

Gi-Chul Yang · Sio-long Ao
Xu Huang · Oscar Castillo *Editors*

IAENG Transactions on Engineering Technologies

Special Issue of the International
MultiConference of Engineers
and Computer Scientists 2012

Lecture Notes in Electrical Engineering

Volume 186

For further volumes:
<http://www.springer.com/series/7818>

Gi-Chul Yang · Sio-Iong Ao
Xu Huang · Oscar Castillo
Editors

IAENG Transactions on Engineering Technologies

Special Issue of the International
MultiConference of Engineers
and Computer Scientists 2012

 Springer

Editors

Gi-Chul Yang
Department of Multimedia Engineering
Mokpo National University
Chonnam
Republic of Korea

Sio-Iong Ao
Unit 1, 1/F
International Association of Engineers
Hong Kong
Hong Kong SAR

Xu Huang
Faculty of Information Sciences
and Engineering
University of Canberra
Bruce, ACT
Australia

Oscar Castillo
Tijuana Institute of Technology
Calzada Tecnológico s/n
Tijuana
Mexico

ISSN 1876-1100

ISBN 978-94-007-5623-6

DOI 10.1007/978-94-007-5651-9

Springer Dordrecht Heidelberg New York London

ISSN 1876-1119 (electronic)

ISBN 978-94-007-5651-9 (eBook)

Library of Congress Control Number: 2012946626

© Springer Science+Business Media Dordrecht 2013

This work is subject to copyright. All rights are reserved 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. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

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.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

A large international conference on Advances in Engineering Technologies and Physical Science was held in Hong Kong, 14–16 March 2012, under the International MultiConference of Engineers and Computer Scientists 2012 (IMECS 2012). The IMECS 2012 is organized by the International Association of Engineers (IAENG). IAENG is a non-profit international association for the engineers and the computer scientists, which was found originally in 1968 and has been undergoing rapid expansions in recent few years. The IMECS congress serves as good platforms for the engineering community to meet with each other and to exchange ideas. The congress has also struck a balance between theoretical and application development. The conference committees have been formed with over three hundred committee members who are mainly research center heads, faculty deans, department heads, professors, and research scientists from over 30 countries with the full committee list available at our congress web site (<http://www.iaeng.org/IMECS2012/committee.html>). The congress is truly international meeting with a high level of participation from many countries. The response that we have received for the congress is excellent. There have been more than 800 manuscript submissions for the IMECS 2012. All submitted papers have gone through the peer review process and the overall acceptance rate is 54.6 %.

This volume contains 26 revised and extended research articles written by prominent researchers participating in the conference. Topics covered include scientific computing, engineering mathematics, engineering physics, control theory, automation, industrial engineering, and industrial applications. The book offers the state-of-the-art of tremendous advances in engineering technologies and physical science and applications, and also serves as an excellent reference work for researchers and graduate students working with/on engineering technologies and physical science and applications.

Gi-Chul Yang
Sio-Iong Ao
Xu Huang
Oscar Castillo

Contents

On Turbulence Transition in Three-Dimensional Boundary Layer Flow	1
J. C. Chen and Weijia Chen	
Efficient Algorithms for the Green’s Function Formalism	17
Jan Jacob, Bodo Krause-Kyora, Lothar Wenzel, Qing Ruan, Darren Schmidt, Vivek Amin and Jairo Sinova	
Performance Evaluation of Check-By-Voting for Colluding Attack in Volunteer Computing Systems	33
Kan Watanabe, Masaru Fukushi, Nobuo Funabiki and Toru Nakanishi	
Fast Slot Planning Using Constraint-Based Local Search	49
Dario Pacino and Rune Møller Jensen	
Sequencing of Arrival Aircraft with Operational Constraints	65
Adriana Andreeva-Mori, Shinji Suzuki and Eri Itoh	
Synchronized Triomineering on Rectangular Boards	81
Alessandro Cincotti	
Robust Portfolio Selection Model with Random Fuzzy Returns Based on Arbitrage Pricing Theory and Fuzzy Reasoning Method . . .	91
Takashi Hasuike, Hideki Katagiri and Hiroshi Tsuda	
Hierarchical Multiobjective Stochastic Linear Programming Problems Through a Fractile Optimization Model Using Reference Membership Intervals	105
Hitoshi Yano	

Subsidy, Revenue and Tragedies in Heterogeneous Communities	121
Mbuyu Sumbwanyambe and Andre L. Nel	
A Game Theoretical Analysis for the Quantity Discount Problem with Weibull Ameliorating Items	137
Hidefumi Kawakatsu, Toshimichi Homma and Kiyoshi Sawada	
Simulation Study of an Integrated Reverse Logistics in Fuzzy Environment	151
Debabrata Das and Pankaj Dutta	
A Hybrid-Heuristics Algorithm for k-Minimum Spanning Tree Problems.	167
Hideki Katagiri and Qingqiang Guo	
An Optimal Model of Adding Relation Between the Top and a Member of a Linking Pin Organization Structure with K Subordinates.	181
Kiyoshi Sawada, Hidefumi Kawakatsu and Takashi Mitsuishi	
An Agri-food Supply Chain Model to Empower Farmers for Supplying Deteriorated Product to Modern Retailer	189
Wahyudi Sutopo, Muh. Hisjam and Yuniaristanto	
Fuzzy Logic Approach to Inventory Lot-Sizing Problem Under Uncertain Environment.	203
Busaba Phruksarphanrat and Thipbodee Tanthatemee	
A Semantic Wiki to Support Knowledge Sharing in Innovation Activities	217
Inaya Lahoud, Davy Monticolo, Vincent Hilaire and Samuel Gomes	
Overall Equipment Effectiveness Measurement: Weighted Approach Method and Fuzzy Expert System.	231
M. Maran, G. Manikandan and K. Thiagarajan	
A Supplier-Manufacturer Model for Procurement Plan in Export Oriented Furniture Industry with Sustainability Considerations	247
Muh. Hisjam, Achmad Habibie, Wahyudi Sutopo and Kuncoro Harto Widodo	
Statistical Analysis of Friction Stir Weld Data	261
Esther T. Akinlabi and Stephen A. Akinlabi	

**Development and Evaluation of an Online Casebook
for Teaching and Learning of Industrial Accidents.** 275
Ke Chen and Alan Hoi Shou Chan

**Comparison of Response to Visual and Auditory
Digit Cues in Man–Machine System.** 289
Annie W. Y. Ng and Alan H. S. Chan

**Gender Differences of Human Response Under
Vibration Condition** 301
Hon Keung Yau

**Extended Kalman Filter (EKF) with Sensitive Equation
for the PEMFC Internal State Estimation** 309
Jichen Liu, Guangji Ji and Su Zhou

**Global Localization and Position Tracking
of Autonomous Transport Vehicles** 325
Christof Röhrig, Christopher Kirsch, Julian Lategahn
and Marcel Müller

**An Intelligent Flow Measurement Technique
by Venturi Flowmeter Using Optimized ANN.** 341
Santhosh Krishnan Venkata and Binoy Krishna Roy

Efficient Video Streaming Over Wireless Mesh Networks 353
Kwok Tung Sze, King Man Ho and Kwok Tung Lo

Subject Index 367

Author Index 383

On Turbulence Transition in Three-Dimensional Boundary Layer Flow

J. C. Chen and Weijia Chen

Abstract Sub-harmonic resonance in zero pressure gradient three-dimensional boundary layer flow occurs in the classical N-type pathway of turbulence transition. Three-dimensionality incurs exorbitant computational demands on the numerical simulations. Imposition of a spectral method and a non-uniform grid countervails the impractical computational demands. Validation of the numerical method versus the three-dimensional OS equation avers confidence in the accuracy of the model. Numerical realizations of the generation, amplification, and interaction of two- and three-dimensional sub-harmonic waves agree qualitatively with classical experiments.

Keywords Boundary layer flow • Combined compact difference method • Non-uniform grid • Spectral method • Sub-harmonic resonance • Three dimensionality • Turbulence transition

1 Introduction

The classical work, Schubauer and Skramstad 1947 [1], conducts a famous experiment to examine the topic of boundary layer turbulence transition. The experiment entails a vibrating ribbon that is placed at the base of the inlet to a flow channel and acts to introduce perturbations into the flow. The perturbations evolve into disturbance waves known as Tollmien–Schillichting (TS) waves that

J. C. Chen (✉) · W. Chen
School of Civil and Environmental Engineering, Nanyang Technological University,
Block N1, Nanyang Avenue, Singapore 639798, Singapore
e-mail: jimchen@ntu.edu.sg

travel downstream. As the disturbance waves propagate downstream, they will begin to interact with one another in progressive stages of transition towards flow turbulence. Initially, the wave interactions are linear in the linear instability stage [2]. Further downstream, the wave interactions become nonlinear. The nonlinear interactions spawn a secondary instability in the flow. The secondary instability eventually becomes unsustainable and break down into turbulence. The stages of transition up to linear instability are well-understood presently. The linear wave interactions can be described accurately with the Orr-Sommerfeld (OS) equation of linear stability theory. Chen and Chen 2010 [3] proffers a scholastic study of the linear stage of transition. However, once the waves undergo nonlinear interactions, the transition phenomenon becomes mysterious and is the subject of much cerebration.

A subsequent classical work, Klebanoff et al. 1962 [4], would shed illuminating insight into the nonlinear stage of transition. As transition to turbulence can occur via multiple pathways, Klebanoff et al. 1962 [4] studies the pathway that has come to bear the namesake of its author, K-type transition. When the amplitude of the initial perturbation exceeds 1 % of the mean flow, the K-type transition mechanism activates to induce an explosive amplification of waves leading to breakdown into turbulence. Klebanoff et al. 1962 [4] observes definitive and reproducible behavior of nonlinear wave interactions beginning with the formation of the first set of waves from the perturbation known as the fundamental waves. The fundamental wave exercises a fecundity that begets second and third harmonics of successively higher wave frequencies. The harmonics would then cluster in wave packets as they traverse downstream. Within the packets, the waves interact and synchronize. The phase synchronization of the waves results in explosive spikes in the observed wave oscillations. These observations have become bespoke signature features of nonlinear turbulence transition [4].

Additional classical works would ensue. Kachanov and Levchenko 1984 [5] and Kachanov 1994 [6] reveal another possible pathway towards turbulence called the N-type transition. The N-type transition is a more controlled pathway to turbulence, evoked by a lower amplitude of the initial disturbance than K-type transition. As such, the N-type transition transpires with measurably exponential amplification of waves as contrasted with the incontinent explosion in the K-type. Also, the N-type transition generates harmonics of lower frequencies than the K-type. The N-type wave interactions are termed sub-harmonic resonance that involves the initial TS disturbance waves of a given frequency β_1 and subsequently generated sub-harmonic waves with frequencies $\beta_{1/2} \approx \beta_1/2$. For more details, Herbert 1988 [7] offers an excellent review of the nonlinear transition stage.

Chen 2009 [8], Chen and Chen 2010 [3], Chen and Chen 2011 [9], Chen and Chen 2012 [10], Chen and Chen 2012 [11], Chen and Chen 2009 [2], Chen and Chen 2010 [12], Chen and Chen 2011 [13], Chen and Chen 2011 [14], Chen and Chen 2012 [15], and Chen and Chen 2012 [16] respectfully anthologize his study of boundary layer turbulence transition.

2 Three Dimensionality

During the transition towards turbulence, the generated waves acquire a three-dimensional characteristic. The formation of three-dimensional waves represents a key development in turbulence transition. Saric et al. 2003 [17] explains that the three-dimensional waves arise from crossflow and centrifugal instabilities occurring in flow regions with pressure gradients. The three-dimensional nature of the flow is the critical element that leads to rapid generation of additional harmonics and their subsequent explosive or exponential amplification. Orszag and Patera 1983 [18] notes that, during wave interactions, the two-dimensional waves are unstable to the presence of even infinitesimal three-dimensional waves and will amplify exponentially from the encounter. Orszag and Patera 1983 [18] systematically illustrates that the combination of vortex stretching and tilting terms in the governing Vorticity Transport Equation accelerates the growth of waves. Both vortex stretching and tilting are required to produce the accelerated growth of waves [18]. Both are three-dimensional phenomena and thus, concurringly underline the important role of three-dimensionality in turbulence transition. Reed and Saric 1989 [19] and Herbert 1988 [7] offer excellent reviews of the formation of three-dimensional waves.

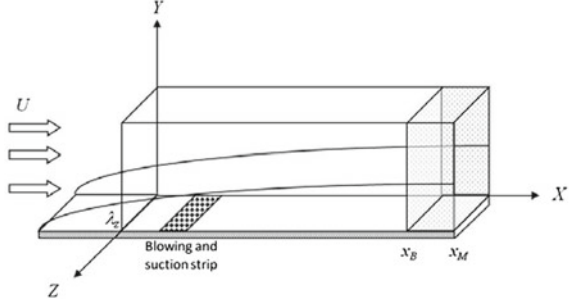
Three-dimensional flows carve a frontier of great interest in cutting-edge fluid dynamics research. However, numerical visualization of three-dimensional waves incurs vast computational demands. The computational demands exponentially reach impractical levels for even typical turbulent flow conditions. Therefore, easing computational demands to within practical limits poses a mandate of utmost importance. Spectral methods offer such a reprieve.

3 Problem Definition

3.1 Flow Domain

Figure 1 depicts the classical three-dimensional boundary layer flow problem of Schubauer and Skramstadt 1947 [1]. A blowing and suction strip generates the disturbances. The spanwise z -direction covers one disturbance wavelength λ_z . A buffer domain before the outflow boundary ramps down the disturbances to prevent wave reflection [10, 12].

Fig. 1 Schematic of the flow domain



3.2 Governing Equation

Non-dimensional variables are used. They are related to their dimensional counterparts, denoted by bars, as follows:

$$Re = \frac{\bar{U}_\infty \bar{L}}{\bar{\nu}}, \quad x = \frac{\bar{x}}{\bar{L}}, \quad y = \frac{\bar{y} \sqrt{Re}}{\bar{L}}, \quad (1a)$$

$$t = \frac{\bar{U}_\infty \bar{t}}{\bar{L}}, \quad u = \frac{\bar{u}}{\bar{U}_\infty}, \quad v = \frac{\bar{v} \sqrt{Re}}{\bar{U}_\infty}, \quad \text{and} \quad w = \frac{\bar{w}}{\bar{U}_\infty} \quad (1b)$$

where the characteristic length $\bar{L} = 0.05$ m, freestream velocity $\bar{U}_\infty = 30$ m/s, kinematic viscosity $\bar{\nu} = 1.5 \times 10^{-5}$ m²/s, Re is the Reynolds number, u , v , and w are the streamwise, transverse, and spanwise flow velocities.

The total flow velocity and vorticity (V, Ω) comprise a steady two-dimensional base flow (V_B, Ω_B) and an unsteady three-dimensional disturbance flow (V', Ω') [20, 21]:

$$V(t, x, y, z) = V_B(x, y, z) + V'(t, x, y, z), \quad (2)$$

$$\Omega(t, x, y, z) = \Omega_B(x, y, z) + \Omega'(t, x, y, z), \quad (3)$$

$$V_B = \{u_B, v_B, 0\}, \quad \Omega_B = \{0, 0, \omega_{zB}\}, \quad (4)$$

$$V' = \{u', v', w'\}, \quad \text{and} \quad \Omega' = \{\omega'_x, \omega'_y, \omega'_z\}. \quad (5)$$

The governing equations for the disturbance flow are the Vorticity Transport Equations [20, 21]:

$$\frac{\partial \omega'_x}{\partial t} + \frac{\partial a}{\partial y} - \frac{\partial c}{\partial z} = \frac{1}{Re} \frac{\partial^2 \omega'_x}{\partial x^2} + \frac{\partial^2 \omega'_x}{\partial y^2} + \frac{1}{Re} \frac{\partial^2 \omega'_x}{\partial z^2}, \quad (6)$$

$$\frac{\partial \omega'_y}{\partial t} - \frac{\partial a}{\partial x} + \frac{\partial b}{\partial z} = \frac{1}{Re} \frac{\partial^2 \omega'_y}{\partial x^2} + \frac{\partial^2 \omega'_y}{\partial y^2} + \frac{1}{Re} \frac{\partial^2 \omega'_y}{\partial z^2}, \quad (7)$$

$$\frac{\partial \omega'_z}{\partial t} + \frac{\partial c}{\partial x} - \frac{\partial b}{\partial y} = \frac{1}{Re} \frac{\partial^2 \omega'_z}{\partial x^2} + \frac{\partial^2 \omega'_z}{\partial y^2} + \frac{1}{Re} \frac{\partial^2 \omega'_z}{\partial z^2}, \quad (8)$$

$$a = v' \omega'_x - u' \omega'_y + v'_B \omega'_x - u_B \omega'_{y'}, \quad (9)$$

$$b = w' \omega'_y - v' \omega'_z + v_B \omega'_z + v' \omega'_{z_{B'}}, \quad (10)$$

$$c = u' \omega'_z - w' \omega'_x + u_B \omega'_z + u' \omega'_{z_{B'}}, \quad (11)$$

$$\omega'_x = \frac{1}{Re} \frac{\partial v'}{\partial z} - \frac{\partial w'}{\partial y}, \quad \omega'_y = \frac{\partial w'}{\partial x} - \frac{\partial u'}{\partial z}, \quad \text{and} \quad \omega'_z = \frac{\partial u'}{\partial y} - \frac{1}{Re} \frac{\partial v'}{\partial x}, \quad (12)$$

In addition, there is a set of Poisson's equations [20, 21]:

$$\frac{\partial^2 u'}{\partial x^2} + \frac{\partial^2 u'}{\partial z^2} = -\frac{\partial \omega'_y}{\partial z} - \frac{\partial^2 v'}{\partial x \partial y'}, \quad (13)$$

$$\frac{1}{Re} \frac{\partial^2 v'}{\partial x^2} + \frac{\partial^2 v'}{\partial y^2} + \frac{1}{Re} \frac{\partial^2 v'}{\partial z^2} = \frac{\partial \omega'_x}{\partial z} - \frac{\partial \omega'_z}{\partial x}, \quad \text{and} \quad (14)$$

$$\frac{\partial^2 w'}{\partial x^2} + \frac{\partial^2 w'}{\partial z^2} = \frac{\partial \omega'_y}{\partial x} - \frac{\partial^2 v'}{\partial y \partial z}. \quad (15)$$

Finally, the continuity equation completes the set:

$$\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} + \frac{\partial w'}{\partial z} = 0. \quad (16)$$

Zero pressure gradient (ZPG) boundary layer flow serves as the base flow. The solution procedure begins by solving the two-dimensional steady base flow followed by the three-dimensional disturbance flow. Chen and Chen 2012 [10] copiously details this solution algorithm.

3.3 Boundary Conditions

3.3.1 Inflow Boundary Condition

The inflow boundary introduces no disturbances; hence, u' , v' , w' , ω'_x , ω'_y , and ω'_z along with their first and second derivatives are all zero there.

3.3.2 Freestream Boundary Condition

Assuming potential flow at the freestream, the vorticity is zero there:

$$\omega'_x = 0, \frac{\partial \omega'_x}{\partial y} = 0, \frac{\partial^2 \omega'_x}{\partial y^2} = 0, \quad (17)$$

$$\omega'_y = 0, \frac{\partial \omega'_y}{\partial y} = 0, \frac{\partial^2 \omega'_y}{\partial y^2} = 0, \quad (18)$$

$$\omega'_z = 0, \frac{\partial \omega'_z}{\partial y} = 0, \frac{\partial^2 \omega'_z}{\partial y^2} = 0, \text{ and} \quad (19)$$

$$\frac{\partial v'}{\partial y} = -\frac{\alpha^*}{\sqrt{Re}} v'. \quad (20)$$

The parameter v' decays much slower, and so its wall-normal derivative remains appreciable with a prescribed wave number of α^* [20, 21].

3.3.3 Wall Boundary Condition

The boundary conditions at the wall are [20, 21]:

$$u' = 0, \quad (21)$$

$$v' = 0, \frac{\partial v'}{\partial y} = 0, \quad (22)$$

$$w' = 0, \quad (23)$$

$$\frac{\partial^2 \omega'_x}{\partial x^2} + \frac{\partial^2 \omega'_x}{\partial z^2} = -\frac{\partial \omega'_y}{\partial x \partial y} + \frac{\partial}{\partial z} \left(\frac{1}{Re} \frac{\partial^2 v'}{\partial x^2} + \frac{\partial^2 v'}{\partial y^2} + \frac{1}{Re} \frac{\partial^2 v'}{\partial z^2} \right), \quad (24)$$

$$\omega'_y = 0, \text{ and} \quad (25)$$

$$\frac{\partial^2 \omega'_z}{\partial x} = \frac{\partial \omega'_x}{\partial z} - \left(\frac{1}{Re} \frac{\partial^2 v'}{\partial x^2} + \frac{\partial^2 v'}{\partial y^2} + \frac{1}{Re} \frac{\partial^2 v'}{\partial z^2} \right). \quad (26)$$

Correct definition of the wall boundary conditions remains a hotly contested matter. Problematic issues include the need to preserve continuity and vorticity. Chen and Chen 2010 [3] engages in a tendentious intellection of the issue of defining the wall boundary condition.

The blowing and suction strip shown in Fig. 1 generates disturbances in spectral space according to [20, 21]:

$$\hat{v}_k = A_k f(x) \sqrt{Re} \sin(\beta_k t), \quad (27)$$

$$f(x) = \begin{cases} 24.96\xi^6 - 56.16\xi^5 + 31.2\xi^4 \\ -24.96\xi^6 + 56.16\xi^5 - 31.2\xi^4 \end{cases}, \quad (28)$$

$$\xi = \frac{x - x_1}{x_{st} - x_1} \quad x_1 < x < x_{st} \text{ for the first case, and} \quad (29)$$

$$\xi = \frac{x_1 - x}{x_2 - x_{st}} \quad x_{st} < x < x_2 \text{ for the second case,} \quad (30)$$

where $k = 0$ or 1 , A_k the disturbance amplitude $A_k = 1.0 \times 10^{-4}$, and the disturbance frequency $\beta_k = 10.0$. The Fourier modes $k = 0$ and 1 correspond to the two- and three-dimensional disturbances, respectively.

3.3.4 Outflow Boundary Condition

A buffer domain located at $x = x_B$ prior to the outflow boundary at $x = x_M$ ramps down the flow disturbances with ramping function [20, 21]:

$$T(L_b) = \frac{0.9778 + 1.3074\cos(\pi L_b) + 0.33187\cos(2\pi L_b) + 0.0022278\cos(3\pi L_b)}{1 + 0.63707\cos(\pi L_b) + 0.17261\cos(2\pi L_b)} \text{ with} \quad (31)$$

$$L_b = \frac{x_B - x}{x_B - x_M}. \quad (32)$$

3.3.5 Spanwise Boundary Condition

The spanwise boundaries located at $Z = 0$ and $Z = \lambda_z$ implement periodic boundary conditions where all disturbance flow parameters and their derivatives at $Z = 0$ are equal to their counterparts at $Z = \lambda_z$ [20, 21].

4 Spectral Method

Spectral methods offer the advantages of exponential convergence, numerical accuracy, and computational efficiency. They have demonstrated superior performance to finite difference methods [22]. Spectral methods approximate the solution to a given set of governing equations, for example, by assuming the solution be a Fourier series or Chebyshev polynomials [22]. In contrast, finite difference methods

approximate the original governing equations and then seek to solve the approximate problem. Spectral methods are especially suited for problems with periodic boundary conditions, which in this study occur at the spanwise boundaries. This study uses a Fourier series approximation for the solution [20, 21, 23]:

$$f'(x, y, z, t) = \sum_{k=-K}^K F_k(x, y, t) \exp(ik\gamma_k z) \quad (33)$$

where f' is a flow variable of interest such as the velocity, $F_k(x, y, t)$ is the Fourier amplitude, $\gamma_k = 2\pi k/\lambda_z$ is the spanwise wave number, and λ_z is the largest spanwise wavelength of flow disturbances. The mode F_k is the complex conjugate of F_{-k} , so the governing equations and boundary conditions can be transformed to $K + 1$ equations and boundary conditions in the two-dimensional x - y plane.

For cases with symmetric flow, the Fourier expansion can be compacted to sines and cosines separately as [20, 21, 23]:

$$(u', v', \omega'_z, b, c) = \sum_{k=0}^K (\hat{u}'_k, \hat{v}'_k, \hat{\omega}'_{zk}, B_k, C_k) \cos(\gamma_k z) \quad \text{and} \quad (34)$$

$$(w', \omega'_x, \omega'_y, a) = \sum_{k=1}^K (\hat{w}'_k, \hat{\omega}'_{xk}, \hat{\omega}'_{yk}, A_k) \sin(\gamma_k z). \quad (35)$$

The parameters $u', v', \omega'_z, b,$ and c are symmetric; $w', \omega'_x, \omega'_y,$ and a are anti-symmetric. This compaction requires half the wavelength, $\lambda_z/2$, hence, reduces computational demands by half [20, 21, 23].

The governing equations in spectral representation are:

$$\frac{\partial \hat{\omega}'_{xk}}{\partial t} + \frac{\partial A_k}{\partial y} + \gamma_k C_k = \frac{1}{Re} \frac{\partial^2 \hat{\omega}'_{xk}}{\partial x^2} + \frac{\partial^2 \hat{\omega}'_{xk}}{\partial y^2} - \frac{\gamma_k^2}{Re} \hat{\omega}'_{xk}, \quad (36)$$

$$\frac{\partial \hat{\omega}'_{yk}}{\partial t} - \frac{\partial A_k}{\partial x} - \gamma_k B_k = \frac{1}{Re} \frac{\partial^2 \hat{\omega}'_{yk}}{\partial x^2} + \frac{\partial^2 \hat{\omega}'_{yk}}{\partial y^2} - \frac{\gamma_k^2}{Re} \hat{\omega}'_{yk}, \quad (37)$$

$$\frac{\partial \hat{\omega}'_{zk}}{\partial t} + \frac{\partial C_k}{\partial x} - \frac{\partial B_k}{\partial y} = \frac{1}{Re} \frac{\partial^2 \hat{\omega}'_{zk}}{\partial x^2} + \frac{\partial^2 \hat{\omega}'_{zk}}{\partial y^2} - \frac{\gamma_k^2}{Re} \hat{\omega}'_{zk}, \quad (38)$$

$$\frac{1}{Re} \frac{\partial^2 \hat{v}'_k}{\partial x^2} + \frac{\partial^2 \hat{v}'_k}{\partial y^2} - \frac{\gamma_k^2}{Re} \hat{v}'_k = \gamma_k \hat{\omega}'_{xk} - \frac{\partial \hat{\omega}'_{zk}}{\partial x}, \quad (39)$$

$$\frac{\partial^2 \hat{w}'_k}{\partial x^2} - \gamma_k^2 \hat{w}'_k = \frac{\partial \hat{\omega}'_{yk}}{\partial x} + \gamma_k \frac{\partial \hat{v}'_k}{\partial y}, \quad \text{and} \quad (40)$$

$$\frac{\partial^2 \hat{u}'_k}{\partial x^2} - \gamma_k^2 \hat{u}'_k = -\gamma_k \hat{\omega}'_{yk} - \frac{\partial^2 \hat{v}'_k}{\partial x \partial y}. \quad (41)$$

5 Overall Numerical Method

The present study uses a spectral Fourier method in the spanwise direction. By that, imagine the flow domain shown in Fig. 1 to be divided into a series of x - y planes. For each x - y plane, the governing equations are given in Eqs. (36–41). Each x - y plane would then need its numerical discretization. The spatial discretization uses 12th-order Combined Compact Difference (CCD) schemes. The temporal discretization uses a 4th-order 5-6 alternating stages Runge–Kutta (RK) scheme. Chen and Chen 2011 [9], Chen and Chen 2012 [10], and Chen and Chen 2012 [11] comprise a graphomaniac series on the development of these numerical methods.

6 Non-Uniform Grid

6.1 Computational Efficiency

Since the numerical realization of turbulence transition exerts impractically onerous computational demands, one would do well to preserve computational resources as much as possible. One method of conservation uses non-uniform grids that concentrate the computational resolution in regions of interest and relax to coarse resolutions in regions of less relevance. For the case of boundary layer turbulence transition, this entails using very fine grids near the wall where the transition process occurs and gradually coarsening the grid away from the wall. In so doing, the computational capacity would be allocated with maximum utility.

Furthermore, micro-scaled wave interactions in turbulence transition can be easily distorted by numerical errors. So, high-order numerical methods would seem to be a logical remedy to control the errors. However, high-order numerical methods present an additional issue at the wall boundary. To properly close a numerical method, the appropriate boundary scheme would generally be at least one order lower than the numerical scheme in the interior domain, in order to prevent numerical instability [24]. The difference in orders between the interior and boundary schemes widens with increasing order of the numerical method [24]. So, even when using a high-order numerical method, the overall order of the numerical method would be diluted by the need for lower-order boundary schemes, hence, posing a threat to the numerical stability [24]. One means to preserve numerical stability of high-order methods at the boundary and combat the dilutive effects of lowering the order of the boundary scheme implements a non-uniform grid that concentrates fine grid spacing near the wall. The solution would first generate a non-uniform grid and then derive a numerical method with coefficients bespoke to the non-uniform grid [24].

6.2 Non-Uniform Grid Generation

The non-uniform grid is generated in the wall-normal y -direction using:

$$y = y_c \left(1 + \frac{a \sin(-\alpha_g \cos(\frac{\pi i}{2c}))}{a \sin(-\alpha_g)} \right) \text{ for } 0 \leq i \leq c \text{ and} \quad (42)$$

$$y = y_c + (y_c - y_{c-1}) \left(\beta_g^{(i-c)} - 1/\beta_g - 1 \right) \text{ for } c + 1 \leq i \leq n \quad (43)$$

where α_g and β_g are the grid stretching parameters, and c is the index for a designated node point where the two piecewise functions meet.

