using higher resolution detailed bathymetry for short reaches between Fort McMurray and the PAD include modelling studies using River1D and River2D (e.g., [26, 27], Northwest hydraulics Consultants Ltd. [4, 19, 20]). Recently, hydraulic models containing more continuous and detailed bathymetry have been completed using Mike11 and EFDC models from Fort McMurray to Old Fort (e.g., [17, 23, 24]). However, these studies focussed on sediment transport, water quality, and toxin transport, rather than navigation.

In 2008, a multi-stakeholder Phase 2 Framework Committee (P2FC) was established to provide recommendations for a Phase 2 Water Management Framework that would prescribe the timing and quantity of water that could be withdrawn from the LAR for cumulative Oil Sands mining water use [5]. For this, modelling work was completed using River2D for four reaches between Fort McMurray and the Athabasca Delta [3]. This modelling exercise was used to understand flow-depth relations for developing the preliminary Aboriginal Navigation Index (ANI). The ANI was developed to indicate navigation conditions from Fort Chipewyan to Fort McMurray. Subsequently, community based monitoring was completed to make updated recommendations [8]. One of the modelling studies that supported this work was the Athabasca Integrated River Model (AIRM) [8], that incorporated land-use, climate, and hydrologic changes. Transport Canada (Dillon [10], in consultation with Indigenous groups also completed 1D modelling using HEC-RAS to assess navigation conditions in the LAR system using surveyed data from 11 navigation reaches between Fort McMurray to Embarrass Portage. These studies enhanced the understanding of navigation conditions in LAR system. The intent of this current study is to predict various probabilities for navigating the LAR system between Old Fort and Fort McMurray (i.e. moving upstream) using a high resolution bathymetry dataset that has not been used previously in other studies focussing on navigation.

3 Methodology

3.1 Numerical Methods, Calibration and Validation

In this study we extended the work of Shakibaeinia et al. [23, 24] and developed a 1D MikeHydro hydrodynamic model extending about 208 km along the thalweg from Fort McMurray to Old Fort (see Fig. 1: Section 0 to Section 208 km). Notable improvements include: adding 20 tributary streamflows, re-analyzed bathymetry from [9], re-analyzing bed roughness, and a new calibration and validation for the period from January 1, 2016 to December 31, 2016. River cross-sections were extracted at approximately 1 km intervals along the channel with an interval of 20 m across the river using the 5 m by 5 m digital elevation model (DEM) data described in [9]. This DEM combined both bathymetry and topography data from Geoswath, Point cloud LiDAR, detailed surveyed sections, rectangular cross sections used in the MRBHM [21], and ADCP (acoustic Doppler current profiler) reaches.

Channel roughness was estimated by analysing the non-cohesive sediment data based on the sediment core samples (Table 1). The analysis suggested a bimodal mixture of sand and gravel upstream of Shott Island, and unimodal sand from Shott Island to Old Fort. The fining of the sediment was calculated by [25]:

$$\ln D = (-a_D)x + \ln D_0 \tag{1}$$

where: D = calculated D_{50} in km. a_D = abrasion coefficient in $\text{km}^{-1} = 0.0028$ [25] x = distance in km in the longitudinal direction.

The Manning’s n for the main channel was calculated using the 1923 Strickler formula [11], based on composite sediment sizes. The resulting Manning’s n ranged from 0.021 (Section 0 km) to 0.011 (Section 135 km), and then increased to 0.012 (Section 208 km).

The model was run under open water conditions throughout the year without including the ice cover and processes. This was deemed acceptable for this study because navigation is only of concern during under open water conditions. The downstream boundary discharge rating curve condition at Old Fort was based on the open-water relationship of the Water Surface Elevation (WSE) at Water Survey Canada (WSC) gauge at Old Fort (07DD011) versus the streamflow at RAMP (WSC)

Table 1 Median sediment sized determined from sediment core samples

| Point location Distance (m) from model u/s boundary | Sample, reference | Sand | Gravel |
|--|-------------------|---------------|---------------|
| | | D_{50} (mm) | D_{50} (mm) |
| 5414 | Sample 1, [7] | 0.250 | 16.000 |
| 5414 | Sample 2, [7] | 0.375 | 16.000 |
| 94,691 | Sample 8, [25] | 0.180 | |
| 168,417 | Sample 9, [25] | 0.250 | |

gauging stations at Embarrass Airport S46 (07DD001) expressed by the following equation [2].

$$WSE = 10^{((\log_{10} Q - \log_{10}(135.326))/1.647)} + 209.3 \tag{2}$$

The flow at the upstream boundary was determined from the flow at the WSC gauge at Fort McMurray (07DA001) minus the flow of Clearwater at Draper (WSC gauge 07CD001). Flows for the 20 tributaries were determined from a combination of flow calculated from a validated VIC hydrological model and WSC gauging station data [2].

Calibration (January 1, 2000 to December 31, 2007) and validation (January 1, 2008 to December 31, 2016) results are shown in Table 2 and Fig. 2, for WSC gauge 07DA001 and RAMP gauge S24 and S46. The final adjusted Manning’s N after calibration ranged from 0.021 to 0.040, and no parameters were adjusted after validation.

4 Analysis

The study objective is to predict, for a given river flow rate, the probabilistic accessible land area for a navigator travelling from the Athabasca River at the start of the Athabasca Delta (Fig. 1, near Old Fort, cross-Section 208 km) to any location in the LAR watershed until the WSC gauge 07DA001 (Fig. 1, cross-Section 7). For this analysis, median water depths were extracted from the model for the validation period, for a given flow rate $Q \pm 5\%$ at WSC gauge 07DA001.

One challenge in modelling navigation is that the navigator is an agent who synthesizes information from their environment to arrive at a probable best course of action for navigating the river. One approach could be to simply model thalweg depth and test if it remains above the navigability threshold. This assumes, however, that a navigator unerringly finds the thalweg in all cases, which is probably not realistic for even the most skilled boat pilots, and may create an unrealistically high expectation of river navigability, particularly for reaches that are close to the navigability threshold.

To model navigator agency in the selection of river channels, we used a simple probabilistic model where the probability of finding a 1.2 m depth scales with a power function that ranges from one, implying completely random selection, to infinity, implying the unrealistically perfect navigator described above.

First, consider the random navigator, moving upstream past one of the cross-sections, ‘i’, who arbitrarily selects a path along the wetted width of the river channel. For a flow rate at WSC gauge 07DA001, ‘Q’, the probability such a “random navigator” would find a path where the depth was more than 1.2 m is:

$$P(i, i, 1, Q) = \frac{\text{Number of nodes at cross - section } i \text{ where depth } > 1.2 \text{ m and flow at } 07DA001 \text{ is } Q}{\text{Number of nodes at cross - section } i \text{ where depth } > 0 \text{ m and flow at } 07DA001 \text{ is } Q} \tag{3}$$

Table 2 Calibration and validation statistics for water surface elevation (during Open Water Seasons, May 1–Oct. 31) and Flow from Jan. 1, 2000 to Dec. 31, 2016

| | WSE calibration | | | WSE validation | | | Flow calibration | | | Flow validation | | |
|----------------|-----------------|------|-----|----------------|------|------|------------------|------|-----|-----------------|-------|-------|
| | 07DA001 | S24 | S46 | 07DA001 | S24 | S46 | 07DA001 | S24 | S46 | 07DA001 | S24 | S46 |
| r | 0.97 | 0.99 | n/a | 0.99 | 0.99 | 0.99 | 0.97 | 0.97 | n/a | 0.99 | 0.98 | 0.99 |
| r ² | 0.94 | 0.97 | n/a | 0.98 | 0.98 | 0.98 | 0.95 | 0.95 | n/a | 0.99 | 0.96 | 0.97 |
| NSE | 0.91 | 0.92 | n/a | 0.94 | 0.94 | 0.94 | 0.95 | 0.95 | n/a | 0.99 | 0.93 | 0.97 |
| ME | 0.24 | 0.30 | n/a | 0.19 | 0.23 | 0.40 | -4.44 | 29.8 | n/a | -7.66 | -4.80 | 37.31 |
| MAE | 0.28 | 0.35 | n/a | 0.24 | 0.33 | 0.41 | 47.8 | 65.5 | n/a | 29.1 | 76.6 | 68.1 |

Where r = Pearson Correlation Coefficient, r² = Coefficient of Determination, NSE = Nash Sutcliffe Efficiency, ME = Mean Error, and MAE = Mean Absolute Error. ME for WSE (Flow) has units of m (m³/s), and MAE for WSE (Flow) has units of m (m³/s)

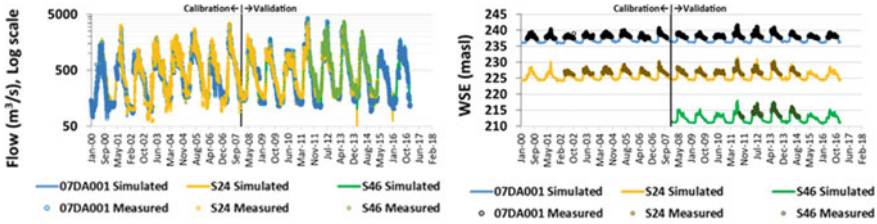


Fig. 2 Simulated and measured streamflow (left) and WSE (right) for calibration (January 1, 2000-December 31, 2007) and validation periods (January 1, 2008-December 31, 2016). *Note* for flow y-axis is a logarithmic scale; measured values for WSE are only shown for open water season (May 1-October 31) as the model does not simulate ice effects

Now consider a situation where the navigator is able to select paths such that their first attempt at cross-section ‘i’ would be equivalent to the deepest of ‘n’ randomly selected paths. For such a navigator, the probability of finding a path with depth of 1.2 m or more is:

$$P(i, i, n, Q) = 1 - [1 - P(i, i, 1, Q)]^n \tag{4}$$

A navigator where $n = 1$ would be the random navigator described above, while a navigator where $n = \infty$ would be the perfect navigator who always find a depth of 1.2 m, if such a depth exists. An alternative way of conceiving of n is the probability that a navigator would select the deepest 10% of a cross-section on their first attempt:

$$P_{10\%} = 1 - 0.9^n \tag{5}$$

For a navigator to have a 50% or greater chance of selecting the deepest 10% of a channel, $n \geq 6.58$. For our purposes here, we call a navigator with $n = 6.58$ the “reference navigator”, and we call the value ‘n’ the “navigator success rate”.

Next, consider a navigator attempting to pass all the cross-sections from ‘i’ to ‘j’, inclusive. The probability of this navigator successfully crossing all these cross-sections in a single attempt is given by the product of the probabilities at each cross-section:

$$P(i, j, n, Q) = \prod_{k=i}^j P(k, k, n, Q) \tag{6}$$

If the upstream starting cross-section is ‘m’, and the incremental drainage area between cross-section ‘j’ and ‘j + 1’ is $A_{inc}(j)$, then the effective accessible land area, ‘ A_{eff} ’, upstream of cross-section ‘i’ is given by, for a navigator success rate ‘n’ as:

$$A_{eff}(i, n, Q) = \sum_{j=i}^m A_{inc}(j)P(j, m, n, Q) \tag{7}$$

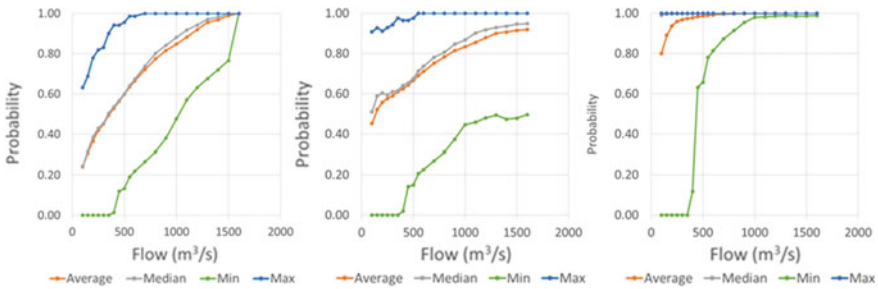


Fig. 3 Statistics for the following probabilities: depth > 1.2 m at any flow (left), the random navigator (n = 1) finds depth > 1.2 m randomly on first try (center), the reference navigator (n = 6.58) finds depth > 1.2 m on first try (right). These statistics are for cross-Sections 7 through to 208

For this paper, we assume that a single failure to find a depth of at least 1.2 m at any location would result in upstream areas being inaccessible. The possibility of a navigator making multiple attempts at a difficult cross-section is not included in this analysis, although it could be considered by increasing the navigator success rate. For example, a navigator with “first attempt” success rate of n , if they were willing or able to make two attempts at a difficult section, would have the same level of access as a single attempt as a navigator with success rate = $2n$.

We applied this model over a range of flows from 100 to 1600 m³/s that spanned the range where navigability is likely to be impacted., We then assessed the probability of a navigator finding a depth >1.2 m relative to 1600 m³/s flow. This probability is provided by the following equation and is shown in Fig. 3, left:

$$P(i, i, Q) = \frac{\text{Number of nodes at cross – section } i \text{ where depth } > 1.2 \text{ m and flow at 07DA001 is } Q}{\text{Number of nodes at cross – section } i \text{ where depth } > 1.2 \text{ m and flow at 07DA001 is } 1600 \text{ m}^3/\text{s}} \tag{8}$$

5 Results and Discussion

As we can see, minimum probabilities are zero up until 450 m³/s, and increase to 1 at 1600 m³/s. These results agree well with other navigation studies that flow should be greater than about 400-500 m³/s for navigation of the main stem, and that at 1600 m³/s, the river could be navigated across most success rates [3, 6, 8].

We then consider the effect of navigator success rate (see Fig. 3, center for success rate $n = 1$, right for success rate $n = 6.58$). We see that the random navigator ($n = 1$) reduces the minimum probabilities, and that the reference navigator ($n = 6.58$) will have a substantially increased minimum probability of finding a depth > 1.2 m.

As the navigator moves up the main stem, the probability of reaching the next successive cross-section decreases (Fig. 4). There are three reaches with similar

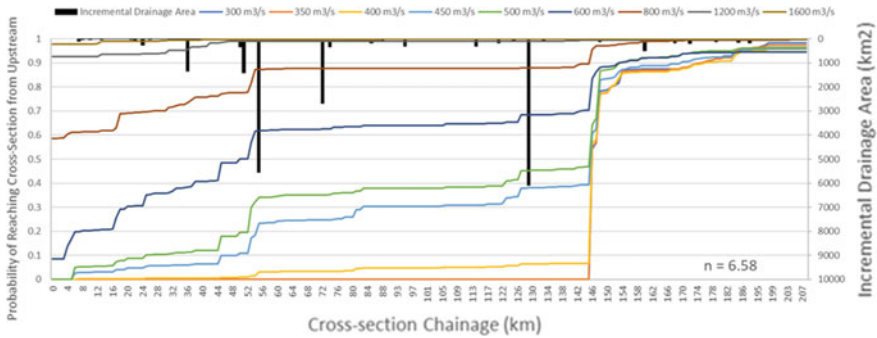


Fig. 4 Probability of reaching a given cross section from upstream (Section 208 km) to downstream (Section 0 km) for the reference navigator ($n = 6.58$). Incremental drainage areas for each cross section are shown on the top

probability slopes (Sections 208–146, Sections 54–146, and Sections 0–54). We see a dramatic decrease in probabilities between Sections 146–155, which corresponds to a navigation trouble area at Poplar Point, where sand dunes have been noted [6]. Similarly, we see another drop at approximately Section 54, which is just upstream of Fort MacKay, and has also been noted to be a navigation pinch point, where depths tend to be shallow [6].

Figure 5 shows the values of $A_{eff}(7, n, Q)$, that is for travel from the Athabasca Delta to Fort McMurray (WSC gauge 07DA001), for various flows and n values. It should be noted that the total drainage area between Sections 208 and 0 km is 22,550 km². This figure allows us to visualize the magnitude by which access to land area off the main channel can decrease with decreasing flow rates. It is interesting to see a dramatic reduction to land access at flows below 500 m³/s, which is consistent with

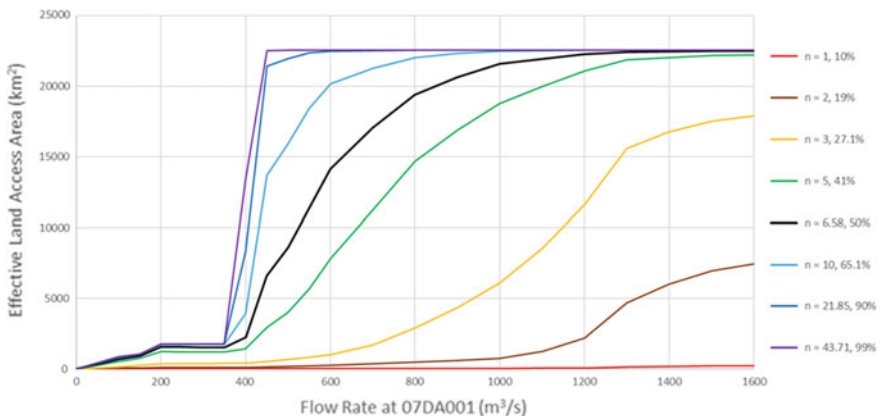


Fig. 5 Effective accessible land access area for various experience levels at the given flow rates at 07DA001

values reported from community based monitoring in the PAD [8]. Below 350 m³/s access is severely limited, primarily only to the lands downstream of the pinch point at 145 km which do not include any of the major tributaries to the lower Athabasca River. In addition, we see that at $n \geq 6.58$ all lands may be accessed at 1600 m³/s.

At this time, we have no quantitative information to calibrate the value of n for typical navigators on the Athabasca River. However, we believe that quantifying this value is not as important as establishing the relationship between navigator agency and flow level. It is interesting to note that a near perfect navigator, with a 90% or higher chance of selecting the deepest 10% of a given channel section ($n > 20$), the loss of access area for flows below 450 m³/s is dramatic while for n between 5 and 10 this loss is much more gradual, and can start as soon as flows drop below ~1300 m³/s. Although both groups of navigators could be expected to have full access at flows above 1600 m³/s and minimal access below 500–400 m³/s, as reported in [6], their respective experiences for flows between 400 and 1600 m³/s would be markedly different.

It should be noted that although the trends from this work broadly agree with what has been observed in the field [6], these analyses are only predictions and are based on output from a numerical model. Such outputs are subject to some uncertainty in both the numerical estimations, as well as the quality of the input data used to prepare the model. In particular, there were uncertainties in reference elevations for some ADCP reaches included in the bathymetry DEM, and at times best judgement calls and analysis were used when setting these.

A future analysis of probabilities in the PAD region between Old Fort and Fort Chipewyan, would be of great interest, and would provide insight into the relationship between flow and access to a region of immense importance for Indigenous communities.

6 Conclusions

We analyzed flow depths output from a 1D Mike Hydro numerical model to predict navigability of the main stem of the LAR (between Old Fort and Fort McMurray) based on a stochastic model of navigator agency at various flow rates at the WSC gauge 07DA001 at Fort McMurray. We considered flow rates at which access to the land and tributaries in the basin would be limited if a depth of 1.2 m is needed to navigate. We found that both flow rates and navigator success rate were important to consider when determining navigability of the river. At flow rates less than or equal to 450 m³/s, it is likely depths at some sections along the main stem would be ≤ 1.2 m. Above this flow however, we found a navigator who found the deepest 10% of the Section 50% of the time (the reference navigator) should not have trouble navigating the river at 1600 m³/s. Two pinch points found from the model (just upstream of Fort McKay and Poplar Point) agreed with what were found to be navigation trouble spots in the field. We also found that at flow rates below 500 m³/s, access to lands off the main stem may be limited substantially. Future analysis on the Peace Athabasca

Delta would provide valuable insight into flow conditions suitable for navigation in another region of critical importance to Indigenous communities.

Acknowledgements The authors would like to acknowledge Environment and Climate Change Canada, particularly Ahmad Shakibaeinia and Yonas Dibike who played a fundamental role in the development of the original Mike11 model. In addition, George Wang, Babak Fajid, Anil Gupta and DSI LLC, who prepared input data some model boundary conditions.

References

1. AECOM Canada Ltd. (2009) Review of water management alternatives on water depth in the Lower Athabasca River. Working Group, Surface Water Working Group, Contract #2009–009, CEMA
2. Alberta Environment and Parks (AEP) (2019) Personnel communication
3. Alberta Environment and Parks (AEP) (2015) Lower Athabasca region, surface water quantity management framework for the Lower Athabasca River. Alberta Government
4. Andrishak R, Abarca JN, Wojtowicz A, Hicks F (2008) Freeze-up study on the Lower Athabasca River (Alberta, Canada). In: Proceedings of the 19th IAHR international symposium on ice, Vancouver, BC, July 6–11
5. Carver M, Candler C, Boag T (2010) Athabasca Chipewyan First Nation (ACFN) Industry Relations Corporation, Mikisew Cree First Nation (MSFN) Government Industry Relations. In: Athabasca Chipewyan first nation and Mikisew Cree first nation review of the phase 2 framework committee recommendations: synthesis report
6. Candler C, Olson R, DeRoy S and the Firelight Group Research Cooperative with ACFN and MCFN (2010) As long as the rivers flow—Athabasca river knowledge, use and change. Parkland Institute, University of Alberta
7. Carson MA (1990) Evaluation of sediment transport data for the Lower Athabasca River Basin, Alberta. Report for the Inland Waters Directorate, Environment Canada
8. Carver M (2018) Assessment of impacts of water quantity, indigenous access and traditional use due to Teck's Frontier mine. Submission to the Joint review Panel for the Frontier Mine, Prepared for: Mikisew Cree First Nation, Aqua Environmental Associates
9. Chowdhury EH (2017) Extending a high resolution DEM for the Lower Athabasca River from Firebag River confluence to old fort. Alberta Environment and Parks
10. Dillon Consulting (2019) Athabasca river navigational study—Phases 3 and 4, Final Report, Transport Canada
11. Garde RJ, Raju KGR (2006) Mechanics of sediment transportation and alluvial stream problems, 3rd edn. New Age International Publishers, Delhi, India
12. Golder (2003) Athabasca river operational discharges (Phase I). Submitted to CONRAD Environmental Research Group. Fort McMurray
13. Golder (1997) Preliminary investigation of potential water release scenarios. Report prepared for Syncrude Canada Ltd
14. Golder (1996) Athabasca river water releases impact assessment. Prepared for Suncor Oil Sands Group. Fort McMurray
15. Gorecki RC (1990) Geomorphology and sedimentology of the lower delta plain and subaqueous delta plain of the Athabasca River Delta, N.E. Alberta, MSc. Thesis. Department of Geography, University of Calgary, Calgary, AB
16. Government of Alberta (1996) Northern River Basins Study Final Report, Report to the Ministers
17. Kashyap S, Dibike Y, Shakibaeinia A, Prowse T, Droppo I (2017) Two-dimensional numerical modelling of sediment and chemical constituent transport within the lower reaches of the Athabasca River. Environ Sci Pollut Res. <https://doi.org/10.1007/s11356-016-7931-3>

18. Kouwen N (2016) Watflood—model setup for MRBHM update. Government of the NWT Contract NO. 84
19. Northwest Hydraulics Consultants Ltd. (2007a) Lower Athabasca River habitat surveys—2007 winter water levels and rating curves, cumulative environmental management association, Project 1–6890, Contract Number 2007–0007
20. Northwest Hydraulics Consultants Ltd. (2007b) Lower Athabasca River habitat surveys—2007 winter flow simulations at poplar point (Reach #3). Cumulative Environmental Management Association, Project Number 1–6890, Contract Number 2007–0007
21. Pietroniro A, Hicks F, Andrishak R, Watson D, Boudreau P, and Kouwen N (2011) Hydraulic Routing of Flows for the Mackenzie River. Report submitted to the Mackenzie River Basin Board (Ed. Jesse Jasper), 20 September 2011
22. Rosner T (2005) Athabasca River model update and reach segmentation, submitted to cumulative environmental management association. Golder Associates
23. Shakibaeinia A, Kashyap S, Dibike Y, Prowse T (2016a) An integrated numerical framework for water quality modelling in cold-region rivers: a case of the Lower Athabasca River. *Sci. Total Environ* 634–646
24. Shakibaeinia A, Dibike YB, Prowse T (2016b) A numerical framework for modelling sediment and chemical constituents transport in the lower Athabasca River. *J Soils Sediments*. <https://doi.org/10.1007/s11368-016-1601-4>
25. Shaw J, Kellerhals R (1982) The composition of recent Alluvial gravels in Alberta River Beds, Bulletin 41. Alberta Research Council
26. Trillium (2004) Open Water Survey of Athabasca River at Bitumount (Reach #4). CEMA Lower Athabasca River Habitat Surveys. Trillium Engineering and Hydrographics Inc. Project Number 04–568
27. Trillium (2005) Flow simulations and fish habitat evaluation for the Athabasca River at Bitumount (Reach #4). Cumulative Environmental Management Association, Registration
28. Westland (2009) Impact hypothesis analysis: effects of water withdrawal on traditional use of the Lower Athabasca River. Report prepared for the Cumulative Environmental Management Association, Contract #2009–0026, Traditional Land Use Study of the Lower Athabasca River – Phase 2, WRG Westland Resource Group Inc

Water Distribution System Leak Detection Using Support Vector Machines



Z. Cai, R. Dziedzic, and S. S. Li

1 Introduction

Water loss in water distribution systems is a growing concern for the water utilities, as it results in the waste of water and unnecessary costs. The most common source of water loss is leakage. Leaks can contribute up to 70% of water loss, with higher percentages expected in undermanaged networks [9]. Therefore, exploring effective approaches for identifying whether leaks in water distribution system is essential.

Acoustic detection remains the primary means of detecting and locating pipeline leaks. These types of methods use mobile acoustic sensors to monitor leak sounds since water leaking from a broken pipe makes a hissing or whooshing sound that can be transmitted through the pipe. However, the use of the sensors can be hampered by local environment and the presence of engineered structures. Sound at the surface can be muffled by deep soil, thick roadway pavement, or heavy local traffic. As a result, there are many uncertainties leading to the low accuracy of leak detection while using this method. Other non-acoustic methods include thermographic methods, tracer gas detection, radiographic methods, ultrasonic and electromagnetic methods and remote sensing. However, these are not feasible for long reaches of water distribution systems, since they are generally user-dependent, slow in progress, influenced by pipe material and labour intensive [12].

With the development of advanced machine learning models and supervisory control and data acquisition (SCADA) systems, real-time pressure and flow measurements can now be used to predict if there is a leak in the water distribution system. Liu et al. [5] applied acoustic sensors and support vector machines (SVM) to detect water pipeline leakage. Coelho et al. [3] used flow sensor measurements and compared the

Z. Cai (✉) · R. Dziedzic · S. S. Li

Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, QC 3G 1M8, Canada

e-mail: caiziyuanca@163.com

© Canadian Society for Civil Engineering 2022

S. Walbridge et al. (eds.), *Proceedings of the Canadian Society of Civil Engineering*

Annual Conference 2021, Lecture Notes in Civil Engineering 250,

https://doi.org/10.1007/978-981-19-1065-4_41

application of different classification methods, specifically random forest, decision trees, neural networks, and SVM. Other studies have used a combination of pressure and flow data as well as hydraulic model calibration and other machine learning methods, such as genetic algorithms [7]. The present study proposes the application of SVMs trained based on pressure and flow measurements from SCADA system, and historic leakage information.

2 Methods

2.1 Support Vector Machine

Support vector machines (SVMs) are a supervised machine learning algorithm, mostly applied to classification problems. The main objective of SVMs is to seek the most suitable separating hyperplane that classifies all data points in a given training set into two classes [13].

A training set is defined as $\{(x_i, y_i): i = 1, 2, \dots, N\}$ containing a large number of training data points, $x_i \in \mathbb{R}^n$ and $y_i \in \{1, -1\}$ [6]. In this definition, x_i are training examples and y_i are the classes separated by a hyperplane. The hyperplane found by a SVM not only separates positive and negative classes by placing all the positive examples on one side of the hyperplane and negative on the other, but also maximizes the distances, i.e. margin, between the nearest training examples and the hyperplane. The SVM is represented by Eqs. 1 and 2, as follows.

$$g(x) = \sum_{i=1}^{L_S} a_i d_i K(x_i, x') + a_0 \quad (1)$$

where L_S is the number of support vectors; x' are support vectors; d_i is the class indicator that is linked to each x' ; a_i are constraints determined from training examples; $K(x_i, x')$ represents kernel function associated with the kernel selected in this study. subject to:

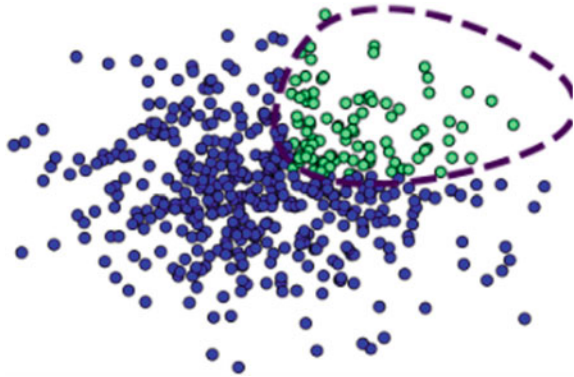
$$y_i(\omega_T X_i + b) \geq 1 \quad (2)$$

where ω and b are parameters of a hyperplane in feature space.

In a 2-dimensional space, training sets containing non-linear training examples cannot be separated by a hyperplane, since this hyperplane is a straight line in this space. In this case, the kernel trick can be applied. The SVM kernel takes low dimensional input space and transforms it to a higher dimensional space so that non-separable training sets can be converted into separable training sets [14]. In the present study, the radial basis function (RBF) kernel is used.

Figure 1 shows an example hyperplane for a training set containing a large number of non-linear training examples denoted by points in the two-dimensional space. A linear hyperplane cannot segregate the two classes, blue and green points. In this

Fig. 1 Illustrative hyperplane shown as the purple dashed line classifying blue and green data points



case, the SVM applies the RBF kernel and adds a new dimension, converting two-dimensional space into three-dimensional space. As a result, the gap between blue and green classes can be observed in three-dimensional space and the SVM can seek to find a hyperplane in this gap as shown by the purple dashed line in Fig. 1.

When an SVM is applied to a classification challenge, it is trained on a large number of training examples (x_i) to fit the classification model. Afterwards, testing examples can be fed into the trained SVM to separate them into two different classes ($y_i = 1$ and $y_i = -1$). Finally, the testing accuracy is evaluated with approaches such as Receiver Operating Characteristic (ROC) curve and cross validation. The higher the classification accuracy, the more reliable the utilisation of an SVM to segregate two classes.

2.2 SVMs Applied to Leak Detection

Due to the occurrence of leakage, pressure and flow measurements next to leaks are expected to become abnormal. The goal of the SVM is to identify these abnormalities. In order to train the SVM, previous SCADA with pressure and flow measurements must be related to historical leakage data, containing the time and location of the leak. That way the SVMs can recognize whether a pipe network has leaks in different areas and at different times based on given pressure and flow measurements.

In the present study, code was developed in Python to organize the available SCADA data and leakage data, develop a SVM to detect leaks, and test the accuracy of the SVM. Figure 2 presents the proposed algorithm for developing a leak detection SVM. First, all needed Python packages, such as pandas, numpy, and SVM packages, are imported. Pressure and flow measurements are taken in SCADA and loaded for analysis. They are organized into the x training data frame, as needed for the SVM. This data should show for each timestep of SCADA measurement, the node ID of the pressure, flow, or level sensor, and the measurement at that time. Leakage data, on the other hand, must be organized for the y training data frame. This should include

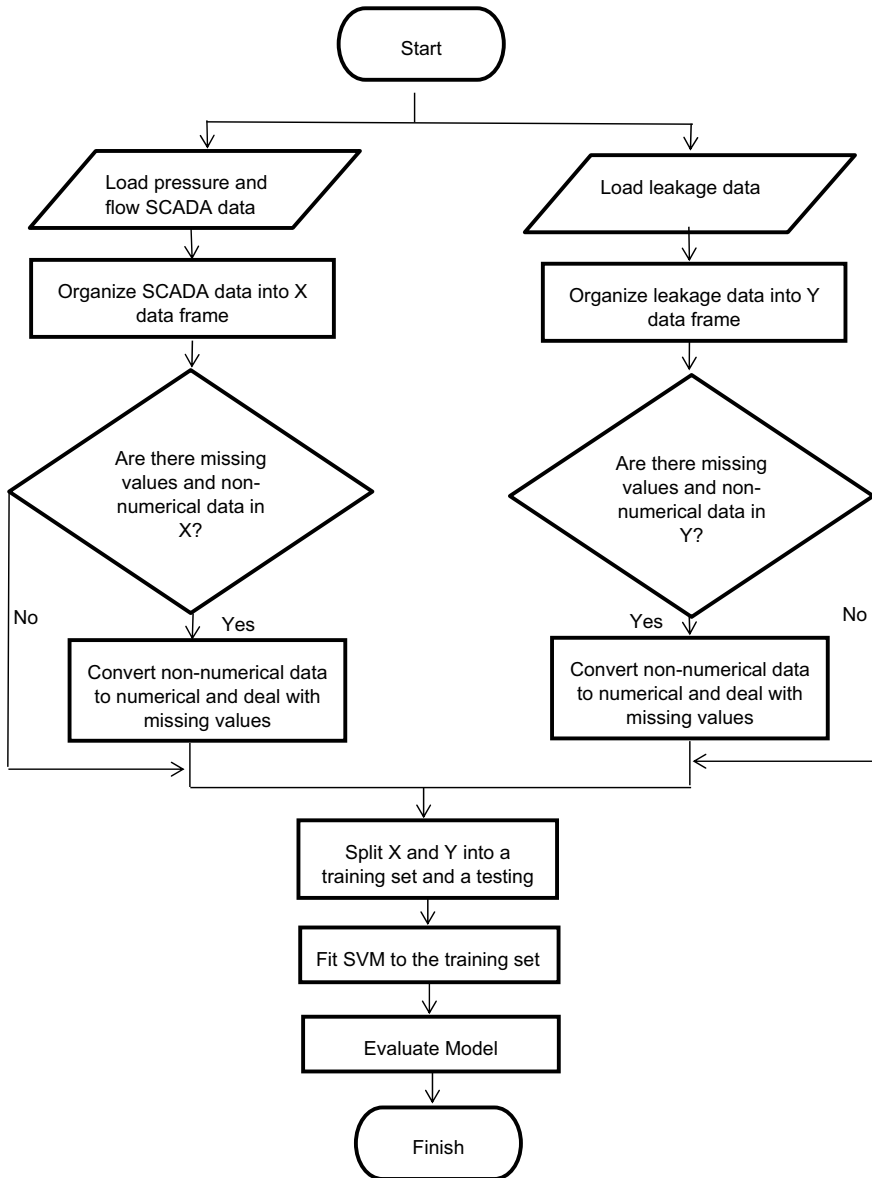


Fig. 2 Algorithm for developing a leak detection SVM

the same timesteps as defined in x and a binary classifier indicating if there was a leak. For each table, non-numerical values are then converted to numerical. Missing values are replaced by the node average.

3 Results

3.1 Case Study System Description

The proposed SVM leak detection method is applied herein to the L-Town hypothetical system. The system and its data were developed for “The Battle of the Leakage Detection and Isolation Methods 2020” [10]. A number of previous studies have proposed different approaches for detecting leaks in this system since it facilitates comparisons [1, 2, 4, 7, 8, 11].

The network is composed of 782 junctions and 905 pipes, as depicted in Fig. 3. The water distribution system receives water from two reservoirs and is divided into three areas, A, B and C. Area B is located in the lower center of the town while Area C is situated in the top left. Pressure Reduction valves (PRVs) are installed downstream of the two main reservoirs to help regulate the pressure and a PRV is at

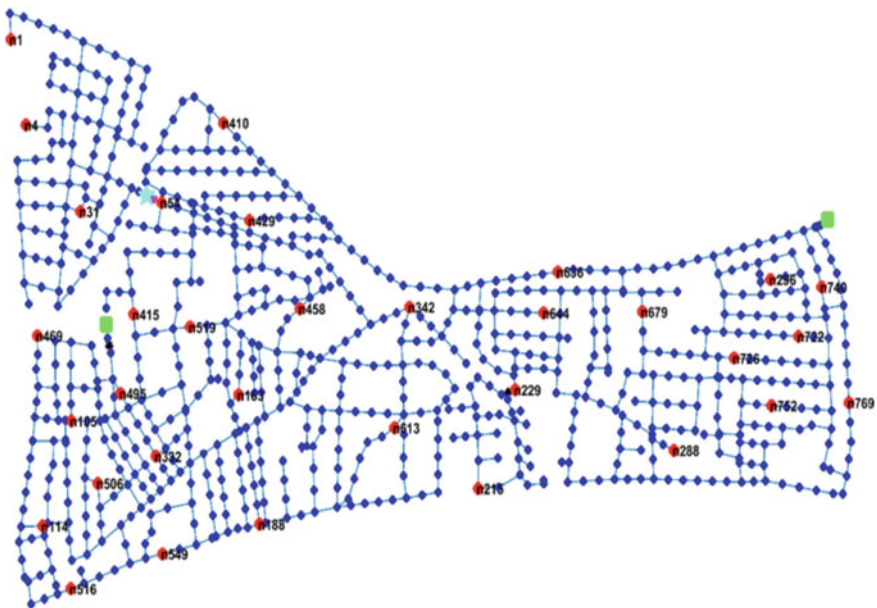


Fig. 3 The L-Town water distribution network, where red nodes (●) indicate pressure sensors, green squares (■) reservoirs, and blue stars (★) tank level sensors, blue dots (●) represent junctions, blue line segments (—) represent pipes

Area B to reduce background leakages. The network is monitored by one tank water level sensor, three flow sensors, and 33 pressure sensors (red nodes) highlighted in Fig. 3. All sensors transmit their measurements every 5 min to the utility's SCADA system. Data is available for the entire year of 2018.

Thus, for each timestep there are 36 measurements of either pressure or flow at different locations. This data was organized into the input X data frame, as shown below in Table 1. Column names are network node IDs. Identifiers beginning with "n" are pressure sensors and the values in their respective columns are of pressure head (m). Nodes beginning with "P" are pumps, to which flow sensors are attached. Values in these columns are flow in m^3/h .

Similarly, the historical information about when and where leakage occurred is represented as another data frame, Y, shown in Table 2. For each timestep a binary leaking indicator was created based on historical leakage information. The indicator is 0 when no leakage was occurring at the time, and 1 when leakage was occurring.

3.2 SVM Results

After loading and cleaning the data, the X and Y data frames were split into a training set containing 80% of data points, and a testing set, containing 20%. A SVM with a RBF kernel was applied to training set, and evaluated with cross validation and ROC. Results of the predicted leaking indicators at each time step throughout the year of available data are shown in Fig. 4.

Cross validation was used to randomly select training and testing sets and evaluate model accuracy based on different splits. In this case, data was split into training and testing five consecutive times. The resulting testing accuracy was found to be 0.977. In order to further evaluate the classification model, a ROC curve was plotted. ROC compares a model's true positive rate (TPR) and false positive rate (FPR), as shown in Fig. 5. If a model is perfectly accurate, then the area under the ROC curve is 1. In this case, the area, i.e., the testing accuracy is 0.891. Thus, both accuracy results measured by cross validation and ROC curve are high.

4 Conclusion

This study applied SVMs to detect leakage in a hypothetical water distribution system. A similar approach can be applied to other networks. The main objective of the SVM is to find a hyperplane separating two classes, leaking and non-leaking. The SVM was trained based on SCADA pressure and flow data, as well as historical leakage data. Resulting accuracies were high. However, multiple steps still need to be taken to improve this method and make it more useful. The proposed algorithm has only identified if there is a leak in the system, not its specific location. Thus, next steps in the study include organizing the data to find areas of influence of each

Table 1 Organized X data frame combining all pressure and flow values at each time

| Sequence number | Time | n1 (m) | n4 (m) | n31 (m) | n54 (m) | ... | n769 (m) | P227 (m ³ /h) | P235 (m ³ /h) | PUMP_1 (m ³ /h) |
|-----------------|---------------------|-----------|-----------|------------|------------|-----|-------------|-----------------------------|-----------------------------|-------------------------------|
| 0 | 2018-01-01 00:00:00 | 28.92 | 33.87 | 37.13 | 37.13 | ... | 48.50 | 77.77 | 83.93 | 44.4 |
| 1 | 2018-01-01 00:05:00 | 28.94 | 33.89 | 37.16 | 37.23 | ... | 48.54 | 72.51 | 76.34 | 44.6 |
| 2 | 2018-01-01 00:10:00 | 28.96 | 33.91 | 37.18 | 37.21 | ... | 48.56 | 71.54 | 78.68 | 44.6 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 105,119 | 2018-12-31 23:55:00 | 28.0.63 | 33.68 | 36.99 | 37.00 | ... | 48.42 | 84.53 | 86.91 | 44.4 |

Table 2 Leaking status at various times from data sorting implemented by using codes

| Sequence number | Time | Leaking indicator |
|-----------------|---------------------|-------------------|
| 0 | 2018-01-01 00:00:00 | 0 |
| 1 | 2018-01-01 00:05:00 | 0 |
| 2 | 2018-01-01 00:10:00 | 0 |
| ... | ... | ... |
| 105119 | 2018-12-31 23:55:00 | 1 |

Fig. 4 Predicted leaking indicators at different times for L-Town using the SVM leak detection model

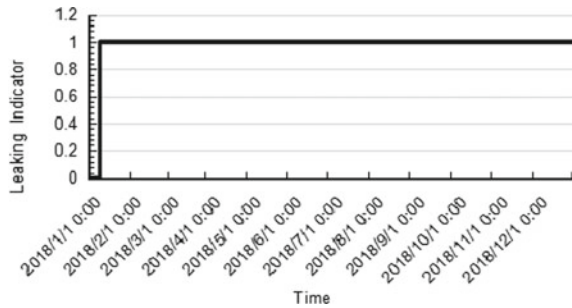
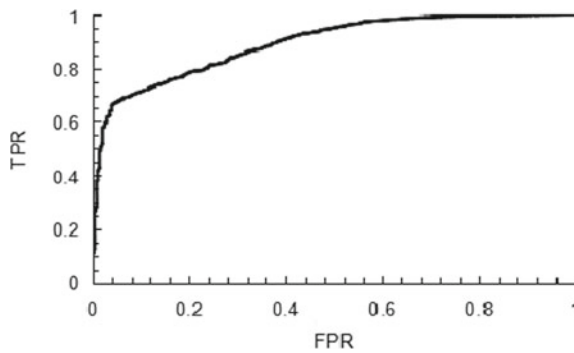


Fig. 5 ROC curve for SVM leak detection model



leak and sensors within those areas. Furthermore, missing sensor measurements were replaced by average yearly pressure measurements. This a simplification that does not account for varying daily and yearly pressures. So, the model will also be improved by better cleaning the data, and replacing missing values based on historical patterns and latest values. Lastly hydraulic models and SCADA based calibration could also be applied to identify baseline pressure and flow values. This could further improve the usefulness of the predictions, specifically for identifying background leaks that can last for months.

Acknowledgements This study received financial support from Concordia University through Faculty Research Support by R. Dziejdzic and S.S. Li.

References

1. Cheng T, Li Y, Harrou F, Sun Y, Gao J, Leiknes T (2020) A hybrid leakage detection and isolation approach based on ensemble multivariate changepoint detection methods. Zenodo. <https://doi.org/10.5281/zenodo.3964167>
2. Daniel I, Pesantez J, Letzger S, Khaksar Fasaee MA, Alghamdi F, Mahinthakumar K, Berglund E, Cominola A (2020) A high-resolution pressure-driven method for leakage identification and localization in water distribution networks. Zenodo. <https://doi.org/10.5281/zenodo.3924632>
3. Coelho JA, Gloria A, Sebastiao P (2020) Precise water leak detection using machine learning and real-time sensor data. *IoT*, **1**(2):474–493
4. Huang L, Du K, Guan M, Wang Q (2020) The combined usage of the hydraulic model calibration residual and an improved vectorial angle method for solving the battledim problem. Zenodo. <https://doi.org/10.5281/zenodo.3925507>
5. Liu Y, Ma X, Li Y, Tie Y, Zhang Y, Gao J (2019) Water pipeline leakage detection based on machine learning and wireless sensor network. *Sensors* **19**(23):5086
6. Mashford J, Silva DD, Burn S, Marney D (2012) Leak detection in simulated water pipe networks using SVM. *Appl Artif Intell* **26**:429–444
7. Saldarriaga J, Solarte L, Salcedo C, Montes C, Martínez L, González M, Cuello M, Ariza A, Galindo C, Ortiz N, Gómez C, Vanegas S (2020) Battle of the leakage detection and isolation methods: an energy method analysis using genetic algorithms. Zenodo. <https://doi.org/10.5281/zenodo.3924536>
8. Steffelbauer DB, Deuerlein J, Gilbert D, Piller OAND, Abraham E (2020) A Dual model for leak detection and localization. Zenodo. <https://doi.org/10.5281/zenodo.3923907>
9. Van Zyl JE, Clayton CRI (2007) The effect of pressure on leakage in water distribution systems. *Proc Inst Civil Eng Water Manage* **160**(2):109–114
10. Vrachimis S, Eliades D, Taormina T, Ostfeld A, Kapelan Z, Liu S, Kyriakou M, Pavlou P, Qiu M, Polycarpou M (2019) Battle of the leakage detection and isolation methods (BattLeDIM 2020). <http://battledim.ucy.ac.cy> (accessed 5th Sept 2020)
11. Wang X, Li J, Yu X, Ma Z, Huang Y (2020) A multistage approach to detect and isolate multiple leakages in district metering areas in water distribution systems. Zenodo. <https://doi.org/10.5281/zenodo.3924109>
12. Zaman D, Tiwari MK, Gupta AK, Sen D (2020) A review of leakage detection strategies for pressurised pipeline in steady-state. *Eng Fail Anal* **109**:104264
13. Zoltan C (2018) SVM and kernel SVM. *Towards data science*. <http://towardsdatascience.com/svm-and-kernel-svm-fed02bef1200> (accessed 12th Nov 2020)
14. Zhang G (2018) What Is the kernel trick? Why is it important? Medium. <https://medium.com/@zxr.nju/what-is-the-kernel-trick-why-is-it-important-98a98db0961d> (accessed 26th Nov 2020)

Comparison of Machine Learning Classifiers for Predicting Water Main Failure



M. Amini and R. Dziedzic

1 Introduction

Water networks are among the most essential infrastructure worldwide, as they convey potable water to billions of end-users. Water distribution networks are reported to make up approximately 80% of total expenditures associated with the water industry [13]. According to [3], approximately 240,000 incidents occur annually in the US, leading to around 1\$ trillion in rehabilitation backlog required to improve the condition of water network components. A recent 2018 study found that 16% of installed pipes were beyond their useful life whereas a similar study in 2012 found 8% were beyond their useful life, and many utilities are lacking adequate fund to replace all of them [24, 25]. Moreover, in US and Canada, overall water main failure is reported to have surged between 2012 and 2018, from 11 to 14 Failures/year/100 mile, respectively [24]. It was also reported that the rate of failure for Cast Iron (CI) and Asbestos Cement (AC) pipes increased by 40% during the aforementioned 6-year period. It should be noted that these two types of pipes account for almost 41% of all installed pipe in US and Canada. Furthermore, in 2017, the ASCE report card was prepared and given grade D to drinking water infrastructure in the USA as opposed to D- in 2009. Canada Infrastructure Report Card (2016) also reported that 29% of potable water infrastructures in very poor, poor or fair condition with a cost of \$60 billion to replace. This is comparable to the 25% found in the latest 2019 report (“Canada Infrastructure Report Card” 2019).

In Canada, 59% of pipes were reported to be less than 40 years old and only 9% above 80 (“Canada Infrastructure Report Card” 2019). However, if reinvestment is

M. Amini · R. Dziedzic (✉)
Concordia University, Montreal, QC, Canada
e-mail: rebecca.dziedzic@concordia.ca

M. Amini
e-mail: mohammad.amini@concordia.ca

not increased in Canada, the condition of core infrastructure may worsen, increasing the cost and risk of service interruption (“Canada Infrastructure Report Card” 2016). Mirza [20] found that Canada had an estimated \$123 billion total infrastructure backlog, \$31 billion of which was related to water and wastewater networks. In 2016, the Canada Infrastructure Report Card reported that 24.2\$ billion is needed to maintain the water network in Canada. Water main deterioration may lead to service interruption, decrease in hydraulic capacity in the network and declining the quality of water flowing within the network [13]. Confronting these consequences, water network agencies are striving to develop new strategies to tackle the challenges pertinent to this important infrastructure. This highlights the importance of predictive models to plan and enhance the more efficient rehabilitation/maintenance operations [8].

In recent decades, many studies have been conducted in order to find an appropriate method to assess the condition of water distribution networks [9]. A variety of physical and statistical models, as well as data-driven and machine learning algorithms have been utilized to predict the deterioration process of water main. For instance, Artificial Neural Networks [1, 12], Gradient Boosting Algorithm (Snider and [23] and Random Forest [22] have been used to evaluate the condition of water networks. Physical models, on the other hand, are more comprehensive, however, acquiring data required for these models may be costly, therefore these models are justifiable only for transmission networks, where cost of failure is significant [9, 13]. Statistical models, however, employ historical records to recognize an explicit failure patterns, and then utilize these patterns to predict the probability or rate of failure in the future [13]. These models can link the finding pattern to the pipe features such as age, diameter, material, etc. [9]. This study focuses on three types of classification models: Decision Tree, Random Forest, and Logistic Regression.

2 Literature Review

Since the deterioration of water mains is an intricate process, attempts to forecast the failure of water pipes focus primarily on statistical models [14]. Statistical models utilize historical failure data in order to define patterns that are assumed to continue in the future. Kleiner and Rajani [13] categorized these models to deterministic and probabilistic models. Using different attributes associated with water pipes, these models may estimate the probability of failure, rate of failure, and age at first and subsequent failures [13, 16]. Thus, comprehensive data would, undoubtedly, increase the accuracy of the models [13]. Nevertheless, the development process of such models for evaluating the condition of water mains is quite complex since failures typically occur as a result of different independent variables [7].

The present study focuses on statistical deterministic models, using machine learning classifiers. Deterministic models predict specific rate of failure, or pipe age at failure based on historical failure rates [13]. They even can predict whether a pipe failed or not by associating them to machine learning models. These models

require pipes to be partitioned into homogeneous groups that have similar characteristics. Such features could be material, size, soil characteristics and pipe vintage. This partitioning, however, imposes a challenge on the analysis process. That is, creating homogeneous groups may lead to unduly small groups. Simultaneously, these groups should be large enough for the statistical analysis to be reliable [13].

There are a variety of machine learning classifiers, among which Decision Tree, Random Forest and Logistic Regression are employed in this study. The application of these models in previous studies are briefly described below.

2.1 Decision Tree

Decision Tree (DT) can be used either for classification or regression problems. Harvey and McBean [10] applied DT to the prediction of sewer pipe failure in Guelph, ON, and compared it to Support Vector Machines. In this case, DT showed a 77% higher accuracy. Syachrani et al. [26] also compared this type of model to regression and neural networks. Again DT outperformed other methods in terms of accuracy. However, DT has not been applied frequently for water distribution failure prediction, being more common for sewer systems [15]. A DT model forecasts target labels with application of some predictive rules that are shaped in a structure similar to a tree [26]. Rule building initiates from the root of the tree where the dataset is assigned. This process continues by roots split into branches, which are known as decision nodes. This splitting process will not typically stop until the detection of only one class in a node which is called leaf. Figure 1 depicts the concept of the DT algorithm [26]

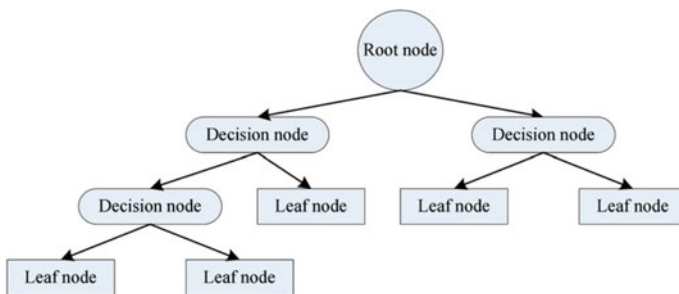


Fig. 1 Concept of decision tree algorithm [26]

2.2 *Random Forests*

Random Forest is a specific classification and regression algorithm based on the mixture of several decision trees (Breiman 2001). This algorithm has been used in several water-related studies [5, 22, 29]. Sadler et al. [19] reported that the traditional DT is prone to over fitting while applying it to the training data. The randomness of the Random Forest prevents over-fitting and leads to better model performance. In the conventional decision tree approach, the dataset is separated into smaller parts by using the best variable splitter, whereas Random Forest changes the splits while choosing random predictors. This makes Random Forest a more robust algorithm than DT. There is no assumption for Random Forest when splitting samples, and it clusters everything automatically [27].

2.3 *Logistic Regression*

Logistic Regression is a specific form of generalized linear model regression [17, 28]. It can be used to calculate the probability of a sample pertaining to one class [4]. Or in this case, can be used to classify binary values, such as broken and none broken pipes.

3 Methodology

3.1 *Data Cleaning*

The first step in developing a statistical predictive model is to clean and prepare the data. Cleaning facilitates the subsequent modelling steps since many outliers and inconsistent entries are removed. Thus, the accuracy of the predictive models can be enhanced. The case study datasets include pipe diameter, pipe length, age at failure, month of failure, and material. In the present study, only pipes with a length of 200 m or less were considered, as well as an age of 80 years or less, since less information is available for older breaks.

Data on water main and breaks is generally available in two separate datasets, an inventory and a main break register. These datasets were merged, allowing for the identification in the inventory of pipes that have failed or not.

In order to account for the impact of different monthly weather condition, the month at time of break was defined as a binary value (0 or 1) for each month. Furthermore, although there is a range of materials within the datasets, the most common materials were used in the analysis, Asbestos Cement (AC), Cast Iron (CI), Ductile Iron (DI), Polyvinyl Chloride (PVC) and Polyethylene (PE), focusing on CI pipes. Each material was also treated as a binary value, and 0 or 1 was assigned

Table 1 Summary of input attributes for Waterloo

| Attribute | Count | Mean | STD | Min value | Max value |
|---|--------|--------|-------|-----------|-----------|
| Diameter (mm) | 28,869 | 200.84 | 64.31 | 40 | 600 |
| Length (m) | 28,869 | 30.49 | 39.61 | 0.20 | 199.97 |
| Age (years) | 28,869 | 27.37 | 19.61 | 0 | 80 |
| Break month: Jan–Dec (Binary variable) | 28,869 | – | – | 0 | 1 |
| PVC | 28,869 | 0.54 | 0.50 | 0 | 1 |
| Asbestos cement | 28,869 | 0.32 | 0.47 | 0 | 1 |
| Cast iron | 28,869 | 0.13 | 0.34 | 0 | 1 |
| PE | 28,869 | 0.00 | 0.05 | 0 | 1 |
| Ductile iron | 28,869 | 0.00 | 0.05 | 0 | 1 |

Table 2 Summary of input attributes for Waterloo

| Attribute | Count | Mean | STD | Min value | Max value |
|---|-------|--------|-------|-----------|-----------|
| Diameter (mm) | 6884 | 202.06 | 70.15 | 25 | 150 |
| Length (m) | 6884 | 46.50 | 52.25 | 0.10 | 5.80 |
| Age | 6884 | 34.44 | 18.10 | 2 | 20 |
| Break month: Jan–Dec (Binary variable) | – | – | – | 0 | 1 |
| PVC | 6884 | 0.58 | 0.49 | 0 | 0 |
| Asbestos cement | 6884 | 0.00 | 0.05 | 0 | 0 |
| Cast iron | 6884 | 0.27 | 0.44 | 0 | 0 |
| PE | 6884 | 0.00 | 0.02 | 0 | 0 |
| Ductile iron | 6884 | 0.15 | 0.36 | 0 | 0 |

to each material. Tables 1 and 2 which are provided in Section 4 show the most important attributes that have been used in this analysis.

3.2 Classification of Broken and Non-broken Pipes

Three different machine learning algorithms were applied to both datasets, Random Forest Classifier, Logistic Regression Classifier, and Decision Tree. Each model was applied to the Saskatoon and Waterloo data separately, which includes five types of materials; AC, CI, PVC, PE, and DI. The datasets were divided into training set (70%) and test set (30%). Accuracy was evaluated on the test set, as well with a fivefold cross validation approach. Such an evaluation (Cross Validation) indicates accuracy among different portions of data. In addition to accuracy, precision, recall and F-1

score were also calculated. These classification metrics present the prediction power of the applied classification models.

Once the models were developed for each system, the Saskatoon model was tested on Waterloo data. This allows for the comparison of the models, and evaluation of their applicability under other conditions. There are many environmental factors that contribute to pipe failure, such as temperature, soil type, etc. which vary among different systems. Thus, Waterloo and Saskatoon have different characteristics within their water networks that are not currently included in the dataset.

Since homogeneity is reported to be significantly important while making a predictive model [21], Cast Iron (CI) pipes were also analyzed separately for Saskatoon and Waterloo.

3.3 Case Study Systems

The data for this analysis was collected from the cities of Saskatoon, Saskatchewan and Waterloo, Ontario. Water main inventories, as well as water main breaks were used to predict pipe failure. Main break data is available for 2000 to 2019. Saskatoon and Waterloo water networks consist of 1,188 KMs and 433 KMs of water mains, respectively. This includes a variety of materials such as Cast Iron (CI), Asbestos Cement (AC), Ductile Iron (DI), Polyvinyl Chloride (PVC), Polyethylene (PE), High-density polyethylene (HDPE) and Flexible Polyvinyl Chloride (FPVC). However, the final materials included in the analysis are Cast Iron (CI), Ductile Iron (DI), PVC, Polyethylene (PE) and Asbestos Cement. Table 3 shows the proportion of each material within both network after data cleaning. AC and CI pipes account for almost 48% of the Saskatoon network, 564 km. In Waterloo, CI pipe makes up around 30% of the network, 128 km. Thus, this report focuses primarily on Cast Iron pipes. The most important attributes employed in the analysis for both city of Saskatoon and Waterloo (Tables 1 and 2).

Table 3 Proportion of each type of material in both networks

| Saskatoon | | | Waterloo | | |
|----------------------|----------------|-------------|----------------------|----------------|-------------|
| Material | Percentage (%) | Length (km) | Material | Percentage (%) | Length (km) |
| Asbestos cement (AC) | 38.12 | 335.5 | Asbestos Cement (AC) | 0.25 | 0.80 |
| Cast iron (CI) | 15.02 | 132.2 | Cast Iron (CI) | 25.05 | 80.17 |
| Ductile iron (DI) | 0.14 | 1.2 | Ductile Iron (DI) | 15.16 | 48.53 |
| Poly ethylene (PE) | 0.13 | 1.1 | Poly Ethylene (PE) | 0.07 | 0.21 |
| PVC | 46.60 | 410.1 | PVC | 59.48 | 190.39 |
| Total | 100 | 880.07 | Total | 100 | 320.10 |

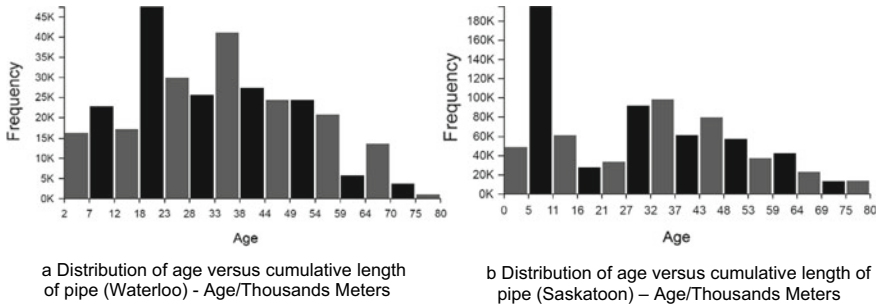


Fig. 2 a Distribution of age versus cumulative length of pipe (Waterloo)—Age/Thousands Meters. b Distribution of age versus cumulative length of pipe (Saskatoon)—Age/Thousands Meters

There are also different pipes with different ages in the inventory. The distribution of age versus cumulative length of pipes for both networks is provided in the given figures (Fig. 2a) and (Fig. 2b). It should be noted that the age distribution has been narrowed down, in order to improve the accuracy of the model.

4 Results

4.1 Classification Metrics

Several primary classification metrics can be employed in order to analyze the reliability of classifiers. These include accuracy, precision, recall and F-1, and are calculated using a confusion matrix. The confusion matrix not only indicates the general accuracy of the model, but demonstrates the incorrect and correct prediction of a classifier. This matrix includes the number of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) as defined in Table 4.

Considering these terms, the following evaluation metrics can be calculated.

- **Accuracy:** This metric can be used to evaluate True results (negative and positive) compared to all results (TP, TN, FP and FN). It is straightforward to understand and typically used for evaluation of datasets including balanced classes (Negative or Positive value of target labels are relatively equal). When imbalanced classes are available in the dataset, accuracy alone could result in misinterpretation [10].

Table 4 Document Classification [11]

| | |
|----------------|--|
| True Positive | True positive labels in the main dataset |
| True Negative | True negative labels in the main dataset |
| False Positive | False prediction of positive value |
| False Negative | False prediction of negative value |

Table 5 Evaluation metrics [26]

| | |
|---------------|---------------------------------|
| Accuracy | $(TP + TN)/(TP + FP + FN + TN)$ |
| Precision (P) | $(TP)/(TP + FP)$ |
| Recall (R) | $(TP)/(TP + FN)$ |
| F-1 | $2 * ((P * R)/(P + R))$ |

- **Precision:** This metric can be used to evaluate Positive results and the accuracy of positive prediction. When positive results are influential in the analysis, this metric would be helpful.
- **Recall:** This metric is employed for evaluation of actual positive classification. When the importance of positive target is important, recall can be used.
- **F-1:** In order to have better understanding of accuracy of the classifiers, precision and recall could be combined. This combination would result in F-1 which is a reliable indicator, showing the accuracy of classification models. This metric also can be used in a case that target label has imbalanced values. Table 5 shows the equations for each of these metrics.

4.2 Classification Results

Evaluation metrics of the classification models applied to Saskatoon and Waterloo are provided in Table 6. The Random Forest classifier outperformed other models in both networks in the first step (dataset including different materials) with general accuracy of 77% and 97%, respectively. However, F-1 score for Waterloo and Saskatoon is higher for Logistic Regression model. Moreover, after applying cross validation to test the models, Logistic regression indicated better accuracy with 60% of accuracy and 83% recall. In waterloo, however, Logistic Regression did not show a significant improve after cross validation method. Random Forest found age and length to be the most important predictors with 43% and 41% contribution, respectively. However, Decision Tree considered age as the most important contributor with 75% weight. The significant difference between F-1 score in both case studies indicates the dependency of the predictive models on each specific site factors. It can be seen F-1 score shows higher accuracy for Saskatoon dataset. The analysis carried on with applying models to Cast Iron (CI) pipes in Saskatoon. Interestingly, with partitioning the dataset into specific material type, the accuracy of the models increased relatively. For instance, Random Forest Accuracy surged from 77% in the first step to 81%, and Decision Tree from 74 to 81%. In this step Random Forest and Decision Tree represented a better performance comparing to Logistic Regression. Recall and Precision are almost same for all three classifiers. In this case, F-1 score for Saskatoon is relatively higher, indicating the higher reliability of classifiers for Saskatoon network. In this preliminary study, age and length found to be the most important attributes affecting the prediction results. Overall, the results demonstrate somewhat reliability of machine learning classifiers for forecasting whether a pipe

Table 6 Comparison of classifiers—(Saskatoon—Waterloo)

| Classification models comparison—(Saskatoon—All types of materials) | | | | | |
|---|--------------|-----------------|------------------|---------------|-----------|
| Model | Accuracy (%) | CV Accuracy (%) | CV Precision (%) | CV Recall (%) | CV F1 (%) |
| Random forest | 77 | 56 | 65 | 77 | 71 |
| Logistic Regression | 66 | 60 | 67 | 83 | 74 |
| Decision tree | 74 | 57 | 65 | 79 | 71 |
| Classification models comparison—(Saskatoon—cast iron pipes) | | | | | |
| Model | Accuracy (%) | CV accuracy (%) | CV precision (%) | CV recall (%) | CV F1 (%) |
| Random forest | 81 | 69 | 75 | 85 | 80 |
| Logistic Regression | 69 | 68 | 74 | 85 | 79 |
| Decision tree | 81 | 70 | 76 | 85 | 80 |
| Classification models comparison—(Waterloo—All types of materials) | | | | | |
| Model | Accuracy (%) | CV accuracy (%) | CV precision (%) | CV recall (%) | CV F1 (%) |
| Random forest | 97 | 94 | 73 | 40 | 38 |
| Logistic Regression | 96 | 95 | 78 | 34 | 41 |
| Decision tree | 96 | 94 | 62 | 35 | 39 |
| Classification models Comparison—(Waterloo—Cast iron pipes) | | | | | |
| Model | Accuracy (%) | CV accuracy (%) | CV precision (%) | CV recall (%) | CV F1 (%) |
| Random forest | 92 | 92 | 77 | 59 | 65 |
| Logistic Regression | 92 | 90 | 73 | 49 | 58 |
| Decision tree | 93 | 90 | 73 | 53 | 60 |

fails in the future. However, according to previous studies [2, 6, 21] many other factors may have significant impact on water main failure, which are not available in this study. These factors could be soil resistivity, temperature, soil type and other operational and environmental features. Additionally, the importance of partitioning data into homogenous groups can be noticed clearly from the result, as the accuracy of the models is clearly improved by grouping them by material. Decision Tree

and Random Forest classifiers seem to be appropriate methods for evaluation and prediction.

In order to evaluate the performance of classifiers, same process has been done for Waterloo, and the results are provided in Table 6. Although the general accuracy of the models is relatively high in the first step (dataset including different materials), after cross validation, the prediction power of models decreased significantly. As it can be seen, recall and F-1 scores for all types of material are not satisfactory. However, after partitioning data and applying the same models to only CI pipes, the accuracy of models in prediction somewhat improved. This again emphasizes the importance of partitioning pipes in homogeneous classes. In this section of analysis length and age with 30% and 36% contribution—according to feature importance analysis in open source tools—seemed to be the most important factors that may affect the prediction results. However, this results are not finalized and they are site-specific. Hence, more analysis is required to prove the reliability of the models. As it can be seen F-1 score for Waterloo after cross-validation method increased, which emphasize the importance of partitioning data into homogeneous groups.

5 Summary and Conclusions

The present study focused on failure prediction of water pipes considering pipe age, length, material, and month of failure. Results show that classifiers can provide useful and moderately accurate predictions of pipe failure. However, there are many other attributes that may contribute to failure and could be taken into consideration to increase the reliability of the model. Other significant factors could include soil type, previous rate of failure, temperature, and soil resistivity. Next steps of this project involve including other variables as well as other Canadian cities to the analysis.

References

1. Al-Barqawi H, Zayed T (2008) Infrastructure management: integrated AHP/ANN model to evaluate municipal water mains' performance. *J Infrastruct Syst* 14(4):305–318. [https://doi.org/10.1061/\(ASCE\)1076-0342\(2008\)14:4\(305\)](https://doi.org/10.1061/(ASCE)1076-0342(2008)14:4(305))
2. Andreou SA (1986) Predictive models for pipe break failures and their implications on maintenance planning strategies for deteriorating water distribution systems
3. ASCE (2017) ASCE drinking water report card
4. Chang M, Maguire M, Sun Y (2019) Stochastic modeling of bridge deterioration using classification tree and logistic regression. *J Infrastruct Syst* 25(1):04018041. [https://doi.org/10.1061/\(ASCE\)IS.1943-555X.0000466](https://doi.org/10.1061/(ASCE)IS.1943-555X.0000466)
5. Chen G, Long T, Xiong J, Bai Y (2017) Multiple random forests modelling for urban water consumption forecasting. *Water Resour Manage* 31(15):4715–4729. <https://doi.org/10.1007/s11269-017-1774-7>
6. Clark et al (1982) Water distribution systems: a spatial and cost evaluation
7. Dawood T, Elwakil E, Novoa HM, Delgado JFG (2019) Pipe failure prediction and risk modeling in water distribution networks: a critical review. *Sci Technol* 7

8. Dawood T, Elwakil E, Novoa HM, Delgado JFG (2020) Water pipe failure prediction and risk models: state-of-the-art review. *Can J Civ Eng* 47(10):1117–1127. <https://doi.org/10.1139/cjce-2019-0481>
9. Giraldo-González MM, Rodríguez JP (2020) Comparison of statistical and machine learning models for pipe failure modeling in water distribution networks. *Water* 12(4):1153. <https://doi.org/10.3390/w12041153>
10. Harvey RR, McBean EA (2014) Comparing the utility of decision trees and support vector machines when planning inspections of linear sewer infrastructure. *J Hydroinf* 16(6):1265–1279. <https://doi.org/10.2166/hydro.2014.007>
11. Ikonomakis M, Kotsiantis S, Tampakas V (2005) Text classification using machine learning techniques 10
12. Jafar R, Shahroui I, Juran I (2010) Application of Artificial Neural Networks (ANN) to model the failure of urban water mains. *Math Comput Model* 51(9–10):1170–1180. <https://doi.org/10.1016/j.mcm.2009.12.033>
13. Kleiner Y, Rajani B (2001) Comprehensive review of structural deterioration of water mains: statistical models. *Urban Water* 3(3):131–150. [https://doi.org/10.1016/S1462-0758\(01\)00033-4](https://doi.org/10.1016/S1462-0758(01)00033-4)
14. Lei J, Saegrov S (1998) Statistical approach for describing failures and lifetimes of water mains 9
15. Oliveira D, Guo W, Soibelman L, Jr Garrett JH, (2007) Spatial data management and analysis in sewer systems' condition assessment: an overview. In *Computing in civil engineering (2007)*. American Society of Civil Engineers, Pittsburgh, Pennsylvania, United States, pp 391–398. [https://doi.org/10.1061/40937\(261\)48](https://doi.org/10.1061/40937(261)48)
16. Park HJ, Agbenowosi N, Kim BJ, Lim K (2011) The proportional hazards modeling of water main failure data incorporating the time-dependent effects of covariates. *Water Resour Manage* 25(1):1–19. <https://doi.org/10.1007/s11269-010-9684-y>
17. Robles-Velasco A, Cortés P, Muñozuri J, Onieva L (2020) Prediction of pipe failures in water supply networks using logistic regression and support vector classification. *Reliab Eng Syst Saf* 196(April):106754. <https://doi.org/10.1016/j.ress.2019.106754>
18. Roiger RJ (2017) Basic data mining techniques. In *Data mining*, Roiger RJ (ed) 2nd edn. Chapman and Hall/CRC, pp 63–102. <https://doi.org/10.1201/9781315382586-3>
19. Sadler JM, Goodall JL, Morsy MM, Spencer K (2018) Modeling urban coastal flood severity from crowd-sourced flood reports using Poisson regression and random forest. *J Hydrol* 559(April):43–55. <https://doi.org/10.1016/j.jhydrol.2018.01.044>
20. Mirza S (2007) Danger ahead: the coming collapse of Canada's municipal infrastructure. Ottawa, Ont.: Federation of Canadian Municipalities. <https://www.deslibris.ca/ID/250220>
21. Shamir U, Howard CDD (1979) An analytic approach to scheduling pipe replacement. *J Am Water Works Assoc* 71(5):248–258. <https://doi.org/10.1002/j.1551-8833.1979.tb04345.x>
22. Shirzad A, Safari MJS (2019) Pipe failure rate prediction in water distribution networks using multivariate adaptive regression splines and random forest techniques. *Urban Water Journal* 16(9):653–661. <https://doi.org/10.1080/1573062X.2020.1713384>
23. Snider B, McBean EA (2018) Improving time-to-failure predictions for water distribution systems using gradient boosting algorithm 8
24. Folkman S (2012) Water main break rates in the USA and Canada a comprehensive study 28
25. Folkman S (2018) Water main break rates in the USA and Canada: a comprehensive study 48
26. Syachrani S, Hyung S, Jeong D, Chung CS (2013) Decision tree-based deterioration model for buried wastewater pipelines. *J Perform Constructed Facil* 27(5):633–645. [https://doi.org/10.1061/\(ASCE\)CF.1943-5509.0000349](https://doi.org/10.1061/(ASCE)CF.1943-5509.0000349)
27. Vitorino D, Coelho ST, Santos P, Sheets S, Jurkovic B, Amado C (2014) A random forest algorithm applied to condition-based wastewater deterioration modeling and forecasting. *Procedia Eng* 89:401–410. <https://doi.org/10.1016/j.proeng.2014.11.205>
28. Vladeanu GJ, Koo DD (2015) A comparison study of water pipe failure prediction models using weibull distribution and binary logistic regression. In *Pipelines 2015*. American Society of Civil Engineers, Baltimore, Maryland, pp 1590–1601. <https://doi.org/10.1061/9780784479360.146>

29. Zhu J, Pierskalla WP (2016) Applying a weighted random forests method to extract karst sinkholes from LiDAR Data. *J Hydrol* 533(February):343–352. <https://doi.org/10.1016/j.jhydrol.2015.12.012>

New Hybrid Search Algorithm for the Capacitated Vehicle Routing Problem



Nayera Elgharably, Ashraf Nassef, Said Easa, and Ashraf El Damatty

1 Introduction

The Vehicle Routing Problem (VRP) is one of the most studied combinatorial optimization problems in operations research [17]. The CVRP is an extension of the well-known Traveling Salesman Problem (TSP) where, a set of minimum distance routes are determined to visit a given set of customers with known demands without violating the capacity constraint of the vehicles used [6]. The VRP is classified as NP-hard. Several exact and approximate solution methods have been used to solve the problem. Exact methods can only solve relatively small size problems while approximate algorithms have been able to reach near optimum solutions [3]. The aim of this paper is to present a new hybrid search algorithm for the vehicle routing problem using a new local search heuristic and an evolutionary algorithm.

2 Literature Review

The vehicle routing problem was first tackled by Dantzig et al. [5], as a variant of the travelling salesman problem. The problem was later refined by adding extra realistic constraints such as the capacitation of the vehicle routes [13]. Algorithms employed for solving the problem can be divided into algorithms seeking exact optimum

N. Elgharably (✉) · A. El Damatty
Western University, London, ON, Canada
e-mail: nelghara@uwo.ca

A. Nassef
American University in Cairo, Cairo, Egypt

S. Easa
Ryerson University, Toronto, ON, Canada

solutions (exact algorithms) and those seeking near optimal solutions (approximate algorithms). In further research both categories were hybridized. Exact algorithms are basically based on dynamic programming, and set-partitioning methods [2]. However, they can solve small sized problems. Later research utilized partial enumeration algorithms such as branch and bound, heuristics or pseudo random search algorithms to arrive at near optimal solutions of larger problem instances. Several heuristics (approximate algorithms) have been proposed for the VRP and are divided into two classes: classical heuristics and metaheuristics. Classical heuristics as the sweep algorithm, the Clarke and Wright algorithm, and the Christofides-Mingozi-Toth two-phase algorithm were addressed by Laporte [13]. Montoya-Torres et al. [14] used Random based heuristic algorithm to design vehicle routes. Faulin et al. [8] introduced the SR1 Simulation based heuristic algorithm that uses initial good solutions from the classical Clarke and Wright Heuristic then a random oriented local search is used to find the list of best solution routes. Weyland et al. [23] proposed a local search Heuristic that assigns different collection points to vehicles to solve a real-world oil collection problem of the VRP. Metaheuristics as Genetic Algorithms (GA), Tabu Search (TS), and Simulated Annealing (SA) are discussed in literature. Baker and Ayechev [1] applied a straightforward genetic algorithm to the VRP and showed that incorporating neighbourhood search into the GA produces significant improvement to the solution. Koo et al. [12] proposed a 2-phase heuristic procedure. The first phase finds the lower bound of the fleet size, while the second phase applies a tabu search to find the solution set of routes. Wassan [22] introduced a Reactive Tabu Search (RTS) with a new escape mechanism to solve the CVRP. A study by Karakatic and Podgorelec [11] presented a survey on genetic algorithms and stated that GA is preferred for solving large NP-hard Problems over exact and other heuristic methods due to their main advantage of the linear scaling with growing problem size. Wang et al. [21] applied a genetic algorithm-based approach to solve a 2-echelon CVRP with stochastic demands with 4 satellites and 20 customers. Biesinger et al. [4] introduced a GA that uses a solution archive to solve the VRP to store all generated solutions and avoid adding duplicates to the population.

A limited number of hybrid search algorithms are proposed in literature. Subramanian et al. [16] proposed ILS-SP hybrid algorithm that combines the Iterated Local Search heuristic with the Set Partitioning approach to find new solutions based on known routes from previous local optimums. Vidal et al. [19, 20] proposed the Unified Hybrid Genetic Search (UHGS) that finds not only good but diverse solutions by applying a continuous diversification procedure to modify the objective function during parents and survivors' selection [17].

In this paper, a new hybrid search algorithm is proposed. The algorithm combines the evolutionary genetic search with a new local search heuristic. In routing decisions, the heuristic considers both locations and demand quantities of the nodes to be visited not just distances travelled. Both the location and the demand associated with each node/customer were considered in routing decisions as implemented in the new local search heuristic. In addition, the GA operators will use the resultant local search heuristic as a tool to adjust the routes created in the process of applying the mutation and crossover operators to guarantee the feasibility of the routing decisions. The

proposed hybrid algorithm utilizes the resultant local search heuristic in applying the GA operators so that the solution produced requires no repairing.

3 Problem Description

3.1 Characteristics of the Problem

The CVRP consists of $n + 1$ points, n customers and a depot. Distances (d_{ij}) between each two points is known. The objective is to determine a set of minimum cost routes to be performed by a homogeneous fleet of vehicles (m) to serve a given set of customers (n) with known demands (q), where each route starts and ends at a single depot.

Each customer must be assigned to only one vehicle and the total demand of all customers assigned to a vehicle does not exceed its capacity (Q). The number of vehicles (routes) to be used is not fixed but to be determined by the solution approach. In some studies, the number of vehicles is fixed, while others define a minimum possible number of vehicle routes (K_{min}). According to [17], there are two reasons for not fixing the number of vehicles used. The first reason is that fixing the number of routes is an indirect way of minimizing the fixed cost associated with the cost per vehicle. In other words, this way ignores the trade-off between the variable and fixed costs associated with the suggested set of routes. The second reason is that in the literature the original CVRP proposed by [5] does not consider fixing the number of routes in the problem as it requires adding the cost of unused capacity to the model which in practice is of minor importance. According to the authors, minimization of the travel distance is independent of the number of vehicles used. Table 1 summarizes the characteristics of the CVRP of the study.

Table 1 Problem characteristics

| Element | Characteristics |
|-----------------------|--|
| Size of fleet | Unbounded |
| Type of fleet | Homogenous |
| Origin of vehicles | Single depot |
| Demand type | Deterministic demand (Known) |
| Location of demand | At the customer (node) |
| Maximum time on route | No constraint |
| Objective | Minimize total distance |
| Constraints | Single visit at customers, Routes start and end at depot, Nodes served by single vehicle, Vehicle capacity cannot be exceeded |

3.2 Mathematical Modeling

The formulation of the problem is presented in a previous publication where integer decision variables are used where the formulation has been validated using several tools [7]. The VRP problem is a generalization of the TSP that introduces more than one salesman (m ; hence, m number of tours can be done; each starting and ending at the depot). For formulating the VRP, the starting customer is considered node 1 (depot), where X_i represents the current visited node and Y_i represents the next node to be visited, where i varies from 1 to n , and n is the number of nodes to be visited by a given vehicle k . Now, m routes are introduced to the model, where distance d_{X_i, Y_j} is associated with each arc and represents the distance travelled from node X_i^k to node Y_j^k on route k , as shown in Fig. 1.

The decision variables are Y_i^k , where Y_i^k determines the value of the next customer i to be visited on route k . The X_i^k variable represents the value of the start node of the arc on route k . The use of loop segments is not allowed (leaving a node then arriving to same node, $X_i^k \neq Y_j^k$), as all nodes must be visited exactly once. The binary variable S_{X_i, Y_j}^k represents all possible arcs connecting any two nodes on route k . S_{X_i, Y_j}^k is given a value of 1 if arc (X_i^k, Y_j^k) belongs to route k ; 0 otherwise. Both X_i^k and S_{X_i, Y_j}^k are considered uncontrollable variables. The problem is formulated as follows:

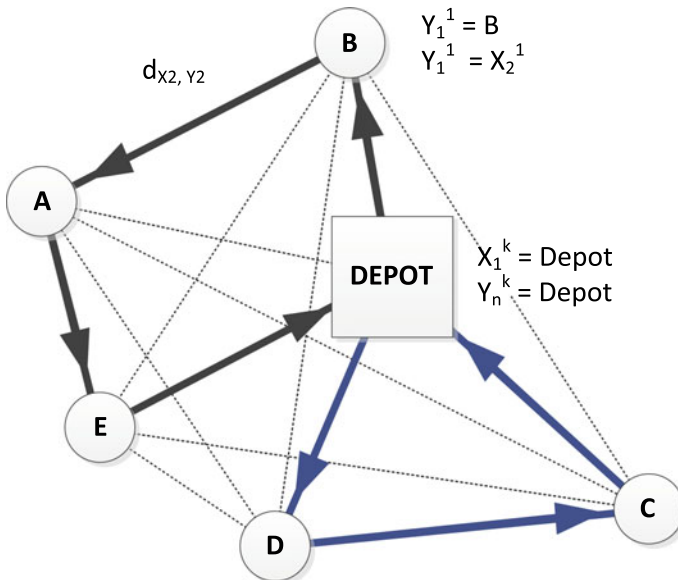


Fig. 1 Illustration of the VRP, [7]

$$\text{Minimize } Z = \sum_{k=1}^m \sum_{i=1}^n \sum_{j=1}^n S_{X_i, Y_j}^k * d_{X_i, Y_j} \quad (1)$$

Subject to

$$X_1^k = 1 \quad \forall k = 1, \dots, m \quad (2)$$

$$Y_n^k = 1 \quad \forall k = 1, \dots, m \quad (3)$$

$$X_i^k = Y_{i-1}^k \quad \forall i = 2, \dots, n, \quad \forall k = 1, \dots, m \quad (4)$$

$$X_i^k \leq n \quad \forall k = 1, \dots, m \quad (5)$$

$$Y_i^k \leq n \quad \forall k = 1, \dots, m \quad (6)$$

$$\sum_{j=1}^n \sum_{k=1}^m S_{X_i, Y_j} = 1 \quad \forall i = 2, \dots, n \quad (7)$$

$$\sum_{i=1}^n \sum_{k=1}^m S_{X_i, Y_j} = 1 \quad \forall j = 1, \dots, n-1 \quad (8)$$

$$\sum_{j=1}^n \sum_{i=1}^n S_{X_i, Y_j}^k * q_{Y_j} \leq Q_k \quad \forall k = 1, \dots, m \quad (9)$$

$$X_i^k, Y_j^k > 0 \text{ and integer} \quad (10)$$

The objective function (1) minimizes the total travel distance on all k routes, where m is the number of routes proposed. Constraints (2) and (3) ensure that each route starts and ends at the depot. Constraint (4) ensures that each route of the k routes is not segmented, that is, if a vehicle arrives at a customer, it eventually leaves the customer again. Constraints (5) and (6) state the range of values given, whereas constraints (7) and (8) state that every customer is visited exactly once. Knowing that at each customer, customers' demand (q_{Y_j}) is present and that each vehicle has limited capacity Q_k . Constraint (9) ensures that the total demand of all customers assigned to a route k does not exceed the vehicle's capacity. Finally, constraint (10) is the non-negativity constraint and assures that the variables can assume integer values only.

4 New Hybrid Search for VRP

The proposed hybrid search algorithm combines the evolutionary genetic search algorithm with a new local search heuristic that calculates a heuristic resultant based on both the distance travelled or the location of the nodes/customers and the demand associated with the given node/customer. Genetic algorithm is considered an approximate solution approach (metaheuristic) that is used to solve NP hard class of problems to obtain not necessarily optimum but near optimum solutions. GA's performance and results on time constraints and limited computer power obtains competitive solutions compared to other metaheuristic approaches. GA is a stochastic adaptive optimization algorithm which a subset of evolutionary algorithms, that adopts Darwin's theory of evolution, consisting of the reproduction, selection and diversity nature basic principles. It was first introduced in 1960 by John Holland [11].

4.1 Resultant Local Search Heuristic

In the implemented local search method, a heuristic resultant for each customer was used as follows:

$$HR_i = \alpha d_i + (1 - \alpha)DR_i \quad (11)$$

where HR_i = heuristic resultant for customer i , α and $(1 - \alpha)$ = weights of the distance and demand (used to achieve diversity and not to be caught in local optimum), d_i = euclidian distance to be travelled from the current node to the expected following node by customer i , and DR_i = demand remainder for customer i , which is the difference between the vehicle's capacity and the demand (i), where demand (i) is the quantity of items to be delivered or picked up by the vehicle at the customer i . For example, at the beginning of constructing the route, the current location would be the depot, while in the middle of the route the current location would be the last visited node/customer. The function identifies the nearest route (heuristic) based on the resultant heuristic function between the remainder of the demand of each node compared to the vehicle capacity and the distance from the current location to the following node. Routes are constructed using the nodes (i) of the nearest heuristic resultant until the vehicle capacity is reached then a new route is initiated. The developed resultant heuristic is used in the initialization process of the population generation and in deterministic operators described in the following subsections. As stated by Baker and Ayechev [1] incorporating neighborhood searches to the GA resulted in more improvements to solution. Therefore, a portion of the evolutionary search population is filled heuristically using the local search resultant heuristic, while the remaining portion is filled randomly.

4.2 Initial Population and Fitness Unction

In GA, the first step is the initialization of population that consists of several solutions to the problem. Each solution is called a chromosome. A fitness function associated to each chromosome is calculated to evaluate the goodness of each solution. In case of CVRP, the lower scores of the fitness function are favored, since CVRP is a minimization problem of the total distance travelled by the vehicles. Each chromosome is a matrix (n, n) , n is the number of nodes/customers to be visited in the given problem of study that represents a feasible solution to the problem. Each row in the chromosome matrix represents a route that starts and ends at the depot with no violations to the capacity constraints. A portion of the initial population is filled heuristically using the local search heuristic developed, the remaining portion is filled randomly to achieve diversity and not to be caught in a local optimum. The random part of the initial population is based only on the vehicle capacity ignoring any distance calculations. A set of operators are then performed to the initial population to mimic the nature of evolution. Operators as selection, mutation and crossover are used to widen the search space and inherit good solutions to the next generations. An elitist selection process is performed, where a portion of the existing population is used to breed the new generation. Individuals are selected based on their fitness function.

4.3 Mutation Operators

To achieve diversity and to widen the span of the search space, a set of one deterministic and four random mutation operators is applied. A deterministic Route Reduction Mutation (RRM) is performed that decreases the number of routes in a solution without violating any constraints. The aim is to lower the number of routes considering only capacity and demand calculations. The routes found in an individual solution are sorted based on the highest remaining demand compared to the vehicle capacity. Routes with maximum remaining capacity are combined with the ones with min demand (Fig. 2). While routes with remaining capacity less than the minimum demand in the route remain unchanged. The routes are then adjusted using the resultant local search heuristic. Comparably, [10], applied a Vehicle Merge operator to a

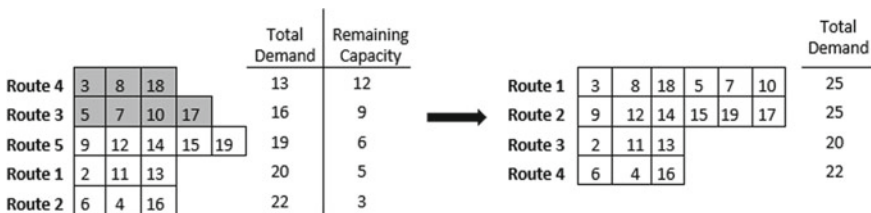


Fig. 2 Illustration of the route reduction mutation

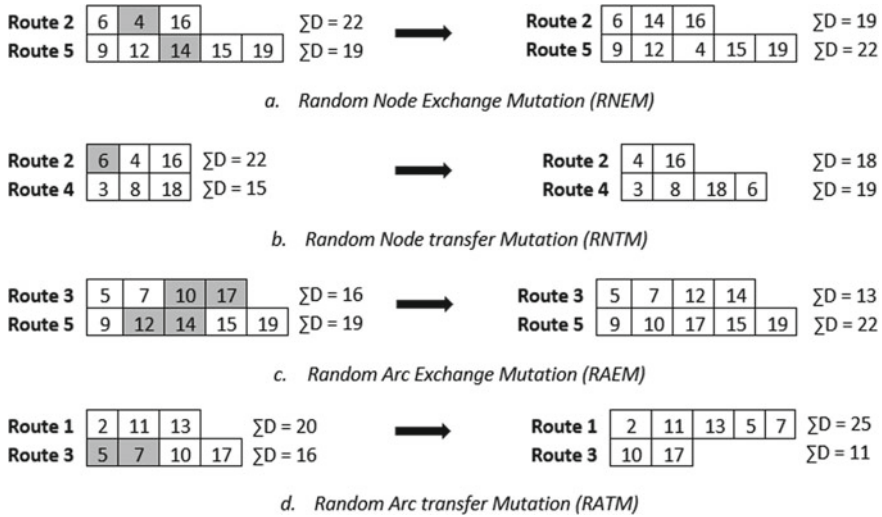


Fig. 3 Illustration of the random mutation operators

pickup and delivery VRP that merges two vehicles selected at random. The nodes of the selected vehicles are placed in a relocation pool and distributed on the existing vehicle routes before constructing new routes.

Random Node Exchange Mutation (RNEM) is a mutation operator that exchanges nodes from randomly selected routes without violating any capacity constraints. Two nodes are selected at random from the previously chosen routes and are then exchanged yielding to different routes with updated total demand for each route (Fig. 3a). The routes are then adjusted using the resultant local search heuristic. The random node exchange mutation was used by Baker and Ayechev [1], and Biesinger et al. [4].

Random Node transfer Mutation (RNTM) is a mutation operator that transfers a randomly selected node from one route to another. The two selected routes are chosen randomly (Fig. 3b) where if a node is transferred from a one-node route the number of routes will decrease by one. The routes are then adjusted using the developed resultant heuristic with no capacity violation. Similarly, a mutation operator called relocation heuristic by Wang et al. [21] and an insertion mutation by Pereira et al. [15] and Ursani et al. [18] were applied in literature. However, these operators remove one customer from its location and reinsert it in a different location whether in the same route or a different one with no demand and vehicle capacity considerations.

Random Arc Exchange Mutation (RAEM) illustrated in Fig. 3c and Random Arc Transfer Mutation (RATM) illustrated in Fig. 3d follows the same process of the RNEM and RNTM but instead of selecting nodes at random, arc within the route are selected at random. Taking into consideration if an arc is transferred from a two-node route the number of routes will decrease by one, in case of the RATM

operator. A route insertion mutation by Garcia-Najera and Bullinaria [9] applies the same concept of the arc transfer mutation.

4.4 Crossover Operators

Crossover is the process which two individual chromosomes act as two parents and are combined to produce two children where the children inherit characteristics from the parents. Two crossover operators are performed, one at random while the other is deterministic that inherits good characteristics from parents.

Hosny and Mumford [10] applied a vehicle copy crossover that copies a random number of good routes from each parent, where good routes are ranked according to the number of nodes served in each route. If the number of routes is similar, then routes are ranked based on the total distance travelled. Wang et al. [21] modified the vehicle copy crossover to use a different insertion heuristic to construct routes for the remaining node in the relocation pool rather than the construction algorithm applied by Hosny and Mumford [10]. The Heuristic Inheritance Crossover (HIC) is a deterministic crossover operator that perform changes to the routes within a given solution inheriting good routes without violating any constraints. The HIC is used for intensification of good solutions in the breeding generation rather than diversification. In each chromosome the heuristic resultant is calculated based on distances and demands for each route and is sorted. The number of good routes to be inherited by each child is predetermined. Then the best predetermined number of routes from Parents 1 and 2 are sent to each child correspondingly. From the other parent, the routes with no common nodes are inherited and sent to each child. The remaining set of nodes that are not present in any of the selected parent routes are considered floating nodes that are found in a relocation pool and are to be distributed among the routes or to form new routes in each child without violating capacity constraints (Fig. 4). The new routes are then adjusted using the resultant local search heuristic.

Random Inheritance Crossover (RIC) follows the same process as the HIC, the only difference is that the routes to be inherited from parent 1 are chosen at random not based on good routes. The RIC operator acts as a diversification operator.

5 Computational Study

To evaluate the performance of the developed algorithm a computational study is conducted. Several benchmark data sets were proposed in literature. Uchoa et al. [17] proposed a new benchmark dataset that provides a more comprehensive and balanced experimental setting to the classic CVRP.

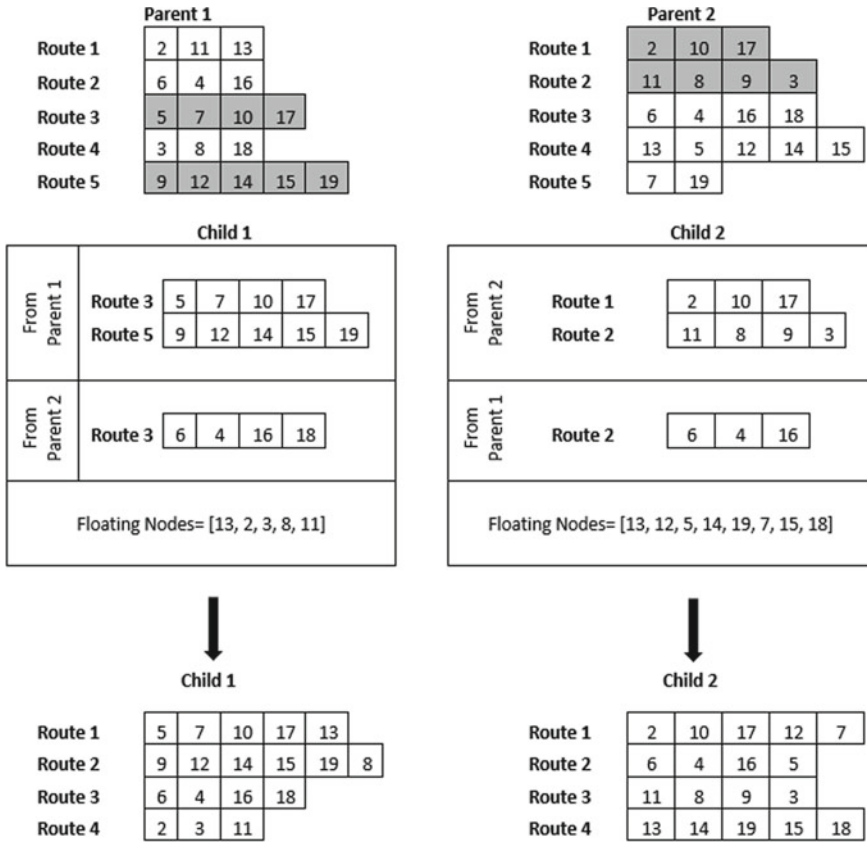


Fig. 4 Illustration of the crossover operators

5.1 The Benchmark Problem Instance

The validity of the proposed solution algorithm is checked using the problem instance: X-n101-k25, taken from [17] new benchmark instances and is implemented in MATLAB. Instance: X-n101-k25 consists of a depot and 100 customers, the number of vehicles to be used is not fixed but the minimum feasible number of vehicles is known ($K_{min} = 25$). The vehicle capacity is 206 units. Demands of customers [0, 100] are deterministic. Euclidian distances are calculated from the given X and Y co-ordinates. The depot and customer positioning of the X-n101-k25 instance is random and the solution of the instance is known (total distance = 27,591). The validation of the proposed solution algorithm is a two-step process. First, the percentage of the heuristic local search used in the initial population of the hybrid algorithm is to be determined (Sect. 5.2). Then, the second step is to determine the best set of model parameters to be used in the evolutionary model (Sect. 5.3).

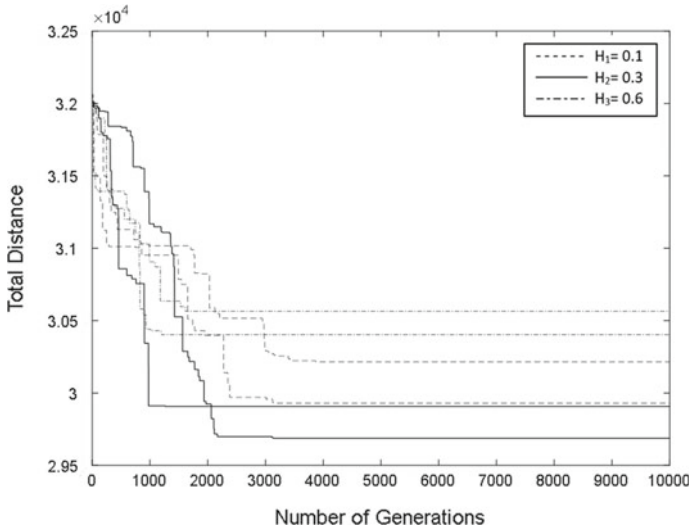


Fig. 5 Sample runs performed to determine the heuristic (H) portion of the initial population

5.2 Effect of Using Local Search Heuristic

In the developed algorithm, a portion of the initial population is filled heuristically using the local search heuristic developed while the remaining portion is filled randomly to achieve diversity. To determine the portion of the initial population to be filled heuristically, a set of runs with different percentages of the local search heuristic are performed.

The algorithm is tested several times at different percentages ranging from 10 to 90% of the population. Figure 5 shows a sample of the runs performed at different levels of the heuristic H at 10, 30 and 60% of the initial population to be filled heuristically using the local search heuristic. H in the figure denotes the percentage of local search heuristic usage in the initial population of the hybrid algorithm. At thirty percent ($H = 0.3$), the figure shows that the model converges to better solutions rather than the ten and sixty percent.

5.3 Evolutionary Model Parameters

To determine the evolutionary model parameters, a second set of runs is performed. Five different trials are performed to determine the number of times to perform the mutation and crossover operators. Each trial is experimented at different levels of α in Eq. 11. The scenario assumed for each trial is as follows:

- Trial 1: reduced crossovers and increased random node exchange and transfer mutations,
- Trial 2: reduced crossovers and increased random arc exchange and transfer mutations,
- Trial 3: reduced crossovers and route reduction mutation, and increased all other operators,
- Trial 4: increased route reduction mutation and reduced all other operators,
- Trial 5: increased random crossover and reduced all other operators.

Table 2 shows the configuration of each trial. The best results of the runs performed for each of the five trials are illustrated in Fig. 6, where Fig. 6c shows that the configuration of Trial 3 achieved the best fitness value; shortest total distance compared to the other trials. More runs were performed on trial 3 configuration. More time and

Table 2 Mutation and crossover operator configuration of each trial

| Trial number | Number of times to perform the operator | | | | | | |
|--------------|---|------------------------|------------------------|-----------------------|-----------------------|----------------------------------|-------------------------------|
| | Route reduction mutation | Node exchange mutation | Node transfer mutation | Arc exchange mutation | Arc transfer mutation | Heuristic inheritance cross-over | Random inheritance cross-over |
| 1 | 12 | 10 | 10 | 4 | 4 | 5 | 5 |
| 2 | 10 | 5 | 5 | 10 | 10 | 5 | 5 |
| 3 | 6 | 10 | 10 | 10 | 10 | 2 | 2 |
| 4 | 20 | 5 | 5 | 5 | 5 | 5 | 5 |
| 5 | 5 | 5 | 5 | 5 | 5 | 5 | 20 |

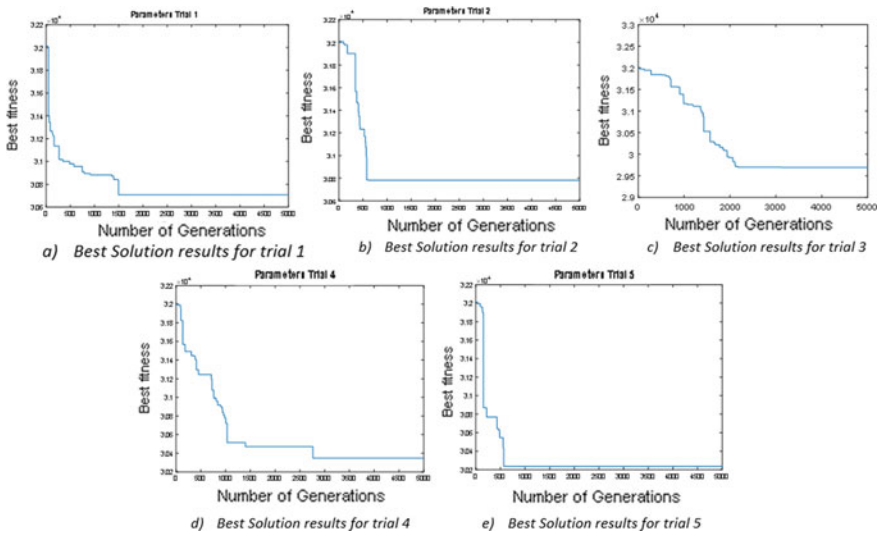


Fig. 6 Best solution reached at each trial

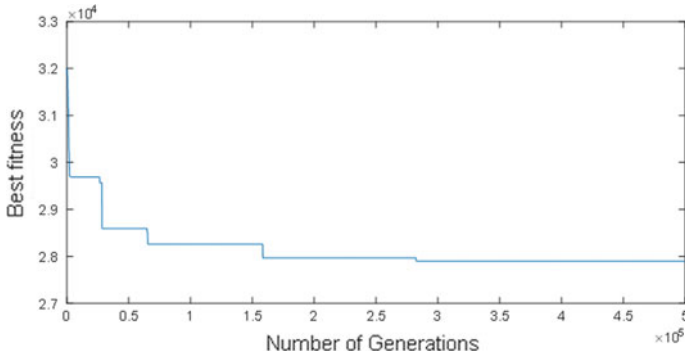


Fig. 7 Validation of the proposed hybrid algorithm

a greater number of generations were used to run the algorithm to test its capability of reaching the best-known solution. The introduced new hybrid search algorithm was capable of finding the best-known solution to the Uchoa et al. [17] benchmark X-n101-k25 data instance as shown in Fig. 7.

6 Conclusions

A new hybrid search algorithm that combines the evolutionary genetic search with a new local search heuristic is developed to solve the capacitated vehicle routing problem. The proposed heuristic calculates a heuristic resultant based on both the distance travelled and the demand associated with the given customer (not only distances as previously considered in the literature). The developed algorithm will be a fundamental tool for the development of a multi-objective green VRP that considers demand quantities in the calculation of fuel consumption rates. For this reason, the demand quantity consideration was included as an aspect in the routing decisions. In addition, a new set of simple genetic operators that requires no further repairing after application were developed and implemented in the algorithm. Several computational experiments were conducted to define the best set of model parameters. The proposed algorithm was validated and found to be satisfactory. The developed algorithm was capable of converging to the optimum solution of the tested benchmark instance. The developed algorithm is considered the base model to be used in later research work to be published by the authors, where the hybrid algorithm will be implemented in solving a multi-objective green vehicle routing problem in both deterministic and stochastic environments.

References

1. Baker MB, Ayechev MA (2003) A genetic algorithm for the vehicle routing problem. *Comput Oper Res* 30:787–800
2. Baldacci R, Hadjiconstantinou E, Mingozzi A (2004) An exact algorithm for the capacitated vehicle routing problem based on a two-commodity network flow formulation. *Oper Res* 52:723–738
3. Baldacci R, Toth P, Vigo D (2010) Exact algorithms for routing problems under vehicle capacity constraints. *Ann Oper Res* 175:213–245
4. Biesinger B, Hu B, Raidl GR (2018) A genetic algorithm in combination with a solution archive for solving the generalized vehicle routing problem with stochastic demands. *Transp Sci* 52(3):673–690
5. Dantzig GB, Ramser JH (1959) The truck dispatching problem. *Manage Sci* 6:80–91
6. Derigs U, Reuter K (2009) A simple and efficient tabu search heuristic for solving the open vehicle routing problem. *J Oper Res Soc* 60:1658–1669
7. Elgharably NE, El-Kilany KS, El-Sayed AE (2013) Optimization using simulation of the vehicle routing problem, world academy of science, engineering and technology. *Int J Indus Manuf Eng* 7(6):1236–1242
8. Faulin J, Gilibert M, Juan AA, Vilajosana X, Ruiz R (2008) SR-1: a simulation based heuristic algorithm for the capacitated vehicle routing problem. Winter simulation conference, Miami, FL
9. Garcia-Najera A, Bullinaria AJ (2011) An improved multi-objective evolutionary algorithm for the vehicle routing problem with time windows. *Comput Oper Res* 38:287–300
10. Hosny, M., and Mumford, C. 2009. Investigating genetic algorithms for solving the multiple vehicle pickup and delivery problem with time windows, MIC: The VIII metaheuristics international conference, Hamburg, Germany
11. Karakatic S, Podgorelec V (2015) A survey of genetic algorithms for solving multi depot vehicle routing problem. *Appl Soft Comput* 27:519–532
12. Koo PH, Lee WS, Jang DW (2004) Fleet sizing and vehicle routing for container transportation. *OR Spectrum* 26:193–209
13. Laporte G (1992) The vehicle routing problem: an overview of exact and approximate algorithms. *Eur J Oper Res* 59:345–358
14. Montoya-Torres JR, Alfonso-Lizarazo EH, Franco EG, Halabi AX (2009) Using randomization to solve the deterministic single and multiple vehicle routing problem with service time constraints. In: Winter simulation conference, Austin, TX, USA, pp 2989–2994
15. Pereira FB, Tavares J, Machado P, Costa E (2002) GVR: a new genetic representation for the vehicle routing problem. In: O’Neill M, Sutcliffe RFE, Ryan C, Eaton M, Griffith NJL (eds) Artificial intelligence and cognitive science. AICS 2002. Lecture Notes in Computer Science, 2464. Springer, Berlin, Heidelberg
16. Subramanian A, Uchoa E, Ochi LS (2013) A hybrid algorithm for a class of vehicle routing problems. *Comput Oper Res* 40:2519–3253
17. Uchoa E, Pecin D, Pessoa A, Poggi M, Vidal T, Subramanian A (2017) New benchmark instances for the capacitated vehicle routing problem. *Eur J Oper Res* 257:845–858
18. Ursani Z, Essama D, Cornforthb D, Stocker R (2011) Localized genetic algorithm for vehicle routing problem with time windows. *Appl Soft Comput* 11:5375–5390
19. Vidal T, Crainic TG, Gendreau M, Lahrichi N, Rei W (2012) A hybrid genetic algorithm for multidepot and periodic vehicle routing problems. *Oper Res* 60(3):611–624
20. Vidal T, Crainic TG, Gendreau M, Prins C (2014) A unified solution framework for multi-attribute vehicle routing Problems. *Eur J Oper Res* 234:658–673
21. Wang K, Lan S, Zhao Y (2017) A genetic-algorithm based approach to the two-echelon capacitated vehicle routing problem with stochastic demands in logistics service. *J Oper Res Soc* 68(11):1409–1421

22. Wassan NA (2006) A reactive tabu search for the vehicle routing problem. *J Oper Res Soc* 57:111–116
23. Weyland D, Salani M, Montemanni R, Gambardella LM (2013) Vehicle routing for exhausted oil collection. *J Traffic Logistics Eng* 1:5–8

Comparison of Evaporative Losses in Alberta Based on Five Evapotranspiration Models



Zahidul Islam, Shalini Kashyap, and Michael Seneka

1 Introduction

Water losses can occur from the earth surface to the atmosphere by two distinct mechanisms: evaporation and transpiration. Evaporation is the process whereby liquid water is converted to water vapour and removed from the evaporating surface, such as lakes, rivers, pavements, soils and wet vegetation. Transpiration is the process of water loss from plants through small openings found on the underside of leaves, namely, stomata. As evaporation and transpiration can occur simultaneously, a collective term, evapotranspiration accounts for all processes through which water in liquid or solid form becomes atmospheric water vapour. Evapotranspiration is the loss of water from the earth's surface through the combined processes of evaporation and plant transpiration. Several terms often used in literature to describe specific manifestations of evaporation and evapotranspiration: potential evaporation, potential evapotranspiration (PET), lake evaporation, and areal or actual evapotranspiration (AET). Potential evaporation is the rate of evaporation, under existing atmospheric conditions, from a surface of water that is chemically pure and has the temperature of the lowest layer of the atmosphere. PET is the amount of water evaporated (both as transpiration and evaporation from the soil) from an area of continuous, uniform vegetation that covers the whole ground and that is well supplied with water. Lake evaporation is the evaporation from the surface of a lake. AET is the amount of water lost to evapotranspiration from soil-plant continuum by an actively growing plant or crop.

In Alberta, provincially-averaged annual precipitation is about 502 mm while the average runoff is only about 98 mm and the average annual groundwater recharge is about 41 mm [2]. Therefore, the average annual AET is about 373 mm, which

Z. Islam (✉) · S. Kashyap · M. Seneka
Alberta Environment and Parks, Edmonton, Canada
e-mail: Zahidul.Islam@gov.ab.ca

© Canadian Society for Civil Engineering 2022
S. Walbridge et al. (eds.), *Proceedings of the Canadian Society of Civil Engineering Annual Conference 2021*, Lecture Notes in Civil Engineering 250,
https://doi.org/10.1007/978-981-19-1065-4_44

529

is about 74% of its total precipitation. In other words, considering the simple water balance, only 26% of total precipitation falling over Alberta goes into the streams as surface and sub-surface runoff. Therefore, proper estimation of evapotranspiration is essential in water balance calculation in Alberta. The spatial distribution of evaporative losses is also quite significant. In general, PET is higher at the southeast region of Alberta. Because of fewer amounts of radiative energy available, northern Alberta shows lower PET. In contrast, as the southeast Alberta is mostly dry, it shows lower AET and highest amount of AET is estimated in south central Alberta.

In winter days, it is often possible to obtain negative values for evapotranspiration since in some locations, the longwave radiation from the surface is large and the vapour pressure deficit is small. Under these conditions, net condensation of water from the atmosphere is possible [4]. Moreover, values calculated for evapotranspiration for night-time occasionally take on negative values [5]. In some situations, negative hourly computed evapotranspiration may indicate some condensation of vapour during periods of early morning dew. In practice, negative values are sometimes set to zero before summing over 24-h period. However, in some situations, negative hourly computed evapotranspiration may be accounted as negative. Even though the impact of hourly negative values on evapotranspiration summed over daily periods is usually less than a few percent, it may be appropriate to retain the negative values [5]. However, we should pay careful consideration before deciding on any of these approaches.

With the aforementioned introduction, objectives of this study are:

- (i) Estimate monthly PET and AET from various models at different locations in Alberta.
- (ii) Compare long term averaged monthly PET and AET from different models.
- (iii) Analyze PET and AET for winter months to investigate the issue of negative evaporative losses.

2 Data and Methodology

2.1 Study Area

Alberta is one of the three Canadian Prairie provinces and is located between 49–60° N latitude and 110–120°W longitude [9]. Located at the leeward side of the Canadian Rocky, Alberta is relatively dry, with an average annual precipitation ranging from about 350 mm in the southeast to over 500 mm in the northwest and significantly higher in mountain and foothills areas [11]. The mean annual temperature ranges from 3.6 to 4.4 °C, with a winter temperature typically varying between –25.1 and –9.6 °C, and a summer temperature between 8.7 and 18.5 °C [13]. Alberta has seven major river basins: Peace/Slave, Athabasca, Hay, North Saskatchewan, South Saskatchewan, Beaver and Milk River basins. Alberta's has six natural regions: Grassland, Parkland, Foothills, Boreal Forest, Rocky Mountains

and Canadian Shield. Alberta’s largest natural region is the Boreal Forest and the smallest is the Canadian Shield [9].

2.2 Data

We used the gridded historical climate data (1955–2019) of 19 stations from the Alberta Climate Information Service (ACIS) of Alberta Agriculture and Forestry [1]. Figure 1 shows a map of Alberta with those stations. The dataset contains daily time series of minimum and maximum temperature, precipitation, solar radiation, and humidity for 6900 townships in Alberta. The Alberta Township Survey (ATS) system is a grid network dividing the province into equal-sized parcels of land and

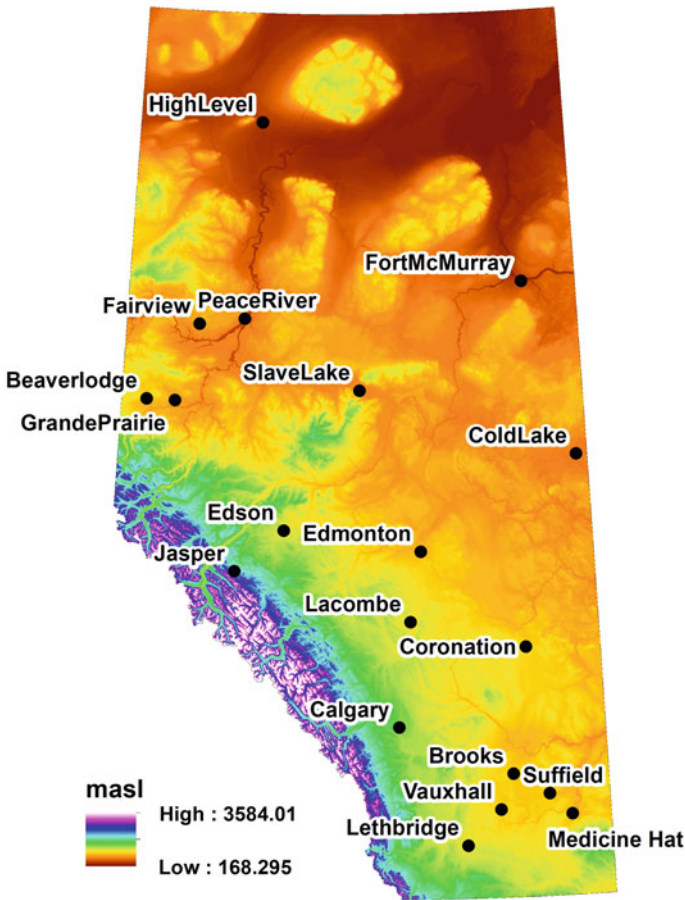


Fig. 1 Map of Alberta with selected 19 stations. The color shading demonstrates elevation band

consists of townships, ranges, and meridians. The ATS defines a land location as being west of the fourth, fifth, or sixth meridians (110° , 114° , 118° west longitude, respectively); ranges are six-mile wide columns between meridians and are numbered consecutively from east to west starting at range 1 west of each meridian. Townships, as numbered consecutively from township 1 at the Montana border to township 126 at the Northwest Territories border, are six-mile-wide rows that intersect ranges [3].

The methodology for interpolating climate data for each township location includes using a mathematical data interpolation procedure that weighted up to the eight nearest station observations. Station observations includes the raw data from Environment and Climate Change Canada, Alberta Environment and Parks, and Alberta Agriculture and Forestry. The interpolation method for precipitation is the Hybrid Inverse Distance cubed weighing (IDW) using a daily search radius out to 60 km, or a maximum of eight closest stations, whichever was satisfied first. In case no stations were available within 60 km of the township centre, interpolation process used the nearest neighbor, regardless of its distance from the township centre. For temperature, humidity and solar radiation the interpolation process includes a linear IDW procedure with a radius of 200 km or eight closest stations whichever is satisfied first. In case no stations were available within 200 km of the township centre, interpolation process used the nearest neighbor, regardless of its distance from the township centre [1].

2.3 Models

Dissipation of solar radiative energy occurs mostly as latent heat, sensible heat and energy heating or cooling the terrestrial or aquatic surface. As potential evaporation from water surface or potential evapotranspiration from terrestrial surfaces is the water equivalent of the latent heat, ideally, we can estimate these by solving the energy balance at water surface or terrestrial surface, respectively. However, the calculation of latent heat in the energy balance is quite complex, although mathematically achievable. To overcome the challenge of solving the energy balance that requires rigorous field observations, scientists and practitioners have developed different operational methodologies of estimating evaporation and evapotranspiration. We can generally group these methods into four categories: energy-based, vapour transfer-based, combination, and complementary relationship-based methods. Energy-based methods use the radiative energy, or the air temperature, or both to estimate the evaporation from water surface or evapotranspiration from terrestrial surfaces. Vapour transfer-based methods use the difference between the actual vapour pressure of evaporative surface and the actual vapour pressure of air to estimate evaporation and evapotranspiration. Combination methods use both the energy budget and vapour transfer to estimate evaporation and evapotranspiration. Complementary relationship-based methods use the complementary relationship between potential and actual relationship estimate evaporation and evapotranspiration.

In this study, we use five evapotranspiration models: Hamon Method, Penman Method, Penman-Monteith FAO-56 Method, Morton’s Complementary Relationship Areal Evapotranspiration Model, and Granger and Gray Method.

The Hamon Method [8] is an energy-based (temperature) method that predicts potential evapotranspiration. We used a version of the method as presented by [14]. They tested the equation on 679 gauged watersheds in the United States.

$$ET = \frac{29 * 8 * Hr * e_s}{(T_a + 273.2)} \tag{1}$$

where, T_a = mean air temperature (°C), Hr = the number of daylight hours (hours), and e_s = the saturation vapour pressure.

The Penman Method [12] is a combination method and estimates open-water potential evaporation.

$$ET = \frac{\Delta}{(\Delta + \gamma)} \left(\frac{R_n}{\lambda} \right) + \frac{(\gamma)}{(\Delta + \gamma)} E_a \tag{2}$$

where, Δ = slope vapour pressure curve, R_n = net radiation, γ = psychrometric constant, and E_a = relative drying power of air.

The Penman–Monteith FAO-56 Method [4] is a combination method that estimates evapotranspiration for a specific reference crop with an assumed height of 0.12 m and a surface resistance of 70 s/m, and having an albedo of 0.23 [6].

$$ET_o = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T+273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \tag{3}$$

where, Δ = slope vapour pressure curve, R_n = net radiation, G = ground heat flux, u_2 = wind speed at 2 m height (m/s), $(e_s - e_a)$ = vapour pressure deficit of the air, γ = psychrometric constant, and T = air temperature (°C).

Morton’s Model (Morton 1983) is a complementary relationship-based method that can estimate potential evaporation, potential evapotranspiration, lake evaporation, and areal or actual evapotranspiration.

$$E_T = 2E_{TW} - E_{TP} \tag{4}$$

where, E_{TP} = Morton’s Potential Evapotranspiration which is derived by solving both energy balance and vapour transfer equations and E_{TW} = Wet area evapotranspiration (i.e., the evapotranspiration that would occur if the soil–plant surface of the area were saturated).

The Granger Gray Method [7] is a complementary relationship-based method estimating actual areal evapotranspiration.

$$ET_{areal} = \frac{\Delta G_g}{(\Delta G_g + \gamma)} \frac{(R_n - G)}{\lambda} + \frac{\gamma G_g}{\Delta G_g + \gamma} E_a \tag{5}$$

where, Δ = slope vapour pressure curve, G = ground heat flux, R_n = net radiation, γ = psychrometric constant, and E_a = relative drying power of air, and G_g = dimensionless relative evapotranspiration.

2.4 Methodology

We adopted the following methodology in this study:

First, run the evapotranspiration models using the daily climate data for 19 locations and simulate PET and AET in daily or monthly scale. Note: Hamon, Penman, Penman-Monteith FAO-56, and Morton methods can estimate PET; Morton and Granger Gray methods can estimate AET.

Second, aggregate the simulated PET and AET into long-term averaged (1955–2019) monthly values.

Third, compare the long-term averaged PET and AET in Alberta.

Finally, compare the long-term averaged PET and AET for winter months.

3 Results and Discussion

3.1 Potential Evapotranspiration (PET)

We summarized simulated monthly PET from four models (Hamon, Penman, Penman-Monteith FAO-56, and Morton) at 19 stations in Table 1. These monthly values represent the minimum, maximum and median PET for a period of 65 years (1955–2019) based on those locations. Figure 2 shows a bar chart of median values of PET in Alberta, with error bar showing the ranges of PET from the median values.

We found that with the exception of the Hamon method, all PET methods could predict negative numbers during the winter months. McMahon et al. [10] describe that such equations may perform better or worse under different environments. While the Morton's method gave a greater variability for monthly median values, it is somewhat more complex and requires more input data compared to the other methods and has the potential for application under greater conditions compared to the other equations [10].

3.2 Actual Evapotranspiration (AET)

We summarized simulated monthly AET from two models (Morton and Granger Gray) at 19 stations in Table 2. These monthly values represent the minimum, maximum and median AET for a period of 65 years (1955–2019) based on those

Table 1 Comparison of simulated potential evapotranspiration in Alberta (in mm)

| Month | Min | | | | | Max | | | | | Median | | | | |
|-------|-------|--------|-----|--------|-------|-------|--------|-----|--------|--------|--------|--------|-----|--------|--|
| | Hamon | Penman | PM* | Morton | Hamon | Hamon | Penman | PM* | Morton | Morton | Hamon | Penman | PM* | Morton | |
| Jan | 4 | 0 | 3 | -5 | 13 | 15 | 19 | 4 | 8 | 3 | 6 | -4 | | | |
| Feb | 6 | 4 | 5 | -2 | 16 | 25 | 25 | 12 | 11 | 11 | 12 | 1 | | | |
| Mar | 15 | 23 | 22 | 12 | 28 | 49 | 45 | 55 | 20 | 31 | 29 | 24 | | | |
| Apr | 37 | 65 | 59 | 81 | 47 | 106 | 92 | 114 | 39 | 75 | 67 | 91 | | | |
| May | 59 | 109 | 97 | 138 | 77 | 148 | 132 | 174 | 71 | 130 | 117 | 157 | | | |
| Jun | 75 | 114 | 102 | 144 | 101 | 165 | 150 | 195 | 91 | 136 | 122 | 169 | | | |
| Jul | 88 | 111 | 101 | 144 | 124 | 187 | 177 | 227 | 103 | 130 | 119 | 168 | | | |
| Aug | 76 | 85 | 78 | 113 | 109 | 156 | 155 | 203 | 87 | 102 | 96 | 138 | | | |
| Sep | 48 | 42 | 42 | 62 | 66 | 104 | 102 | 124 | 53 | 57 | 56 | 81 | | | |
| Oct | 24 | 12 | 16 | 11 | 39 | 70 | 71 | 52 | 30 | 27 | 30 | 28 | | | |
| Nov | 8 | -1 | 2 | -5 | 19 | 23 | 27 | 13 | 13 | 3 | 6 | 0 | | | |
| Dec | 4 | -4 | -1 | -6 | 14 | 14 | 19 | 4 | 8 | -2 | 3 | -5 | | | |

* PM = Penman Monteith (FAO-56)

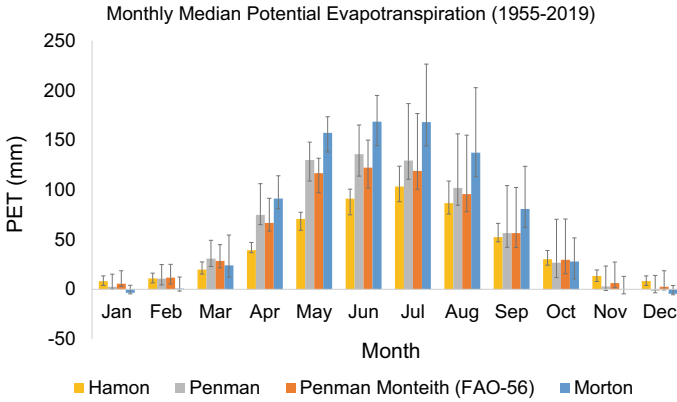


Fig. 2 Comparison of long-term averaged (1955–2019) monthly median potential evapotranspiration in Alberta simulated by four models: Hamon, Penman, Penman-Monteith FAO-56 (PM), and Morton. The error bar shows ranges of potential evapotranspiration from the median values

Table 2 Comparison of actual evapotranspiration in Alberta (in mm)

| Month | Min | | Max | | Median | |
|-------|--------|--------------|--------|--------------|--------|--------------|
| | Morton | Granger gray | Morton | Granger gray | Morton | Granger gray |
| Jan | −5 | −1 | 2 | 2 | −4 | 0 |
| Feb | −2 | 1 | 9 | 5 | 1 | 3 |
| Mar | 9 | 12 | 24 | 22 | 15 | 18 |
| Apr | 30 | 35 | 48 | 48 | 37 | 43 |
| May | 57 | 66 | 82 | 77 | 67 | 74 |
| Jun | 85 | 79 | 107 | 90 | 95 | 85 |
| Jul | 86 | 81 | 120 | 97 | 101 | 89 |
| Aug | 49 | 57 | 87 | 73 | 67 | 66 |
| Sep | 16 | 19 | 35 | 35 | 22 | 30 |
| Oct | 9 | 0 | 18 | 11 | 13 | 5 |
| Nov | −5 | −2 | 6 | 1 | 0 | −1 |
| Dec | −6 | −4 | 1 | 2 | −5 | −2 |

locations. Figure 3 shows a bar chart of median values of AET in Alberta, with error bar showing the ranges of AET from the median values.

Here we again observe the Morton method shows greater variability in median monthly values. Scientists have reported both these equations successful for estimating water balances. Morton’s model however shows strength in predicting lake evaporation [10].

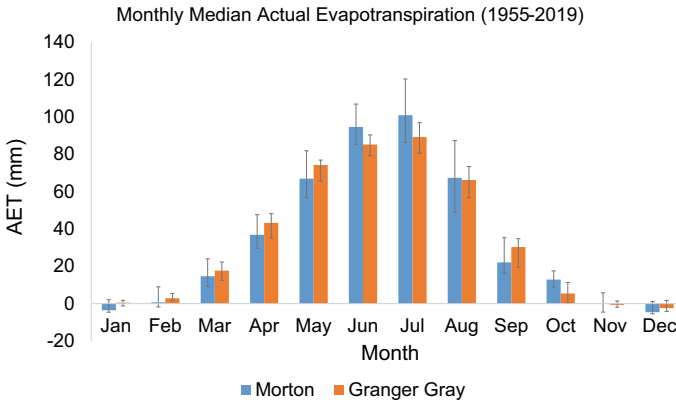


Fig. 3 Comparison of long-term averaged (1955–2019) monthly median actual evapotranspiration in Alberta simulated by two models: Morton and Granger Gray. The error bar shows ranges of actual evapotranspiration from median values

3.3 Negative Evaporative Losses During Winter Months

As shown in Figs. 2 and 3, a few models simulated negative evaporative losses in winter months, especially in January and December. Table 3 and Fig. 4 shows January and December PET from four different models in 19 locations in Alberta. Table 4 and Fig. 5 shows January and December AET from two different models for the same locations.

We observed the Morton model consistently provides negative PET values for January and December in most of the locations, except Lethbridge and Vauxhall. January and December PET can be as low as -5 mm and -6 mm, respectively. The Penman Monteith model does not provide any negative PET for any of those stations for January; however, December PET at Calgary and High Level was slightly negative. The Penman model does not provide any negative PET for any of the stations in January; however, for some stations (about 10) December PET values were estimated as negative. Hamon model does not provide any negative PET.

For AET, we only compared Morton and Granger Gray methods. Morton’s model consistently provides negative AET values for January and December in most of the location, except Lethbridge. January and December AET can be as low as -5 mm and -6 mm, respectively. The Granger Gray method estimated negative AET values for 8 stations in January and 13 stations in December.