6.3 Numerical Method Bespoke to Non-Uniform Grid

The numerical scheme would be derived bespoke to the non-uniform grid. High-order CCD schemes provide the advantages of accuracy of simulations and control of numerical errors. The CCD scheme combines the discretization for the function, f , its first derivative, F , and second derivative, S , with a , b , and c as the coefficients of the scheme and h as the grid size:

$$h \sum_{j=j_1}^{j_2} a_{1,j} F_{i+j} + h^2 \sum_{j=j_1}^{j_2} b_{1,j} S_{i+j} + \sum_{j=j_1}^{j_2} c_{1,j} f_{i+j} = 0 \text{ and} \quad (44)$$

$$h \sum_{j=j_1}^{j_2} a_{2,j} F_{i+j} + h^2 \sum_{j=j_1}^{j_2} b_{2,k} S_{i+j} + \sum_{j=j_1}^{j_2} c_{2,j} f_{i+j} = 0. \quad (45)$$

The coefficients of the CCD scheme are derived using Lagrange polynomial interpolation of a function $y(x)$ [24]:

$$y(x) = \sum_{i=1}^n l_i(x) f(x_i) \quad (46)$$

where $l_i(x)$'s are Lagrange basis polynomials [24]:

$$l_i(x) = \prod_{j=1, j \neq i}^n \frac{(x - x_j)}{(x_i - x_j)}. \quad (47)$$

The Lagrange polynomial interpolation can be expanded to [24]:

$$y(x) = \sum_{d=0}^D \sum_{i \in I_n} \rho_{d,i}(x) f^{(d)}(x_i) + \sum_{i \in I_m} r_i(x) f(x_i) \quad (48)$$

where $f^{(d)}(x_i)$ is the D^{th} -order derivative of the function $f(x_i)$, I_n is the set of points defining $f^{(d)}(x_i)$ up to the D^{th} -order derivative, I_m is the set of points defining only

the function values of $f(x_i)$, and $\rho_{d,i}(x)$, and $r_i(x)$ are additional interpolation polynomials. The numerical schemes can be derived by differentiating Eq. (48) D times to obtain the expressions for $y^{(p)}(x)$ as [24]:

$$y^{(p)}(x) = \sum_{d=0}^D \sum_{i \in I_n} \rho_{d,i}^{(p)}(x) f^{(d)}(x_i) + \sum_{i \in I_m} r_i^{(p)}(x) f(x_i) \quad (49)$$

for $p = 1, 2, \dots, D$. The coefficients are derived from Eq. (49). In this study, the numerical scheme derived is a 12th-order 5-point non-uniform CCD scheme. The concomitant boundary schemes are 10th- and 11th-order.

7 Stability of Numerical Method

With the objective of customizing the numerical method to a non-uniform grid for strengthening numerical stability, a logical evaluation of the method would consider its ranges of stability. The stability of a numerical method entails its temporal and spatial discretizations. Both aspects must be numerically stable. Mathematical theory decrees that the eigenvalues of a spatial discretization define its range of numerical stability [25]. The eigenvalue analysis begins with applying the numerical method, a 12th-order 5-point non-uniform CCD scheme with 10th and 11th-order boundary schemes, to a reference governing equation, the classical one-dimensional convective diffusion equation. The theory mandates that the real part of the eigenvalues must be negative to ensure stability of the spatial discretization.

For the temporal discretization, the theory examines its amplification factor [25]. A temporal discretization will be stable if the absolute value of the amplification factor is less than one. Since the temporal discretization integrates the spatial discretization over time, the amplification factor is a function of the eigenvalue. This linkage allows for concurrent examination of both discretizations. The overall stability condition would require first that the real part of the eigenvalue be negative. Next, it selects the eigenvalues that limit the amplification factor to less than one.

Applying this two-step analysis to the one-dimensional convective diffusion equation indeed yields a set of eigenvalues for which the numerical method will remain stable, as illustrated in Fig. 2. The stability of the numerical method when applied to this reference case provides indications as to how it will fare on the actual flow problem.

8 Validation of Numerical Method

The stable numerical method must be validated for accuracy. An accurate numerical model would agree well with the Orr-Sommerfeld (OS) equation of linear stability theory. Figure 3 affirms agreement amongst the present study, the

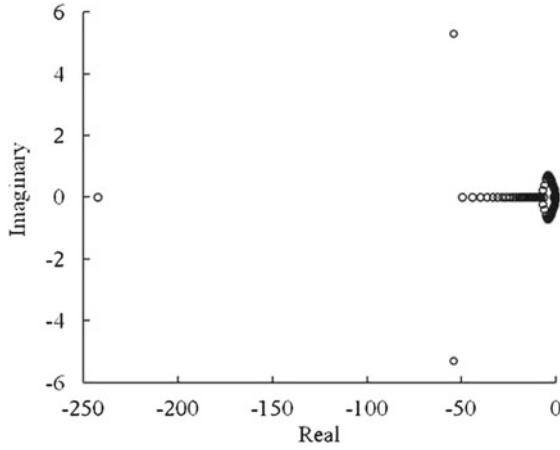


Fig. 2 Stability range of eigenvalues

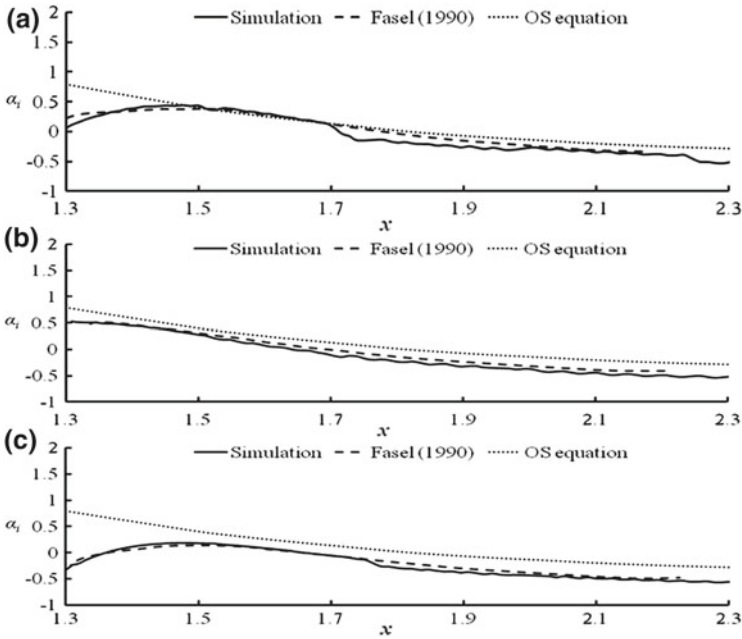


Fig. 3 Comparison of model, three-dimensional OS equation, and Fasel et al. 1990 [26] for downstream amplification rates α_i of velocities. (a) u' (b) v' and (c) w'

three-dimensional OS equation, and the numerical study of Fasel et al. 1990 [26] for the downstream amplification rates α_i of the disturbance velocities u' , v' , and w' :

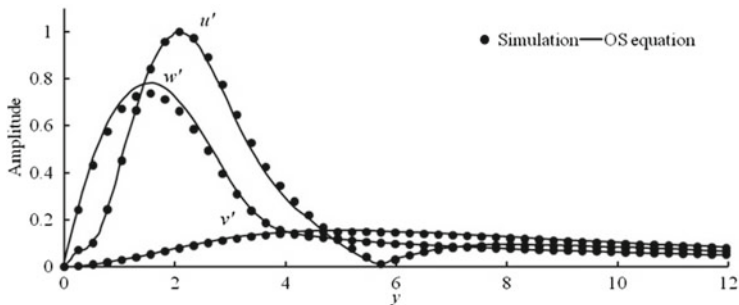


Fig. 4 Transverse profiles of disturbance velocities u' , v' , w' from Fig. 3

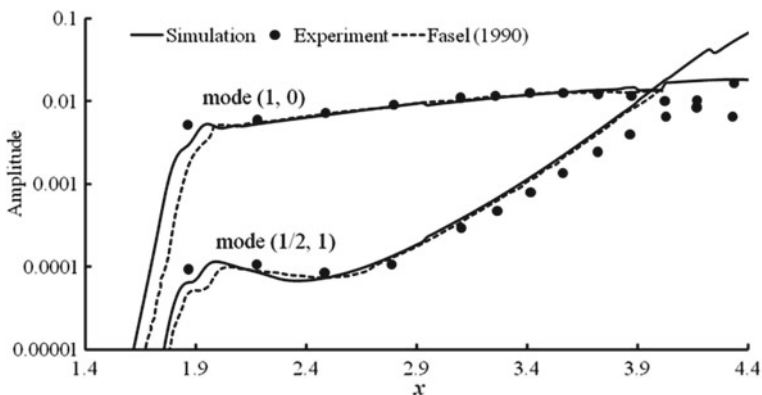


Fig. 5 Comparison amongst present study, Kachanov and Levchenko 1984 [5], and Fasel et al. 1990 [26] for downstream amplification of waves: modes (1, 0) and (1/2, 1)

$$\alpha_i = \frac{d}{dx} \ln(f')$$
(50)

where f' is the flow variable of interest. Figure 4 asserts further averment with a near complete overlap between the present study and the three-dimensional OS equation for the transverse profiles of the disturbance velocities u' , v' and w' . With confidence in the model accuracy, it is now ready for investigation of turbulence transition.

9 Sub-Harmonic Resonance

The classical works Kachanov and Levchenko 1984 [5] and Kachanov 1994 [6] reveal the N-type turbulence transition involving sub-harmonic resonance of the propagating waves in ZPG boundary layer flow. The present study respectfully

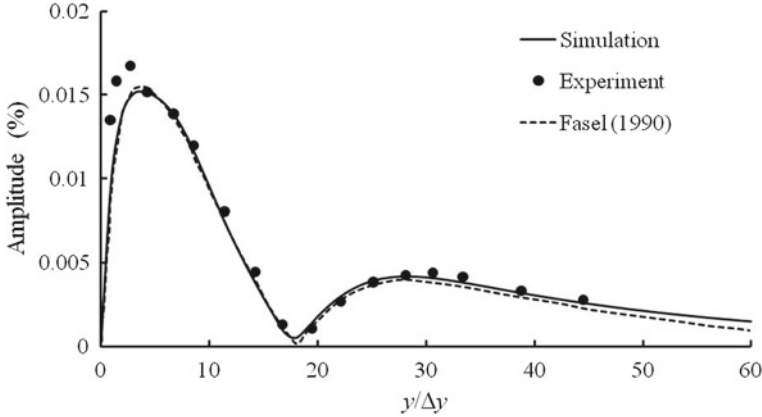


Fig. 6 Transverse profiles of wave amplitudes of the mode (1, 0) from Fig. 5

simulates the experiments presented forth in Kachanov and Levchenko 1984 [5] using three Fourier modes in the spectral method with lowest wave number 31.47. The blowing and suction strip generates two-dimensional disturbances with $\beta_0 = 12.4$ and $A_0 = 1.2 \times 10^{-4}$ and three-dimensional disturbances with $\beta_1 = 6.2$ and $A_1 = 5.1 \times 10^{-6}$.

Figure 5 shows comparison of the present study, Kachanov and Levchenko 1984 [5], and Fasel et al. 1990 [26] for downstream amplification of the $(u'^2)^{1/2}/U_\infty$ waves: modes (1, 0), two-dimensional initial TS waves, and (1/2, 1), subsequently generated three-dimensional sub-harmonic waves. The first entry in the brackets stands for multiples of the TS wave frequency, and the second entry is multiples of the spanwise wave number. Figure 6 shows the transverse profile of the wave amplitude for the mode (1, 0).

10 Conclusion

This is a study of sub-harmonic resonance in three-dimensional ZPG boundary layer flow. Sub-harmonic wave generation, amplification, and interaction drive the N-type pathway in turbulence transition made famous by the classical works of Kachanov and Levchenko 1984 [5] and Kachanov 1994 [6]. At the onset of turbulence transition, the wave dynamics become three-dimensional. Investigation of three-dimensional flows propels the forefront of cutting-edge fluid dynamics research. Three-dimensionality incurs exorbitant computational demands on the numerical simulations that exponentially increase to impractical levels. The present study countervails such onerous computational demands with a spectral method and a non-uniform grid. The symmetrical characteristic of the flow in the present study requires only sine and cosine expansions in the spectral method,

thereby reducing the computational demands by half. The non-uniform grid raises numerical stability issues at the wall boundary. Eigenvalue analysis ascertains ranges of stability of the numerical method. Validation of the numerical method versus the three-dimensional OS equation avers confidence in the accuracy of the model. The present study respectfully emulates the classical experiments of Kachanov and Levchenko 1984 [5].

References

1. Schubauer G, Skramstad HK (1947) Laminar boundary layer oscillations and stability of laminar flow. *J Aeronaut Sci* 14(2):69–78
2. Chen J, Chen W (2009) Turbulence transition in two-dimensional boundary layer flow: linear instability. Proceedings of 6th international conference on flow Dyn, Sendai, Japan, 2009, pp 148–149
3. Chen J, Chen W (2010) The complex nature of turbulence transition in boundary layer flow over a flat surface. *Intl J Emerg Multidiscip Fluid Sci* 2(2–3):183–203
4. Klebanoff PS, Tidstrom KD, Sargent LM (1962) The three-dimensional nature of boundary-layer instability. *J Fluid Mech* 12(1):1–34
5. Kachanov YS, Levchenko VY (1984) The resonant interaction of disturbances at laminar-turbulent transition in a boundary layer. *J Fluid Mech* 138:209–247
6. Kachanov YS (1994) Physical mechanisms of laminar-boundary-layer transition. *Annu Rev Fluid Mech* 26:411–482
7. Herbert T (1988) Secondary instability of boundary layers. *Annu Rev Fluid Mech* 20: 487–526
8. Chen J (2009) The law of multi-scale turbulence. *Intl J Emerg Multidiscip Fluid Sci* 1(3):165–179
9. Chen J, Chen W (2011) Two-dimensional nonlinear wave dynamics in blasius boundary layer flow using combined compact difference methods. *IAENG Intl J Appl Math* 41(2):162–171
10. Chen W, Chen J (2012) Combined compact difference methods for solving the incompressible Navier-Stokes equations. *Intl J Numer Methods Fluids* pp 1234–1256
11. Chen J, Chen W (2012) Nonlinear wave dynamics in two-dimensional boundary layer flow. In: Ao S, Chan A, Katagiri H, Xu L (eds) *IAENG transactions on engineering technologies*, vol 7, World Scientific, Hong Kong, pp 130–143
12. Chen J, Chen W (2010) Combined compact difference method for simulation of nonlinear wave generation, interaction, and amplification in boundary layer turbulence transition. Proceedings of 7th International Conference on Flow Dyn, Sendai, Japan, pp 82–83
13. Chen J, and Chen W (2011) Numerical realization of nonlinear wave dynamics in turbulence transition using Combined Compact Difference methods. In: Proceedings of the international multiconference of engineers and computer scientists 2011, Lecture notes in engineering and computer science, vol 2, Hong Kong, pp 1517–1522
14. Chen J, Chen W (2011) Non-uniform grids in numerical simulations of boundary layer turbulence transition. Proceedings of 8th International Conference on Flow Dyn., Sendai, Japan, pp 168–169
15. Chen J, Chen W (2012) Sub-harmonic resonance in three-dimensional boundary layer flow. In: Proceedings of the international multiconference of engineers and computer scientists 2012, Lecture notes in engineering and computer science, vol 2, Hong Kong, pp 1635–1640
16. Chen J, Chen W (2012) How flow becomes turbulent. *IAENG Intl J Appl Math* 42(2):99–110
17. Saric WS, Reed HL, White EB (2003) Stability and transition of three-dimensional boundary layers. *Annu Rev Fluid Mech* 35:413–440

18. Orszag SA, Patera AT (1983) Secondary instability of wall-bounded shear flows. *J Fluid Mech* 128:347–385
19. Reed HL, Saric WS (1989) Stability of three-dimensional boundary layers. *Annu Rev Fluid Mech* 21:235–284
20. Meitz HL, Fasel HF (2000) A compact-difference scheme for the Navier-Stokes equations in vorticity-velocity formulation. *J Comput Phys* 157(1):371–403
21. Rist U, Fasel H (1995) Direct numerical simulation of controlled transition in a flat-plate boundary layer. *J Fluid Mech* 298:211–248
22. Canuto C, Hussaini MY, Quarteroni A, Zang TA (2006) *Spectral methods: Fundamentals in single domains*. Springer, Berlin
23. Liu C, Maslowe SA (1999) A numerical investigation of resonant interactions in adverse-pressure-gradient boundary layers. *J Fluid Mech* 378:269–289
24. Zhong X, Tatineni M (2003) High-order non-uniform grid schemes for numerical simulation of hypersonic boundary-layer stability and transition. *J Comput Phys* 190(2):419–458
25. Hirsch C (2007) *Numerical computation of internal and external flows: fundamentals of computational fluid dynamics*, 2nd edn. Butterworth-Heinemann, London
26. Fasel HF, Rist U, Konzelmann U (1990) Numerical investigation of the three-dimensional development in boundary-layer transition. *AIAA J* 28(1):29–37

Efficient Algorithms for the Green's Function Formalism

Semiconductor Transport Simulations on CPUs and GPUs

Jan Jacob, Bodo Krause-Kyora, Lothar Wenzel, Qing Ruan,
Darren Schmidt, Vivek Amin and Jairo Sinova

Abstract We present efficient implementations of the non-equilibrium Green's function method for numeric simulations of transport in semiconductor nanostructures. The algorithms are implemented on CPUs and GPUs using LabVIEW 2011 64-Bit together with the Multicore Analysis and Sparse Matrix Toolkit and the GPU Analysis Toolkit.

Keywords Algorithm · GPU · Green's function · Inversion · LabVIEW · Simulations

J. Jacob (✉)

Institute of Applied Physics, University of Hamburg, Jungiusstraße 11,
22359 Hamburg, Germany
e-mail: jjacob@physnet.uni-hamburg.de

B. Krause-Kyora

PHYSnet Computing Center, University of Hamburg, Jungiusstraße 9,
22359 Hamburg, Germany
e-mail: krause@physnet.uni-hamburg.de

L. Wenzel · Q. Ruan · D. Schmidt

National Instruments, 11500 N Mopac Expwy, Austin, TX 78759-3504, USA
e-mail: lothar.wenzel@ni.com

Q. Ruan

e-mail: qing.ruan@ni.com

D. Schmidt

e-mail: darren.schmidt@ni.com

V. Amin · J. Sinova

Physics Department, Texas A&M University, 4242 TAMU, College Station,
TX 77843-4242, USA
e-mail: aminvp@physics.tamu.edu

J. Sinova

e-mail: sinova@physics.tamu.edu

1 Introduction

The continuing circuit miniaturization leads to device dimensions where quantum effects can be detrimental to the operation of standard CMOS devices. At the same time these effects yield the potential for novel energy efficient devices, since their channels no longer have to be depleted completely for switching [2, 9]. An electric field across the channel that is generated by a gate can influence the electron’s spin precession length—and thereby its spin orientation with respect to a spin-sensitive detector [25]. Besides optical generation and detection of spin-polarized currents [18] and by ferromagnetic electrodes [23] also all-semiconductor devices are considered [6, 7] to pave the way to possible spin-based computing.

While charge-based devices are well understood, the realization of their spin-based cousins is still challenging. Analytical predictions for these nano structures often cannot be directly compared with experiments due to oversimplification. Numerical simulations bridge the gap between analytical descriptions and experiments allowing a thorough understanding of the quantum effects in such semiconductor devices by accounting for donor impurities, lattice imperfections, and interactions within a sample that are inaccessible to most analytical methods. Simulations with both realistic dimensions and appropriate grid sizes provide significant computational challenges, due to the large processing and memory load.

The non-equilibrium Green’s function (NEGF) method [1] is a common approach to transport simulations. Other approaches, like the transfer matrix algorithm [24], can be translated to the NEGF method exhibiting similar mathematical structure and numerical techniques with the same computational challenges. They can be broken down to eigenvalue problems as well as multiplications and inversions of large, but often sparse matrices. We have explored implementations of the Green’s function method with respect to their memory footprint and scalability over multiple threads, as well as their portability to general purpose graphics processing units (GPU) [8] and now expand our work to multi-GPU and multi-node implementations.

2 Mathematical and Physics Background

We employ the Green’s function method to simulate transport in mesoscopic structures [1] with dimensions smaller than the coherence length. Here the conductance is obtained via the Landauer formula from the transmission probability T describing the probability of carriers to be transmitted from the input to the output contact $\mathbf{G} = \frac{2e^2}{h} T$, where e is the elementary charge and h is Planck’s constant. The transport in these systems is through different quantum-mechanical modes, each contributing $2e^2/h$ to the total conductance for a perfect transmission $T = 1$. In the case of coherent transport, where the conductor is smaller than the phase-relaxation length, the total transmission probability can be decomposed into the sum of the individual channels:

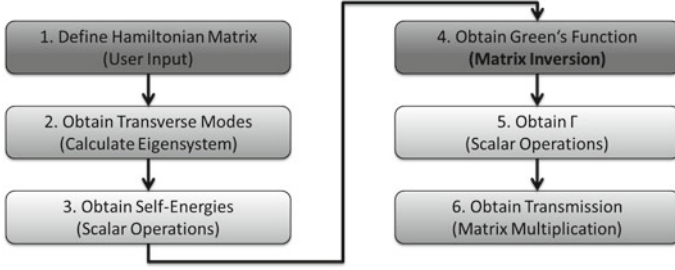


Fig. 1 Flowchart for the basic green's function algorithm

$$\mathbf{G} = \frac{2e^2}{h} \sum_{m,n} T_{mn}, \quad \text{where } T_{mn} = |s_{mn}|^2. \quad (1)$$

Green's functions provide a convenient way to obtain the s -matrix of the microscopic sample. Not going into the detailed relationship between Green's functions and the s -matrix, we will show how to calculate the Green's function for a particular sample and how to obtain T from the Green's function. Starting from the Schrödinger equation $H\Psi = E\Psi$ we define a Green's function, for a system with the Hamiltonian $\hat{H}(r)$

$$[E - \hat{H}(r)]G(r, r') = \delta(r - r'). \quad (2)$$

by rewriting the Schrödinger equation with an added source term. The Green's function can be interpreted as the wave function in terms of the position vector r of the source. Now the device is discretized on a grid in a tight-binding model of finite differences, such that $G(r, r') \rightarrow G_{ij}$, where i and j are denoting different lattice positions corresponding to r and r' . The differential equation becomes a matrix equation $[EI - H]G = I$, where each row or column represents a particular lattice site and each element defines the hopping between the two sites of its row and column. For a system of N_x horizontal and N_y vertical sites the matrix is of dimension $N_x N_y \times N_x N_y$. Considering realistic dimensions of some micrometer and a fine grid spacing of a few nanometer to precisely simulate all quantum effects in such nano structures this leads to huge matrices. Converting the Hamiltonian to a matrix operator also requires discretized derivative operators that are now given by

$$\left[\frac{dF}{dx} \right]_{x=(j+\frac{1}{2})a} \rightarrow \frac{1}{a} [F_{j+1} - F_j] \quad (3)$$

$$\left[\frac{d^2F}{dx^2} \right]_{x=ja} \rightarrow \frac{1}{a^2} \{F_{j+1} - 2F_j + F_{j-1}\}. \quad (4)$$

For a one-dimensional system, where only a simple kinetic and a potential term are considered, the Hamiltonian matrix is given by

$$H = \begin{pmatrix} \dots & -t & 0 & 0 & 0 \\ -t & U_{-1} + 2t & -t & 0 & 0 \\ 0 & -t & U_0 + 2t & -t & 0 \\ 0 & 0 & -t & U_1 + 2t & -t \\ 0 & 0 & 0 & -t & \dots \end{pmatrix}, \quad (5)$$

where $t = \hbar^2/2ma^2$ is the hopping parameter and U_i denotes the potential at each lattice site. With appropriate labeling such a matrix can also be given for two or three dimensional systems. The Green's function can now be compute through matrix inversion.

$$G = [EI - H]^{-1}. \quad (6)$$

There are two independent solutions for G , the retarded and advanced Green's function; often an imaginary parameter is added to the energy in Eq. 6 in order to force one of the solutions. Solutions like Eq. 6 only yield information about scattering inside the sample. The leads are included by connecting them to the sample at various lattice site. Only directly neighboring sites are consider to be relevant to compute the lead's full effect on the transmission of semi-infinite, homogeneous, and reflection-less leads. Considering a sample c represented by a $N_x \times N_y$ grid and fully connected to two leads on either vertical side, labeled p and q , one can rewrite the Green's function in block matrix form as

$$G = \begin{pmatrix} G_c & G_{cp} & G_{cq} \\ G_{pc} & G_p & 0 \\ G_{qc} & 0 & G_q \end{pmatrix}. \quad (7)$$

Carriers enter or leave the sample via G_{cp} , G_{cq} or G_{pc} , G_{qc} and travel through the sample via G_c and in the leads via G_p and G_q . As there is no connection between the leads, carriers must transmit through the sample to travel between p and q . We assume this structure for G :

$$G = \begin{pmatrix} EI - H_c & \tau_p & \tau_q \\ \tau_p^\dagger & EI - H_p & 0 \\ \tau_q^\dagger & 0 & EI - H_q \end{pmatrix}^{-1}, \quad (8)$$

where the dimension of each element of the block matrix depends on the number of lattice sites corresponding to the portion they describe, and $[\tau_{p(q)}]_{ij} = t\delta_{ij}$, assuming that carriers may only enter the sample through a site adjacent to a lead. One can show, after some algebra, that

$$G_c = [EI - H_c - \Sigma]^{-1}, \quad \text{where} \quad (9)$$

$$\Sigma = \begin{pmatrix} t^2 g_p & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & t^2 g_q \end{pmatrix} = \Sigma_p + \Sigma_q \quad \text{and} \quad (10)$$

$$g_{p(q)} = [EI - H_{p(q)}]^{-1}. \quad (11)$$

The $\Sigma_{p(q)}$ are $N_x N_y \times N_x N_y$ matrices and g_p and g_q are $N_y \times N_y$ matrices. In the final step the transmission probability is then calculated by

$$T = \sum_{m,n} T_{mn} = \text{Tr}[\Gamma_p G_c \Gamma_q G_c^\dagger], \quad \text{where } \Gamma_{p(q)} = i[\Sigma_{p(q)} - \Sigma_{p(q)}^\dagger] \quad (12)$$

3 Basic Implementation

The implementation of this algorithm includes six steps (see Fig. 1): First the Hamiltonian H for the system is created. As the matrix H is of size $(N_x N_y) \times (N_x N_y)$ it can cause memory issues, when implemented as a dense matrix. The second step determines the eigensystem for the transverse Hamiltonian H_y of the size $N_y \times N_y$ and can take significant computing resources for large systems. This is used in step three to define the self-energies of the leads by scalar operations. The fourth step calculates the Green's function by inversion of a $(N_x N_y) \times (N_x N_y)$ matrix representing the bottleneck of the algorithm. Step five creates the Γ matrices, requiring only scalar operations on the elements of Σ^A and Σ^R . This can be done in parallel to the inversion in step four. In the sixth step the transmission probability is calculated from Eq. 12 that includes the product of four $(N_x N_y) \times (N_x N_y)$ matrices.

4 Optimizations

The matrix H is of the size $(N_x N_y) \times (N_x N_y)$ scaling with the system size. However, the matrix's structure allows memory optimizations:

$$H_{(N_x N_y) \times (N_x N_y)} = \begin{pmatrix} \ddots & Y & 0 & 0 \\ Y & X & Y & 0 \\ 0 & Y & X & Y \\ 0 & 0 & Y & \ddots \end{pmatrix}, \text{ with} \quad (13)$$

$$X_{N_y \times N_y} = \begin{pmatrix} \ddots & B & 0 & 0 \\ B & A & B & 0 \\ 0 & B & A & B \\ 0 & 0 & B & \ddots \end{pmatrix} \text{ and } Y_{N_y \times N_y} = \text{diag}[C], \quad (14)$$

where $A = 4t + V(x, y)$, $B = C = -t$, with the potential energy at a given site $V(x, y)$. The transverse Hamiltonian has an even simpler structure.

$$H_y = \begin{pmatrix} \ddots & B & 0 & 0 \\ B & D & B & 0 \\ 0 & B & D & B \\ 0 & 0 & B & \ddots \end{pmatrix}, \quad (15)$$

where $D = 2t$. While a sparse implementation of the matrices would help, it is most efficient to create only $N_y \times N_y$ matrix blocks, when needed in the algorithm. Solving the Eigenproblem for H_y in step two yields the modes and wave functions for the self energies and has potential for optimizations. However, benchmarks have shown that the compute time is much smaller than for step four. Therefore, we do not focus on optimizations of the Eigensolver. The inversion in Eq. 9 executed in step four is most demanding and therefore the focus of our optimizations. We denote the basic implementation as Algorithm A shown in Fig. 2. Four of the eight multiplications of $N_y \times N_y$ matrices in step five to obtain Eq. 12 can be done in parallel, which is already ensured by the parallel nature of the LabVIEW development environment. The same is true for the second set of four multiplications that need the results of the first set. As the computational load of this part is small compared to the inversion no further optimizations have been done to this part.

Optimized linear algebra functions. The Multicore Analysis and Sparse Matrix Toolkit (MASMT) [12] provides Intel's Math Kernel Library's (MKL) [5] linear algebra functions in LabVIEW, optimized execution on multi-core processors and operation on large matrices. Replacing the matrix inversion with the corresponding MKL function led to Algorithm B, that however, still is limited by memory inefficiency due to the dense matrix representation as the benchmarks in Sect. 6 show.

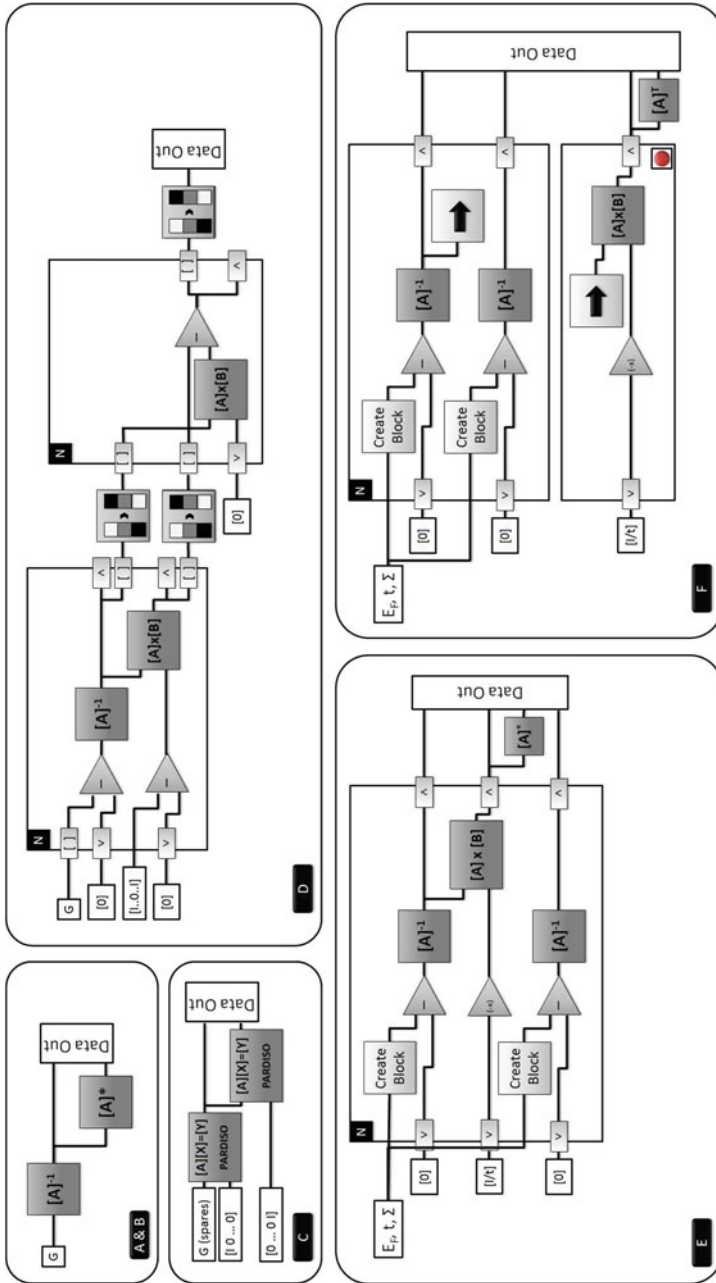


Fig. 2 Visualization of the different algorithms for the matrix inversion

Table 1 Summary of the benchmark results for the CPU-based algorithms

$N_{x,y}$	A (s)	B (s)	C (s)	D (s)	E (s)	F (s)
10	0.007	0.017	0.002	0.001	0.001	0.001
20	0.192	0.407	0.006	0.004	0.004	0.003
30	2.096	2.684	0.013	0.016	0.013	0.008
40	11.745	13.261	0.026	0.038	0.024	0.017
50	49.714	47.328	0.054	0.081	0.048	0.038
60	148.369	138.163	0.088	0.154	0.072	0.058
70	346.183	339.151	0.134	0.215	0.114	0.094
80	769.706	730.780	0.201	0.371	0.151	0.127
90	1,647.595	1,543.517	0.241	0.468	0.214	0.187
100	2,964.965	2,949.634	0.357	0.715	0.279	0.236
200	o.o.m.	o.o.m.	2.194	8.662	2.428	1.765
300	o.o.m.	o.o.m.	7.560	42.750	7.804	5.767
400	o.o.m.	o.o.m.	18.323	130.317	20.709	14.643
500	o.o.m.	o.o.m.	39.306	311.673	57.965	33.411
600	o.o.m.	o.o.m.	72.519	595.367	102.021	61.147
700	o.o.m.	o.o.m.	125.120	o.o.m.	168.005	109.006
800	o.o.m.	o.o.m.	o.o.m.	o.o.m.	263.918	191.874
900	o.o.m.	o.o.m.	o.o.m.	o.o.m.	389.083	297.420
1000	o.o.m.	o.o.m.	o.o.m.	o.o.m.	538.907	422.620

Version A is the original direct inversion algorithm. Version B uses the optimized LabVIEW high-performance computing libraries, Version C makes use of the matrices' sparsity, Version D is the first implementation of the block-tridiagonal solver, Version E is the optimized block-tridiagonal solver, and Version F is the optimized block-tridiagonal solver with pipelining for improved thread utilization (o.o.m. stands for out of memory—this benchmark could not be performed on the test machine).

Sparse matrices. Also part of MASMT is the PARDISO direct sparse linear solver [20–22] that we implemented as Algorithm C and that is faster than the dense solver, but still memory limited above $N_x = N_y = 700$ as the solver creates large intermediate data.

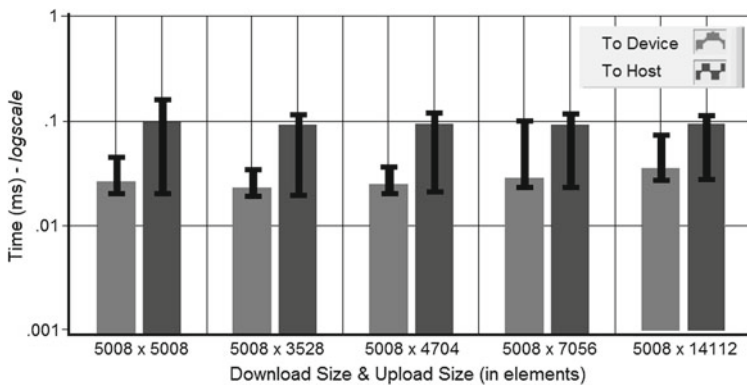
Block-Tridiagonal solver. Utilizing the block tri-diagonal structure of H we replaced the PARDISO solver in Algorithm D with the generalized Thomas algorithm [19]. Our block tri-diagonal linear system

$$\begin{bmatrix} A_1 & B_1 & & & & 0 \\ C_1 & A_2 & B_2 & & & \\ & C_2 & A_3 & \ddots & & \\ & & \ddots & \ddots & B_{N_x-1} & \\ 0 & & & C_{N_x-1} & A_{N_x} & \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_{N_x} \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_{N_x} \end{bmatrix}, \quad (16)$$

where A_k , B_k and C_k are all $N_y \times N_y$ blocks, can be computed in two steps. Step 1: for k from 1 to N_x

Table 2 Benchmark results for the GPU implementation of the pipelined and optimized block-tridiagonal matrix inversion solver

$N_{x,y}$	Matrixsize (elem.)	GPU pipelined BT solver (s)
128	16,384	2.463
256	65,536	0.691
384	147,456	2.936
512	262,144	8.887
640	409,600	21.255
768	589,824	43.610
896	802,816	80.244
1024	1,048,576	136.685
1280	1,638,400	332.707
1536	2,359,296	688.338
1792	3,211,264	1,272.800
2048	4,194,304	2,170.260
2560	6,553,600	5,290.440
3072	9,437,184	10,964.600
3584	12,845,056	20,297.700
4096	16,777,216	34,616.500
5120	26,214,400	84,462.700

**Fig. 3** Execution time for data transfers of single precision elements

$$\bar{B}_k = (A_k - C_{k-1}\bar{B}_{k-1})^{-1}B_k \quad (17)$$

$$\bar{Y}_k = (A_k - C_{k-1}\bar{B}_{k-1})^{-1}(Y_k - C_{k-1}\bar{Y}_{k-1}). \quad (18)$$

Step 2: for k from N_x to 1

$$X_k = \bar{Y}_k - \bar{B}_k X_{k+1}. \quad (19)$$

While the algorithm takes advantage of the sparse matrix structure it still requires quite large memory to store $3N_y N_y \times N_y$ matrices between step 1 and step 2. This results in about 48 GB RAM for $N_x = N_y = 1000$ forcing slow storage on hard disk media. Since only the four $N_y \times N_y$ corners of the inverse matrix are relevant for the transmission, we are solving two linear systems with the right hand sides

$$[I_{N_y} \ 0 \ \dots \ 0]', \text{ and } [0 \ \dots \ 0 \ I_{N_y}]', \quad (20)$$

where I_{N_y} is an $N_y \times N_y$ identity matrix, and only the first and last blocks are of interest. As the last block of each system is already computed after the first step in the Thomas algorithm, we propose another method to compute the first block. Denote

$$K = \begin{bmatrix} 0 & & & I_{N_y} \\ & & & I_{N_y} \\ & & \ddots & \\ & & & \\ I_{N_y} & & & 0 \end{bmatrix}. \quad (21)$$

where K satisfies $K^T = K$ and $K^2 = I$. Furthermore, if

$$A = \begin{bmatrix} A_1 & B_1 & & & 0 \\ C_1 & A_2 & B_2 & & \\ & C_2 & A_3 & \ddots & \\ & & \ddots & \ddots & B_{N_x-1} \\ 0 & & & C_{N_x-1} & A_{N_x} \end{bmatrix}, \text{ then} \quad (22)$$

$$KAK = \begin{bmatrix} A_{N_x} & B_{N_x-1} & & & 0 \\ C_{N_x-1} & A_{N_x-1} & B_{N_x-2} & & \\ & C_{N_x-2} & A_{N_x-2} & \ddots & \\ & & \ddots & \ddots & B_1 \\ 0 & & & C_1 & A_1 \end{bmatrix}. \quad (23)$$

Since $(KAK)^{-1} = KA^{-1}K$, the upper left(right) corner of A^{-1} is equal to the lower right(left) corner of $(KAK)^{-1}$ and the first step of Thomas algorithm with KAK gives the upper left(right) corner of A^{-1} . This new Algorithm E saves memory because the second step is omitted by the cost of an extra matrix inversion. As this additional inversion can be computed in parallel a significant performance gain can be seen in [Sect. 6](#) and the memory efficiency allows very large systems. For further performance gain on parallel architectures we pipelined sequential linear

algebra operations in Algorithm F and adjusted each group of operations to roughly the same complexity ensuring a constant high core utilization during the full inversion algorithm while at the same time keeping the memory usage below 3 GB for systems of up to $N_x = N_y = 1000$.

5 Implementation on GPUs

The optimized pipelined block-tridiagonal solver is still the most demanding part of the code. To further improve the performance we employed GPUs as they yield a high performance potential for the matrix multiplications and inversions of the algorithm. The rest of the code is executed exclusively on the host as it lacks computational complexity. For the implementation on the GPUs we used a prototype of the LabVIEW GPU Analysis Toolkit (LVGPU) [11]. The small memory footprint of the solver allows us to download the entire problem to the GPU and invoke the solver on the device, retrieving only the final result and thereby minimizing communication between the host and the GPU. Also multiple independent simulation steps (i.e. as part of a sweep of e.g. potential or Fermi Energy) could be executed if multiple GPUs are available. To control GPU devices from LabVIEW, LVGPU includes multiple interfaces for calling CUDA-based functions for execution on NVIDIA GPUs: LVCUDA encompasses functions based on the CUDA APIs for selecting and managing NVIDIA GPUs safely from the LabVIEW environment. LVCUBLAS and LVCUFFT functions call into CUBLAS and CUFFT available from NVIDIA's website [13, 14]. All GPU computations performed using LVGPU execute in parallel as if they were native LabVIEW operations. The current block-tridiagonal solver on the GPU uses LabVIEW and multi-core CPUs to orchestrate parallel execution of multiple LVGPU functions within an iteration of the solver. Results for this solver are consistent with performance gains achieved from LVGPU to solve other demanding problems. For example, an NVIDIA Tesla C2070 computes large FFTs (e.g. 512 K elements) in a few milliseconds including data transfers to and from the host. This meets real-time constraints for real-world problems such as quench detection of superconducting magnets. Performing an FFT on this extreme signal size overwhelms the core and cache size of present day multi-core CPUs. Experiments on multiple GPU devices were equally successful. The same GPU solver producing the benchmarks in Table 2 was deployed to two GPUs simultaneously. For a 1 K system size, the increase in execution time was 250 ms—adding only a fraction of the single GPU target time of 136.685 s. Distribution across multiple GPU devices is trivial when using LVGPU and potentially important to evolving algorithm variations such as Algorithm F from Fig. 2 as described in Sect. 7. To ensure that distribution to multiple GPU devices is feasible at multiple implementation levels, the jitter in data transfers to and from a GPU device was benchmarked. In Fig. 3, the mean execution time for I/O to (light bar) and from (dark bar) an NVIDIA

Tesla C1060 is shown. Error bars for each reflect the variation in transfer time recorded over several hundred iterations. Jitter on the order of $100\mu\text{s}$ affords implementations of the current solver which deploy GPU kernels to multiple devices at virtually any level.

6 Benchmarks

We ran code implementing the direct inversion of the Green's function matrix (Algorithm A) and the different optimizations (Algorithm B–F) on an IBM i-dataplex dx360 M3 [3] with two Intel Xeon X5650 six-core processors, running at 2.67 GHz, 48 GB RAM, and two NVIDIA Tesla M2050 GPUs with 3 GB RAM [17]. All code was written in LabVIEW 2011 using functionality from MASMT and LVGPU. Internally, MASMT called Intel's Math Kernel Library (MKL) v10.3 for execution on the CPU's multiple cores. LVGPU invokes routines from NVIDIA's CUDA Toolkit v4.0 [15]. Benchmarks were performed with LabVIEW 2011 64-Bit running under Windows 2008 Server Enterprise, with the NVIDIA Tesla GPU in TCC mode. Results from the CPU-based implementations are shown in Table 1. Results for the code in version F which executed primarily on GPUs are shown in Table 2. The results include just the execution of the inversion in Algorithm F. The initialization and postprocessing are not taken into account as they represent just a fraction of the computation time. However, the presented benchmarks include the time for transferring the data to and from the GPUs. To visualize the performance we summarize the results and show the number of system slices along the x -direction that can be simulated per hour on a single node or GPU in Fig. 4. These timings are dependent on the number of system sites in the y direction. This number gives a good description of the performance related to the system size and shows the applicability of our CPU and GPU-based implementations to systems with realistic dimensions. While the information is given for a two-dimensional system, where the transversal slice is one-dimensional, the same holds for three dimensional systems, where the number of sites is the product of height and width of the system.

7 Multiple GPU Implementation

For larger simulations and to further speed up the algorithm we present a scheme for multi-GPU distribution. The limited GPU-memory size can be circumvented by distributing the matrices over two or more GPUs. This in particular is convenient, when multiple operations on independent matrices can be executed in parallel as in our Algorithm F. Since CUDA-Version 4.0 multiple GPUs can share the same memory-environment allowing GPU-Memory addressing in a single namespace. Algorithm F consists of three threads where the third thread depends

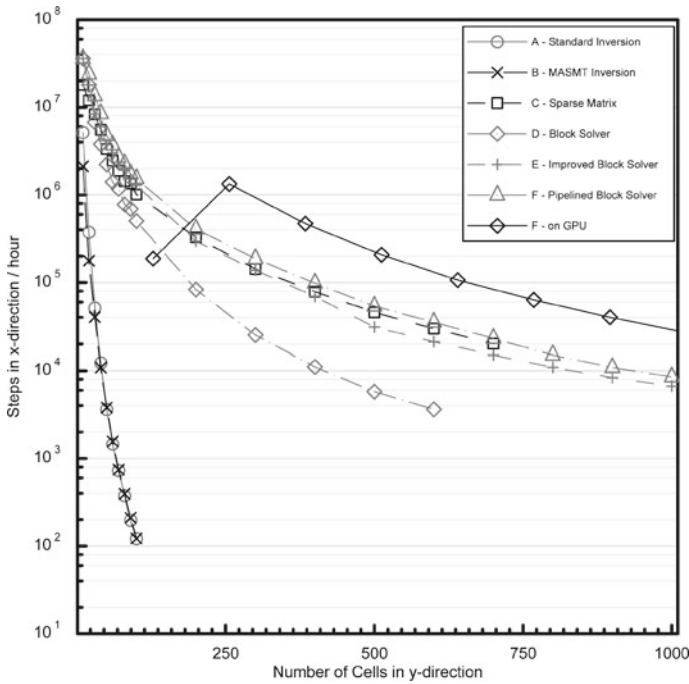


Fig. 4 Benchmark results in terms of simulation steps in x -direction per hour in dependence of the system size in y -direction

on previous results as shown in Fig. 5. One has to design a synchronized pipeline of these different threads to fully utilize the total available computing power.

While the third GPU is calculating the matrix multiplication, the first and second GPU can already compute their next step. A deployment on just two GPUs would lead to an asymmetric load: While both GPUs would start calculating the two inversions in parallel the GPU1 would have to calculate the multiplication in the end while the GPU2 would be idle. However, as all necessary results for the next set of inversions are already present, the GPU2 could take over the thread of inversion plus multiplication of the second step. As soon as the first multiplication is done the GPU1 starts calculating the other inversion of the second step. As this one does not require the final multiplication both GPUs would be in sync again at the end of the second step and the procedure would start over for the next step. As the CPU is not computing while the GPUs do, one can leverage the CPU-Power to compute the initial part of the next problem within a sequence (e.g. when varying a gate voltage, etc.) or cope with the communication. As the employed servers are equipped with two GPUs, the next logical step is to distribute the problem to several nodes. In this case multiple scenarios can be considered: First each of the three tasks in Fig. 5 could be run on a dedicated node with a hybrid CPU/multi-GPU implementation with algorithm as developed for example in the MAGMA

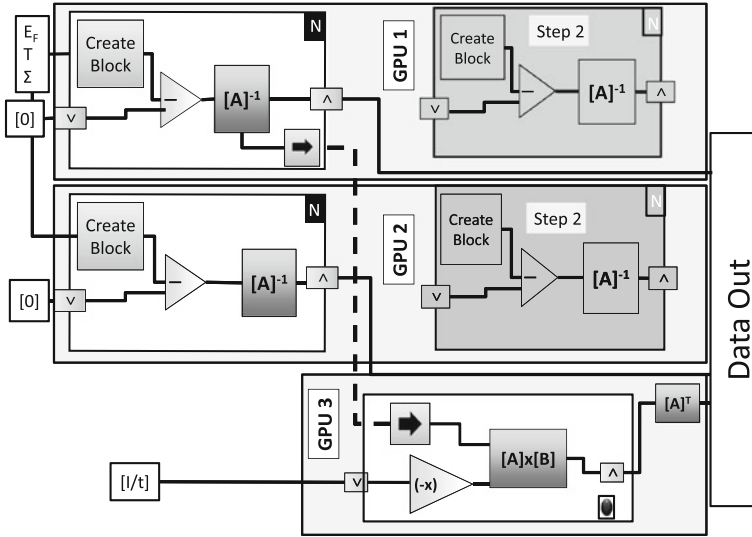


Fig. 5 Visualization of the multi-GPU setup

toolkit [10]. Second several steps of a parameter sweep can be distributed to several nodes as such a sweep is embarrassingly parallel. Third nodes could be specialized to distinct tasks: One node will create the Hamiltonian and solve the Eigenproblem and afterwards move the data to another set of nodes calculating the Green's function including the distribution of this subproblem to multiple nodes as described before. While the parallel execution of multiple simulations within a sweep is extremely easy, it is often desirable to receive results from a single step as fast as possible, giving the other two approaches additional priority. However, these two approaches will only be useful if a high-bandwidth-connection system like infiniband [4] is provided between the nodes as the information passed between the nodes would be matrices of size $(N_x N_y) \times (N_x N_y)$ in the first and still matrices of the size $N_x \times N_x$ in the third approach. To speed-up the transfer of data between GPUs on different nodes remote direct memory access RDMA, which will be available in the upcoming CUDA toolkit 5.0 release [16], will be very beneficial.

8 Conclusion

Simulations of transport in semiconductor nanostructures relying on the Green's function algorithm require—in their direct implementation—the inversion of gigantic matrices if realistic device sizes and sufficient small grid spacings are applied, making such a solution impractical. By exploiting the underlying sparse structure of the matrices we have presented an optimized implementation that

avoids the direct inversion by employing a block-diagonal solver to reduce the memory load from $(N_x N_y) \times (N_x N_y)$ matrices to $N_y \times N_y$ matrices. We enhanced the parallelism of the algorithm and balanced the computational load using pipelining to maximize the performance yield. The small memory footprint allows to implement the whole algorithm on a NVIDIA Tesla M2050 GPU without transferring data between the host and the GPU during the calculation. With the above summarized techniques *we were able to increase the system size by a factor of 100 compared to the primitive algorithm and even beyond on CPUs with another performance gain by a factor of three on the GPUs*. Taking into account the further parallelization approaches outlined in Sect. 7 that allow multi-GPU and multi-node implementations further significant performance gain can be achieved. The simulation of the transport in dependence on one varied parameter (e.g. gate voltage) with 1000 steps for a device of $1 \mu\text{m}$ by $1 \mu\text{m}$ and a grid spacing of 1 nm takes a total time of approximately 19 h. If further spread over multiple GPUs in multiple nodes, complete simulations can be obtained within very short time spans allowing almost just-in-time direct comparison of numeric results with experiments. Having demonstrated that it is feasible to perform such simulations in reasonable times, we are now exploring codes for three-dimensional structures with multiple bands contributing to the transport. While adding bands increases the matrix size to $(N_x N_y N_s) \times (N_x N_y N_s)$, where N_s is the number of bands, the addition of a third dimension increases the matrix size to $(N_y N_z) \times (N_y N_z)$ for each "slice" of the system letting the matrices again grow very fast. Therefore we will intensify the research on multi-GPU and multi-node implementations as well as on further optimizations of the algorithm.

Acknowledgments This work was supported by the Deutsche Forschungsgemeinschaft via GK 1286 and Me916/11-1, the City of Hamburg via the Center of Excellence "Nanospintronics", the Office of Naval Research via ONR-N00014110780, and the National Science Foundation by NSF-MRSEC DMR-0820414, NSFDMR-1105512, NHARP

References

1. Datta S (1999) Electronic transport in mesoscopic systems. Cambridge University Press, Cambridge
2. Datta S, Das B (1990) Appl Phys Lett 56(7):665
3. IBM. idataplex dx360 M3 datasheet. <http://www-03.ibm.com/systems/x/hardware/idataplex/dx360m3/index.html>
4. Infiniband TA. Introduction to Infiniband. http://members.infinibandta.org/kwspub/Intro_to_IB_for_End_Users.pdf
5. Intel. Intel Math Kernel Library. <http://software.intel.com/en-us/articles/intel-mkl/>
6. Jacob J, Lehmann H, Merkt U, Mehl S, Hankiewicz E (2011) DC-biased InAs spin-filter cascades. J Appl Phys 112:013706
7. Jacob J, Meier G, Peters S, Matsuyama T, Merkt U, Cummings AW, Akis R, Ferry DK (2009) Generation of highly spin-polarized currents in cascaded InAs spin filters. J Appl Phys 105:093714

8. Jacob J, Wenzel L, Schmidt D, Ruan Q, Amin V, Sinova J (2012) Numerical transport simulations in semiconductor nanostructures on CPUs and GPUs. Lecture notes in engineering and computer science: proceedings of the international multicongress of engineers and computer scientists 2012, IMECS
9. Koo HC, Kwon JH, Eom J, Chang J, Han SH, Johnson M (2009) Control of spin precession in a spin-injected field effect transistor. *Science* 325(5947):1515–1518
10. MAGMA. Magma—matrix algebra on gpu and multicore architectures. <http://icl.cs.utk.edu/magma/>
11. National instruments (2012) LabVIEW GPU Analysis Toolkit. beta version
12. National instruments (2012) LabVIEW multicore analysis and sparse matrix toolkit. <https://decibel.ni.com/content/docs/DOC-12086>
13. NVIDIA. CUDA BLAS implementation description. <http://developer.nvidia.com/cuBLAS>
14. NVIDIA. CUDA version 4.0 datasheet. <http://developer.nvidia.com/cuFFT>
15. NVIDIA. CUDA version 4.0 datasheet. <http://developer.nvidia.com/cuda-toolkit-40>
16. NVIDIA. CUDA version 5 RDMA feature. <http://developer.nvidia.com/gpudirect>
17. NVIDIA (2011) Tesla M2050 GPGPU datasheet
18. Oestreich M, Bender M, Hubner J, Hägele D, Rühle WW, Hartmann Th, Klar PJ, Heimbrodt W, Lampalzer M, Volz K, Stolz W (2002) Spin injection, spin transport and spin coherence. *Semicond Sci Technol* 17(4):285–297
19. Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1999) Numerical recipes in C, vol 123. Cambridge University Press, Cambridge, p 50
20. Schenk O, Bollhoefer M, Roemer R (2008) *SIAM Rev* 50:91–112
21. Schenk O, Waechter A, Hagemann M (2007) *J Comput Optim Appl* 36(2–3):321–341
22. Schenk O, Gärtner K (2004) *Journal of Future Generation Computer Systems* 20(3):475–487
23. Schmidt G, Ferrand D, Molenkamp LW, Filip AT, van Wees BJ (2000) Fundamental obstacle for electrical spin injection from a ferromagnetic metal into a diffusive semiconductor. *Phys Rev B* 62(8):R4790–R4793
24. Usuki T et al (1994) *Phys Rev B* 50:7615–7625
25. Wunderlich J, Park B-G, Irvine AC, Zarbo LP, Rozkotov E, Nemeč P, Novak V, Sinova J, Jungwirth T (2010) *Science* 330(6012):1801–1804

Performance Evaluation of Check-By-Voting for Colluding Attack in Volunteer Computing Systems

Kan Watanabe, Masaru Fukushi, Nobuo Funabiki
and Toru Nakanishi

Abstract Volunteer computing (VC) is a type of Internet based parallel computing paradigm, which allows any participants in the Internet to contribute their idle computing resources towards solving large problems. In VC, sabotage-tolerance mechanism is mandatory to improve the reliability of computations because some malicious participants may sabotage their jobs by returning incorrect results. Then, this paper evaluates current sabotage-tolerance methods such as voting and presents an impact of colluding attack on the sabotage-tolerance performance. The colluding attack is one of the most important issues to realize reliable modern VC systems because it may ruin the fundamental assumption of voting, i.e., “the majority is correct”. This becomes a big threat to the voting-based sabotage-tolerant methods which eliminate malicious participants based on whether their results are the majority or not. In this paper, we propose a colluding attack model, and conduct a detailed study for two major voting-based methods (i.e., *M*-first voting and check-by-voting) through a Monte Carlo simulation to reveal the sabotage-tolerance performance against colluding attacks. Our model introduces a colluding probability to cover various sabotaging and colluding scenarios. The

K. Watanabe (✉) · N. Funabiki · T. Nakanishi
Graduate School of Natural Science and Technology, Okayama University,
3-1-1 Tsushimanaka, Okayama 700-8530, Japan
e-mail: can@cne.okayama-u.ac.jp

N. Funabiki
e-mail: funabiki@cne.okayama-u.ac.jp

T. Nakanishi
e-mail: nakanisi@cne.okayama-u.ac.jp

M. Fukushi
Graduate School of Science and Engineering, Yamaguchi University,
2-16-1, Tokiwadai, Ube, Yamaguchi 755-8611, Japan
e-mail: mfukushi@yamaguchi-u.ac.jp

simulation study shows that check-by-voting works well even if colluding attacks happen, while M -first voting does not.

Keywords Check-by-voting · Colluding attacks · Desktop grids · Majority voting · Sabotage-tolerance · Volunteer computing

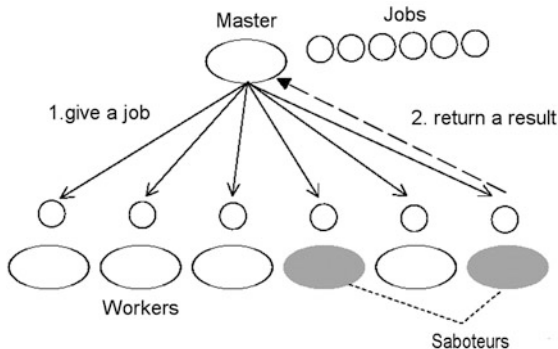
1 Introduction

Volunteer computing (VC) is a type of Internet based parallel computing paradigm, which allows any participants in the Internet to contribute their idle computing resources towards solving large problems in parallel. By making it easy for anyone on the Internet to join a computation, VC makes it possible to build very large and high performance global computing environment with a very low cost. The most popular example of VC is SETI@home [1]. It is currently employing hundreds of thousands of volunteer participants. Nowadays, VC has a major role in scientific computations in various areas such as Biology [2], Astronomy [3, 4], Physics [5], and Mathematics [6].

In VC, a management node (master) divides a computation into small independent jobs and allocates them to participant nodes (workers). Then, the workers execute the allocated jobs in parallel and return their results to the master. Different from the common grid computing systems where computation nodes are appropriately-managed by the system administrators, workers in VC systems may not be reliable since they are personal computers managed by volunteer participants. Those workers may behave erratically due to hardware/software failures or virus infections, or may behave maliciously to falsify the computation, each of which results in sabotage to the computation. Hence, sabotage-tolerance mechanism is mandatory for VC systems to improve the reliability of computations.

The basic mechanism for the sabotage-tolerance of VC is voting, which collects multiple candidate results for one job to compare the value. The collected candidates are classified into two groups, majority and minority groups. Usually, the result in the majority group is accepted as the final one because the majority seems to be correct, whereas results in the minority groups are eliminated as incorrect ones.