Table 3 Comparison of simulated potential evapotranspiration in Alberta during winter (in mm)

| Station | January | | | | December | | | |
|----------------|---------|--------|-------|--------|----------|--------|-------|--------|
| | Hamon | Penman | PM* | Morton | Hamon | Penman | PM* | Morton |
| Beaverlodge | 8.13 | 0.20 | 3.45 | -3.63 | 7.99 | -1.60 | 2.59 | -4.60 |
| Brooks | 10.22 | 12.15 | 14.01 | -1.36 | 10.64 | 10.94 | 14.90 | -1.48 |
| Calgary | 6.27 | 1.57 | 3.84 | 3.56 | 6.45 | -3.11 | -0.74 | 2.78 |
| Cold Lake | 11.93 | 13.97 | 18.68 | -3.59 | 11.70 | 12.95 | 18.68 | -4.72 |
| Coronation | 7.84 | 5.84 | 7.93 | -4.48 | 8.23 | 0.60 | 4.19 | -5.26 |
| Edmonton | 8.14 | 2.57 | 5.60 | -3.33 | 8.25 | 0.09 | 4.20 | -4.09 |
| Edson | 9.17 | 2.99 | 6.36 | -4.00 | 8.60 | -1.83 | 1.80 | -4.81 |
| Fairview | 6.68 | 1.28 | 4.57 | -4.38 | 6.68 | -2.94 | 1.05 | -5.29 |
| Fort McMurray | 5.17 | 0.46 | 3.26 | -4.07 | 5.41 | -2.25 | 1.28 | -5.20 |
| Grande Prairie | 7.52 | 0.08 | 3.05 | -4.33 | 7.48 | -1.51 | 2.39 | -5.16 |
| High Level | 3.90 | 0.56 | 3.12 | -3.28 | 3.97 | -3.59 | -0.64 | -4.59 |
| Jasper | 9.98 | 1.04 | 4.57 | -3.90 | 9.49 | -3.59 | 0.11 | -5.03 |
| Lacombe | 8.91 | 5.84 | 8.99 | -3.02 | 8.99 | 1.10 | 5.54 | -3.64 |
| Lethbridge | 13.44 | 10.65 | 15.00 | 4.06 | 13.51 | 7.63 | 12.85 | 3.79 |
| Medicine Hat | 11.13 | 11.00 | 13.77 | -0.40 | 11.68 | 9.68 | 14.47 | -0.30 |
| Peace River | 6.32 | 0.97 | 4.35 | -4.35 | 6.40 | -2.88 | 1.18 | -5.47 |
| Slave Lake | 7.00 | 1.08 | 4.25 | -4.64 | 7.00 | -2.87 | 0.72 | -5.52 |
| Suffield | 11.61 | 15.20 | 17.07 | -1.03 | 11.98 | 13.90 | 18.06 | -1.08 |
| Vauxhall | 11.64 | 14.35 | 16.59 | 0.44 | 12.02 | 13.02 | 17.56 | 0.48 |

* PM = Penman Monteith (FAO-56)

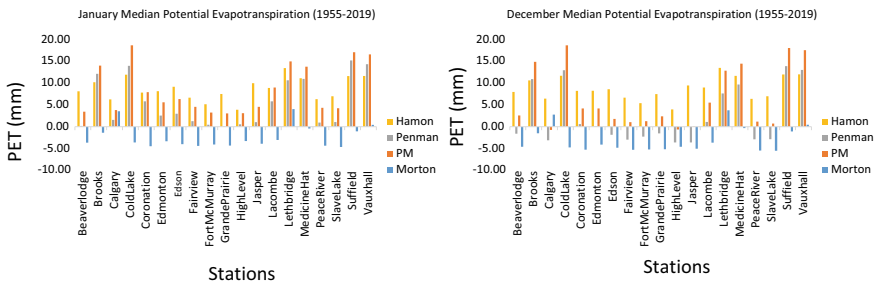


Fig. 4 Comparison of long-term averaged (1955–2019) January and December median potential evapotranspiration at 19 locations in Alberta simulated by four models: Hamon, Penman, Penman-Monteith FAO-56 (PM), and Morton

Table 4 Comparison of simulated actual evapotranspiration in Alberta during winter (in mm)

| Station | January | | December | |
|----------------|---------|--------------|----------|--------------|
| | Morton | Granger gray | Morton | Granger gray |
| Beaverlodge | -3.68 | -1.26 | -4.60 | -2.68 |
| Brooks | -1.60 | 1.48 | -1.77 | 1.48 |
| Calgary | 2.07 | 1.64 | 1.04 | -3.46 |
| Cold Lake | -3.59 | 1.35 | -4.72 | 1.72 |
| Coronation | -4.48 | 1.42 | -5.26 | -0.79 |
| Edmonton | -3.33 | 0.44 | -4.12 | -1.02 |
| Edson | -4.00 | -0.26 | -4.81 | -2.56 |
| Fairview | -4.38 | -0.23 | -5.29 | -3.47 |
| Fort McMurray | -4.07 | -0.18 | -5.20 | -2.69 |
| Grande Prairie | -4.33 | -0.97 | -5.16 | -2.39 |
| High Level | -3.28 | 0.17 | -4.59 | -3.87 |
| Jasper | -3.90 | -1.18 | -5.03 | -4.18 |
| Lacombe | -3.02 | 1.06 | -3.67 | -0.37 |
| Lethbridge | 2.08 | 0.82 | 1.10 | 0.18 |
| Medicine Hat | -0.55 | 1.77 | -0.79 | 1.32 |
| Peace River | -4.35 | 0.03 | -5.47 | -3.33 |
| Slave Lake | -4.64 | -0.34 | -5.52 | -3.55 |
| Suffield | -1.12 | 1.59 | -1.44 | 1.34 |
| Vauxhall | -0.31 | 1.41 | -0.53 | 1.28 |

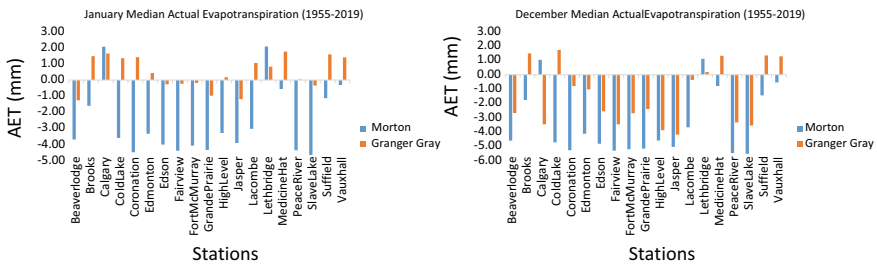


Fig. 5 Comparison of long-term averaged (1955–2019) January and December median actual evapotranspiration at 19 locations in Alberta simulated by two models: Morton and Granger Gray

3.4 Analysis of Negative Evaporative Losses in the Morton Model

Since Morton model consistently provides negative AET and PET for January and December months, we further investigate the model equations to see what is causing

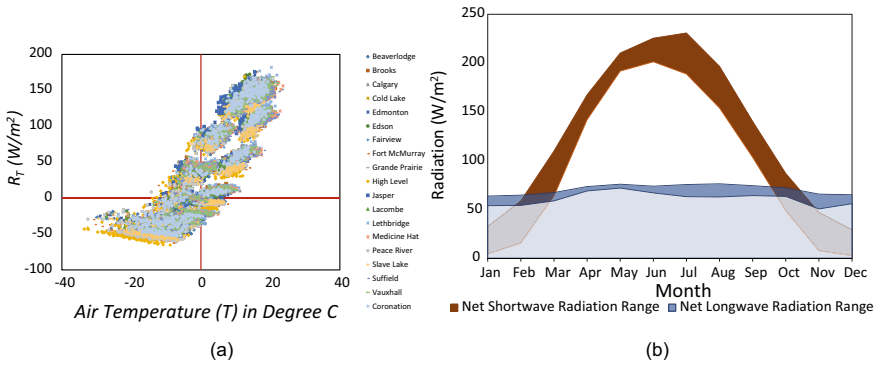


Fig. 6 **a** Monthly net radiation versus air temperature ($^{\circ}C$) and **b** Net shortwave and Net Longwave radiation in Alberta based on 65 years (1955–2019) of data

such values. As shown in Eq. 4, Morton’s model estimates PET by solving both energy balance and vapour transfer equations. Moreover, the Morton model estimates AET from the PET and wet area evapotranspiration. After analysis using the model equations, it is evident the only reason PET or AET in Morton’s model can result in negative numbers is if the net radiation is negative.

Figure 6a shows the net radiation as estimated from the Morton model versus air temperature plot, based on the climate data from 19 locations in Alberta for the 65 years (1955–2019). It is evident that during winter months, when air temperature becomes negative, net radiation also becomes negative. Morton’s model estimates net radiation by subtracting the net longwave radiation from net shortwave radiation. Figure 6b shows a monthly plot of net shortwave and net longwave radiation in Alberta based 65 years (1955–2019) of data at 19 locations. The orange shaded area shows the range of net shortwave radiation, estimated based on albedo and incident solar radiation. The blue shaded area shows net longwave radiation. The average net longwave radiation and shortwave radiation for December-January months, estimated for 65 years of period, range from 54–65 W/m^2 , to 3–33 W/m^2 , respectively. Therefore, it is inferred that during winter months, especially in January and December, the net longwave radiation offsets the net shortwave radiation and results in negative net radiative energy.

4 Conclusions

We conclude that Morton’s model provides negative PET during winter months in Alberta, especially in January and December. Other PET methods, for example Penman and Penman Monteith, often provide negative PET at some locations during December and January. Morton and Granger Gray methods provide negative AET at various locations in Alberta, especially in January and December. Negative PET

and AET values are more evident in the Morton model compared to other models attributed to negative net radiation during winter months. Further analysis, especially application of these models to analyze the water balance for a watershed, may be necessary to understand the implication of such negative evaporative losses. In addition, comparing these results from calculations from a numerical model using the energy balance method could be useful.

References

1. Alberta Agriculture and Forestry (2019) Alberta Climate Information Service (ACIS). January. <http://agriculture.alberta.ca/acis/>
2. Alberta Environment and Parks (2013) Evaporation and evapotranspiration in Alberta. <https://open.alberta.ca/dataset/46557167-126e-4bb3-b84f-615ead212b3f/resource/93686041-152d-400d-854e-b12d4d3a5481/download/8938.pdf>
3. Alberta Environment and Parks (2019) Alberta environment and parks. 1 28. <http://aep.alberta.ca/recreation-public-use/recreation-on-agricultural-public-land/alberta-township-survey-system.aspx>
4. Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration—guidelines for computing crop water requirements. FAO Irrigation & Drainage Paper 56. FAO
5. Allen RG, Walter IA, Elliott RL, Howell TA, Itenfisu D, Jensen ME, Snyder RL (2005) The ASCE standardized reference evapotranspiration equation
6. FAO (2017) Chapter 2—FAO Penman—Monteith Equation. Website: <http://www.fao.org/docrep/X0490E/x0490e06.htm>, downloaded on 1 Sept 2017
7. Granger RJ, Gray DM (1989) Evaporation from natural nonsaturated surfaces. *J. Hydrology* 111:21–29
8. Hamon WR (1963) Computation of direct runoff amounts from storm rainfall. *Int Assoc Sci Hydrol Publishing* 63:52–62
9. Jiang R, Gan TY, Xie J, Wang N, Kuo C-C (2017) Historical and potential changes of precipitation and temperature of Alberta subjected to climate change impact: 1900–2100. *Theor Appl Climatol* 127(3–4):725–739
10. McMahan TA, Peel MC, Lowe L, Krikanthan R, McVicar TR (2013) Estimating actual, potential, reference crop and pan evaporation using standard meteorological data: a pragmatic synthesis
11. Mwale D, Gan TY, Devito K, Mendoza C, Silins U, Petrone R (2009) Precipitation variability and its relationship to hydrologic variability in Alberta. *Hydrological Processes Int J* 23(21):3040–3056
12. Penman HL (1948) Natural evaporation from open water, bare soil and grass. *Proceedings of the Royal Society of London. Series A Math Phys Sci* 193:120–145
13. Sekhon NS, Hassan QK, Sleep RW (2010) Evaluating potential of MODIS-based indices in determining “snow gone” stage over forest-dominant regions. *Remote Sensing* 2(5):1348–1363
14. Shaw SB, Riha SJ (2011) Assessing temperature-based PET equations under a changing climate in temperate, deciduous forests. *Hydrol Process* 25:1466–1478

Numerical Modelling of Variable Density Shallow Water Flows with Friction Term



Amine Hanini, Abdelaziz Beljadid, and Driss Ouazar

1 Introduction

Numerical modelling of the scalar transport process is of a great importance for surface water pollution risk assessment which is used in a large variety of applications for environmental protection and management [8, 15]. The coupled system of shallow water equations and solute transport model are used to predict the dynamics of both water flow and contaminant transport in rivers and coastal regions [2, 5]. In shallow flows, horizontal scales are predominant over vertical ones and the pressure is nearly hydrostatic [17]. The shallow water equations can be derived by integrating vertically the three-dimensional incompressible Navier-Stokes equations and assuming hydrostatic pressure distribution where vertical acceleration and viscous effects are neglected.

The design of efficient numerical schemes for solving the coupled model of shallow water flow and scalar transport remains major challenge which requires robust numerical techniques [13] to ensure important physical properties of schemes such as the well-balanced property [1]. Among the difficulties that may be encountered, the resulting nonlinear hyperbolic system can lead to non-smooth solutions which can exhibit singularities, and in the presence of source terms, computational

A. Hanini (✉) · A. Beljadid
Mohammed VI Polytechnic University, Ben Guerir, Morocco
e-mail: amine.hanini@um6p.ma

A. Beljadid
e-mail: abdelaziz.beljadid@um6p.ma

A. Beljadid
University of Ottawa, Ottawa, ON, Canada

D. Ouazar
Mohammadia School of Engineering, Rabat, Morocco
e-mail: ouazar@emi.ac.ma

techniques may lead to numerical oscillations caused by the imbalance between source and flux terms. For instance, the friction source term which is non-linear can lead to numerical issues [18]. The important challenge here for the proposed numerical method is to discretize properly this term. In our approach, we used the implicit techniques proposed by [19] to discretize the friction term. These techniques can be implemented explicitly, and our aim is to efficiently solve the coupled system of shallow water flow and solute transport with friction term.

Many advantages allow the finite volume method to be convenient for solving system of conservation or balance laws, for instance, this method is conservative [1, 3, 4], and deals with solutions with large gradients. One of the most successful numerical approach to solve hyperbolic PDE in the finite volume framework is due to Godunov [6] and is based on representing the solution at cell centers on both sides of each cell interface of the whole mesh by piecewise constant states. To have more accurate results, instead of using constant states, we may consider polynomial ones for the reconstruction variable in each cell. For instance, linear piecewise approximation [18] can be used to design second order accurate schemes. Thereafter, many methods have been developed such as central-upwind schemes [9, 10] which are based on characteristic information on local speeds of propagation on the two sides of cell interfaces. These numerical schemes have been used in many studies to solve the shallow water system and have the advantage of simplicity of implementation and lead to accurate results for this system.

The outline of this paper is as follows. In Sect. 2, we present the coupled system of shallow water equations and scalar transport model and the semi-discrete form of the numerical scheme used to solve this system. In Sect. 3 the well-balanced discretization of the source term due to variable bottom topography is discussed as well as the discretization of friction source term. We present numerical experiments in Sect. 4 to test the performance of the proposed scheme. Finally, we provide some concluding remarks in Sect. 5.

2 Governing Equations

In this study, we used the following coupled system of shallow water equations (SWEs) with variable density and transport model. The SWEs, also called the Saint-Venant equations can be derived from the depth-averaged three-dimensional incompressible Navier-Stokes equations [17]. To model the solute transport in water, we used the depth-averaged scalar transport equation.

$$\left\{ \begin{array}{l} \partial_t hr + \partial_x hur + \partial_y hvr = 0 \\ \partial_t hur + \partial_x (hu^2r + \frac{1}{2}gh^2r) + \partial_y huvr = -ghr\partial_x Z - gn^2(h)^{-1/3}u\sqrt{(u^2 + v^2)} \\ \partial_t hvr + \partial_y huvr + \partial_x (hv^2r + \frac{1}{2}gh^2r) = -ghr\partial_y Z - gn^2(h)^{-1/3}v\sqrt{(u^2 + v^2)} \\ \partial_t hc + \partial_x huc + \partial_y hvc = 0, \end{array} \right. \tag{1}$$

where x and y are the spatial coordinates, t is time, h is the mixture depth, u and v are the depth-averaged velocities of the flow in the x - and the y - directions, c is the scalar depth-averaged volumetric concentration, Z is the elevation of the bottom topography, g is the acceleration due to gravity. The variable r is the relative density of the mixture to that of the clean water and n is the Manning coefficient [13].

Equation 1 can be rewritten into the following vectorial form

$$\partial_t \mathbf{U} + \partial_x \mathbf{F}(\mathbf{U}) + \partial_y \mathbf{G}(\mathbf{U}) = \mathbf{S}_b + \mathbf{S}_f, \tag{2}$$

where we use the vector of variables $\mathbf{U} = (q_1, q_2, q_3, q_4)^T$, $q_1 := hr$, $q_2 := hur$, $q_3 := hvr$, and $q_4 := hc$. We denote by $w := h + Z$ the water surface elevation. The friction source term \mathbf{S}_f is expressed using the manning formula and \mathbf{S}_b is the source term due to variable bottom topography.

The fluxes in the x - and y -directions are as follows:

$$\mathbf{F}(\mathbf{U}) = \left(q_2, \frac{q_2^2}{q_1} + \frac{1}{2}gq_1(q_1 - \Delta q_4), \frac{q_2q_3}{q_1}, \frac{q_2q_4}{q_1} \right)^T$$

$$\mathbf{G}(\mathbf{U}) = \left(q_3, \frac{q_2q_3}{q_1}, \frac{q_3^2}{q_1} + \frac{1}{2}gq_1(q_1 - \Delta q_4), \frac{q_3q_4}{q_1} \right)^T.$$

The bottom topography and friction source terms are, respectively, expressed as

$$\mathbf{S}_b = (0, -g(w - Z)r\partial_x Z, -g(w - Z)r\partial_y Z, 0)^T$$

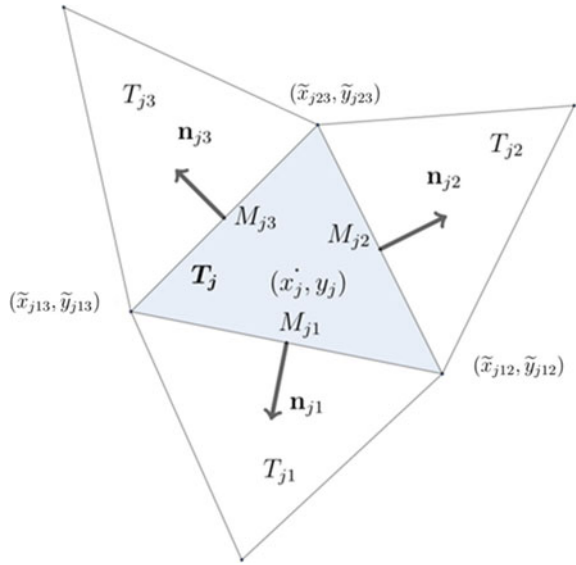
$$\mathbf{S}_f = \left(0, -gn^2(w - Z)^{-\frac{7}{3}}\frac{q_2}{r^2}\sqrt{q_2^2 + q_3^2}, -gn^2(w - Z)^{-\frac{7}{3}}\frac{q_3}{r^2}\sqrt{q_2^2 + q_3^2}, 0 \right)^T,$$

We note that in the case of small water depth $h \approx 0$, the friction source term \mathbf{S}_f becomes stiff which is very challenging in terms of numerical discretization.

The relative density of the mixture to that of the clean water is related to the volumetric concentration via the following formula

$$r = 1 + \Delta c, \tag{3}$$

Fig. 1 Typical triangular control volume and three neighbors



where $\Delta = \frac{\rho - \rho_w}{\rho_w}$ is the relative density of the constituent, ρ is its density, $\rho_w r$ is the mixture density and ρ_w is the density of water.

We used a triangulation $\mathcal{T} := \bigcup_j T_j$ of the computational domain, using non-overlapping triangular cells T_j of measure $|T_j|$. We denote by (x_j, y_j) the coordinates of the center of mass G_j of triangle T_j , $M_{jk} = (x_{jk}, y_{jk})$ is the midpoint of the k -th side of the triangle, T_{j1} , T_{j2} and T_{j3} are the three neighboring triangles sharing a common side with T_j , and $n_{jk} := (\cos(\theta_{jk}), \sin(\theta_{jk}))^T$ is the outer unit normal vector to the k -th side Γ_{jk} of T_j , of length $|\Gamma_{jk}|$, with $k = 1, 2, 3$ as shown in Fig. 1.

The variable $\bar{U}_j(t)$ represents the approximation of the cell average of the solution at time t

$$\bar{U}_j(t) \approx \frac{1}{|T_j|} \int_{|T_j|} \mathbf{U}(t, \mathbf{x}, y) dx dy.$$

We consider the semi-discrete form of numerical scheme [3], which is described by the following ODE

$$\begin{aligned} \frac{d\bar{U}_j}{dt} = & -\frac{1}{|T_j|} \sum_{k=1}^3 \frac{|\Gamma_{jk}| \cos(\theta_{jk})}{a_{jk}^{in} + a_{jk}^{out}} \left[a_{jk}^{in} \mathbf{F}(\mathbf{U}_{jk}(M_{jk})) + a_{jk}^{out} \mathbf{F}(\mathbf{U}_j(M_{jk})) \right] \\ & - \frac{1}{|T_j|} \sum_{k=1}^3 \frac{|\Gamma_{jk}| \sin(\theta_{jk})}{a_{jk}^{in} + a_{jk}^{out}} \left[a_{jk}^{in} \mathbf{G}(\mathbf{U}_{jk}(M_{jk})) + a_{jk}^{out} \mathbf{G}(\mathbf{U}_j(M_{jk})) \right] \end{aligned}$$

$$+ \frac{1}{|\mathbf{T}_j|} \sum_{k=1}^3 |\Gamma_{jk}| \frac{a_{jk}^{\text{in}} a_{jk}^{\text{out}}}{a_{jk}^{\text{in}} + a_{jk}^{\text{out}}} [\mathbf{U}_{jk}(\mathbf{M}_{jk}) - \mathbf{U}_j(\mathbf{M}_{jk})] + (\bar{\mathbf{S}}_b)_j + (\bar{\mathbf{S}}_f)_j. \quad (4)$$

The values $\mathbf{U}_j(\mathbf{M}_{jk})$ and $\mathbf{U}_{jk}(\mathbf{M}_{jk})$ of the variables of the system [1] at the midpoints of cell interfaces, respectively at the left and right side of each interface of the computational domain are approximated using the linear piecewise reconstructions $\tilde{\mathbf{U}}$

$$\mathbf{U}_j(\mathbf{M}_{jk}) = \lim_{(x,y) \rightarrow \mathbf{M}_{jk}, (x,y) \in \mathbf{T}_j} \tilde{\mathbf{U}}(x, y),$$

$$\mathbf{U}_{jk}(\mathbf{M}_{jk}) = \lim_{(x,y) \rightarrow \mathbf{M}_{jk}, (x,y) \in \mathbf{T}_{jk}} \tilde{\mathbf{U}}(x, y),$$

In this study, we used a linear reconstruction of the discrete variables of the system [3] defined by

$$\tilde{\mathbf{U}}(x, y) := \bar{\mathbf{U}}_j + (\nabla_x \mathbf{U})_j (x - x_j) + (\nabla_y \mathbf{U})_j (y - y_j). \quad (x, y) \in \mathbf{T}_j. \quad (5)$$

In Eq. 5, $\bar{\mathbf{U}}_j$ corresponds to the value $\mathbf{U}_j := \tilde{\mathbf{U}}(x_j, y_j)$ at the center of mass of the cell \mathbf{T}_j .

The directional local speeds at interfaces between two cells are approximated based on the eigenvalues of the Jacobian matrix \mathbf{J}_{jk} of the system:

$$\begin{cases} a_{jk}^{\text{in}} = -\min\{\lambda_{\min}[\mathbf{J}_{jk}(\mathbf{U}_j(\mathbf{M}_{jk}))], \lambda_{\min}[\mathbf{J}_{jk}(\mathbf{U}_{jk}(\mathbf{M}_{jk}))], 0\} \\ a_{jk}^{\text{out}} = \max\{\lambda_{\max}[\mathbf{J}_{jk}(\mathbf{U}_j(\mathbf{M}_{jk}))], \lambda_{\max}[\mathbf{J}_{jk}(\mathbf{U}_{jk}(\mathbf{M}_{jk}))], 0\} \end{cases}$$

where λ_{\min} and λ_{\max} are respectively the minimum and maximum eigenvalues of \mathbf{J}_{jk} and

$$\mathbf{J}_{jk} = \frac{\partial \mathbf{F}}{\partial \mathbf{U}} \cos(\theta_{jk}) + \frac{\partial \mathbf{G}}{\partial \mathbf{U}} \sin(\theta_{jk}).$$

The expressions of the eigenvalues and eigenvectors of the matrix \mathbf{J}_{jk} are given in [7]. For simplicity, the following normal velocities at the midpoints \mathbf{M}_{jk} are introduced

$$\mathbf{u}_j^\theta(\mathbf{M}_{jk}) = \cos(\theta_{jk}) \mathbf{u}_j(\mathbf{M}_{jk}) + \sin(\theta_{jk}) \mathbf{v}_j(\mathbf{M}_{jk}), \quad \mathbf{u}_{jk}^\theta(\mathbf{M}_{jk}) = \cos(\theta_{jk}) \mathbf{u}_{jk}(\mathbf{M}_{jk}) \\ + \sin(\theta_{jk}) \mathbf{v}_{jk}(\mathbf{M}_{jk}),$$

which are used to express the directional local speeds as follows

$$\begin{cases} a_{jk}^{in} = -\min \left\{ u_j^\theta(M_{jk}) - \tilde{c}_j(M_{jk}), u_{jk}^\theta(M_{jk}) - \tilde{c}_{jk}(M_{jk}), 0 \right\} \\ a_{jk}^{in} = -\min \left\{ u_j^\theta(M_{jk}) + \tilde{c}_j(M_{jk}), u_{jk}^\theta(M_{jk}) + \tilde{c}_{jk}(M_{jk}), 0 \right\} \end{cases}$$

where $\tilde{c}_j(M_{jk}) = \sqrt{gh_j(M_{jk})}$ and $\tilde{c}_{jk}(M_{jk}) = \sqrt{gh_{jk}(M_{jk})}$.

To improve the accuracy of the numerical method, we used a linear reconstruction for the variables of the system where the gradient of the i th component of \mathbf{U} is computed using Green’s formula

$$\nabla \mathbf{U}^i \approx \frac{1}{|T_j|} \int_{T_j} \nabla \mathbf{U}^i dx dy = \frac{1}{|T_j|} \int_{\partial T_j} \mathbf{U}^i \mathbf{n} \, d\sigma \approx 1 |T_j| \sum_{k=1}^3 |\Gamma_{jk}| \mathbf{U}^i(M_{jk}) \mathbf{n}_{jk}, \quad (6)$$

where \mathbf{n} is the unit normal vector to the boundary ∂T_j .

In Eq. 6, we choose $\mathbf{U}^i(M_{jk}) = \frac{1}{2} (\mathbf{U}_j^i + \mathbf{U}_{jk}^i)$ to obtain the approximation of the gradients of the variables in each cell of the computational domain. To avoid oscillations in numerical solutions, we introduce a minmod limiter function [14] which is applied to the components of the computed cell gradients. In the x – direction, we use the following correction:

$$\frac{\partial^{lim} \mathbf{U}_j}{\partial x} = \frac{1}{2} \left(\min_{k \in \mathcal{N}(j)} \text{sign} \left[\frac{\partial \mathbf{U}_k}{\partial x} \right] + \max_{k \in \mathcal{N}(j)} \text{sign} \left[\frac{\partial \mathbf{U}_k}{\partial x} \right] \right) \min_{k \in \mathcal{N}(j)} \text{sign} \left| \frac{\partial \mathbf{U}_k}{\partial x} \right|,$$

where $sign$ is the standard sign function, $\mathcal{N}(j)$ is the set of neighboring cells sharing common side with the cell T_j . Similar steps are used to obtain the component of the cell gradients in the y – direction $\frac{\partial^{lim} \mathbf{U}_j}{\partial y}$.

The new approximation of the cell gradient is used in Eq. 5.

3 Discretization of the Topography and Friction Source Terms

The cell average of the source term due to bottom topography $\bar{\mathbf{S}}_j \approx \frac{1}{|T_j|} \int_{T_j} \mathbf{S}(t, x, y) \, dx dy$ is discretized using the approach developed in [7] where variable density is taken into account. The semi-discrete form of the numerical scheme [3] is well-balanced and it exactly preserves the following steady-states solutions

$$h \equiv h_0 = \left(\frac{n^2 q_0^2}{r_0^3 C} \right)^{3/10} \quad q_2 \equiv q_0 \quad q_3 \equiv 0 \quad r \equiv r_0 \quad \partial_x Z = -C \quad \partial_y Z = 0 \quad c \equiv c_0, \quad (7)$$

where $q_0, r_0, C(C > 0)$ and c_0 are constants. In the following, we will prove that [6] is an exact solution of the ODE [3].

We note that since we are dealing with stationary solutions, the term in the left-hand side of Eq. 4 is zero. From Eq. 7 $q_3 \equiv 0$, then all the components of the flux \mathbf{G} introduced in Sect. 2 vanish and consequently the second term of the right-hand side of Eq. 4 is zero. For the steady state solution [6] we have $\mathbf{U}_{jk}(M_{jk}) = \mathbf{U}_j(M_{jk})$, then the third term on the right-hand side of Eq. 4 vanishes. Finally, Eq. 4 becomes

$$0 = -\frac{1}{|T_j|} \sum_{k=1}^3 \frac{|\Gamma_{jk}| \cos(\theta_{jk})}{a_{jk}^{in} + a_{jk}^{out}} \left[(a_{jk}^{in} + a_{jk}^{out}) \mathbf{F}(\mathbf{U}_j(M_{jk})) \right] + (\bar{\mathbf{S}}_b)_j + (\bar{\mathbf{S}}_f)_j,$$

which gives the following formulation

$$0 = -\frac{1}{|T_j|} \sum_{k=1}^3 |\Gamma_{jk}| \cos(\theta_{jk}) \left[\frac{q_0^2}{q_1} + \frac{1}{2} g h_0^2 r_0 \right] + \frac{g}{2|T_j|} r_0 \sum_{k=1}^3 |\Gamma_{jk}| \cos(\theta_{jk}) h_0^2 + g r_0 h_0 C - g n^2 (h_0)^{-\frac{7}{3}} \frac{q_0}{r_0^2} |q_0|$$

then by simplification, we obtain

$$0 = -\frac{1}{|T_j|} \sum_{k=1}^3 |\Gamma_{jk}| \cos(\theta_{jk}) \left[\frac{q_2^2}{q_1} \right] + g r_0 h_0 C - g n^2 (h_0)^{-\frac{7}{3}} \frac{q_0}{r_0^2} |q_0|$$

This leads to

$$0 = -\frac{1}{|T_j|} \sum_{k=1}^3 |\Gamma_{jk}| \cos(\theta_{jk}) \left[\frac{q_2^2}{q_1} \right] = -\frac{1}{|T_j|} \left[\frac{q_2^2}{q_1} \right] \sum_{k=1}^3 |\Gamma_{jk}| \cos(\theta_{jk})$$

This equation is valid since in every triangle the relation $\sum_{k=1}^3 |\Gamma_{jk}| \cos(\theta_{jk}) = 0$ is verified, which can be obtained using divergence theorem.

Following the same steps for the second momentum equation in Eq. 4, one obtains

$$0 = -\frac{1}{|T_j|} \sum_{k=1}^3 |\Gamma_{jk}| \sin(\theta_{jk}) \left[\frac{q_3^2}{q_1} \right].$$

The friction source term \mathbf{S}_f is discretized using the approach developed in [19] which allows us to obtain a semi-implicit scheme. The obtained system is solved explicitly to avoid the use of computationally expensive iterative methods.

4 Numerical Examples

Here, we perform two numerical examples to test the performance of the numerical scheme. A first-order forward Euler method for temporal discretization is adopted with the following stability CFL condition

$$\Delta t \leq \frac{1}{6} \frac{\min\{H_{jk}\}}{\max\{a_{jk}^{out}\}},$$

where H_{jk} are the heights of the triangle T_j . In all numerical tests the constant of gravitation is $g = 9.81.m.s^{-2}$ and $CFL = 0.8$.

In the case when the water depth is very small or even zero, to calculate the velocities u, v and depth averaged concentration c we used the desingularisation methodology detailed in [11].

4.1 Example 1: Steady Flow Over a Slanted Surface

This example has been adopted to verify the ability of our numerical scheme to preserve the steady state solutions [6] in the presence of variable topography as shown in Fig. 2. We consider a computational domain $[0, 10] \times [0, 5]$ discretized with a non-uniform triangular cells with a cell average area 2.10^{-3} subject to the constant initial data [6] with $h_0 = 0.25119m, q_0 = 0.1m^2.s^{-1}, n = 0.1m^{-1/3}.s, C = -0.01, \Delta = 0.1$ wich gives $r = 1.1$. Wall boundary conditions are set in the y - direction while outflow type-condition are set upstream and downstream of the domain. The numerical simulations are performed until $t = 100s$.

Table 1 shows the computed errors and Fig. 3 presents the computed solution for the water surface level w, x - discharge q_2 , concentration c and x - velocity u at time

Fig. 2 Bottom and water surface level for example 1

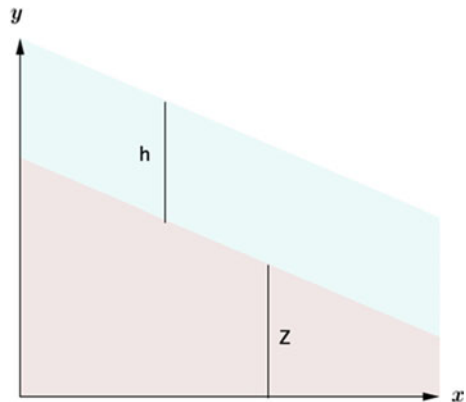


Table 1 Errors in computing the steady states [6] for Example 1

| | L_1 - error | L_2 - error | L_∞ - error |
|-------|-----------------------|-----------------------|-----------------------|
| h | 1.99×10^{-4} | 2.29×10^{-4} | 2.42×10^{-4} |
| q_0 | 8.37×10^{-4} | 1.1×10^{-3} | 1.4×10^{-3} |
| c | 1.3×10^{-3} | 1.3×10^{-3} | 3.9×10^{-3} |

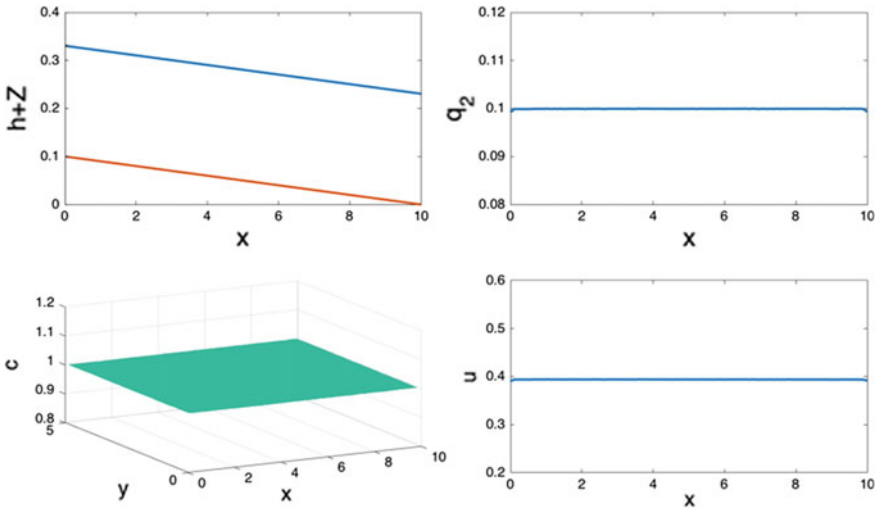


Fig. 3 Solution (w, Z, q_2, c, u) computed at time $t = 100s$

$t = 100 s$. The results show that the computed solution is free of any oscillation. The numerical solution remains stationary which confirm the robustness of the proposed numerical scheme.

4.2 Example 2: Partial Variable Density Dam-Break Over Dry Bed

In this numerical example, we consider a 2D partial dam-break with a symmetrical breach. We use the computational domain 500×300 as shown in Fig. 4. This numerical test is performed using wall-boundary condition at all sides of the computational domain which is discretized using triangular meshes with an average cell area $|T_j| = 4.68$. Initially the upstream water depth is 10 and volumetric concentration is 1, and they are both equal to 0 downstream. The relative density is $\Delta = 0.1$ which gives a density of the mixture of 1.1. The Manning roughness coefficient is set to 0.1. The snapshots of the results displaying the water height and volumetric concentration at time $t = 10000 s$ after dam failure. The numerical solution is free of any oscillation and remains symmetric. The scheme remains stable despite the

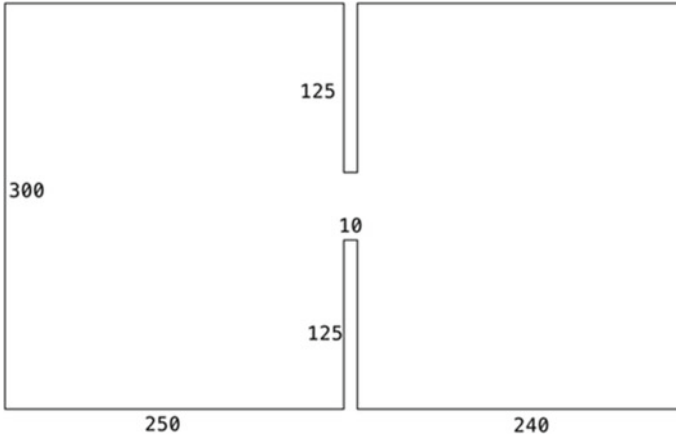


Fig. 4 Sketch of the domain for Example 2

stiffness of the friction source term for small values of the water depth. The concentration remains constant in the wet and dry region where $c = 1$ for $h > 0$, and $c = 0$ for $h = 0$ m. At the final simulation time, the concentration is constant $c = 1$ which confirms that the numerical scheme is well-balanced in terms of concentration (Fig. 5).

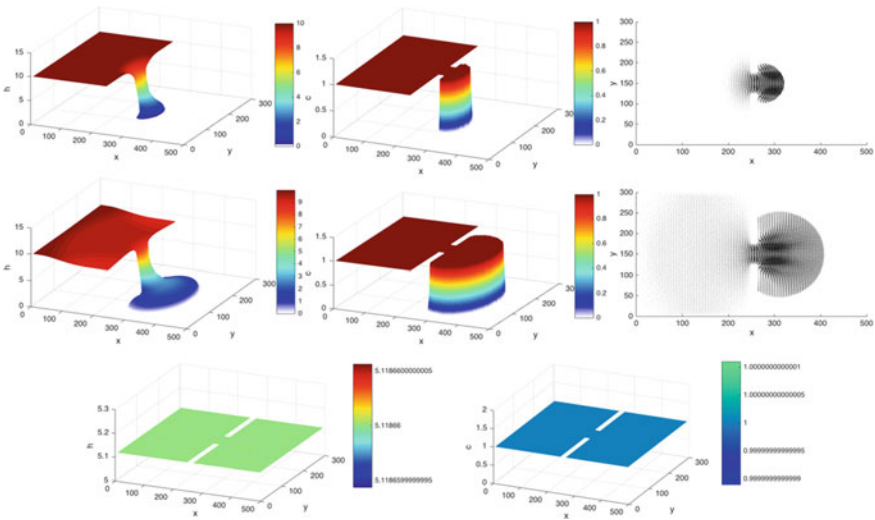


Fig. 5 Snapshots of the solution of the water height h , volumetric concentration c and velocity field at $t = 5s$ (first row), $t = 25s$ (second row) and $t = 10000s$ (long-time simulation, third row)

5 Conclusions

In this paper, we have studied a coupled 2D shallow water system and solute transport model with variable topography and friction source term. We proposed accurate numerical techniques for modelling this coupled system using triangular grids. A central-upwind scheme is used to solve the hyperbolic shallow water equation. A well-balanced discretization technique is used for the bottom topography. For accurately discretize the friction source term, we applied the techniques proposed by [19]. Our approach performs very well for nontrivial stationary flow in the presence of variable topography. The proposed scheme is well-balanced, and we prove that it exactly preserves the nontrivial steady-state solutions. Our numerical experiments demonstrate the robustness of the proposed techniques and confirm that the proposed method is stable, well-balanced and accurate in the modelling of the coupled system of shallow water equation and solute transport model with friction term.

References

1. Audusse E, Bouchut F, Bristeau M-O, Klein R, Perthame B (2004) A fast and stable well-balanced scheme with hydrostatic reconstruction for shallow water flows. *SIAM J Sci Comput* 25:2050–2065
2. Begnudelli L, Sanders BF (2006) Unstructured grid finite-volume algorithm for shallow-water flow and scalar transport with wetting and drying. *J Hydraul Eng* 132(4):371–384
3. Beljadid A, Mohammadian A, Kurganov A (2016) Well-balanced positivity preserving cell-vertex central-upwind scheme for shallow water flows. *Comput Fluids* 136:193–206
4. Beljadid A, Mohammadian A, Qiblawey H (2012) Numerical simulation of rotation dominated linear shallow water flows using finite volume methods and fourth order Adams scheme. *Comput Fluids* 62:64–70
5. Bryson S, Epshteyn Y, Kurganov A, Petrova G (2011) Well-balanced positivity preserving central-upwind scheme on triangular grids for the Saint-Venant system. *M2AN Math Model Numer Anal* 45:423–446
6. Cea L, Vazquez-Cendon ME (2012) Unstructured finite volume discretisation of bed friction and con- vective flux in solute transport models linked to the shallow water equations. *J Comput Phys* 231(8):3317–3339
7. Godunov SK (1959) Finite difference method for numerical computation of discontinuous solutions of the equations of fluid dynamics. *Math Sbornik* 47(3):271–306
8. Hanini A, Beljadid A, Ouazar D (2021) A well-balanced positivity-preserving numerical scheme for shallow water models with variable density. *Comput Fluids* 231:105156
9. Horn AL, Rueda FJ, Hormann G, Fohrer N (2004) Implementing river water quality modelling issues in mesoscale watershed models for water policy demands-an overview on current concepts, deficits, and future tasks. *Phys Chem Earth, Parts A/B/C* 29(11–12):725–737
10. Kurganov A (2018) Finite-volume schemes for shallow-water equations. *Acta Numerica* 289–351
11. Kurganov A, Levy D (2002) Central-upwind schemes for the Saint-Venant system. *ESAIM: Math Model Numeri Anal* 36(3):397–425
12. Kurganov A, Petrova G (2007) A second-order well-balanced positivity preserving scheme for the Saint-Venant system. *Commun Math Sci* 5(1):133–160
13. Liang D, Lin B, Falconer RA (2007) Simulation of rapidly varying flow using an efficient TVD–MacCormack scheme. *Int J Numer Meth Fluids* 53(5):811–826

14. Manning R (1891) On the flow of water in open channels and pipes. *Trans Inst Civil Eng Ireland* 20:161–207. Supplement 1895, 24:179–207
15. Murillo J, LatorreBorja B, Garcia-Navarro P (2012) A Riemann solver for unsteady computation of 2D shallow flows with variable density. *J Comput Phys* 231(14):4775–4807
16. Nessyahu H, Tadmor E (1990) Non-oscillatory central differencing for hyperbolic conservation laws. *J Comput Phys* 87(2):408–463
17. Rekolainen S, Kamari J, Hiltunen M, Saloranta T (2003) A conceptual framework for identifying the need and role of models in the implementation of the water framework directive. *Int J River Basin Manage* 1(4):347–352
18. Roe PL (1981) Approximate Riemann solvers, parameter vectors, and difference schemes. *J Comput Phys* 43:357–372
19. Vreugdenhil CB (1994) *Numerical methods for shallow-water flow*. Springer, Netherlands
20. Van Leer B (1979) Towards the ultimate conservative difference scheme. V. A second-order sequel to Godunov's method. *J Comput Phys* 32:101–136
21. Xia X, Liang Q, Ming X, Hou J (2017) An efficient and stable hydrodynamic model with novel source term discretization schemes for overland flow and flood simulations. *Water Resour Res* 53(5):3730–3759
22. Xia X, Liang Q (2018) A new efficient implicit scheme for discretising the stiff friction terms in the shallow water equations. *Adv Water Resour* 117:87–97

A Short-Term Flood Forecasting Model Using Markov Chain



Arpita Islam, Maysara Ghaith, Sonia Hassini, and Wael El-Dakhakhni

1 Introduction

Flood forecasting and warning have become paramount strategies in hydrometeorology by providing more preparation time preceding this event, facilitating evacuations from floodplains and impeding the menace and human casualties [16]. In river hydrodynamics, forecasting can be done either by anticipating streamflow or water level/stage [7, 10]. However, researchers have to depend only on the stage data for forecasting purposes if an absence of a flow measuring station or scarcity of data at the location of interest is found. Therefore, future water level state (based on depth) prediction is the prime concern in this paper.

Some hydrologic and hydrodynamic models such as HEC-HMS, WATFLOODTM, HEC-RAS are used for flood forecasting purposes [8, 18]. The water level prediction model and usage of a threshold water level value are also used for flood warning purposes [15, 19]. Statistical models such as least square methods [9], autoregressive moving average (ARMA) & Autoregressive Integrated Moving Average (ARIMA) models [17] were used for streamflow forecasting to facilitate flood warning purpose. Artificial neural network (ANN) approach is also applied in flood forecasting based on water level fluctuations [1, 2, 5]. However, none of the abovementioned models capture the uncertain nature of hydrologic variables.

To overcome the limitations, some studies adopted other statistical approaches such as the Gray MC model, combined ARIMA-MC model, Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models for flood forecasting purposes [3, 4, 20]. However these models can handle the inherent uncertainties and anticipate the river stage, MC model is the simplest and efficient method for predicting drought

A. Islam (✉) · M. Ghaith · S. Hassini · W. El-Dakhakhni
McMaster University, Hamilton, Canada
e-mail: islama24@mcmaster.ca

duration and flood forecasting [13, 14]. This model is one of the easiest stochastic approaches widely used for predicting future states depend only on the current states or one or more preceding states for rainfall, water level, and streamflow generation purposes [6, 11]. To the best of the authors' knowledge, no studies have been done using MC in a seasonal and monthly manner for generating states. Therefore, in this study, the first-order MC model is chosen as an efficient alternative to other statistical models to generate future water level states on a monthly and seasonal basis. This model conserves both statistical and stochastic properties of data series. Finally, the generated states will facilitate flood forecasting purposes to avoid potential risk.

2 Methodology

2.1 Study Area and Data Collection

According to Fraser Basin Council, communities located near the Fraser River have been experiencing flooding since 2012. These areas are vulnerable to flooding if the water level goes beyond 5.5 m, and the limiting condition is found on the Fraser River flood protection webpage. Due to the abovementioned reasons and data availability, two stations along the Fraser River in British Columbia are chosen as a study area. Eighteen years of continuous daily water level data of 'Fraser River at Mission' (1997–2014) and 'Fraser River at Hope' (2002–2019) are retrieved from Environment Canada to develop a monthly and seasonal markovian flood forecasting model.

2.2 Markov Chain

The probability of occurrence of the future state in the first-order MC model is independent of the previous states and relies only on the current state. Transition matrix (TM) formulation is a prerequisite to generate these future states. Five states are considered in this study based on the maximum and minimum water level value of each station, and the state range is kept at an equal interval for particular stations. TM matrix containing five states can be represented as:

$$P = \begin{bmatrix} P_{1,1} & P_{1,2} & \dots & \dots & P_{1,5} \\ P_{2,1} & P_{2,2} & \dots & \dots & P_{2,5} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ P_{5,1} & P_{5,2} & \dots & \dots & P_{5,5} \end{bmatrix} \quad (1)$$

Here $P_{1,1}$ denotes the probability of the next state to be 1 if the current state is at 1, and this is determined by estimating the frequencies of the states using the following equation:

$$P_{1,1} = \frac{n_{1,1}}{\sum_5 n_{1,5}} \tag{2}$$

The summation of transition probabilities in each row should be equal to one, which means

$$P_{1,5} = \sum_5 P_{1,1} = 1 \tag{3}$$

Though the daily river stage has high order dependence between successive days, a second order TM is also developed where the current and the previous state are used to predict future state. Cumulative TM is developed for both first and second order MC model and applied for the station ‘Fraser River at Hope.’ Similar results are observed in both cases as the future states are derived based on the current state’s observed data for the first-order MC model and both current (predicted) and previous states (observed) data for the second order. Therefore, in further analysis, only the first-order Markov model is applied as it is simpler and no statistical tests are carried out in support of high-order MC model application.

2.3 Markov Property Test

Markovian properties such as time series dependency and spatial homogeneity need to be checked to assess this model’s validity to be used in the study. In the dependency test, the null hypothesis is considered for independent successive events, and statistical α is calculated as:

$$\alpha = 2 \sum_{i,j}^m n_{ij} \ln \frac{P_{ij}}{p_j} \tag{4}$$

where, m = number of states; p_j = marginal probability.

In the homogeneity test, the null hypothesis is considered as transition probabilities of successive events are heterogeneous in nature and the test-static Υ is computed as:

$$\Upsilon = \sum_s^S \sum_{i,j}^m n_{ij}(s) \frac{(P_{ij}(s) - P_{ij})^2}{P_{ij}} \tag{5}$$

where S = number of stations.

Table 1 Statistical test results of MC model

| Test static | Fraser river at mission | Fraser river at hope |
|--|-------------------------|----------------------|
| Dependency test-static, α | 12,522.42 | 15,596.4 |
| Spatial homogeneity static, Υ | 1023.72 | |

Both test-statics α and Υ follow an asymptotic chi-square (χ^2) distribution having $(m - 1)^2$ and $(S - 1) * m * (m - 1)$ degrees of freedom (DF), respectively [12]. If the test static values are greater than the critical χ^2 value, the null hypothesis will be rejected, and it can be concluded that MC is applicable to be used in this study.

2.4 Future Stage Prediction

Cumulative probability TM is developed, and the initial state, i , is selected based on the first observed stage data. Uniform Random numbers (URN) are generated, and a comparison is made between this number and the i th row of the cumulative TM. If the URN is smaller than the cumulative probability of the next state but larger than the previous state, the next state will be taken as the future state and the whole process is repeated for generating future states.

3 Results and Discussions

3.1 Statistical Test-Results

The results of Markov property tests are shown in Table 1. The limiting χ^2 value for the dependency and homogeneity test is 26.4 with 16 DF and 36.4 with 24 DF at a 5% significance level. The results depict that, both test statics show a higher value compared to the critical χ^2 . So, it can be concluded that the null hypothesis is rejected in both cases, which means the transitions of daily water level poses first-order MC property and the property is homogeneous at both stations.

3.2 Monthly Stage Prediction

The retrieved stage dataset illustrates that the stage level inhibits a higher value at 'Fraser River at Hope' (2.56m–9.67 m) than the 'Fraser River at Mission' station (0.35m–6.33 m). Tables 2 and 3 illustrate the monthly overall prediction percentages

Table 2 Prediction of future water level states at ‘Fraser River at Mission’

| Predicted state (%) | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---------------------|------|------|------|------|------|------|------|------|------|------|------|------|
| Correct prediction | 73.8 | 88.1 | 95.2 | 73.5 | 73.0 | 88.0 | 77.8 | 85.1 | 90.4 | 80.8 | 71.3 | 81.0 |
| Over-estimation | 13.1 | 6.2 | 2.2 | 13.2 | 13.8 | 5.4 | 10.4 | 7.4 | 4.2 | 9.0 | 14.0 | 9.0 |
| Under-estimation | 13.1 | 5.8 | 2.7 | 13.3 | 13.3 | 6.7 | 11.8 | 7.5 | 5.4 | 10.2 | 14.6 | 10.0 |

Table 3 Prediction of future water level states at ‘Fraser River at Hope’

| Predicted state (%) | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---------------------|------|------|------|------|------|------|------|------|------|------|------|------|
| Correct prediction | 90.7 | 81.2 | 87.3 | 74.4 | 76.5 | 73.7 | 83.3 | 79.4 | 76.3 | 82.3 | 76.3 | 83.5 |
| Over-estimation | 4.7 | 9.3 | 5.56 | 11.5 | 11.7 | 12.2 | 8.6 | 10.2 | 11.3 | 9.1 | 10.6 | 7.7 |
| Under-estimation | 4.7 | 9.5 | 7.17 | 14.1 | 11.9 | 14.1 | 8.1 | 10.4 | 12.4 | 8.6 | 13.2 | 8.8 |

(correct, overestimation, and underestimation) of future water level states to ensure the first-order MC model’s accuracy. The results show that the model can predict future states with an accuracy ranging from 71.3% to 95.2% for both stations. The overestimation and underestimation percentages are almost equal in both stations, which is 9.5% approximately.

Based on the observed datasets of both stations, the flood occurs from May to July in both stations. Therefore, accurate prediction of future higher stage levels is necessary to forecast flood warnings. Figures 1 and 2 depict the accuracy of predicting future states (%) for both stations.

In the MC model, only state 5 represents flood occurrence for ‘Fraser River at Mission’ whereas state 2 to state 5 indicates this hazardous event during this period

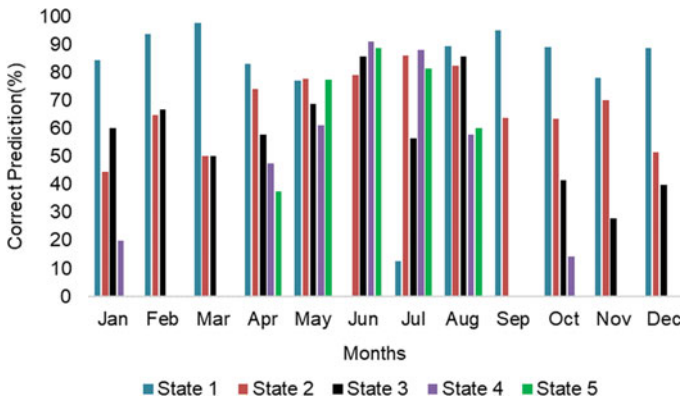


Fig. 1 Accurate prediction (%) of water stage states at ‘Fraser River at Mission’

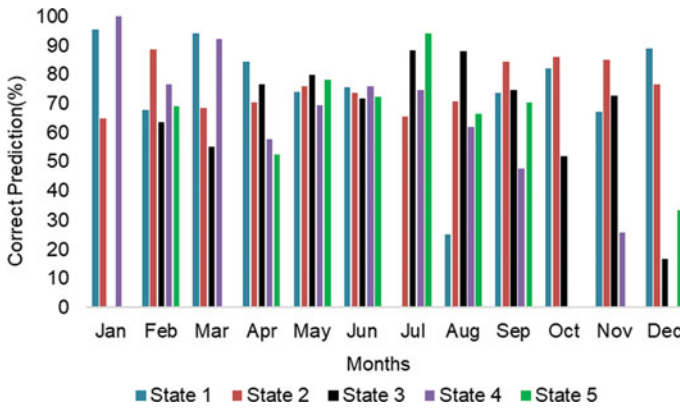


Fig. 2 Accurate prediction (%) of water stage states at ‘Fraser River at Hope’

for the other station. The results show that this model can provide flood warning (stage > 5.5 m) with an accuracy ranging from 77.36%–88.80% at ‘Fraser River at Mission’ and 69.44%–94.23% in the other station. However, the MC model cannot provide better flood forecasting in May and June for ‘Mission and ‘Hope’ respectively due to the presence of larger underestimation percentages, which are 13.3% and 14.1%, respectively.

3.3 Seasonal Stage Prediction

To further evaluate the Markovian model’s relevance to generate higher water level states, element frequencies for both observed and predicted data series are compared on a seasonal basis and shown in Figs. 3 and 4. Based on the flooding water level criteria and timing of its occurrence, it can be said that the period (summer to spring) is the most vulnerable period for this hazardous eventuality. The following two figures show that the MC model can forecast flood warnings with an accuracy varying from 53.77% to 62.50% at ‘Fraser River at Mission,’ whereas for the other station, it lies between 79.43% and 92.40%. Though, the frequency of both predicted and observed water level states shows a similarity during the flooding seasons; predicting the flooding states is low during the winter and fall seasons, as the probability of other states is more dominant.

To better comprehend the results, the correct prediction percentage is calculated in conjunction with overestimation and underestimation percentages for both stations and shown in Table 4. The analysis reveals that the model accuracy lies between 82.83% and 92.44% at ‘Fraser River at hope’ while it ranges from 77.23% to 81.39% at ‘Fraser river at Mission.’ The overestimation and underestimation percentage of future states are almost equal in both stations and seasons. However, during the spring and summer seasons, the model shows a large underestimation percentage at ‘Fraser

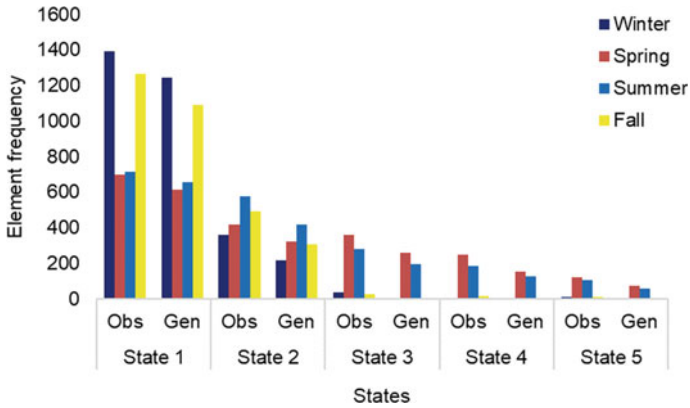


Fig. 3 Frequency of seasonal transition matrix elements at 'Fraser River at Mission'

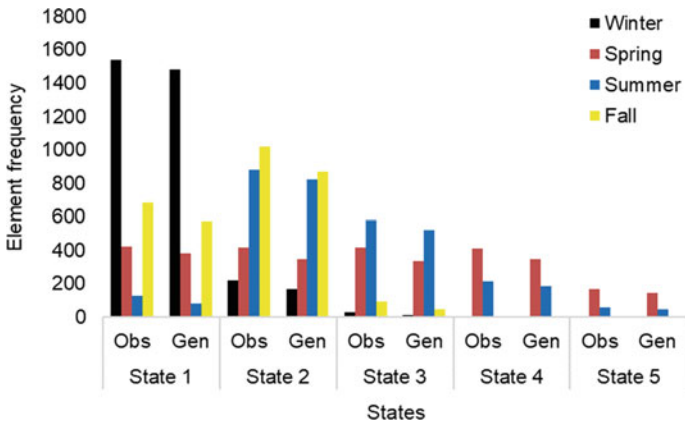


Fig. 4 Frequency of seasonal transition matrix elements at 'Fraser River at Hope'

Table 4 Analysis of correct prediction of future states (seasonal basis)

| Stations | Season | Predicted state (%) | | |
|-------------------------|--------|---------------------|---------------|----------------|
| | | Correct prediction | Overestimated | Underestimated |
| Fraser River at Hope | Winter | 92.44 | 4.00 | 3.56 |
| | Spring | 84.73 | 6.90 | 8.37 |
| | Summer | 89.03 | 5.65 | 5.32 |
| | Fall | 82.83 | 8.5 | 8.67 |
| Fraser River at Mission | Winter | 81.39 | 9.00 | 9.61 |
| | Spring | 77.23 | 7.39 | 15.38 |
| | Summer | 77.90 | 5.65 | 16.45 |
| | Fall | 77.72 | 10.72 | 11.56 |

river at Mission,' which is 15.38% and 16.45%. The analysis outcome exhibits a risk of flooding without a previous warning at this station if the spring and summer MC model is used. So, it can be concluded that the MC model for 'Hope' station can be used for all seasons while it can be used only during the winter and fall season for 'Mission station.

4 Conclusion

Flood is one of the natural jeopardies in Canada, impacting the livelihood of people and the economy. Hydrometric datasets such as rainfall, streamflow, and water-level help forecast vulnerable events such as floods and drought. For the locations, inhibit the absence of streamflow measuring station or lack of long period flow data; synthetic water level state generation can be beneficial. For this purpose, a first-order MC model is used in this study to generate future water level states on a monthly and seasonal basis to facilitate short-term flood forecasting. Eighteen years of stage data is used for 'Fraser River at Mission' and 'Fraser River at Hope' to prosecute the analysis. The statistical test result of the Markovian property also shows that the time series datasets are dependent on each other and inhibit homogeneity nature in both stations. Anticipating future state accuracy is evaluated by examining and comparing the percentages of correct, underestimated, and overestimated predictions for both stations. The analysis reveals that the prediction accuracy varies between 71.3% and 95% on a monthly and seasonal basis for both stations. An accurate forecast of individual states is also conducted for both stations to validate the model's accuracy in predicting higher states. The analysis shows that this model can provide flood warnings in a monthly basis with an accuracy ranging from 77.36% to 88.80% at 'Mission' and 69.44%–94.23% for the other station. However, the larger underestimation percentages on May (13.3%) at 'Mission' and June (14.1%) at 'Hope' indicate a risk of flood occurrence if the corresponding month's MC model is used. Based on this concept, it is also found that the spring and summer MC model of the 'Mission' station can not predict flooding states precisely. In contrast, the seasonal MC model of the 'Hope' station is suitable for generating both flooding and other states with an accuracy ranging from 82.83% to 92.44%. Although the prediction accuracy covers a wide range, the first-order MC model can be used as a preeminent alternative to generate future stage states at both stations and facilitate short-term flood warnings. In the future, the transition matrix can be updated incorporating climate change impacts, and also, by using the probability distribution of observed datasets, actual stage values can be determined.

References

1. Campolo M, Soldati A, Andreussi P (2003) Artificial neural network approach to flood forecasting in the River Arno. *Hydrol Sci J* 48(3):381–398
2. Coulibaly P, Anctil F, Aravena R, Bobée B (2001) Artificial neural network modeling of water table depth fluctuations. *Water Resour Res* 37(4):885–896
3. Dong S, Chi K, Zhang Q, Zhang X (2012) The application of a Grey Markov Model to forecasting annual maximum water levels at hydrological stations. *Journal of ocean university of China* 11(1):13–17
4. Du H, Zhao Z, Xue H (2020) ARIMA-M: a new model for daily water consumption prediction based on the autoregressive integrated moving average model and the Markov chain error correction. *Water* 12(3):760
5. Elsafi SH (2014) Artificial neural networks (ANNs) for flood forecasting at Dongola Station in the River Nile, Sudan. *Alex Eng J* 53(3):655–662
6. Fadhil RM, Rowshon MK, Ahmad D, Fikri A, Aimrun W (2016) A stochastic rainfall generator model for simulation of daily rainfall events in Kurau catchment: model testing. In: III International conference on agricultural and food engineering, ISHS, Budapest, Hungary, vol 1152, pp 1–10
7. Givati A, Fredj E, Silver M (2016) Operational flood forecasting in Israel. In: *Flood forecasting*. Academic Press, pp 153–167
8. Hostache R, Lai X, Monnier J, Puech C (2010) Assimilation of spatially distributed water levels into a shallow-water flood model. Part II: Use of a remote sensing image of Mosel River. *J Hydrol* 390(3–4):257–268
9. Noor MSFM, Sidek LM, Basri H, Husni MMM, Jaafar AS, Kamaluddin MH et al (2016) Development of flood forecasting using statistical method in four River Basins in Terengganu, Malaysia. In: *Earth and Environmental Science*, IOP, Orlando, Florida, USA, vol 32, p 012023
10. Palash W, Jiang Y, Akanda AS, Small DL, Nozari A, Islam S (2018) A streamflow and water level forecasting model for the Ganges, Brahmaputra, and Meghna Rivers with requisite simplicity. *J Hydrometeorology* 19(1):201–225
11. Pereira AGC, Sousa FAS, Andrade BBD, Campos VSM (2018) Higher order Markov chain model for synthetic generation of daily streamflows. *TEMA (São Carlos)* 19(3):449–464
12. Shamshad A, Bawadi MA, Hussin WW, Majid TA, Sanusi SAM (2005) First and second order Markov chain models for synthetic generation of wind speed time series. *Energy* 30(5):693–708
13. San THH, Khin MM (2016) River flood prediction using Markov model. In: *Genetic and evolutionary computing*, vol 387. Springer, Cham, New York, Colorado, USA, pp 435–443
14. Sharma TC, Panu US (2012) Prediction of hydrological drought durations based on Markov chains: case of the Canadian prairies. *Hydrol Sci J* 57(4):705–722
15. Song Y, Park Y, Lee J, Park M, Song Y (2019) Flood forecasting and warning system structures: procedure and application to a small urban stream in South Korea. *Water* 11(8):1571–1594
16. Thiémig V, De Roo A, Gadain H (2011) Current status on flood forecasting and early warning in Africa. *Intl. J. River Basin Manage* 9(1):63–78
17. Valipour M, Banihabib ME, Behbahani SMR (2012) Parameters estimate of autoregressive moving average and autoregressive integrated moving average models and compare their ability for inflow forecasting. *J Math Stat* 8(3):330–338
18. Zahmatkesh Z, Kumar Jha S, Coulibaly P, Stadnyk T (2019) An overview of river flood forecasting procedures in Canadian watersheds. *Can Water Resour J/Revue canadienne des ressources hydriques* 44(3):213–229
19. Zhang X, Zhang Q, Zhang G, Nie Z, Gui Z (2018) A hybrid model for annual runoff time series forecasting using Elman neural network with ensemble empirical mode decomposition. *Water* 10(4):416
20. Zhao JH, Dong ZY, Zhao ML (2009) A statistical model for flood forecasting. *Australas J Water Resour* 13(1):43–52

Modelling of Unsaturated Flow Through Porous Media Using Meshless Methods



Mohamed Boujoudar, Abdelaziz Beljadid, and Ahmed Taik

1 Introduction

The process of infiltration through porous media is an important part of hydrological cycle. The modelling of this process has important practical applications in engineering such as water resources management and agriculture. Most numerical models that describe the unsaturated flow in soils use the Richards model [16] which is a highly nonlinear equation. This equation is obtained from Darcy's law and the conservation of mass [2]. The strong non-linearity of the unsaturated conductivity and the capillary pressure as functions of saturation and the presence of both advection and diffusion terms make the Richards equation more challenging in terms of numerical approximations and require the development of efficient numerical techniques. The unsaturated conductivity and capillary pressure are correlated using empirical models and experiment data such as the van Genuchten [20], Brooks-Corey [4] and Gardner [6] models. Analytical solutions of the Richards equation can only obtained for some cases with special initial and boundary conditions [7, 17–19]. Therefore, different numerical techniques are developed to efficiently solve the Richards equation such as, finite-difference, finite-element, and finite-volume

M. Boujoudar (✉) · A. Beljadid
Mohammed VI Polytechnic University, Green City, Morocco
e-mail: mohamed.boujoudar@um6p.ma

A. Beljadid
e-mail: abdelaziz.beljadid@um6p.ma

A. Beljadid
University of Ottawa, Ottawa, Canada

A. Taik
FST-Mohammedia, University Hassan II, Casablanca, Morocco
e-mail: ahmed.taik@fstm.ac.ma

methods. For instance, Celia et al. [5] used the mixed form of the Richards equation and proposed a general mass-conservative numerical scheme, and Bause and Knabner [1] developed an adaptive mixed hybrid finite element discretization for the Richards equation. Manzini and Ferraris [15] developed a mass conservative finite volume method using two-dimensional unstructured grids. Although many numerical techniques have been developed to numerically solve the Richards equation, there is still a strong need for more robust numerical techniques for modelling flows in unsaturated soils.

The aim of this work is to develop a new technique based on the localized radial basis function method and the Kirchhoff transformation in order to solve Richards equation in one and two-dimensional homogeneous medium. The proposed technique allows us to avoid mesh generation, which makes the numerical method less expensive in terms of computational cost. The use of localized meshless method has the advantage of flexibility in dealing with complex geometries [3]. The proposed method performs well in terms of accuracy and efficiency for modelling unsaturated flow through soils.

To handle the nonlinearity of the Richards equation, we use the Kirchhoff transformation which allows us to reduce the nonlinearity of the studied problem. We used Picard iterations to solve the problem with the Kirchhoff variable where we used the backward Euler method for temporal discretization. Other numerical techniques using the Kirchhoff transformation to solve the Richards equation can be found in [8]. The performance of the proposed numerical method is assessed using different test cases.

The outline of the paper is as follows. In Sect. 2, we introduce the governing equation and the proposed system using the Kirchhoff transformation. In Sect. 3, we present the proposed meshless method. Numerical simulations are performed in Sect. 4 for modelling water flow through one and two-dimensional unsaturated porous media. Finally, we provide some conclusions in Sect. 5.

2 Governing Equation

2.1 The Mathematical Model

Infiltration of water in unsaturated soils is described by the Richards equation [16] which can be derived from Darcy's law and the conservation of mass. This equation is given by:

$$\frac{\partial \theta}{\partial t} + \nabla \cdot (K \nabla h) + \frac{\partial K}{\partial z} = s(\mathbf{x}, t), \quad \mathbf{x} \in \Omega, 0 \leq t \leq T, \quad (1)$$

where $\theta[\text{L}^3/\text{L}^3]$ is the moisture content, $h[\text{L}]$ is the pressure head, $K[\text{L}/\text{T}]$ is the unsaturated hydraulic conductivity, $\mathbf{x} = (x, y, z)^T$ is the coordinate vector, $x[\text{L}]$

and $y[L]$ denote the horizontal dimensions and $z[L]$ denotes the vertical dimension positive down (coordinate in the direction of gravity) and $s(\mathbf{x}, t)$ is a source or sink term which can depend on space and time and can include evaporation, plant root extraction, etc. In this study, we assume that $s(\mathbf{x}, t) = 0$, Ω is an open set of $\mathbb{R}^d (d = 1, 2, 3)$, and T is the final simulation time.