Two major voting-based sabotage-tolerant methods are M -first voting and check-by-voting. M -first voting simply collects M matching results for each job. Because of its simplicity, M -first voting is actually used in major VC middle-ware such as [7]. Check-by-voting [8–10] is the upgraded version of M -first voting, which introduces the idea of a weighted voting for performance improvement. Check-by-voting can guarantee that the error rate of computation ε is less than a given value ε_{acc} by determining the voting weight for the returned results appropriately. Both of these two voting methods follow a fundamental assumption that the majority is correct.

Fig. 1 Model of VC systems

However, the majority is not always correct as pointed out in [10, 11]. In the recent VC systems which allow worker-to-worker communication on a P2P communication network, all workers, including malicious ones (i.e., saboteurs), can communicate with each other. The saboteurs can generate incorrect results with the same values which may win the voting. This is called the colluding attack. Due to the capability of worker-to-worker communications, it is anticipated that such modern VC systems gain more popularity in distributing large-size jobs to mitigate the master's load or supporting more general parallel applications which need inter-process communication. Therefore, the need for the sabotage-tolerance of the colluding attacks has been grown.

In this paper, we explore the sabotage-tolerance performance of two voting-based methods, M -first voting and check-by-voting, against the colluding attacks. First, we propose a simple model for colluding attacks introducing a colluding probability c into the existing saboteurs' model [10, 12]. Although this model is simple, it covers the existing scenario (i.e., no colluding attack happens) and the worst case scenario (i.e., all saboteurs always collude), as a special case. Then, we conduct simulations of VC by changing the parameter of c from 0 to 1, and evaluate the throughput and error rate of the two methods to see the impact of colluding attacks on the reliability of VC systems. Also, we perform functional partitioning of check-by-voting and simulate the variant methods of check-by-voting to explore each function in details.

2 VC Model

The computation model of VC in this paper is the well-known work-pool-based master-worker model. Figure 1 illustrates the model. Details of the model are described as follows.

- A VC system consists of a management node (master) and W different participant nodes (workers).

- A computation to be executed in the VC system is divided into N independent jobs.
- The master gives a job to each idle worker in each time step. Each worker executes an allocated job and returns the result to the master.
- The computation finishes when the time step reaches a given deadline P or all N jobs finish.

In the modeling of workers' behaviors, the presence of saboteur is one of the most important components. To discuss the sabotage-tolerance problem of VC, Sarmenta [12] proposed the following sabotage model.

- A certain faulty fraction f of W workers is assumed to be saboteurs.
- Each saboteur is assumed to return an incorrect result with a constant probability s , which is known as the sabotage rate.
- The values of f and s are unknown to the master.

In VC systems with saboteurs, the performance is evaluated by two metrics, the throughput and the error rate. The throughput T is given as the number of the finished jobs when the computation reaches the final time step P . If the master accepts an incorrect result as the final one for a job, this job is called an incorrect job. The error rate ε is given as the ratio of incorrect jobs to finished T jobs. Using no sabotage-tolerance methods, each job is finished with one result whether the result is correct or not. If all W workers function at the same speed and execute a job in a unit time, T is given by $W \times P$ and ε is given by $W \times f \times s \times P/T = f \times s$. It is clear that ε is proportional to the number of saboteurs and sabotage rate s . Therefore, to reduce the error rate, some sabotage-tolerance methods must be used.

3 Sabotage-Tolerance Methods

3.1 Basic Approach

The most basic approach to the sabotage-tolerance of VC is to adopt some sort of redundant computation. Voting is a typical redundant computation technique, which has been widely used in many areas such as fault-tolerant computing.

Two major voting methods are M -first voting and check-by-voting. In these methods, each job is replicated and allocated to multiple workers so that the master can collect multiple results and compare their values for the majority decision. The results collected for a job are then classified into groups (called result groups) according to their value. The master decides which result group should be accepted as the final result through the voting.

In the remaining part of this section, we describe M -first voting and check-by-voting, which are target methods in our evaluation.

3.2 *M-First-Voting*

In *M*-first voting, the result group which collects *M* matching (the same value) results first is accepted as the final result. Because of its simplicity, *M*-first voting is used in major VC middle-wares such as BOINC [7]. However, *M*-first voting has a serious drawback on the performance because the master always needs to collect multiple (at least *M*) results for each job, whether each result is reliable or not. Compared to a VC system with no sabotage-tolerance method, *M*-first voting degrades the throughput less than $1/M$. Because the minimum number of *M* is 2, *M*-first voting always degrades the throughput less than half.

3.3 *Check-By-Voting*

Check-by-voting [8, 9] is the upgraded version of *M*-first voting, which introduces the idea of a weighted voting. In check-by-voting, the master gives a credibility value to each worker which represents how the worker is reliable. The master updates the credibility value after every voting, based on the workers' behavior in the computation, e.g., the number of being majority in voting. The credibility is used as its weight in voting so that the number of results needed to be collected for one job can be flexible. Note that, if the credibility of a worker is enough high, a job can be finished with just one result. This implies that the throughput of check-by-voting can get above the upper bound of the performance of *M*-first voting, i.e., the half throughput.

The problems of check-by-voting are how to check workers and how to calculate the proper value of the credibility. For these problems, the authors have developed the following technique and formula [8]. As a feature, the credibility in the developed formula is given as a conditional probability so that the expected value of the error rate ε is less than an arbitrary value ε_{acc} for any cases of *s* and *f*. This condition $\varepsilon \leq \varepsilon_{acc}$ is called reliability condition.

Checking technique is the method to check whether a worker is reliable or not in each voting. When a job is finished through a voting, the collected results are classified into two groups, i.e., the majority results and minority ones. In this technique, a worker *w* who returned majority results is regarded as "a reliable worker" and the parameter k_w is counted up, which represents how many times *w* becomes the majority ($k_w = 0$ at the start of the computation). On the other hand, workers who returned minority results are regarded as saboteurs and the following methods can be used as a penalty.

- Backtracking: the master backtracks (invalidates) all results returned by the worker because each might be an incorrect one.
- Blacklisting: the master puts the worker's identification information (e.g. IP address) on a blacklist to prevent the worker from returning results or getting any more jobs.

Using the above checking technique, a worker who returns correct results continuously gets larger k_w and gains credibility. The results returned from such reliable workers tend to be accepted, which leads to the reduction in the error rate. Also, after a worker gains enough credibility, the throughput of VC can be improved because the worker will produce some reliable results, which have high credibility enough to finish the jobs with just one result.

After checking the worker w , the credibility $C_W(w)$ is given based on the parameter k_w . Equation (1) shows $C_W(w)$ when all W workers execute a job in a unit time. Because s and f are unknown to the master, the credibility should be given by the upper bound for s and f . The parameter f_{max} is the assumed ratio of saboteurs and given by the master so that the condition $f \leq f_{max}$ is satisfied. Using $C_W(w)$, the credibility of other VC elements such as result group and job are also defined (see [8] for details).

$$C_W(w) = \begin{cases} 1 - f_{max} & \text{if } k_w = 0, \\ 1 - \frac{f_{max}}{1 - f_{max}} \times \max\left(\frac{1}{k_w e^{(1 - \epsilon_{acc})}}, \epsilon_{acc}^{k_w}\right) & \text{otherwise} \end{cases} \quad (1)$$

4 Colluding Attack

4.1 The Definition

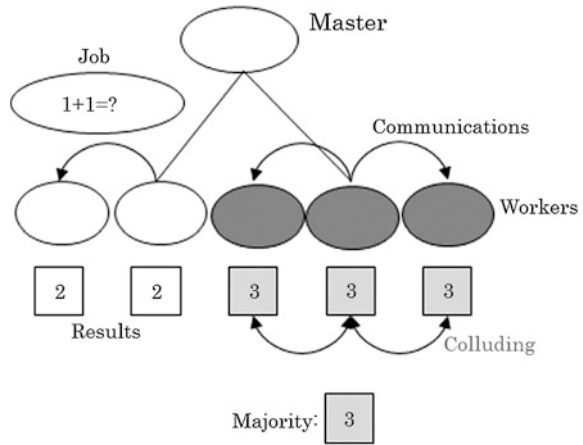
In the recent VCs, workers can communicate each other for several reasons [11]. For example, VC which distributes large-size jobs by a P2P communication technique requires worker-to-worker communications. Saboteurs in such VC also communicate with each other using this network. They can generate and return the incorrect results having the same value to increase the opportunity of being accepted.

In this paper, the colluding attack is defined as the attack to VC systems by returning incorrect results which have the same value. Figure 2 shows an example of colluding attack. For a job to find the answer for “ $1 + 1 = ?$ ”, saboteurs can communicate with each other and generate the same incorrect results “3”, while the correct result is “2”. In this case, the saboteurs’ results “3” becomes majority, which are to be accepted by the master. As shown in this case, the majority is not always correct if colluding attack happens.

4.2 Colluding Method

The type of colluding attacks is classified into the following three categories depending on the way to determine incorrect values.

Fig. 2 An example of colluding attack



- Direct colluding method: this method allows a saboteur to communicate with other saboteurs directly and determine the incorrect value for every sabotaging. In this method, the number of saboteurs in each colluding group tends to be small since each saboteur must know information on other saboteurs for direct communications.
- Indirect colluding method: this method extends the direct colluding method to enable to build an extensive colluding group of saboteurs. Instead of the direct communication, the saboteurs access to a third party, e.g., a colluding server, and obtain a value which should be set to their results. Thus, saboteurs can communicate with each other indirectly (via the colluding server) without knowing information on other saboteurs. This method has a possibility of allowing all saboteurs to collude together, and increasing the error rate of a computation dramatically.
- Accidental colluding method: this happens by accident. For example, the same incorrect values may be returned when a multiple-choice question is given as a job. This implies a colluding can happen depending on the type of jobs even if workers have no malicious intent.

4.3 Effect on Sabotage-Tolerance

The current sabotage-tolerance methods do not assume the presence of colluding attacks because they focus on the traditional VC, where there is no worker-to-worker communications, and incorrect values are random which do not match each other. If colluding happens, the performance metrics may be significantly affected by the following reasons.

- Error rate ϵ : the error rates of a computation will be increased since the incorrect results may become the majority and be accepted as final results. Especially, in

check-by-voting, such saboteurs may gain the credibility illegally by returning colluding results. This implies the occurrence of serious problem that the reliability condition $\varepsilon \leq \varepsilon_{\text{acc}}$ cannot be guaranteed. This is because the credibility formula (i.e., Eq. (1)) does not deal with the possibility of colluding attacks.

- Throughput T: the throughput of check-by-voting will be degraded since the correct results returned from non-saboteurs may be invalidated. This invalidation happens when non-saboteurs' results become the minority by the colluding attacks.

4.4 Model of Colluding Attack

To evaluate the effect of colluding attack shown in Sect. 4, they should be modeled. In this paper, we propose a colluding model based on the colluding rate c . The details of the proposed model are described in the followings:

- When a saboteur generates an incorrect result with the sabotage rate s , the saboteur performs colluding attack with the probability c and performs random attack with the probability $1 - c$.
- In case of the colluding attack, the saboteur generates a colluding result having a predetermined incorrect value.
- In case of the random attack, the saboteur generates an incorrect result having a random value.

Note that, a saboteur randomly generates a correct result with the probability $1 - s$.

The behavior of a saboteur is summarized as follows. For each job, a saboteur returns a colluding result with the probability sc , a random incorrect result with the probability $s(1 - c)$ and a correct result with the probability $1 - s$, respectively. This model covers any case of sabotaging and colluding. If $c = 0$, the computation model is the same as the existing one. On the other hand, if $c = 1$, the model generates the worst case, i.e., all saboteurs always collude and return the same incorrect results.

5 Simulation

5.1 Evaluation Method

The error rate ε and the throughput T are evaluated under the proposed colluding model. The proposed model allows to simulate any situation of VC by changing the colluding probability c from 0 to 1, including the existing VC model ($c = 0$). The target sabotage-tolerance methods in the evaluation are the following two.

Table 1 Simulation parameters

Computation deadline (P)	100
The number of workers (W)	100
The maximum number of jobs (N)	10000
The assumed ratio of saboteurs (f_{max})	0.35
The actual ratio of saboteurs (f)	f_{max}
Sabotage rate (s)	0–1
Acceptable error rate (ϵ_{acc})	0.01–0.05

- 2-first-voting: this is an instance of M -first voting where M is set to the minimum number 2. As M becomes larger, the throughput tends to be smaller since each job is replicated more. Thus, 2-first-voting shows the best throughput in M -first voting.
- Check-by-voting: this is an improved version of M -first voting shown in Sect. 4. As a job scheduling method, the same existing method is used as in [9].

5.2 Simulation Condition

Table 1 shows the simulation parameters in our evaluation.

The assumed ratio of saboteurs f_{max} is set to 0.35 based on the real VC experiments [13]. The simulation conditions are as follows. All W workers are assumed to have the same speed and execute a job in a unit time as in [9]. In check-by-voting, two penalties (i.e., backtracking and blacklisting) are employed to minority workers [9].

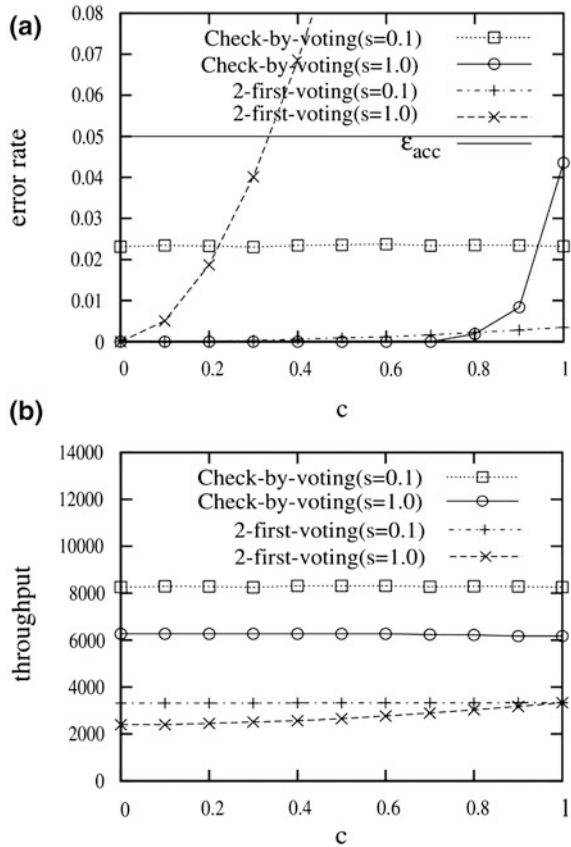
5.3 Simulation Results

5.3.1 Colluding Rate c

Figure 3 shows the error rate ϵ and the throughput T for colluding rate c . Figure 3a shows that the error rate of 2-first voting increases with the colluding rate c . Since larger c increases the probability of returning colluding results which tend to be the majority. Especially, in case of $s = 1$, the error rate exceeds the given parameter $\epsilon_{acc} = 0.05$ if c is greater than 0.4. This indicates that 2-first voting cannot guarantee the required reliability condition when saboteurs collude.

On the other hand, the error rate of check-by-voting is almost constant for any c (except $c \geq 0.8$ for $s = 1$), and is always smaller than the given ϵ_{acc} . To understand this result, remember the simulation conditions that all workers produce the same number of results and the ratio of incorrect results is $f \times s$. As long as f is smaller than 0.5, all incorrect results are less than half of all produced

Fig. 3 Error rate and throughput for colluding rate c ($\epsilon_{acc} = 0.05$). **a** Error rate. **b** Throughput

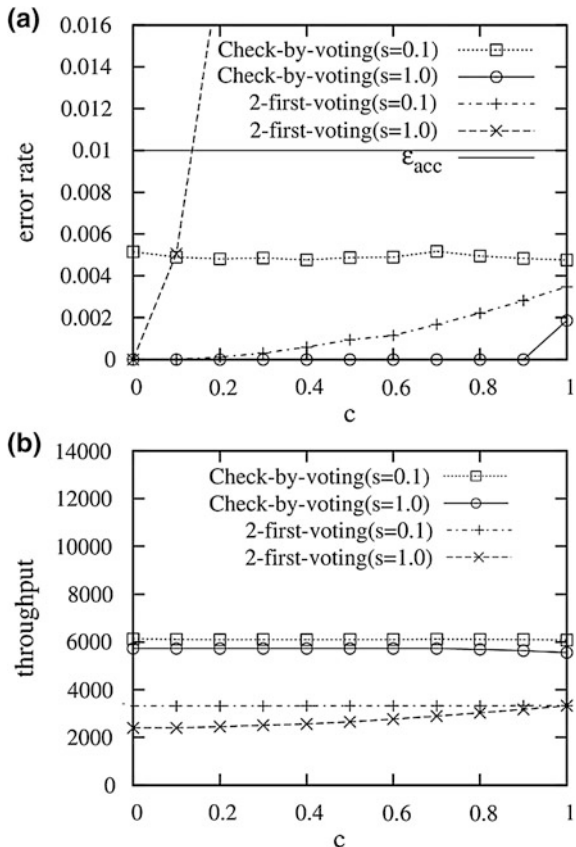


results, regardless of colluding. Even if some incorrect results returned by colluding saboteurs win the voting in some jobs, they cannot win in other jobs. Once a saboteur fails to win a voting, due to the penalty of backtracking, all results returned from the saboteur are invalidated. Then, the past all voting related to the invalidated results are re-evaluated, which gives the opportunity for correct results to win. This is the reason why the error rate of check-by-voting becomes smaller.

Figure 3b shows that the throughput of check-by-voting is almost constant for any c . This is because almost all saboteurs become the minority in any voting and take the penalty by blacklisting. In this situation, the remaining few saboteurs do not affect the throughput even if s and c are large. On the other hand, the throughput of 2-first voting depends on c (from 2,500 to 3,200 for $s = 1$) because all saboteurs survive until the end of the computation. Note that the throughput of 2-first voting is always less than $W \times P/2 = 5,000$ (i.e., upper-bound), while check-by-voting achieves over $T = 6,000$ and outperforms 2-first-voting for any c .

Figure 4 shows the simulation results in case of $\epsilon_{acc} = 0.01$. The result of 2-first-voting is the same as in Fig. 3 because 2-first-voting does not use ϵ_{acc} . As

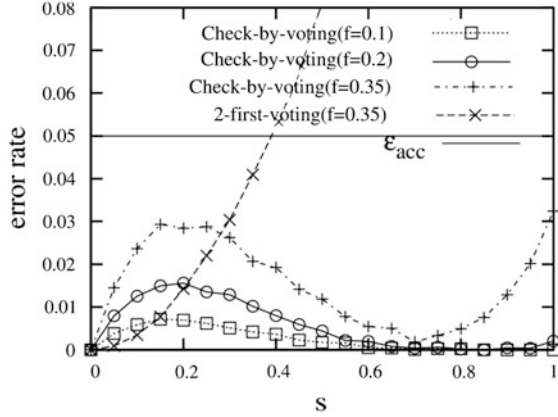
Fig. 4 Error rate and throughput for colluding rate c ($\epsilon_{acc} = 0.01$). **a** Error rate. **b** Throughput



shown in Figs. 3a and 4a, the error rate of check-by-voting changes depending on ϵ_{acc} to guarantee the reliability condition $\epsilon \leq \epsilon_{acc}$. For instance, for $s = 1$, $\epsilon = 0.023$ at $\epsilon_{acc} = 0.05$ and $\epsilon = 0.005$ at $\epsilon_{acc} = 0.01$. Also, as shown in Figs. 3b and 4b, the throughput of check-by-voting decreases depending on ϵ_{acc} . For instance, for $s = 1$, $T = 6,500$ at $\epsilon_{acc} = 0.05$ and $T = 5,800$ at $\epsilon_{acc} = 0.01$. These results imply that check-by-voting obtains less error rate to satisfy $\epsilon \leq \epsilon_{acc}$ at the expense of throughput. This is an important feature of check-by-voting as a sabotage-tolerant method.

In check-by-voting, the credibility changes depending on given ϵ_{acc} (Eq. (1)) and then the number of results to be collected (redundancy) changes depending on the credibility. Generally, if ϵ_{acc} becomes smaller, the credibility values become smaller and the required redundancy becomes larger to eliminate incorrect results through voting. On the other hand, in M -first-voting, the redundancy is always M regardless of ϵ_{acc} . These results indicate that check-by-voting can change the redundancy adaptively to guarantee the reliability condition for given ϵ_{acc} .

Fig. 5 Error rate for sabotage rate s ($\epsilon_{\text{acc}} = 0.05$, $c = 1.0$)



5.3.2 Sabotage Rate s

Figure 5 shows the error rate for the sabotage rate s . In this figure, the colluding rate c is set to 1 to assume the worst case. This figure shows that the error rate of check-by-voting has two local maximal values at $s = 0.2$ and $s = 1.0$. This result implies the difficulty of guaranteeing the reliability condition, which means the worst case is not always at $s = 1.0$. To guarantee $\epsilon \leq \epsilon_{\text{acc}}$ for any s , we must investigate the reason why these local maximal points arise.

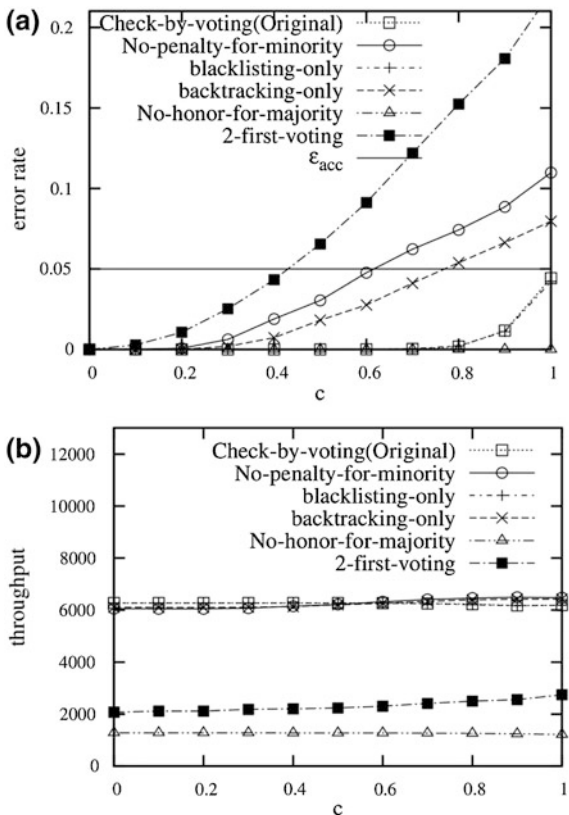
One possible reason is that increasing s has the two opposite effects for saboteurs; it increases the probability of producing incorrect results and also increases the probability of being detected as saboteurs in the checking technique. Thus, after s reaches a certain value ($s = 0.2$ in this case), the error rate becomes smaller due to the invalidation of incorrect results by backtracking. After s reaches 0.7, the ratio of incorrect results to all results for a job becomes large enough to win a voting; hence, the error rate increases again with s . The above speculation seems to be true in our simulation from the analysis of log data. However, the condition of arising such local minimal is not known and will be in our future work.

5.3.3 The Effect of Each Function of Check-By-Voting

As shown in Figs. 3, 4 and 5, check-by-voting works well even if saboteurs collude. We further explore the function of check-by-voting in details.

Check-by-voting is composed of several functions, counting up of parameter k_w , blacklisting and backtracking. In the original check-by-voting, when worker w wins a voting, the parameter k_w is counted up as a honor and two penalties, i.e., blacklisting and backtracking, are applied for loser. The variant methods of check-by-voting are used in the simulation as follows:

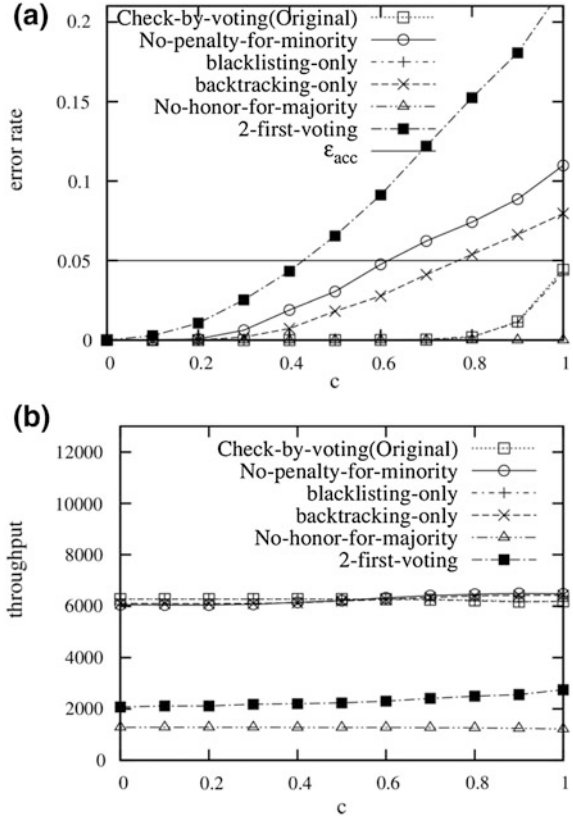
Fig. 6 The effect of each function of check-by-voting ($\epsilon_{acc} = 0.05$). **a** Error rate. **b** Throughput



- No-penalty-for-minority: losers take two penalties, that is, the results returned from the losers are valid and the losers continue to join the computation (no-backtracking and no-blacklisting).
- Blacklisting-only: the losers take only blacklisting (no-backtracking).
- Backtracking-only: the losers take only backtracking (no-blacklisting).
- No-honor-for-majority: even if w wins a voting, the parameter k_w does not counted up, while losers take two penalties.

Figures 6a and 7a show the errors for colluding rate c when $\epsilon_{acc} = 0.05$ and $\epsilon_{acc} = 0.01$, respectively. These figures show that the methods which guarantee the reliability condition always contain blacklisting (original check-by-voting, blacklisting-only and no-honor-for-majority). This result indicates that blacklisting is an essential function of check-by-voting. By comparing no-penalty-for-minority and backtracking-only, we can observe the effect of backtracking. For example, backtracking reduces the error rate from 0.11 to 0.08 at $c = 1$ in Fig. 6a. However the effect of backtracking is relatively small compared to blacklisting because once a worker takes the penalty of blacklisting, the worker will not produce any more

Fig. 7 The effect of each function of check-by-voting ($\epsilon_{\text{acc}} = 0.01$). **a** Error rate. **b** Throughput



results which will be invalidated by backtracking. Thus, the error rates of original check-by-voting and blacklisting-only (without backtracking) are almost the same.

Figures 6b and 7b show the throughputs for colluding rate c when $\epsilon_{\text{acc}} = 0.05$ and $\epsilon_{\text{acc}} = 0.01$, respectively. These figures show that the throughput of no-honor-for-majority is very small, and lower than that of 2-first-voting. Without using the parameter k_w and giving credibility to workers (no-honor-for-majority), check-by-voting is reduced to simple voting methods such as M -first voting. Also, two penalties, i.e., blacklisting and backtracking, eliminate workers and results from the system, each of which just degrades the throughput. These figures indicate that giving credibility based on k_w is also an essential function for check-by-voting to achieve higher performance.

Through these simulations, we have observed that the effect of backtracking is relatively small compared to that of blacklisting. On the other hand, when backtracking is enforced to a saboteur, re-evaluations of past all voting related to the saboteur is required. Such re-evaluations require a number of re-calculations of credibility and impose a strain on the master node. Thus, the variant methods, i.e., check-by-voting without backtracking, maybe available in the case that the master node does not have enough resources to perform original check-by-voting.

6 Conclusion and Future Work

In this paper, we have conducted Monte Carlo simulation-based performance evaluation for M -first voting and check-by-voting. For the evaluation, we have proposed a simple colluding model which introduces a colluding probability c into the existing saboteurs' model. The proposed model is general and it covers existing scenario (i.e., no colluding attack happens at $c = 0$) and worst case scenario (i.e., all saboteurs always collude at $c = 1$), as a special case. The simulation results have shown that check-by-voting guarantees the reliability condition $\varepsilon \leq \varepsilon_{\text{acc}}$ for any s and c , while achieving better throughput than 2-first voting.

There are still unknown parts of colluding. As shown in Fig. 5, the error rate of check-by-voting has two local maximal values for sabotage rate s . This result implies the difficulty of guaranteeing the reliability condition $\varepsilon \leq \varepsilon_{\text{acc}}$. We must investigate the reason why these local maximal points arise. One of our future works is analyzing these behaviors for guaranteeing the reliability condition in any case of saboteurs' behavior.

Acknowledgments This work was supported by KAKENHI 23800041.

References

1. SETI@home: a scientific experiment that uses Internet-connected computers in the Search for Extraterrestrial Intelligence (SETI). <http://setiathome.ssl.berkeley.edu/>
2. Rosetta@home: a scientific project that determines the 3-dimensional shapes of proteins. <http://boinc.bakerlab.org/rosetta/>
3. Cosmology@home: a project to search for the model that best describes our Universe and to find the range of models. <http://www.cosmologyathome.org/>
4. Knispel B et al (2010) Pulsar discovery by global volunteer computing. *Science* 329(5994):927–930
5. LHC@home: a project in theoretical physics computations for the Large Hadron Collider at CERN. <http://lhcatome2.cern.ch/test4theory/>
6. SAT@home: a research project that uses Internet-connected computers to solve hard and practically important problems. <http://sat.isa.ru/pdsat/>
7. BOINC: open-source software for volunteer computing. <http://boinc.berkeley.edu/>
8. Watanabe K, Fukushi M (2010) Generalized spot-checking for reliable volunteer computing. *IEICE Trans Inf Syst* E93-D(12):3164–3172
9. Watanabe K, Fukushi M, Kameyama M (2011) Adaptive group-based job scheduling for high performance and reliable volunteer computing. *J Inf Process* 19:39–51
10. Watanabe K, Funabiki N, Nakanishi T, Fukushi M (2012) Modeling and performance evaluation of colluding attack in volunteer computing systems. In: Proceedings of the international multiconference of engineers and computer scientists 2012, IMECS 2012, 14–16 March, 2012. Lecture notes in engineering and computer science, Hong Kong, pp 1658–1663
11. Araujo F, Farinha J, Domingues P, Silaghi GC, Kondo D (2011) A maximum independent set approach for collusion detection in voting pools. *J Parallel Distrib Comput* 71(10):1356–1366

12. Sarmenta LFG (2002) Sabotage-tolerance mechanisms for volunteer computing systems. *Futur Gener Comput Syst* 18(4):561–572
13. Kondo D, Araujo F, Malecot P, Domingues P, Silva LM, Fedak G, Cappello F (2007) Characterizing error rates in internet desktop grids. In: 13th European conference on parallel and distributed computing, pp 361–371

Fast Slot Planning Using Constraint-Based Local Search

Dario Pacino and Rune Møller Jensen

Abstract Due to the economic importance of stowage planning, there recently has been an increasing interest in developing optimization algorithms for this problem. We have developed a 2-phase approach that in most cases can generate near optimal stowage plans within a few 100 s for large deep-sea vessels. This paper describes the constraint-based local search algorithm used in the second phase of this approach where individual containers are assigned to slots in each bay section. The algorithm can solve this problem in an average of 0.18 s per bay, corresponding to a runtime of 20 s for the entire vessel. The algorithm has been validated on a benchmark suite of 133 industrial instances for which 86 % of the instances were solved to optimality.

Keywords Container · Stowage · Constraint · Local search · Slot planning · Optimization · Heuristic · Transportation

1 Introduction

Cost efficient and reliable container shipping is essential to today's global supply chains. In liner shipping, container vessels sail on a fixed route like a city bus. Once a vessel reaches a terminal, loading and unloading operations are performed

This research is sponsored in part by the Danish Maritime Fund under the BAYSTOW project.

D. Pacino (✉) · R. M. Jensen
IT-University of Copenhagen, Rued Langgaards Vej 7, 2300 Copenhagen, Denmark
e-mail: dpacino@itu.dk

R. M. Jensen
e-mail: rmj@itu.dk

by *quay cranes*. A key objective is to reduce time at port, and since the number of quay cranes available is fixed by contract, the main instrument for a liner shipper to reduce the vessel's port stay is to devise a time efficient *stowage plan*.

Stowage plans define where export containers should be stowed on the vessel. Such plans are, however, hard to produce in practice. First, they are made under time pressure by human stowage coordinators just hours before the vessel calls the port. Second, deep-sea vessels are large and often require thousands of container moves in a port. Third, complex interactions between low-level stacking rules and high-level stress limits and stability requirements make it difficult to minimize the makespan of cranes and, at the same time, avoid that containers block each other (*overstowage*). Finally, according to our industrial partner, runtimes of more than 10 min are impractical, since stowage coordinators possibly need to run several forecast scenarios.

A 2-phase hierarchical decomposition for stowage planning optimization is presented in Pacino et al. [16]. As depicted in Fig. 1, first the *multi-port master planning phase* distributes container groups into *locations* (subsections of the vessel) along the ship satisfying stress and stability requirements. Second, the *slot planning phase* solves the assignment problem of each location according to the master plan, thus generating a complete stowage plan. This method is able to generate phase-wise optimal stowage plans for 75 % of the tested instances in less than 400 s.

This chapter is a revised version of the work presented in Pacino and Jensen [15], and describes the Constraint Based Local Search (CBLS) algorithm that is used as a heuristic fallback of an exact Constraint Programming (CP) model [8, 9] to solve the slot planning phase. Slot planning first attempts to solve the instance with the CP model optimally within 1 s, should it fail (or should the instance be infeasible), the CBLS algorithm presented in this chapter is used to find high-quality solutions heuristically and can also handle infeasibilities by rolling out containers. A slot planning instance must be solved for each location in a vessel. Container vessels consist of up to 100 locations, and since master planing historically has been time consuming, we aim at solving each individual slot planning instance within 1 s in order to achieve total runtimes of less than 10 min as needed by the industry. We have evaluated the CBLS algorithm experimentally using 133 real instances provided by our industrial partner. We compare our results with the CP model presented in [9]. Each instance is solved in an average of 0.18 s. Thus, if only using the CBLS algorithm for slot planning, the total time needed for the slot planning phase is less than 20 s in average.

The remainder of the chapter is organized as follows. [Section 2](#) describes the problem. [Section 3](#) introduces related work. [Section 4](#) presents the approach taken. [Sections 5](#) and [6](#) present the experiments and draw conclusions.

2 Background and Problem Statement

A liner shipping vessel is a ship that transports box formed containers on a fixed, cyclical route. Containers typically have a width of 8', a length of either 20' or 40' and are 8.6' high. There exists, however, longer containers such as 45' and 50', and



Fig. 1 Hierarchical decomposition of stowage planning into master and slot planning

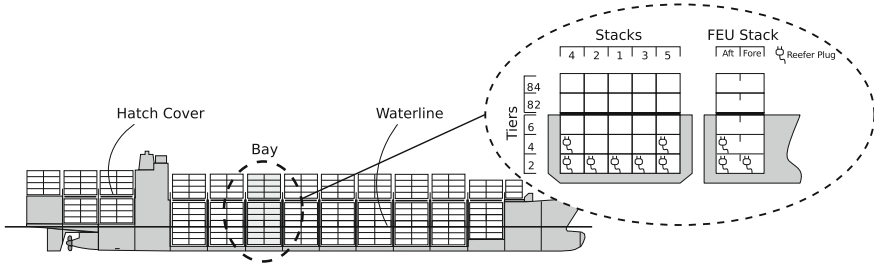


Fig. 2 The arrangement of bays in a small container vessel

higher 40’ containers (called *high-cubes*) that are 1 foot taller. Some containers are refrigerated (*reefers*) and require access to special power plugs. Other special types of containers are pallet-wide containers, where a standard European pallet can be stored, and IMO containers, which are used to store dangerous goods. In addition, there are out-of-gauge (OOG) containers with cargo sticking out in the top or at the side (e.g., a yacht) and non-containerized break-bulk like wind turbine blades.

The cargo space of a vessel is composed of a number of *bays*, which are a collection of container *stacks* along the length of the ship. Each bay is divided into an *on-deck* and *below-deck* part by a *hatch cover*, which is a flat, watertight structure that prevents the vessel from taking in water. An overview of a vessel layout is shown in Fig. 2.

Figure 2 also shows how each stack is composed of two Twenty foot Equivalent Unit (TEU) stacks and one Forty foot Equivalent Unit (FEU) stack, which hold vertically arranged *cells* indexed by *tiers*. The TEU stack cells are composed of two *slots*, which are the physical positioning of a 20’ container. The *aft* slot toward the stern and the *fore* slot toward the bow of the vessel. Some of the cells have access to power plugs.

The loading and unloading of containers is carried out by quay cranes that can access containers on top of the stacks. The primary objective of stowage planning is to minimize port stay. An important secondary objective is to generate stowage plans that are robust to changes in the cargo forecast.

When a set of containers to stow in a bay has been decided, the positioning of those containers has limited interference with containers in other bays. For this reason, it is natural to divide the constraints and objectives of the stowage planning problem into high-level inter bay constraints and objectives and low-level intra bay constraints and objectives.

High-level constraints mainly consider the stability of the vessel as defined by its trim, metacentric height, and stress moments such as shear, bending and torsion.

In addition, the weight and volume capacity limits as well as capacity limits of the different container types must be satisfied. High-level objectives include the minimization of the crane makespan and of the overstowage between on and below deck.

Since this chapter focuses on slot planning, we describe the low-level constraints and objectives in more details. Low-level constraints are mainly stacking rules. Containers must be stowed, forming stacks, taking into consideration that two 20' containers cannot be stowed on top of a 40' container. Each stack has a maximum allowed height and weight which cannot be exceeded. Each cell can have restrictions regarding what kind of containers it can hold. Some cells can be, for example, reserved for only 40' containers. Reefer containers require power and can only be stowed in cells with access to special power plugs. Dangerous goods (IMO containers) must be stowed following predefined security patterns, while OOG containers can only be stowed where sufficient space is available. On deck, special rules also apply due to the use of lashing rods. On deck container weight must decrease with stack height. Wind also imposes special stacking patterns on deck, and line of sight rules restrict the stack heights. Low-level objectives reflect rules of thumb from stowage coordinators in order to produce stowage plans that are robust to changes in forecasted demands. The objectives include maximizing the number of empty stacks, clustering of containers' discharge port in stacks, minimizing the number of reefer slots used for non-reefer containers, and minimizing overstowage between containers in the same stack.

Taking all industrial constraints and objectives into account is desirable, however, it makes an academic study of the problem impractical. To that end, we have defined a representative problem with our industrial partner where the number of constraints and objectives is reduced, but the structural complexity is kept. Here we give a formal definition of this representative problem using the Integer Programming (IP) model given in Delgado et al. [9]. Given a set of stacks S and the sets of tiers per stacks T_s of the locations for slot planning, we indicate a cell by a pair of indexes $s \in S$ and $t \in T_s$. Let $x_{stc} \in \{0, 1\}$ be a decision variable indicating if the cells indexed by $s \in S$ and $t \in T_s$ contains the container $c \in C$, where C is the set of all containers to be stowed in the location.

$$\frac{1}{2} \sum_{c \in C^{20}} x_{s(t-1)c} + \sum_{c \in C^{40}} x_{s(t-1)c} - \sum_{c \in C^{40}} x_{stc} \geq 0 \quad \forall s \in S, t \in T_s \quad (1)$$

$$\sum_{c \in C^{20}} x_{s(t-1)c} + \sum_{c \in C^{40}} x_{stc} \leq 1 \quad \forall s \in S, t \in T_s \quad (2)$$

$$\frac{1}{2} \sum_{c \in C^{20}} x_{stc} + \sum_{c \in C^{40}} x_{stc} \leq 1 \quad \forall s \in S, t \in T_s \quad (3)$$

$$x_{stc} = 1 \quad \forall (s, t, c) \in P \quad (4)$$

$$\sum_{s \in S} \sum_{t \in T} x_{stc} = 1 \quad \forall c \in C \quad (5)$$

$$\sum_{c' \in C^{20}} x_{stc'} - 2x_{stc} \geq 0 \quad \forall s \in S, t \in T_s, c \in C^{20} \quad (6)$$

$$\sum_{c \in C} R_c x_{stc} - R_{st} \leq 0 \quad \forall s \in S, t \in T_s \quad (7)$$

$$\sum_{t \in T_s} \sum_{c \in C} W_c x_{stc} \leq \mathcal{W}_s \quad \forall s \in S \quad (8)$$

$$\sum_{t \in T_s} \left(\frac{1}{2} \sum_{c \in C^{20}} H_c x_{stc} + \sum_{c \in C^{40}} H_c x_{stc} \right) \leq \mathcal{H}_s \quad \forall s \in S \quad (9)$$

$$\sum_{t'=1}^{t-1} \sum_{d'=2}^{d-1} \sum_{c \in C} A_{cd'} x_{st'c} - 2(t-1)\delta_{std} \leq 0 \quad \forall s \in S, t \in T_s, d \in D \quad (10)$$

$$A_{cd} x_{stc} + \delta_{std} - o_c \leq 1 \quad \forall s \in S, t \in T_s, c \in C \quad (11)$$

$$e_s - x_{stc} \geq 0 \quad \forall s \in S, t \in T_s, c \in C \quad (12)$$

$$p_{sd} - A_{cd} x_{stc} \geq 0 \quad \forall s \in S, t \in T, d \in D, c \in C \quad (13)$$

Constraints (1) and (2), where $C^{20} \subseteq C$ and $C^{40} \subseteq C$ are respectively the set of 20' and 40' containers, forces containers to form valid stacks where 20' containers cannot be stowed on top of 40' ones. Constraint (3) ensures that either 20' or 40' containers can be stowed in a cell, but not both at the same time. The set P of containers already onboard is composed of triples (s, t, c) indicating that container $c \in C$ is stowed in stack $s \in S$ and tier $t \in T_s$. Constraint (4) forces those containers to their preassigned cell. Cells holding 20' containers must be synchronized, meaning that they cannot hold only one 20' container. Such constraints are handled by inequality (6). Constraint (7), where $R_c \in \{0, 1\}$ indicates if container $c \in C$ is a reefer container, and R_{st} holds the reefer capacity of a cell, enforcing the capacity of reefer cells. Given W_c as the weight of container $c \in C$, constraint (8) limits the weight of a stack $s \in S$ to not exceed the maximum weight \mathcal{W}_s . Similarly, the height limit is enforced by constraint (9), where H_c indicates the height of container $c \in C$ and \mathcal{H}_s is the maximum height of stack $s \in S$. Constraint (10) defines the indicator variable δ_{std} which indicates if a container below the cell in $s \in S$ and $t \in T_s$ is to be unloaded before port $d \in D$, where D is the set of

discharge ports. The constant $A_{cd} \in \{0, 1\}$ indicates if container $c \in C$ must be discharged at port $d \in D$. Constraint (11) then uses the δ_{std} variable to count overstocking into the variable o_c . The number of non-empty stacks is stored in the variable e_s via the constraint (12). The number of discharge ports used within a stack is then calculated into the variable p_{sd} in constraint (13). An optimal solution to slot planning minimizes the following objective:

$$\begin{aligned} & C^O \sum_{c \in C} o_c + C^P \sum_{s \in S} \sum_{d \in D} p_{sd} + C^S \sum_{s \in S} e_s \\ & + C^R \sum_{s \in S} \sum_{t \in T_s} \left(\mathcal{R}_{st} \sum_{c \in C^{40}} x_{stc} (1 - R_c) + \sum_{c \in C^{20}} x_{stc} \left(\frac{1}{2} \mathcal{R}_{st} - R_c \right) \right) \end{aligned} \quad (14)$$

where the cost weights C^O, C^C, C^S and C^R represent the priority given by the stowage coordinators to each objective: overstocking, the number of different discharge ports in a stack, the number of non-empty stacks and the number of reefer cells used by non-reefer container.

3 Literature Survey

The number of publications on stowage planning has grown substantially within the last few years. Contributions can be divided into two main categories: single-phase and multi-phase approaches. Multi-phase approaches decompose the problem hierarchically. 2-phase [12, 16, 20, 22] and 3-phase approaches [1, 21] are currently the most successful in terms of model accuracy and scalability. Single-phase approaches represent the stowage planning problem (or parts of it) in a single optimization model. Approaches applied include IP [3, 6, 11, 13], CP [2, 8], GA [7, 10], placement heuristics [4], 3D-packing [18], simulation [5], and case-based methods [14]. Slot planning models and algorithms in the work above, however, either do not solve sufficiently representative versions of the problem or lack experimental evaluation. In Delgado et al. [8, 9] a CP approach is used to solve the representative problem described in this chapter. The CP model was shown to greatly outperform both the IP model given in this chapter and a column generation approach.

4 Solution Approach

As mentioned in Sect. 1, slot planning is part of a larger decomposition approach. Since high-level constraints and objectives are already solved during the master planning phase, our hypothesis is that the generated slot planning instances are under-constrained, and as such, high-quality solutions can be efficiently found using heuristic methods. We thus propose a placement heuristic that relaxes a subset of the constraints. Feasible solutions are then found and improved using CBLS.

4.1 Constraint Based Local Search (CBLs)

CBLs is an architecture for combinatorial local search algorithms based on constraint satisfaction concepts. The constraint satisfaction part of a combinatorial optimization problem is transformed into an optimization problem where the objective is to minimize constraint violations.

At the core of the architecture we find the concept of *invariants* or one-way constraints. Invariants are represented by *incremental variables* which express a relationship that must be maintained once a new assignment is made to the decision variables. Consider for example $v = \sum_{i=1}^n x_i$, where x_i is a decision variable and v is an incremental variable. The variable v is an invariant, and each time a variable x_i assumes a new value, v must be updated accordingly.

Once invariants are available it is natural to create compound objects that maintain properties incrementally. Examples of those are constraints and objectives maintaining properties such as satisfiability, violations and the contribution of each variable to the violation. In order to incrementally maintain the system, it is enough to implement an incremental algorithm for the basic invariants based on the neighborhood operator of choice. These objects can now be used to define formal models of combinatorial problems, and their properties (such as violations) can be used to define search strategies.

We refer the reader to Van Hentenryck and Michel [19] for a in-depth description of the architecture and the concepts behind CBLs.

4.2 A Logic Based Model

Given the set $\mathcal{P} \in \{1, 2\}$ where $1 = Aft$ and $2 = Fore$, we propose a logic based model where we define the decision variable $x_{stp} \in C \cup \{\perp\}$ to be the assignment of container $c \in C$ to the slot in stack $s \in S$, tier $t \in T_s$ and position $p \in \mathcal{P}$, or the empty assignment \perp :

$$\begin{aligned} & \forall s \in S, t \in T_s. \neg f(x_{st1}) \wedge (f(x_{st2}) \Rightarrow \perp(x_{st1})) \\ \forall s \in S, t \in T_s \setminus \{1\}, p \in \mathcal{P}. \neg \perp(x_{stp}) & \Rightarrow \end{aligned} \quad (15)$$

$$(t(x_{s(t-1)1}) \wedge t(x_{s(t-1)2})) \vee f(x_{s(t-1)1}) \quad (16)$$

$$\forall s \in S, t \in T_s \setminus \{N_s^T\}, p \in \mathcal{P}. f(x_{st1}) \Rightarrow \neg t(x_{s(t+1)p}) \quad (17)$$

$$\forall c \in C. |\{x_{stp} = c \mid s \in S, t \in T_s, p \in \mathcal{P}\}| = 1 \quad (18)$$

$$\forall c \in C^P. x_{s_c t_c p_c} = c \quad (19)$$

$$\forall s \in S, t \in T_s, p \in \mathcal{P}. r(x_{stp}) \wedge t(x_{stp}) \Rightarrow \mathcal{R}_{stp} \quad (20)$$

$$\forall s \in S, t \in T_s. r(x_{st1}) \wedge f(x_{st1}) \Rightarrow \mathcal{R}_{st1} \vee \mathcal{R}_{st2} \quad (21)$$

$$\forall s \in S. \sum_{t \in T_s} (w(x_{st1}) + w(x_{st2})) \leq \mathcal{W}_s \quad (22)$$

$$\forall s \in S. \sum_{t \in T_s} \max(h(x_{st1}), h(x_{st2})) \leq \mathcal{H}_s. \quad (23)$$

The assignment convention for the 40' containers is maintained by constraint (15), where $f(c)$ indicates if container $c \in C \cup \{\perp\}$ is a 40' container and $\perp(c)$ indicates it is an empty assignment ($c = \perp$). Physical support from below is guaranteed to all containers via constraint (16), where $t(c)$ is true if container $c \in C \cup \{\perp\}$ is a 20' container. Stacking rules between 20' and 40' containers are modelled by constraint (17), where N_s^T is the index of the top tier in stack $s \in S$. Each container is assigned to exactly one slot (18). Given a set of release containers C^P constraint (19) ensures the original assignment (s_c, t_c, p_c) is maintained. Reefer slots are indicated with the constant $\mathcal{R}_{stp} = true$, and 20' reefer containers are forced to be stowed in those slots by constraint (20), while (21) ensures that 40' reefer containers are stowed in cells where either one of the two slots has a power-plug. Reefer containers are identified by the function $r(c)$. Stacks weights and heights are maintained within limits by constraints (22) and (23), where $w(c)$ and $h(c)$ are respectively the weight and height of container $c \in C \cup \{\perp\}$.

An optimal solution to a slot planning problem optimizes the following weighted sum:

$$\min \quad C^O o_{os} + C^C o_{ps} + C^S o_{us} + C^R o_{ur} \quad (24)$$

which is composed by the following objective components:

$$o_{os} = \sum_{s \in S} \sum_{t \in T_s} \sum_{p \in \mathcal{P}} o_{stp} \quad (25)$$

$$o_{ur} = \sum_{s \in S} \sum_{t \in T_s} \sum_{p \in \mathcal{P}} ur_{stp} \quad (26)$$

$$o_{ps} = \sum_{s \in S} | \{d(x_{stp}) \mid t \in T_s, p \in \mathcal{P}, \neg \perp(x_{stp})\} | \quad (27)$$

$$o_{us} = \sum_{s \in S} us_s \quad (28)$$

In (25) one unit cost is counted for each container that is overstacking another one below in the stack, where o_{stp} defines whether a container $c \in C$ in stack $s \in S$, tier $t \in T_s$ and slot position $p \in \mathcal{P}$ overstacks another container in the stack. Thus, $o_{stp} = 1$ if $\neg \perp(x_{stp})$ and there exists a tier $t' \in \{t_s, \dots, t-1\}$ below t with an overstacked container $d(x_{st'p}) < d(x_{stp})$ ($d(c)$ returns the discharge port of container c), otherwise $o_{stp} = 0$. Objective (26) counts one unit cost for each misused reefer slot, where $ur_{stp} = 1$ if \mathcal{R}_{stp} and $(f(x_{st1}) \wedge \neg r(x_{st1}) \vee t(x_{stp}) \wedge \neg r(x_{stp}))$, and 0 otherwise. Notice that a 40' non-reefer container will add a unit cost for each reefer slot it covers. In order to favor stacks stowing containers with the same port of destination, one unit cost is counted for each discharge port present in the stack (27). In order to minimize the number of used stacks, one unit cost for each stack used is counted by (28), where $us_s = 1$ if there exists a $t \in T_s$ and $p \in \mathcal{P}$ such that $\neg \perp(x_{stp})$, and 0 otherwise.

The algorithm uses a neighborhood generated by swapping containers. A swap is an exchange of some containers between a pair of cells. Formally, a swap γ is a pair of tuples $\gamma = (\langle s, t, c \rangle, \langle s', t', c' \rangle)$ where the containers c in the cell at stack s and tier t exchange positions with the containers c' in the cell at stack s' and tier t' . The sets c and c' can contain at most two containers.

The constraints of the model are defined in terms of violations. Violations are numerical evaluations of how far the current solution is from satisfying the constraint under consideration. Once a constraint has no violation it is considered satisfied, thus a solution where all constraint have no violations is a feasible solution. Let π_{stp} be the value of the decision variables for assignment π . Taking constraint (1) as an example, the violations for a single slot can be defined as $v_{stp}^\pi = \sum_{t'=0}^{t-1} \neg(t(\pi_{stp} \implies \neg f(\pi_{st'Aft}))$. For simplicity we consider boolean variables as $\{0, 1\}$ and allow arithmetic operations over them. The total violation for constraint (1) is thus defined as $\sum_{s \in S} \sum_{t \in T} \sum_{p \in \mathcal{P}} v_{stp}^\pi$. Violations are also used during the search as heuristic guidance, as it will soon be shown in the description of the search algorithm.

Objectives are also defined in terms of violations like the constraints. A solution minimizing objective violations is optimizing the objective function. Violations for the overstackage objective (11) are, for example, defined for each slot as $o_{stp}^\pi = \exists t' \in \{0, \dots, t-1\}, p' \in \mathcal{P}. dsp(pi_{stp}) > dsp(\pi_{st'p'})$ where $dsp(c)$ indicates the discharge port of container c . Thus, the total number of overstacking containers is $\sum_{s \in S} \sum_{t \in T} \sum_{p \in \mathcal{P}} o_{stp}^\pi$. Other constraints and objectives are defined in a similar fashion. A comprehensive definition is found in Pacino and Jensen [17].

4.3 Search Strategy

Constraint satisfaction and objective optimization tend to drive the search in opposite directions, ultimately generating a poor heuristic. For this reason we decided to split the main search into a feasibility and optimality phase. We start

with a greedy initial solution, generated by relaxing the stack height and weight constraint. Containers are then stowed using the lexicographical order (reefers \prec discharge port \prec 20' container), designed to optimize the objective function. Containers that cannot be stowed, due to the constraints, are then sequentially stowed at the end of the procedure. This *placement heuristic* produces high quality search states that are slightly infeasible. Such states become the input of the *feasibility phase* where only the violation of the constraints are minimized. Feasible solutions are then passed to the *optimality phase* which will now only consider feasible neighborhoods and optimize the objective function. Both the feasibility and optimality phases use a *min/max heuristic* where swaps are chosen by selecting a slot with the maximum number of violations and swapping it with a slot producing the minimum violations (in other words, the most improving swap). This is how violations become the essential building blocks of the search heuristic.

Consider the following algorithm:

```

CBLSP_SlotPlanning ():
1:  $\pi = \text{placementHeuristic}()$                                 /*Placement Heuristic*/
2: while ( $\neg \text{satisfy}(\text{constraints})$ )
3:    $\pi' \leftarrow \pi$                                           /*Feasibility phase*/
4:   select  $s_1 \leftarrow \text{most violated slot}$ 
5:     select  $s_2 \leftarrow \text{most improving slot to swap}$ 
6:        $\pi \leftarrow \text{swap}(s_1, s_2)$ 
7:     end select
8:   end select
9:   if ( $\pi' = \pi$ ) perform side move
10  end while
11: while (there is objective improvement)
12:    $\pi' \leftarrow \pi$                                           /*Optimality phase*/
13:   select  $s_1 \leftarrow \text{most objective violating slot}$ 
14:     select  $s_2 \leftarrow \text{most objective improving feasible slot to swap}$ 
15:        $\pi \leftarrow \text{swap}(s_1, s_2)$ 
16:     end select
17:   end select
18:   if ( $\pi' = \pi$ )  $\pi \leftarrow \text{perform tie-breaking swap on } \pi$ 
19: end while
20: return  $\pi$ 

```

The placement heuristic is used in line 1 to generate an initial solution π . Lines 2–10 are the implementation of the feasibility phase. While the constraints are not satisfied (line 2), we perform swaps (line 6) selecting the most violated slot (line 4) and the most improving slot (line 5). If the swap generates a non-improving

(but not worse) state, we accept the state as a side move. Side moves are allowed only after a number of failed improving moves.

Lines 11–19 are the implementation of the optimality phase. With π now being a feasible solution, we perform feasible swaps (line 12) so long as an objective improvement can be generated. Similarly to the feasibility phase, lines 13–15 select the most violating and most improving slots. The selection, however, is limited to feasible swaps (swaps that do not generate any violations on the constraints). Should the swap generate a non-improving solution, a limited number of side moves are allowed where solutions are chosen using tie-breaking rules (line 18). For the overstockage objective (11) tie-breaking is defined as the number of containers overstocked by each container, which will make the algorithm choose a solution a container that only overstocks a single container rather than many containers over one. For the non-empty stack objective (12) the tie is broken by the number of containers in the stack, so that the search rather swaps containers that are in almost empty stacks. The clustering objective (13) calculates the tie-breaker by summing the quadratic product of the number of containers with the same discharge port, which will favor those stacks that have the most containers with the same discharge port.

4.4 Incremental Computations

In order to efficiently compute and evaluate the neighborhoods, we designed incremental calculations for each constraint and objective. Incremental calculations (or delta evaluations) allow the algorithm to evaluate and apply swaps efficiently without having to recompute all of the constraint and objective violations. A delta evaluation is based on a simple construction/destruction principle, where the violations of the original assignment are first removed from the total violation, are then reinserted according to the new assignment. A delta evaluation can be made volatile, and thus only be used as an evaluation, or be made persistent if the change has to be applied to the solution.

Consider a single $20'$ to $20'$ swap $\gamma = (\langle s, t, c \rangle, \langle s', t', c' \rangle)$ between two distinct stacks for the overstockage objective. Starting from stack s we first remove the known violations of container c . Then we look at all the containers stowed above it, removing one violation for each of the ones that only overstock container c . Now all the violations connected to container c are removed. Swapping c with c' , we now have to do an opposite operation, where we add 1 violation to all container that only overstock c' in stack s at tier t and 1 violation if c' is overstocking any container below it. The same operation is then performed on the other stack s' where c and c' exchange roles. It is easy to show that given the correct data structures such operations can be performed linearly with the size of the tiers. More details and formal definitions for all incremental evaluations can be found in Pacino and Jensen [17].

Table 1 Test set characteristics

Class	40'	20'	Reefer	HC	DSP > 1	Inst.
1	✓					6
2		✓				18
3	✓	✓				4
4	✓			✓		42
5	✓	✓		✓		27
6	✓		✓	✓		8
7	✓	✓	✓	✓		7
8	✓			✓	✓	7
9	✓	✓		✓	✓	10
10	✓		✓	✓	✓	2
11	✓	✓	✓	✓	✓	2

The first column is an instance class ID. Column 2, 3, 4, and 5 indicate whether 40-foot, 20-foot, reefer, and high-cube containers are present. Column 6 indicates whether more than one discharge port is present. Finally, column 7 is the number of instances of the class

5 Computational Results

The algorithm has been implemented in C++ and all experiments have been conducted on a Linux system with 8 Gb RAM and 2 Opteron quad core 1.7 GHz CPUs, each with 2Mb of cache. The slot planning instances have been derived from real stowage plans used by our industrial collaborator for deep-sea vessels. We use test data set of 133 instances, including locations between 6 TEUs and 220 TEUs. Table 1 gives a summarized overview of these instances. Notice that many instances only have one discharge port. This is a feature of the instances generated after a multi-port master planning phase, thus it reflects the actual problem we are facing. We report average results over ten runs per instance where parallel random restarts of the algorithm are used to achieve more robust results. The experimental results on the test data set, shown in Table 2a, show the percentage of solutions solved within a specific optimality gap (optimal solutions have been generated using the CP algorithm described in Delgado et al. [9]). It is easy to see that only few instances diverge from near optimality and that in 86 % of the cases the algorithm actually reached the optimal solution.