We note that the Richards equation can be expressed using the water saturation $S = \left(\frac{\theta - \theta_r}{\theta_s - \theta_r}\right)$ and the parameter $\phi = \theta_s - \theta_r$ where θ_s and θ_r are respectively the saturated and residual moisture contents. The unsaturated hydraulic conductivity is given by:

$$K = K_s k_r, \tag{2}$$

where k_r is the water relative permeability, which accounts for the effect of partial saturation and the saturated hydraulic conductivity is as follows:

$$K_s = \frac{\rho g k}{\mu}, \tag{3}$$

where ρ is the water density, g is the gravitational acceleration, k is the intrinsic permeability of the medium, and μ is the fluid dynamic viscosity. The Richards equation can be rewritten in the following form:

$$\phi \frac{\partial S}{\partial t} + \nabla \cdot (K_s k_r \nabla h) + \frac{\partial (K_s k_r)}{\partial z} = 0, \quad \mathbf{x} \in \Omega, 0 \leq t \leq T, \tag{4}$$

Equation (4) is highly non-linear due to the nonlinearity of the hydraulic conductivity and the capillary pressure function. Constitutive relationships are available for the functions $S[L^3/L^3]$ and $K[L/T]$ based on experiment. In our study, the numerical techniques will be developed based on Eq. (4) where we will introduce the Kirchhoff transformation in order to reduce the nonlinearity of the equation.

2.2 Capillary Pressure

The pressure head can be expressed as a function of saturation in the following form:

$$h(S) = h_{cap} J(S), \tag{5}$$

where $J(S)[-]$ is a dimensionless capillary pressure function and $h_{cap}[L]$ is the capillary rise which is given by the classical Leverett scaling [13]:

$$h_{cap} \sim \frac{\gamma \cos \theta}{\rho g \sqrt{\frac{k}{\phi_p}}}, \tag{6}$$

γ is the surface tension between the fluids, θ is the contact angle and ϕ is the medium porosity.

2.3 Kirchhoff Transformation

The Kirchhoff integral transformation is defined as:

$$\varphi(h) = \int_{+\infty}^h k_r(s) ds. \tag{7}$$

By applying this transformation, we can rewrite the Richards equation using the variable φ , as explained below:

$$\nabla \cdot (K_s k_r \nabla h) = K_s \nabla^2 \varphi, \tag{8}$$

$$\frac{\partial \varphi}{\partial t} = k_r \frac{\partial h}{\partial t}, \quad \frac{\partial \varphi}{\partial z} = k_r \frac{\partial h}{\partial z}. \tag{9}$$

By transforming the derivative terms $\frac{\partial S}{\partial t}$ and $\frac{\partial}{\partial z}(K_s k_r)$ using the variable φ , we obtain:

$$\frac{\partial S}{\partial t} = \frac{\partial S}{\partial h} \frac{\partial h}{\partial t} = \left(k_r^{-1} \frac{\partial S}{\partial h} \right) \frac{\partial \varphi}{\partial t}, \tag{10}$$

$$\frac{\partial}{\partial z}(K_s k_r) = K_s \frac{\partial k_r}{\partial h} \frac{\partial h}{\partial z} = \left(K_s k_r^{-1} \frac{\partial k_r}{\partial h} \right) \frac{\partial \varphi}{\partial z}. \tag{11}$$

We consider the variables:

$$\begin{cases} A = \frac{\phi}{K_s} \left(k_r^{-1} \frac{\partial S}{\partial h} \right), \\ B = \left(k_r^{-1} \frac{\partial k_r}{\partial h} \right). \end{cases} \tag{12}$$

This leads to the following equation:

$$A \frac{\partial \varphi}{\partial t} + \nabla^2 \varphi + B \frac{\partial \varphi}{\partial z} = 0. \tag{13}$$

Finally, by applying the Kirchhoff transformation, we reduced the nonlinearity of Eq. (4) and obtain Eq. (13) which has many benefits in terms of convergence of the proposed numerical method.

2.4 Initial and Boundary Conditions

For the initial condition, we assume that the pressure head is $h(\mathbf{x}, 0) = h_0$ for each point \mathbf{x} on the computational domain Ω , which can be expressed using the Kirchhoff variable as $\varphi(\mathbf{x}, 0) = \varphi(h_0)$.

We transform the boundary conditions using the Kirchhoff variable in a similar way:

Dirichlet: $h(\mathbf{x}, t) = h_D$ for each $\mathbf{x} \in \partial\Omega$ leads to $\varphi(\mathbf{x}, t) = \varphi(h_D)$.

Neumann: $n_i \frac{\partial h}{\partial x_i} = h_N$ implies $n_i \frac{\partial \varphi}{\partial x_i} = k_r h_N$, with h_D and h_N are given functions and n_i is the unit normal vector to the boundary.

3 The Materials and Proposed Techniques

In this section, we propose an efficient computational technique based on radial basis function collocation method [9, 10]. This method has recently become very popular due to its advantages in terms of approximation properties of solutions and its less computational cost since it does not require mesh generation.

Equation (13) is solved using the localized RBF collocation method and the Picard iteration technique. The temporal discretization of Eq. (13) using the backward Euler method is given by:

$$A^{n+1} \frac{\varphi^{n+1} - \varphi^n}{\Delta t} + \nabla^2 \varphi^{n+1} + B^{n+1} \frac{\partial \varphi^{n+1}}{\partial z} = 0, \tag{14}$$

where φ^{n+1} , A^{n+1} and B^{n+1} are the approximate values of φ , A and B at $t = t^{n+1}$, respectively. $\Delta t = t^{n+1} - t^n$ is the time setup and the solution is assumed to be known at t^n and unknown at t^{n+1} .

Equation (14) is linearized using the Picard iteration method which involves sequential estimation of the unknown φ^{n+1} using the latest estimates of A^{n+1} and B^{n+1} . If m identifies iteration levels, then the Picard iteration steps can be written as:

$$A^{m,n+1} \frac{\varphi^{m+1,n+1} - \varphi^n}{\Delta t} + \nabla^2 \varphi^{m+1,n+1} + B^{m,n+1} \frac{\partial \varphi^{m+1,n+1}}{\partial z} = 0. \tag{15}$$

For the sake of simplicity, we consider the following notations:

$$\mathcal{L}^m = \left(\frac{A^{m,n+1}}{\Delta t} \cdot + \nabla^2 \cdot + B^{m,n+1} \frac{\partial \cdot}{\partial z} \right), \tag{16}$$

$$f^{m,n+1} = A^{m,n+1} \frac{\varphi^n}{\Delta t}, \tag{17}$$

\mathcal{L}^m is a linear operator for each Picard iteration m . Subject to boundary and initial conditions, Eq. (15) can be rewritten in the following form:

$$\begin{cases} \mathcal{L}^m \varphi^{m+1,n+1}(\mathbf{x}) = f^{m,n+1}(\mathbf{x}), & \mathbf{x} \in \Omega, \\ \mathcal{B} \varphi^{m+1,n+1}(\mathbf{x}) = q(\mathbf{x}), & \mathbf{x} \in \partial\Omega, \\ \varphi^{m+1,0}(\mathbf{x}) = \varphi_0^{m+1}(\mathbf{x}), & \mathbf{x} \in \Omega, \end{cases} \tag{18}$$

\mathcal{B} is a border operator, q is the given function associated with the boundary conditions. For each iteration n , Eq. (18) is solved using localized RBF meshless method at each Picard iteration m until the stop condition is verified which is given by:

$$\delta^m = |\varphi^{m+1,n+1} - \varphi^{m,n+1}| \leq Tol, \tag{19}$$

with Tol is the error tolerance.

3.1 Localized RBF Meshless Method

In this section, we present the local multiquadric (LMQ) method [11]. This approach is different from the traditional global multiquadric approximation since only local configuration of nodes are used. To recall the localized RBF techniques, let $\{\mathbf{x}_j\}_{j=1}^{n_i}$ and $\{\mathbf{x}_j\}_{j=n_i+1}^N$ be the collocation points in Ω and $\partial\Omega$, respectively. n_i is the number of interior points and N the total number of collocation points distributed over the computational domain. For each $\mathbf{x}_s \in \Omega$, we create a localized domain $\Omega^{[s]}$ that contains n_s nearest neighbors interpolation points $\{\mathbf{x}_k^{[s]}\}_{k=1}^{n_s}$ to \mathbf{x}_s .

In each localized domain $\Omega^{[s]}$, the approximate solution can be written as a linear combination of n_s multiquadric functions in the following form:

$$\varphi_{[s]}^{m+1,n+1}(\mathbf{x}_s) = \sum_{k=1}^{n_s} \alpha_k^{m+1,n+1} \Phi_k \left(\|\mathbf{x}_s - \mathbf{x}_k^{[s]}\| \right), \tag{20}$$

where $\{\alpha_k^{m+1,n+1}\}_{k=1}^{n_s}$ unknown coefficients to be determined, $\|\cdot\|$ is the Euclidian norm and Φ_k are the multiquadric radial basis functions defined as:

$$\Phi_k(\mathbf{x}) = \Phi(r_k) = \sqrt{1 + (\varepsilon r_k)^2}, \quad (21)$$

where $\varepsilon > 0$ is the shape parameter and $r_k = \|\mathbf{x} - \mathbf{x}_k\|$. Equation (21) can be presented in the matrix form:

$$\varphi_{[s]}^{m+1,n+1} = \Phi^{[s]} \alpha_{[s]}^{m+1,n+1}, \quad (22)$$

where $\varphi_{[s]}^{m+1,n+1} = [\varphi_{[s]}^{m+1,n+1}(\mathbf{x}_1^{[s]}), \varphi_{[s]}^{m+1,n+1}(\mathbf{x}_2^{[s]}), \dots, \varphi_{[s]}^{m+1,n+1}(\mathbf{x}_{n_s}^{[s]})]^T$, $\alpha_{[s]}^{m+1,n+1} = [\alpha_{[s]}^{m+1,n+1}(\mathbf{x}_1^{[s]}), \alpha_{[s]}^{m+1,n+1}(\mathbf{x}_2^{[s]}), \dots, \alpha_{[s]}^{m+1,n+1}(\mathbf{x}_{n_s}^{[s]})]^T$ and $\Phi^{[s]}$ is an $n_s \times n_s$ real symmetric coefficient matrix defined as $\Phi^{[s]} = [\Phi(\|\mathbf{x}_i^{[s]} - \mathbf{x}_j^{[s]}\|)]_{1 \leq i, j \leq n_s}$. The vector $\alpha_{[s]}^{m+1,n+1}$ can be obtained as the following equation:

$$\alpha_{[s]}^{m+1,n+1} = (\Phi^{[s]})^{-1} \varphi_{[s]}^{m+1,n+1}. \quad (23)$$

For $\mathbf{x}_s \in \Omega$, we apply the differential operator \mathcal{L}^m to Eq. (20) to obtain the following equation:

$$\begin{aligned} \mathcal{L}^m \varphi_{[s]}^{m+1,n+1}(\mathbf{x}_s) &= \sum_{k=1}^{n_s} \alpha_k^{m+1,n+1} \mathcal{L}^m \Phi_k(\|\mathbf{x}_s - \mathbf{x}_k^{[s]}\|) = \sum_{k=1}^{n_s} \alpha_k^{m+1,n+1} \Psi^m(\|\mathbf{x}_s - \mathbf{x}_k^{[s]}\|), \\ \Theta_{[s]}^m \alpha_{[s]}^{m+1,n+1} &= \Theta_{[s]}^m (\Phi^{[s]})^{-1} \varphi_{[s]}^{m+1,n+1} = \Lambda_{[s]}^m \varphi_{[s]}^{m+1,n+1} = \Lambda^m \varphi^{m+1,n+1}, \end{aligned} \quad (24)$$

where $\varphi^{m+1,n+1} = [\varphi^{m+1,n+1}(\mathbf{x}_1), \varphi^{m+1,n+1}(\mathbf{x}_2), \dots, \varphi^{m+1,n+1}(\mathbf{x}_N)]^T$, $\Theta_{[s]}^m = [\Psi^m(\|\mathbf{x}_s - \mathbf{x}_1^{[s]}\|), \Psi^m(\|\mathbf{x}_s - \mathbf{x}_2^{[s]}\|), \dots, \Psi^m(\|\mathbf{x}_s - \mathbf{x}_{n_s}^{[s]}\|)]^T$, $\Psi^m = \mathcal{L}^m \Phi_k$ and $\Lambda_{[s]}^m = \Theta_{[s]}^m (\Phi^{[s]})^{-1}$.

In order to extend Eq. (24) to be able to use $\varphi^{m+1,n+1}$ instead of $\varphi_{[s]}^{m+1,n+1}$, we consider Λ^m as the expansion of $\Lambda_{[s]}^m$ which can be obtained by padding the local vector with zeros.

Similarly, for $\mathbf{x}_s \in \partial\Omega$, we create an influence domain $\Omega^{[s]}$ containing \mathbf{x}_s . Then we have:

$$\begin{aligned} \mathcal{B}\varphi^{m+1,n+1}(\mathbf{x}_s) &= \sum_{k=1}^{n_s} \alpha_k^{m+1,n+1} \mathcal{B}\Phi_k(\|\mathbf{x}_s - \mathbf{x}_k^{[s]}\|) = (\mathcal{B}\Phi^{[s]}) \alpha_{[s]}^{m+1,n+1}, \\ &= (\mathcal{B}\Phi^{[s]})(\Phi^{[s]})^{-1} \varphi_{[s]}^{m+1,n+1} = \sigma^{[s]} \varphi_{[s]}^{m+1,n+1} = \sigma \varphi^{m+1,n+1}, \end{aligned} \quad (25)$$

where $\sigma^{[s]} = (\mathcal{B}\Phi^{[s]})(\Phi^{[s]})^{-1}$, and σ is the expansion of $\sigma^{[s]}$ obtained by completing the local vector with zeros.

We substitute the Eqs. (24) and (25) into Eq. (18) to obtain the following system:

$$\begin{cases} \mathcal{L}^m \varphi^{m+1,n+1}(\mathbf{x}_s) = \Lambda^m(\mathbf{x}_s) \varphi^{m+1,n+1} = f^{m+1,n+1}(\mathbf{x}_s), \\ \mathcal{B} \varphi^{m+1,n+1}(\mathbf{x}_s) = \sigma(\mathbf{x}_s) \varphi^{m+1,n+1} = q(\mathbf{x}_s), \end{cases} \quad (26)$$

which leads to the following sparse system:

$$\begin{pmatrix} \Lambda^m(\mathbf{x}_1) \\ \Lambda^m(\mathbf{x}_2) \\ \vdots \\ \vdots \\ \Lambda^m(\mathbf{x}_{n_i}) \\ \sigma(\mathbf{x}_{n_i+1}) \\ \vdots \\ \vdots \\ \sigma(\mathbf{x}_N) \end{pmatrix} = \begin{pmatrix} \varphi^{m+1,n+1}(\mathbf{x}_1) \\ \varphi^{m+1,n+1}(\mathbf{x}_2) \\ \vdots \\ \vdots \\ \varphi^{m+1,n+1}(\mathbf{x}_{n_i}) \\ \varphi^{m+1,n+1}(\mathbf{x}_{n_i+1}) \\ \vdots \\ \vdots \\ \varphi^{m+1,n+1}(\mathbf{x}_N) \end{pmatrix} \begin{pmatrix} f^{m+1,n+1}(\mathbf{x}_1) \\ f^{m+1,n+1}(\mathbf{x}_2) \\ \vdots \\ \vdots \\ f^{m+1,n+1}(\mathbf{x}_{n_i}) \\ q(\mathbf{x}_{n_i+1}) \\ \vdots \\ \vdots \\ q(\mathbf{x}_N) \end{pmatrix}. \quad (27)$$

The matrix generated by the localized RBF is sparse due to the presence of the local configuration in the solution approximation. This allows us to avoid ill-conditioned issues that arise in dense systems of equations generated by the global approach. By solving Eq. (27), we obtain the approximate values of $\varphi^{m+1,n+1}$ at all nodes $\varphi^{m+1,n+1}(\mathbf{x}_s)$, $s = 1, 2, \dots, N$ of the computational domain.

4 Numerical Tests

In this section, we perform numerical experiments for solving the Richards equation by using the obtained Eq. (27) in one and two-dimensional systems. We used the localized RBF method based on the multiquadric radial basis function. For the temporal discretization, we used the backward Euler method.

4.1 One Dimensional Infiltration Problem

In this numerical test, we used the Brooks-Corey model [4] which describes the pressure head and the power law for the relative permeability.

$$J(S) = S^{-1/\lambda}, \quad k_r = S^\beta = \begin{cases} \left(\frac{h}{h_{cap}}\right)^{-\lambda\beta}, & \text{if } h \geq h_{cap}, \\ 1, & \text{if } h < h_{cap}. \end{cases} \quad (28)$$

Table 1 Hydraulic property parameters of 2 types of soil

| Texture | θ_r | θ_s | θ_0 | K_s | h_{cap} | λ | $\lambda\beta$ |
|------------|------------|------------|------------|-------|-----------|-----------|----------------|
| Sandy clay | 0.109 | 0.321 | 0.121 | 0.002 | 29.15 | 0.168 | 2.504 |
| Loam | 0.027 | 0.463 | 0.040 | 0.022 | 11.15 | 0.220 | 2.660 |

where λ and β are respectively the parameters related to the Brook-Corey model and the power law for the relative permeability. The second (inequality) condition for the capillary pressure in Eq. (28) is introduced to avoid numerical issues [12, 14].

We consider two different types of soils with a depth L and their hydraulic parameters are shown in Table 1. We simulate a one-dimensional infiltration problem using the proposed method. In order to verify the effectiveness of the developed numerical model, we compare our numerical results with the numerical solutions of 1D-Hydrus where we consider the following initially and boundary conditions:

$$\begin{cases} \theta(z, 0) = \theta_0, \\ \theta(0, t) = \theta_s, \\ \theta(L, t) = \theta_0, \end{cases} \quad (29)$$

Fig. 1 shows the numerical solutions obtained using the proposed method and 1D-Hydrus solutions where we observe good agreement between the solutions.

Table 2 presents the root mean squared error (RMSE), the relative error (L_{er}^1) between the numerical solutions and the solutions simulated by 1D-Hydrus. The results confirm the accuracy of the proposed method in modelling unsaturated flow in soils.

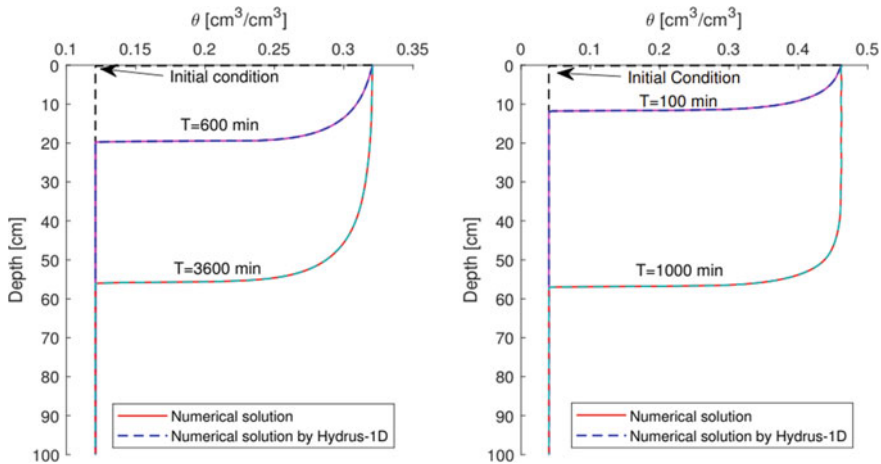


Fig. 1 Time evolution of moisture content for the soils given in Table 1. Left (Sandy clay), right (Loam)

Table 2 The RMSE and the L_{er}^1 between the numerical solutions and 1D-Hydrus solutions

| Soils | $T(\text{min})$ | RMSE | L_{er}^1 |
|------------|-----------------|----------------------|-----------------------|
| Sandy clay | 600 | 5×10^{-3} | 3.5×10^{-3} |
| | 3500 | 5.8×10^{-3} | 4.3×10^{-3} |
| Loam | 100 | 4.8×10^{-3} | 1.08×10^{-3} |
| | 1000 | 6×10^{-3} | 7.2×10^{-3} |

4.2 Two-Dimensional Infiltration Problem

In this example, we perform numerical simulations using the proposed method for a two-dimensional infiltration problem where we consider a rectangular domain $[0, l] \times [0, L]$. We used the same hydraulic parameters of test 1 (Table 1) and $l = L = 100$ cm. We consider the following initial and boundary conditions:

$$\begin{cases} \theta(x, z, 0) = \theta_0, \\ \theta(x, 0, t) = \theta_s, \\ \theta(x, L, t) = \theta_0, \end{cases} \quad (30)$$

and no-flux boundary conditions are imposed on the sides $x = 0$ and $x = l$ of the domain.

The numerical simulations are performed using $N_x = 200, N_z = 200, \Delta t = 0.05$, and the localized RBF parameters $\varepsilon = 0.6$ and $n_s = 5$. The sandy clay and loam soils are selected in this numerical test to simulate unsaturated flow through a two-dimensional homogeneous medium. The time evolution of the total mass per unit of length of the 2D numerical solutions and the solutions simulated by 1D-Hydrus for a computational domain of unit length (1D problem) are displayed in Fig. 2. We observe good agreement between the solutions, which demonstrates the accuracy of

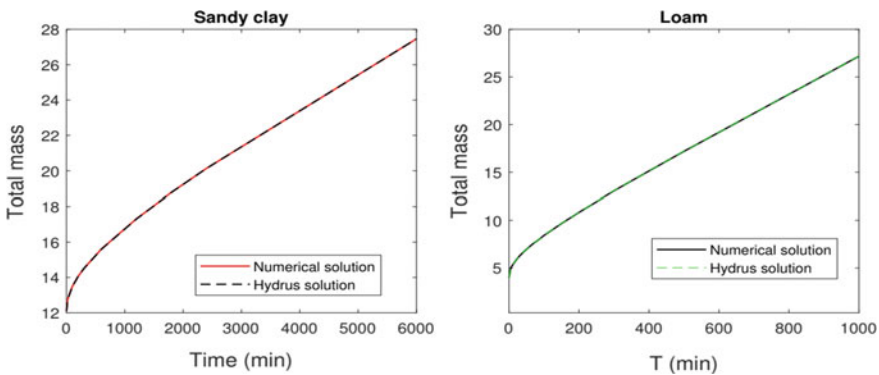


Fig. 2 Time-evolution of the total mass per unit of length of the sandy clay and loam soils

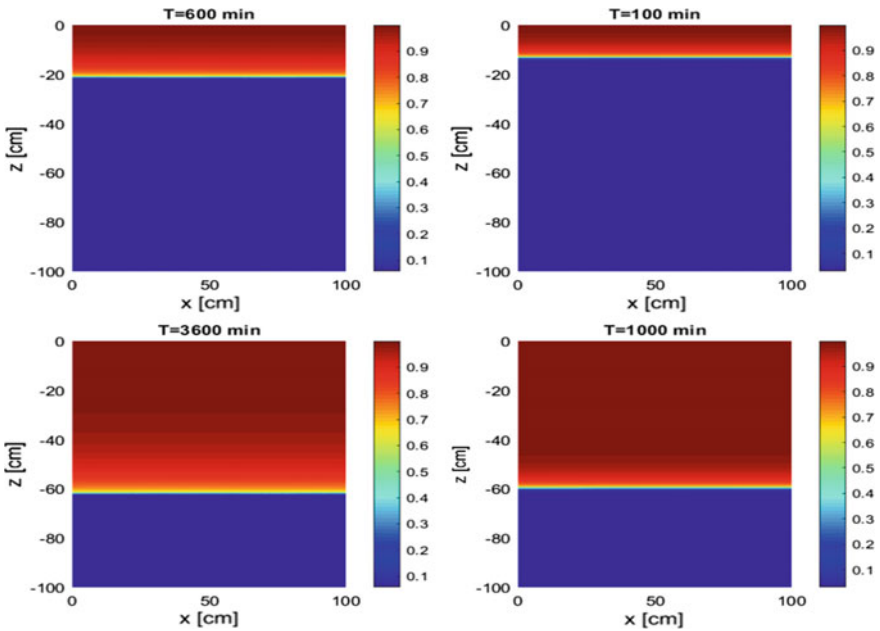


Fig. 3 The time evolution of saturation for the sandy clay (left) and loam (right) soils

the proposed method in modelling 2D unsaturated flow in soils. Figure 3 shows the time evolution of saturation for the sandy clay and loam considered soils.

The proposed method is efficient and accurate for solving the Richards equation. The method can be used for modelling unsaturated flow through homogeneous soils.

5 Conclusion

This paper focused on the infiltration process in porous media and introduced computational techniques for efficiently solving the Richards equation in one- and two-dimensional homogeneous medium. The proposed techniques using the Kirchhoff transformation allow us to reduce the nonlinearity of the obtained system from the Richards equation. Our approach using a localized radial basis function method avoiding mesh generation allows us to reduce the computational cost. The accuracy of the proposed method was validated using comparison between the numerical solutions and the results of 1D-Hydrus. Our results confirm the accuracy of the proposed techniques and their efficiency in terms of computational cost for solving the Richards equation. The numerical techniques proposed in this study for modelling unsaturated flow through homogeneous porous media is a first step toward developing efficient and accurate numerical methods for modelling unsaturated flows through heterogeneous soils.

References

1. Bause M, Knabner P (2004) Computation of variably saturated subsurface flow by adaptive mixed hybrid finite element methods. *Adv Water Resour* 27:565–581
2. Bear J (1972) Dynamics of fluids in porous media. In: Courier Corporation. American Elsevier Publishing Company, New York
3. Boujoudar M, Beljadid A, Taik A (2021) Localized MQ-RBF meshless techniques for modeling unsaturated flow. *Engineering Analysis With Boundary Elements* 130:109–123
4. Brooks R, Corey T (1964) HYDRAU uc properties of porous media. *Hydrology Papers*, Colorado State University 24:37
5. Celia MA, Bouloutas ET, Zarba RL (1990) A general mass-conservative numerical solution for the unsaturated flow equation. *Water Resour Res* 26:1483–1496
6. Gardner WR (1958) Some steady-state solutions of the unsaturated moisture flow equation with application to evaporation from a water table. *Soil Sci* 85:228–232
7. Huang RQ, Wu LZ (2012) Analytical solutions to 1-D horizontal and vertical water infiltration in saturated/unsaturated soils considering time-varying rainfall. *Comput Geotech* 39:66–72
8. Ji SH, Park YJ, Sudicky EA, Sykes JF (2008) A generalized transformation approach for simulating steady-state variably-saturated subsurface flow. *Adv Water Resour* 31:313–323
9. Kansa EJ (1990) Multiquadrics—A scattered data approximation scheme with applications to computational fluid-dynamics—II solutions to parabolic, hyperbolic and elliptic partial differential equations. *Comput Math Appl* 19(1):147–161
10. Kansa EJ (1990) Multiquadrics—a scattered data approximation scheme with applications to computational fluid-dynamics—I surface approximations and partial derivative estimates. *Comput Math Appl* 19:127–145
11. Lee CK, Liu X, Fan SC (2003) Local multiquadric approximation for solving boundary value problems. *Comput Mech* 30:396–409
12. Lenhard RJ, Parker JC, Parker MS (1989) On the correspondence between Brooks-Corey and van Genuchten models. *J Irrig Drainage Eng* 115:744–751
13. Leverett M (1941) Capillary behavior in porous solids. *Trans AIME* 142:152–169
14. Ma Q, Hook JE, Ahuja LR (1999) Influence of three-parameter conversion methods between van Genuchten and Brooks-Corey functions on soil hydraulic properties and water-balance predictions. *Water Resour Res* 35:2571–2578
15. Manzini G, Ferraris S (1990) Mass-conservative finite volume methods on 2-D unstructured grids for the Richards' equation. *Adv Water Resour* 27:1483–1496
16. Richards LA (1931) Capillary conduction of liquids through porous mediums. *Physics* 1:318–333
17. Sander GC, Parlange JY, Kühnel V, Hogarth WL, Lockington D, O'kane JPJ (1988) Exact nonlinear solution for constant flux infiltration. *J Hydrol* 97:341–346
18. Srivastava R, Yeh TCJ (1991) Analytical solutions for one-dimensional, transient infiltration toward the water table in homogeneous and layered soils. *Water Resour Res* 27:753–762
19. Tracy FT (2011) Analytical and numerical solutions of Richards' equation with discussions on relative hydraulic conductivity. In: *BoD—Books on demand*
20. Van Genuchten MT (1980) A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci Soc Am J* 44:892–898

Digital Twin: A City-Scale Flood Imitation Framework



Maysara Ghaith, Ahmed Yosri, and Wael El-Dakhakhni

1 Introduction

Floods have been highlighted as the costliest natural hazards in Canada for decades, as they often caused fatalities and destroyed the city infrastructures. Floods can impact the city both directly and indirectly as they do not only affect the city's buildings and roads but also can hinder the economic and social activities in the community. Flood hazards are getting more frequent and severe over the globe, which has lately been attributed to climate change [2, 10]. In Canada, during the past decade, several extreme flood events occurred and caused major damages across the country. For example, the economic loss due to the 2013 flood in the city of Calgary was around \$5 billion [14]. Similar devastating floods impacted other Canadian regions (e.g., Toronto in 2013 and 2018, Ottawa-Gatineau in 2017, 2018, and 2019, Montreal in 2019, and Fredericton in 2019) [3, 6]. These examples show that Canadian cities are vulnerable to floods and are far from being resilient under such hazards. Proper, efficient, and rapid flood management tools are therefore essential to avoid, reduce and mitigate future flood hazards.

Flood estimation and behaviour is an evergoing focus for hydrologists, where behaviour tools are evolving as the data availability and computer power increase. Simulating flood impacts is typically carried out using 0-D or 1-D river simulation tools. 0-D river simulation relies on assuming a constant water level in the river and subsequently inundate the surrounding areas that are lower than this level. 0-D models are computationally efficient and can be used for identifying the vulnerable

M. Ghaith (✉) · A. Yosri · W. El-Dakhakhni
Department of Civil Engineering, McMaster University, Hamilton, ON L8S 4L7, Canada
e-mail: ghaithm@mcmaster.ca

A. Yosri · W. El-Dakhakhni
INTERFACE Institute for Multi-Hazard Systemic Risk Studies, McMaster University, Hamilton,
ON L8S 4L7, Canada

communities when a flood happens; however, such models do not reflect the actual impacts of the flood event and do not provide accurate measures. On the other hand, 1-D river simulations (e.g., Hec-Ras 1-D) assume that lateral flows are constant. A 1-D river model typically result in highly accurate flow values; however, inundation areas are still not exact as the model do not distinguish the properties of the main channel from those of the overland flooded areas. Lately, 2-D/3-D flood mapping tools have become more common due to the increasing computer power that makes the simulation speed rapid and convenient to use. Examples of such tools include Hec-Ras 2-D and 3-D, Iber 2D, Flood Modeller 2D, Delft 3D and PCSWMM. 2-D and 3-D flood mapping tools have been successful in simulating the actual flow extend and inundation depth efficiently [7, 9, 11, 15], making such models to be effective candidates for flood simulation. However, the development of a 2-D or 3-D flood mapping model requires collecting significant amount of data which is most often prohibited due to limited resources. Nonetheless, 2-D and 3-D flood mapping tools are key for policy and decision makers to find the city's vulnerable areas under expected flood projections and upgraded these areas accordingly, which can ultimately lead to reducing the city vulnerability and boost its resilience under flood hazards.

Despite the existence of numerous flood simulation tools, the use of the resulting flood maps in identifying the city's vulnerable locations and devising proper mitigation strategies is not quite accurate as the system simulated (i.e., the city) typically consists of a network of interdependent infrastructure networks (e.g., buildings, roads, power grids, waterlines, stormwater networks, wastewater facilities). For example, the failure of roads and buildings can cause an indirect failure of the power grid, traffic signals, and finance infrastructures. As such, effective modelling of the interdependence between the different city infrastructures is key for the development of accurate and reliable development, preparedness, and mitigation plans under natural hazards such as floods. So lately, this has been achieved through digitizing and mimicking the city behaviour by virtually replicating the city in a digital twin [5]. A city digital twin contains all the city infrastructures, intra-dependencies within the same infrastructure system, and inter-dependencies between the different systems, as well as human-infrastructure interactions. Applying the hazard to such digital twins enables the real-life simulation and management of what happens under such hazard and can aid the decision-makers in trying different responding plans to minimize the impact of the hazard [1, 8]. The development of a city digital twin requires a lot of input data that can be collected from several sources (e.g., sensors, internet of things, surveillance cameras, and satellites). Several studies have been employed the concept of city digital twins for different purposes. For example, a digital twin of the city of Zurich has been under-development since 1990 and is proposed to facilitate the urban planning under expected population growth [13]. Another example is a digital twin for the city of Helsinki that was developed in 1980 w for urban development purposes, but is currently used as a platform for managing and designing the new life of the city [12]. These examples, in addition to others, show that digital twins are the new promising platforms to evaluate, enhance, and control the city at all aspects in order to boost its resilience under natural and anthropogenic hazards.

In this paper, a framework describing a systematic way for building a city digital twin is developed. To demonstrate the applicability and viability of this framework, a digital twin of the city of Calgary is built based on such framework and used to replicate the impacts of the flood happened in June 2013. It is worth noting that building a river flood inundation digital twin is the main focus of this study, and other aspects of the city digital twin can be included in future studies.

2 City Digital Twin

A *City Digital Twin* is a replica of the integrated multi-physical behaviour of a city consists of several complex, interacting infrastructure systems (e.g., buildings, power, water, roads). Each of these systems can in fact be represented through a network, and subsequently a network digital twin. For example, a digital twin for the network of smart buildings can be built and be used to monitor the surrounding environmental and social aspects (e.g., temperature, air quality, light, people). Another digital twin of the transportation network can be built to simulate the traffic flow, congestions, and accidents under expected scenarios of population growth. In addition, a digital twin for the river network can be emulated to simulate water stages under different scenarios for the efficient design of water management and flood risk strategies. Also, digital twins for water, stormwater, and wastewater networks can be built to simulate their interdependence as well as their interaction with other infrastructure networks (e.g., the river network to maintain a good water quality, and transportation and building networks to reduce the risk of flooding). Furthermore, a power network digital twin can be built as the power network has been proven to be important for the functionality of other city's infrastructures [4].

In order for developing a network digital twin, raw data characterizing the corresponding infrastructure system are required. A physical, analytical, or numerical model has to be subsequently implemented to simulate the real-time behaviour of the system. Finally, the behaviour of the different systems can be integrated and merged with a city information model (CIM), and subsequently be visualized on the city digital twin. It should be emphasized that building a city digital twin is an ongoing process as new data can be continuously acquired and used to update the state of the digital twin in order to enhance its reliability. Such new data can be collected from various sources such as: (1) surveillance cameras, that can provide a lot of rich information about the traffic, weather conditions, and incidents; (2) sensors, recording climatic and traffic conditions, air and water quality, and electricity and gas flow; (3) geographical positioning systems (GPS) attached to cars or cell phones, that can provide useful information about human behaviour across the city; (4) Internet of Things and satellites, that can provide more data about areas not covered by other sources; (5) open smart buildings, that can provide a full data about the surrounding environment. It should be noticed that the resolution of the city digital twin is determined based on which infrastructure system is more important in the city under consideration.

3 City Digital Twin Development Framework

To start building a digital twin for a specific area, the first step is to identify the most important infrastructure systems that are most valuable to the city. As shown in Fig. 1, the main infrastructure systems are, for example, the water distribution and collection, transportation, power, and river networks. These networks are usually the most important infrastructures; however, other systems can be added depending on the city interest and the hazard under consideration. The second step in developing a city digital twin is to select a suitable model (e.g., analytical, numerical, data-driven) to simulate the response of each network. The model chosen to replicate the actual response of the network must be chosen based on multiple criteria, including reliability, efficiency, and accuracy. In order to rely on the model selected, it has to be well calibrated. As such, the third step in developing a city digital twin is to determine and acquire the data needed to build, calibrate, and validate the selected models. Once the models are calibrated and validated, they should be hooked up and merged with a CIM to form the city digital twin. Accordingly, the effect of any infrastructure system disturbance on other systems and the whole city can be immediately monitored.

Even it may seem that building a city digital twin is a relatively easy task, integrating all of its comprising infrastructure systems is challenging. In real-life, city

Fig. 1 Framework of city digital twin development

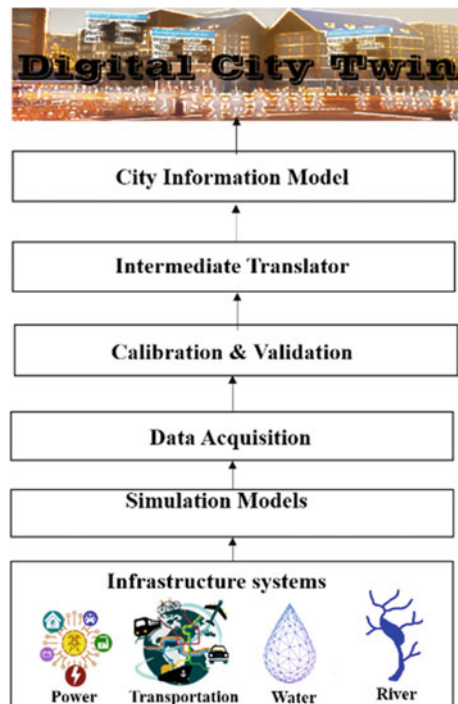
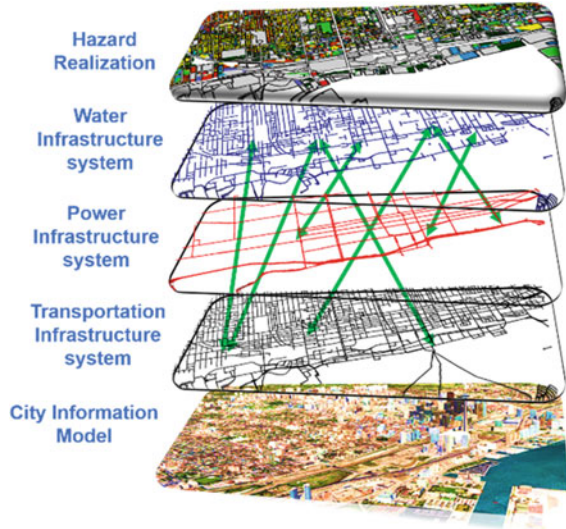


Fig. 2 Infra-structure system interdependency



infrastructure systems have interdependencies between them (Fig. 2), and consequently the failure of a component in one system can affect the performance of other systems. To interrupt the interdependencies between the data and models describing the city infrastructure systems, another software is needed to translate and define rules between them as shown in Fig. 1. Once all the infrastructure systems are integrated with the interdependency rule, corresponding models will run iteratively until a stable state is reached. Subsequently, the final results will be visualized on the city digital twin.

4 Operating the Framework

A city digital twin is a real-time operating platform that can be used to simulate the city behavior under a disturbance (e.g., hazard realization) considering the interdependencies between its different infrastructure systems. For example, as shown in Fig. 3, when the flow gage along the river reads high level than expected, it is determined as a flood hazard. The river simulation model will run independently to simulate the effect of this flood hazard. The resulting inundation map have to be translated to all infrastructure simulation models that are already calibrated. Each model will also run independently and evaluate the performance of the corresponding infrastructure system. As prementioned, since the city infrastructure systems are interdependent in real-life, the effect of such interdependency has to be taken into consideration. So, simulation results will be sent back and forth between the simulation models and the translator software (e.g., GIS) until a steady state results are achieved.

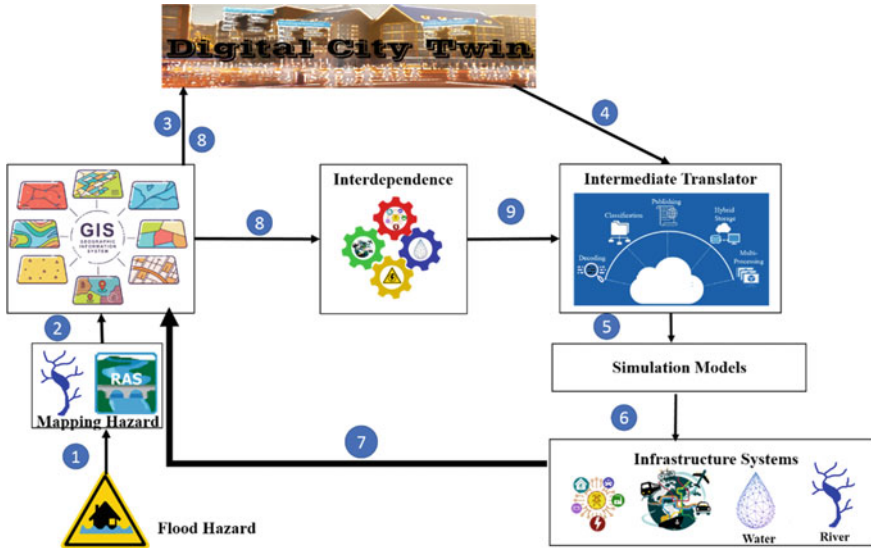


Fig. 3 Operation of the developed framework under a flood event

Following the aforementioned procedures, direct and indirect effects of any hazard can be visualized on the city digital twin. Decision makers can accordingly (1) interfere with the system (i.e., providing maintenance crews) to reduce the risk at any time instance; and, (2) can simulate potential hazards on the city digital twin to identify vulnerable areas, assess potential projects, and devise preparedness plans to avoid catastrophic events. At all cases, the action taken can be linked to the city digital twin through the translator to see the new effect.

5 Demonstration Example

5.1 Study Area, Infrastructure Systems Identification, and Simulation Model Selection

The main goal of this study is to build a city digital twin for Canadian cities in order to simulate the effects of flood hazards and to reduce casualties and economic loss resulting from such hazards. The city of Calgary was therefore chosen as a testbed for building the first Canadian city digital twin as it was impacted by the most devastating flood hazard in the last decade. The main infrastructure networks that will be integrated in the Calgary digital twin are the river, water, power, and transportation networks. Since, as previously explained, building a city digital twin is a complex process and is based on integrating multiple interdependent infrastructure

systems, the main focus of this demonstration is to build a river system digital twin for the city of Calgary as a first step towards building the whole city digital twin. As such, Hec-Ras software was selected to simulate the effects of flood events. It is noteworthy that other river analysis software can be used; however, Hec-Ras was used in this study as it contains an efficient and accurate 2-D/3-D flood mapping module which is required to map the flood inundation for big cities. The data required to build the Hec-Ras model are: a Digital Elevation Model (DEM), historical inflow hydrographs, historical outflow hydrograph, and/or a historical flood extend. The DEM was obtained from the Government of Canada website (<https://open.canada.ca>) under the Open Government License. To get the inflow, it can be done by either running hydrological model to translate meteorological conditions, land use, land cover, soil characteristics into a hydrograph or by finding a flow monitoring station. In this study, flow records during 2013 were obtained from Government of Canada-Environment and natural resources website (<http://wateroffice.ec.gc.ca>) and were used for simplicity. To calibrate the river simulation model, flow monitoring stations at the downstream are required to identify the optimum manning coefficient values in the study area and a flood extend is required to make sure that model is well performing based on calibration points. The flood extent during 2013 was obtained from the City of Calgary's Open Data Portal (<https://data.calgary.ca/>).

5.2 Model Building and Calibration

To build a Hec-Ras model, a DEM of the area is obtained, the boundaries of the study area are identified, and the boundaries of existing river systems are drawn. Subsequently, the model boundaries are defined, and a numerical mesh is generated accordingly as shown in Fig. 4. The mesh size has to be chosen wisely as a bigger mesh size reduces the model accuracy, whereas a smaller mesh size renders the need for more computational power and time. In this demonstration example, a square mesh of a 250 m side length is selected, producing approximately 9000 cells as shown in Fig. 4. Boundary conditions are then defined in order to enable solving the equation governing the underlying physical processes. Three input flow gages were used as inflow boundary conditions at the upstream of the city of Calgary, whereas a normal depth is selected at the location of the downstream boundary condition (Fig. 4). The only parameter to be calibrated in the Hec-Ras model is the Manning's coefficient. As such, the model was run from April 2013 to August 2013, and two downstream head gages are used together with the measured flood extend from the June 2013 flood event (Fig. 6a). To run the model, a suitable timestep has to be chosen such that the model is numerically stable. As such, a timestep ranging between 0.0035 and 47 s was used and corresponded to a Courant number between 0.45 and 2.3. The automatic calibration tool built in the Hec-Ras software resulted in two regions of different Manning's Coefficients of 0.2 and 0.08 as presented in Fig. 5, and the corresponding numerical model replicated the actual water levels at the calibration locations as shown in Fig. 6b and c. In addition, flood extend produced from the Hec-Ras mode

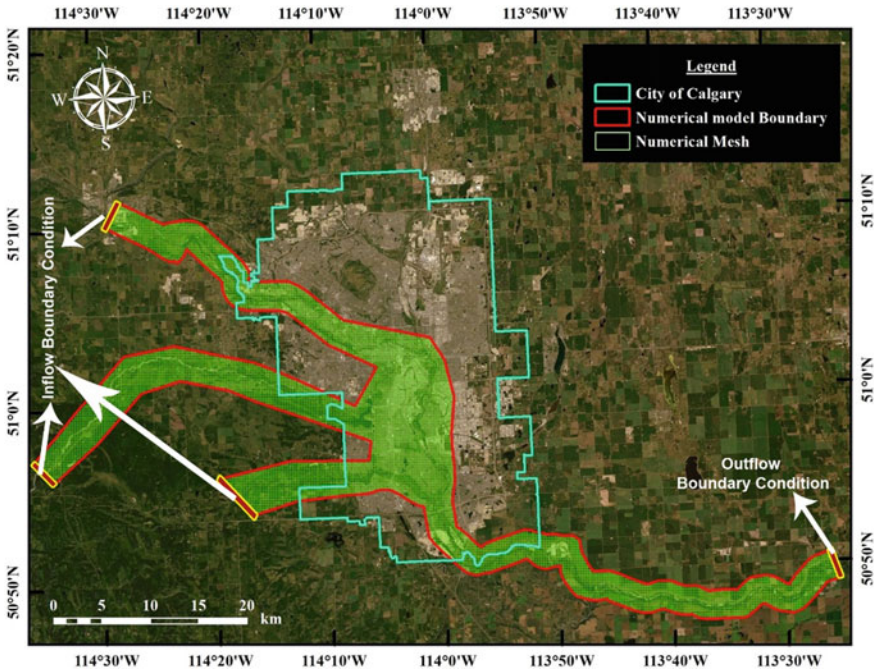


Fig. 4 Hec-Ras model setup

encloses the actual one with some over-inundated areas. These results support the model capabilities in producing an accurate flood inundation map for the city of Calgary.

5.3 The City of Calgary Digital Twin

The Hec-Ras model results were integrated with a CIM of the city of Calgary in a 3D visualization of the 2013 flood event between June 15th and June 31st. The CIM was built using open street map data, DEM, and satellite images. As shown in Fig. 7, the CIM can be used to visualize the flooded area at any specified time as a bird-eye or human-eye view. Also, the building information and the hydrograph at any building location can be visualized as shown in Fig. 7c. The CIM provides the decision makers with the real feel of the city from different perspectives which can enable them to find proper solutions under catastrophic events.

In a GIS software, a complementary analysis as well as visualization can be done. For instance, Fig. 8 shows the maximum inundation depth resulted from the 2013 flood in Calgary and the affected building are colored based on the corresponding inundation depth. The total number of building affected by the flood event were

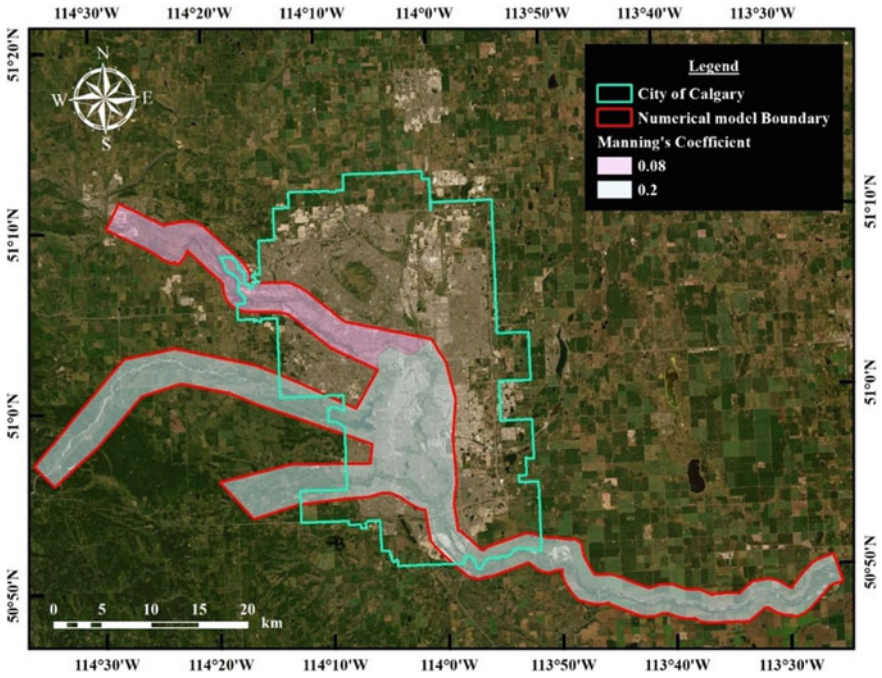


Fig. 5 Manning coefficient map

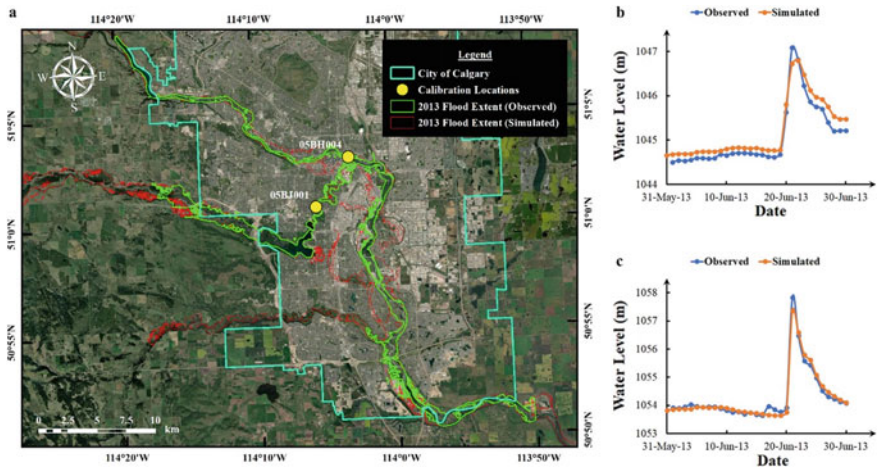


Fig. 6 Comparison between observed and simulated 2013 flood in terms of a extent; b hydrographs at station 05BH004; c hydrographs at station 05BJ001

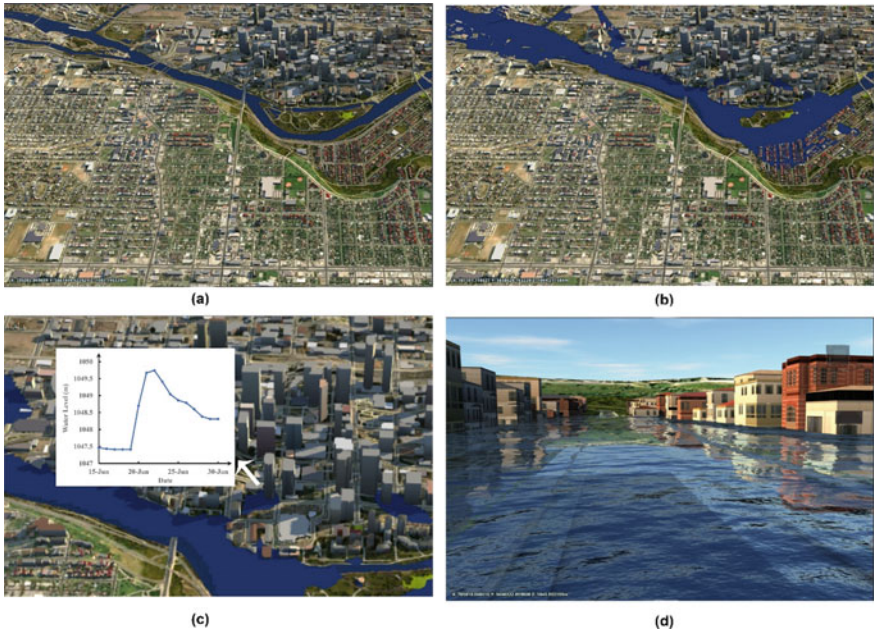


Fig. 7 City digital twin visualization: **a** Bird-eye view of part of the city on June 15th before flood **b** Bird-eye view of same part of the city on June 23rd **c** hydrograph at one of the buildings **d** Human-eye view of the flooded buildings

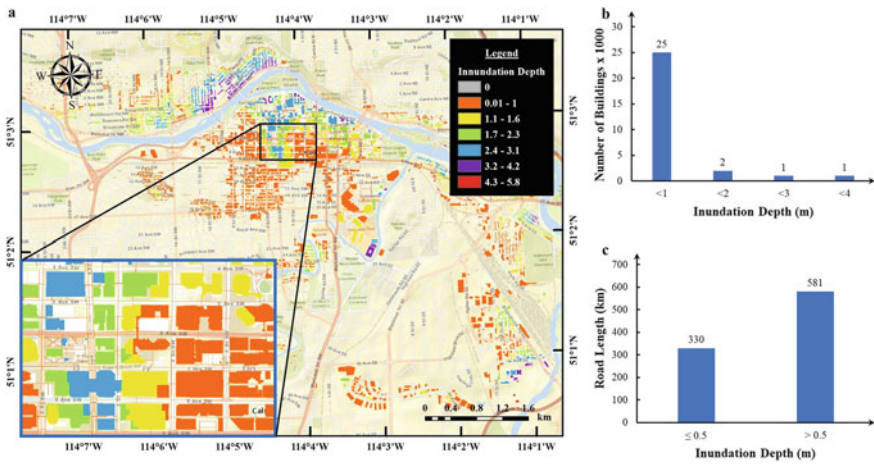


Fig. 8 Flood depth infographic map for June 2013 flood event

29,000 which were subsequently divided into six categories for spatial analysis and four categories for statistics (Fig. 8a and b, respectively). The number of building that are flooded with less than one meter, between 1 and 2 m, between 2 and 3 m, and more than 3 m are 25,000, 2000, 1000, and 1000, respectively. Other insights can be inferred through the use of a GIS-based analysis such as identifying the state of a car following a flood event for insurance purposes. A car is considered a total loss and beyond repair if it is affected by a flood depth of around 50 cm, depending on the car type. During the 2013 flood event, the length of Calgary's streets flooded by more than 50 cm were 581 km (Fig. 8c). Many more insights can be inferred from both the city digital twin and the GIS-based analysis such as the total monetary loss due to the flood event; however, this is beyond the focus of this study.

6 Conclusion

Smart city digital twin is an emerging tool that will provide a better future. However, building a city digital twin is not an easy task due to the immense data required as well as the interdependence between the city's infrastructure systems. The present study presents a framework that can be used to facilitate the development of city digital twins through the integration of emerging technologies (e.g., sensors, Internet of Things, satellites), infrastructure modelling, and interoperability translators. The framework was applied to build a digital twin for the city of Calgary as testbed. However, the ongoing technological advancement and data availability highlight that having digital twins for large cities will be reached in the near future. A city digital twin built based on the framework presented in this study will provide the decision and policy makers with the complete actual figure for urban development, potential locations for investment, and a platform for devising risk mitigation strategies under future hazards.

References

1. Ford DN, Ph D, Asce M, Wolf CM, Asce M (2020) Smart cities with digital twin systems for disaster management. *Am Soc Civil Eng* 36(4):1–10. [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0000779](https://doi.org/10.1061/(ASCE)ME.1943-5479.0000779)
2. Gaur A, Gaur A (2018) Future changes in flood hazards across Canada under a changing climate. *Water* 10(1441):1–21. <https://doi.org/10.3390/w10101441>
3. Gaur A, Gaur A, Simonovic SP (2018) Modelling of future flood risk across Canada due to climate change. *WIT Trans Eng Sci* 121:149–159. <https://doi.org/10.2495/RISK180131>
4. Haggag M, Ezzeldin M, El-Dakhkhni W, Hassini E (2020) Resilient cities critical infrastructure interdependence: a meta-research. *Sustain Resilient Infrastruct* 1–22. <https://doi.org/10.1080/23789689.2020.1795571>
5. Ivanov S, Nikolskaya K, Radchenko G, Sokolinsky L, Zymbler M (2020) Digital twin of city: concept overview. In: *Global smart industry conference*, pp 178–86. <https://doi.org/10.1109/GloSIC50886.2020.9267879>

6. Kokas T, Simonovic SP, Binns A (2016) Water resources research report flood risk management in Canadian urban environments: a comprehensive framework for water resources modeling and decision-making
7. Lamb R, Crossley M, Waller S (2009) A fast two-dimensional floodplain inundation model. *Inst Civil Eng Water Mang* 162:363–70. <https://doi.org/10.1680/wama.2009.162.6.363>
8. Lu Q, Asce AM, Parlikad AK, Woodall P, Ranasinghe GD, Xie X, Liang Z, Konstantinou E, Heaton J, Schooling J (2021) Developing a digital twin at building and city levels: case study of west Cambridge campus. *Am Soc Civil Eng* 36(2017):1–19. [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0000763](https://doi.org/10.1061/(ASCE)ME.1943-5479.0000763)
9. Merwade V, Cook A, Coonrod J (2008) GIS techniques for creating river terrain models for hydrodynamic modeling and flood inundation mapping. *Environ Model Softw* 23:1300–1311. <https://doi.org/10.1016/j.envsoft.2008.03.005>
10. Nofal OM, Van De Lindt JW (2020) Understanding flood risk in the context of community resilience modeling for the built environment: research needs and trends built environment: research needs and trends. *Sustain Resilient Infrac* 00(00):1–17. <https://doi.org/10.1080/23789689.2020.1722546>
11. Pinos J, Timbe L (2019) Performance assessment of two-dimensional hydraulic models for generation of flood inundation maps in mountain River Basins. *Water Sci Eng* 12(1):11–18. <https://doi.org/10.1016/j.wse.2019.03.001>
12. Ruohomaki T, Airaksinen E, Huuska P, Kesaniemi O, Martikka M, Suomisto J (2018) Smart city platform enabling digital twin. In: *International conference on intelligent systems*, vol 0, pp 155–161
13. Schrotter G, Hürzeler C (2020) The digital twin of the City of Zurich for urban planning. *PFG J Photogrammetry Remote Sens Geoinformation Sci* 88(1):99–112. <https://doi.org/10.1007/s41064-020-00092-2>
14. Tanganelli M, Viti S, Mariani V, Pianigiani M (2016) Ensembles compatible to EC8 and NTC 2008 seismic assessment of existing rc buildings under alternative ground motion ensembles compatible to EC8 and NTC 2008. *Bull Earthq Eng* (October) <https://doi.org/10.1007/s10518-016-0028-z>
15. Teng J, Jakeman AJ, Vaze J, Croke BFW, Dutta D, Kim S (2017) Flood inundation modelling: a review of methods, recent advances and uncertainty analysis. *Environ Model Softw* 90:201–203

Modelling of Coupled Surface and Subsurface Water Flows



Hasan Karjoun and Abdelaziz Beljadid

1 Introduction

Modelling of coupled systems of surface and subsurface water flows has many applications in hydrology, water management and agricultural engineering [6, 15, 16]. These coupled models allow understanding the dynamic processes of surface water flows and their interactions with the infiltration process in soils, and can be used for mass balance estimation. The water mass balance analysis is used to quantify the amount of surface water for water management purposes [14]. In our study, the physically based model used for surface water flows is the shallow water equations (SWEs), also called the Saint Venant system [5]. The SWEs are widely used to study systems of shallow water flows where the water depth is much smaller than the horizontal length scale of these systems and the variations of the flow in the vertical direction is negligible compared to the variations in the horizontal one [4, 12, 19]. For subsurface water flow, the Green-Ampt model [8] is considered to describe the infiltration process through soils. This simplified physically based hydrodynamic model can be used to predict infiltration capacity, ponding time, and cumulative infiltration for layered soils.

Different physical processes occur during rainfall event including the infiltration process in soils, the dynamics of flows over variable topography with friction effect and their interactions, which make their simulations very challenging and require robust numerical techniques. In the present study, to better capture these

H. Karjoun (✉) · A. Beljadid
Mohammed VI Polytechnic University, Green City, Morocco
e-mail: hassan.karjoun@um6p.ma

A. Beljadid
e-mail: abdelaziz.beljadid@um6p.ma

A. Beljadid
University of Ottawa, Ottawa, Canada

physical processes, the HLL-Riemann solver finite volume method on unstructured triangular grids [1, 9, 19] is implemented for solving the resulting nonlinear-coupled system of SWEs and the Green-Ampt infiltration model. This infiltration model leads to nonlinear system, which is solved using implicit iterative methods. Appropriate discretization techniques are used for the bottom topography and Manning friction source terms to ensure the underlying physical properties. Moreover, piecewise linear reconstructions of the solutions are used at the discrete level in order to achieve the second-order accuracy of the numerical scheme. For subsurface water, we proposed accurate method for computing the infiltration rate under rainfall-runoff processes. The developed numerical techniques can be used to predict surface and subsurface water flows under rainfall events over complex bottom topography involving wet and dry areas and large bed slop where we obtain accurate results in terms of mass balance.

The paper is organized as follows. In Sect. 2, we present the coupled model of surface and subsurface water flows based on the SWEs and Green-Ampt model for one- and two-layers soil. To numerically solve this model, we propose to use the second-order HLL finite volume scheme in Sect. 3. In Sect. 4, we test the robustness and accuracy of the proposed numerical techniques for simulating rainfall-runoff over variable bottom topography with wet and dry areas where we perform three numerical experiments. Finally, some concluding remarks are given in Sect. 5.

2 Model Equation

In this study, we focus on the following shallow water system over variable bottom topography with different source terms such as friction effects, precipitation and infiltration, modelling coupled surface and subsurface water flows:

$$\begin{cases} \frac{\partial h}{\partial t} + \frac{\partial hu}{\partial x} + \frac{\partial hv}{\partial y} = R - I, \\ \frac{\partial hu}{\partial t} + \frac{\partial}{\partial x} \left(hu^2 + \frac{g}{2} h^2 \right) + \frac{\partial}{\partial y} (huv) = -gh \frac{\partial B}{\partial x} - \frac{\tau_x}{\rho}, \\ \frac{\partial hv}{\partial t} + \frac{\partial}{\partial x} (huv) + \frac{\partial}{\partial y} \left(hv^2 + \frac{g}{2} h^2 \right) = -gh \frac{\partial B}{\partial y} - \frac{\tau_y}{\rho}, \end{cases} \quad (1)$$

where h represents the water depth, $\mathbf{u} := (u, v)^T$ is the depth averaged velocity field of the flow, g is the gravitational constant, the function $B(x, y)$ represents the bottom elevation, and ρ is the water density. The source terms R and I are respectively the rainfall intensity and the infiltration rate into soil. We consider the following Manning formula for the computation of the friction source terms:

$$\begin{cases} \frac{\tau_x}{\rho} = g \frac{n^2}{h^{1/3}} \|\mathbf{u}\|u, \\ \frac{\tau_y}{\rho} = g \frac{n^2}{h^{1/3}} \|\mathbf{u}\|v, \end{cases} \tag{2}$$

with n is the Manning coefficient and $\|\mathbf{u}\|$ is the norm of the velocity field of the flow. The Manning formula is based on the assumption of uniform flow [11] and it is initially developed for flow in open channels and pipes. The Manning formula is widely used in many applications to compute the friction stresses in runoff models [4, 6]. In this study, we assume a constant Manning coefficient. While with this assumption we obtain a good agreement between experiments and simulations for the numerical tests performed in this study, a Manning formula with variable coefficient n could be used, as shown in the comparative study by [12].

2.1 Green-Ampt Infiltration Model

In this section, we introduce the Green-Ampt infiltration model [8] which is widely used [12, 16] to compute the infiltration capacity of water in soils. This model is a simplified form of the Richards equation where the wetting front is considered as a sharp boundary between saturated and dry zones of the soil. To design an efficient coupled surface–subsurface numerical model in terms of computational cost, the Green-Ampt model is used in our study because of its simplicity and performance in computing the cumulative infiltration in soils.