Studying the algorithm performance closer, we were able to gain some insight on the quality of the different phases. The heuristic placement does not take the height and weight constraint into account, leaving space for improvements. However Table 2c, d show that the solution found by the heuristic placement is not far from being feasible in most of the cases. This is supported by the fact that often feasibility is reached within 20 iterations and that 61 % of the time, the objective value is not compromised. The quality of the objective value of the first feasible solution is also optimal in 74 % of the cases (Table 2b), suggesting that the heuristic placement procedure is performing well. These results also support our hypothesis that slot planning instances generated by the master planning phase are under-constrained.

Table 2 Algorithm analysis

Opt. gap ^a		Opt. gap (feas.) ^b		Feas. iter. ^c		Feas. worse ^d	
Gap. (%)	Freq. (%)	Gap. (%)	Freq. (%)	It. ≤	(%)	Obj (%)	(%)
0	86	0	74	0	29	-20	2
1	2	5	6	5	23	-10	2
2	2	10	2	10	18	-5	4
3	2	15	5	15	8	0	61
4	1	20	3	20	6	5	8
10	1	25	3	25	3	10	2
15	4	35	3	30	6	20	7
20	1	40	1	35	1	> 20	15
30	2	> 100	3	40	2		
				45	2		
				50	1		
				65	2		

^aCost gap between returned solution and optimal solution

^bCost gap between first feasible solution and optimal solution

^cNumber of iterations needed to find the first feasible solution

^dWorsening of the cost of the heuristic placement when searching for the first feasible solution

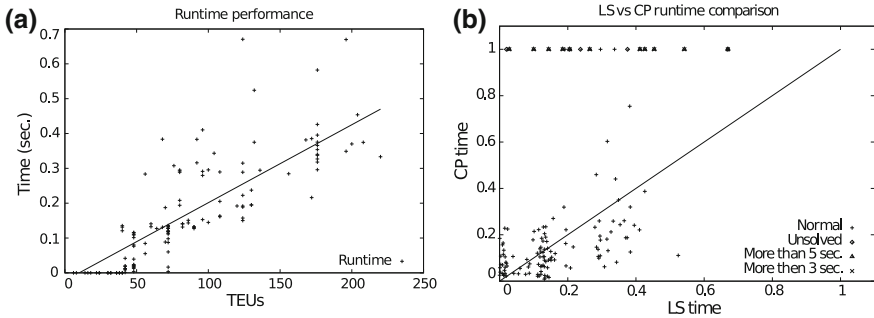


Fig. 3 **a** Execution time as a function of instance size measured in TEUs. **b** Execution time comparison between our algorithm and the complete constraint programming approach used to generate optimal solutions

The average runtime of the instances is 0.18 s, with a worst case of 0.65 s. Figure 3a shows the runtime of the algorithm as a function of the size of the instance measured in TEUs. As depicted, the execution time scales well with the instance size. Figure 3b compares the execution time between our algorithm and the exact CP approach used for generating optimal solutions.

As depicted, the CP approach is highly competitive within the set of instances that it can solve in 1 s. However, our approach can solve the problematic instances for CP fast as well. This result indicates that the under-constrained nature of the problem might force exact methods to spend an excessive amount of time proving optimality. A pragmatic solution would be to execute the two approaches in

parallel and rely on the optimal CP solution for the locations where it can be generated fast.

6 Conclusion

We have developed a representative slot planning model together with a major liner shipping company and implemented a CBLs algorithm. Our experimental results show that these problems are easy in practice even though the general problem is NP-Complete. Our work shows that when these problems are combinatorially hard in practice, the main challenge is to distribute containers to locations in vessel bays rather than assigning them to individual slots in these bays.

References

1. Ambrosino D, Anghinolfi D, Paolucci M, Sciomachen A (2010) An experimental comparison of different heuristics for the master bay plan problem. In: Proceedings of the 9th international symposium on experimental algorithms, pp 314–325
2. Ambrosino D, Sciomachen A (1998) A constraint satisfaction approach for master bay plans. *Marit Eng Ports* 36:175–184
3. Ambrosino D, Sciomachen A (2003) Impact of yard organization on the master bay planning problem. *Marit Econ Logist* 5:285–300
4. Avriel M, Penn M, Shpirer N, Witteboon S (1998) Stowage planning for container ships to reduce the number of shifts. *Ann Oper Res* 76:55–71
5. Aye WC, Low MYH, Ying HS, Jing HW, Min Z (2010) Visualization and simulation tool for automated stowage plan generation system. In: Proceedings of the international multicongference of engineers and computer scientists (IMECS 2010), vol 2. Hong Kong, pp 1013–1019
6. Botter RC, Brinati MA (1992) Stowage container planning: a model for getting an optimal solution. In: Proceedings of the 7th international conference on computer applications in the automation of shipyard operation and ship design, pp 217–229
7. Davidor Y, Avihail M (1996) A method for determining a vessel stowage plan. Patent Publication WO9735266
8. Delgado A, Jensen RM, Schulte C (2009) Generating optimal stowage plans for container vessel bays. In: Proceedings of the 15th international conference on principles and practice of constraint programming (CP-09), LNCS series, vol 5732, pp 6–20
9. Delgado A, Jensen RM, Janstrup K, Rose TH, Andersen KH (2011) A constraint programming model for fast optimal stowage of container vessel bays. *Eur J Oper Res* (accepted for publication)
10. Dubrovsky O, Levitin G, Michal P (2002) A genetic algorithm with a compact solution encoding for the container ship stowage problem. *J Heuristics* 8:585–599
11. Giemesch P, Jellinghaus A (2003) Optimization models for the containership stowage problem. In: Proceedings of the international conference of the german operations research society
12. Kang JG, Kim YD (2002) Stowage planning in maritime container transportation. *J Oper Res Soc* 53(4):415–426

13. Li F, Tian C, Cao R, Ding W (2008) An integer programming for container stowage problem. In: Proceedings of the international conference on computational science, Part I, LNCS, vol 5101. Springer, pp 853–862
14. Nugroho S (2004) Case-based stowage planning for container ships. In: The international logistics congress
15. Pacino D, Jensen RM (2012) Constraint-based local search for container stowage slot planning. In: Proceedings of the international multicongference of engineers and computer scientists (IMECS 2012). Hong Kong, pp 1467–1472
16. Pacino D, Delgado A, Jensen RM, Bebbington T (2011) Fast generation of near-optimal plans for eco-efficient stowage of large container vessels. In: ICCL, pp 286–301
17. Pacino D, Jensen RM (2010) A 3-phase randomized constraint based local search algorithm for stowing under deck locations of container vessel bays. Technical report TR-2010-123, IT-University of Copenhagen
18. Sciomachen A, Tanfani A (2003) The master bay plan problem: a solution method based on its connection to the three-dimensional bin packing problem. *IMA J Manag Math* 14:251–269
19. Van Hentenryck P, Michel L (2009) Constraint-based local search. The MIT Press, Cambridge
20. Wilson ID, Roach P (1999) Principles of combinatorial optimization applied to container-ship stowage planning. *J Heuristics* 5:403–418
21. Yoke M, Low H, Xiao X, Liu F, Huang SY, Hsu, WJ, Li Z (2009) An automated stowage planning system for large containerships. In: Proceedings of the 4th virtual international conference on intelligent production machines and systems
22. Zhang W-Y, Lin Y, Ji Z-S (2005) Model and algorithm for container ship stowage planning based on bin-packing problem. *J Mar Sci Appl* 4(3):1269

Sequencing of Arrival Aircraft with Operational Constraints

Adriana Andreeva-Mori, Shinji Suzuki and Eri Itoh

Abstract At most airports around the world landings are assigned based on the first come, first served rule. Limited airport capacity and environmental issues have called for better sequencing strategies. New technologies can be helpful, but hardware approaches require significant time and resources before the actual implementation. This research investigates possible fuel savings obtained only by re-sequencing of arrival aircraft. This novel re-sequencing method consists of a set of rules obtained after analyzing the statistical performance of several suggested guidelines. It is concluded that refining only the landing sequencing operations can lead to average fuel savings of 2.8 % of the total fuel consumed during the descent.

Keywords Arrival sequence · Descent · Fuel saving · Guidelines · Optimization · Terminal area

A. Andreeva-Mori (✉) · S. Suzuki · E. Itoh
Department of Aeronautics and Astronautics, University of Tokyo, 7-3-1 Hongo,
Bunkyo-ku, Tokyo 113-8656, Japan
e-mail: adiha2002@yahoo.com

S. Suzuki
e-mail: tshinji@mail.ecc.u-tokyo.ac.jp

E. Itoh
Air Traffic Management Department, Electronic Navigation Research Institute (ENRI),
7-42-23, Jindaiji-Higashi, Chofu, Tokyo 182-0012, Japan
e-mail: eri@enri.go.jp

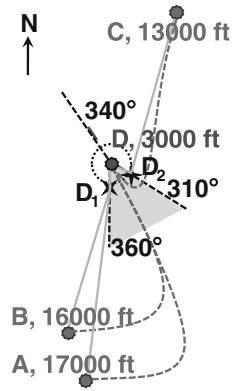
1 Introduction

In 1903 the Wright brothers invented and built the first airplane [1] and since then aviation has transformed into a major means of transportation. By year 2030, Airbus [2] forecasts an annual airline traffic growth of 4.8 %, while according to Boeing [3] this figure is 5.1 %. If realized, these forecasts will see air traffic double every fifteen years. Such a growth will result in numerous economic and social benefits but it would also put some burden on the environment because of the greenhouse gasses, CO₂ in particular, emitted due to increasing fuel burn. Furthermore, fuel burn is also accountable for a substantial part of the airline's costs. According to IATA, fuel has been the largest single cost item for the airline industry representing about 32.3 % of all operating costs in 2010 [4].

Numerous measures which reduce aircraft fuel burn have already been taken. Advanced materials used in new airplanes have made aircraft lighter so that less fuel is burnt per passenger per kilometer. New engines have made it possible to make the most of the propulsion technology improvements. These two examples require that the airline purchase new aircraft, though, so their implementation and introduction into service requires both time and financial resources. As promising as the new technologies might be, we believe that there is enough room for improvement by optimizing operations only, both ground and airborne in air traffic management. Such a strategy requires less investment on the airline side and can be beneficial for all parties involved. A lot of research work has been done on fuel saving aircraft scheduling. The sequencing problem can be described as a job scheduling problem. Mak and Sun [5] consider cranes scheduling problem and develop a new genetic algorithm which successfully solves small-scale problems with sufficient accuracy but might be impractical for large-scale problems. Balakrishnan and Chandran [6] implement constrained position shifting to the aircraft sequencing problem and look for a real-time application. These conventional approaches suggest implementations of new automation tools which generate optimal scheduling. They, however, do not provide any user-friendly rules and procedures to air traffic controllers who are the people in charge of sequencing operation. In this research, we take a new sequencing approach aiming to provide sequence assignment rules to air traffic controllers.

Here, a new re-sequencing method for arrival air traffic descending to an airport is proposed. Every flight can be divided into several stages: taxing, take-off, climb, cruise, descent, final approach, landing and taxing. In general, climb is performed at close to maximum thrust so that the aircraft leaves the vicinity of the airport as soon as possible. This is done because the constantly increasing air traffic poses pressure to the airports. Cruising is performed at a flight speed close to the optimal. Descent, on the other hand, is thought to be the stage which still requires a lot of optimization as it allows for changes without jeopardizing the safety of the flight within the scope of the technologies available at present. Therefore, we focus on the descent stage, whose control is in the hands of air traffic controllers at the terminal area. Our research considers the most efficient plausible arrival sequence

Fig. 1 Terminal area assumptions. Aircraft are usually navigated along the dotted lines connecting waypoints A, B or C with D, but RNAV would allow more efficient descents like the ones shown in solid lines



in terms of combined fuel burnt by all aircraft involved. This sequence is subject to operational constraints such as minimum separation and available arrival time window.

This paper is organized as follows: [Section 2](#) briefly explains our early studies on optimizing descent trajectories for arrival air traffic at Tokyo International Airport. In [Sect. 3](#) an overview of the sequencing problem and the questions that need to be addressed are discussed. [Section 4](#) describes the novel re-sequencing approach and evaluates it by numerical simulations without available time window constraints, with these being introduced in the [Sect. 5](#). Here, various sequencing rules and their effect on fuel burn are also examined. The research is summarized in [Sect. 6](#) with some discussions and conclusions.

2 Terminal Area and Single Aircraft Trajectory Optimization

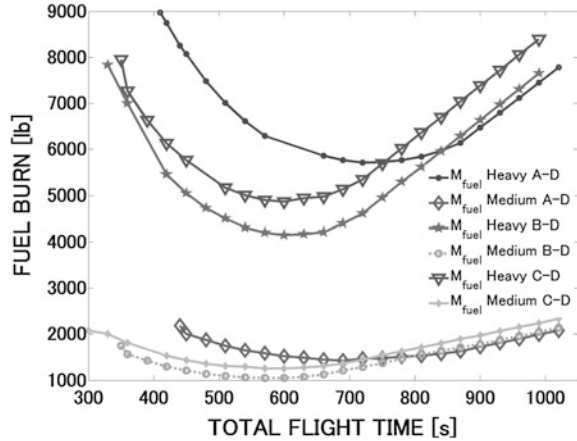
In our past studies [7–9], optimal descent trajectories with time constraints were investigated through numerical simulations. The considered terminal area is a model of Tokyo International Airport with the air traffic management operations which used to be executed until recently (Fig. 1). Actually, operations were changed but for the purposes of this research the past model is considered sufficient.

The assumed coordinates and waypoint altitudes are shown in Table 1. After the aircraft enter the terminal area at one of the three waypoints A, B or C, air traffic controllers have to merge the traffic coming from south and north in the terminal area while keeping the separation minima, which will be explained in detail later in Section III. The landing sequence is also decided at this point, as usually no reordering occurs once the aircraft is directed to the final approach waypoint D where it is transferred to the tower air traffic control and considered out of the scope of the terminal control.

Table 1 Waypoint coordinates

Waypoint	Distance east (nm)	Distance north (nm)	Altitude (ft)
A	-5.96	-52.50	17,000
B	-11.68	-40.10	16,000
C	15.99	36.81	13,000
D	0	0	3,000

Fig. 2 Fuel burn versus descent time. Around the optimal descent time the fuel burn can be adequately approximated by a quadratic function



Optimizations were performed with the sequential quadratic programming method (SQP) and included two types of aircraft- Boeing 737, a short-to-medium range airplane with a maximum take-off weight of 180,000 lb and a standard seating capacity of 137; and Boeing 747, an aircraft with a maximum take-off weight of 875,000 lb and standard seating capacity of 366.

Suppose the optimal flight time for minimum fuel burn is the expected time of arrival (ETA) which would have been the arrival time had the aircraft followed their optimal descent profile in the best time without taking into account the other aircraft. However, not all aircraft can follow their optimal descent profile all the time due to other traffic, weather conditions or after-landing issues. The minimum fuel burn under certain descent time constraints is shown in Fig. 2.

The aircraft was required to be either early (negative time shift) or late (positive time shift). It was proven that close to the optimal descent time the fuel burn changes can be described by a quadratic function with a maximum deviation of 3.4 lb.

$$f = a(t - t_{opt})^2$$

where f is the fuel burn increase, a is a parameter related to the entry waypoint and aircraft type and t_{opt} is the optimal flight time and t -the actual flight time. More on the simulated terminal area and flight procedures for single aircraft descent trajectories can be found in our previous work [7, 8].

Table 2 ICAO separation minima

Lead	Heavy	Follower medium	Light
Heavy (nm)	4	5	6
Medium (nm)	3	3	5
Light (nm)	3	3	3

3 Sequencing Problem Formulation

3.1 Minimum Aircraft Separation

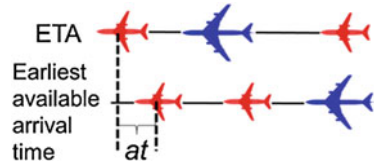
The optimal aircraft sequencing depends greatly on the minimum aircraft separation, which is in turn determined by the strength of wake vortex formed behind aircraft. Each aircraft creates turbulence behind itself as it moves through the air. This turbulence is referred to as wake turbulence and its main component is the turbulence generated by the wingtip vortices. The region behind the aircraft is where wake turbulence occurs. The strength of the turbulence is directly proportional to the weight, wingspan and speed of the aircraft [10]. The main factor determining the wake vortex is the weight, though, i.e. heavier aircraft produce stronger vortices than lighter ones. Aircraft are categorized in three groups based on their maximum take-off weight (MTOW).

- Heavy aircraft types of 136,000 kg (300,000 lb) or more;
- Medium aircraft types less than 136,000 kg (300,000 lb) and more than 7,000 kg (15,500 lb);
- Light aircraft types of 7,000 kg (15,500 lb) or less; and

Considering the nature at which vortex turbulence diminishes, ICAO has set minimum distance separation [11]. It should be applied when the difference between the altitudes of the pair of aircraft is less than 1,000 ft and depends on the aircraft maximum take-off weight. The separation minimum standard is shown in Table 2.

The aircraft considered in this research are Boeing 737 and Boeing 747, which fall into the medium and heavy categories, respectively. We have assumed a certain velocity at the terminal area exit waypoint for all aircraft, which allowed us to convert the distance separation requirement into a time separation requirement. In our simulations the separation standard is simplified, i.e. when a medium aircraft follows a heavy one, the required separation was considered to be 90 s, and in all other case- 60 s. When the difference between ETA is not enough, the aircraft should be shifted forward or backward.

Fig. 3 Available arrival time



3.2 Available Arrival Time

Because only one aircraft can land or depart from a runway at the same time, and because aircraft must be separated by a certain time interval to avoid collisions, every airport has a finite capacity; it can only safely handle so many aircraft per hour. This capacity depends on many factors, such as the number of runways available, layout of taxi tracks, availability of air traffic control, but also on current or anticipated weather. Especially the weather can cause large variations in capacity because strong winds may limit the number of runways available, and poor visibility may necessitate increases in separation between aircraft. Air traffic control can also be limiting, there are only so many aircraft an air traffic control unit can safely handle. Staff shortages, radar maintenance or equipment faults can lower the capacity of a unit. This can affect both airport air traffic control as well as en-route air traffic control centers.

These issues reflect the aircraft sequence greatly. To describe them mathematically, we consider an available arrival time window defined by the earliest and latest arrival time at which aircraft is allowed to cross the terminal area exit waypoint. The latest arrival time is determined by the fuel available onboard and possibly by any subjective constraints induced by the airlines. In this research, however, the latest available time constraint is always weaker than the earliest available time constraint, because we are trying to not only achieve minimum fuel burn, but also have as short arrival time of the last aircraft as possible, thus maximizing the airport's capacity. Therefore, in the rest of the paper by available arrival time (AAT) we will mean the earliest available arrival time (Fig. 3).

3.3 Problem Formulation

The nature of the sequencing problem imposes a lot of constraints. As air traffic control is human-centered, we aim at developing an easy-to-implement procedure which will reduce the total fuel consumed even if it will not make it minimum. In other words, a trade-off between simplicity and fuel savings is acceptable to a certain extent.

Throughout this research, two questions should be answered:

- (1) What is the optimal sequencing? How is it different from the first come, first served sequence?
- (2) How are the ETAs changed to achieve this optimal sequence?

4 Scenario Simulations Without Available Arrival Time Constraints

4.1 Batches of Three Aircraft

First, simulations without available arrival time constraints were considered, i.e. the first aircraft was allowed to arrive as early or as late as suited and the rest of the aircraft in the batch were to keep the separation minimum.

Consider three aircraft entering the terminal area. The first series of our calculations confirmed our expectations that if all aircraft are of the same type, they should be shifted equally to provide sufficient separation. Without loss of generality, we considered the ETA of one aircraft to be zero and allowed for negative as well as positive ETA of the other two aircraft. The rest of the simulations were divided in two groups- (A) two heavy (indices 1 and 2) and a medium aircraft (index 3, $ETA_3 = 0$) and (B) two medium (indices 1 and 2) and a heavy aircraft (index 3, $ETA_3 = 0$).

For three aircraft there are six possible sequences. The horizontal axis shows the difference between the estimated arrival times of aircraft 1 and aircraft 3. If $ETA_2 - ETA_3$ is negative, aircraft 2 arrives earlier than aircraft 3. Therefore, below the horizontal axis the sequence of aircraft 2 and 3 is always 23. Similarly, on the left side of the vertical axis the sequence is always 13. If no changes in the sequence occur, the arrival sequence should depend only on ETAs as shown in the right graph of Fig. 4. We call this sequence “intuitional” sequence, or the FCFS sequence.

However, it can be proven that the optimal sequence depends not only on ETA, but also on the time of aircraft and the distance between the entry waypoint and the exit waypoint. The optimal sequence for a certain scenario from group A (a_i ($i = 1, 2, 3$)) is shown in Fig. 5. The main difference between Figs. 4 and 5a is the change in the sequence when the medium aircraft is “squashed” between two heavy aircraft relatively close together. A set of results from group B is shown in Fig. 6. Interestingly, the optimal sequence differs from the intuitional one when the heavy aircraft’s ETA precedes both medium aircraft, which fly relative close to each other. The difference in the fuel burn between the optimal sequence and the intuitional one is shown in Figs. 5b and 6b. Such graphs can help us decide on the optimal sequence, but the associated time shifts are a complicated function of (a_i ($i = 1, 2, 3$)) and ETA_i ($i = 1, 2, 3$)). Therefore, such a calculation cannot be done manually by air traffic controllers in real time.

4.2 “Mind the Size” Rule

If the obtained results from our simulations are to be used by air traffic controllers, they should be simplified. A lengthy analysis and numerous simulations led to the following rules.

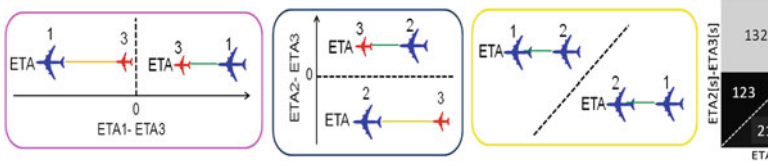


Fig. 4 The first come, first served sequence in respect to the estimated times of arrival

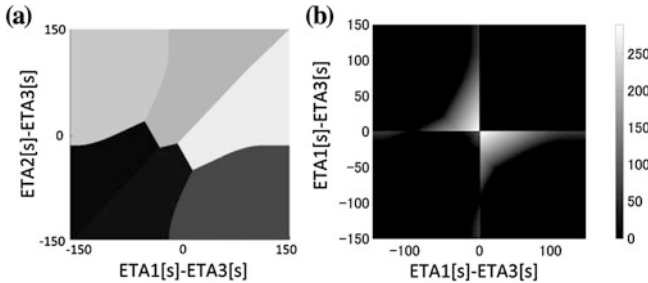


Fig. 5 Sequencing for two heavy and one medium aircraft. **a** Optimal sequence. **b** Fuel burn difference

Sequence:

- Two aircraft (one heavy and one medium)-no changes in the sequence unless the heavy aircraft follows by less than 15 s
- Three aircraft (two heavy and one medium)-no changes in the sequence unless the medium aircraft is between the heavy ones and $|ETA1-ETA2| < k$, where $k = 60$ s but is subject to further analysis.
- Three aircraft (one heavy and two medium)-no changes in the sequence unless the heavy aircraft precedes both medium ones, which are flying soon after, i.e. $|ETA1-ETA3 + ETA2-ETA3| < j$, where $j = 60$ s but may be subject to change.

Adjusted time of arrival (ATA, flight time shift):

- Two aircraft (one heavy and one medium)- no adjustments to ETA of the heavy aircraft, considering the optimal sequence the medium aircraft’s ETA is changed to give ATA.
- Three aircraft (all cases)- ETA of the medium aircraft in the optimal sequence is not changed. ETAs of the other aircraft are adjusted to meet the separation minima.

The essence of these rules is shown in Fig. 7.

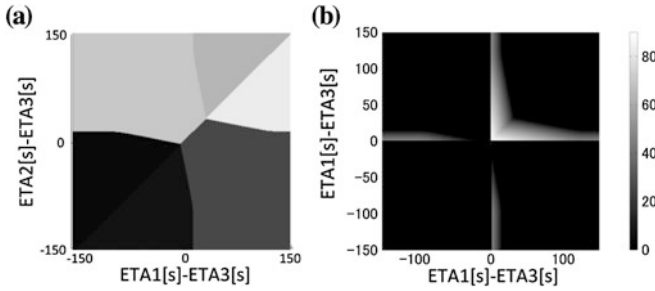


Fig. 6 Sequencing for two medium and one heavy aircraft. **a** Optimal sequence. **b** Fuel burn difference

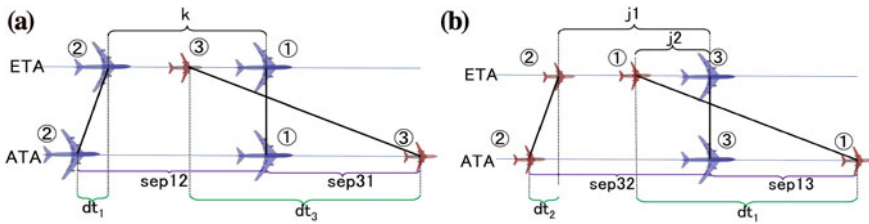


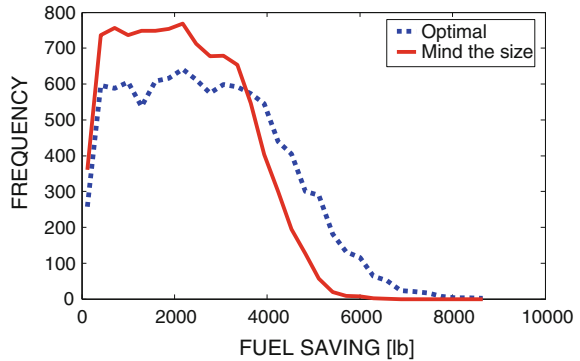
Fig. 7 Sequence change and adjusted time of arrival. **a** Two heavy and a medium aircraft. **b** Two medium and a heavy aircraft

4.3 Monte-Carlo Simulations for “Mind the Size” Rule

To verify the suggested guidelines, medium-congested terminal airspace was considered. Monte-Carlo simulations for 10^4 scenarios of 100 aircraft were performed. The aircraft were to enter the terminal area following a normal random distribution, but there was enough time between each batch of three aircraft. First, we considered the fuel savings obtained by following optimal sequences and optimal ATAs. The fuel burn was compared to that of FCFS case when the flight times of the following aircraft are adjusted based on the first one. A histogram of the results is shown in Fig. 8. The horizontal axis shows the total fuel savings and the vertical axis shows how frequent such savings were observed. The average fuel saving is 2,701 lb (1.18 % of the total fuel).

Applying our “Mind the Size” rule, we obtained the analogous results. The average fuel saving was 2,144 lb (1.01 % of the total fuel burn). Therefore, even though the new rule does not lead to minimum fuel burn, it results in substantial reduce in fuel burn.

Fig. 8 Monte-Carlo simulations for “Mind the Size” rule



5 Scenario Simulations with Available Arrival Time Constraints

5.1 Combination of Aircraft in Longer Batches

The results from the previous section provided us with hints for the direction of the research. In order to be able to extend the simulations to longer batches of aircraft, the earliest available time was implemented in the simulations. The aircraft were still divided into groups of three based on their ETA. The number of aircraft in each group was set to three because of the number of entry waypoints. Besides by keeping this number small, the simulation can easily be approximated to a real-time one.

5.1.1 Two Heavy and One Medium Aircraft

The first group of simulations included two heavy and one medium aircraft. The result in one particular case is shown in Fig. 9a. ETA of the two heavy aircraft were set at 0 and 40 s. The vertical axis shows the ETA of the medium aircraft. The horizontal axis shows the arrival time of the first aircraft to pass through the final waypoint relative to the ETA. The FCFS sequence will not depend on AAT and will always be as the sequence shown in the small window in Fig. 9a top. The white zones represent sequence Medium-Heavy-Heavy, the grey zone shows Heavy-Medium-Heavy and the black zones show Heavy-Heavy-Medium aircraft. It should be noted that depending on the ETA of the heavy aircraft, the grey zone might appear in the optimal sequencing, too. However, this region is relatively small. Also, the border line between MHH and HHM areas also depends on ETA of the three aircraft.

Our goal, however, is to find a rule simple enough to be applied in practice, so instead of searching for high accuracy and analyze every single case, we decided

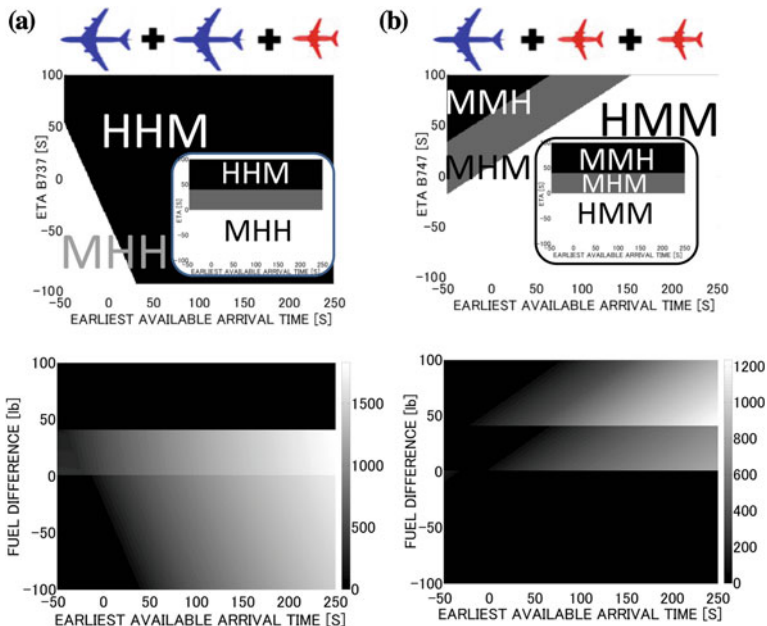


Fig. 9 Sequencing for three aircraft. **a** Two heavy and one medium aircraft. **b** Two medium and one heavy aircraft





































to propose some rules regardless of the ETA of the three aircraft in the group. As presented in Sect. 4, the medium aircraft should go in front when “squeezed” between two large aircraft had there been no arrival time constraints. Looking at the fuel gains shown in Fig. 9a down, however, we changed the above rule with the following one- move the medium aircraft after the heavy ones. Obviously, the more the aircraft are delayed, the more will be gained by the changed sequence.

5.1.2 Two Medium and One Heavy Aircraft

The second group of simulations included two medium and one heavy aircraft. Similarly to the results in Sect. 5.1.1, the fuel gains obtained by optimizing the arrival sequence in respect to the available arrival time and ETA are shown in Fig. 9b. Since the white region occupies most of the optimal sequence graph and leads to the highest fuel gains, the new rule was set to be as follows: have the heavy aircraft land first and then clear the medium ones.

At first sight this rule differs significantly from the one introduced in the previous section. In reality, however, the two rules do not contradict with each other. The main idea behind the rule which placed the heavy aircraft in the middle of the sequence was to move its descent time as little as possible. Actually, this holds

Table 3 Mind the Size revised

Rule No.	FCFS Sequence	Changed Sequence
1	  	  
2	  	  
3	  	  
4	  	  
5	  	  
6	  	  

here, too. Because of the introduction of the available arrival time, delays are to be compensated and the first aircraft will suffer the least, i.e. its descent time will be adjusted the slightest. Therefore, the heavy aircraft should go first in most cases.

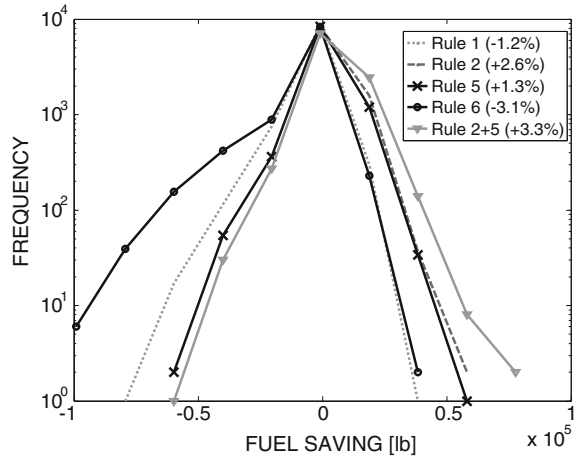
5.2 “Mind the Size” Revised

Based on the results obtained by the simulations of scenarios including the available arrival time, the rules proposed in the previous section were revised as shown in Table 3. The arrival time of the first aircraft in the new sequence should be as close as possible to its optimal one and as early as possible in the available window. The highlighted cases represent a need of change in the sequence.

To examine the effect of each of the rules suggested above, Monte-Carlo simulations were conducted. 10^4 scenarios in which every 60 min 50 aircraft entered the terminal area were conducted. The ration of medium to heavy aircraft was 1:1. Besides, in accordance with the actual traffic at Tokyo International Airport, 50 % of the aircraft entered the terminal area at waypoint A, 20 %- at waypoint B and 30 % at waypoint C. At each entry point the aircraft met the separation requirements. Their exact ETA were distributed randomly. The results from out Monte-Carlo simulations are shown in Fig. 10. Since the mean improvement was observed only in the case of rule 2 and rule 5, a combined scenario where both rules were applied simultaneously was also investigated.

The total fuel burn was decreased when applying two out of the four rules proposed. When the aircraft was moved by more than one slot in the sequence, the mean fuel burn increased (rules 1 and 6). The reason may be twofold. First, the bigger position shift means bigger average fuel burn increase. Second, if we look closely at these cases, they have increased the necessary time separation between the aircraft within the group from $60 + 60$ s to $90 + 60$ s. Indeed, the separation required before and after each group might have shrunk, but this happened only in some cases so the fuel burn increased overall. Therefore, by applying rules 2 and 5 an average fuel saving of 3.3 % can be achieved.

Fig. 10 Distribution of the fuel savings obtained when the new rules were applied. The average saving per each scenario in percentage of the total fuel burn is shown in the legend. Plus indicates positive fuel saving (fuel burn decrease) and minus the vice versa. Note that the vertical graph is in logarithmic scale for better visualization



5.3 Simple Swap

A major disadvantage of the guidelines discussed above is the grouping into three. When looking at the rules which produced the best results, we concluded that changes in the FCFS sequence by one would be sufficient. The sequence should be changed only when the aircraft are in the order heavy-medium-heavy or medium-heavy-medium, as follows:

heavy-medium-heavy → heavy-heavy-medium
 medium-heavy-medium → heavy-medium-medium

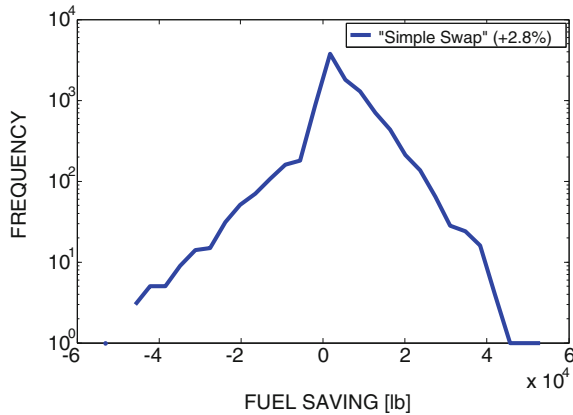
Actually, both cases can be described by one swap between heavy and medium aircraft, i.e. the heavy aircraft goes before the medium one when there is a “squeezed-in” aircraft. Here by a squeezed-in aircraft we mean an aircraft between two other aircraft of different type, i.e. either HMH (M is squeezed-in) or MHM (H is squeezed-in). Besides, when there are more than one possible swap, the earliest one in the FCFS sequence is to be done. This is illustrated in Table 4. The first three aircraft form a “squeezed-in” group, so the heavy one is moved before the medium one. Next, a group can be formed by both (5, 6, 7) or (6, 7, 8). In such a case, consider only (5, 6, 7) as it is before (6, 7, 8) and swap aircraft 5 and 6. Last, consider aircraft (10, 11, 12) and swap 11 and 12. This procedure is referred to as “Simple Swap” rule.

To verify the contribution of the rule “Simple Swap”, Monte Carlo simulations under the same conditions as the ones described in Section V(C) were conducted. A histogram of the results is shown in Fig. 11. The average fuel saving was 2.8 % and the maximum fuel saving for one scenario of 50 aircraft was 34 %. The result is extremely promising considering the simplicity of the rule.

Table 4 Simple Swap

	1	2	3	4	5	6	7	8	9	10	11	12
FCFS	M	H	M	M	M	H	M	H	H	H	M	H
Simple Swap	H	M	M	M	H	M	M	H	H	H	H	M

Fig. 11 Distribution of the fuel savings obtained with the “Simple Swap” rule. The average fuel saving is 2.8 %. Plus indicates positive fuel saving (fuel burn decrease). Note that the vertical graph is in logarithmic scale for better visualization



6 Summary and Conclusion

This research laid the grounds for new sequence assignment rules for aircraft entering the terminal area of a hub airport. Taking into account the human-centered nature of air traffic control, the aim was to develop simple yet efficient guidelines. The fuel burn modeling was based on data from previous research on optimized descent trajectories. Aircraft were then divided in batches in which the best order of arrival was investigated. Interestingly, the average fuel gain was higher when available arrival time window was considered. This facts suggests that the effect of the proposed rules will be even more significant at more congested airports or/and time slots. An average fuel saving of 2.8 % was observed when applying the simple steps summarized below:

- (1) Keep the first come-first served sequenced unless there is a “squeezed-in” aircraft, i.e. a medium aircraft between two heavy aircraft or a heavy aircraft between two medium aircraft. In such a case swap the heavy aircraft with the medium one before it.
- (2) Assign the earliest arrival time plausible for the sequence.

Since the maximum fuel saving was as much as 34 %, the rules can further be improved by looking into the structure of the scenarios which were mostly affected by “Simple Swap”. Nevertheless, it was demonstrated that by a very simple change in the air traffic operations fuel improvement of 2.8 % of the total fuel burn

can be easily achieved. We believe that this research gives a valuable insight into the importance of air traffic operation procedures and their potential contribution to the environmental impact abatement of aviation.

Acknowledgments This work been partially supported by the University of Tokyo Global COE Program, Global Center of Excellence for Mechanical Systems Innovation.

References

1. Kelly F (1943) The Wright brothers. Harcourt, Brace & Co., New York
2. Airbus, Global market forecast 2011–2030(Online). Available. <http://www.airbus.com/company/market/forecast/>
3. Boeing, Boeing's current market outlook (Online). Available. <http://www.boeing.com/commercial/cmo>
4. IATA Economic Briefing- Airline Fuel and Labour Cost Share (Online). Available. http://www.iata.org/whatwedo/Documents/economics/Airline_Labour_Cost_Share_Feb2010.pdf
5. Mak KL, Sun D (2009) Scheduling yard craned in a container terminal using a new genetic approach. *Eng Lett* 17(4):274–280
6. Balakrishnan H, Chandran B (2006) Scheduling aircraft landings under constrained position shifting. In: AIAA guidance, navigation and control conference and exhibit (AIAA2006–6320)
7. Andreeva-Mori A, Suzuki S, Itoh E (2011) Scheduling of arrival aircraft based on minimum fuel burn descents. *ASEAN Eng J* 1(1)
8. Andreeva A, Suzuki S, Itoh E (2009) Flight management based on optimized descent trajectories for minimal environmental impact. In: 47th aircraft symposium, Gifu, Japan 4–6 November 2009 (in Japanese)
9. Andreeva-Mori A (2012) Aircraft resequencing with available arrival time window constraints. In: Proceedings of the international multiconference of engineers and computer scientists 2012, Hong Kong, IMECS 2012, , 14–16 March 2012. Lecture notes in engineering and computer science pp 1513–1517
10. FAA, Pilot and air traffic controller guide to wake turbulence (Online). Available. http://www.faa.gov/training_testing/training/media/wake/04SEC2.PDF
11. International Civil Aviation Organization (1996) Rules of the air and air traffic services (Online). Available. <http://lewczuk.com.pl/materialy/prawo/4444.pdf>

Synchronized Triomineering on Rectangular Boards

Alessandro Cincotti

Abstract In synchronized games players make their moves simultaneously rather than alternately. Synchronized Triomineering is the synchronized version of Triomineering, a variant of a classical two-player combinatorial game called Domineering. New theoretical results for the $n \times 11$ board are presented.

Keywords Combinatorial game · Domineering · Rectangular board · Synchronized game · Synchronized Triomineering · Winning strategy

1 Introduction

The game of Domineering is a typical two-player game with perfect information, proposed around 1973 by Göran Andersson [3, 12, 13]. The two players, usually denoted by Vertical and Horizontal, take turns in placing dominoes (2×1 tile) on a checkerboard. Vertical is only allowed to place its dominoes vertically and Horizontal is only allowed to place its dominoes horizontally on the board. Dominoes are not allowed to overlap and the first player that cannot find a place for one of its dominoes loses. After a time the remaining space may separate into several disconnected regions, and each player must choose into which region to place a domino.

Berlekamp [2] solved the general problem for $2 \times n$ board for odd n . The 8×8 board and many other small boards were recently solved by Breuker et al. [5] using

A. Cincotti (✉)

School of Information Science, Japan Advanced Institute of Science and Technology,
1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan
e-mail: cincotti@jaist.ac.jp

a computer search with a good system of transposition tables. Subsequently, Lachmann et al. solved the problem for boards of width 2, 3, 5, and 7 and other specific cases [14]. Finally, Bullock solved the 10×10 board [6].

The game of Triomineering was proposed in 2004 by Blanco and Fraenkel [4]. In Triomineering Vertical and Horizontal alternate in tiling with a straight triomino (3×1 tile) on a checkerboard. Blanco and Fraenkel calculated Triomineering and values for boards up to 6 squares and small rectangular boards.

2 Synchronized Games

For the sake of self containment, we recall the previous results concerning synchronized games. Initially, the concept of synchronism was introduced in the games of Cutcake [8], Maundy Cake [9], and Domineering [1, 10] in order to study combinatorial games where players make their moves simultaneously.

As a result, in the synchronized versions of these games there exist no zero-games (fuzzy-games), i.e., games where the winner depends exclusively on the player that makes the second (first) move. Moreover, there exists the possibility of a draw, which is impossible in a typical combinatorial game.

In the game of Synchronized Triomineering [7, 11], a general instance and the legal moves for Vertical and Horizontal are defined exactly in the same way as defined for the game of Triomineering.

There is only one difference: Vertical and Horizontal make their legal moves simultaneously, therefore, triominoes are allowed to overlap if they have a 1×1 tile in common. We note that 1×1 overlap is only possible within a simultaneous move.

At the end, if both players cannot make a move, then the game ends in a draw, else if only one player can still make a move, then he/she is the winner.

For each player there exist three possible outcomes:

- The player has a *winning strategy* (ws) independently of the opponent's strategy, or
- The player has a *drawing strategy* (ds), i.e., he/she can always get a draw in the worst case, or
- The player has a *losing strategy* (ls), i.e., he/she does not have a strategy either for winning or for drawing.

Table 1 shows all the possible cases. It is clear that if one player has a winning strategy, then the other player has neither a winning strategy nor a drawing strategy. Therefore, the cases $ws-ws$, $ws-ds$, and $ds-ws$ never happen. As a consequence, if G is an instance of Synchronized Triomineering, then we have six possible legal cases:

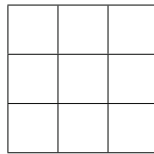
Table 1 The possible outcomes in Synchronized Triomineering

	Horizontal ls	Horizontal ds	Horizontal ws
Vertical ls	$G = VHD$	$G = HD$	$G = H$
Vertical ds	$G = VD$	$G = D$	–
Vertical ws	$G = V$	–	–

- $G = D$ if both players have a drawing strategy, and the game will always end in a draw under perfect play, or
- $G = V$ if Vertical has a winning strategy, or
- $G = H$ if Horizontal has a winning strategy, or
- $G = VD$ if Vertical can always get a draw in the worst case, but he/she could be able to win if Horizontal makes an unlucky move, or
- $G = HD$ if Horizontal can always get a draw in the worst case, but he/she could be able to win if Vertical makes an unlucky move, or
- $G = VHD$ if both players have a losing strategy and the outcome is totally unpredictable.

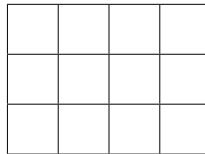
3 Examples of Synchronized Triomineering

The game



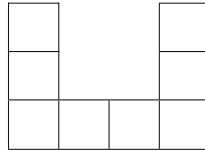
always ends in a draw, therefore $G = D$.

In the game

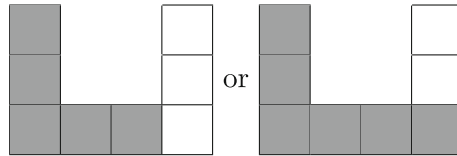


Vertical has a winning strategy moving in the second (or in the third) column, therefore $G = V$.

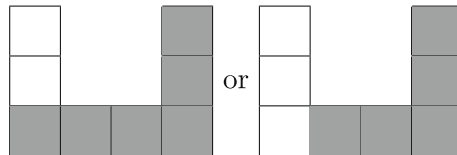
In the game



if Vertical moves in the first column we have two possibilities

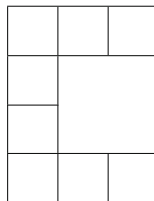


therefore, either Vertical wins or the game ends in a draw. Symmetrically, if Vertical moves in the third column we have two possibilities



therefore, either Vertical wins or the game ends in a draw. It follows $G = VD$.

Symmetrically, in the game



either Horizontal wins or the game ends in a draw therefore, $G = HD$.

In the game

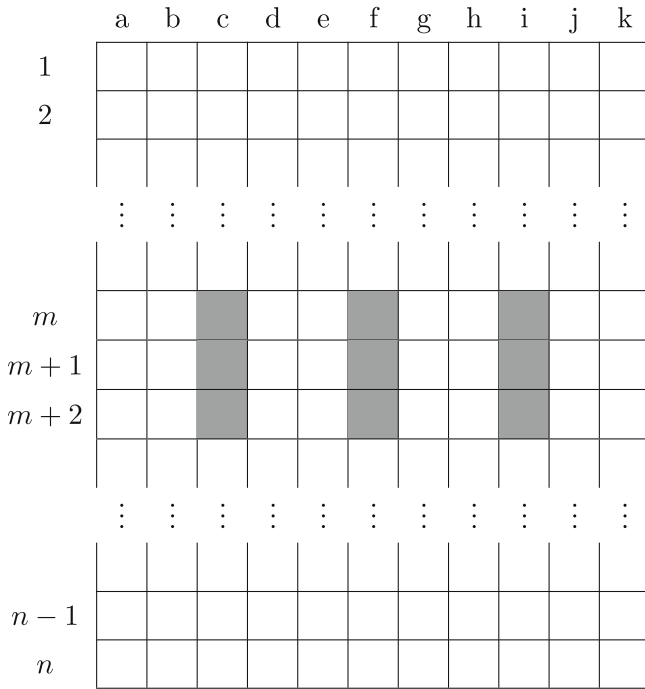
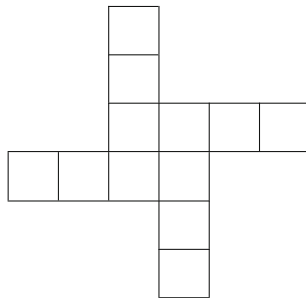


Fig. 1 Vertical strategy on the $n \times 11$ board of Synchronized Triomineering



each player has 4 possible moves. For every move of Vertical, Horizontal can win or draw (and sometimes lose); likewise, for every move by Horizontal, Vertical can win or draw (and sometimes lose). As a result it follows that $G = VHD$.

Table 2 The 22 classes for the 3×11 sub-rectangles

Class	vt	ht	vm	hm
A	3	0	$8 A $	0
B	3	1	$6 B $	0
C	3	2	$4 C $	0
D	3	3	$2 D $	0
E	2	1	$5 E $	$2 E $
F	2	2	$3 F $	$2 F $
G	2	3	$ G $	$2 G $
H	2	4	0	$ H $
I	2	5	0	0
J	1	2	$2 J $	$4 J $
K	1	3	$[3 K /4]$	$4 K $
L	1	4	0	$3 L $
M	1	5	0	$2 M $
N	1	6	0	$ N $
O	1	7	0	0
P	0	3	$[3 P /4]$	$6 P $
Q	0	4	0	$5 Q $
R	0	5	0	$4 R $
S	0	6	0	$3 S $
T	0	7	0	$2 T $
U	0	8	0	$ U $
V	0	9	0	0

4 New Results

In the previous works [7, 11] the $n \times 4$, $n \times 5$, $n \times 7$, and $n \times 8$ boards have been solved. In this section the solution for the $n \times 11$ board is presented.

Theorem 1 *Let G be a $n \times 11$ board of Synchronized Triomineering with $n \geq 41$. Then, Vertical has a winning strategy.*

Proof In the beginning, Vertical will always move into the third, the sixth, and the ninth column of the board, i.e., (m, c) , $(m + 1, c)$, $(m + 2, c)$, (m, f) , $(m + 1, f)$, $(m + 2, f)$, (m, i) , $(m + 1, i)$, and $(m + 2, i)$, where $m \equiv 1 \pmod{3}$, as shown in Fig. 1.

When Vertical cannot move anymore into the third, the sixth, and the ninth column, let us imagine that we divide the main rectangle into 3×11 sub-rectangles starting from the top of the board (by using horizontal cuts). Of course, if $n \not\equiv 0 \pmod{3}$, then the last sub-rectangle will be of size either 1×11 or 2×11 , and Horizontal will be able to make respectively either three more moves or six more moves.

We can classify all these sub-rectangles into 22 classes according to:

- The number of vertical triominoes already placed in the sub-rectangle (vt),

- The number of horizontal triominoes already placed in the sub-rectangle (ht),
- The number of moves that Vertical is able to make in the worst case, in all the sub-rectangles of that class (vm),
- The number of moves that Horizontal is able to make in the best case, in all the sub-rectangles of that class (hm),

as shown in Table 2. We denote with $|A|$ the number of sub-rectangles in the A class, with $|B|$ the number of sub-rectangles in the B class, and so on. The value of vm in all the sub-rectangles belonging to the class E, F, G, J, K, and P considered as a group is

$$5|E| + 3|F| + |G| + 2|J| + \lceil 3|K|/4 \rceil + \lceil 3|P|/4 \rceil$$

The last statement is true under the assumption that Vertical moves first into the sub-rectangles of class E, F, G, and J as long as they exist, and after into the sub-rectangles of the class K and P. When Vertical cannot move anymore into the third, the sixth, and the ninth column, both Vertical and Horizontal have placed the same number of triominoes, therefore

$$\begin{aligned} 3|A| + 2|B| + |C| + |E| &= |G| + 2|H| + 3|I| + |J| \\ &+ 2|K| + 3|L| + 4|M| + 5|N| \\ &+ 6|O| + 3|P| + 4|Q| + 5|R| \\ &+ 6|S| + 7|T| + 8|U| + 9|V| \end{aligned} \tag{1}$$

Let us prove by contradiction that Vertical can make a larger number of moves than Horizontal. Assume therefore

$$moves(V) \leq moves(H)$$

using the data in Table 2

$$\begin{aligned} 8|A| + 6|B| + 4|C| + 2|D| \\ + 5|E| + 3|F| + |G| + 2|J| \\ + \lceil 3|K|/4 \rceil + \lceil 3|P|/4 \rceil \leq 2|E| + 2|F| + 2|G| + |H| \\ + 4|J| + 4|K| + 3|L| + 2|M| \\ + |N| + 6|P| + 5|Q| + 4|R| \\ + 3|S| + 2|T| + |U| + 6 \end{aligned}$$

and applying Eq. 1

$$\begin{aligned} 2|A| + 2|B| + 2|C| + 2|D| + |E| + |F| + |G| + 3|H| \\ + 6|I| + \lceil 3|K|/4 \rceil + 3|L| + 6|M| + 9|N| + 12|O| \\ + \lceil 3|P|/4 \rceil + 3|Q| + 6|R| + 9|S| + 12|T| + 15|U| + 18|V| \leq 6 \end{aligned}$$

which is false because

$$\begin{aligned} &|A| + |B| + |C| + |D| + |E| + |F| + |G| \\ &+ |H| + |I| + |J| + |K| + |L| + |M| + |N| \\ &+ |O| + |P| + |Q| + |R| + |S| + |T| + |U| + |V| = \lfloor n/3 \rfloor \end{aligned}$$

and by hypothesis $n \geq 41$. Therefore, $moves(V) \leq moves(H)$ does not hold and consequently $moves(H) < moves(V)$. We observe that if $n \equiv (3)$, then the theorem holds for $n \geq 22$ and if $n \equiv (3)$, then the theorem holds for $n \geq 3$.

□

By symmetry the following theorem holds.

Theorem 2 Let G be a $11 \times n$ board of Synchronized Triomineering with $n \geq 41$. Then, Horizontal has a winning strategy.

References

1. Bahri S, Kruskal CP (2010) New solutions for synchronized domineering. In: van den Herik HJ, Iida H, Plaat A (eds) Proceedings of the conference on computers and games 2010, Springer, Kanazawa, 24–26 Sept 2010, pp 211–229
2. Berlekamp ER (1988) Blockbusting and domineering. J Comb Theory Ser A 49:67–116
3. Berlekamp ER, Conway JH, Guy RK (2001) Winning way for your mathematical plays. AK Peters, Wellesley
4. Blanco SA, Fraenkel AS (2004) Triomineering, Tridomineering, and L-Tridomineering. Technical report MCS04-10, Department of Computer Science and Applied Mathematics, The Weizmann Institute of Science. [http:// wisdomarchive.wisdom.weizmann.ac.il:81/archive/00000367/](http://wisdomarchive.wisdom.weizmann.ac.il:81/archive/00000367/)
5. Breuker DM, Uiterwijk JWHM, van den Herik HJ (2000) Solving 8×8 domineering. Theor Comput Sci 230:195–206
6. Bullock N (2002) Domineering: solving large combinatorial search space. ICGA J 25(2):67–84
7. Cao T, Cincotti A, Iida H (2012) New results for Synchronized Triomineering. In: Ao SI, Castillo O, Douglas C, Feng DD, Lee J (eds) Lecture notes in engineering and computer science: proceedings of the international multiConference of engineers and computer scientists 2012, IMECS 2012, Newswood Limited, Hong Kong, 14–16 March 2012, pp 1482–1485
8. Cincotti A, Iida H (2007) The game of synchronized cutcake. In: Proceedings of the IEEE symposium on computational intelligence and games, IEEE Press, Honolulu, 1–5 April 2007, pp 374–379
9. Cincotti A, Iida H (2008) The game of synchronized Maundy Cake. In: Proceedings of the 7th annual Hawaii international conference on statistics, mathematics and related fields, Honolulu, 17–19 Jan 2008, pp 422–429
10. Cincotti A, Iida H (2008) The game of synchronized domineering. In: van den Herik HJ, Xu X, Ma Z, Winands MHM (eds) Proceedings of the conference on computers and games 2008, Springer, Beijing, 29 Sept–1 Oct 2008, pp 241–251
11. Cincotti A, Komori S, Iida H (2008) The game of synchronized triomineering and Synchronized Tridomineering. Int J Comput Math Sci 2:143–148

12. Conway JH (2001) On numbers and games. A K Peters
13. Gardner M (1974) Mathematical games. *Sci Am* 230:106–108
14. Lachmann M, Moore C, Rapaport I (2000) Who wins domineering on rectangular boards. In: Nowakowski RJ (ed) *More games of no chance*. Cambridge University Press, Cambridge, pp 307–315

Robust Portfolio Selection Model with Random Fuzzy Returns Based on Arbitrage Pricing Theory and Fuzzy Reasoning Method

Takashi Hasuike, Hideki Katagiri and Hiroshi Tsuda

Abstract This paper considers a robust-based random fuzzy mean-variance portfolio selection problem using a fuzzy reasoning method, particularly a single input type fuzzy reasoning method. Arbitrage Pricing Theory (APT) is introduced as a future return of each security, and each factor in APT is assumed to be a random fuzzy variable whose mean is derived from a fuzzy reasoning method. Furthermore, under interval inputs of fuzzy reasoning method, a robust programming approach is introduced in order to minimize the worst case of the total variance. The proposed model is equivalently transformed into the deterministic nonlinear programming problem, and so the solution steps to obtain the exact optimal portfolio are developed.

Keywords Portfolio selection problem • Arbitrage pricing theory (APT) • Random fuzzy programming • Fuzzy reasoning method • Robust programming • Equivalent transformation • Exact solution algorithm

T. Hasuike (✉)

Graduate School of Information Science and Technology, Osaka University,
2-1 Yamadaoka, Suita, Osaka 565-0871, Japan
e-mail: thasuike@ist.osaka-u.ac.jp

H. Katagiri

Graduate School of Engineering, Hiroshima University, 1-4-1 Kagamiyama,
Higashi-Hiroshima, Hiroshima 739-8527, Japan
e-mail: katagiri-h@hiroshima-u.ac.jp

H. Tsuda

Department of Mathematical Sciences, Faculty of Science and Engineering,
Doshisha University, 1-3 Tatara Miyakodani, Kyotanabe, Kyoto 610-0321, Japan
e-mail: htsuda@mail.doshisha.ac.jp

1 Introduction

The decision of optimal asset allocation among various securities is called portfolio selection problem, and it is one of the most important research themes in investment and financial research fields since the mean-variance model was proposed by Markowitz [20]. Then, after this outstanding research, numerous researchers have contributed to the development of modern portfolio theory (cf. Elton and Gruber [1], Luenberger [19]), and many researchers have proposed several types of portfolio models extending Markowitz model; mean-absolute deviation model (Konno [13], Konno et al. [14]), safety-first model [1], Value at Risk and conditional Value at Risk model (Rockafellar and Uryasev [23]), etc. As a result, nowadays it is common practice to extend these classical economic models of financial investment to various types of portfolio models because investors correspond to present complex markets. In practice, many researchers have been trying different mathematical approaches to develop the theory of portfolio model. Particularly, Capital Asset Pricing Model (CAPM), which is a single factor model proposed by Sharpe [25], Lintner [16] and Mossin [22], has been one of the most useful tools in the investment fields and also used in the performance measure of future returns for portfolios and the asset pricing theory. Hasuike et al. [5] proposed a CAPM-based robust random fuzzy mean-variance model. However, CAPM is a single factor model, and some researchers showed that real economic phenomena were different from the result of CAPM. Therefore, in this paper, we extend the previous model [5] to an uncertain portfolio model based on Arbitrage Pricing Theory (APT) which is a multi-factor model proposed by Ross [24]. APT is one of the most general theories of asset pricing that holds that expected security returns are modeled as a linear function of various macro-economic factors or theoretical market indices, where sensitivity to changes in each factor is represented by a factor-specific beta coefficient.

In such previous researches, expected future return and variance of each asset are assumed to be known. Then, in previous many studies in the sense of mathematical programming for the investment, future returns are assumed to be continuous random variables according to normal distributions. However, investors receive ineffective information from the real markets and economic analysts, and ambiguous factors usually exist in it. Furthermore, investors often have the subjective prediction for future markets which are not derived from the statistical analysis of historical data, but their long-term experiences of investment. Then, even if investors hold a lot of information from the real market, it is difficult that the present or future random distribution of each asset is strictly set. Consequently, we need to consider not only random conditions but also ambiguous and subjective conditions for portfolio selection problems.