We assume that water is ponded at the surface with depth h_p , the Green-Ampt model is used to compute the infiltration capacity I_r as follows:

$$I_r = K_s \left(\frac{(\psi + h_p)\Delta\theta}{I_c} + 1 \right), \tag{3}$$

where K_s is the saturated hydraulic conductivity of soil, $\Delta\theta = \theta_s - \theta_i$ with θ_s and θ_i are respectively the initial and saturated moisture contents, ψ is the suction head, $I_c = \int_0^t I_r(\tau)d\tau$ is the cumulative infiltrated water. The wetting front reaches the depth $d_f = I_c/\Delta\theta$ at time t where we assume $d_f = 0$ at $t = 0$. By integrating Eq. 3 over $[t, t + \Delta t]$ we obtain the following equation for the cumulative infiltration:

$$I_c^{t+\Delta t} = I_c^t + K_s\Delta t + \Delta\theta(\psi + h_p) \ln \left(\frac{((\psi + h_p)\Delta\theta + I_c^{t+\Delta t})}{(\psi + h_p)\Delta\theta + I_c^t} \right). \tag{4}$$

2.2 Two Layers Green-Ampt Infiltration Model

Similarly to one-layer soil, we extend the Green-Ampt model for two-layers soil where the upper layer has a thickness d_1 with hydraulic conductivity $K_s = K_1$, suction head $\psi = \psi_1$ and $\Delta\theta = \Delta\theta_1$, the lower layer corresponds to the ground below the first one, with hydraulic conductivity $K_s = K_2$, suction head $\psi = \psi_2$ and $\Delta\theta = \Delta\theta_2$. The infiltration capacity for this system of two layers soil can be expressed as follows:

$$I_r = K_e \left(\frac{\psi + h_p}{d_f} + 1 \right), \tag{5}$$

where K_e is the effective hydraulic conductivity given by:

$$K_e = \begin{cases} K_1, & \text{if } d_f \leq d_1, \\ \frac{d_f}{\frac{d_1}{K_1} + \frac{d_f - d_1}{K_2}}, & \text{if } d_f > d_1. \end{cases} \tag{6}$$

The cumulative infiltration for two layers soil denoted by \tilde{I}_c is computed as:

$$\tilde{I}_c = I_c + a, \tag{7}$$

where the cumulative infiltration $I_c = d_f \Delta\theta$ is obtained using the following equation:

$$I_r = \frac{dI_c}{dt}. \tag{8}$$

By integrating over $[t, t + \Delta t]$, we get:

$$I_c^{t+\Delta t} = I_c^t + K_s \Delta t + A \times \ln \left(\frac{((\psi + h_p)\Delta\theta + I_c^{t+\Delta t})}{(\psi + h_p)\Delta\theta + I_c^t} \right), \tag{9}$$

where K_s , ψ and $\Delta\theta$ are given according to the soil layer. The constants a and A are given by:

$$\begin{cases} a = 0, & A = \Delta\theta_1(\psi + h_p), & \text{if } d_f \leq d_1, \\ a = d_1 \Delta\theta_1, & A = \Delta\theta_2 \left(\psi + h_p - d_1 \left(\frac{K_2}{K_1} - 1 \right) \right), & \text{if } d_f > d_1. \end{cases} \tag{10}$$

3 Numerical Method

In this section, we derive the numerical method for the coupled model of surface and subsurface water flows. In order to satisfy the well-balanced property of the proposed numerical method, we used the vector of solution $U = (w, p, q)$ with new variables of the system. The variable $w = h + B$ is the water surface elevation, $p = hu$ and $q = hv$ are respectively the discharges in x - and y - directions. By expressing the system in Eq. 1 with vector variable U , we obtain its following matrix form:

$$\frac{\partial U}{\partial t} + \frac{\partial H_x(U, B)}{\partial x} + \frac{\partial H_y(U, B)}{\partial y} = S_b(U, B) + S_f, \tag{11}$$

where the vector of fluxes $F = (H_x, H_y)^T$, and source terms S_b and S_f are:

$$\begin{aligned} H_x &= \left(p, \frac{p^2}{w - B} + \frac{g}{2}(w - B)^2, \frac{pq}{w - B} \right)^T, \\ H_y &= \left(q, \frac{pq}{w - B}, \frac{q^2}{w - B} + \frac{g}{2}(w - B)^2 \right)^T, \\ S_b &= \left(R - I, -g(w - B) \frac{\partial B}{\partial x}, -g(w - B) \frac{\partial B}{\partial y}, 0 \right)^T, \\ S_f &= \left(0, -\frac{\tau_x}{\rho}, -\frac{\tau_y}{\rho}, 0 \right)^T. \end{aligned} \tag{12}$$

3.1 Surface Water Computation

We assume that the domain is partitioned into a set of triangular cells E_j of area $|E_j|$. Let E_{jk} , $k = 1, 2, 3$ be the neighbouring cells of E_j with common edges e_{jk} of length l_{jk} , and define $n_{jk} = (\cos(\theta_{jk}), \sin(\theta_{jk}))$ the unit vector normal to the edge e_{jk} and points toward the cell E_j . Denote by $G_j = (x_j, y_j)$ the coordinates of the center of mass of the cells E_j and $P_{jk} = (x_{jk}, y_{jk})$ the midpoint of the edge e_{jk} with corresponding vertices $P_{jki}, i = 1, 2, 3$, as shown in Fig. 1.

The second-order HLL-Riemann solver scheme on unstructured triangular grids [19] is applied to the coupled model of surface and subsurface water flows, which can be written as follows:

$$\frac{\partial \bar{U}_j}{\partial t} = \frac{1}{|E_j|} \sum_{k=1}^3 F_{jk} \cdot n_{jk} l_{jk} + \bar{S}_{b_j} + \bar{S}_{f_j}, \tag{13}$$

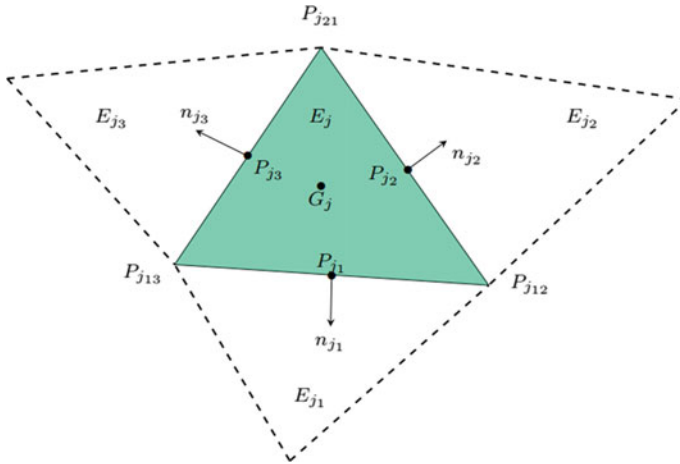


Fig. 1 The triangular grids

where $\bar{U}_j = \frac{1}{|E_j|} \int_{E_j} U(t, x, y) dx dy$ is the approximation of the cell average of the computed solution, and $\bar{S}_b = \frac{1}{|E_j|} \int_{E_j} S_b(U, B) dx dy$ and $\bar{S}_f = \frac{1}{|E_j|} \int_{E_j} S_f(t, x, y) dx dy$ are the cell averages of the source terms.

The advective numerical fluxes F_{jk} are:

$$F_{jk} \cdot n_{jk} = \begin{cases} \mathbf{F}_{jk}^L(U_{jk}^L, B_{jk}) \cdot n_{jk}, & \text{if } S_{jk}^L \geq 0, \\ \frac{S_{jk}^R \mathbf{F}_{jk}^L(U_{jk}^L, B_{jk}) \cdot n_{jk} - S_{jk}^L \mathbf{F}_{jk}^R(U_{jk}^R, B_{jk}) \cdot n_{jk} + S_{jk}^L S_{jk}^R (U_{jk}^R - U_{jk}^L)}{S_{jk}^R - S_{jk}^L}, & \text{if } S_{jk}^L \leq 0 \leq S_{jk}^R \\ \mathbf{F}_{jk}^R(U_{jk}^R, B_{jk}) \cdot n_{jk}, & \text{if } S_{jk}^R \leq 0. \end{cases} \tag{14}$$

where S_{jk}^L and S_{jk}^R are respectively the left and right values of wave speeds of propagation, which are estimated in our study using the extreme eigenvalues based on the left and right values of the water depth and velocities. The values $U_{jk}^L := U_j(P_{jk})$ and $U_{jk}^R := U_j(P_{jk})$, correspond respectively to the left and right reconstructed values of the vector variable U at the midpoints P_{jk} , which are obtained using the following linear approximation:

$$\tilde{U}(x, y) = \bar{U}_j + (U_x)_j(x - x_j) + (U_y)_j(y - y_j). \tag{15}$$

In order to ensure the positivity of the water depth h , this reconstruction is corrected for the water surface elevation w by introducing the parameter $\alpha \in [0, 1]$ [2]:

$$\tilde{w}(x, y) = \bar{w}_j + \alpha(w_x)_j(x - x_j) + \alpha(w_y)_j(y - y_j). \tag{16}$$

The parameter α is computed by requiring $w_j(P_{jk}) > 0$ and $w_{jk}(P_{jk}) > 0$. The numerical gradients in the piecewise linear reconstruction Eqs. 15 and 16 are obtained using the Green-Gauss theorem:

$$\nabla U_j = \frac{1}{|E_j|} \int_{E_j} \nabla U_j(x, y) dx dy = \frac{1}{|E_j|} \sum_{s=1}^3 \int_{E_j} \tilde{U}_{js} n_{js} de, \quad (17)$$

where the values \tilde{U}_{js} are estimated at the cell interface adopting the approach procedure developed in [2]. The required values of the bottom topography at the midpoints B_{jk} are computed from its known values at the vertices, obtained using a continuous piecewise linear approximation:

$$B_{jk} = \frac{B_{jk1} + B_{jk2}}{2}. \quad (18)$$

We used linear reconstruction for the bottom topography at each computational cell. The estimated values of the bottom topography at the midpoints are used to compute the elevation at the center of mass of the cell:

$$B_j = \frac{B_{j1} + B_{j2} + B_{j3}}{3}, \quad (19)$$

3.2 Discretization of the Source Terms

In order to guarantee the well-balance property of the numerical scheme, appropriate discretization of the cell averages of the source term \overline{S}_{b_j} is needed, to exactly balance the numerical fluxes. To this end, the well-balance discretization of the source term \overline{S}_{b_j} is given by [3]:

$$\begin{aligned} \overline{S}_{b_j}^{(1)} &= R_j - I_j, \\ \overline{S}_{b_j}^{(2)} &= \frac{g}{2|E_j|} \sum_{k=1}^3 l_{jk} (w_{jk}^L - B_{jk})^2 \cos(\theta_{jk}) - g(w_x)_j (\overline{w}_j - B_j), \\ \overline{S}_{b_j}^{(3)} &= \frac{g}{2|E_j|} \sum_{k=1}^3 l_{jk} (w_{jk}^L - B_{jk})^2 \sin(\theta_{jk}) - g(w_y)_j (\overline{w}_j - B_j). \end{aligned} \quad (20)$$

For the discretization of the Manning's friction term \overline{S}_{f_j} , we used the methodology developed in [18].

3.3 Subsurface Water Computation

In the previous section, we used the Green-Ampt equation describing the infiltration of water flow into soil where the water is available at the surface at all times. However, during rainfall event the water ponded at the surface if the rainfall intensity is greater than the infiltration capacity of the soil, otherwise all the water infiltrates into the soil. In this study, the infiltration rate is computed using the following techniques:

- If $R_j^n > I_{r,j}^n$: water is ponded at the surface. The infiltration rate is computed from Eqs. 3 to 10 by:

$$I_j^n = \min\left(\frac{\bar{h}_j^n}{\Delta t}, I_{r,j}^n\right). \quad (21)$$

- If $R_j^n \leq I_{r,j}^n$: no ponding occurs at the surface, then we compute the infiltration rate by:

$$\begin{cases} I_j^n = R_j^n, \\ I_{c,j}^{n+1} = I_{c,j}^n + I_j^n \times \Delta t. \end{cases} \quad (22)$$

4 Numerical Experiments

In this section, we will test the performance of the proposed numerical scheme and its ability to simulate surface and subsurface water flows. In our simulations, we used the gravitational constant $g = 9.81 \text{ m/s}^2$. In the first example, we perform a validation of our numerical techniques using experimental data for a simple laboratory non infiltrating slop surface. The second example focuses on the well-balanced and conservative properties of the scheme where we perform a numerical test over infiltrating surface using one-and two-layers soil. In the last example, we test the capability of the proposed numerical method to simulate surface–subsurface water flow over complex bottom topography.

4.1 Rainfall-Runoff Over a Slop Surface

This numerical example is considered to validate the proposed numerical techniques using experimental data [7]. The computational domain is a plane of length $L = 21.945 \text{ m}$ and width $l = 1 \text{ m}$ with a slop $S_x = 0.04$ which is discretized using triangular grids with an average cell area $|\bar{E}_j| = 3.9 \times 10^{-3}$. The soil surface is initially dry and the rainfall with an uniform intensity R is imposed over the domain. The Manning's coefficient $n = 0.5 \text{ m}^{-1/3} \text{ s}$ is used by [7]. Hydrographs of computed

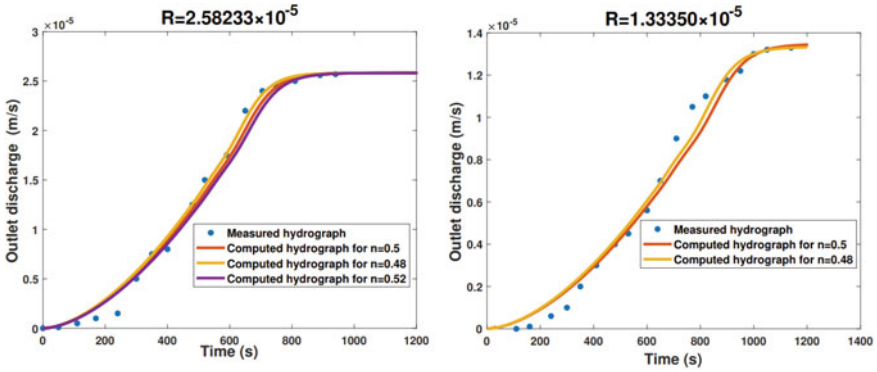


Fig. 2 Hydrographs of computed and measured outlet discharges for constant rainfalls over slope surface

and measured outlet discharges are presented in Fig. 2 (left) for different values of the Manning coefficient to show the impact of uncertainty of this coefficient on the results. A suitable value of the parameter n should be used to calibrate the proposed numerical model. We first calibrate the model where we compare the computed and measured outlet discharges to obtain the appropriate value of the Manning coefficient by minimizing the root mean square error (RMSE) defined as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{k=1}^N \left(q_k^{(s)} - q_k^{(o)} \right)^2}, \tag{23}$$

where $q_k^{(s)}$ and $q_k^{(o)}$ are the simulated and observed outlet discharges, and N is the total number of observations. We obtain $n = 0.48 \text{ m}^{-1/3} \text{ s}$ for the calibrated model and Fig. 2 (left) shows the corresponding results. To validate the calibrated model against other experimental data, numerical simulations are performed using another value of rainfall intensity as shown in Fig. 2 (right). We observe good overall agreement between experiments and simulations.

4.2 Well-Balanced and Conservative Properties

In this example, we test the well-balanced property and the conservation of the proposed numerical scheme. The proposed method is applied to solve the system Eq. 1 for one-and two-layers soil, where we considered the Silt loam soil of 30 percent initial effective saturation and Sandy loam soil of initial effective saturation of 40 percent. The Green-Ampt parameters of the used soils are given in Table 1 [17]. The computational domain $[0, 2] \times [0, 1]$ is discretized using triangular grids with an average cell area $|E_j| = 6.3735 \times 10^{-4}$. The surface water is initially constant

Table 1 Green-Ampt parameters

| | Soil | Suction at the wetting front ψ (cm) | Saturated hydraulic conductivity K_s (cm/h) | $\Delta\theta = \theta_s - \theta_i$ | d_1 (cm) |
|-------------|------------|--|---|--------------------------------------|------------|
| Upper layer | Sandy loam | 11.01 | 1.09 | 0.247 | 0.1 |
| Lower layer | Silt loam | 16.7 | 0.65 | 0.340 | |

everywhere with depth $h(0) = 0.1\text{ m}$ and zeros field velocity $\mathbf{u} = 0\text{ m/s}$. The wall conditions are applied at the boundaries of the domain.

Figure 3, represents the cross section along the x – axis of the computed water depth at different times over infiltrating surface and the corresponding total mass balance for one-and two-layers soil. The results show that the well-balanced property of the scheme is satisfied where water surface elevation decreases with time and remains constant in space. Moreover, the computed total mass of water which includes the surface water and the cumulative infiltrated water is preserved with time, that is $h(t) + I_c(t) = 0.1 = h(0)$.

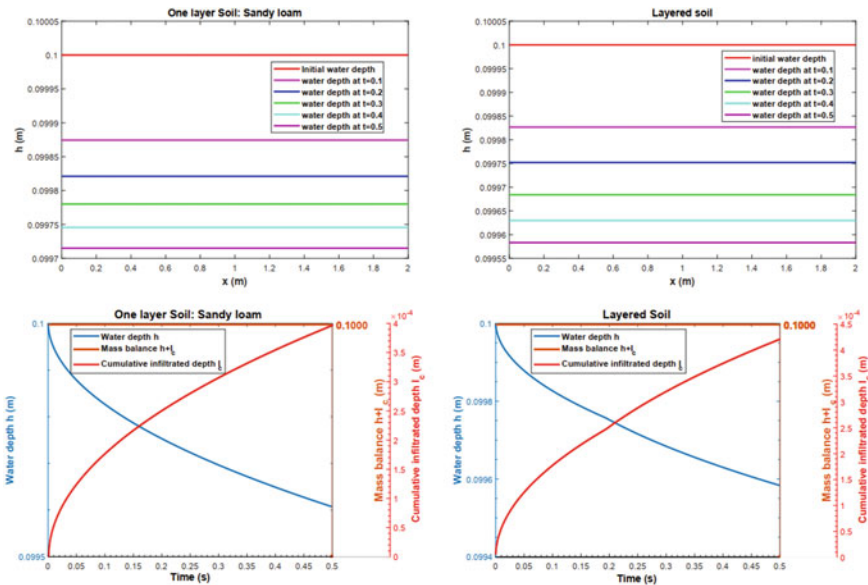


Fig. 3 Computed water depth and the corresponding total mass for one layer soil: Sandy loam (left) and two-layers soil (right)

4.3 Surface–Subsurface Flows Over Complex Basin

We consider a 2D modified numerical example used in [13], to test the ability of the proposed scheme for simulating runoff processes. In this example, we used the computational domain $[0, 10] \times [0, 8]$ and the following variable topography:

$$B(x, y) = 0.01y + 0.01|x - 0.5| - 0.01 \sin\left(\frac{\pi x}{2}\right) - 0.01 \sin\left(\frac{\pi y}{2}\right) + 1. \quad (23)$$

The Manning’s roughness coefficient is set to $n = 0.013 \text{ m}^{-1/3} \text{ s}$, and wall boundary conditions are imposed at the sides and upstream boundaries, while outflow condition is used at the downstream of the domain. We used the hydraulic conductivity of the soil $K_s = 7 \text{ mm/h}$, suction head $\psi = 50 \text{ mm}$ and $\Delta\theta = 0.125$. A constant rainfall with intensity $R = 500 \text{ mm/h}$ is applied over the domain for 5 min.

Figure 4 illustrates the evolution of the water surface elevation w , the water depth h and the component of the velocity in y-direction at different times. Initially, the domain is dry and at time $t = 1 \text{ min}$, the water starts ponding at the surface and filled ponds in the upper areas, the excessive water overflows the ponds and the lower areas inside the domain. At time $t = 5 \text{ min}$, the rain stops while water keeps moving at

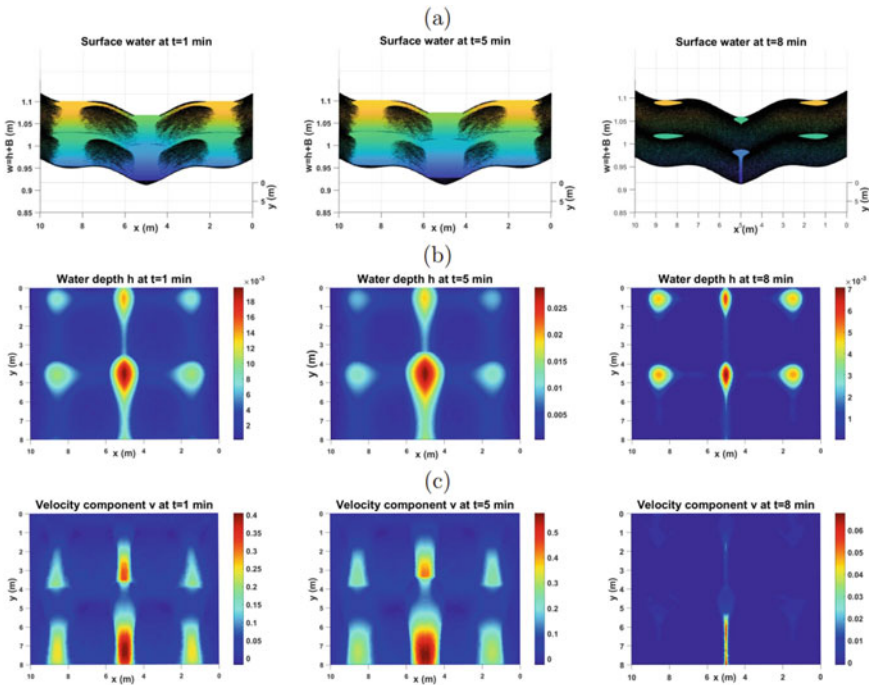
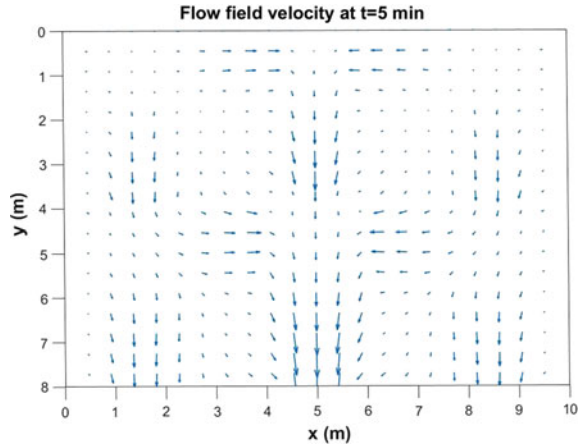


Fig. 4 Evolution of the computed surface water (a), water depth (b) and component of the velocity in y-direction (c) at times $t = 1 \text{ min}$, $t = 5 \text{ min}$ and $t = 8 \text{ min}$

Fig. 5 Flow field velocity vector of the computed solution at time $t = 5$ min



the surface to the downstream of the domain, and most areas became flooded with time. At time $t = 8$ min the upper areas of the domain become almost dry except inside the ponds. In Fig. 5, we show the flow field velocity vectors of the computed solution at time $t = 5$ min. We observe a rapid increase of the flow velocity in the downstream. Finally, we conclude that our numerical simulations using the proposed numerical techniques reproduce physically reasonable rainfall-runoff processes over complex bottom topography with wet and dry area and large bed slop.

5 Conclusion

We proposed numerical techniques for modelling coupled surface and subsurface water flows. The shallow water equations are used for surface water flow over variable bottom topography with source terms due to friction effect, precipitation and infiltration. For subsurface water flow, we used the Green-Ampt infiltration model with one-and two-layers soil. HLL Riemann solver finite volume schemes with linear reconstruction of the solutions are implemented for solving the coupled nonlinear system. Appropriate discretization techniques are used for the bottom topography and Manning friction source terms to ensure the underlying physical properties of the proposed numerical scheme. Moreover, we used accurate techniques for subsurface water computation under rainfall-runoff processes. Our numerical results confirm the accuracy and the capability of the proposed numerical techniques for modelling surface–subsurface water flows over variable bottom topography involving wet and dry areas and large bed slop.

References

1. Aliparast M (2009) Two-dimensional finite volume method for dam-break flow simulation. *Int J Sedim Res* 24(1):99–107
2. Beljadid A, Mohammadian A, Kurganov A (2016) Well-balanced positivity preserving cell-vertex central-upwind scheme for shallow water flows. *Comput Fluids* 136:193–206
3. Bryson S, Epshteyn Y, Kurganov A (2011) Well-balanced positivity preserving central-upwind scheme on triangular grids for the Saint-Venant system. *ESAIM Math Model Numer Anal* 45(3):423–446
4. Cea L, Vazquez-Cendon ME (2012) Unstructured finite volume discretisation of bed friction and convective flux in solute transport models linked to the shallow water equations. *J Comput Phys* 3317–3339
5. De St Venant B (1871) Theorie du mouvement non-permanent des eaux avec application aux crues des rivières et a l'introduction des Mares dans leur lit. *Academic de Sci Comptes Redus* 148–154
6. Fernández-Pato J, Gracia JL, García-Navarro P (2018) A fractional-order infiltration model to improve the simulation of rainfall/runoff in combination with a 2D shallow water model. *J Hydroinf* 20(4):898–916
7. Gottardi G, Venutelli M (2008) An accurate time integration method for simplified overland flow models. *Adv Water Resour* 31(1):173–180
8. Green WH, Ampt GA (1911) *Studies on soil physics*. Cambridge University Press, Cambridge
9. Harten A, Lax PD, Van Leer B (1983) On upstream differencing and Godunov-type schemes for hyperbolic conservation laws. *SIAM Rev* 25(1):35–61
10. Liu G, Craig JR, Soulis ED (2011). Applicability of the Green-Ampt infiltration model with shallow boundary conditions. *J Hydrol Eng* 16(3):266–273
11. Marcus WA, Roberts K, Harvey L, Tackman G (1992) An evaluation of methods for estimating Manning's n in small mountain streams. *Mountain Res Dev* 227–239
12. Mugler C, Planchon O, Patin J, Weill S, Silvera N, Richard P, Mouche E (2011) Comparison of roughness models to simulate overland flow and tracer transport experiments under simulated rainfall at plot scale. *J Hydrol* 63(8):25–40
13. Park S, Kim B, Kim DH (2019) 2D GPU-accelerated high resolution numerical scheme for solving diffusive wave equations. *Water* 11(7):1447
14. Pham TT, Mai TD, Pham TD, Hoang MT, Nguyen MK, Pham TT (2016) Industrial water mass balance as a tool for water management in industrial parks. *Water Resour Ind* 13:14–21
15. Taccone F, Antoine G, Delestre O, Goutal N (2020) A new criterion for the evaluation of the velocity field for rainfall-runoff modelling using a shallow-water model. *Adv Water Resour* 140:103581
16. Tatard L et al (2008) Measurement and modelling of high-resolution flow-velocity data under simulated rainfall on a low-slope sandy soil. *J Hydroinf* 348(1–2):1–12
17. Te Chow V, Maidment DR, Mays LW (1988) *Applied hydrology*. McGraw Hill Inc., New
18. Xia X, Liang Q (2018) A new efficient implicit scheme for discretising the stiff friction terms in the shallow water equations. *Adv Water Resour* 117:87–97
19. Yoon TH, Kang SK (2004) Finite volume model for two-dimensional shallow water flows on unstructured grids. *J Hydraul Eng* 130(7):678–688

Development of New Extreme Rainfall Maps for Urban Infrastructure Design in Canada Using the Scale-Invariance Generalized Extreme Value Distribution



Truong-Huy Nguyen and Van-Thanh-Van Nguyen

1 Introduction

Information on the spatio-temporal variability of extreme rainfall characteristics is of critical importance for many types of hydrologic studies related to the estimation of runoffs for planning, design, and management of various water resources systems [1, 2]. In particular, for urban and small rural watersheds that are generally characterized by fast response, the designs of various hydraulic structures such as small dams, culverts, storm sewers, detention basins, and so on, require extreme rainfall input with very short time durations (e.g., few minutes or hours) for runoff simulation models [1]. More specifically, this information is extracted from the “intensity–duration–frequency” (IDF) relations, which provides extreme rainfall intensities for various durations and return periods at a given site of interest [2, 3].

In current engineering practice, the IDF relations are derived based on statistical frequency analyses of annual maximum (AM) rainfall series for different durations where available rainfall records of adequate lengths could be used to estimate the parameters of a selected appropriate probability distribution model. However, these AM rainfall records for short durations (e.g., less than one day) are often limited or unavailable at the location of interest because of the high measurement costs involved while those for the daily scale are widely available. Hence, there exists an urgent need to develop new methods for modeling extreme rainfall processes over a wide range of time scales such that information related to sub-daily extreme rainfalls could be inferred from the daily extreme rainfalls available at a given site.

T.-H. Nguyen (✉) · V.-T.-V. Nguyen
McGill University, Montreal, Canada
e-mail: huy.nguyen5@mail.mcgill.ca

T.-H. Nguyen
The University of Danang – University of Science and Technology, Da Nang, Vietnam

In Canada, Environment Canada (EC) provides short-duration extreme rainfall data for nine different rainfall durations ($D = 5, 10, 15, 30, 60, 120, 360, 720,$ and 1440 min) and the IDF relations for approximately 650 stations across Canada with at least 10-year record [4]. The estimated extreme rainfalls for six different return periods ($T = 2, 5, 10, 25, 50,$ and 100 years) were obtained by fitting the two-parameter Gumbel distribution to the AM series for each rainfall duration independently using the method of moments. However, it has been widely known that the three-parameter Generalized Extreme Value (GEV) distribution using the L-moment estimation method can provide more accurate extreme rainfall estimates [5–7]. In addition, rather than fitting the GEV distribution to AM rainfall series for each duration independently as commonly used in the traditional method, the scale-invariance GEV model has been shown to provide more robust estimates of extreme rainfalls since it can take into account the relationship between the statistical properties of the extreme rainfall processes over different durations [8, 9]. The scale-invariance concept has increasingly become a promising methodology for modeling of various extreme hydrological processes across a wide range of time scales and has been applied to updating the IDF curves considering climate change impacts in recent years [8–14].

In view of the above-mentioned issues, this study presents the development of new at-site IDF relations as well as regional extreme rainfall maps for urban water infrastructure design for Canada using the scale-invariance GEV model. All available historical AMS data for nine different rainfall durations from a network of 651 raingauges located across Canada were used in this study. It has been found that the newly developed IDF relations and regional extreme rainfall maps based on the proposed scaling GEV model can provide more accurate and more robust extreme design rainfalls than those given by the traditional method using the Gumbel model by the EC. Details regarding the study sites and data are described in Sect. 2. The methodology used in the development of these new products is presented Sect. 3. Section 4 provides the results and discussions. Research findings are summarized in Sect. 5.

2 Study Sites and Data

In the present study, all available AM historical rainfall data for nine different rainfall durations ranging from 5 min to 24 h from a network of 561 raingauges located across Canada were selected as shown in Fig. 1a. These AM data were obtained through the IDF_v3-10 program of the engineering climate dataset from the website of the Government of Canada [4]. The record length for these AM rainfall series varies from 10 to 82 years. Most of the data are recorded after 1960 and updated to 2017. Approximately, two third of these stations are less than 30 years of records in which stations between 10 to 20 years are composed of 40% and between 20 and 30 years cover 27%. Stations with very long record (at least 50 years) only occupy 6%. In total, more than 5000 AM rainfall series were considered in this study.

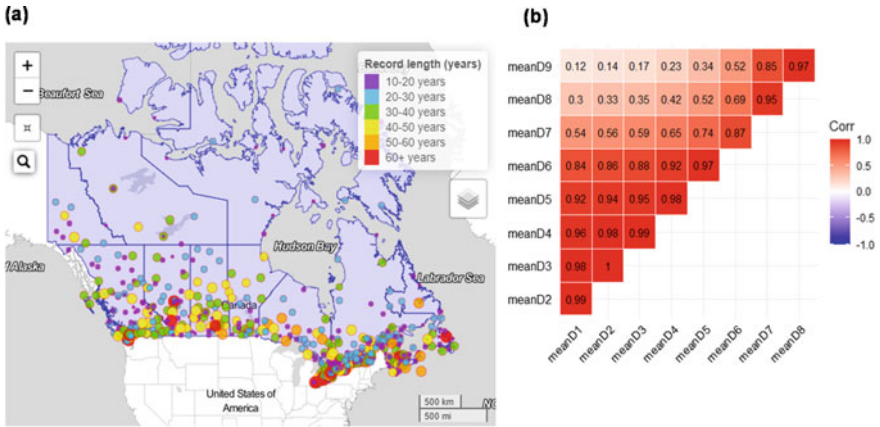


Fig. 1 a Locations and record lengths of 561 stations. b Correlogram of means of annual maximum (AM) rainfalls of nine different rainfall durations (D = 5, 10, 15, 30, 60, 120, 360, 720, and 1440 min). Note: meanD1 denotes mean of 5-min AM rainfall series and meanD9 denotes mean of 1440-min AM rainfall series

As shown in Fig. 1a, the selected stations are located across the vast land of Canada to capture possibly the true image of the spatial variations of extreme rainfall events induced by the diverse climatic regions and the complex topography of Canada. However, the raingauge density is not equal between these different Canadian regions. In fact, only 5% of the stations are scattered in the three Northern territories whereas 95% of stations are mainly located in the Southern provinces of Canada where most people live. Even between these provinces, raingauge density also varies significantly. The three largest provinces including the Pacific province (British Columbia) and Central provinces (Ontario and Quebec) hold 66% of the number of stations while the three Prairie provinces (Alberta, Saskatchewan, and Manitoba) cover 18%, and the remaining four Atlantis provinces (New Brunswick, Nova Scotia, Prince Edward Island, and Newfoundland and Labrador) only occupy 11%.

The mean extreme rainfall values of the AM series available at different locations were computed for each rainfall duration. The Pearson correlation coefficients (CC) between these computed mean values for all nine different rainfall durations were analyzed and plotted in Fig. 1b. In general, results show that there exist strong correlations between these mean values; especially, between any pair of nearest successive rainfall durations (CC larger than the minimum value of 0.87 between 120-min and 360-min extreme rainfalls). In particular, the correlations between extreme rainfalls for very short durations (from 5 min to 1 h) are very strong (CC larger than the minimum value of 0.92 between 5-min and 1-h extreme rainfalls). Therefore, the strong correlations between extreme rainfalls over different rainfall durations as identified in this study should be taken into consideration when performing frequency analyses of these extreme rainfall series rather than treating them independently as in existing traditional methods.

3 Methodology

3.1 Point-Scale Rainfall IDF Estimation

The scaling-invariance approach based on the GEV distribution [8, 9] was used to derive point-scale extreme rainfall IDF relations. The procedure consists of the following steps:

- (i) calculate probability weighted moments (PWMs) of rainfall amounts for each AMS from 5-min to 1-day duration,
- (ii) fit regression models to PWMs over different rainfall durations,
- (iii) estimate the scaling exponents and derive moments of sub-daily and sub-hourly durations from those of daily duration,
- (iv) derive the distribution of each sub-daily and sub-hourly annual maximum series (AMS),
- (v) use the derived distribution to compute rainfall intensity for a given rainfall duration and return period of interest.

First, PWMs of each AMS from 5-min to 24-h rainfall duration are calculated for a given location of interest. Next, regression models are constructed based on the relationships between the computed PWMs of rainfall amounts for different rainfall durations. Based on these statistical relationships, scaling exponents can be then estimated. Some previous studies have indicated that extreme rainfall processes for durations ranging from a few minutes to several days could be well approximated by only one or two scaling regime(s) (see, e.g., [8, 9, 12, 14, 15]). By investigating the log–log plot of the relationships between rainfall PWMs versus rainfall durations, one can identify one scaling rainfall regime (if the plot displays a straight line and does not indicate any break point) or two different scaling rainfall regimes (if the plot shows a break point) over the selected time scales. For a large number of stations, it is simpler to assume one break point and then check the scaling exponents of the two scaling regimes. If they are equal or approximately equal, then there exists only one scaling regime. A numerical criterion was employed to automate the procedure of identification of the breakpoint location and to avoid the subjectivity [8]. The procedure first fits a piecewise (or segmented) linear regression model to the log–log plot of the first three statistical moments versus rainfall durations near one end and compute model residuals. It then iterates through all middle points to other end and finally compares the ranking of the residuals to determine the best fit such that the group-wide and the total residuals are minimized. Only the first three statistical moments are used in the search because it is sufficient to estimate the three parameters of the GEV distribution. Once scaling exponents are computed, a scaling model is constructed which allows to derive moments of rainfall amount of sub-hourly and sub-daily durations from those of daily values. From the derived moments, a theoretical distribution model and its parameters are estimated based on the method of L-moment for each sub-daily and sub-hourly AMS. Rainfall intensity for a given rainfall duration and return period of interest can be easily obtained based on the

derived distribution, even for an “unobserved” duration between 5-min to 1-day duration. Since this scaling GEV model considers all extreme rainfall data for all different durations available at a site and the relationships between these rainfall durations, the results are more consistent and more robust as compared to those given by the traditional technique that fit the chosen probability model to observed AM data for each duration independently [3, 7].

3.2 Grid-Scale Extreme Rainfall Estimation

At-site IDF estimation procedure provides the computed extreme design rainfall quantiles for a given location with available extreme rainfall data (i.e., a gauged site). However, given the low density of IDF sites for its very large territory, it is a common situation in Canada that observed data are unavailable at the location of interest (i.e., an ungauged site). Consequently, it is necessary to transfer information from the gauged locations to ungauged locations. Spatial interpolation techniques are frequently used in practice to tackle this issue by fitting a surface (represented by a raster dataset) to all IDF sample measurements points (i.e., control points). Based on the fitted surface, value at any given location in the output raster can be predicted. The number and distribution of control points can greatly influence the accuracy of the computed spatial interpolation.

There are several procedures to perform spatial interpolation and each method is often referred to as a geospatial fitting model. Each model applies different computational scheme and there are also different assumptions associated with each model. In general, these spatial interpolation techniques are categorized into the non-geostatistical and geostatistical approaches [16]. The non-geostatistical methods assign values to a target location (or cell) based on the surrounding measured values using a specific mathematical equation that determine the smoothness of the resulting surface. Several popular methods in this category include the inverse distance weighting, proximity (or Thiessen method), natural neighbor, and spline. The geostatistical methods rely on the statistical relationship among the measured points (often known as semi-variogram or variogram in short) to estimate the value for a target location (or cell). With the variogram, geostatistical techniques are not only able to predict value at a target site but also provide some measure of uncertainties associated with the calculation. Several popular techniques in this category include the ordinary kriging and universal kriging methods.

In this study, the authors use the inverse square distance weighting (IDW) approach for the spatial interpolation. This simpler approach has been commonly used in many comparative studies for spatial interpolation of extreme rainfalls [17, 18]. It has been shown to provide reasonable results compared to the more complicated approaches such as ordinary kriging. The IDW method assumes that each measured point has a local influence that diminishes with distance. It uses a simple mathematical model to predict value for an unmeasured location based on the values of control points surrounding the target location and the distances between them. The control points

closer to the prediction location have more influence on the predicted value than those farther away.

3.3 Model Assessment Criteria

In this study, two dimensionless indices are used for comparing the performance of the selected models. These criteria include the coefficient of determination (R^2) and the mean absolute relative deviation (MADr) as given in the following equations:

$$R^2 = \left[\frac{\sum \{(x_i - \bar{x})(y_i - \bar{y})\}}{\left\{ \sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2 \right\}^{\frac{1}{2}}} \right]^2 \quad (1)$$

$$\text{MADr} = \frac{1}{n} \sum \left\{ \frac{|x_i - y_i|}{x_i} \right\} \quad (2)$$

where $x_i, i = 1, 2, \dots, n$ are the observed values and $y_i, i = 1, 2, \dots, n$ are the estimated values; n is the sample length; \bar{x} and \bar{y} denote the average value of the observed and estimated quantiles, respectively.

4 Results and Discussion

4.1 Model Performance Assessment

To assess the performance of the scaling model in deriving the distribution of sub-daily and sub-hourly AM rainfalls, the estimated short-duration extreme rainfall quantiles of different durations ($D = 5$ min to 12 h) and different return periods ($T = 2$ to 50 years) were compared with the observed data obtained from the at-site frequency analysis using the GEV distribution. Since the average record length of all stations is 25.5 years, only extreme rainfall quantiles of return periods up to $T = 50$ years were chosen for comparisons to obtain reliable rainfall estimates. The estimation of rainfall quantiles for return periods higher than twice the sample length involves high degree of uncertainties due to the extrapolation and should be used with caution. Two indices including R^2 and MADr were used to investigate model performance for different rainfall durations, different return periods, different geographical locations, and different sample record lengths.

For different rainfall durations and return periods, Table 1 shows that the estimated extreme rainfall quantiles given by the proposed GEV scaling model agree very well with the observed data. In particular, for rainfall quantiles of return periods within the average record length, the lowest R^2 is 0.93 and the highest MADr is 7.3%. For

Table 1 R² and MADr results for different rainfall durations (5 min to 12 h) and different return periods (2–50 years)

| Return period | R ² (dimensionless) for different rainfall durations | | | | | | | |
|---------------|---|--------|--------|--------|------|------|------|------|
| | 5 min | 10 min | 15 min | 30 min | 1 h | 2 h | 6 h | 12 h |
| 2 years | 0.99 | 0.98 | 0.97 | 0.97 | 0.97 | 0.96 | 0.95 | 0.98 |
| 5 | 0.99 | 0.98 | 0.98 | 0.98 | 0.98 | 0.96 | 0.95 | 0.98 |
| 10 | 0.98 | 0.97 | 0.97 | 0.98 | 0.98 | 0.96 | 0.94 | 0.97 |
| 25 | 0.97 | 0.94 | 0.95 | 0.96 | 0.96 | 0.93 | 0.88 | 0.95 |
| 50 | 0.94 | 0.90 | 0.91 | 0.93 | 0.94 | 0.89 | 0.82 | 0.91 |
| Return period | MADr (%) for different rainfall durations | | | | | | | |
| | 5 min | 10 min | 15 min | 30 min | 1 h | 2 h | 6 h | 12 h |
| 2 years | 3.1 | 5.5 | 5.9 | 4.7 | 4.2 | 4.7 | 5.8 | 4.9 |
| 5 | 3.1 | 4.6 | 5 | 4.4 | 4 | 4.7 | 5.6 | 4.6 |
| 10 | 3.4 | 4.8 | 4.9 | 4.3 | 4.3 | 5.4 | 6.2 | 4.8 |
| 25 | 4.8 | 6.6 | 6.4 | 5.8 | 5.8 | 7.3 | 8.6 | 6.3 |
| 50 | 6.5 | 8.8 | 8.7 | 7.9 | 7.8 | 9.7 | 11.3 | 8.1 |

Color scales are used to aid result visualization with the darkest representing the worst (lowest R² or highest MADr) and brightest representing the best result (highest R² or lowest MADr)

those of return periods equal twice the average record length, R² is 0.89 and MADr is 9.7%. Except for the 6-h duration, R² and MADr values are slightly lower. This could be due to the complicated atmospheric conditions that cause mixed extreme rainfall process for these rainfall durations. That is, severe thunderstorm events of very short durations lasting minutes to hours are embedded in a low pressure or frontal system lasting from few hours to a day. Notice that these extreme rainfall quantiles are also often underestimated in the traditional approach when using a linear regression model between rainfall intensities over different rainfall durations.

Comparison between the estimated and observed data for each province and territory of Canada indicate that the scaling approach works well for different climate regions in Canada. In general, the two indices show that the proposed scaling GEV model provides similar performance for all provinces and territories and none of them is exceptional. In particular, R² values are very high (at least 0.95) and MADr values are relatively low (less than 12%) for most of stations. Except a few stations have slightly lower accuracy with R² and MADr approximately 0.9 and 16%, respectively. These stations mainly less than 20 years of record. Investigation of the effect of record lengths on the agreement between the estimated and observed data also show that when the sample length increases, the correlation tends to be higher and the difference between them reduces. In detail, R² gradually increases from 0.9 (for stations with less than 20-year record) to 0.98 (for stations more than 50-year record) while MADr decreases from 16 to 7%.

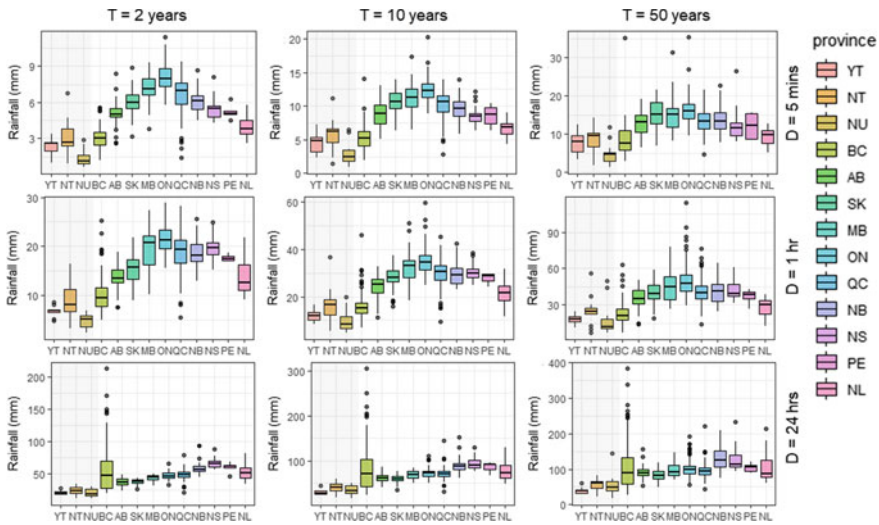


Fig. 2 Boxplots of extreme design rainfalls of different durations ($D = 5$ min, 1 h, 24 h in rows) and return periods ($T = 2, 10, 50$ years in columns) for each Canadian province and territory. Note: YT = Yukon, NT = Northwest Territories, NU = Nunavut, BC = British Columbia, AB = Alberta, SK = Saskatchewan, MB = Manitoba, ON = Ontario, QC = Quebec, NB = New Brunswick, NS = Nova Scotia, PE = Prince Edward Island, NL = Newfoundland and Labrador

4.2 Point- and Grid-Scale IDF Relations

The point-scale IDF relations were generated for all study stations across Canada. Results for several selected durations and return periods frequently used in practice are summarized in Fig. 2. Similar to the distributions of the means of extreme rainfalls in different provinces and territories across Canada, results show the three main patterns of spatial distribution of the extreme rainfall events in the range of (i) 5-min to less than 1-h duration, (ii) 1-h to less than 6-h duration, and (iii) 6-h to 24-h duration.

For extreme rainfalls of very short durations ranging from 5 min to less than 1 h, results show that, moving from the South to the North, the extreme design rainfalls decline (see Figs. 2 and 3a). Whereas, moving from the West to the Central, extreme rainfalls increase and then decrease when moving further from Central to the East. In fact, the highest values occur in Northwest Territories for the Northern Canada and in Ontario province for the Southern part. For instance, on average, the 5-min design rainfalls of 10-year return period increase from 4 mm in Yukon to 6 mm in Northwest Territories and then decrease to 2.5 mm in Nunavut. Whereas, in the South, they rise from 5 mm in British Columbia to 12.5 mm in Ontario and then decline to 7 mm in Newfoundland and Labrador. Similarly, the 10-min, 15-min, and 30-min extreme rainfalls display the same behaviors to what were observed in the 5-min design values.

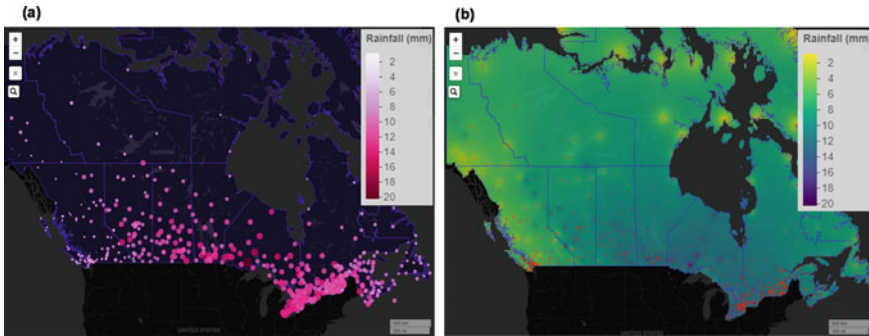


Fig. 3 An example of **a** point-scale and **b** grid-scale extreme design rainfalls of 5-min duration and of 10-year return period. For **(a)**, at-site rainfall depths are represented in the form of circle markers with continuous diameter and color scales (larger diameters and darker colors represent higher values). For **(b)**, rainfall depths are plotted using continuous color scale (darker colors represent higher values) while locations of control points are represented with red dots

For extreme rainfalls of short durations ranging from an hour to less than 6 h, results show a similar pattern to what observed in shorter durations events for the Northern Territories. The Southern provinces, however, display a relative increase in extreme rainfall amounts in the Atlantic region compared to the other regions. For example, for the Northern part, on average, 1-h extreme design rainfalls of 10-year return period increase from 12 mm in Yukon to 16 mm in Northwest Territories and then reduce to 8 mm in Nunavut. For the Southern regions, there are two peaks of extreme design rainfalls when moving from the West to the East Coast. For instance, on average, 1-h design rainfalls of 10-year return period go up from 15 mm in British Columbia to the first peak of 35 mm in Ontario province and then dwindle to 30 mm in Quebec and New Brunswick. They then increase back and reach the second peak of 31 mm in Nova Scotia (slightly higher than New Brunswick) and then go down to 22 mm in Newfoundland and Labrador province.

For extreme rainfalls of durations ranging from 6 to 24 h, similar pattern was found for the Northern Canada with the peak in Northwest Territories but the difference in rainfall depths between the three territories are much lower. For the Southern part, the largest extreme design rainfalls shift to Nova Scotia. Except British Columbia, other provinces follow a distinct pattern. For example, the 24-h design rainfalls of 10-year return period increase from 60 mm in Alberta to 90 mm in Nova Scotia and then decrease to 75 mm in Newfoundland and Labrador. Compared to other regions, British Columbia displays an exceptional behavior with a very large variation (i.e. very large box and long whiskers) of design rainfall values between its gauged stations (i.e. design rainfalls range between 25 and 175 mm). Furthermore, the largest station values also occurred in British Columbia rather than in Nova Scotia. Several stations show the values as large as 200–300 mm compared to the largest value in Nova Scotia which is only 130 mm. These stations mainly located on the Northwest of

Moresby Island, on the Southwest side of Vancouver Island, and in the South of North Vancouver.

For ungauged sites where the developed point-scale IDF relations are unavailable, spatial interpolation based on the inverse square distance weighting method were used to estimate the values at these locations by transferring information from the neighboring IDF stations. The computed extreme design rainfall atlas for each particular rainfall duration and return period was presented in the form of a raster. These rasters have the same resolutions and spatial extents with the ANUSPLIN daily rainfall series product produced by Natural Resources Canada [19]. For illustrative purposes, Fig. 3b shows the computed grid-scale extreme design rainfalls for 5-min duration and for 10-year return period. It can be seen that in the Southern part of Canada where the gauged density is high, the interpolated values are very consistent and highly agree with the general trend of spatial distribution and variation of extreme rainfalls from the West to East Coast discussed earlier based on the sampling values at the control points. However, in the Northern Canada, due to a low density of the observed stations, most of the interpolated values do not follow the general trend and they are often overestimated. More control points are required to enhance the accuracy of the estimation for these regions. These could come from other data sources and products such as radar rainfalls. Merging these products could be another challenge. In addition, other geospatial interpolation techniques, such as thin-plate spline, ordinary kriging, could be used and compared with the current IDW-based product to improve the results.

5 Conclusions

Extreme rainfall IDF relations are considered as a critical tool for the design of various urban water infrastructures. In Canada, these IDF relations are derived based on statistical frequency analyses of annual maximum rainfall series (AMS) for nine different durations (ranging from 5 min to one day) by fitting the two-parameter Gumbel distribution to the observed extreme rainfall data independently for each rainfall duration. The present study proposed a new procedure for developing new at-site IDF relations as well as regional extreme rainfall maps for urban infrastructure design for Canada. The proposed procedure was based on an improved three-parameter scaling GEV model to provide more accurate and more robust estimates of extreme design rainfalls.

In this study, a data set of more than 5000 annual maximum rainfall series (AMS) for nine different rainfall durations from a network of 651 stations located across Canada was selected. First, a detailed analysis of these AMS was performed to identify the scaling behaviour of these extreme rainfall processes. In general, it was found that the extreme rainfall events exhibit different scaling regimes over different regions with diverse climatic conditions. On the basis of this scaling investigation, the scaling GEV distribution was selected to describe the distribution of extreme rainfalls over a wide range of time scales (from 5 min to one day). By accounting for the strong

correlation between statistical moments of rainfall amounts over different rainfall durations and based on the PWM method, it has been shown that the proposed scaling GEV model can provide more accurate and more robust extreme rainfall estimates than those values given by traditional methods. More specifically, the computed extreme rainfalls for different rainfall durations, return periods, geographical locations, and sample lengths were found to be highly agree with the observed data. For example, for rainfall quantiles of return periods within the average record length, the lowest R^2 is 0.93 and the highest MADr is 7.3%.

The present study provided the at-site IDF relations for all 561 study sites located across Canada. General trends of spatial distribution and variation of extreme design rainfalls were also investigated. In general, there are three main patterns associated with extreme events of less than 1 h, from 1 h to less than 6 h, and from 6 h to 1 day. In particular, moving from South to North, extreme rainfalls decrease for all three patterns. However, moving from the West Coast to Central regions, extreme rainfalls of less than 1-h duration increase and then decrease when moving further from Central land to the East Coast. For those of 1-h to less than 6-h duration, after first peaking in Ontario, they rise back and reach the second peak in Nova Scotia. For extreme rainfalls of 6-h duration and longer, the peaks shift from Ontario to Nova Scotia. However, British Columbia is the place where many largest extreme events occurred.

Grid-scale regional extreme rainfall maps were also developed for locations without observed data in Canada. The spatial interpolation technique based on the inverse square distance weighted model was applied to transfer the values from control points (i.e. observed IDF stations) to the ungauged sites. Results show that for the Southern regions where the control point density is high, the results highly agree with the spatial distribution and variation trend from West to East. However, for the Northern territories where the gauged density is low, the interpolated values are often over-estimated. It is suggested that more data and other interpolation techniques should be considered in order to improve the estimation results at these regions.

Finally, it is expected that the development of improved at-site IDF relations and regional extreme rainfall maps as presented in this study could provide more cost-effective design of urban water infrastructures in Canada.

Acknowledgements The authors would like to acknowledge the funding supports provided by NSERC (Discovery Grant and Canadian FloodNet Strategic Network Grant) for this project.

References

1. CSA (2019) Technical guide: development, interpretation and use of rainfall intensity-duration-frequency (IDF) information: guideline for Canadian water resources practitioners. Canadian Standard Association, Toronto
2. WMO (2009) Guide to hydrological practices. Volume 2: management of water resources and application of hydrological practices. World Meteorological Organization

3. Nguyen T-H, Nguyen V-T-V (2019) Decision-support tool for constructing robust rainfall IDF relations in consideration of model uncertainty. *J Hydrol Eng* 24(7):06019004. [https://doi.org/10.1061/\(asce\)he.1943-5584.0001802](https://doi.org/10.1061/(asce)he.1943-5584.0001802)
4. Environment Canada (2020) Rainfall intensity-duration-frequency (IDF) tables and graphs. Version 3–10. Released date: March 2020. Retrieved from https://climate.weather.gc.ca/prods_servs/engineering_e.html. Accessed on 1 Oct 2020
5. Ball J et al (2016) Australian rainfall and runoff: a guide to flood estimation. Commonwealth of Australia
6. Perica S, Dietz S et al (2014) NOAA Atlas 14. Precipitation-frequency Atlas of the United States, California, Version 2.3, vol 6. National Weather Service, Silver Spring
7. Nguyen T-H, El Outayek S, Lim SH, Nguyen V-T-V (2017) A systematic approach to selecting the best probability models for annual maximum rainfalls—a case study using data in Ontario (Canada). *J Hydrol* 553:49–58. <https://doi.org/10.1016/j.jhydrol.2017.07.052>
8. Nguyen T-H, Nguyen V-T-V (2018) A novel scale-invariance generalized extreme value model based on probability weighted moments for estimating extreme design rainfalls in the context of climate change. In: World environmental and water resources congress
9. Nguyen T-H, Nguyen V-T-V (2020) Linking climate change to urban storm drainage system design: an innovative approach to modeling of extreme rainfall processes over different spatial and temporal scales. *J Hydro-environ Res* 29:80–95. <https://doi.org/10.1016/j.jher.2020.01.006>
10. Blanchet J, Ceresetti D, Molinié G, Creutin JD (2016) A regional GEV scale-invariant framework for intensity–duration–frequency analysis. *J Hydrol* 540:82–95. <https://doi.org/10.1016/j.jhydrol.2016.06.007>
11. Ghanmi H, Bargaoui Z, Mallet C (2016) Estimation of intensity-duration-frequency relationships according to the property of scale invariance and regionalization analysis in a Mediterranean coastal area. *J Hydrol* 541:38–49. <https://doi.org/10.1016/j.jhydrol.2016.07.002>
12. Nguyen VT, Nguyen TD, Ashkar F (2002) Regional frequency analysis of extreme rainfalls. *Water Sci Technol J Int Assoc Water Pollut Res* 45(2):75–81
13. Nguyen VTV, Nguyen TD, Cung A (2007) A statistical approach to downscaling of sub-daily extreme rainfall processes for climate-related impact studies in urban areas. In Fang HHP, Lee JHW (eds) *Water science and technology: water supply*
14. Yu P-S, Yang T-C, Lin C-S (2004) Regional rainfall intensity formulas based on scaling property of rainfall. *J Hydrol* 295(1–4):108–123. <https://doi.org/10.1016/j.jhydrol.2004.03.003>
15. Li J, Heap AD (2014) Spatial interpolation methods applied in the environmental sciences: a review. *Environ Model Softw* 53:173–189
16. Das S (2019) Extreme rainfall estimation at ungauged sites: Comparison between region-of-influence approach of regional analysis and spatial interpolation technique. *Int J Climatol* 39(1):407–423. <https://doi.org/10.1002/joc.5819>
17. Haberlandt U (2007) Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. *J Hydrol* 332(1):144–157. <https://doi.org/10.1016/j.jhydrol.2006.06.028>
18. Hopkinson RF et al (2011) Impact of aligning climatological day on gridding daily maximum-minimum temperature and precipitation over Canada. *J Appl Meteorol Climatol* 50(8):1654–1665. <https://doi.org/10.1175/2011jamc2684.1>
19. Bairwa AK, Khosa R, Maheswaran R (2016) Developing intensity duration frequency curves based on scaling theory using linear probability weighted moments: A case study from India. *J Hydrol* 542:850–859

Assessing Drainage System Impacts Due to Urban Intensification in Rurally-Serviced Residential Areas



M. Senior and R. Scheckenberger

1 Introduction

Historically, urban development in Southern Ontario emanates from existing core developed areas. This involves the conversion of rural/un-developed areas (typically referred to as “greenfields”) into urban development areas (generally residential, with some proportion of commercial, institutional, or other usages). Larger block greenfield development is preceded by numerous supporting planning studies to ensure that the development is properly planned, including secondary plans, subwatershed studies and master servicing studies. These studies ensure that engineering and environmental opportunities and constraints are identified, providing guidance to developers to ensure that expectations and design requirements are clearly defined. Stormwater management (generally defined as the broad suite of measures to collect, convey, treat and control stormwater runoff) is a key consideration in this regard. As part of the supporting studies, stormwater management requirements are established, including flow rate targets, sensitivity of watercourse receivers (including flooding, erosion, and water quality, among other considerations), capacity constraints, water balance/budget requirements, and numerous other related considerations. For well-planned greenfield developments, stormwater management is effectively integrated into the planning process, such that the ultimate development is properly designed and considers the requirements of the overall drainage system and environment.

With recent changes in the Ontario Provincial planning direction (Provincial Policy Statement, Growth Plan etcetera) over the past decade, there is an increasing focus away from greenfield development towards infill and intensification within existing developed limits of the community. The objective is largely to manage urban sprawl and optimize the use of existing infrastructure. This trend is attributable

M. Senior (✉) · R. Scheckenberger
Wood Environment and Infrastructure Solutions, Burlington, Canada
e-mail: matt.senior@woodplc.com

to numerous factors, including limited remaining greenfield development area (i.e. where a municipality has effectively maximized the development of such areas and reached its urban or municipal boundary, or is bounded by another regulatory limit such as the Greenbelt in Ontario), or potentially by developer and consumer interest in existing “desirable” neighbourhoods. This may include re-development in the urban core (“downtown”) area, which may result in a negligible change in impervious coverage, and associated impacts to stormwater management (SWM). In some instances however, infill and intensification may involve the re-development of lower density urban areas, including older detached residential areas in “desirable” neighbourhoods. In such older neighbourhoods, existing densities are typically lower, given historically smaller homes and larger lot sizes. As such, when these areas are intensified to reflect current consumer interests, the corresponding increase in impervious area, and associated stormwater and drainage system impacts, can be substantial.

Unlike large-scale greenfield developments, where the development limits are well defined, and the potential stormwater management impacts inherently assessed as part of the supporting planning process, infill and intensification is typically much more irregular, and correspondingly more challenging for municipalities to manage. Infill and intensification may occur opportunistically as properties become available, and land developers may pursue larger land assemblies for re-development. Where smaller scale infill and intensification is proposed (such as the re-development of an existing detached residence to a larger structure) many municipalities may not mandate stormwater management controls, given the perceived difficulties in incorporating such controls on a single private site, and the perceived minimal impact of development from a single property. Notwithstanding, the potential cumulative impact of such individual developments can be substantial with respect to runoff impacts, given the additional number of these properties with no SWM controls.

This paper reviews the work undertaken by the City of Hamilton in assessing the potential drainage system impacts of infills and intensification to a specific rurally-serviced existing residential area in the Community of Ancaster. The project was undertaken to not only assess the magnitude of the expected impacts both to local drainage system and downstream receivers, but also to review potential mitigation measures to address these expected impacts. The results of the analyses are discussed, along with considerations with respect to the ultimately identified preferred management strategy.

2 Background

2.1 *Intensification in the City of Hamilton*

The City of Hamilton, like other municipalities in Southern Ontario, has experienced an increasing trend of infills and intensification in recent years. The City has considered the impacts of intensification as part of its overall Growth Related Integrated Development Strategy (GRIDS), an integrated planning process which identified land use structure, infrastructure requirements, and economic development considerations for various growth options leading to a preferred alternative. The original GRIDS program was commenced in 2006 and considered growth to 2031. The current program update (GRIDS 2) considers growth to 2041. The planning guidance from GRIDS2 is currently being integrated into the City of Hamilton's Water, Wastewater and Stormwater Master Plan Update Study, which is assessing infrastructure requirements at a higher Master Plan level.

In general, some of the current intensification in the City of Hamilton has occurred in, and around, the downtown core, however other areas of the City have experienced intensification pressures, including areas along the Lake Ontario shoreline in the Stoney Creek community. More recently, the Community of Ancaster has been experiencing an increased level of infills and intensification. The community is generally considered to be a "desirable" neighbourhood with high value detached residential properties prevalent. Older properties in these areas, which tend to have relatively larger lot areas and smaller existing residences, have become targets for re-development/intensification. This is particularly the case in areas zoned "Existing Residential" (ER), which are detached residential areas with rural servicing (i.e. roadside ditching). Typically in these areas, the existing residence is demolished and replaced by a newer residence which is constructed to the maximum lot coverage permissible by municipal policies, termed "as-of-right" development (35% for the subject area). Notably however, this accounts only for the building itself; other impervious areas such as driveways, walkways, patios and other hardscaping tends to significantly increase the resulting final impervious coverage, as compared to existing conditions. Uncontrolled, this increased amount of hard surface can result in negative impacts to receiving drainage systems, both localized and further downstream.

Wood Environment and Infrastructure Solutions (Wood; then operating as Amec Environment and Infrastructure) prepared the "Pilot Study Assessment of Increase in Lot Coverage in Rurally Serviced Roadway Neighbourhoods, Community of Ancaster" [3], which involved an assessment of a Pilot area within the Community of Ancaster, with the objective to analyze and assess the potential for impacts on flooding, and to a lesser extent erosion and water quality. Based upon the outcomes of the Pilot Study Area and the analytical modelling specifically, significant potential increases in both peak flows and runoff volumes would be anticipated, depending on the extent of coverage, location within the development area and intensity of the storm. In terms of mitigation, the Pilot Study (termed Phase 1) examined a number of alternatives, including source controls through Low Impact Development

Best Management Practices (LID BMPs) which could be implemented on the individual lots proposing to redevelop or sever. Notwithstanding, other more holistic neighbourhood-based alternatives were also cited, which could be considered at a broader study scale (i.e. upsizing conveyance infrastructure with neighbourhood scale stormwater management).

Subsequently, Wood was retained by the City to undertake a follow-up (Phase 2) study which was intended as an extension of the Phase 1 study area limits to include all of the Existing Residential (ER) neighbourhoods in the Community of Ancaster with rural drainage servicing, and specifically assess the Level of Service (LOS) associated with these drainage systems. This study [9] is the primary basis for the current paper.

2.2 Study Area Characteristics

The general study area is defined as those areas within the Community of Ancaster zoned as “Existing Residential” (ER); detached residential areas with rurally-serviced roadways (ditching). Thirteen (13) separate rurally serviced areas (A to L) were delineated accordingly, totalling an area of 326.30 ha (excluding the Phase 1 Pilot “C+” Area). Study area limits are presented in Fig. 1.

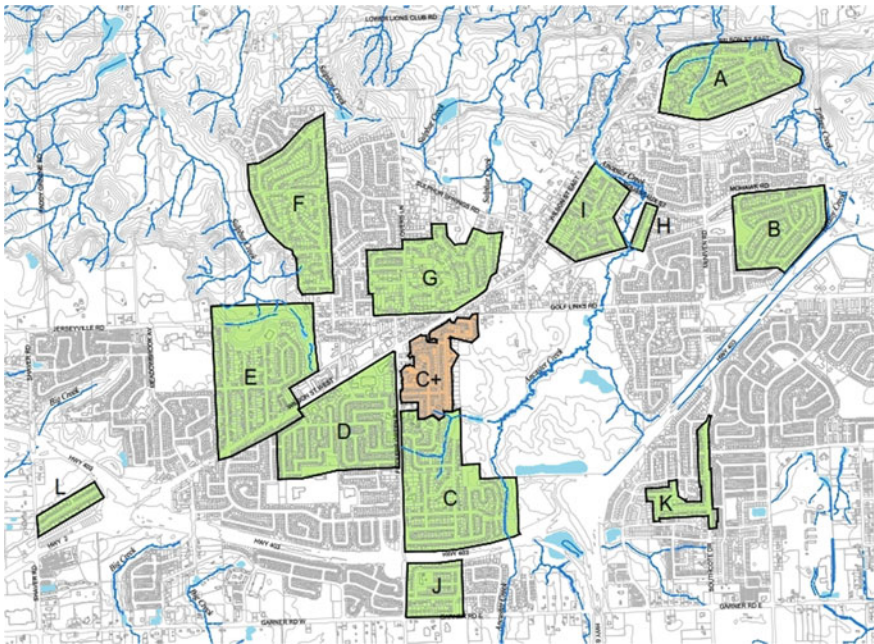


Fig. 1 General study area limits and drainage networks

Despite the predominance of rural servicing (i.e. ditching), the subject areas also include some “hybrid” servicing areas, namely those with rural servicing but some storm sewer collection systems. These features are generally found in isolated locations, often in areas where there is not a suitable outlet for the stormwater runoff or where standing water would likely occur without the storm sewer.

With respect to the study areas presented in Fig. 1, the majority of the study area drains northerly, part of the Ancaster Creek subwatershed, which is part of the overall watershed jurisdiction of the Hamilton Conservation Authority (HCA). Outlets for the study area are approximately evenly split between discharge to urbanized systems (i.e. a storm sewer) and direct outlets to open channel or watercourse receivers. Ultimately, storm sewer receivers would also eventually discharge to watercourses within the study area.

In general, many of the receiving watercourses are characterized as ravine systems, portions of which are located along the Niagara Escarpment, and are largely contained by the Dundas Valley Conservation Area. As such, consideration of downstream impacts is a key consideration, in addition to potential localized conveyance capacity issues.

3 Methodology

3.1 Base Hydrologic Modelling

In order to undertake the required assessment, a robust modelling platform is required, which is capable of simulating both hydrology (simulated flow responses to both discrete and continuous rainfall inputs) and hydraulics [minor (storm sewers and culverts) and major (ditches, channels and overflows) systems as well as storage reservoirs and impacts of tailwater from downstream receivers]. Based on the preceding requirements, PCSWMM was selected, which is a graphical user interface (GUI) and decision support system for the US EPA Stormwater Management Model (SWMM).

A processed 2 m horizontal resolution Digital Elevation Model (DEM) was applied to generate subcatchment boundaries using PCSWMM’s watershed delineation tool. The preliminary drainage boundaries were subsequently reviewed and refined based on other sources of data, including field reconnaissance where necessary. This approach resulted in a highly discrete subcatchment resolution with an average area of 0.43 ha, and a total of 764 subcatchments.

Consistent with the previous Phase 1 Pilot Study [3], the United States Soil Conservation Service (US SCS) Curve Number (CN) approach was applied for the simulation of infiltration for pervious areas. Representative CN values were applied based on surficial soils mapping (Ontario Base Soils Mapping, Soil Survey Report 32—Soils of Hamilton-Wentworth). This base mapping information was further validated using available local geotechnical reports. Overall, approximately 60% of the study

area consists of soils with an SCS hydrological classification of “AB” which represents well-drained sandy type soils. The balance reflects SCS Type “C” soils, which are more poorly drained clay type soils (which are more common within the broader City of Hamilton). Representative SCS CN values were determined based on the hydrological soil group classification, as well as pervious area type (grass/lawn or forested area).

Base impervious coverage was determined using on the City of Hamilton’s available zoning mapping, with representative values applied for different classifications. For the specific “Existing Residential” (ER) zoning designation of interest, imperviousness was more directly estimated based on the division of this area into rooftop/building areas (based on GIS data supplied by the City of Hamilton), the roadway right-of-way and the remaining greenspace/amenity areas (imperviousness values for both were generalized and estimated using sample test areas from available aerial photography).

3.2 Base Hydraulic Modelling

Given the relatively refined scale of the base hydrologic modelling, corresponding refined hydraulic modelling elements are also required as part of the integrated modelling. A key consideration for area hydraulics relates to the local rurally-serviced roadway network, and the associated ditching system. A field reconnaissance was completed for area roadways to categorize the existing roadside ditches into five (5) categories based on depths and extents, ranging from Type ‘A’ (Poorly Defined) to Type ‘E’ (Large Ditch). Typical sections for each ditch type were developed accordingly, in part complemented by a scoped topographic survey. Ditch invert elevations were then estimated based on either topographic survey data (where available) or DEM data (where not). Ditch profiles were then reviewed manually and adjusted as necessary to ensure reasonable and representative results. Each roadside ditch was modelled separately, given potential differences in ditch type and drainage direction. As such, additional spill conduits were incorporated within the modelling to link each section via the centreline of road elevation.

Given the number of driveway culverts in the study area, it was determined that it was not feasible to explicitly include these features within the modelling (estimated to exceed $3000\pm$). Notwithstanding, it is noted that the impact of storage behind blocked driveway culverts could impact the modelling results. Ultimately, it was decided not to incorporate any corrections or adjustments for such storage, in part due to the potential impacts with respect to simulated flow attenuation. Municipal (City-owned) culverts (typically at roadway crossings) were however all included within the modelling. A culvert inventory was undertaken to survey and confirm these features (155 culverts in total); these data were used as part of the model building effort. Although many such culverts were noted to be crushed, damaged or sedimented (partially or totally), for the purpose of the base model full opening area was assumed, to avoid attenuating flows. Local storm sewer systems were also

incorporated into the modelling where present, based on available GIS inventory data and record drawings, with site reconnaissance completed, as required, to confirm or validate modelling assumptions.

3.3 *Model Calibration (Validation)*

In order to further validate model output, a model calibration effort was undertaken. A localized flow monitoring program was conducted at three (3) locations within the study area. Contributing drainage areas to the monitoring sites ranged between 15 and 33 ha. Data were collected for five (5) months (\pm), between June and November of 2018. Although there were a total of seventeen (17) events with a rainfall depth greater than 10 mm during that period [and three (3) events greater than 30 mm], observed flow responses were generally muted. The most frequent flow response was observed at one (1) gauge which contained a “hybrid” drainage system, or a component of storm sewer servicing. Zero or minimal runoff was noted for the majority of the monitoring events at other locations, both in areas with more permeable soils, as well as those with generally less permeable soils.

Notwithstanding the preceding, the available monitoring data were applied to complete a model calibration/validation. Based on the resulted from a hydrologic sensitivity analysis, the most sensitive parameters were identified as subarea routing, subcatchment slope, subcatchment flow length, and pervious area depression storage. Although imperviousness was identified as a sensitive parameter, given that impervious values were directly measured (and the need to calculate expected adjustments under future conditions), this value was not considered for the model calibration effort. Overall, the “subarea routing” parameter was identified as a key calibration parameter. This parameter defines the percentage of the impervious subcatchment segment that is routed across the pervious segment (providing flow attenuation and an opportunity for infiltration). Given the rurally-serviced nature of the study area, and the highly permeable soils in much of the study area, this parameter was therefore considered appropriate for calibration in this study area. Ultimately, the preferred model calibration involved increasing the subarea routing from an initially assumed value of 40%, to a calibrated value of 90%, indicating that the majority of the impervious land segment would be routed across pervious area, as would be the case in a rurally-serviced area (i.e. with grassed ditches). Other calibration adjustments were applied to slope (reduced by 40%), flow length (increased by 20%) and pervious area depression storage (increased from 5 to 10 mm) to reflect the pervious nature of the study area. Relevant calibration statistics from the typically applied scatter plot comparison of simulated versus observed parameters of interest are presented in Table 1.

As evident from the results presented in Table 1, the calibration effort notably reduces the over-estimation of peak flow and runoff volumes, although variability of the data (scatter—as evident from the R^2 results) remains high. By undertaking event screening, the results further improved, although based on a condensed dataset, the

Table 1 Model calibration summary (discrete event simulation)

| Calibration parameter | Modelling scenario | Linear trend line slope | Coefficient of determination (R^2) |
|-----------------------|---------------------------------|-------------------------|--|
| Peak flow | Uncalibrated | 1.92 | -0.93 |
| | Calibrated | 1.09 | -0.38 |
| | Calibrated with event screening | 1.17 | 0.64 |
| Runoff volume | Uncalibrated | 5.58 | -0.46 |
| | Calibrated | 3.75 | 0.56 |
| | Calibrated with event screening | 1.23 | 0.53 |

model calibration effort did validate the model and generate more realistic estimates of hydrologic outputs.

3.4 Simulation Scenarios

Consistent with the approach of the Phase 1 Pilot Study [3], hydrologic simulations have adopted the design storm event approach, with a particular focus on the SCS 24-h Type-II design storm distribution for the 1 in 2 year storm event (53 mm in 24 h), 1 in 5 year storm event (72 mm in 24 h) and the 1 in 100 year storm event (123 mm in 24 h). The typically applied water quality storm event in Ontario (25 mm in 4 h) was also simulated.

In addition to the preceding, potential increases in rainfall due to Climate Change were considered. Climate change influenced rainfall intensity–duration–frequency (IDF) data are readily available through public sources. One example is the University of Western Ontario’s IDFCC Tool, available at <http://www.idf-cc-uwo.ca> [8]. Climate change-altered rainfall has been extracted from this tool for 2080, using the RCP 4.5 emission forcing scenario (as per [5]). Adjusted data were also obtained from the Ontario Climate Change Data Portal (OCCDP) IDF rainfall tool, available at <http://ontarioccdp.ca/>, which is run by the Institute for Energy, Environment and Sustainable Communities (IEESC) at the University of Regina. Data have been extracted from the OCCDP for the same scenario and year for the grid cell (25 km × 25 km) encompassing the study area. Lastly, data were obtained through the Ontario Ministry of Transportation (MTO) IDF Curve Lookup tool (http://www.mto.gov.on.ca/IDF_Curves), for which data have been extracted based on the point location representative of the study area. Modified versions of the 5 and 100-year design storm events were developed accordingly.

A continuous simulation modelling exercise was also undertaken as part of the hydrologic simulations, in order to better assess changes to runoff volume and in particular erosion impacts to sensitive downstream ravine systems. Overall, a 55-year

hourly rainfall dataset was assembled from Environment Canada’s Royal Botanical Gardens (RBG) gauge (1962–2016). Monthly average evaporation was also incorporated into the modelling. The findings of this additional simulation are however not discussed in this paper.

4 Results

4.1 Existing Conditions

The simulated performance of the drainage system under existing conditions serves as a baseline for the assessment of future conditions, and the development of associated stormwater management (SWM) mitigation strategies. Local drainage system performance was quantified by two (2) primary metrics: simulated peak flows at the primary drainage network outlets, and by the simulated hydraulic performance of the ditch conveyance system. The latter is summarized in Table 2, on the basis of the simulated conveyance extents [within the ditch, beyond the ditch (but within the roadway right-of-way) or beyond the roadway right-of-way (on to private property)]. Percentages are based on the corresponding total and individual ditch segment lengths. Based on City of Hamilton design standards and associated level-of-service, the 5-year storm event should be conveyed within the ditch, while the 100-year storm event should be conveyed within the ROW (i.e. not flood on to adjacent private properties).

The simulated results indicate the majority of the area ditches (>90%) are able to contain up to the 2-year flow within the ditch. For the 5-year storm event, this amount is reduced slightly, to 82%. For the 100-year storm event, only 7% ± of ditches indicate exceedance of flow beyond roadway ROW (onto private property). Although not presented in Table 2, the performance under existing conditions under climate change (CC) was also assessed, for the 5 and 100-year storm events. Overall, the simulated results indicated that performance would be worsened under both scenarios, but much more so for the 100-year storm event (increase in ROW flooding from 7% to between 9 and 18%, depending on the scenario).

Table 2 Simulated existing ditch performance under existing conditions

| Storm event | Within ditch (%) | Within ROW (%) | Beyond ROW (%) |
|-------------|------------------|----------------|----------------|
| 25 mm | 97.9 | 2.1 | 0.0 |
| 2-year | 90.8 | 8.6 | 0.6 |
| 5-year | 82.0 | 16.3 | 1.7 |
| 100-year | 59.4 | 33.7 | 6.9 |

4.2 *Future Conditions*

As discussed in Sect. 2.1, urban intensification in the City of Hamilton, and in particular within the Community of Ancaster, is generally permitted to the “As of Right” limits, which for the ER Zoning represents building coverage equal to 35% of the lot area. In addition, the increase in imperviousness associated with lot amenity areas (driveways, walkways, patios, etcetera) must be considered. To assess this potentially significant factor, a review of approximately 100 residential properties was undertaken from available aerial photography, to calculate the building footprint area, and the associated amenity impervious area. Based on this analysis, amenity area was found to average 91% of the building footprint area, representing a notable additional impervious area under future conditions.

Under future conditions, it was conservatively assumed that the additional development (building and amenity areas) would occur in areas of existing greenspace (pervious areas). As such, the subcatchments developed under existing conditions were split into two (2) subcatchments: the portion representing the existing impervious areas and remaining pervious areas, and the portion of existing greenspace expected to be developed under “as-of-right” conditions. Additional impervious area was calculated on a network by network basis, based on the relative expected increase in building area (as calculated based on the existing building area relative to available lot area). The increases were then applied to individual subcatchments, along with the corresponding expected increase in amenity area. Based on the preceding approach, overall study area imperviousness increased from 41.6 to 57.2%, or an absolute increase of 15.6%, under a full development scenario (“as-of-right”).

As would be expected, without SWM controls, increases in peak flows and decreases in ditch conveyance performance would be expected. Increases in simulated peak flows were most notable for smaller, more frequent storm events; an average increase in peak flows of 89% was indicated for the 25 mm storm event, reflecting the impact of the loss of infiltration potential for smaller rainfalls. For design storm events, peak flows increases averaged between 34% for the 2-year storm event, to 15% for the 100-year storm event, reflecting the reduced infiltration capacity for pervious areas during more formative storm events. For downstream receivers along watercourses, peak flows for the 5 and 100-year storm events increased by an average of 12% for the 5-year storm event, and 6% for the 100-year storm event, reflecting some of the effects associated with flow routing and hydrograph timing.

Based on the preceding results, a stormwater management (SWM) strategy was clearly warranted to mitigate the primary impacts related to runoff quantity, including worsened conveyance performance (i.e. roadside ditches and culverts, including spills beyond the right-of-way onto private property), and potential downstream (off-site) flooding impacts. Other related impacts (although not discussed as part of the current paper) also required consideration, including erosion, water budget, and water quality impacts. A long-list of alternatives was reviewed; the short-listed alternative involved the implementation of source controls, including Low Impact Development Best Management Practices (LID BMPs). Such measures involve the capture, and

typically infiltration and filtration of runoff, at source/on site. This may include both surface (bio-retention areas, enhanced grass swales, permeable pavement, rainwater harvesting, green roofs) and sub-surface (soakaway pits, exfiltration pipes) measures. Such measures are considered advantageous, since if appropriately sized, they would be expected to manage impacts associated with both runoff peaks and runoff volumes, with secondary benefits related to water balance, erosion mitigation, and potentially water quality, depending on how the controls are designed.

A key consideration relates to where such source controls are located: within the public realm (within the roadway right-of-way) or within the private realm (incorporated into the development site). While placing such controls within the public ROW ensures access for the Municipality (to facilitate Operations and Maintenance), it can be problematic for a number of different reasons. With respect to implementation timing, it would require the Municipality to construct the works in advance of the local development, which may be problematic in the case of multiple distributed developments (i.e. lot intensification). Alternatively, if the private developer constructs the works as part of the development, this necessitates municipal review and oversight, and utilizes available municipal property to provide controls for private development, which is generally contrary to standard development practices. By contrast, requiring the controls to be placed on the private side shifts the responsibility to the developer, while also allowing the developer greater flexibility to design the source controls to suit the specific considerations of the proposed site, and to ensure the works are scheduled and constructed as part of the development. Such an approach must however adequately ensure that the municipality is granted access through some type of legal instrument or arrangement to inspect the works in the future and ensure that the developer or ultimate owner is responsible for maintaining the works in perpetuity. Lastly, this approach preserves land and space availability within the public realm (ROW) to implement future SWM controls (including source controls and LID BMPs) to address existing drainage system deficiencies, unrelated to future intensification.

The modelling was applied to iteratively determine the required capture rate of future source controls. Given the variability in surficial soils and associated infiltration (and runoff) potential, different capture rates were assessed for different drainage network areas (as per Fig. 1). Based on this approach, source control capture rates were determined to be between 55 and 70 mm per impervious hectare (or 550–700 m³ per impervious hectare). While these capture rates are notably higher than typical industry values for source controls and LID BMP measures (which typically range from 5 to 30 mm), it should be understood that the source controls advocated for this situation are intended to provide quantity/flood control up to and including the 100-year storm event; thus an inherently higher capture depth is required. A comparison of source control volumes from the current study to those for more typical end of pipe stormwater management ponds for greenfield developments (i.e. from subwatershed or secondary planning studies) is presented in Table 3. The results generally compare favourably, particularly when considering the generally more permeable soils in the study area, and the lack of explicit extended detention erosion storage.

Table 3 Comparable quantity control storage volumes for greenfield SWM facilities

| Municipality | Study area | Extended detention and erosion control (m ³ /imp ha) | 100-year flood control including ED (m ³ /imp ha) |
|------------------|-----------------------|---|--|
| Town of Milton | Bristol survey [6] | 430 | 795 |
| | Sherwood survey [7] | 375 | 855 |
| | Boyne survey [2] | 375–400 | 795–975 |
| City of Brampton | Mount pleasant [1] | 200–325 | 900–1500 |
| City of Markham | Future urban area [4] | 300–500 | 800–1100 |

The precise form of the source controls to be applied would vary by site and would need to be determined by the designer in consultation with the City. Given the high degree of estimated amenity area through the current study, it is expected that sub-surface measures may be preferable, such as soakaway pits for roof drains. Permeable pavement (concrete, asphalt or paving stones) could also be considered for driveways and walkways, particularly given the additional contaminant loading associated with impervious roadways and driveways. Given the general stormwater management quantity control principle of post-development to pre-development peak flow control, the preceding criteria would also only be applicable to the additional impervious area, not the total impervious area, which would clearly reduce storage volume requirements as compared to development for a greenfield site.

The simulation modelling indicates that with the proposed source controls in place, peak flows to network outlets would be controlled to pre-development values ($\pm 2\%$). Peak flows within downstream receivers (nodes along receiving watercourses) would similarly be within $\pm 1\%$ of existing condition values. Ditch conveyance performance would similarly be within $\pm 1\%$ of the simulated existing performance. Overall, comparison metrics confirm that the proposed source control strategy would be effective in maintaining existing drainage system performance values. Although not discussed in detail as part of this paper, the continuous simulation modelling assessment also confirmed that metrics related to water budget and erosion threshold analyses would be generally satisfied by the proposed approach.

As noted in Sect. 3.4, climate-change altered rainfall was generated for the study area and used to assess the potential changes in drainage patterns both under existing conditions, as well as under the expected future development scenario (“as-of-right”). Of particular interest was the additional source control volume requirements to mitigate the potential additional impacts associated with climate change. In the absence of a formal City of Hamilton policy regarding climate change, the previously applied three (3) climate change altered rainfall IDF datasets have again been employed. Of the three (3) scenarios presented, the University of Western Ontario (UWO) climate change altered 100-year design storm generated the highest flows and greatest degree of storage exceedance. The UWO 100-year design storm event reflects an approximate 60 mm increase in total rainfall depth, and a 48% increase in peak intensity, as compared to base (non-climate change adjusted) IDF data. This storm event

is the most formative of the three (3) climate change scenarios and was therefore applied to estimate additional source control requirements. In order to estimate associated storage requirements, the same iterative approach as described previously was employed, however climate change-altered rainfall was applied only to the areas of change; all other areas applied the existing IDF rainfall scenarios in order to minimize potential differences. Based on this approach, an additional 30–45 mm of capture depth (300–450 m³) per impervious hectare would be required to fully mitigate the direct impacts associated with climate change. As noted this would only apply to the additional areas of change; separate mitigation measures would be required for existing drainage areas not influenced by intensification. Notwithstanding, this represents a significant additional storage requirement, and confirms the potential for significant impacts associated with climate change.

5 Summary and Future Conditions

The results presented in this paper outline the magnitude of the potential drainage system impacts associated with urban intensification. While the impact of the re-development of a single residential property may be relatively minor, the cumulative impact on a neighbourhood scale is substantial, affecting not only local conveyance systems and roadway flooding, but also downstream receiving water-courses, including flooding and erosion impacts. Although not directly assessed as part of the study, water quality impacts are also a potential concern.

Based upon a review of mitigation alternatives, the preferred stormwater management strategy was determined to be private-side source controls, including Low Impact Development Best Management Practices. This approach enforces the general “polluter pay” principle, while providing the developer with options and flexibility in how best to implement such measures on site, through site design. Given the value of surface features and the significant amount of amenity area associated with re-development, it is generally expected that sub-surface measures such as soakaway pits will be preferred, however this will ultimately be at the discretion of the developer. Hydrologic simulation results indicate that source control volumes of between 55 and 70 mm (550–700 m³) per new/additional impervious hectare would be required to ensure that peak flows both to immediate receivers, as well as downstream nodes are generally maintained to existing condition levels. While the preceding target volumes appear substantial compared to typical LID BMP targets (i.e. 5–30 mm), the target volumes are comparable to those for “greenfield” developments in other areas of Southern Ontario, and reflects the proposed source control function with respect to quantity and flood control for storms up to, and including, the 100-year event. Based on a preliminary assessment, if climate change impacts are considered, an additional 30–45 mm (300–450 m³) of storage per additional impervious hectare would be required.

It is expected that a number of future studies and works will result from the current study. A particular focus will likely be upon providing sufficient guidance to developers to ensure that the developed criteria are easily implemented. The modelling developed as part of this study may also be applied in this regard to validate/verify the function of actual development SWM strategies accordingly. The City of Hamilton is also advancing a defined municipal strategy to comprehensively address the potential impacts associated with climate change. The preliminary work completed as part of this study may therefore warrant revision in light of the pending direction from this work. Lastly, the focus of the current study (and paper) have been upon ensuring that future drainage system impacts, associated with future urban intensification are appropriately controlled to existing condition values and level of service. Notwithstanding, a number of drainage system deficiencies have been identified under existing conditions. It is expected that these issues will be addressed opportunistically by the City of Hamilton, in particular as part of planned future capital works, in particular roadway reconstructions.

Acknowledgements The authors of this paper gratefully acknowledge the participation and co-operation of the City of Hamilton for its contributions to this paper and the project as a whole. Supporting information was provided by Cole Engineering Limited (flow monitoring program data) and J. D. Barnes Limited (culvert inventory and topographic survey).

References

1. Amec Earth and Environmental (2011) Huttonville and Fletcher's Creeks subwatershed study—City of Brampton. Amec Earth and Environmental, Burlington
2. Amec Environment and Infrastructure (2015) Sixteen mile creek areas 2&7 subwatershed update study—technical appendix: functional stormwater and environmental management strategy, Boyne survey secondary plan area—Town of Milton. Amec Environment and Infrastructure, Burlington
3. Amec Environment and Infrastructure (2016) Pilot study assessment of increase in lot coverage in rurally serviced roadway neighbourhoods, community of Ancaster—City of Hamilton. Amec Environment and Infrastructure, Burlington
4. Amec Foster Wheeler Environment and Infrastructure (2019) North Markham future urban area: Berczy, Bruce and Robinson Creeks. Amec Foster Wheeler Environment and Infrastructure, Burlington
5. Intergovernmental Panel on Climate Change (2014) Climate change 2014: synthesis report. In: Pachauri RK, Meyer LA (eds) Contribution of working groups I, II and III to the fifth assessment report of the intergovernmental panel on climate change. IPCC, Geneva
6. Philips Planning and Engineering Limited (2000) Sixteen mile creek subwatershed planning study areas 2 and 7. Philips Planning and Engineering Limited, Burlington
7. Philips Engineering Limited (2004) Indian creek/sixteen mile creek sherwood survey subwatershed management study. Philips Engineering Limited, Burlington
8. Simonovic SP, Schardong A, Sandink D, Srivastav R (2016) A web-based tool for the development of intensity duration frequency curves under changing climate. *Environ Modell Softw J* 81:136–153

9. Wood Environment and Infrastructure Solutions (2019) Summary report (draft final)—detailed drainage assessment study (phase 2) of rurally serviced existing residential neighbourhoods, community of Ancaster, City of Hamilton. Wood Environment and Infrastructure Solutions, Burlington

Review and Updated Guidance for Embankment Overtopping Dam Breach Dimensions



Mayari Bernard-Garcia and Tew-Fik Mahdi

1 Introduction

Restraint by the limits of knowledge in various fields (e.g. sediment transportation, etc.) and by the actual computational capacities (e.g. computer time calculation, etc.), most of the current worldwide standard (legislative or normative) frameworks of hydraulic dams' safety have been formulated using statistical, empirical or semi-empirical approaches. The "accuracy" of these standard specifications thus remains, as a result, intimately linked to the amount of "detailed" historical dam failure case studies (i.e. recording failure parameters, e.g. final breach dimensions: breach depth, top width, bottom width, etc., breach formation time, etc.). Since these guidelines provide dam breach parameters (e.g. final breach dimensions: breach depth, top width, bottom width, etc., breach formation time, etc.) based on dams' type (embankment dams, concrete dams, ...) and the failure mode (e.g. overtopping failures, seepage failures, ...). Therefore, to provide "useful" and "efficient" results for the dam industry (i.e. dam owners, standard frameworks, etc.), the authors have strived to obtain new knowledge for dam hydraulic safety frameworks.

In this paper, an overview of overtopping dam failure parameters guidelines, currently available and commonly used in practice, is first presented. An updated portrait formulated for embankment overtopping dam breach dimensions is thus

M. Bernard-Garcia: Presenting author.

M. Bernard-Garcia · T.-F. Mahdi (✉)
Département des génies Civil, Géologique et des Mines (CGM), Polytechnique Montreal, 2900
Boul. Édouard-Montpetit, University of Montreal Campus, 2500 Chemin de Polytechnique,
Montreal, QC 3T 1J4, Canada
e-mail: tewfik.mahdi@polymtl.ca

M. Bernard-Garcia
e-mail: mayari.bernard-garcia@polymtl.ca

provided using the largest and latest known database currently available in the literature [2].

2 Overview of Overtopping Dam Failure Parameters Guidance

In this section, five (5) overtopping dam failure guidance specifications, adapted from Marche [4, Appendix 1] and the Hydrologic Engineering Center [3, Table 3], are summarized in Table 1.

Overall, Table 1 presents overtopping (OT) dam failure parameters (i.e. breach dimensions and time formation) for embankment dams (EFL + RFL) and concrete dams (CON) breach studies, also as (for most guidance) distinguish guidance specifications for embankment faced dams (EFL + RFL; FD) and concrete arc dams (CON; ARC). The **breach dimensions**, identified in Table 1, are illustrated in Fig. 1.

Finally, as highlighted in bold in Table 1 and presented in the next section, notice that **embankment overtopping dam breach dimensions ratios** (W_{ave}/H_D and W_{BB}/H_D) **guidelines** stand between 1 and 5.

3 Review and Updated Embankment Dam Breach Dimensions Guidance

In this paper, the overtopping embankment dam breach dimensions are studied through a breach dimensions ratio (i.e. the ratio W_i/H_D of a breach width (W_{ave} , W_{BB} and/or W_{TB}) on the dam height (H_D) or breach depth (H_B)), using the latest and largest compilation currently known and available in the literature [2]. To provide a scalable portrait, results adapted from Wahl [5] database are also provided in this paper for comparison. Therefore, a brief portrait of the samples considered in this study is presented in this section.

First, among the 355 historical embankment dam failures (EFL + RFL) recording at least one of the three final breach widths (i.e. top, bottom and/or average breach widths), compiled among the 2769 worldwide Man-Made dam failures (MM-D) recorded in Bernard-Garcia and Mahdi [2] database, only the case studies failed by overtopping (OT), excluding the embankment faced dams (no FD), have been considered and were all subject to a double verification. In the same vein, among the 108 historical embankment dam failures recorded in Wahl [5] database, the same selection criteria (i.e. OT, no FD) have been considered in the case studies selection in order to ensure consistency in the results presented. Overall, considering the “Best reliable” information recorded in each database (see columns “Cote” and “Remarks” in Bernard-Garcia and Mahdi [2] database and “Best Reliable Information” in Wahl

Table 1 Overtopping dam failure standards adapted from Marche [4, Annex 1] and the Hydrologic Engineering Center [3, Table 3]

| Dam type | Overtopping failure parameters | HQ (1) | MEQ-CEHQ (1) | NWS (2) | FERC (2) | Switzerland (1) |
|--------------------------------------|---|------------|--------------|----------------------|----------------------|-----------------|
| Concrete Gravity dam (CON; GR) | Bottom breach width; W_{BB} | 4*H | 4*H | Usually $\leq 0.5*L$ | Usually $\leq 0.5*L$ | Entire dam |
| | Breach average width; W_{ave} | 4*H | 4*H | | | Entire dam |
| | Formation Time; t_f (hrs) | 0.1 | 0.1 | 0.1–0.2 | 0.1–0.3 | 0 |
| Concrete Arch dam (CON; ARC) | Bottom breach width; W_{BB} | 4*H | | 0.8*L to L | Entire dam | Entire dam |
| | Breach average width; W_{ave} | 4*H | | | | Entire dam |
| | Formation Time; t_f (hrs) | 0.1 | | ≤ 0.1 | ≤ 0.1 | 0 |
| Embankment dam (EFL + RFL) | Bottom breach width; W_{BB} | 4*H | 3*H | 2*H–5*H | 1*H–5*H | 2*H |
| | Breach average width; W_{ave} | 5*H | 3*H | | | 3*H |
| | Formation Time; t_f (hrs) | 0.5 | 0.5 | 0.1–1 | 0.1–1 | 0 |
| Embankment faced dam (EFL + RFL; FD) | Bottom breach width; W_{BB} | 4*H | 3*H | | | 2*H |
| | Breach average width; W_{ave} | 5*H | 3*H | | | 3*H |
| | Formation Time; t_f (hrs) | 0.5–2 | 0.5 | | | 0 |

Notes

H dam height; L length of the crest dam; HQ hydro-Québec (QC, CAN); $CEHQ$ centre d'expertise hydrique du Québec (QC, CAN); NWS national weather service (USA); $FERC$ federal energy regulatory commission (USA)

(1): Adapted from Marche [4, Appendix 1]

(2): Adapted from Hydrologic Engineering Center [3, Table 3]

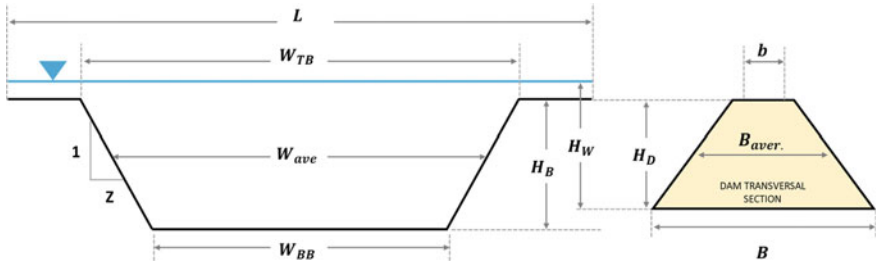


Fig. 1 Idealized trapezoid dam breach geometry [2]

[5] database) has led the authors to a sample of 30 “verified” and “detailed” historical embankment overtopping dam failures from Wahl [5] database and to a sample of 177 “verified” and “detailed” historical embankment overtopping dam failures from Bernard-Garcia and Mahdi [2] database.

Highlight that, in this study, when two (or three) widths were available for one case study, the order in which the breach width (W_i) has been considered to calculate the numerical value of the ratio W_i/H_D or W_i/H_B is: (1) the breach average width (W_{ave}), (2) the breach bottom width (W_{BB}) and (3) the breach top width (W_{TB}).

4 Results

Since some of the case studies included in the samples considered do not record the dam breach height and others do not record the dam height, Fig. 2 first illustrates the distribution of these cases in Wahl [5] and Bernard-Garcia and Mahdi [2] databases.

As illustrated in Fig. 2:

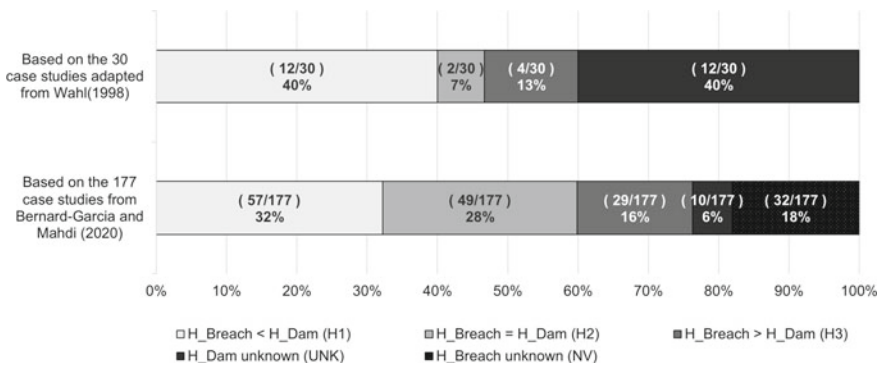


Fig. 2 Distribution of the number of case studies recording at least one breach dimensions recorded in Bernard-Garcia and Mahdi [2] and Wahl [5] databases

- *The breach depth (H_B) or the dam height (H_D) is not recorded (UNK + NV):*
 - For 12 case studies (OT, EFL + RFL) of the 30 historical dam failures adapted from Wahl [5] database, thus **40%** of the compilation.
 - For 42 case studies (OT, EFL + RFL) of the 177 historical dam failures adapted from Bernard-Garcia and Mahdi [2] database, thus **24%** of the sample.
- *The breach depth (H_B) is lower (H1) or equal (H2) to the dam height (H_D):*
 - For a total of 14 case studies (OT, EFL + RFL) adapted from Wahl [5] database, thus **78%** of the 18 case studies recording both dimensions.
 - For a total of 106 case studies (OT, EFL + RFL) adapted from Bernard-Garcia and Mahdi [2] database, thus **79%** of the 135 case studies recording both dimensions.

Figure 3 illustrates the comparisons between the dam height (H_D) and the breach height (H_B) for the 135 “verified” case studies adapted from Bernard-Garcia and Mahdi [2] database, also as the 18 embankment dams failed by overtopping adapted from Wahl [5] recording both variables.

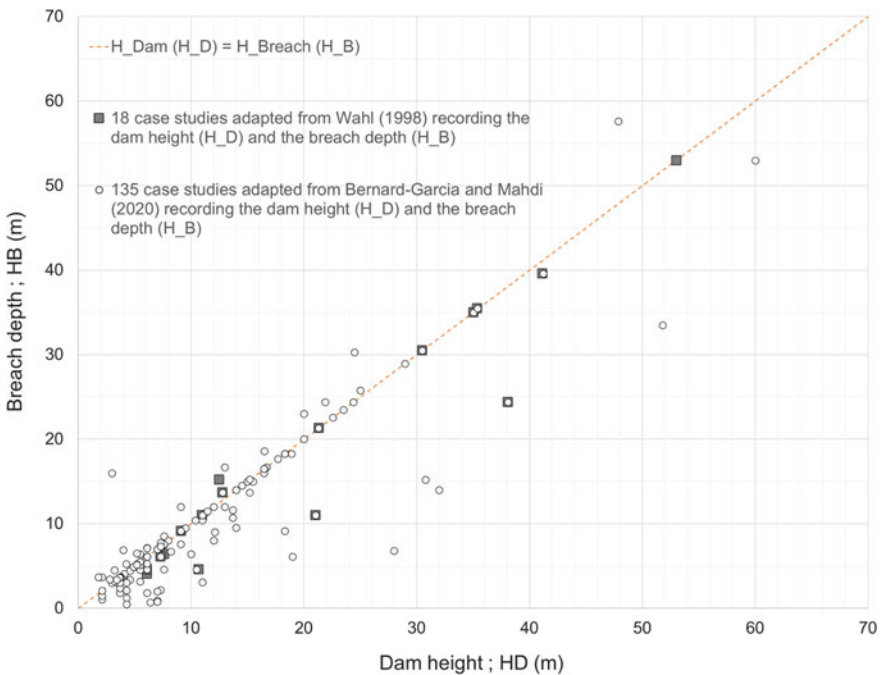


Fig. 3 Observed breach depth (H_B) and dam height (H_D) recorded for embankment dams (EFL + RFL) failed by overtopping (OT) recording both dimensions in Bernard-Garcia and Mahdi [2] database and Wahl [5] database

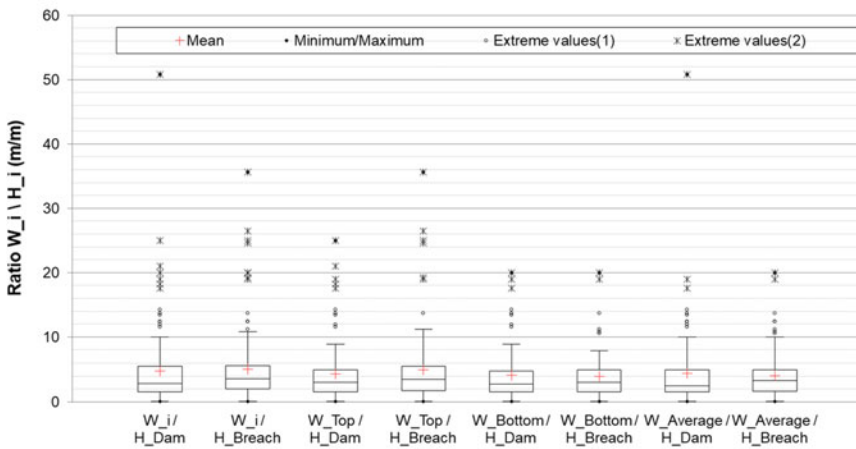
As highlighted in Fig. 3, most of the cases adapted from Wahl [5] database compose Bernard-Garcia and Mahdi [2] database, since Bernard-Garcia and Mahdi [2] database has notably been built on Wahl [5] database. Also, as illustrated in Fig. 3, most of the case studies recorded refer to dam height (H_D) smaller or equal than 20 m.

- For a total of 10 case studies (OT, EFL + RFL) adapted from Wahl [5] database, thus **56%** of the 18 case studies recording both dimensions.
- For a total of 114 case studies (OT, EFL + RFL) adapted from Bernard-Garcia and Mahdi [2] database, thus **84%** of the 135 case studies recording both dimensions.

Figure 4 illustrates the statistical influence of considering the dam height (H_D) instead of the breach height (H_B) for the average breach width (W_{ave}), the bottom breach width (W_{BB}) and/or the top breach width (W_{TB}), using the 135 “verified” case studies adapted from Bernard-Garcia and Mahdi [2] database.

As illustrated in Fig. 4,

- Comparing the number of breach width variables recorded in Bernard-Garcia and Mahdi [2], **the average breach width (W_{ave}) is the “most” recorded width** (i.e.



Descriptive statistics (Quantitative data) :

| Descriptive statistics | W _i /H _{Dam} | W _i /H _{Breach} | W _{Top} /H _{Dam} | W _{Top} /H _{Breach} | W _{Bottom} /H _{Dam} | W _{Bottom} /H _{Breach} | W _{Average} /H _{Dam} | W _{Average} /H _{Breach} |
|--------------------------|----------------------------------|-------------------------------------|------------------------------------|---------------------------------------|---------------------------------------|--|--|---|
| Nbr of data | 135 | 135 | 105 | 105 | 84 | 84 | 111 | 111 |
| Minimum | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| Maximum | 50.8 | 35.6 | 25.0 | 35.6 | 20.0 | 20.0 | 50.8 | 20.0 |
| 1st Quartile | 1.6 | 2.0 | 1.6 | 1.7 | 1.6 | 1.6 | 1.5 | 1.7 |
| Median | 2.9 | 3.6 | 3.0 | 3.5 | 2.7 | 3.0 | 2.5 | 3.3 |
| 3rd Quartile | 5.5 | 5.6 | 5.0 | 5.5 | 4.8 | 5.0 | 5.0 | 5.0 |
| Mean | 4.8 | 5.1 | 4.4 | 4.9 | 4.1 | 3.9 | 4.4 | 4.1 |
| Borne inf. on mean (95%) | 3.7 | 4.1 | 3.5 | 3.8 | 3.2 | 3.1 | 3.3 | 3.4 |
| Borne sup. on mean (95%) | 5.8 | 6.0 | 5.3 | 6.0 | 5.1 | 4.7 | 5.5 | 4.7 |

Fig. 4 Descriptive statistics (box plot graph adapted from Addinsoft [1] toolbox) of the breach ratio W_i/H_B compared to the ratio W_i/H_D observed for embankment dam (EFL + RFL) failed by overtopping (OT) adapted from the 135 cases recorded in Bernard-Garcia and Mahdi [2] database

111 records/135 cases). Notice that this is also true for the sample adapted from Wahl [5] (i.e. 25 records/30 cases).

- Overall, considering the dam height (H_D) instead of the breach depth (H_B) tends to underestimate the ratio values.

Figure 5 illustrates the distributions of the dam height (or, when unknown, the breach height) on the breach widths (W_i) for the 177 case studies adapted from Bernard-Garcia and Mahdi [2] database and the 30 case studies adapted from Wahl [5] database.

As illustrated in Fig. 5:

- Based on Wahl [5] sample, illustrated in Fig. 5a, the overall range/domain of “possible” dam breach dimensions ratio obtained (W_i/H_D) stands between **0.5 and 20**.
- Based on Bernard-Garcia and Mahdi [2] sample, illustrated in Fig. 5b, the overall range/domain of “possible” dam breach dimensions ratio obtained (W_i/H_D) stands between **0.25 and 20**.

Finally, to also provide a statistical description of the breach dimensions ratios (W_i/H_D) adapted from Bernard-Garcia and Mahdi [2] and Wahl [5] databases, Fig. 6 illustrates the “Box Plots” graphs obtained using the XLSTAT Student 2019 [1].

As illustrated in Fig. 6, for the cases where the breach depth (H_B) recorded is higher than the dam height (H_D), i.e. **H3** sample, the range of “ W_i/H_D ” ratios observed is wider and higher than the others (e.g. “H1 + H2” sample), which is true for both databases. Moreover, as highlighted in Fig. 6:

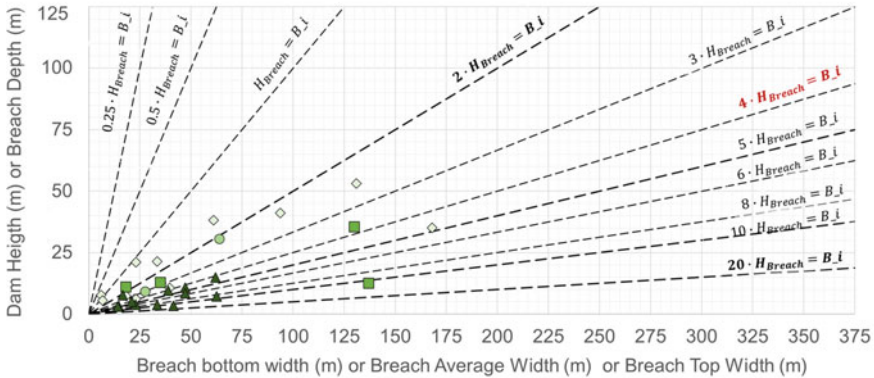
- The *interquartile range* of “possible” ratio “ W_i/H_D ” obtained for “H1 + H2 + H3” sample:
 - Based on Wahl [5] compilation stands between **2 and 4**.
 - Based on Bernard-Garcia and Mahdi [2] database stands between **2 and 6**.
- Based on the *lower and upper boundary (95%) of the mean* “ W_i/H_D ” ratio values obtained for “H1 + H2 + H3” sample:
 - The mean value of this breach dimensions ratio, adapted from Wahl [5], also stands between **2 and 4**.
 - The mean value of the breach dimensions ratio adapted from Bernard-Garcia and Mahdi [2] database stands between **4 and 6**.

5 Discussion

First, in this paper, an *overview of overtopping dam failures parameters guidelines* has been drawn in Table 1 using five (5) overtopping dam failure parameters guidelines adapted from Marche [4, Appendix 1] and the Hydrologic Engineering Center [3, Table 3]. Thus, the overall range of “possible” embankment overtopping breach

a) Adapted from Wahl (1998) database

- ◇ 12 case of embankment dam (EFL+RFL) overtopping failures (OT), where $H_{Breach} < H_{Dam}$ (H1), adapted from Wahl (1998)
- 2 case of embankment dam (EFL+RFL) overtopping failures (OT), where $H_{Breach} = H_{Dam}$ (H2), adapted from Wahl (1998)
- 4 case of embankment dam (EFL+RFL) overtopping failures (OT), where $H_{Breach} > H_{Dam}$ (H3), adapted from Wahl (1998)
- ▲ 12 case of embankment dam (EFL+RFL) overtopping failures (OT), where the dam height is unknown (UNK), adapted from Wahl (1998)



b) Adapted from Bernard-Garcia and Mahdi (2020)

- ◇ 57 case of embankment dam (EFL+RFL) overtopping failures (OT), where $H_{Breach} < H_{Dam}$ (H1), recorded in this database
- 55 case of embankment dam (EFL+RFL) overtopping failures (OT), where $H_{Breach} = H_{Dam}$ (H2), recorded in this database
- 23 case of embankment dam (EFL+RFL) overtopping failures (OT), where $H_{Breach} > H_{Dam}$ (H3), recorded in this database
- ▲ 42 case of embankment dam (EFL+RFL) overtopping failures (OT) (where $H_{Breach} = H_{Dam}^*$) but NOT VERIFIED (NV) recorded in Bernard-Garcia and Mahdi (2020) database

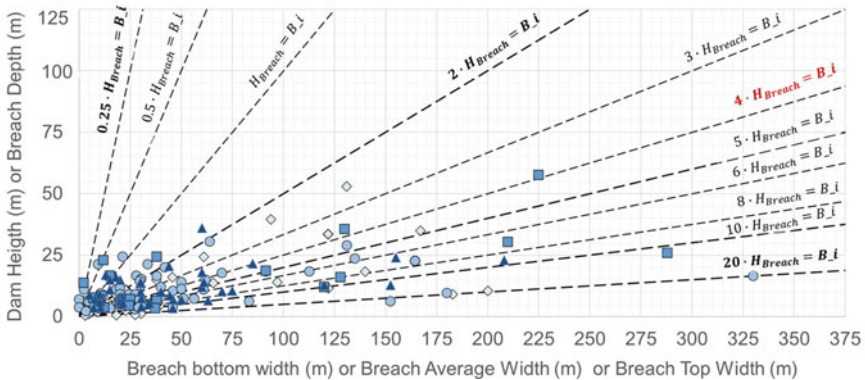
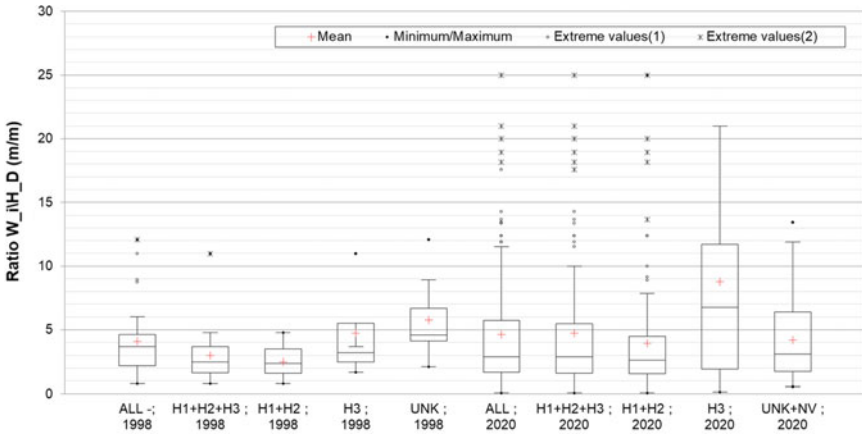


Fig. 5 Distribution of the breach width ($W_i \in [W_{ave}, W_{BB}$ or $W_{TB}]$) and dam height (H_D) observed for embankment dam (EFL + RFL) failed by overtopping (OT) **a** for 30 cases adapted from Wahl [5] **b** for 177 adapted from Bernard-Garcia and Mahdi [2] databases



Descriptive statistics (Quantitative data) :

| Descriptive statistics | adapted from Wahl (1998) database | | | | | adapted from Bernard-Garcia and Mahdi (2020) database | | | | |
|--------------------------|-----------------------------------|-----------------|--------------|-----------|------------|---|-----------------|--------------|-----------|---------------|
| | ALL - 1998 | H1+H2+H3 - 1998 | H1+H2 - 1998 | H3 - 1998 | UNK - 1998 | ALL - 2020 | H1+H2+H3 - 2020 | H1+H2 - 2020 | H3 - 2020 | UNK+Nv - 2020 |
| Nbr of data | 30 | 18 | 14 | 4 | 12 | 177 | 135 | 112 | 23 | 42 |
| Minimum | 0.8 | 0.8 | 0.8 | 1.7 | 2.1 | 0.1 | 0.1 | 0.1 | 0.2 | 0.6 |
| Maximum | 12.1 | 11.0 | 4.8 | 11.0 | 12.1 | 50.8 | 50.8 | 25.0 | 50.8 | 13.4 |
| 1st Quartile | 2.2 | 1.6 | 1.6 | 2.5 | 4.1 | 1.6 | 1.6 | 1.6 | 1.9 | 1.8 |
| Median | 3.7 | 2.5 | 2.4 | 3.2 | 4.6 | 2.9 | 2.9 | 2.6 | 6.8 | 3.1 |
| 3rd Quartile | 4.6 | 3.7 | 3.5 | 5.5 | 6.7 | 5.7 | 5.5 | 4.5 | 11.7 | 6.4 |
| Mean | 4.1 | 3.0 | 2.5 | 4.8 | 5.8 | 4.6 | 4.8 | 3.9 | 8.8 | 4.2 |
| Borne inf. on mean (95%) | 3.0 | 1.8 | 1.8 | -2.0 | 4.0 | 3.8 | 3.7 | 3.2 | 4.1 | 3.2 |
| Borne sup. on mean (95%) | 5.1 | 4.1 | 3.2 | 11.5 | 7.5 | 5.4 | 5.8 | 4.7 | 13.5 | 5.2 |

Fig. 6 Descriptive statistics (box plot graph) of the breach dimensions ratio W_i/H_D observed for embankment dam (EFL + RFL) failed by overtopping (OT) obtained based on the 177 cases adapted from Bernard-Garcia and Mahdi [2] database and from the 30 cases adapted from Wahl [5]

dimensions (illustrated in Fig. 1) W_{ave}/H_D and W_{BB}/H_D ratios specified in Table 1 guidelines stand between 1 and 5.

Figures 2 and 6 have thus been used to provide technical findings adapted for the guidelines framework and an **“updated” embankment overtopping dam breach dimensions** (W_i/H_D) adapted from Bernard-Garcia and Mahdi [2] database. Notice that to provide scalable results, Wahl [5] database has also notably been used for comparison, in Figs. 3, 5 and 6. These notably highlight that despite the recent “progress” in recording and sharing historical dam failures data, *the “accuracy” of these statistical/empirical standard specifications is still intimately linked to the amount of “detailed” historical dam failure case studies.*

Since the breach depth (H_B) is generally assumed equal to the dam height (H_D) in most guidance (see Table 1), Fig. 2 illustrates that this assumption is conservative (i.e. $H_B \leq H_D$) for almost 80% of the historical dam failures recorded in Wahl [5] and Bernard-Garcia and Mahdi [2] databases. Moreover, Fig. 3 illustrates that this “true” with no regards to the dam height (notably used to classify the dam, as a “small” or a “large” dam).

In the same vein, since breach dimensions ratios guidelines (W_{ave}/H_D or W_{BB}/H_D) are formulated using the dam height (H_D) instead of the breach depth (H_B), Fig. 4 highlights that this assumption tends to underestimate the ratio value W_i/H_i independently of the breach width considered (bottom, average or top breach width). Figure 4 also highlights that for the ratios W_{ave}/H_i and W_{BB}/H_i , the values stand between **2 and 5** independently of the height dimensions (H_i) considered, i.e. dam height (H_D) or breach depth (H_B). Also, by comparing the ratios graphically observed in Fig. 5a, b it is possible to highlight that the overall range of “possible” ratios (W_i/H_D) value obtained by Bernard-Garcia and Mahdi [2] (resp. Wahl [5]) database stands between **0.25 and 20** (resp. **0.5 and 20**).

Finally, as presented in Fig. 6, according to Bernard-Garcia and Mahdi [2] database (resp. Wahl [5]), the “**updated**” **interquartile range of breach dimensions ratio W_i/H_D value** stands between **2 and 6** (resp. **2 and 4**). *Therefore, even if overtopping embankment dam breach dimensions guidelines cover Wahl [5] interquartile range, it does not cover the updated range deduced using the largest and latest historical dam failures database available [2].* Figure 6 also highlights that **the cases, where the breach depth (H_B) is higher ($H3$) than the dam height (H_D)**, traduce a wider and higher range of “ W_i/H_D ” ratios, i.e. standing between **2 and 12**. In other words, depending on the foundation quality of the dam (for example), a wider and/or higher range of breach ratio dimension (W_i/H_D) might be justified for a specific dam breach study.

6 Conclusion

In conclusion, this paper aims to provide an overview of dam breach guidelines formulated for overtopping dam failures and an up-to-date portrait of the current embankment overtopping dam breach dimensions, based on the largest historical dam failures available in the literature [2]. To provide scalable results, Wahl [5] database has also been presented for comparison. Indeed, even if overtopping embankment dam breach dimensions guidelines cover Wahl [5] interquartile ratio range, it does not cover the updated breach dimensions ratio range obtained using Bernard-Garcia and Mahdi [2] database.

Since more “updated” portrait, for embankment dam failures, can easily be drawn from Bernard-Garcia and Mahdi [2] database, e.g. formulated for other dam failure modes (e.g. piping ...) and/or other dam breach parameters (e.g. time failure, ...). Indeed, the results presented in this paper not only represents a considerable asset for a wide audience (e.g. researchers, dams’ owner, civil engineers, ...) but, furthermore, foresees the possibility of improving safety studies and ensure better dams’ management.

Overall, despite the recent progress in the recording and sharing of historical dam failures case studies (e.g. Bernard-Garcia and Mahdi [2] database), highlight that historical embankment dam failures are the “dam type” case studies recording most (i.e. at least more than 10 case studies) “detailed” historical dam breach data.

Therefore, a special care thus needs to be considered in the future to record, share and investigate other dam type failure, as concrete overtopping dam failures.

7 Data Availability Statement

This database (V1), an Excel spreadsheet file, is available to the public and can be downloaded: <https://doi.org/10.5683/SP2/E7Z09B> which is hosted by Scholars Portal Dataverse and deposit by Polytechnique Montreal. Also, this database has finally been released free of charge for *open access* data retrieval and for unrestricted use under a CC0 License in February 2020.

Acknowledgements This research was supported in part by a National Science and Engineering Research Council (NSERC) Discovery Grant, application No.: RGPIN-2016-06413.

References

1. Addinsoft (2019) XLSTAT statistical and data analysis solution. Boston, USA. <https://www.xlstat.com>
2. Bernard-Garcia M, Mahdi T-F (2020) A worldwide historical dam failure's database. Scholars Portal Dataverse. V1. <https://doi.org/10.5683/SP2/E7Z09B>
3. Hydrologic Engineering Center (2014) Using HEC-RAS for dam break studies. Aug 2014. <https://www.hec.usace.army.mil/publications/TrainingDocuments/TD-39.pdf>
4. Marche C (2008) Barrages: crues de rupture et protection civile. 2e éd. Presses Internationales Polytechnique. Annexe 1, Montréal
5. Wahl TL (1998) Prediction of embankment dam breach parameters: a literature review and needs assessment. Dam safety research report DSO-98-004. US Bureau of Reclamation (USBR)

Robust Cross-Validated Feature Selection for Machine Learning Streamflow Forecasting Models: A Case Study



A. Gharib, E. G. R. Davies, and N. Ilich

1 Introduction

The value of a streamflow forecasting model lies in its ability to reliably inform water managers up to months in advance of a potential event. Two characteristics are important in this regard: the reliability of model forecasts, and the ability to accurately estimate the model's generalization performance. Machine learning (ML) data-driven models are increasingly being used due to their ability to model the inherent complexity in streamflow. Such models are based on establishing statistical inference between inputs, in the form of antecedent conditions such as observed streamflow and temperature as well as global climate indices. A vital step in building an ML model is feature selection, which identifies the set of input variables that produces the best model performance as well as reducing model complexity. However, feature selection is often applied by using the training set both to build the model and to select the important features, which leads to the selection of irrelevant features and increases the model complexity. The likelihood of encountering this problem increases with short historical records, large numbers of variables and complex ML models, all of which are commonly encountered when applying ML algorithms to streamflow forecasting. Recently, Gharib and Davies [4] proposed the use of a cross-validated feature selection algorithm that can incorporate feature selection as a part of the model building process. The algorithm resamples the feature selection process as an external cross-validation loop to apply the feature selection algorithm and identify the feature set with the minimum residual sum of squares, and an inner cross-validation loop for application of a grid search to optimize the hyperparameters. Using two

A. Gharib · E. G. R. Davies (✉)
Civil and Environmental Engineering Department, University of Alberta, Edmonton, Canada
e-mail: evan.davies@ualberta.ca

N. Ilich
Optimal Solutions Ltd., Calgary, Canada

resampling loops can significantly reduce the chances of over-fitting the predictors [5]. This study aims to implement this algorithm to build a streamflow forecasting model for the Oldman River Basin (ORB), Alberta, Canada, using Radial Basis Function Neural Network algorithm driven by local hydrometeorological conditions and large-scale climate indices.

2 Study Area

The Oldman River Basin is located in a semiarid zone and drains an area of approximately 28,000 km². It stretches from the Rocky Mountains eastward through Lethbridge to its confluence with the Bow near Medicine Hat, Alberta. With over 1,427,660 ac-ft of water available for allocation to irrigation, irrigated agriculture accounts for most of the water use in the basin [1]. The Oldman River has more flow variability and has historically been more prone to droughts than the Bow River. In addition, climate projections show that natural streamflow in the ORB is generally declining [2], which will increase the challenge of responding to projected increases in irrigation, municipal, and environmental water demands [3]. With accurate streamflow forecasts, better water allocation and drought management can be accomplished. In this study, naturalized summer-season streamflow data for the Oldman River above the Little Bow River confluence was used. Other input features included in the study are linked to streamflow and drivers of streamflow variability during winter and summer months in Western Canada. At the local scale, the hydrometeorological variables tested were monthly snow water equivalent (SWE) from snow courses, monthly precipitation, average daily maximum and minimum temperatures, monthly maximum and minimum temperatures, estimated solar radiation, and average humidity. Naturalized streamflow and SWE data were obtained from Alberta Environment and Parks. Other local hydrometeorological data was obtained from Alberta Agriculture and Forestry. At the global scale, we tested the monthly time series of four global climate indices including the Southern Oscillation Index (SOI), the Pacific Decadal Oscillation (PDO), the Atlantic Multidecadal Oscillation (AMO), and the Pacific-North American pattern (PNA). PDO, AMO, and PNA data was obtained from the National Oceanic and Atmospheric Administration (NOAA) database, and SOI data was obtained from the Climate Research Unit at the University of East Anglia. The input dataset had a historical record of 46 years from 1970 to 2015.

3 Methods

This section describes the techniques used to apply the two-resampling-loops algorithm to build the ORB summer-season streamflow forecasting model. In general, model training and validation were applied using inner and outer resampling loops

to properly resample the variable selection process. The inner loop was used for hyperparameter tuning and the outer loop was used for applying the variable selection algorithm. In each loop we used 10 times repeated 10-fold cross validation to estimate the model performance for 100 different validation sets, which can give a robust estimate of the model generalization performance. The input dataset was split into training and test sets. We used 20% of the input dataset for testing. Input monthly features were lagged by 1–12 months to represent conditions up to one year ahead of the forecast issue date (the end of April every year). Input data was standardized to improve the calculations' numerical stability, and was included inside the cross-validation loops, to prevent revealing information about the validation sets [5]. For feature selection, we used a stepwise variable selection algorithm which directly evaluates model performance for the combination of predictors. A radial basis function neural network (RBF) model was trained. RBF has input, hidden, and output layers, along with a Gaussian activation function. The number of hidden nodes in RBF was tuned by grid search for values from two up to the number of samples in the training data, and the optimum number was based on correspondence to the minimum root mean square error (RMSE). The model was evaluated using the root mean square error (RMSE), normalized root mean square error (NRMSE), mean absolute percent error (MAPE), and the correlation coefficient (r). The selected final model corresponds to the best generalization performance and has the optimal input predictors.

4 Results and Discussion

The feature selection algorithm in the outer cross-validation loop successfully reduced the number of potential input variables from 130 to 11 variables that were selected more than 10% of the time, as shown in Fig. 1. The top selected variables included the maximum, minimum average temperatures, and precipitation in the winter, fall, and summer months preceding the forecast. In addition, a single SWE variable from snow course 14A03 in April was chosen. The PDO in December was the only global climate index selected among the top variables. The two resampling loops reduced the number of possible candidate models by reducing the number of potential input predictors. They also gave valuable insight into model complexity through the probability mass functions of the number of input predictors and the number of hidden nodes selected by the models fitted using the two resampling loops, as shown in Fig. 2. The number of input predictors ranged from 2 to 8 predictors with an average of 4 predictors, and the number of selected hidden nodes ranged from 2 to 11 nodes.

The performance indicators of the streamflow forecasting model were estimated using the training and test sets, as shown in Table 1. The bias in estimating the performance indicators using the training was small when compared with the values estimated using the test set, which suggests that the model was not overfitted to the training data. The final model had a correlation coefficient of 0.715. A comparison

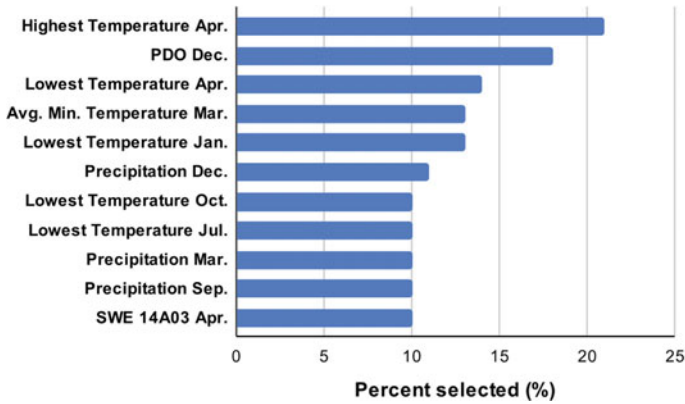


Fig. 1 The top input variables selected by the feature selection algorithm in the outer cross-validation loop as well as the percentage of time each was selected

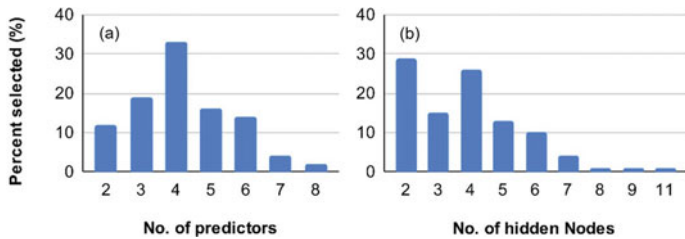


Fig. 2 The probability mass functions of (a) the number of input predictors selected by the outer cross-validation loop and (b) the number of hidden nodes selected by the inner cross-validation loop

Table 1 Performance indicators using root mean square error (RMSE), normalized root mean square error (NRMSE), mean absolute percent error (MAPE), and correlation coefficient (r) for the training and test sets

| Dataset | RMSE (m ³ /s) | NRMSE | MAPE | r |
|----------|--------------------------|-------|-------|-------|
| Training | 72.41 | 0.250 | 0.235 | 0.787 |
| Testing | 71.32 | 0.246 | 0.189 | 0.715 |

between the predicted and observed streamflow for the training and test sets, along with the relative importance of the model input predictors, are shown in Fig. 3. The highest temperature in April and the average minimum temperature in March are the two most important local hydrometeorological predictors and represented nearly 70% of the explained variation. The PDO in December had a very important contribution to the model performance and represented 25% of the explained variation.

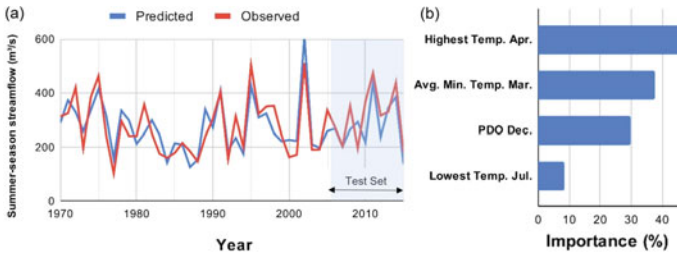


Fig. 3 (a) A comparison of the streamflow forecasting model predictions and the observed streamflow for the training period (1970–2005) and the testing period (2006–2015), and (b) The relative importance of input predictors for the final streamflow forecasting model

5 Conclusions

In this study, we implemented a cross-validated algorithm capable of considering feature selection as a part of the model building process. The algorithm resamples the feature selection process as an external cross-validation loop to apply the feature selection algorithm and identify the feature set with the minimum residual sum of squares, along with an inner cross-validation loop for application of the grid search to optimize the hyperparameters. The results suggest that the algorithm can significantly reduce the number of input predictors and give valuable insight into the final model complexity and number of input predictors. In addition, using the algorithm led to the selection of a parsimonious and more interpretable model, and avoided overfitting to the training set.

Acknowledgements This work was supported financially by the University of Alberta, Mitacs Accelerate Fellowship, and Optimal Solutions Ltd.

References

1. Alberta Agriculture and Forestry (2020) Alberta irrigation information 2019
2. AMEC (2009) South Saskatchewan River Basin in Alberta water supply study. Available from [http://www1.agric.gov.ab.ca/\\$Department/deptdocs.nsf/all/irr13053/\\$FILE/ssrb_main_report.pdf](http://www1.agric.gov.ab.ca/$Department/deptdocs.nsf/all/irr13053/$FILE/ssrb_main_report.pdf)
3. Byrne J, Kienzle S, Johnson D, Guy Duke V, Gannon BS, Thomas J (2006) Current and future water issues in the Oldman River Basin of Alberta, Canada. *Water Sci Tech* 53(10):327–334
4. Gharib A, Davies EGR (2021) A workflow to address pitfalls and challenges in applying machine learning models to hydrology. *Adv Water Resour* 152:103920
5. Kuhn M, Johnson K (2013) *Applied predictive modeling*. Springer, New York

A Multiyear Infrastructure Planning Framework for Connected and Automated Vehicles



Fehintola Sanusi, Juyeong Choi, and Ren Moses

1 Introduction

Connected and Automated Vehicles (CAVs) offer great potential in improving the safety, mobility, and environmental sustainability of the transportation system. However, a lack of adequate infrastructure to support safe and reliable CAV operations may stall the anticipated benefits. Some field evaluations of CAV technology have indicated that the cameras and other sensors may have difficulty correctly detecting faded pavement markings or traffic signs, which can affect CAV operations [12]. CAV technologies are faced with several challenges of coping with declining infrastructure conditions such as faded pavement markings [9], obscured traffic signage [12], poor pavement surface conditions, and inconsistent traffic control devices across different states and regions [7]. CAV technologies also face other roadway challenges, including inconsistencies in the traffic signs across different regions, obscured and occluded traffic signs, failure to detect the presence of obscured and overlapping pavement markings [13]. Just as poorly maintained roadway conditions may impede human drivers' (HDVs') perception and increase their reaction time, CAV operations may be hampered given poor infrastructure conditions.

Also, the recent infrastructure report card by the American Society of Civil Engineers (ASCE) shows that about 43% of US roads are in poor or mediocre conditions. Also, the majority of the public roadways have suffered a backlog of infrastructure preservation [2]. Clearly, the existing infrastructure conditions are far from ready to support CAVs. Planning agencies and other researchers are also beginning to recognize the need for adequate infrastructure to support safe and reliable CAV operations.

F. Sanusi · J. Choi (✉) · R. Moses
Department of Civil and Environmental Engineering, FAMU-FSU College of Engineering,
Tallahassee, FL, USA
e-mail: jchoi@eng.famu.fsu.edu

© Canadian Society for Civil Engineering 2022
S. Walbridge et al. (eds.), *Proceedings of the Canadian Society of Civil Engineering Annual Conference 2021*, Lecture Notes in Civil Engineering 250,
https://doi.org/10.1007/978-981-19-1065-4_54

Thus, transportation planners are tasked with the responsibility of fully exploiting the anticipated CAV benefits by making their roadway infrastructure CAV-ready.

Prior research shows that it may take several decades before achieving higher CAV market penetration [10, 11] and that human-driven vehicles (HDVs) and CAVs will coexist in a mixed traffic environment. Therefore, as CAV market penetration changes over time, their infrastructure needs will also evolve. For instance, in the short-term with low CAV market penetration, infrastructure that can support the needs of all road users (both CAVs and HDVs) should be prioritized [7]. To accommodate these changing CAV infrastructure needs, transportation planners would require making investment decisions to properly plan for a dynamic transportation system within the agencies' budget. So far, there have been limited studies that seek to evaluate the performance of the future transportation system while considering evolving infrastructure needs over a specified time horizon. To address this gap, this study develops a conceptual framework to guide a holistic approach to multi-year infrastructure planning for CAVs.

The objective of this study is to develop a conceptual framework that can guide transportation planners and decision-makers in a holistic approach to multi-year infrastructure planning for CAVs while considering the dynamic nature and the long-term goals of their transportation system. The framework consists of five major steps which include: (i) development of a knowledge base to identify a group of infrastructure options to support CAV system functions (ii) prioritization of roadway locations needing infrastructure improvement based on the potential contribution of CAVs to their local traffic through a geographical information system analysis, (iii) predicting future infrastructure needs using a Bass Diffusion model to forecast CAV adoption rates for short-, mid-, and long-term planning phases, (iv) evaluation of the performance of the future mixed traffic environment under various infrastructure planning scenarios and (v) evaluation of the benefits and costs associated with each project to guide effective selection infrastructure systems for each planning term. The rest of this paper is organized as follows: Sect. 2 gives a detailed explanation of the different components of the proposed conceptual framework for multiyear CAV infrastructure plan, Sect. 2.1 discusses the development of the knowledge that is required to understand the evolution of CAV infrastructure needs, while in Sect. 3, the proposed framework is demonstrated through a case study of a hypothetical network scenario in the state of Florida. Lastly, the conclusion and discussion are presented in Sect. 4.

2 Conceptual Framework for Multi-year Infrastructure Plan for CAVs

For the development of a framework, this study chooses freeway merge segments for CAV infrastructure planning. Please note that freeway merge segments are more vulnerable to road crashes and/or traffic jams compared to other freeway segments

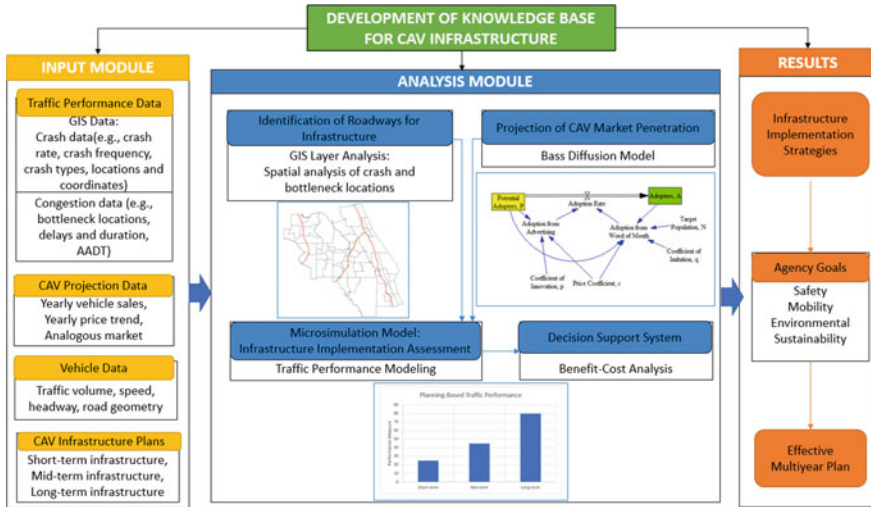


Fig. 1 Conceptual framework for multiyear infrastructure planning for CAVs

[1], which makes them suitable for CAV deployment and thus proper infrastructure planning. Given that infrastructure implementation requires significant capital investments, it is important for planners to prioritize different freeway segments within their budget constraints. As such, instead of implementing infrastructure modification/maintenance throughout the entire freeway of interest, this study assumes that such infrastructure projects selectively take place for the most-needed roadway sections. Also, since the infrastructure needs of the roadway environment will evolve, the framework assumes three distinct planning phases based on the anticipated CAV market projection. The first phase (short-term) spans from 0 to 5 years and the mid-term spans from 5 to 15 years, and the long-term will span 15 years or more. Figure 1 shows the proposed conceptual framework for a multi-year infrastructure planning for a dynamic transportation system, which consists of (i) development of a knowledge base to understand the available infrastructure options that can support CAV system functions, (ii) identification of roadway locations that require infrastructure improvement, (iii) forecasting of CAV market penetration rates to define infrastructure planning timeline, (iv) evaluation of the impacts of infrastructure implementation on CAV performance in terms of safety and mobility, and (v) analysis of associated benefit–cost to guide effective infrastructure investment options.

2.1 *Development of a Knowledge Base for CAV Infrastructure Requirements*

To achieve safe and reliable CAV driving and maneuvering operations, a certain level of infrastructure readiness must be achieved. The objective of a knowledge base of CAV infrastructure requirements is to (i) identify the different infrastructure that would be required at different CAV market penetration levels and (ii) what CAV system functions would be supported by different infrastructure options. To achieve this, the proposed framework uses a systematic literature review approach to extract information regarding the infrastructure that can support CAV system functions (i.e., platooning, merging, intersection movement, and dynamic routing) in various contexts, including safety, mobility, and environmental sustainability. Based on the literature review, two categories of CAV infrastructure were identified: existing and future infrastructure.

2.1.1 Existing Infrastructure

As previously defined, the existing infrastructure is the basic infrastructure that continues to support HDVs but would require adequate maintenance to support safe and reliable CAV operations. This group of infrastructure options is required to be in well-maintained conditions. For instance, pavement markings need to be highly visible to be easily detected by CAV sensors, traffic signage should be free from dirt, unobscured and consistent across different regions. Also, traffic signals should be upgraded with advanced technologies to enable communication with CAVs so that the vehicles can receive traffic information through the broadcast of signal phasing and timing (SPaT).

2.1.2 Future Infrastructure

As CAV market penetration increases, there will be more demand for infrastructure that can support more CAV and infrastructure communications, while some of the existing infrastructure options will be less required. For instance, it is expected to see the digital infrastructure that supports wireless technology and enabling infrastructure on the majority of the roadways to facilitate an increasing demand in vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications. Examples of these infrastructure options include roadside units (RSU) and dedicated short-range communications (DSRC) enabling infrastructure, and fiber optic cables.

2.1.3 CAV System Functions

All the identified CAV infrastructure in both categories (i.e., existing and future infrastructure) are also grouped in terms of the driving functions that are activated by the implementation of specific CAV infrastructure. Examples of CAV functions include CAV platooning, freeway merging, intersection movement, and dynamic route guidance. Each of these system functions requires an adjustment of the driver's behaviors such as reduction in vehicle speed, increase in inter-vehicle headways, or changing of lanes.

2.2 *Input Module*

Depending on the analysis level that is being completed/conducted, the proposed framework requires four different data categories that serve as input to the analysis module: traffic performance data, CAV projection data, vehicle data, and CAV infrastructure plans. The importance of each data category to the study is discussed in the following section.

2.2.1 Traffic Performance Data

The traffic performance data are necessary to perform corridor screening for locations requiring more attention for prioritization. A network screening allows for the comparison of traffic performance across different segments of the roadways (e.g., freeway mainline, ramps, or intersections). Depending on the planning agency's goals for their transportation system, different performance measures are applicable for use. The framework is also designed in such a way that integrate as much data as possible to be used as input to the analysis module as needed. To be more specific, a 5-year crash data set is used to analyze for safety and thus to identify high crash risk locations. Bottleneck locations can be used to identify regions with high traffic congestion for mobility. The data can be collected from state agencies' GIS crash database, and other telemetered traffic monitoring sites or by field surveys.

2.2.2 CAV Projection Data

Projection of CAVs market is important to define the planning analysis period and to determine what percentage of CAVs is anticipated. The market projection helps transportation planners determine the population of CAV users and what infrastructure investment decisions should be made based on the evolution of the traffic environment. For instance, in the short-term, infrastructure should be adapted in such a way that both HDVs and CAVs will be supported. Since CAVs are yet to be broadly deployed, however, there is little information about the historical data that can be

used for sales projection; hence, data from an analogous market (i.e., technology with similar market characteristics) can be used in predicting the sales of CAVs.

2.2.3 Vehicle Data

Traffic data including vehicle speed, roadways geometric, traffic volume (e.g., average annual daily traffic), vehicle arrival rates, and other traffic-related data would be required to develop the microsimulation model to be used to conduct traffic analysis and evaluate the effectiveness of CAV infrastructure options on the future roadway (i.e., the performance of CAVs and HDVs). These data can be obtained from state or local transportation agencies' databases or traffic monitoring sites.

2.3 Analysis Module

The analysis module comprises four main steps which represent a holistic approach to guiding the appropriate infrastructure investment decisions in order to achieve an effective multiyear infrastructure plan for CAVs.

2.3.1 Identification of Roadway for Infrastructure Improvement

In this framework, we choose safety and mobility as major areas for improvement through CAV infrastructure planning. With that, this section aims to identify roadway segments with high crash locations and congestion, which have the potential for improvement through CAVs, and to prioritize them based on the benefits achieved through CAV infrastructure planning. Different methods such as average crash frequency, crash rate, equivalent property damage, and empirical Bayes method [6] can be employed in ranking critical locations for safety. We, first, collect the crash data and bottleneck data as described in the traffic performance data section of the input module. Next, pre-selection of roadway segments is done to exclude other roadway segments that are not required for the analysis. The crash data and congestion data can then be merged to identify the intersection of the bottleneck location and the associated crash locations. The roadway segments can then be ranked based on the combined high crash and bottleneck ranking to select the top-ranked road segments for infrastructure implementation.

2.3.2 CAV Market Projection

The objective of CAV market projection is to understand the anticipated CAV market penetration and the implication on infrastructure planning. Different factors such as

socio-demographic factors (e.g., age and socio-economic wealth) [14], price, technology awareness, and perception may influence people's willingness to adopt CAV technology. As such, it is very important to determine the market size when planning for CAV infrastructure. While various forecasting models have been used for projecting the dissemination of CAV technologies, the Bass Diffusion model is one of the most common methods found in the literature as well as has been widely applied to projections for other automotive technologies. The Bass diffusion model consists of an innovation parameter, an imitation parameter, and a market size which can be used to forecast the sales of a new product [5]. Other models such as logistic models and time series models can also be used to forecast new product sales [4].

2.3.3 Traffic Performance Modeling

This framework proposes a microsimulation approach for the analysis of the infrastructure implementation on traffic performance to evaluate its benefits. Note that different infrastructure options can support different CAV system functions (e.g., freeway merging). Also, depending on the proportion of CAVs and HDVs within the traffic environment, preferred infrastructure options may vary across different planning phases. The microsimulation model is used to evaluate the influence of different combinations of infrastructure options on the performance of mixed roadway environments (e.g., in terms of lane changing, speed adjustment, cooperative braking, and time headway distance) to inform the optimal infrastructure plans for different planning terms. For instance, in the case of co-located traffic signs (e.g., freeway and exit ramp speed signs), CAVs can confuse speed signs for exit ramps and react by reducing their speed, which can affect the traffic flow. The impact of such an abrupt change in CAV driving behaviors will be incorporated into the simulation model for evaluation. Other HDVs and CAV driving behaviors that would be affected by infrastructure implementation will be incorporated within the traffic model as well.

2.3.4 Decision Support System

The proposed framework uses a Benefit–Cost Analysis (BCA) approach to conduct an economic appraisal for evaluating and comparing the costs and benefits of different infrastructure implementation plans for a specific roadway location under different planning timelines. The BCA will also help the public agencies to prioritize their infrastructure investments by comparing the benefits and costs across multiple locations. For the BCA, the base case is defined as a “do-nothing approach” versus different alternatives for the selected roadway segments. The alternatives are defined as strategies or countermeasures that would be implemented.

Table 1 List of CAV infrastructure for different planning timeline

| Infrastructure recommendations | Performance measures | Planning phase |
|--|----------------------|----------------|
| High-visibility lane marks and traffic signs, advanced traffic signal controller, Variable Message Signs (VMS) and RSUs | S, M, E | Short-term |
| Dedicated lane, high-visibility lane marks, and improved pavement surface condition, machine-readable signs, advanced traffic signal controller, data centers and RSUs | S, M, E | Mid-term |
| A combination of previously deployed infrastructure across all regions | S, M, E | Long-term |

3 Case Study

This section demonstrates the implementation of the proposed framework in developing a multi-year infrastructure plan for CAVs. In this case study, Florida District 5 is selected. Initially, all three interstates (i.e., I-4, I-95, and I-75) within the region were included for pre-evaluation. Based on the selection criteria, however, only freeway merge sections of I-4 will be considered for further analysis within the framework.

3.1 Development of Knowledge Base for CAV Infrastructure

Table 1 shows a list of different infrastructure options that are identified for different planning phases (i.e., short-term, mid-term, and long-term). The performance measures to be achieved with the implementation of different infrastructure options are also identified. Please note that depending on the ratio of HDVs to CAVs, feasible infrastructure options may vary across different planning terms. For instance, in the short-term, high visibility pavement markings and traffic signage would be preferred to support both HDVs and the small number of CAVs in the traffic stream.

3.2 Identification of Critical Roadway Locations

Crash rate analysis records for 5 years are taken from the Florida crash database repository of the State Highway System (SHS). Data for bottleneck locations within Florida Department of Transportation (FDOT) district 5 were also collected from the FDOT traffic database. Systematic identification of crashes occurring on freeway merging areas was conducted. The technique used to identify merge segments was to search for road geometry with “on-ramp”, “exit”, and “entry” in the data attributes. Locations satisfying the defined criteria were then selected and identified if they were within the interstates in FDOT District 5. Next, ArcGIS geoprocessing tool

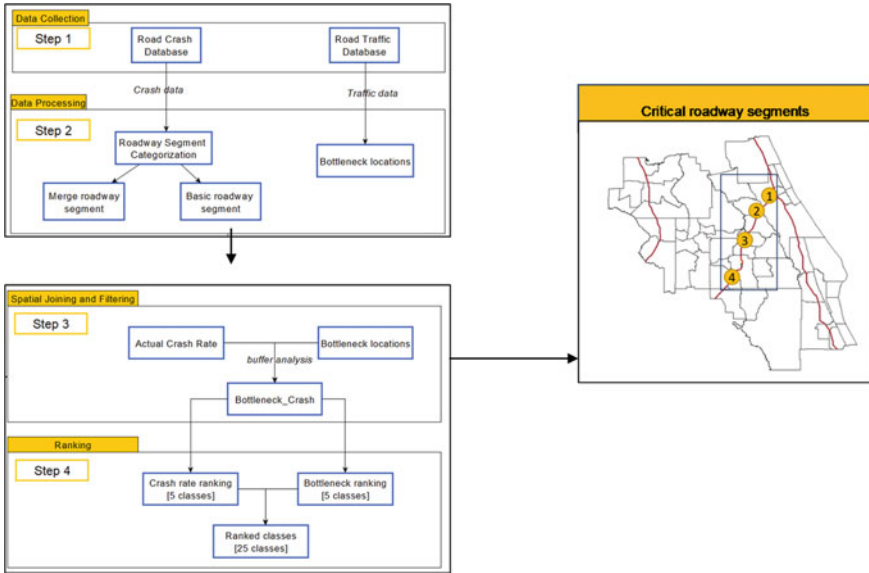


Fig. 2 Critical roadway segments for CAV Infrastructure Implementation

(i.e., intersect tool) was used to find the overlap between the crash locations and the bottleneck locations for the selected freeway merge segments. Also, not all bottlenecks can be associated to a specific crash location; as such, a buffer analysis was used to narrow down only bottleneck locations that are within a specified distance to the crash locations. To achieve this, buffer zones of 1500 ft were selected since this distance can be considered to be within freeway merge influence area [3, 8]. A polygon of intersections between the high crash rate segments and freeway bottleneck segments were obtained and used for ranking purposes in order to determine critical locations. Ranking criteria were defined using a combined ranking method for crash rate and bottleneck locations. The selected critical locations based on the ranking criteria are shown in Fig. 2 with all the critical locations occurring on I-4.

3.3 Generalized Bass Diffusion Model for CAV Market Projection

The Generalized Bass Diffusion Model (GBDM) for this study assumes (i) a fixed market size (m) throughout the analysis period and (ii) a homogenous market (i.e., potential adopters and adopters) with no repeated purchase. Since there is no historical sales data for CAVs, a similar product with available market sales data was chosen. In this case, the historical sales of hybrid electric vehicles (HEVs) were used

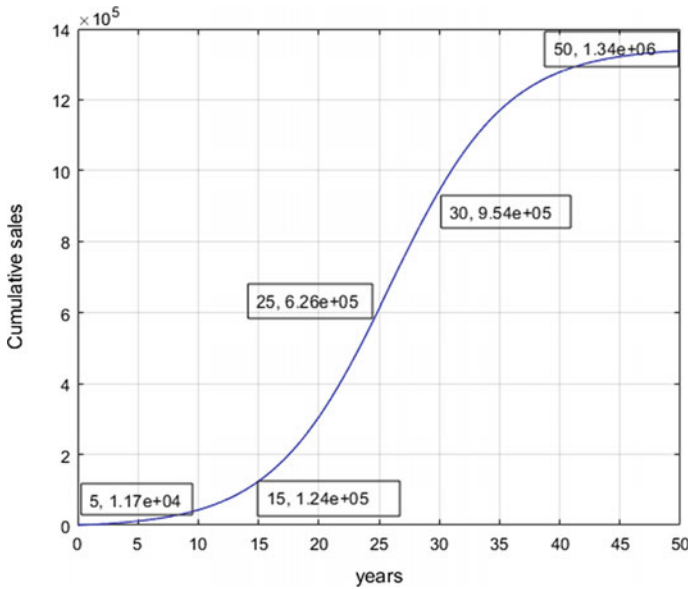


Fig. 3 Forecasted CAV market penetration for infrastructure planning

given that both CAVs and HEVs have similar benefits (i.e., mobility and environmental benefits); however, the HEVs may not be as revolutionary as CAVs [10]. To obtain a more realistic and reasonable innovation parameter for CAVs, the parameter can be calibrated using a technology that is similar in terms of adoption nature. In this case study, renewable energy seemed to be a close analogous market because it requires policies and government interventions to facilitate its adoption. HEV sales data between 2011 to 2019 in the state of Florida and hypothetical market size of 1,727,651 (the number of households in Florida district 5) were used as the model parameters. The nonlinear least-squares method was used to estimate the GBDM parameters. The predicted CAV market penetration rates are shown in Fig. 3.

3.4 Traffic Modeling

For a short-term infrastructure plan, as shown in Fig. 4, examples of infrastructure options are roadside units, clear and consistent traffic signage (e.g., merge sign), and enhanced pavement markings for clearly distinguishing traffic lanes (e.g., ramp acceleration lane). In the short-term, since a low CAV market penetration rate is anticipated, existing infrastructure options, such as the improvement in traffic signage may be preferred over the future infrastructure options. In the mid-term planning phase, as more CAVs will emerge, there will be a need for more communication infrastructure to maximize CAV benefits. Also, traditional traffic signage (e.g., freeway merge sign)

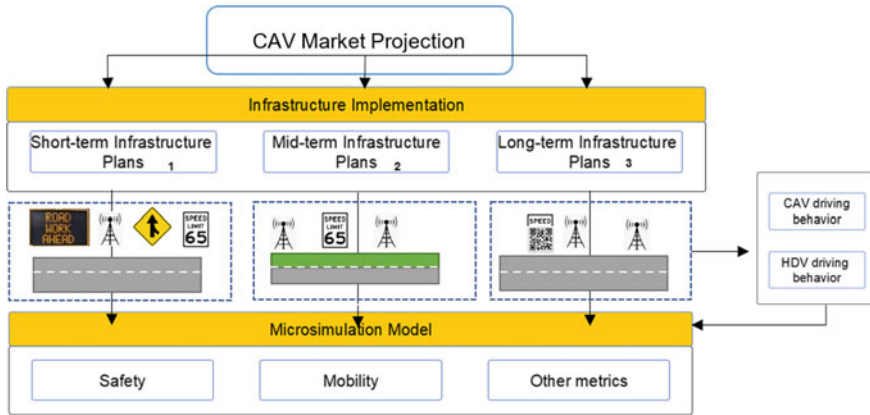


Fig. 4 Infrastructure modeling scenarios for freeway merge

can be integrated with machine-readable barcodes (i.e., machine-readable traffic signs) that can display dynamic traffic information to CAV sensors while meeting the needs of HDVs. In the long-term, the need for additional RSUs will be used to meet the increasing communication demands of a higher CAV market penetration to ensure a wider communication coverage area and continuous connectivity between CAVs and other roadway infrastructure (Table 1). For all the planning terms, it is anticipated that these infrastructure options will have varying impacts on both CAV and HDV driver behaviors. For example, the roadside units can receive real-time traffic information from the traffic management center or other vehicles on a freeway mainline. This information can then be used to coordinate the merging of vehicles on the ramp. In this case study, we propose using VISSIM, a commonly used and flexible microsimulation tool that allows the modeling of mixed traffic environments and has the capability to define different driving behaviors (e.g., car-following) through the simulation of CAVs and HDVs behaviors. Within VISSIM, the driving behavior parameters of HDVs and CAVs can be modified to evaluate the effectiveness of different combinations of infrastructure options.

3.5 Decision Support System

In this case study, the baseline or do-nothing scenario is compared with different planning alternatives for the planning terms. Table 2 shows the infrastructure options or improvements from the base case that are proposed for different planning terms. According to the knowledge base development, there are two infrastructure categories for safe and reliable CAV operations: (i) existing and (ii) future infrastructure. Transportation planners can choose combinations of these two categories across different planning terms (i.e., based on infrastructure needs of the mixed traffic environment).

Table 2 Proposed CAV infrastructure investment options for different planning phase

| Short-term | Mid-term | Long-term |
|--|---|--|
| <ol style="list-style-type: none"> 1. Repaint pavement markings 2. Increase the width of pavement markings 3. Implement unified pavement markings | <ol style="list-style-type: none"> 1. Re-paint pavement markings 2. Implement unified pavement markings | <ol style="list-style-type: none"> 1. Maintain pavement markings |
| <ol style="list-style-type: none"> 1. Repair damaged traffic signage face 2. Re-position traffic signage for visibility/Change location of traffic sign 3. Replace traffic signage sheeting materials with highly retroreflective materials | <ol style="list-style-type: none"> 1. Install machine-readable traffic signage 2. Maintain existing traffic signage | <ol style="list-style-type: none"> 1. Maintain existing traffic signage 2. Maintain machine-readable traffic signage |
| <ol style="list-style-type: none"> 1. Install roadside units at one of the selected locations | <ol style="list-style-type: none"> 1. Maintain roadside unit 2. Install roadside units at two of the selected locations | <ol style="list-style-type: none"> 1. Install roadside units on all selected location |

For example, planners may choose to repaint the pavement markings than can be easily detected by CAV sensors and HDVs (i.e., human drivers), and also install RSU to enhance CAV operations in the short-term. Similarly, in the long-term, the agency may choose to install more communication infrastructure instead of focusing on maintaining the pavement markings since there will be a larger number of CAVs deployed and a less number of HDVs in the traffic stream, thereby resulting in more communication needs. For this study, travel time savings will be used to quantify mobility-related benefits while conflict reduction will be used as a measure for safety-related benefits. The major form of associated costs for infrastructure implementation will be installation, maintenance, and operations costs.

4 Discussion and Conclusion

With the emergence of CAVs technology, transportation planners have faced the task of new design and policy challenges to ensure the readiness of their transportation infrastructure to facilitate safe and reliable CAV operations. To better prepare for the implementation of these technologies, planners need to effectively plan for the anticipated transformation that CAVs will bring into the transportation system. This study proposed a conceptual framework for multi-year CAV infrastructure planning that guides transportation planners in making appropriate infrastructure investments. The proposed framework can be used by transportation agencies to identify CAV infrastructure options available during different planning terms and develop effective plans based on the projection of future infrastructure needs. For a state like Florida, transportation planners may find the framework beneficial in guiding economic evaluation

of multiple infrastructure scenarios to support CAV implementation. First, groups of CAV infrastructure options for a specific traffic environment (e.g., freeways) are identified from the developed knowledge base. Next, based on the agency's long-term goals for their transportation system (e.g., zero crash, zero congestion and zero emissions), the transportation planners can perform GIS layer analysis to identify critical roadway locations that require infrastructure improvements. The framework accounts for uncertainty in CAV market penetration by developing a Bass diffusion model that can help transportation planners to project their CAV market structure. The result of the projected CAV market penetration rates can be used to define the different planning timelines (i.e., short, mid-, and long terms). Furthermore, transportation planners can then define multiple scenarios of infrastructure improvements for their selected roadway locations. For instance, for low CAV market penetration rates (i.e., short-term), planners may consider the maintenance of the roadway by implementing CAV-ready pavement markings and positioning traffic signage in appropriate locations that are easily detected by the CAV sensors. Similarly, for a medium CAV market penetration rate (i.e., mid-term), planners can consider the use of CAV dedicated lanes to improve traffic operations and maximize CAV performance. Analysis through traffic simulation approach can be used to evaluate the impacts of the infrastructure improvement on traffic performance. Using a benefit–cost analysis approach, the user cost components (e.g., travel time savings and crash reduction savings) can be identified, quantified, and compared to the benefits of infrastructure implementation in order to determine the best CAV infrastructure plan. While this framework is holistic to guide transportation planners for the multiyear infrastructure planning based on the evolution of CAV infrastructure needs, more research is still necessary to improve the robustness of the framework. For example, if the framework goes hand-in-hand with an optimization model that facilitates finding optimal infrastructure plans under varying planning scenarios (i.e., with different planning terms and performance metrics [e.g., safety, mobility, and environmental sustainability]), the results of the planning framework will become more robust and effective.

References

1. Ahammed MA, Hassan Y, Sayed TA (2008) Modeling driver behavior and safety on freeway merging areas. *J Transp Eng* 134(9):370–377. [https://doi.org/10.1061/\(ASCE\)0733-947X\(2008\)134:9\(370\)](https://doi.org/10.1061/(ASCE)0733-947X(2008)134:9(370))
2. American Society of Civil Engineers (2021) 2021 infrastructure report card. <https://infrastructurereportcard.org/cat-item/roads/>
3. Atamo MA (2012) Safety assessment of freeway merging and diverging influence areas based on conflict analysis of simulated traffic. ProQuest Dissertations and Theses. <https://search.proquest.com/docview/1012347854?accountid=188395>
4. Ayyadi S, Maaroufi M (2018) Diffusion models for predicting electric vehicles market in Morocco. In: *EPE 2018—proceedings of the 2018 10th international conference and expositions on electrical and power engineering*, pp 46–51. <https://doi.org/10.1109/ICEPE.2018.8559858>
5. Bass FM, Krishnan TV, Jain DC (1994) Why bass model fits without decision variables

6. Bham GH (2009) Identification and analysis of high crash segments on interstate, US, and State highway systems of Arkansas. *Mbtc Dot* 2099/3006, no. 2099
7. Chapman S, Agashe N (2019) Traffic signs in the evolving world of autonomous vehicles. <https://reflectives.averydennison.com/content/dam/averydennison/reflective-responsive/documents/english/white-papers/traffic-signs-evolving-world-autonomous-vehicles.pdf>
8. Chen H, Liu P, Lu JJ, Behzadi B (2009) Evaluating the safety impacts of the number and arrangement of lanes on freeway exit ramps. *Accid Anal Prev* 41(3):543–551. <https://doi.org/10.1016/j.aap.2009.01.016>
9. Hallmark S (2019) Preparing local agencies for the future of connected and autonomous vehicles. <http://mndot.gov/research/reports/2019/201918.pdf>
10. Lavasani M, Jin X, Du Y (2016) Market penetration model for autonomous vehicles on the basis of earlier technology adoption experience. *Transp Res Rec* 2597:67–74. <https://doi.org/10.3141/2597-09>
11. Litman T (2021) Autonomous vehicle implementation predictions: implications for transport planning. In: *Transportation research board annual meeting*, vol 42
12. Roper Y, Rowland M, Chakich Z, McGill W, Nanayakkara V, Young D, Whale R, Jones C, Zhou R (2018) Implications of traffic sign recognition (TSR) systems for road operators
13. Sage A (2020) Where's the lane? self-driving cars confused by Shabby US roadways. <https://www.reuters.com/article/us-autos-autonomous-infrastructure-insig/wheres-the-lane-self-driving-cars-confused-by-shabby-u-s-roadways-idUSKCN0WX131>
14. Shabanpour R, Shamshiripour A, Mohammadian A (2018) Modeling adoption timing of autonomous vehicles: innovation diffusion approach. *Transportation* 45(6):1607–1621. <https://doi.org/10.1007/s11116-018-9947-7>

The Influence of Socioeconomic Variables on Cycling Infrastructure Preferences in Calgary



H. Smith and F. Sadeghpour

1 Introduction

1.1 Background

Active transportation (AT) as defined by the government of Canada is any form of human-powered travel [7]. Cycling is among the most common forms of active transportation. AT has numerous benefits over vehicle transportation. AT encourages increased physical activity that can lead to numerous positive health public health outcomes such as a decreased risk of heart disease, high blood pressure, and type 2 diabetes [1]. Because cycling is a self-propelled form of transportation, it generates less air pollution and is less carbon intensive than motor vehicle transport [18]. Economically, AT has been shown to decrease costs related to traffic congestion, healthcare, employee absenteeism, and transportation infrastructure [1]. It also directly decreases the money spent on car maintenance and insurance. Because cycling is one of the mode affordable modes of transportation, it is among the most equitable of all transport modes, meaning it should be equally available to the entire population [14]. The benefits related to cycling as a form of active transportation is the motivation behind why many cities, including Calgary, are aiming to increase the frequency of cycling within the city.

Numbers from the 2016 census show that there is room to improve in regard to cycling frequency. Calgary has a low cycling mode share at 1.5% compared to other cities [16]. This can be attributed to the challenges that arise with winter cycling. A study done in Calgary showed that the number of frequent cyclists drop by approximately one third in cold weather conditions [2]. Wind, snow, cold temperatures, and unfavourable surface conditions as a result of winter are all reasons cited for

H. Smith (✉) · F. Sadeghpour
Schulich School of Engineering, Calgary, Canada
e-mail: hannah.smith@ucalgary.ca

decreased cycling frequency [2, 19]. Because there are costs associated with winter cycling infrastructure maintenance, widespread implementation needs to be justified.

However, countries in Europe have found ways to keep cycling numbers high, even in colder weather [10]. Cycling participation in North America is significantly lower compared to European cities. In the Netherlands, Germany, and Denmark, cycling levels are more than 10 times higher than in Canada and the United States [14]. Government policies, that favour cycling, are the reason why the cycling rates are much higher. These countries have focused on making their cities “people-friendly” as opposed to the “car-friendly” cities seen in North America [14]. In Northern Europe, cities that share a similar winter climate to Calgary, found that the two main factors that influence winter cycling rates are (1) the quality of cycling infrastructure and (2) the degree of winter bike path maintenance [10]. By controlling these factors, Northern European countries have been able to keep cycling rates constant year round.

1.2 Problem Statement and Objectives

As a low-cost transportation alternative, cycling is among the most equitable transport modes [14]. There is an opportunity to increase the cycling frequency in Calgary as the current numbers are low. Among many variables, cycling infrastructure has also been shown to affect cycling frequency, especially in the winter [19]. Currently, the influence that different cycling infrastructure has on people of different socioeconomic backgrounds is unknown. For the City of Calgary to recognize the benefits of cycling as a form of active transportation, cycling infrastructure preferences of different socioeconomic groups need to be understood.

Specifically, the objective of this study is to explore the application of using utility functions to determine the preferences that different socioeconomic groups have towards cycling infrastructure, specifically cycling facilities and surface conditions, within the City of Calgary. The socioeconomic variables that will be looked at in this study are age, vehicle ownership, and household income. The results of this study would ideally be used to influence policy and infrastructure changes. This study will inform a greater understanding about where policy and infrastructure changes need to be made to increase the accessibility and desirability of cycling to different populations in Calgary. This would help provide the required justification to make these changes and ultimately lead to an increase in cycling frequency in Calgary.

2 Literature Review

Many variables have been shown to effect cycling frequency. Two groups of variables that have been studied extensively in literature are socioeconomic variables and characteristics of cycling infrastructure. Some socioeconomic variables that have been

previously looked at in relation to cycling frequency are age, vehicle ownership, and household income. Gender is also an important factor and has been previously assessed in a different study by the authors [9] and therefore will not be looked at in this study. Studies that investigated the relationship between age and cycling show that cycling frequency declines with age [6, 11, 15], with younger age groups being more likely to continue to cycle in the wintertime [3]. Studies investigating the relationship between vehicle ownership and cycling frequency found that individuals with more cars are less likely to ride a bicycle [8, 12, 13]. An overview study of over 100 different studies stated that the relationship between income and cycling frequency is unclear as different studies have found positive, negative, and no correlations between household income and cycling [8].

Facility type and surface conditions are two cycling infrastructure variables that have been explored in literature in relation to cycling frequency. Many studies have found that cyclists prefer facilities that are separated from traffic [5, 17, 19]. When looking at cycling data from 90 large American cities it was found that cities with a greater supply of bike path and bike lanes have a significantly higher bicycle commuter rate [4]. A study that investigated motivators and deterrents of bicycling found that unsafe surface conditions are a top deterrent in making the decision to cycle [2, 19]. In a study regarding winter cycling in Calgary it was found that more than half of cyclists view ice as a major concern [2]. Snow and gravel removal off bike paths and lanes were also expressed as concerns [2].

The aforementioned studies have examined the relationship between socioeconomic variables and cycling frequency as well as the relationship between cycling frequency and cycling infrastructure preferences such as cycling facility type and surface conditions. The effect that different socioeconomic variables have on cycling infrastructure preferences is currently unknown. This study specifically aims to investigate the link between socioeconomic variables and cycling infrastructure preferences.

The previous studies have mainly investigated the existence of relationships using data collected through questionnaires. The limitation of using questionnaires is that it has a qualitative nature in measuring preferences regarding cycling infrastructure options. The utility analysis approach used in this study offers alternative real-life scenarios to participants and observes their response to different cycling infrastructure alternatives. These preferences inform the utility functions that provide numerical indications that quantify the relative impacts of different factors of cycling facilities and surface conditions.

3 Methodology

3.1 Data Collection

3.1.1 Survey

This study used pre-existing data that was collected between October 2014 and March 2015. A survey was conducted for 1797 Calgarians where they were asked about their behaviours and attitudes related to cycling. The respondents were asked to provide information about the following three measures of socioeconomics:

- Year born
- Household annual before-tax income
- Household automobile holdings.

3.1.2 Stated Preference Experiment

The survey included a game where respondents were asked to indicate their preference among four different hypothetical alternatives for a winter cycling route. During the experiment, respondents were asked to imagine that they were cycling to a specific destination, that remained the same in each case, in the weather conditions that were present at noon on the day of the interview using the conditions presented in the each of the hypothetical routes. The hypothetical alternatives were described in terms of three specific attributes: cycle time, facility type, and winter related surface condition, that varied among each alternative. The cycling facilities considered in this study include “Mixed with Traffic”, “Mixed Traffic with Bike Signs”, “Bike Lane (separated by paint)”, “Cycle Track (separated by physical barrier)”, “Bike Path Shared with Pedestrians”, and “Bike Path Without Pedestrians”. The different surface conditions considered in this study were “Clear with No Gravel”, “Clear with Gravel”, “Fresh Snow (but No Ice)”, “Fresh Snow on Existing Ice”, “Snow and Ice from Previous Day”, “Ice (but No Snow)”, and “Some Snow Piles from Clearing”. The study participants were then asked to rank their preferences from most to least preferred between each of the four alternative routes, along with providing a rating from 0 to 10. This process was repeated four times for each respondent, usable data was obtained for a total of 7112 games. The combination of attributes was different (randomly selected) in each of the trials.

The design of this experiment is intended to make the respondents consider the influence of relatively good and relatively bad attributes and how they make trade offs between them. This allows responses that can be used to estimate the function parameters of discrete choice, or logit, models that can accurately represent this behaviour.

3.2 Data Analysis Approach

A utility analysis was used to determine how different socioeconomic variables effect cycling infrastructure preferences. As previously mentioned, the socioeconomic variables specifically looked at are age, household size, vehicle ownership, and household income. The indication of preferences among the four hypothetical routes in the stated preference experiment as described in the previous section were used to estimate function parameters in a discrete choice (or logit) model that represents behaviour. The function parameters serve to quantify the influence of the corresponding attributes. Equation (1) shows the utility function that represents the discrete choice model used to measure the preferences of different cycling alternatives.

$$V_a = \beta_r * RT_a + \sum_{m \in M} k_m^F * x_{a,m}^F + \sum_{w \in W} k_w^S * x_{a,w}^S \tag{1}$$

where:

a = index for alternative hypothetical cycle routes, A being the full set of routes considered by respondent

m = index for cycling facility types, M being the full set of cycling facility types considered by all respondents

w = index for surface conditions, W being the full set of surface conditions considered by all respondents

V_a = utility for alternative a (utils)

RT_a = ride time for alternative a (utils)

β_r = utility sensitivity to ride time (utils/minute)

k_m^F = utility for facility type m (utils)

k_w^S = utility for surface condition type w (utils)

$x_{a,m}^F$ = 1 when alternative a has utility type m and = 0 otherwise

$x_{a,w}^S$ = 1 when alternative a has surface condition type w and = 0 otherwise.

To obtain these function parameters, β_r , k_m^F , k_w^S , maximum likelihood techniques were used with the stated preference game data. Many different software that can perform this technique are commonly available. In this study PandasBiogeme software was used.

Parameters were estimated for each alternative. Within each alternative, the function parameter, the absolute value of the t-ratio, and equivalent ride time in minutes were also calculated. A t-ratio with a higher absolute magnitude indicates more confidence in the result of the estimated parameter. The equivalent ride time is the function parameter divided by the sensitivity to ride time, β_r . This is equivalent to the change in ride time that would need to occur to have the same impact on utility as the associated attribute. It also provides a basis for comparing the estimated function parameters in a way that is easy to understand.

The estimates for k_m^F and k_w^S parameters must be calculated relative to a singular alternative, meaning that one alternative must be fixed at 0 in order to act as a reference. In this study, the alternatives that were fixed at 0 were “Mixed with Traffic” when considering facility types and “Clear and No Gravel” when considering surface conditions. This means that the estimate for each k_m^F and k_w^S parameters is the difference in utility that corresponds to the associated facility type or surface condition, respectively. It can then be said that the more positive the estimated parameter is the more attractive the associated alternative is and the more negative the estimated parameter is the less attractive the associated alternative is, all with respect to the fixed alternatives.

4 Results

The results from the utility analysis are plotted below for ease of understanding. The results presented first are those representing the utility functions indicative of the entire sample population. Below are results from different utility analyses done with respect to the socioeconomic variables age, vehicle ownership, and household income. Figures 1 and 2 summarize the function parameters found from the utility analysis for sensitivities to cycling facilities and surface conditions for the whole population of Calgary.

As previously mentioned, the function parameters can be converted to time equivalents by dividing by the sensitivity to ride time obtained from the utility analysis. For example, when you divide the utility for cycling in “Mixed Traffic with Bike Signs” ($k_m^F = 0.214$) for the whole population, by the sensitivity to ride time ($\beta_r = -0.056$) you get -3.8 min. This indicates that adding bike signage to mixed traffic is equivalent to a reduction in ride time to 3.8 min.

“Cycle Track” and “Bike Path without Pedestrians” are the options with the highest utility for the full population, with “Cycle Track” being the most attractive option. “Bike Lanes” and “Bike Paths Shared with Pedestrians” are more attractive options

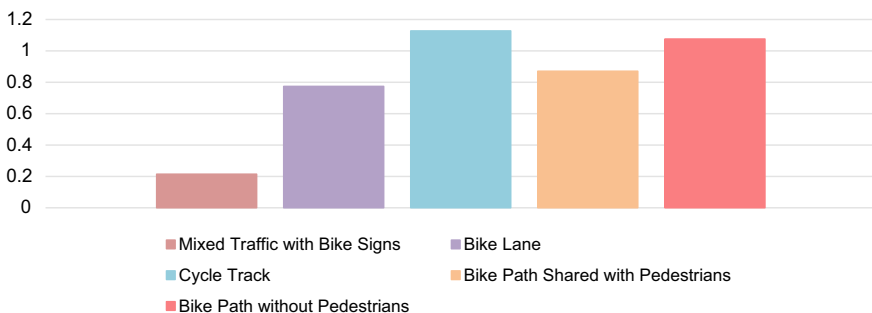


Fig. 1 Sensitivity to cycling facilities for whole population

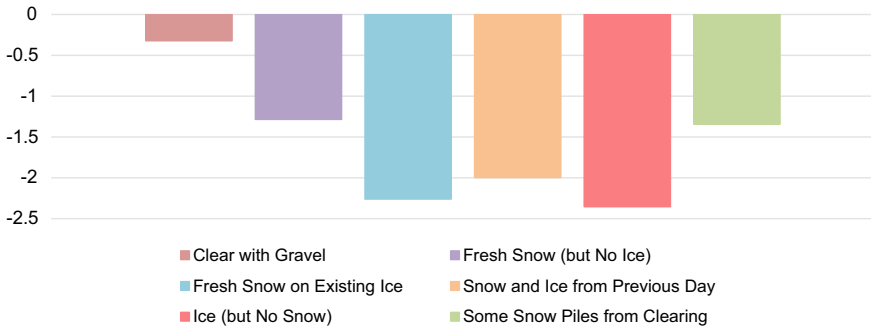


Fig. 2 Sensitivity to surface conditions for whole population

but do not have utilities as high as the former mentioned alternatives. Adding signage to the “Mixed Traffic” condition has the smallest effect on utility.

“Ice (but No Snow)”, “Fresh Snow on Existing Ice”, and “Snow and Ice from Previous Day” are the least attractive surface conditions with all utility values less than -2 . This indicates that the presence of ice has the most negative impact on utility. “Some Snow Piles from Clearing” and “Fresh Snow (but No Ice)” have more moderate impacts on utility. The condition “Clear with Gravel” has the least negative impact on utility out of all surface conditions.

4.1 Age

Figures 3 and 4 summarize the function parameters found from the utility analysis for sensitivities to cycling facilities and surface conditions for different age groups within Calgary to determine if cycling infrastructure preferences varied between age groups.

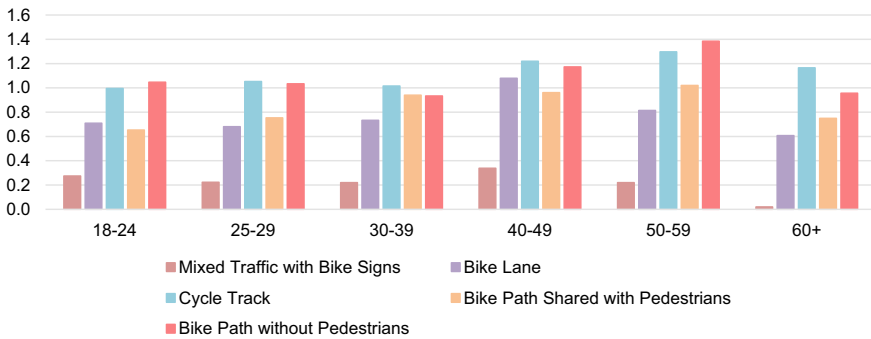


Fig. 3 Sensitivity to cycling facilities across age groups

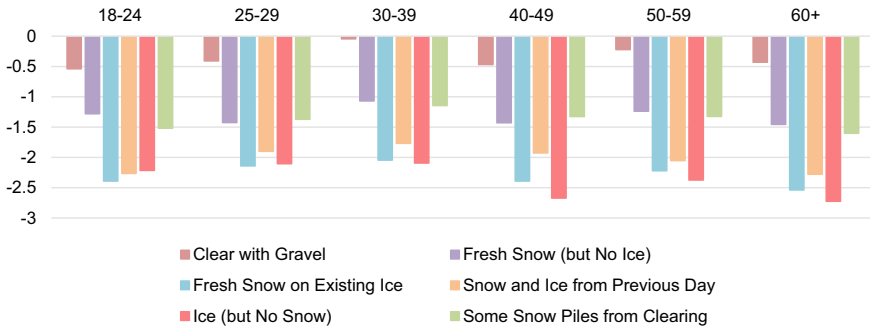


Fig. 4 Sensitivity to surface conditions across age groups

“Cycle Track” and “Bike Path without Pedestrians” are the options with the highest utility consistently across different age groups, except for 30–39. The utility for using “Bike Path Shared with Pedestrians” is slightly higher than the utility for “Cycle Track” making it the second most attractive option for this age group. Adding signs to “Mixed with Traffic” have the smallest effect on utility for all age groups, and almost having no effect at all on the age group of 60+. However, the t-ratio for the utility for “Mixed Traffic with Bike Signs” is 0.3 which indicates this is a weak estimate and cannot be taken with more than 95% confidence. Sensitivity to cycling facilities tends to increase from age groups 18–24 to 50–59, then decreases for the age group of 60+.

“Fresh Snow on Existing Ice” and “Ice (but No Snow)” consistently have the most negative impact on utility across all age groups, except for 18–24. “Snow and Ice from Previous Day” has a slightly more negative utility than “Ice (but No Snow)”. Adding “Gravel” has the least negative impact on utility across all age groups. Generally, the sensitivity to surface conditions increases across the increasing age groups. However, 18–24 year old’s show a higher sensitivity to surface conditions out of all age groups.

4.2 Vehicle Ownership

Figures 5 and 6 summarize the function parameters found from the utility analysis for sensitivities to cycling facilities and surface conditions for different vehicles per household within Calgary to determine if cycling infrastructure preferences varied between vehicle ownership.

For 0–3 vehicles per household, “Cycle Track” and “Bike Path without Pedestrians” are the two most attractive options. For households with 4 or more vehicles, “Bike Paths without Pedestrians” remain the top option and “Bike Paths Shared with Pedestrians” are the second most attractive option while “Cycle Track” drops down to third place. The sensitivity to cycling facilities remains relatively constant for households that have 0–3 vehicles. Households with 4 or more vehicles show

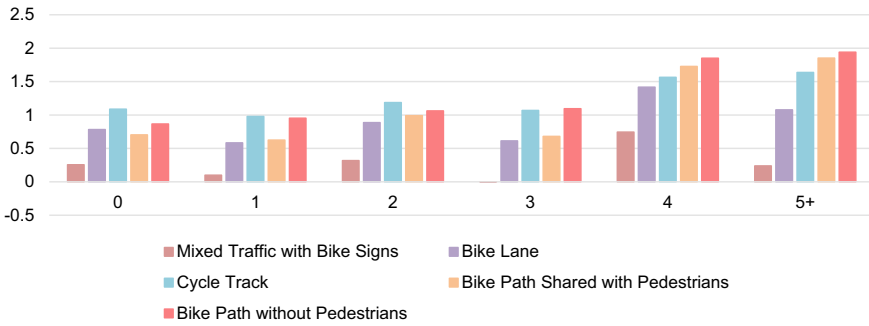


Fig. 5 Sensitivity to cycling facilities across number of vehicles per household

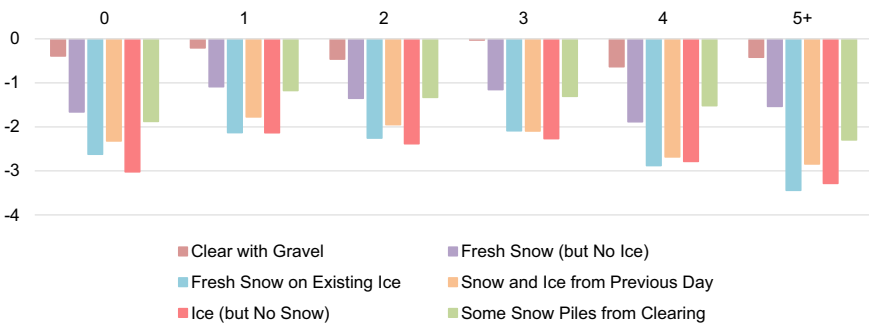


Fig. 6 Sensitivity to surface conditions across number of vehicles per household

increased sensitivity to cycling facilities. The positive impact on utility that “Bike Paths without Pedestrians” almost doubles between households with 0–3 cars to households with 4 or more cars. The estimate for “Mixed Traffic with Bike Signs” for households with 3 cars shows no impact on utility on the graph. The function parameter estimate is equal to -0.00065 , with a t-ratio of 0. This indicates that this result cannot be taken with confidence.

“Fresh Snow on Existing Ice” and “Ice (but No Snow)” remain the surface conditions with the most negative impact on utility across all household vehicle holdings. Households with 4 or more cars tend to have the highest sensitivity to surface conditions except for households with 0 cars having an increased sensitivity to “Ice (but No Snow)”. The estimate for “Clear with Gravel” for households with 3 cars shows a small negative impact on utility on the graph. The function parameter estimate is equal to -0.02926 , with a low t-ratio of 0.3. This indicates that this result cannot be taken with confidence.

4.3 Household Income

Figures 7 and 8 summarize the function parameters found from the utility analysis for sensitivities to cycling facilities and surface conditions for different household incomes within Calgary to determine if cycling infrastructure preferences varied between different household income levels.

“Cycle Track” and “Bike Paths without Pedestrians” are consistently the two most attractive cycling facility alternatives across all levels of income except for people who earn 0–29,999 dollars per year. The utility estimates for income group 0–29,999 dollars indicates that “Cycle Track” and “Bike Paths Shared with Pedestrians” are respectively the first and second most attractive options. Income levels over 90,000 dollars per year generally show higher sensitivity for cycling facility type apart from income level 125,000–149,000 dollars per year. The utility estimate for income level 125,000–149,000 dollars per year shows that the addition of signage to biking in mixed traffic has a negative impact on utility, meaning that this option is less attractive than cycling in mixed traffic. The t-ratio for this estimate is 2.3 which indicates that

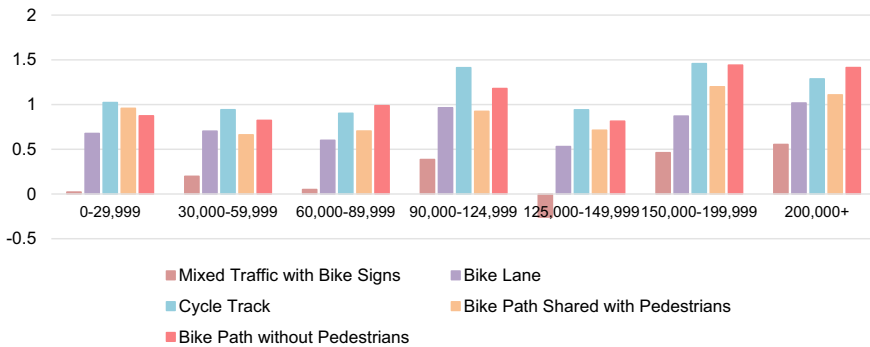


Fig. 7 Sensitivity to cycling facilities across household income

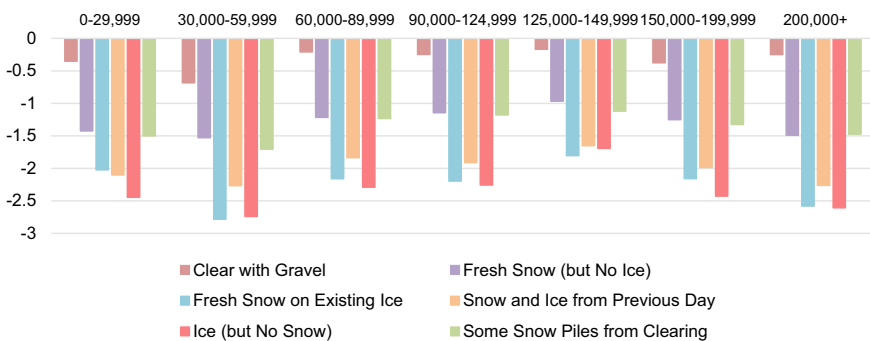


Fig. 8 Sensitivity to surface conditions across household income

this estimate can be taken at least with 95% confidence. However, compared to the other t-ratio values it is relatively low.

“Fresh Snow on Existing Ice” and “Ice (but No Snow)” remain consistent for the top two surface conditions that have the most negative impact on utility except for income level 0–29,000 dollars. In this case, “Ice (but No Snow)” and “Snow and Ice from Previous Day” are the top two least attractive surface conditions. Sensitivity to surface conditions generally remains consistent across all income groups with few exceptions. The utility estimates show that people who make between 30,000–59,999 dollars a year are the most sensitive to surface conditions and people who make between 125,000–149,999 dollars per year are the least sensitive to surface conditions.

5 Discussion and Conclusion

The objective of this study was to explore the application of a utility function analysis to determine the preferences that different socioeconomic groups have towards cycling infrastructure within the City of Calgary. With this type of analysis, it was able to be determined which cycling infrastructure conditions different socioeconomic groups were most sensitive to based on their preferences. Therefore, it can be determined that infrastructure changes made to benefit the different socioeconomic groups would ultimately increase the frequency of cycling in Calgary.

Before addressing the findings specific to each socioeconomic condition, some findings that remained consistent over all socioeconomic groups, with few exceptions mentioned in the results section, and was found to be true for the whole population was that “Bike Paths without Pedestrians” and “Cycle Tracks” were the most preferred option regarding cycling facilities. “Ice (but No Snow)” and “Fresh Snow on Existing Ice” consistently had the most negative impact on utility overall. This result shows that the implementation of more desired cycling facilities and improvements to winter bike path maintenance would lead to increased cycling frequency overall. However, some groups displayed higher sensitivities to cycling facilities and surface conditions. These results will be discussed below.

Sensitivity to cycling facilities increases from age groups 18–24 to 50–59 and then declines for age group 60+. The sensitivity to surface conditions was also found to increase over increasing age groups. It was found that households that have 4 or more cars have the highest sensitivity to both cycling facilities and surface conditions. Households without any cars also display a high sensitivity to surface conditions. In respect to household income, the utility analysis showed that individuals who had an income higher than 90,000 dollars per year had higher sensitivities to cycling facilities, except for individuals who made between 125,000 and 149,999 dollars per year. Sensitivity to surface conditions remained relatively consistent over income levels. These results indicate that making positive changes to cycling facilities and surface condition maintenance would encourage these groups of people to cycle more frequently, increasing the overall cycling frequency in Calgary.

It is worthy to note that this study does not consider the many other variables that have been proven to effect cycling behaviour. Therefore, we cannot know if there were external factors that effected the results of the study. A couple examples of other variables include network connectivity and weather conditions. Finally, these results are only indicative of Calgary and further studies would need to be done to determine if the results can be generalized to larger populations.

In conclusion, “Cycle Tracks” and “Bike Paths without Pedestrians” generally had the highest overall positive impact on utility, indicating that these were the most preferential alternatives. “Fresh Snow on Existing Ice” and “Ice (but No Snow)” were the two surface conditions that had the most negative impact on utility, making these the top deterrents to cyclists. This indicates that including more cycle tracks and bike paths without pedestrians and increasing winter bike path maintenance to avoid the accumulation of ice would influence all groups to cycle more. This can in turn increase the cycling frequency in Calgary overall. The use of the utility function to determine this provided numerical indications that quantify the relative impacts that improving cycling infrastructure would have on different socioeconomic groups in Calgary. The results of this study can help planners and policy makers to make informed decisions for funding allocation and expenditure on cycling facility and surface condition maintenance. Utility function analysis can help the development of cycling policy in cities for which they are conducted in. In the future, it would be interesting to perform this study across Canada to determine if cities are behaving the same in terms of cycling infrastructure preferences. This would aid in the development of a national cycling policy.

References

1. Alberta Centre for Active Living (n.d.) Benefits of active transportation. Retrieved 2 Oct 2020, from https://www.centre4activeliving.ca/media/filer_public/07/77/077713ef-e097-4bc5-8278-1fb455d26630/2017-active-transportation-factsheet.pdf
2. Amiri M, Sadeghpour F (2014) Cycling characteristics in cities with cold weather total green house gas emissions in Canada. *Sustain Cities Soc* 14:397–403
3. Bergström A, Magnusson R (2003) Potential of transferring car trips to bicycle during winter. *Transp Res Part A Policy Prac* 37(8):649–666
4. Buehler R, Pucher J (2012) Cycling to work in 90 large American cities: new evidence on the role of bike paths and lanes. *Transp (Dordrecht)* 39(2):409–432
5. Caulfield B, Brick E, McCarthy OT (2012) Determining bicycle infrastructure preferences—a case study of Dublin. *Transp Res Part D Transp Environ* 17(5):413–417
6. Dill J, Voros K (2007) Factors affecting bicycling demand: initial survey findings from the Portland region. Transportation Research Board, Washington, DC
7. Government of Canada (2014) Mobilizing knowledge on active transportation. Retrieved 12 Oct 2020 from <https://www.canada.ca/en/public-health/services/health-promotion/healthy-living/physical-activity/mobilizing-knowledge-on-active-transportation.html>
8. Heinen E, Van Wee B, Maat K (2010) Commuting by bicycle: an overview of the literature. *Transp Rev* 30(1):59–96
9. Hunt JD, Sadeghpour F (2019) Assessing impacts on the utility of both winter cycling and fair-weather cycling in Calgary. In: World conference on transportation research. Elsevier B.V., Mumbai

10. Jaffe E (2016) How to keep cyclists riding even in frigid snowy winter. Citylab. Retrieved 10 Oct 2020 from, <https://www.citylab.com/transportation/2016/01/winter-bike-riding-seasonal-cycling/426960/>
11. Moudon AV, Lee C, Cheadle AD, Collier CW, Johnson D, Schmid TL, Weather RD (2005) Cycling and the built environment: a US perspective. *Transp Res Part D* 10:245–261
12. Parkin J, Wardman M, Page M (2008) Estimation of the determinants of bicycle mode share for the journey to work using census data. *Transportation* 35(1):93–109
13. Plaut PO (2005) Non-motorized commuting in the US. *Transp Res Part D Transp Environ* 10(5):347–356
14. Pucher J, Buehler R (2008) Making cycling irresistible: lessons from the Netherlands, Denmark and Germany. *Transp Rev* 28(4):495–528
15. Pucher J, Komanoff C, Schimek P (1999) Bicycling renaissance in North America? Recent trends and alternative policies to promote bicycling. *Transp Res Part A* 33(7/8):625–654
16. Statistics Canada (2017) Census in brief—commuters using sustainable transportation in census metropolitan areas. Retrieved 9 Dec 2020 from <https://www12.statcan.gc.ca/census-recensement/2016/as-sa/98-200-x/2016029/98-200-x2016029-eng.cfm>
17. Taylor D, Mahmassani H (1996) Analysis of stated preferences for intermodal bicycle-transit interfaces. *Transp Res Rec* 1556:86–95
18. Transport Canada (2011) Active transportation in Canada. Retrieved 12 Oct 2020, from http://publications.gc.ca/collections/collection_2011/tc/T22-201-2011-eng.pdf
19. Winters M, Davidson G, Kao D, Teschke K (2010) Motivators and deterrents of bicycling: comparing influences on decisions to ride. *Transportation* 38:153–168

Riverine Hydrokinetic Energy Extraction: Investigation into a Location's Suitability for Turbine Deployment



K. Kirby, C. Rennie, J. Cousineau, S. Ferguson, and I. Nistor

1 Introduction

Run-of-the-river hydroelectric energy systems have received recent and increasing interest as a means to generate renewable energy in Canada. In addition to large availability of hydrokinetic energy across the nation [12], run-of-the-river systems are particularly attractive because of their low environmental impact in comparison to stored hydropower solutions. Run-of-the-river hydrokinetic energy generation essentially requires the introduction of a turbine into a river's flow, and, through the rotation of the turbine rotors, kinetic energy is extracted from the river's velocity. The amount of energy available in the flow is dictated by the velocity of the flow as illustrated by the governing hydrokinetic power equation:

$$P_K = \frac{1}{2} \rho A v^3 \quad (1)$$

where P_K is the hydrokinetic power, ρ is the density of the fluid (i.e., water), A is the cross-sectional area, and v is the flow velocity. Developers must also adhere to practical constraints related to water depth, which must be sufficient to accommodate the turbine structure. Researchers and developers often seek information regarding both flow velocity and depth to support investigation of site suitability. Generally, a suitable location has a flow velocity greater than 1.5 m/s and a minimum depth of 3 m [22].

K. Kirby (✉) · C. Rennie · I. Nistor
Department of Civil Engineering, University of Ottawa, Ottawa, Canada
e-mail: kkirb087@uottawa.ca

J. Cousineau · S. Ferguson
National Research Council Canada, Ottawa, Canada

Depending on the goal of the experiment, various methods of hydrokinetic resource estimation have been employed. These include GIS techniques and hydrological modelling [1, 2], hydrodynamic modelling using bathymetry and flow data [7], numerical modelling using flow data [4], utilization of winter satellite imagery in Canada [3, 5], regression between water level and velocity data [15], generation of velocity duration curves (VDC) using historical data [18, 19], field measurement surveys [13, 21], and combined field measurement surveys and numerical modelling [14, 26].

The goal of hydrokinetic resource assessment experiments can vary from high level, such as identification of reaches with a high likelihood of fast-moving water [3, 5], or more directed, such as the development of a flow velocity field for a specific reach using field measurements [7, 14, 21, 26]. The so-called middle ground between these two types of studies would be the generation of a VDC or similar velocity representation for the reach to calculate the amount of available power [1, 2, 5, 13, 15]. Studies that use the field measurement approach can yield a more accurate estimation of the available power throughout the reach. Among these experiments, there has yet to be a study whose primary aim is to evaluate the optimal turbine placement within a reach. An extensive literature search using the terms “hydrokinetic”, “assessment”, “river”, “potential”, and/or “survey” resulted in few studies, the oldest of which was published in 2010. This speaks to the novelty of this study and to hydrokinetic research in general.

Studies that have attempted to characterize the velocity flow field and bathymetry of a river reach have used flow meters or acoustic Doppler current profilers (ADCPs) in either a point survey or cross-section survey configuration. Petrie et al. [21] measured the flow field across two cross-sections with an ADCP with four repeat transects at each cross-section. Kalnacs et al. [13] did not specify what type of equipment was used to measure the flow field but used a point survey approach throughout the reach. Kasman et al. [14] used a flow meter (type and model unspecified) with a point survey that measured velocity at two cross-sections with three points per cross-section (left, right, and middle) and at three depths in the water column at each point location. Santos et al. [26] used an ADCP to study two reaches where 61 and 64 cross-sections were measured, and it was not specified how many transects were measured per cross-section. The use of ADCP to map velocity flow fields and bathymetry in rivers from a moving boat is now quite common [11, 17, 20, 23, 24, 28, 30]. Also, to reduce uncertainty in ADCP velocity measurements, specifically when the study is concerned with secondary flow mechanisms, it has been suggested to measure 5–6 transects at each cross-section [6, 23, 27]. In addition to prior hydrokinetic assessments, a standard was published in 2019 that describes the procedure to generate a VDC at the specific location that the renewable energy converter (i.e., turbine) will be installed for the purpose of characterizing the hydrokinetic energy available [10]. The survey in this experiment would satisfy one of the five velocity-discharge measurements required at the desired turbine deployment site as per the [10] standards.

This study evaluates field measurement protocols to determine best turbine installation location(s) within a river reach. A reach on the Rivière-Rouge, Quebec that

is suitable for hydrokinetic energy extraction and turbine deployment is used as a case study site. The site was chosen because of its ice-free conditions during winter, indicating high flow velocity conditions [3, 5]. A survey was conducted on the reach using an ADCP and a global positioning system (GPS). The survey was completed November 5, 2020 during an approximate mean flow condition. The data collected during the survey were plotted to determine bathymetry and velocity fields in the reach. Ultimately, the most suitable locations for a hydrokinetic turbine deployment within the reach were established based on observed velocities and depths. This work is expected to aid researchers and developers to improve these types of assessments and perhaps lead to deployment of a turbine and hydrokinetic energy extraction from the flow of the Rivière-Rouge.

2 Methodology

2.1 Study Reach

The study reach is located at the south-most point of the Rivière-Rouge where it discharges into the Ottawa River north of L'Orignal, Quebec. The river drains approximately 5500 km² of southeastern Quebec, Canada. Investigation of Landsat and Sentinel winter imagery (i.e., December, January, and February) revealed that the study location remains ice-free during the winter, suggesting that that the study location produces high-velocity flows and could be a desirable candidate for hydrokinetic energy generation [3, 5]. Rapids are located upstream of the study area and, as the velocity slows down as the flow enters the Ottawa River, sand deposits can be seen downstream. As such, measurements were limited any further upstream than the extent shown in Fig. 1a for to safety reasons. The bed of the study reach was observed to be composed primarily of rock and gravel. Data collection was restricted approximately 5 m from the bank on the east side of the study reach, due to shallow conditions and the presence of rocks as illustrated in Fig. 1c. The west bank was mostly sheer rock face upstream, as shown in Fig. 1b and diminished to conditions similar to that of the east bank as flow travelled downstream. The entire study reach from an aerial view is shown in Fig. 1a. The average river width was 60 m and the reach was approximately 190 m long.

2.2 Measurements

The ADCP used in this experiment was a SonTek M9 RiverSurveyor, and it was mounted on a Oceanscience trimaran riverboat which was towed approximately 1.5 m to the port side of the boat, as illustrated in Fig. 2. The ADCP apparatus was positioned as such to be outside the wake of the boat. The vessel used in the survey was a 4.6 m

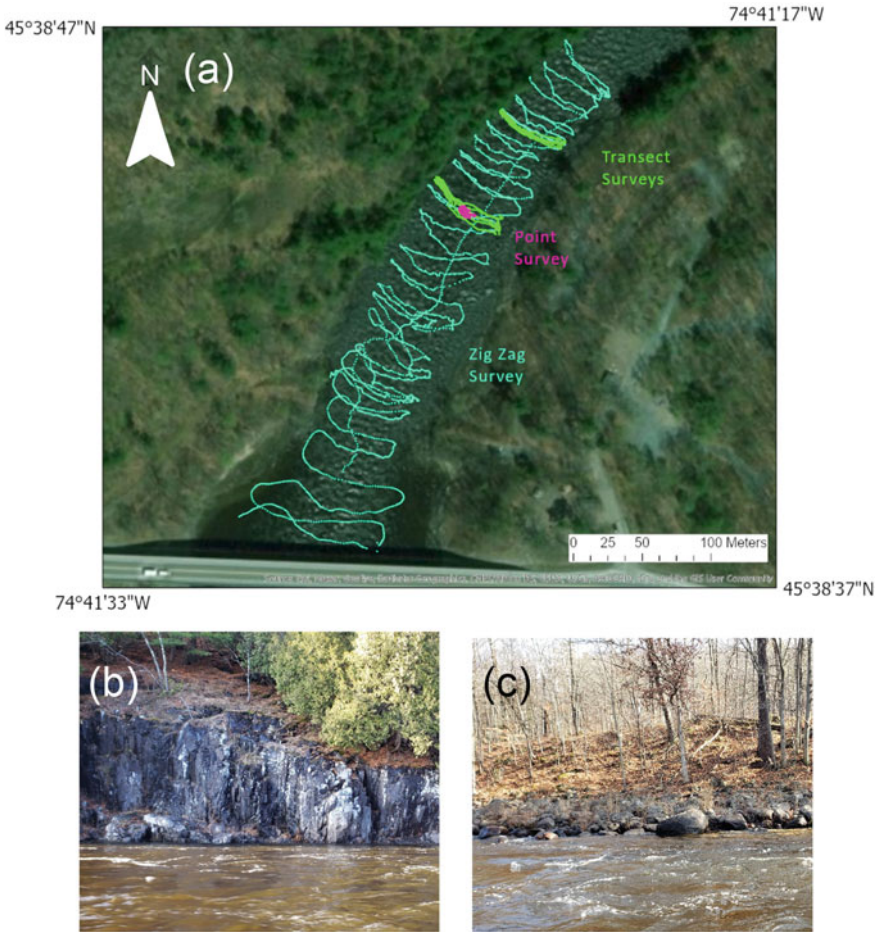
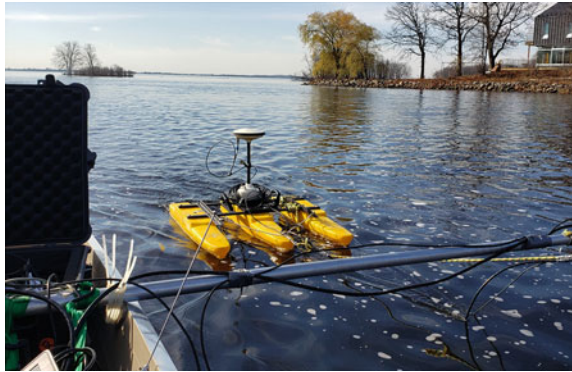


Fig. 1 **a** Location of the study area on Rivière-Rouge and ADCP data tracks with primary zig zag survey, secondary transect surveys, and tertiary point survey shown; **b** northwest bank of study area; **c** east bank of the study area

(15') aluminum boat with a 25 HP motor. The boat held the field crew and computers for data acquisition. The ADCP compass was calibrated in situ.

The principal of operation of the ADCP is that it measures flow velocity using the frequency shift of acoustic pulses emitted from the device transducers, reflected off of particles entrained in the flow, and collected back to the device receiver to determine the flow velocity using the Doppler principle. The sampling frequency was 1 Hz. Depth bins were 0.2 m. To obtain the depth-averaged velocity, all velocity measurements within the water column were averaged. To correct for the boat's velocity, both bottom track on the ADCP and a real-time kinematic differential global positioning system (RTK-DGPS) were used. The reach bathymetry can be estimated

Fig. 2 Experimental setup with ADCP and GPS receiver tethered approximately 1.5 m to the port side of the boat. The photo was taken at the mouth of the Rivière-Rouge into the Ottawa River



either through the ADCP's bottom track measurement or, if the bed is not stationary, differential GPS data can be used to correct for the boat's velocity [23]

Various survey configurations have been used in the past, namely, point and transect surveys. Zig-zag surveys are another type of spatial survey that can be used where data are collected continuously as the mobile instrument vessel travels diagonally from bank to bank upstream or downstream the river reach. The spacing between transects for zig-zag surveys should not be more than 1/5 of the river width [23], with smaller spacing being more desirable, particularly if collected data are to be used for calibration and validation of numerical models [20, 29]. The zig-zag survey was the first conducted, and the track of the survey is shown in Fig. 1a. This initial survey of the entire reach had two purposes: (1) to generate a spatially intensive field of measurement data from which velocity fields and reach bathymetry could be estimated and (2) to determine the approximate location of the highest velocity at which further transect and point surveys could be taken. The boat faced upstream at all times to minimize compass errors at the turning points. The boat could not come within approximately five meters of the East shore because of shallow conditions and the presence of rocks. Two transect surveys were conducted at locations where high flow velocities were observed from the preceding zig-zag survey; each transect survey consisted of six repeat cross-section surveys across the width of the channel. The transect surveys ensure that the locations of highest interest are surveyed with more spatial intensity to reduce inaccuracies and to generate more reliable assessments of secondary velocity components [16, 27]. The locations of the two transects are shown in Fig. 1a. Lastly, point surveys measure velocity at a stationary or semi-stationary position by anchoring or fixing the position of the measurement device or by holding the mobile instrument vessel (e.g., a boat) stationary. The point survey was recorded at the area of highest observed flow velocity which was located near the middle of the downstream transect survey. The position, shown in Fig. 1a, was held approximately stationary, using the propulsion of the motor to hold the boat in place against the flow for 15 min. Point surveys can be used to generate stationary time series of flow depths and water velocity profiles [25]. The entire survey was completed on November 5, 2020.

2.3 Data Processing

Site suitability was determined using depth-averaged flow velocity. The three-dimensional flow velocity field recorded during the survey was converted into a two-dimensional depth-averaged flow field by averaging the easting and northing velocity components throughout the water column at each sample location. To correct for the boat velocity, bottom track was used because the bed was observed to be stationary. Post-processing of the data collected was completed using an in-house MATLAB code [20, 23].

To visually represent the river bathymetry and velocity field to determine the optimal turbine location, interpolation of the data was conducted using empirical Bayesian kriging in ArcGIS Pro 2.4.0. Approximately 3600, 1230, and 640 data points were collected from the zig-zag survey, both transect surveys, and the point survey, respectively. The standard error of prediction for the interpolation was generated.

3 Results

3.1 Flow Depth Mapping

Measured flow depth in the study reach ranged from 0.5 to 4.2 m. Few data points were measured with less than 0.5 m depth due to the risk of damaging the instrument vessel in shallow areas. Interpolation by empirical Bayesian kriging revealed two deeper sections within the reach and a shallow location in the middle of the measured reach, as illustrated in Fig. 3a. The data were interpolated within the measurement locations' edges as to not extrapolate beyond the data's extent. Depths were shallower at the banks of the river except the northwestern area of the reach. This is the location of the river bank with the sheer rock face as shown in Fig. 1b. There is a small shallow area in the center of the reach, approximately mid-way down the length of the reach generated by three data points with depths of 1.05, 0.77, and 1.45 m. Comparatively, the surrounding data points measured a depth of approximately 2.2 m. The field crew observed this to be caused by a large rock.

The standard error of prediction was generated for the interpolation. As expected, the error is minimized in a pattern that follows the measurement data track. Errors are largest at the edge of the data extent where measurements were less spatially intense. These locations tend to coincide with shallow depths, which is favorable for this study as the optimal turbine location within the reach will have a greater water depth.

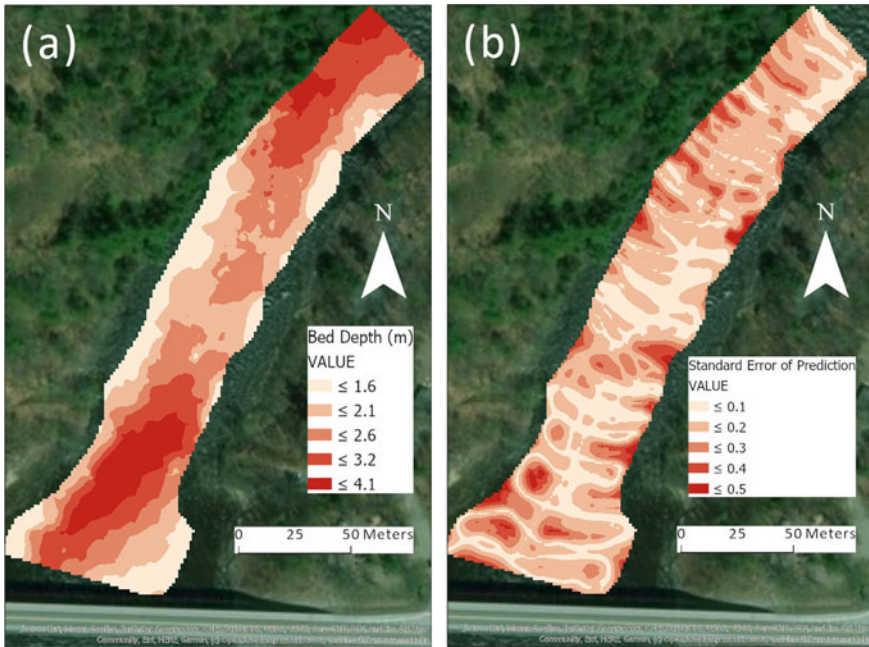


Fig. 3 **a** Bottom track flow depth results using empirical Bayesian kriging to interpolate the data; **b** standard error of prediction of the depth data interpolation

3.2 Depth-Averaged Velocity Distribution

The direction and magnitude of the depth-averaged flow velocity were calculated using the northing and easting velocity components. The magnitude of the depth-averaged velocity was interpolated using empirical Bayesian kriging, and the result is illustrated in Fig. 4a. The velocity vector field was also generated using the interpolation of the velocity direction and magnitude. The depth-averaged velocity in the measured reach ranged from less than 0.3 m/s toward the southwestern bank to greater than 2.3 m/s at locations scattered throughout the center of the reach. Of the 35 datapoints with depth-averaged velocity greater than 2.3 m/s, 24 were located within the point survey. The greatest velocity values were observed in the center of the measured reach where the depth was shallower.

The standard error of prediction of the interpolation for depth-averaged velocity showed smaller error being associated with the track of the data measurements. However, greater error was also found in the middle of the reach where velocities were higher compared to some of the bank locations. Interestingly, a small area of relatively higher error is located directly north of the point survey. This could be due to the abrupt change in measurement intensity from the point survey to a gap in the zig-zag survey.

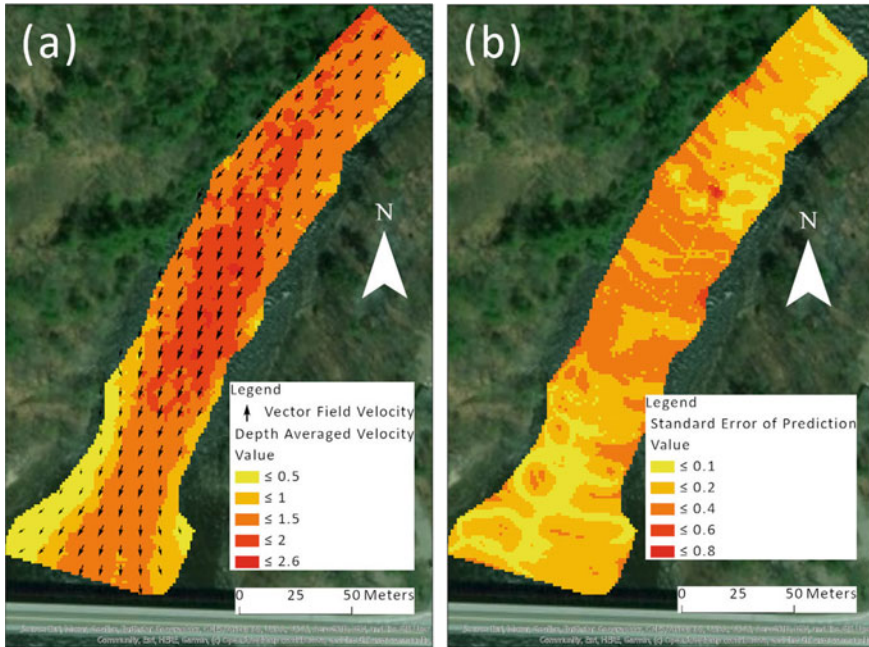


Fig. 4 **a** Depth-averaged velocity and velocity vector field results using empirical Bayesian kriging to interpolate the data; **b** standard error of prediction of the depth-averaged velocity data interpolation

4 Discussion

4.1 Optimal Turbine Placement

To simplify consideration of the optimal hydrokinetic turbine placement, only flow depth and depth-averaged velocity were considered, although many other factors can influence the suitability of a location for turbine deployment such as secondary flow mechanisms, presence of debris or sediment, or ecological, environmental, and economic considerations. The depth will be the first consideration, as the turbine cannot operate in any capacity if the flow depth cannot accommodate the presence of the turbine. Where depth constraints are satisfied, areas with greater flow velocities will contain more available hydrokinetic energy. However, the velocity at a given location must also exceed the cut-in velocity of the turbine. The flow depth required by a hydrokinetic turbine to operate is determined by the design of the turbine. The required depth range varies from 3 to 10 m [22] and greater. Some ultra-low head turbines can operate in flow depths less than 3 m [9]. The study reach is relatively shallow, thus the depth constraint was set at 3 m. An example of a turbine that can operate in this depth range is the Hydroquest 40 kW turbine [8, 22]. Areas with flow depths greater than or equal to 3 m are shown in Fig. 5a. There are only two locations

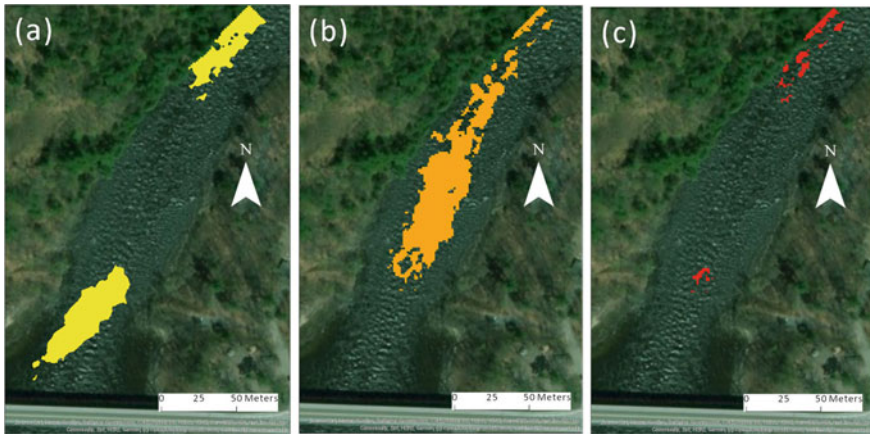


Fig. 5 **a** Depth constraint for turbine deployment set at 3 m with sections deeper than 3 m shown in yellow. **b** Velocity constraint for the turbine deployment set at 1.5 m/s with sections faster than 1.5 m/s shown in orange. **c** Combined constraint of depth greater than 3 m and velocity greater than 1.5 m/s for turbine deployment. Areas shown in red represent areas meeting the constraint criteria

within the reach that meet this criterion: the northwestern bank with the sheer rock face and the southern middle area of the reach.

Experts in the area of riverine hydrokinetic energy extraction were consulted, and a flow velocity of 1.5 m/s is an accepted minimum velocity for consideration of turbine deployment. Below this value, it is more likely that the turbine would not be able to produce a feasible amount of power. Areas with depth-averaged velocities greater than 1.5 m/s are shown in Fig. 5b. There is a fairly well-connected expanse of acceptable velocities spanning from the northern portion of the study reach to the mid-section of the study reach. Acceptable velocities are concentrated near the west bank of the upstream portion of the study reach, and in the mid-channel region of the mid-section of the study reach. The areas of highest velocity are located in the middle of the reach where the depth is comparatively shallower and does not meet the depth constraint. Figure 5c is a visual representation of areas that meet both criteria. Two locations exist for optimal turbine placement: a small section of the southern deep area where velocities are high enough to warrant turbine deployment and the northwestern bank with deep, fast flow next to the sheer rock face.

4.2 Implications for the Future of Hydrokinetic Potential Assessments and Future Research

To the authors' knowledge, there has not been a study yet that has considered optimal placement of a turbine in a reach to this extent. Most studies that determine whether a reach is suitable for energy extraction consider the whole reach with daily average

velocities [4, 18, 19]. Some studies that employ similar field survey techniques have not yet considered depth and velocity constraints within the region of interest as directly as this study [21, 26]. The findings in this study suggest that this spatially intense method for estimating the bathymetry and velocity flow field is beneficial to determine which, if any, locations within a reach are optimal for turbine deployment. As such, this study illustrates the need for spatially intense surveys, rather than assuming one average flow value for a reach, when determining feasibility of a site for hydrokinetic power production. The antagonism between flow depth and velocity constraints cannot be captured in a reach without mapping the bathymetry and the velocity flow field. Additionally, it is important to understand the deployment criterion thoroughly before conducting the site survey. Prior to the survey, the depth constraint was thought to be 2 m. Upon further research, a 3 m constraint was more acceptable for realistic turbine deployment. The point survey conducted in this survey was chosen based on the 2 m constraint. More meaningful data could have been collected at depths greater than 3 m if these criteria were clearly defined, based on research, prior to the survey.

Suggested further research for this study includes estimation of available power, flow variability characterization, uncertainty characterization, and fuzzy logic or multi-criteria decision analysis (MCDA). Estimation of available power throughout the reach should be conducted. Secondary flow mechanisms should be considered in this estimation, as complex, non-linear flows could contribute to lower turbine efficiency. Thus, additional field campaigns should be conducted at the same reach under different flow conditions to understand if and how the variation of flow can influence the change in secondary flow mechanisms, the flow velocity field, and optimal turbine placement. As this study was conducted during an approximate mean flow condition in late fall, two other field campaigns should occur in spring during a high flow condition and in summer during a low flow condition. Locations with high flow velocity and depths greater than 3 m should be measured more heavily to reduce error and uncertainty. As such, uncertainty in the field data should be considered in future studies. Lastly, rather than considering the deployment constraints as true or false, fuzzy logic or multi-criteria decision analysis (MCDA) could be employed in future assessments, which would lead to a refined understanding of suitable locations within the reach.

5 Conclusions

This experiment was the first of its kind to explicitly determine the most optimal location for hydrokinetic turbine deployment in a river reach using ADCP measurements. A survey was conducted on a 190 m long section of the Rivière-Rouge during an approximate mean flow condition to locate areas of optimal turbine deployment. The survey was conducted with an ADCP and a GPS, and the resulting flow depth and depth-averaged velocity were mapped. A depth constraint of minimum 3 m and a velocity constraint of minimum 1.5 m/s were applied and two locations of optimal

turbine placement were identified within the reach: one near the southern (downstream) portion of the study reach and one near the northern (upstream) extent of the study reach. However, the survey extent was limited due to rapids north of the study area precluding safe data collection with available field resources. As such, a turbine deployment may be comparatively more feasible in areas beyond the northern extent of the study reach. This novel method can be applied to any river reach where an ADCP can be deployed and GPS measurements can be taken, but more investigation needs to be done on if and how seasonal flow variability can change the location of optimal turbine placement within a river reach.

Acknowledgements This work was funded in part by the Office of Energy Research and Development (OERD) of Natural Resources Canada and in part by National Research Council Canada.

References

1. Adeogun AG, Ganiyu HO, Ladokun LL, Ibitoye BA (2020) Evaluation of hydrokinetic energy potentials of selected rivers in Kwara State, Nigeria. *Environ Eng Res* 25(3):267–273. <https://doi.org/10.4491/eer.2018.028>
2. Ali F, Srisuwan C, Techato K, Bennui A, Suepa T, Niammuad D (2020) Theoretical hydrokinetic power potential assessment of the U-Tapao river basin using GIS. *Energies* 13(7):1749. <https://doi.org/10.3390/en13071749>
3. Birjandi AH, d'Auteuil S, Ridd C, Bibeau EL (2015) An innovative low cost hydrokinetic site selection technique for cold climate regions. <https://doi.org/10.1109/OCEANS-Genova.2015.7271660>
4. Cruz JDS, Blanco CJC, Brasil Jr ACP (2020) Flow-velocity model for hydrokinetic energy availability assessment in the Amazon. *Acta Sci Technol* 42(1). <https://doi.org/10.4025/actascitechnol.v42i1.45703>
5. d'Auteuil S, Birjandi A, Bibeau E, Jordan S, Soviak J, Friesen D (2019) Riverine hydrokinetic resource assessment using low cost winter imagery. *Renew Sustain Energy Rev* 105:293–300. <https://doi.org/10.1016/j.rser.2019.01.057>
6. Dinehart RL, Burau JR (2005) Averaged indicators of secondary flow in repeated acoustic doppler current profiler crossings of bends. *Water Resour Res* 41(9). <https://doi.org/10.1029/2005WR004050>
7. Holanda PDS, Blanco CJC, Mesquita ALA, Brasil Jr ACP, de Figueiredo NM, Macêdo EN, Secretan Y (2017) Assessment of hydrokinetic energy resources downstream of hydropower plants. *Renew Energy* 101:1203–1214. <https://doi.org/10.1016/j.renene.2016.10.011>
8. HydroQuest (2020) RIVER power. HydroQuest. <https://www.hydroquest.fr/river-power/>
9. HydroQuest (2020) ULH power. HydroQuest. <https://www.hydroquest.fr/ulh-power/>
10. IEC (2019) Marine energy—wave, tidal and other water current converters—part 301: river energy resource assessment. Technical Specification 62600-301, Edition 1.0. International Electrotechnical Commission, Geneva, Switzerland
11. Jamieson EC, Rennie CD, Jacobson RB, Townsend RD (2011) 3-D flow and scour near a submerged wing dike: ADCP measurements on the Missouri river. *Water Resour Res* 47(7). <https://doi.org/10.1029/2010WR010043>
12. Jenkinson W, Bombhof J (2014) Assessment of Canada's hydrokinetic power potential. Phase III report: resource estimation. National Research Council Canada, Ottawa

13. Kalnacs A, Kalnacs J, Mutule A, Persis U (2014) Methods for estimation of the Riverflow potential for hydrokinetic power generation. *Latvian J Phys Tech Sci* 51. <https://doi.org/10.2478/lpts-2014-0008>
14. Kasman RH, Eptanto IL (2019) Study potency of hydrokinetic energy in discharge Balambano hydro electric power plant, vol 2088. <https://doi.org/10.1063/1.5095320>
15. Mohd Saupi AF, Mailah NF, Radzi MAM, Mohamad KB, Ahmad SZ, Soh AC (2018) An illustrated guide to estimation of water velocity in unregulated river for hydrokinetic performance analysis studies in East Malaysia. *Water* 10(10):1330. <https://doi.org/10.3390/w10101330>
16. Moradi G, Vermeulen B, Rennie CD, Cardot R, Lane SN (2019) Evaluation of ADcp processing options for secondary flow identification at River junctions. *Earth Surf Proc Land* 44(14):2903–2921. <https://doi.org/10.1002/esp.4719>
17. Muste M, Kim D, Merwade V (2012) Modern digital instruments and techniques for hydrodynamic and morphologic characterization of river channels. In: *Gravel-bed rivers*. Wiley, pp 315–341. <https://doi.org/10.1002/9781119952497.ch24>
18. Nhabetse T, Cuamba B, Kucel S, Mungoi N (2017) Assessment of hydrokinetic potential in the Umbeluzi Basin, Mozambique, pp 1455–1466. <https://doi.org/10.18086/swc.2017.23.04>
19. Nordino MM (2016) An assessment of hydrokinetic potential in the Maputo Basins, Mozambique. <http://makir.mak.ac.ug/handle/10570/6423>
20. Parsapour-Moghaddam P, Rennie CD (2018) Calibration of a 3D hydrodynamic meandering river model using fully spatially distributed 3D ADCP velocity data. *J Hydraul Eng* 144(4):04018010. [https://doi.org/10.1061/\(ASCE\)HY.1943-7900.0001424](https://doi.org/10.1061/(ASCE)HY.1943-7900.0001424)
21. Petrie J, Diplas P, Gutierrez M, Nam S (2014) Characterizing the mean flow field in rivers for resource and environmental impact assessments of hydrokinetic energy generation sites. *Renew Energy* 69:393–401. <https://doi.org/10.1016/j.renene.2014.03.064>
22. Ranjitkar G (2020) River hydrokinetic energy systems overview. Marine Energy Group, CanmetENERGY, Ottawa
23. Rennie CD, Church M (2010) Mapping spatial distributions and uncertainty of water and sediment flux in a large gravel bed river reach using an acoustic Doppler current profiler. *J Geophys Res Earth Surf* 115(F3). <https://doi.org/10.1029/2009JF001556>
24. Rennie CD, Millar RG (2004) Measurement of the spatial distribution of fluvial bedload transport velocity in both sand and gravel. *Earth Surf Proc Land* 29(10):1173–1193. <https://doi.org/10.1002/esp.1074>
25. Rennie C, Millar R, Church M (2002) Measurement of bed load velocity using an acoustic Doppler current profiler. *J Hydraul Eng ASCE* 128. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2002\)128:5\(473\)](https://doi.org/10.1061/(ASCE)0733-9429(2002)128:5(473))
26. Santos IFSD, Camacho RGR, Tiago Filho GL, Botan ACB, Vinent BA (2019) Energy potential and economic analysis of hydrokinetic turbines implementation in rivers: an approach using numerical predictions (CFD) and experimental data. *Renew Energy* 143:648–662. <https://doi.org/10.1016/j.renene.2019.05.018>
27. Szupiany RN, Amsler ML, Best JL, Parsons DR (2007) Comparison of fixed- and moving-vessel flow measurements with an ADp in a large river. *J Hydraul Eng* 133(12):1299–1309. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2007\)133:12\(1299\)](https://doi.org/10.1061/(ASCE)0733-9429(2007)133:12(1299))
28. Venditti JG, Rennie CD, Bomhof J, Bradley RW, Little M, Church M (2014) Flow in bedrock canyons. *Nature* 513(7519):534–543
29. Williams RD, Brasington J, Hicks M, Measures R, Rennie CD, Vericat D (2013) Hydraulic validation of two-dimensional simulations of braided river flow with spatially continuous ADcp data. *Water Resour Res* 49(9):5183–5205. <https://doi.org/10.1002/wrcr.20391>
30. Williams RD, Rennie CD, Brasington J, Hicks DM, Vericat D (2015) Linking the spatial distribution of bed load transport to morphological change during high-flow events in a shallow braided river. *J Geophys Res Earth Surf* 120(3):604–622. <https://doi.org/10.1002/2014JF003346>

Integrating Activity-Based Modelling with Agent-Based Transit Assignment



Patrick Meredith-Karam, Siva Srikukenthiran, James Vaughan,
Eric J. Miller, and Amer Shalaby

1 Introduction

Accurate modelling of travellers' behavioural choices is imperative for policy-makers to make informed transportation planning decisions. Activity-based travel demand modelling is used to represent travellers and their interactions with transportation infrastructure around them, and to then identify experienced service levels and travel times. One such model, called *GTAModel*, has been developed at the University of Toronto [8]. This open source agent-based model forecasts travel demand in the Greater Toronto-Hamilton Area (GTHA) and is used by municipalities for transportation policy analysis. *GTAModel* currently uses an aggregate transit assignment model, which is inconsistent with the agent-based structure of the rest of the model system.

Agent-based transit assignment models aim to address this gap in travel behavioural modelling by representing individual travellers and allowing them to make travel choices independently. The *Nexus* microsimulation framework performs transit assignment, allowing agents to consider factors including congestion, fare prices, and real-time information in their route choice [12].

This leads to the research question, “*How can agent-based transit assignment models be leveraged to improve the performance of an activity-based travel behaviour model, and how may this improved performance be used to better inform transportation planning decisions?*” To address this challenge, this project aims to meet two primary goals. First, to develop an integration module within XTMF (the *GTAModel* software platform) which can be used to perform *GTAModel* transit

P. Meredith-Karam (✉)
Massachusetts Institute of Technology, Cambridge, USA
e-mail: p.meredith.karam@gmail.com

P. Meredith-Karam · S. Srikukenthiran · J. Vaughan · E. J. Miller · A. Shalaby
University of Toronto, Toronto, Canada

assignment with *Nexus*. Second, to assess the integrated system by evaluating its performance relative to current aggregate assignment methods, and applying it to a case study which highlights its practical use.

This paper first provides a review of literature (Sect. 2) on activity-based travel demand modelling and transit assignment, including both a broad overview of the state of the industry and detailed information on *GTAModel* and *Nexus*. This background work is used to identify a research gap which is addressed by the project. The methods employed in this project are described in Sect. 3, including the overall process of integration between *GTAModel* and *Nexus*, and a case study application of the integrated model. Project results and discussion are provided in Sect. 4, highlighting both the numerical results of the integrated model and the qualitative success of the project. Ultimately, conclusions are discussed in Sect. 5, highlighting the significance of this work in the field of travel demand modelling, and directions for future research which may help overcome limitations of this first attempt at integration.

2 Literature Review

A review of relevant literature was conducted to obtain an understanding of both concepts and specific models used in this project. This review helps to develop an understanding of the gap addressed by the project, namely the integration of microsimulation-based transit assignment procedures into travel demand modelling software. This literature review includes: (1) general overview of travel demand modelling and *GTAModel*; (2) description of transit assignment modelling; (3) limitations of current *GTAModel* transit assignment; and (4) description of *Nexus* as a means of overcoming these limitations.

2.1 Travel Demand Modelling

Travel demand forecasting models are consistently used around the world for governing agencies to plan and make decisions regarding transportation policy and infrastructure. These models are typically developed specifically for a given context, but the overall modelling process is fairly consistent across implementations [8]. The standard approach to travel demand model development includes the following key components: a set of analysis zones which spatially divide the area of interest, estimates of population and employment within each zone, a representation of road and transit network infrastructure, and a connected model which predicts the trips made from every origin zone to every destination zone over the time period of interest, then assigns these trips to specific paths along the transportation network infrastructure [8].

As discussed in detail in [5], the four key components of this standard modelling process include trip generation, trip distribution, mode split, and trip assignment. In this process, “trip generation” estimates volumes of trips which originate and terminate at each zone in the model. “Trip distribution” predicts trip flows between origins and destinations, through the linking of the trip ends estimated by the trip generation stage. Next, mode choice estimates the proportion of trips between each origin and destination which will use each available mode of travel (e.g. transit, auto, walk, etc.). Finally, trip assignment estimates flow volumes along specific routes of travel through the network, such as streets for driving and subway, bus, and streetcar lines for transit. The result of the model is typically a set of predicted flows on each segment of the transportation network, with associated travel times, costs, or other measures of trip-maker level of service.

2.1.1 GTAModel

GTAModel, developed by the Travel Modelling Group (TMG) at the University of Toronto, is an operational activity-based travel demand model used for the Greater Toronto-Hamilton Area (GTHA). The model is used as the standard travel demand model system for the cities of Toronto, Mississauga, Brampton and Vaughan and the Regions of Durham, Halton and Peel.

Data used to forecast demand using *GTAModel* is obtained through the Transportation Tomorrow Survey (TTS), which has been conducted every five years since 1986 [2]. The TTS is performed through both phone and web, targeting a sample of 5% of all households in the area of study. The survey data includes detailed information on trips made by members of the household on a single weekday [2]. The TASHA (Toronto Area Scheduling Model for Household Agents) activity scheduling microsimulation model uses information collected in the TTS to generate activity schedules, attempting to determine which activities are performed by which agents, at what time and location [6]. Trips used in *GTAModel* are generated from these activity schedules.

GTAModel currently uses the “*Emme*” commercial software package [8] to assign trips to specific paths along the road and transit networks, hereafter referred to as the “current aggregate assignment method”. However, the current model structure does allow other road and transit network assignment packages to be used as well, with, for example, Aimsun and MATSim implementations being currently under development. Through the current process, trips are assigned between zones on an aggregate basis, as opposed to being considered for individual agents. This presents an area of potential development on which this project focusses, discussed further in Sect. 2.3.

2.2 *Transit Assignment Overview*

The trip assignment problem is the final step in the general travel demand modelling process, wherein trips are routed along specific paths. In particular, this project focusses on transit assignment, which refers to the process of “determining route choice for transit riders in order to simulate flows over transit networks” [11]. Fu, Liu, and Hess define two main purposes of transit assignment models: “(a) predict the passenger volumes choosing different services connecting any pairs of origin and destination; and (b) validate the effectiveness of operation schemes for transit systems” [3]. This process of transit assignment can thus help transit agencies in their planning of service supply [4]. Furthermore, the underlying function for transit assignment contains three aspects: “characteristics of the supply on transit networks and services; information about the supply that passengers could have before and during journeys; and passengers’ responses towards current situations given related travel information” [3]. Several factors may influence a rider’s choice of transit route, including travel time, route load, transit fares, real-time information provision, service reliability, and mode perception (e.g. view of subways as opposed to buses) [4].

A transit trip includes several steps, each of which develops an associated “cost” or “disutility” in the transit assignment process defined by the specific model used. The total disutility for potential routes is then calculated by summing the costs of each component for every journey (applying appropriate weights), and is used as a means of evaluating alternatives.

2.3 *Identified Gap: Transit Assignment in GTAModel*

This section describes the current limitations of transit assignment in *GTAModel*, and identifies areas which may be improved through replacement with a schedule-based approach using microsimulation methods. The significance of this work to advance the body of knowledge in demand modelling is also highlighted.

In the current *GTAModel* process, trips are modelled as originating or terminating at zone centroids [4], which presents an immediate challenge to the effectiveness of the model. This assumption would make consideration of congestion and crowding effects difficult, since in reality passengers may have access to several different transit stops within a zone, and may attend these stops based on their proximity to an actual trip origin or destination location.

Additionally, the current aggregate transit assignment method in *GTAModel* is limited in the manner in which it considers congestion. It does not enforce hard vehicle capacity constraints in transit assignment [4], but rather uses a congested assignment procedure in which the disutility of a transit trip increases as vehicle occupancy increases, similar to how roadway congestion effects are typically modelled. However, this presents a limited consideration of the overall impacts of crowding,

since congestion on station platforms and in pedestrian walkway channels may also impact travel choices.

The current Emme frequency-based assignment method has also presented challenges in its treatment of low-frequency routes. It does not consider individual vehicle runs, assuming that passengers will not arrive with the intention of boarding a particular vehicle run. This has presented a challenge to the accuracy of *GTAModel* for modelling passenger arrivals and subsequent wait times at low-frequency routes.

A knowledge of the entire trip history is also currently unavailable, and would be necessary to correctly determine the total transit fare in consideration of the various transfer-fare discounts provided between the various transit agencies in the GTHA. This challenge has been identified as a significant area of interest for *GTAModel* developers.

While agent-based simulation has become crucial for small-scale analysis of corridors and project study areas, this study contributes to the broader body of knowledge through its scale of implementation. The majority of city-level travel demand models continue to use aggregate assignment processes, which to some extent limits the efficacy of agent-based techniques which are increasingly used in the trip generation and mode split portions of these models. By integrating agent-based microsimulation for transit assignment on the scale of a region-wide activity-based demand model. This study demonstrates an opportunity to push the general state of practice further ahead.

2.4 Chosen Transit Assignment Model: Nexus Microsimulation Framework

To fill the gap in current modelling capabilities identified in Sect. 2.3, this project proposes to use the *Nexus* microsimulation framework to perform transit assignment in *GTAModel*. *Nexus* uses agent-based microsimulation, to consider the travellers' decisions on an individual basis, and may enable *GTAModel* to improve upon many of the limitations presented by traditional transit assignment modelling frameworks.

The *Nexus* microsimulation framework enables detailed simulation of agents in a transit network, to allow modelling with greater functionality than aggregate-level simulation. *Nexus* enables consideration of the impacts of congestion both in transit vehicles and on station platforms and other areas of bottleneck pedestrian flow through its interface with MassMotion, a commercial pedestrian microsimulation software [12]. The impacts of crowding are of considerable interest to transit agencies, particularly at times when a disruption in transit service causes abnormally large flows of pedestrians in enclosed areas such as station platforms [12]. The *Nexus* framework presents an opportunity to greatly improve the performance of the transit network with respect to crowding, through its innovative linked analysis of crowd flow and system service, which have traditionally been considered separately [12]. As compared with other micro-simulation software such as Aimsun or VISSIM which

have traditionally focused strongly on vehicle traffic simulation, *Nexus* enables a focus on transit microsimulation [12].

3 Methods

The study is implemented within the *GTAModel Version 4* activity-based model framework. A high-level view of the model process is shown in Fig. 1 (left), while a more detailed schematic is provided in Appendix. *GTAModel* is a fully disaggregated agent-based microsimulation model system in which all out of home activities and associated trips for a typical weekday are simulated for a complete set of synthesized persons and households that are resident within the study region for the forecast year. The overall process of *GTAModel* involves first probabilistically assigning: work and school locations to all workers and students (based on doubly-constrained entropy/gravity spatial interaction models), individual driver’s licenses to persons 16 years or older (binary logit model) and household vehicle ownership level (ordered logit model). Individual daily out-of-home activity/travel schedules are then simulated using the TASHA (Travel/Activity Scheduler for Household Agents) model, which includes accounting for household-level activities and constraints [6, 10]. Given these person schedules, trip modes are determined using a tour- and household-based multinomial probit random utility model, including resolution of conflicts in household drivers’ demands for the household’s cars and explicitly modelling within-household ridesharing [7, 9]. Trips are extracted from the tours generated by TASHA and aggregated into origin–destination (O-D) trip matrices by time period and assigned to the auto network using Emme’s deterministic static

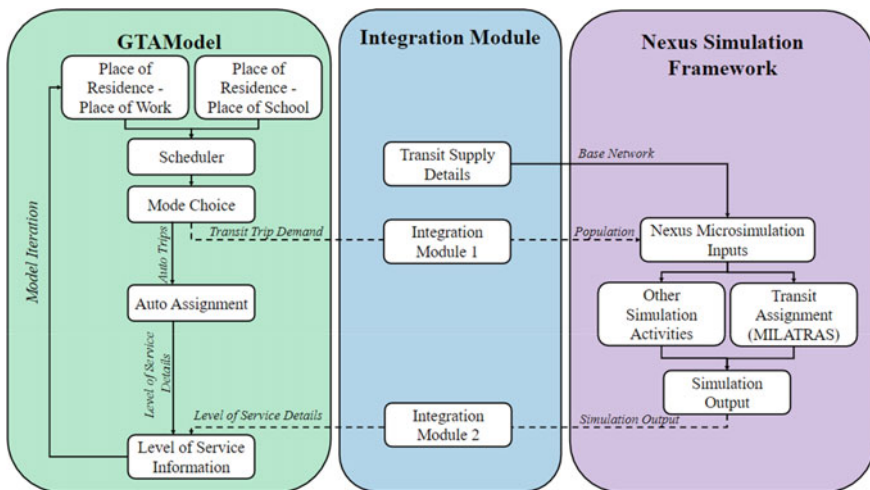


Fig. 1 GTAModel-Nexus integration process diagram

user equilibrium procedure, and to the transit network using an extended version of Emme's aggregate assignment package that includes vehicle average crowding effects and uses a generalized path utility function that includes both time and fare factors. For a more thorough technical overview of *GTAModel*, please see [15]. *GTAModel V4* is implemented within a custom software environment developed by TMG called XTMF (eXtensible Travel Modelling Framework), a highly modularized and optimized software environment within which to develop and use advanced travel demand model systems, and which includes a robust user interface for model system customization and control of model system runs [16].

Thus, the very disaggregated, agent-based representation of travel generated by TASHA is "lost" through the aggregation of individual person-tours into aggregate O-D trip flows for input into the aggregate assignment procedures. The purpose of this study is to investigate an alternative solution to the current aggregate transit network assignment process in *GTAModel*, using an agent-based simulation procedure with *Nexus*.

The Integration process between *GTAModel* and *Nexus* is illustrated in Fig. 1, with components developed in this project listed under the "Integration Module" section. Of particular importance in the integration process is the information which is passed from *GTAModel* to *Nexus*, and vice versa. Upon completion of the "Mode Choice" component of the *GTAModel* process, "Transit Trip Demand" is passed from *GTAModel*. This information is extracted from *GTAModel* and passed to *Nexus* in the form of a "Population" for simulation, through the first integration module. In parallel to this process, "Transit Supply Details" are provided to *Nexus* in the form of specifically-formatted GTFS files which define the routes, schedules, and other service characteristics of the transit network. Once the *Nexus* simulation process is complete, the results are extracted by the second integration model and passed back to *GTAModel* in the form of "Level of Service Details". This includes travel time, total fare cost, wait time, walk time, and perceived travel time.

3.1 Case Study: Toronto's Downtown Relief Line

As described in Sect. 2.4, one of the primary advantages of the *Nexus* microsimulation platform is its ability to model pedestrian congestion on station platforms and along walkways in the transit network. In order to showcase a practical application for this benefit, the integrated model is applied to a case study of Toronto's proposed Downtown Relief Line (DRL, currently "rebranded" in somewhat modified form as the "Ontario Line"). The line primarily aims to alleviate congestion at the Yonge-Bloor subway station, which, as of 2015 serves an estimated 216,200 passengers per day [14]. The DRL is expected to decrease morning peak period transfers from Line 2 Westbound to Line 1 Southbound from a projected 10,000 in 2031 to only 7300 [1]. Furthermore, the line is projected to attract 13,400 new riders to the TTC, and relieve congestion on the crowded King Streetcar by 74% and the Queen Streetcar by 69% [1]. To evaluate whether these benefits are likely to be realized, detailed modelling

of pedestrian crowding will be required. This case study provides an opportunity to showcase a real-world case where *Nexus* may enable improved analysis of the impacts of a transit infrastructure change on experienced rider level of service.

For application to this project, the case study requires a change to the “Transit Supply Details” input of the integrated model, but may otherwise be run as normal due to the robust nature of the model. DRL-specific GTFS files were obtained from a previous DRL modelling project conducted by [13]. These files included a complete TTC network with the DRL, which were simply used to replace the TTC network from the previous base network files. Second, the *GTAModel-Nexus* integrated model could simply be applied to the updated base network to consider the DRL line. This was a simple manner of adjusting the parameters of the model within XTMF through the program’s graphical user interface. Generally, the ease of application of the integrated model to this case study serves to demonstrate the model’s robust nature, and its potential for use on other transit infrastructure decisions.

4 Results and Discussion

The integrated model described in Sect. 3 was successfully developed as an XTMF module in C#, consisting of the two main components described. The main takeaway from this process is that the integration process was demonstrated to be feasible. The method proposed was successful, and satisfied the project objective of obtaining trips from *GTAModel* and performing transit assignment using agent-based microsimulation in *Nexus*. This being said, the current state of operations is subject to some limitations, which are identified and considered for future development in Sect. 5.3.

Ultimately, the integration fulfils the original project goal of successfully linking *GTAModel* with *Nexus*. This process shows positive results for this initial study, and highlights the opportunity for further pursuit of agent-based microsimulation to perform transit assignment for activity-based models. Although the program developed for this project is specific to *GTAModel* and *Nexus*, the process may be broadly applicable to other systems on a conceptual level.

4.1 Numerical Results Analysis

Numerical results were obtained from completed integrated model runs for a single iteration and time period, with a complete set of agents travelling to and from Toronto’s downtown core. Trips were restricted to only include trips to and from the downtown core so that runtimes would remain reasonable for testing and troubleshooting purposes. The key input parameters of the run are as follows: a population of 389,964 agents was simulated, over the morning peak period from 7:00 to 9:00am. A single iteration was performed, considering trips between all zones (2375 total)

and a set of 32 downtown zones. Level of service matrices were generated to a level of granularity of 1 h.

4.1.1 Comparison of Nexus and Current Model Results

From this model run, numerical results for key level of service metrics (the output of the *Nexus* simulation) were obtained for all combinations of origin–destination zones both to and from the downtown core. Although this output is far too large to show in this report, aggregate values and comparisons are provided in Table 1. Note that two travel directions are provided, including “ALL-DT” and “DT-ALL”. The first is reflective of a model run with all zones (2375 total) as the origin zones and downtown-only zones (32 total) as the destination zones, and the second is the opposite case. “*Base*” refers to a base run of *GTAModel* using the current aggregate transit assignment process, while “*Nexus*” refers to a run of *GTAModel* using *Nexus* microsimulation for transit assignment.

The run output provided demonstrates that the integrated model yields intelligible results which appear practical in the real world. This is an extremely positive indication, especially considering the limitations of using only one iteration. However, there seem to be clear and systematic differences between travel times, walking times, and

Table 1 Output results for *Nexus* model run and model run with current process

| Travel direction | Metric | Cost (\$) | Perceived time (min) | Travel time (min) | Walk time (min) | Transfer time (min) |
|------------------|--------------------------|-----------|----------------------|-------------------|-----------------|---------------------|
| ALL-DT | Average (<i>Base</i>) | 5.23 | 154.95 | 58.61 | 25.12 | 7.13 |
| | Average (<i>Nexus</i>) | 9.55 | 127.32 | 103.72 | 0.85 | 66.45 |
| | Average difference | 4.33 | −27.63 | 45.11 | −24.28 | 59.33 |
| | Variance of Difference | 20.06 | 1772.82 | 1044.29 | 193.06 | 1336.23 |
| | Average % difference | 108% | −16% | 128% | −96% | 1029% |
| DT-ALL | Average (<i>Base</i>) | 4.65 | 155.13 | 57.90 | 23.49 | 7.78 |
| | Average (<i>Nexus</i>) | 8.97 | 129.32 | 106.79 | 0.88 | 85.43 |
| | Average difference | 4.32 | −25.81 | 48.90 | −22.61 | 77.65 |
| | Variance of difference | 17.15 | 2981.72 | 1311.94 | 406.46 | 1987.78 |
| | Average % difference | 109% | −13% | 144% | −96% | 1289% |

transfer times predicted by the current assignment and the *Nexus* assignment. First, discrepancies between travel time are likely because *Nexus* measures door-to-door total travel time, while the current process reports the in-vehicle travel time only. Low walk times and high transfer times in *Nexus* may result from the use of zone centroids as origin and destination points, or the inclusion of walking times as part of the transfer time. This highlights an area which will need to be calibrated in a full integration, to reconcile precise discrepancies in model reporting. The limitations of these results for practical use are therefore clear. To obtain practically usable results and truly evaluate the effectiveness of the integrated model, it would be essential to develop a more nuanced understanding of the spatial correlations between results obtained with *Nexus* and the current process, and to scale the model up to perform a full set of four iterations in *GTAModel*.

Furthermore, the high variance of the differences between *Nexus* and the current process in perceived travel time, travel time, and transfer time indicate that the differences in these variables are spatially dependent. It is possible that this difference lies primarily in external zones with low transit mode share (and very high transit times), which may be skewing the aggregated data. In an attempt to better understand these spatial relationships, further study of trips within a restricted downtown geographical zone is performed. These results indicate that the values between *Nexus* and base model runs are much closer together along each of the level of service measures identified. The magnitude for all “Average Difference” values is significantly lower for downtown-downtown trips than when all zones are used for either origin or destination, when the previously mentioned reporting differences are considered. This is particularly relevant because transit mode shares tend to be higher in downtown areas, and lower in areas with more limited transit infrastructure and thereby longer travel times.

4.1.2 Case Study Results Analysis

The integrated base model was also applied to the Downtown Relief Line case study, to showcase the ability of *Nexus* to predict the impact of a change in transit infrastructure. Similar to the model results discussed in Sect. 4.1.1, model runs were performed for both trips made to the downtown core from all zones, and trips made to all zones from the downtown core. Two model runs were conducted using the integrated model, one which included all current transit lines in the GTHA (indicated “*Current*”), and one which included all of these lines in addition to the proposed Downtown Relief Line (indicated “*DRL*”). Aggregated results for these two model runs are provided in Table 2.

As anticipated, it was found that the average perceived travel time and actual travel time decreased marginally between the base network and DRL model runs, thereby improving passenger level of service. However, additional considerations may need to be taken to fully measure the impact of the DRL. First, the accuracy of transfer times and travel times may be further improved by addressing the model limitations (described in Sect. 5.3), and by adding a MassMotion simulation run to

Table 2 Comparison of downtown relief line model run output with integrated current model run

| Travel direction | Metric | Cost (\$) | Perceived time (min) | Travel time (min) | Walk time (min) | Transfer time (min) |
|------------------|----------------------------|-----------|----------------------|-------------------|-----------------|---------------------|
| ALL-DT | Average (<i>Current</i>) | 9.42 | 125.46 | 102.25 | 0.87 | 65.16 |
| | Average (<i>DRL</i>) | 9.37 | 125.35 | 102.14 | 0.84 | 65.53 |
| | Average difference | -0.05 | -0.10 | -0.11 | -0.02 | 0.37 |
| | Variance of difference | 9.38 | 741.64 | 601.16 | 0.01 | 495.93 |
| | Average % difference | 9% | 4% | 4% | -2% | 12% |
| DT-ALL | Average (<i>Current</i>) | 8.79 | 127.41 | 105.43 | 0.87 | 84.26 |
| | Average (<i>DRL</i>) | 8.83 | 126.23 | 104.24 | 0.85 | 83.68 |
| | Average difference | 0.04 | -1.18 | -1.19 | -0.03 | -0.58 |
| | Variance of difference | 6.43 | 855.16 | 711.02 | 0.01 | 678.16 |
| | Average % difference | 7% | 2% | 2% | -3% | 7% |

the *Nexus*-based transit assignment. In the case of Bloor-Yonge station, integration with pedestrian simulation would enable more realistic transfer times to be obtained. This expansion is feasible and would require minimal development, as pedestrian simulation has been used with *Nexus* for several Toronto-based simulations of transit, such as one project by [13].

Overall, the use of the integrated model for the DRL showcases the ability of the *Nexus* microsimulation platform to comprehensively evaluate the impacts of pedestrian congestion in station platforms and walkways. This may enable consideration of a greater breadth of factors in the analysis of proposed transportation infrastructure changes, thereby better-informing policy decisions.

5 Conclusion

In summary, this project presented the successful development of an integrated activity-based model system using agent-based microsimulation for transit assignment. After performing background research on relevant topics and models, a gap was identified in the current aggregate transit assignment process used in *GTAModel*.

This gap was addressed through the integration of the *Nexus* microsimulation platform with *GTAModel*, and a robust method was created for this integration. Model results were evaluated both in comparison with the current model and in application to a case study of Toronto's Downtown Relief Line.

Ultimately, this project successfully addressed the original research question. The first component was answered through the development of a method for integrating an activity-based model with an agent-based transit assignment model, and evaluation of its performance relative to current methods. Additionally, the opportunity to improve analysis and thereby better-inform transportation planning decisions was highlighted through the application of the model to a case study. The key takeaway from the project is that the *Nexus* platform was demonstrated to be viable for integration to *GTAModel*, and presents one potential opportunity to address the limitations of current transit assignment methods.

5.1 Significance of the Research Performed

This research is significant to both the field of travel demand modelling and real-world transportation planning. Activity-based models enable comprehensive agent-based analysis of travel patterns, but their value may be limited when aggregate assignment processes are employed. By demonstrating the ability to use agent-based microsimulation for transit assignment, this project highlights a key development which could enable the future progression of a truly agent-based end-to-end travel demand model.

Furthermore, this research presents an opportunity to develop more useful results for application to real-world transportation policy and infrastructure decisions. Generally, agent-based models have been found to yield more comprehensive results than aggregate models. The *Nexus* microsimulation platform, in particular, provides a unique ability to coordinate several different models for transit assignment, thereby allowing for consideration of many key factors. For instance, *Nexus* allows detailed simulation of pedestrian congestion in subway stations and on platforms, which can aid in the evaluation of transit infrastructure options which are designed to reduce these areas of congestion. Therefore, the integrated model developed in this project may pave the way for more informed models which better capture the implications of policy and infrastructure changes, enabling more evidence-based policy decisions.

5.2 Project Lessons Learned

Through consideration of the experience in this project, one key lesson was illuminated: "*Network simulation models developed independently of overall travel demand models may require significant work to integrate with a comprehensive, full-scale demand model system*". This lesson stems from experience throughout

the project, primarily due to the level of custom development of the *Nexus* system which was required to integrate with *GTAModel*. The original *Nexus* system was not capable of producing full zone-to-zone level of service matrices, and this function had to be implemented by *Nexus* developers as part of this project. For reference, a full morning peak volume of transit trips is approximately 629,000 in Toronto, while the number of zone-to-zone combinations in *GTAModel* is 5,640,625. This change caused a significant increase on the computational requirements of the *Nexus* system, necessitating efficiency improvements for operation at this greater scale.

This lesson is broadly applicable to travel demand modelling. In this project, it was fortunate that the microsimulation model used for transit assignment was developed at the University of Toronto. Since *Nexus* developers were available to perform necessary development to enable full level of service matrix outputs, the project was feasible to complete despite its many difficulties. However, in a project which seeks to integrate an activity-based model with commercial off-the-shelf microsimulation software, these challenges may be significantly harder. Custom development may be quite limited for commercial software, thus restricting the ability to tailor the software to produce the outputs required for an activity-based model.

5.3 Limitations and Directions for Future Research

Upon examination of the results of the integrated model, four key limitations of the current integrated model are identified, including centroid to centroid travel, single iteration of *GTAModel*, lack of integration with auto assignment, and runtime considerations. Each is discussed in more detail below. Future research could focus on developing solutions to these areas, to progress the prototype-level integrated model developed in this project to a fully functional version of *GTAModel* which may be used in practice.

The assumption of centroid to centroid travel was made in the development of the integrated model, as the origin and destination points of trips extracted from *GTAModel* were provided as zone centroids instead of individual household locations. This could have limited the results by artificially creating congestion and thereby increasing travel times in some popular transit stations near zone centroids. This issue may be remedied through a simple extension to the project, where trip origins and destinations are spatially distributed across zones, either uniformly (as a simplifying assumption) or according to the density of residential and employment location across particular zones. This is a standard process which has been completed by both developers of *GTAModel* and developers of *Nexus* in the past, and could be readily implemented without significant development effort.

One of the most important challenges encountered in this project is the required runtime for the *Nexus* microsimulation process. The model required approximately 24 h to complete under consideration of a full daily population in a restricted area (approximately 390,000 transit trips). Two significant bottleneck processes are present in the *Nexus* portion of the integrated model, namely path set generation

(approximately 24 h) and level of service estimation (approximately 6 h per run). The path finding process could potentially be performed only once, and have results cached for several runs of the integrated model. However, the level of service output process will be necessary for every iteration and time period. This presents a major bottleneck, which can only feasibly be solved through additional development of the *Nexus* system. Feasible improvements have been identified in consultation with *Nexus* system developers, which may be made to improve the efficiency of the level of service export process. Specifically, the model was found to produce many zone-to-zone trips which would not be feasible to complete with transit. In place of providing specific but impractically high travel time measures, it could be useful to identify the origin–destination pairs that would produce this result, and assign some default “very large” value to their level of service measures in place of performing unnecessary calculations.

One iteration of *GTAModel* is performed in this integrated model process, while four iterations are typically performed in a full *GTAModel* run. When a full run is attempted, it will be important to leverage programing efficiencies to maintain a reasonable model runtime. Specifically, this could be achieved by restricting the number of times which *Nexus* performs path-finding operations, which can currently take over 24 h to complete. The path-finding process includes iterating over each trip, and identifying a set of the seven most feasible routes from the origin to the destination. The results of this process may be cached and stored for use in future iterations, which would significantly reduce the marginal runtime increase when scaling to a full integration. Additionally, as the model progresses from one iteration to the next, it is possible that the fastest path through the network may not be in within the original path set. This would be due to the re-updating of surface transit route speeds through iterations. To determine the appropriate frequency of path set updating, the likelihood of the fastest path not being in the original set would need to be investigated, in conjunction with the number of paths stored for each trip.

Finally, while this project focussed on the transit assignment modelling process, it is important to note that the *Nexus* microsimulation platform does not currently have the capability of assigning automobile trips. Therefore, in order to complete a full run of *GTAModel* using *Nexus*, it would still be necessary to use the current process of auto assignment. This would increase the overall runtime requirement of the integrated model, and would limit the opportunity to entirely replace current trip assignment methods in *GTAModel*. However, it is possible that *Nexus* may develop the capability to perform auto assignment in the future. If this development proceeds, auto assignment could be integrated with *Nexus* in a similar process to that performed in this project.

Appendix: GTAModel V4.0 Overview

Figure 2 provides a high-level overview of the GTAModel activity-based modelling system, to expand upon the summary schematic provided in Sect. 3. Note that “TASHA” refers to the activity-scheduling component of the model.

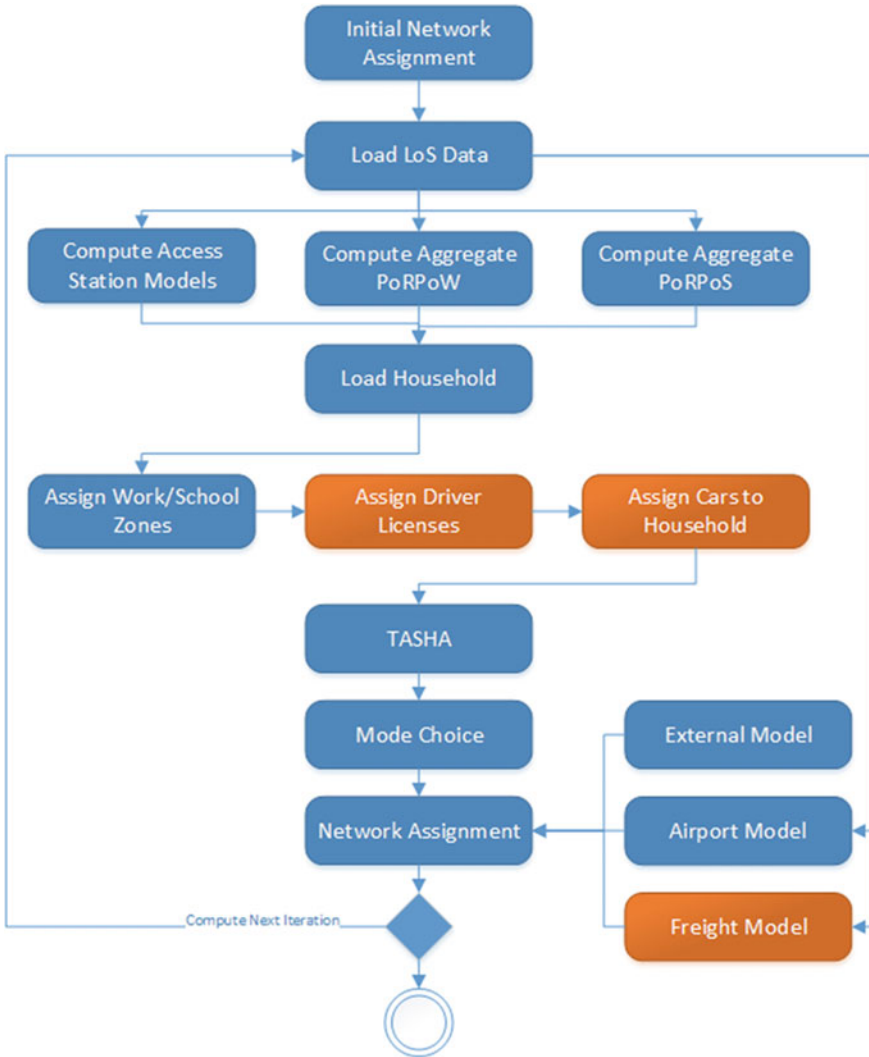


Fig. 2 Overview of the GTAModel process [15].

References

1. City of Toronto (2016) Relief line subway: initial business case. City of Toronto, Toronto
2. Data Management Group (2018) Design and conduct of the survey. Transportation Tomorrow Survey 2016, Toronto
3. Fu Q, Liu R, Hess S (2012) A review on transit assignment modelling approaches to congested networks: a new perspective. *Proc Soc Behav Sci* 54
4. Kucirek P (2012) Comparison between MATSim and EMME: developing a dynamic, activity-based microsimulation transit assignment model for Toronto. Masters thesis. University of Toronto, Toronto
5. Meyer MD, Miller EJ (2003) *Transportation planning: a decision-oriented approach*, 3rd edn. McGraw Hill, New York
6. Miller EJ, Roorda MJ (2003) Prototype model of household activity-travel scheduling. *Transp Res Rec* 1831:114–121
7. Miller EJ, Roorda MJ, Carrasco JA (2005) A tour-based model of travel mode choice. *Transportation* 32(4):399–422
8. Miller EJ, Vaughan J, Nasterska M (2016) Smarttrack ridership analysis: project final report. University of Toronto, Toronto
9. Roorda MJ, Passmore D, Miller EJ (2009) Including minor modes of transport in a tour-based mode choice model with household interactions. *ASCE J Transp Eng* 135(12):935–945
10. Roorda MJ, Miller EJ, Habib KMN (2008) Validation of TASHA: a 24-hour activity scheduling microsimulation model. *Transp Res A* 42:360–375
11. Spiess H, Florian M (1989) Optimal strategies: a new assignment model for transit networks. *Transp Res Part B Methodol* 23:83–102
12. Srikukenthiran S, Shalaby A (2017) Enabling large-scale transit microsimulation for disruption response support using the Nexus platform. *Public Transport*
13. Talker J, Srikukenthiran S, Morrow E, Shalaby A (2017) Potential effects of the downtown relief line on Toronto's subway network: capacity and performance analysis using integrated rail and crowd simulation in Nexus. University of Toronto, Toronto
14. Toronto Transit Commission (2015) *Subway ridership, 2015*. TTC, Toronto
15. Travel Modelling Group (2021) GTAModel V4 documentation. <http://tmg.utoronto.ca/doc/1.6/gtamodel/index.html>
16. Travel Modelling Group (2021) XTMF documentation. <http://tmg.utoronto.ca/doc/1.6/xtmf/index.html>