As recent studies in mathematical programming, some researchers have proposed various types of portfolio models under randomness and fuzziness. These portfolio models with probabilities and possibilities are included in stochastic programming problems and fuzzy programming problems, respectively, and there

are some basic studies using stochastic programming approaches, goal programming approaches, and fuzzy programming approaches to deal with ambiguous factors as fuzzy sets (Inuiguchi and Ramik [9], Leon et al. [15], Tanaka and Guo [27], Tanaka et al. [28], Vercher et al. [29], Watada [30]). Furthermore, some researchers have proposed mathematical programming problems with both randomness and fuzziness as fuzzy random variables (for instance, Katagiri et al. [11, 12]). In the studies [11, 12], fuzzy random variables were related with the ambiguity of the realization of a random variable and dealt with a fuzzy number that the center value occurs according to a random variable. On the other hand, future returns may be dealt with random variables derived from the statistical analysis, whose parameters are assumed to be fuzzy numbers due to the decision maker's subjectivity, i.e., random fuzzy variables which Liu [17] defined. There are a few studies of random fuzzy programming problem (Hasuike et al. [3, 4], Huang [8], Katagiri et al. [10]). Most recently, Hasuike et al. [4] proposed several portfolio selection models including random fuzzy variables and developed the analytical solution method.

However, in [4], each membership function of fuzzy mean values of future returns was set by the investor, but the mathematical detail of setting the membership function is not obviously given. Of course, it is also important to determine the fuzzy mean values of future returns with the investor's long-term experiences and economical analysts' effective information. Therefore, in order to involve the necessary information into mean values of future returns mathematically, we introduce a fuzzy inference or reasoning method based on fuzzy if-then rules. The fuzzy reasoning method is the most important approach to extract and decide effective rules under fuzziness mathematically. Since outstanding studies of Mamdani [21] and Takagi and Sugeno [24], many researchers have extended these previous approaches, and proposed new fuzzy reasoning methods. Particularly, we focus on a single input type fuzzy reasoning method proposed by Hayashi et al. [6, 7]. This method sets up rule modules to each input item, and the final inference result is obtained by the weighted average of the degrees of the antecedent part and consequent part of each rule module. Nevertheless this approach is one of the simplest mathematical approaches in fuzzy reasoning methods, the final inference result is similar to the other standard approaches. Therefore, in this paper, we proposed a random fuzzy mean-variance model introducing APT-based future returns and Hayashi's single input type fuzzy reasoning method for the mean value of market portfolio of APT.

The proposed random fuzzy mean-variance model is not formulated as a well-defined problem due to fuzziness, we need to set some certain optimization criterion so as to transform into well-defined problems. In this paper, assuming the interval values as a special case of fuzzy numbers and introducing the concept of robust programming, we transform the main problem into a robust programming problem. Recently, the robust optimization problem becomes a more active area of research, and there are some studies of robust portfolio selection problems determining optimal investment strategy using the robust approach (For example, Goldfarb and Iyengar [2], Lobo [18]). In robust programming, we obtain the exact optimal portfolio.

This paper is organized in the following way. In Sect. 2, we introduce mathematical concepts of random fuzzy variables, Arbitrage Pricing Theory, and a single input type fuzzy reasoning method. In Sect. 3, we propose a random fuzzy portfolio selection problem with mean values derived from the fuzzy reasoning method. Performing the deterministic equivalent transformations, we obtain a fractional programming problem with one variable. Finally, in Sect. 4, we conclude this paper.

2 Mathematical Definition and Notation

In many existing studies of portfolio selection problems, future returns are assumed to be random variables or fuzzy numbers. However, since there are few studies of them treated as APT with random fuzzy variables and fuzzy reasoning method, simultaneously. Therefore, in this section, we explain definitions and mathematical formulations of random fuzzy variable, APT, and single input type fuzzy reasoning method proposed by Hayashi et al. [6, 7].

2.1 Random Fuzzy Variables

First of all, we introduce a random fuzzy variables defined by Liu [17] as follows.

Definition 1 (Liu [17]) A random fuzzy variable is a function ξ from a collection of random variables R to $[0, 1]$. An n -dimensional random fuzzy vector $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ is an n -tuple of random fuzzy variables $\xi_1, \xi_2, \dots, \xi_n$.

That is, a random fuzzy variable is a fuzzy set defined on a universal set of random variables. Furthermore, the following random fuzzy arithmetic definition is introduced.

Definition 2 (Liu [17]) Let $\xi_1, \xi_2, \dots, \xi_n$ be random fuzzy variables, and $f : R^n \rightarrow R$ be a continuous function. Then, $\xi = f(\xi_1, \xi_2, \dots, \xi_n)$ is a random fuzzy variable on the product possibility space $(\Theta, P(\Theta), \text{Pos})$, defined as $\xi(\theta_1, \theta_2, \dots, \theta_n) = f(\xi_1(\theta_1), \xi_2(\theta_2), \dots, \xi_n(\theta_n))$ for all $(\theta_1, \theta_2, \dots, \theta_n) \in \Theta$.

From these definitions, the following theorem is derived.

Theorem 1 (Liu [17]) Let ξ_i be random fuzzy variables with membership functions μ_i , $i = 1, 2, \dots, n$, respectively, and $f : R^n \rightarrow R$ be a continuous function. Then, $\xi = f(\xi_1, \xi_2, \dots, \xi_n)$ is a random fuzzy variable whose membership function

is $\mu(\eta) = \sup_{\eta_i \in R_i, 1 \leq i \leq n} \left\{ \min_{1 \leq i \leq n} \mu_i(\eta_i) \mid \eta = f(\eta_1, \eta_2, \dots, \eta_n) \right\}$ for all $\eta \in R$, where $R = \{f(\eta_1, \eta_2, \dots, \eta_n) \mid \eta_i \in R_i, i = 1, 2, \dots, n\}$.

2.2 Arbitrage Pricing Theory

As a useful pricing theory in portfolio models, Capital Asset Pricing Model (CAPM) proposed by Sharpe [25], Lintner [16], and Mossin [22] is well-known and used in many practical investment cases by not only researchers but also practical investors. However, CAPM is a single factor model, and some researchers showed that real economic phenomena were different from the result of CAPM. Therefore, in this paper, we focus on Arbitrage Pricing Theory (APT) which is a multi-factor model proposed by Ross [24] as follows:

$$r_j = a_{0j} + \sum_{i=1}^m a_{ij}f_i$$

where each f_i , ($i = 1, 2, \dots, m$) is the value of factors related to the economic and financial market such as diversified stock indexes and surprises in inflation and GNP. Then, a_{0j} and a_{ij} , ($i = 1, 2, \dots, m$) are inherent values derived from historical data in investment fields. This model includes the CAPM in the case that $a_{ij} = 0$, ($i = 2, \dots, m$) and $f_1 = r_m$. Therefore, the APT is the more extended and versatile pricing model than CAPM.

However, in the case that the decision maker predicts the future return using APT, it is obvious that each factor f_i also occurs according to a random distribution with the investor's subjectivity. Therefore, in these situations, we propose a random fuzzy APT model. In this model we assume that \tilde{f}_i is a random fuzzy variable, and the "dash above" and "wave above", i.e., "-" and "~", denote randomness and fuzziness of the coefficients, respectively. In this paper, \tilde{f}_i occurs according to a random distribution with fuzzy mean value \tilde{f}_i and constant variance σ_i^2 . We assume that each factor is independent of each other. To simplify, we also assume that each fuzzy expected return \tilde{f}_i is an interval values $\tilde{f}_i = [f_i^L, f_i^U]$ derived from a fuzzy reasoning method in the next subsection.

2.3 Single Input Type Fuzzy Reasoning Method

Many researchers have proposed various fuzzy inference and reasoning methods based on or extending Mamdani [21] or Takagi and Sugeno's [26] outstanding studies. In this paper, as a mathematically simple approach, we introduce a single input type fuzzy reasoning method proposed by Hayashi et al. [6, 7]. In this method, we consider the following K rule modules:

$$\text{Rule-}k: \{ \zeta_{ik} = A_s^{ik} \rightarrow f_{ik} = f_s^{ik} \}_{s=1}^{S_k}, (i = 1, 2, \dots, m, k = 1, 2, \dots, K)$$

where ζ_{ik} and f_{ik} are the i th input and consequent data, respectively. Then, f_s^{ik} is the real value of output for the consequent part. A_s^{ik} is the fuzzy set of the s th rule of

the Rules- k , and S_k is the total number of membership function of A_s^{ik} . The degree of the antecedent part in the s th rule of Rules- k is obtained as $h_s^{ik} = A_s^{ik}(\zeta_{ik}^0)$. In Hayashi's single input type fuzzy reasoning method, the inference result f_i^0 is calculated as follows:

$$f_i^0 = \frac{\sum_{s=1}^{S_1} h_s^{i1} f_s^{i1} + \cdots + \sum_{s=1}^{S_K} h_s^{iK} f_s^{iK}}{\sum_{s=1}^{S_1} h_s^{i1} + \cdots + \sum_{s=1}^{S_K} h_s^{iK}} = \frac{\sum_{k=1}^K \sum_{s=1}^{S_k} h_s^{ik} f_s^{ik}}{\sum_{k=1}^K \sum_{s=1}^{S_k} h_s^{ik}} \quad (1)$$

Particularly, if membership functions of all fuzzy sets A_s^{ik} are triangle fuzzy numbers, formula (1) is a linear fractional function on input column vector ζ_i^0 . In this paper, using this fuzzy reasoning method, we obtain the mean value of each factor f_i . We assume that input column vector ζ_i^0 means important financial and social factors to decide the mean value of market portfolio. However, it is difficult to set input column vector ζ_i^0 as a constant value vector. Therefore, we set each input ζ_{ik}^0 as an interval value, and in order to obtain the maximum and minimum values of f_i under interval values of $\left[\underline{\zeta}_{ik}^0, \bar{\zeta}_{ik}^0 \right]$, we introduce the following mathematical programming:

$$\begin{aligned} & \text{Maximize(Minimize)} \quad \frac{\sum_{k=1}^K \sum_{s=1}^{S_k} h_s^{ik} f_s^{ik}}{\sum_{k=1}^K \sum_{s=1}^{S_k} h_s^{ik}} \\ & \text{subject to } \underline{\zeta}_{ik}^0 \leq \zeta_{ik}^0 \leq \bar{\zeta}_{ik}^0, k = 1, 2, \dots, K \end{aligned} \quad (2)$$

Each problem is a fractional linear programming problem under triangle fuzzy numbers A_s^{ik} , and so we obtain the optimal solutions. Let f_i^U and f_i^L be the optimal solution maximizing and minimizing the object, respectively.

3 Formulation of Portfolio Selection Problem with Random Fuzzy Returns

The previous studies on random and fuzzy portfolio selection problems often have considered standard mean-variance model or safety first models introducing probability or fuzzy chance constraints based on modern portfolio theories (e.g. Hasuike et al. [4]). However, there is no study to the random fuzzy mean variance model using the fuzzy reasoning method to obtain the interval mean value of market portfolio. Therefore, in this paper, we extend the previous random fuzzy

mean-variance model to a robust programming-based model using the fuzzy reasoning method.

First, we deal with the following most simple portfolio selection problem involving the random fuzzy variable based on the standard asset allocation problem to maximize total future returns:

$$\begin{aligned} & \text{Maximize } \sum_{j=1}^n \bar{r}_j x_j \\ & \text{subject to } \sum_{j=1}^n x_j = 1, x_j \geq 0, j = 1, 2, \dots, n \end{aligned} \quad (3)$$

where the notation of parameters used in this paper is as follows:

- \bar{r} Future return of the j th financial asset assumed to be a random fuzzy variable, whose fuzzy expected value is \tilde{m}_j and variance-covariance matrix is \mathbf{V} , respectively. Then, we denote randomness and fuzziness of the coefficients by the “dash above” and “wave above”, i.e., “-” and “~”, respectively.
- r_G Target total return
- n Total number of securities
- x_j Budgeting allocation to the j th security

In [4], we consider several models and solution approaches based on standard safety-first models of portfolio selection problems. However, in order to solve the previous models analytically, we must assume that each return occurs according to the normal distributions in the sense of randomness. This assumption is a little restricted. Therefore, in this paper, we do not assume certain random distributions for future returns. Alternatively, we introduce the following portfolio model minimizing the worst total variance, i.e., maximizing the total variance, as a robust portfolio model:

$$\begin{aligned} & \text{Minimize } \max_{f_i \in [f_i^L, f_i^U]} \left\{ \frac{1}{2} V \left(\sum_{j=1}^n \bar{r}_j x_j \right) \right\} \\ & \text{subject to } E \left(\sum_{j=1}^n \bar{r}_j x_j \right) \geq r_G, \\ & \sum_{j=1}^n x_j = 1, x_j \geq 0, j = 1, 2, \dots, n \end{aligned} \quad (4)$$

By assuming this robust programming problem, the investor may be able to avoid the latent risk including the worst case of future return. In order to solve this problem, we fix the value of interval mean value of random fuzzy market portfolio \bar{f}_i as $\bar{f}_i(\omega) \in [f_i^L, f_i^U]$, i.e., each future return is also only a random variable as

$\bar{r}_j(\omega) = a_{0j} + \sum_{i=1}^m a_{ij}\bar{f}_i(\omega)$. Therefore, problem (4) is transformed into the following standard mean-variance portfolio model:

$$\begin{aligned} & \text{Minimize } \frac{1}{2} V \left(\sum_{j=1}^n \bar{r}_j(\omega) x_j \right) \\ & \text{subject to } E \left(\sum_{j=1}^n \bar{r}_j(\omega) x_j \right) \geq r_G, \\ & \sum_{j=1}^n x_j = 1, x_j \geq 0, j = 1, 2, \dots, n \\ & \bar{r}_j(\omega) = a_{0j} + \sum_{i=1}^m a_{ij}\bar{f}_i(\omega) \end{aligned} \quad (5)$$

This problem is a convex quadratic programming problem due to positive definite matrix, and so we obtain the exact optimal portfolio by using the following steps in nonlinear programming.

First we introduce the Lagrange function for problem (5) as follows:

$$L = \frac{1}{2} V \left(\sum_{j=1}^n \bar{r}_j(\omega) x_j \right) + \lambda \left(r_G - E \left(\sum_{j=1}^n \bar{r}_j(\omega) x_j \right) \right) + \zeta \left(1 - \sum_{j=1}^n x_j \right) \quad (6)$$

where λ and ζ are Lagrange multiples. Variance $V \left(\sum_{j=1}^n \bar{r}_j(\omega) x_j \right)$ and $E \left(\sum_{j=1}^n \bar{r}_j(\omega) x_j \right)$ are equivalently transformed into the following forms from the assumption of APT in Sect. 2.2:

$$\left\{ \begin{array}{l} V \left(\sum_{j=1}^n \bar{r}_j(\omega) x_j \right) = \sum_{j=1}^n \hat{\sigma}_j^2 x_j^2 \\ \hat{\sigma}_j^2 = \sum_{i=1}^m a_{ij}^2 \sigma_i^2 \end{array} \right\}, \left\{ \begin{array}{l} E \left(\sum_{j=1}^n \bar{r}_j(\omega) x_j \right) = \sum_{j=1}^n \hat{r}_j(\omega) x_j \\ \hat{r}_j(\omega) = a_{0j} + \sum_{i=1}^m a_{ij}\bar{f}_i(\omega) \end{array} \right.$$

Therefore, by using Karush-Kuhn-Tucker (KKT) condition, we obtain the following equation on each variable x_j :

$$\begin{aligned} \frac{\partial L}{\partial x_j} &= \hat{\sigma}_j^2 x_j - \lambda \hat{r}_j(\omega) - \zeta = 0, (j = 1, 2, \dots, n) \\ \sum_{j=1}^n \hat{r}_j(\omega) x_j &= r_G, \sum_{j=1}^n x_j = 1 \end{aligned} \quad (7)$$

In order to solve equations derived from KKT condition, we set the vector notation, and obtain the solution of x as follows:

$$\begin{aligned}
& \mathbf{V}\mathbf{x} - \lambda\hat{r}(\omega) - \xi\mathbf{I} = \mathbf{0} \\
& \Leftrightarrow \mathbf{V}\mathbf{x} - \mathbf{A}' \begin{pmatrix} \lambda \\ \xi \end{pmatrix} = \mathbf{0}, \\
& \left(\mathbf{V} = \begin{pmatrix} \hat{\sigma}_1^2 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \hat{\sigma}_n^2 \end{pmatrix}, \mathbf{A} = \begin{pmatrix} \hat{r}_1(\omega) & \cdots & \hat{r}_n(\omega) \\ 1 & \cdots & 1 \end{pmatrix} \right) \quad (8) \\
& \Leftrightarrow x = \mathbf{V}^{-1}\mathbf{A}' \begin{pmatrix} \lambda \\ \xi \end{pmatrix}, \left(\mathbf{V} = \begin{pmatrix} 1/\hat{\sigma}_1^2 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & 1/\hat{\sigma}_n^2 \end{pmatrix} \right)
\end{aligned}$$

By substituting this solution into second and third equations in KKT condition (7), we obtain the following optimal values of Lagrange multiples:

$$\begin{aligned}
& \mathbf{A}\mathbf{V}^{-1}\mathbf{A}' \begin{pmatrix} \lambda \\ \xi \end{pmatrix} = \begin{pmatrix} r_G \\ 1 \end{pmatrix} = \hat{\mathbf{r}}_G \\
& \Leftrightarrow \begin{pmatrix} \lambda \\ \xi \end{pmatrix} = (\mathbf{A}\mathbf{V}^{-1}\mathbf{A}')^{-1}\hat{\mathbf{r}}_G
\end{aligned} \quad (9)$$

Consequently, we obtain the optimal portfolio \mathbf{x}^* and the objective value $\sum_{j=1}^n \hat{\sigma}_j^2 (x_j^*)^2 = (\mathbf{x}^*)' \mathbf{V} \mathbf{x}^*$ as follows:

$$\begin{aligned}
\mathbf{x}^* &= \mathbf{V}^{-1}\mathbf{A}'(\mathbf{A}\mathbf{V}^{-1}\mathbf{A}')^{-1}\hat{\mathbf{r}}_G \\
(\mathbf{x}^*)' \mathbf{V} \mathbf{x}^* &= \left(\hat{\mathbf{r}}_G' (\mathbf{A}\mathbf{V}^{-1}\mathbf{A}')^{-1} \mathbf{A}\mathbf{V}^{-1} \right) \mathbf{V} \left(\mathbf{V}^{-1}\mathbf{A}'(\mathbf{A}\mathbf{V}^{-1}\mathbf{A}')^{-1}\hat{\mathbf{r}}_G \right) \\
&= \hat{\mathbf{r}}_G' (\mathbf{A}\mathbf{V}^{-1}\mathbf{A}')^{-1} \mathbf{A}\mathbf{V}^{-1}\mathbf{A}' (\mathbf{A}\mathbf{V}^{-1}\mathbf{A}')^{-1} \hat{\mathbf{r}}_G \\
&= \hat{\mathbf{r}}_G' (\mathbf{A}\mathbf{V}^{-1}\mathbf{A}')^{-1} \hat{\mathbf{r}}_G
\end{aligned} \quad (10)$$

From x^* and the optimal objective value, we secondly consider a robust programming-based portfolio model, i.e. the worst case of the total variance:

$$\begin{aligned}
& \text{Maximize } \hat{\mathbf{r}}_G' (\mathbf{A}\mathbf{V}^{-1}\mathbf{A}')^{-1} \hat{\mathbf{r}}_G \\
& \text{subject to } f_i \in [f_i^L, f_i^U], (i = 1, 2, \dots, m)
\end{aligned} \quad (11)$$

In objective function $\hat{\mathbf{r}}_G' (\mathbf{A}\mathbf{V}^{-1}\mathbf{A}')^{-1} \hat{\mathbf{r}}_G$, inverse matrix $(\mathbf{A}\mathbf{V}^{-1}\mathbf{A}')^{-1}$ is calculated as the following form:

$$\mathbf{A}\mathbf{V}^{-1}\mathbf{A}^t = \begin{pmatrix} \sum_{j=1}^n \frac{\hat{r}_j^2(\omega)}{\hat{\sigma}_j^2} & \sum_{j=1}^n \frac{\hat{r}_j(\omega)}{\hat{\sigma}_j^2} \\ \sum_{j=1}^n \frac{\hat{r}_j(\omega)}{\hat{\sigma}_j^2} & \sum_{j=1}^n \frac{1}{\hat{\sigma}_j^2} \end{pmatrix}$$

$$\left\{ \begin{array}{l} (\mathbf{A}\mathbf{V}^{-1}\mathbf{A}^t)^{-1} = \frac{1}{\Delta(\mathbf{r})} \begin{pmatrix} \sum_{j=1}^n \frac{1}{\hat{\sigma}_j^2} & -\sum_{j=1}^n \frac{\hat{r}_j(\omega)}{\hat{\sigma}_j^2} \\ -\sum_{j=1}^n \frac{\hat{r}_j(\omega)}{\hat{\sigma}_j^2} & \sum_{j=1}^n \frac{\hat{r}_j^2(\omega)}{\hat{\sigma}_j^2} \end{pmatrix} \\ \Delta(\mathbf{r}) = \left(\sum_{j=1}^n \frac{1}{\hat{\sigma}_j^2} \right) \left(\sum_{j=1}^n \frac{\hat{r}_j^2(\omega)}{\hat{\sigma}_j^2} \right) - \left(\sum_{j=1}^n \frac{\hat{r}_j(\omega)}{\hat{\sigma}_j^2} \right)^2 \end{array} \right. \quad (12)$$

Therefore, objective function $\hat{\mathbf{r}}_G^t (\mathbf{A}\mathbf{V}^{-1}\mathbf{A}^t)^{-1} \hat{\mathbf{r}}_G$ is also calculated as follows:

$$\begin{aligned} & \hat{\mathbf{r}}_G^t (\mathbf{A}\mathbf{V}^{-1}\mathbf{A}^t)^{-1} \hat{\mathbf{r}}_G \\ &= \frac{1}{\Delta(\mathbf{r})} \left(\sum_{j=1}^n \frac{\hat{r}_j^2(\omega)}{\hat{\sigma}_j^2} - 2r_G \sum_{j=1}^n \frac{\hat{r}_j(\omega)}{\hat{\sigma}_j^2} + r_G^2 \sum_{j=1}^n \frac{1}{\hat{\sigma}_j^2} \right) \end{aligned} \quad (13)$$

Consequently, problem (11) is equivalently transformed into the following problem:

$$\begin{aligned} & \text{Maximize } \frac{1}{\Delta(\mathbf{r})} \left(\sum_{j=1}^n \frac{r_j^2}{\hat{\sigma}_j^2} - 2r_G \sum_{j=1}^n \frac{r_j}{\hat{\sigma}_j^2} + r_G^2 \sum_{j=1}^n \frac{1}{\hat{\sigma}_j^2} \right) \\ & \text{subject to } r_j = a_{0j} + \sum_{i=1}^m a_{ij}f_i, f_i \in [f_i^L, f_i^U], (i = 1, 2, \dots, m) \\ & \left(\text{where } \Delta(\mathbf{r}) = \left(\sum_{j=1}^n \frac{1}{\hat{\sigma}_j^2} \right) \left(\sum_{j=1}^n \frac{\hat{r}_j^2(\omega)}{\hat{\sigma}_j^2} \right) - \left(\sum_{j=1}^n \frac{\hat{r}_j(\omega)}{\hat{\sigma}_j^2} \right)^2 \right) \end{aligned} \quad (14)$$

In this problem, we substitute $r_j = a_{0j} + \sum_{i=1}^m a_{ij}f_i$ derived from APT, and the numerator and denominator of objective function are calculated as follows:

$$\begin{aligned} & \sum_{j=1}^n \frac{r_j^2}{\hat{\sigma}_j^2} - 2r_G \sum_{j=1}^n \frac{r_j}{\hat{\sigma}_j^2} + r_G^2 \sum_{j=1}^n \frac{1}{\hat{\sigma}_j^2} \\ &= \sum_{h=1}^m \sum_{i=1}^m \sum_{j=1}^n \frac{a_{hj}a_{ij}hf_i}{\hat{\sigma}_j^2} - 2 \sum_{i=1}^m \sum_{j=1}^n \left(\frac{a_{ij}f_i}{\hat{\sigma}_j^2} r_G + \frac{a_{0j}a_{ij}f_i}{\hat{\sigma}_j^2} \right) + \text{const.} \\ &= \sum_{h=1}^m \sum_{i=1}^m \sum_{j=1}^n p_{hij}^{(1)} f_i hf_i - \sum_{i=1}^m \sum_{j=1}^n p_{ij}^{(2)} f_i + p_c \end{aligned}$$

$$\begin{aligned}
 \Delta(r) &= \left(\sum_{j=1}^n \frac{1}{\hat{\sigma}_j^2} \right) \left(\sum_{h=1}^m \sum_{i=1}^m \sum_{j=1}^n \frac{a_{hij} a_{ij} f_i f_i}{\hat{\sigma}_j^2} \right) - \left(\sum_{j=1}^n \frac{a_{0j} + \sum_{i=1}^m a_{ij} f_i}{\hat{\sigma}_j^2} \right)^2 \\
 &+ \left(\sum_{j=1}^n \frac{1}{\hat{\sigma}_j^2} \right) \left(\sum_{i=1}^m \sum_{j=1}^n \frac{2a_{0j} a_{ij} f_i}{\hat{\sigma}_j^2} \right) + const. \\
 &= \sum_{h=1}^m \sum_{i=1}^m \sum_{j=1}^n q_{hij}^{(1)} f_i f_i - \sum_{i=1}^m \sum_{j=1}^n q_{ij}^{(2)} f_i + q_c
 \end{aligned}
 \tag{15}$$

Therefore, the main problem is represented as the following mathematical programming problem with variable $f_i, (i = 1, 2, \dots, m)$:

$$\begin{aligned}
 \text{Maximize } & \frac{\sum_{h=1}^m \sum_{i=1}^m \sum_{j=1}^n p_{hij}^{(1)} f_i f_i - \sum_{i=1}^m \sum_{j=1}^n p_{ij}^{(2)} f_i + p_c}{\sum_{h=1}^m \sum_{i=1}^m \sum_{j=1}^n q_{hij}^{(1)} f_i f_i - \sum_{i=1}^m \sum_{j=1}^n q_{ij}^{(2)} f_i + q_c} \\
 \text{subject to } & f_i \in [f_i^L, f_i^U], (i = 1, 2, \dots, m)
 \end{aligned}
 \tag{16}$$

This problem is a nonlinear fractional programming problem, and so it is generally difficult to obtain the optimal solution. However, this problem has variables $f_i, (i = 1, 2, \dots, m)$, and so we can use the quadratic convex programming approaches, and obtain the strict optimal portfolio.

4 Conclusion

In this paper, we have proposed a robust-based mean-variance portfolio selection problem with random fuzzy APT using a single input type fuzzy reasoning method. In order to deal with each factor in APT as a random interval variable, and to perform the deterministic equivalent transformations, the proposed model has been nonlinear programming problem with only one variable. Therefore, we have obtained the exact optimal portfolio using standard nonlinear programming approaches.

As future studies, we need to develop the solution algorithm in cases of general fuzzy numbers including interval values. Furthermore, we also need to consider random fuzzy portfolio models derived from not only a single input type fuzzy reasoning method but also more general fuzzy reasoning methods.

References

1. Elton EJ, Gruber MJ (1995) *Modern portfolio theory and investment analysis*. Wiley, New York
2. Goldfarb D, Iyengar G (2003) Robust portfolio selection problems. *Math Operat Res* 28:1–38
3. Hasuike T, Katagiri H, Ishii H (2009) Multiobjective random fuzzy linear programming problems based on the possibility maximization model. *J Adv Comput Intell Intell Inform* 13(4):373–379
4. Hasuike T, Katagiri H, Ishii H (2009) Portfolio selection problems with random fuzzy variable returns. *Fuzzy Sets Syst* 160:2579–2596
5. Hasuike T, Katagiri H, Tsuda H (2012) Robust-based random fuzzy mean-variance model using a fuzzy reasoning method. In: *Lecture notes in engineering and computer science: proceedings of the international multi conference of engineers and computer scientists 2012, IMECS 2012, 14–16 March, 2012, Hong Kong*, pp 1461–1466
6. Hayashi K, Otsubo A, Shiranita K (1999) Realization of nonlinear and linear PID control using simplified direct inference method. *IEICE Trans Fundam (Japanese Edition)*, J82-A(7), pp 1180–1184
7. Hayashi K, Otsubo A, Shirahata K (2001) Improvement of conventional method of PI fuzzy control. *IEICE Trans Fundam E84-A(6)*, pp 1588–1592
8. Huang X (2007) Two new models for portfolio selection with stochastic returns taking fuzzy information. *Eur J Oper Res* 180:396–405
9. Inuiguchi M, Ramik J (2000) Possibilistic linear programming: A brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem. *Fuzzy Sets Syst* 111:3–28
10. Katagiri H, Hasuike T, Ishii H, Nishizaki I (2008) Random fuzzy programming models based on possibilistic programming. In: *Proceedings of the 2008 IEEE international conference on systems, man and cybernetics (to appear)*
11. Katagiri H, Ishii H, Sakawa M (2004) On fuzzy random linear knapsack problems. *CEJOR* 12:59–70
12. Katagiri H, Sakawa M, Ishii H (2005) A study on fuzzy random portfolio selection problems using possibility and necessity measures. *Scientiae Mathematicae Japonicae* 65:361–369
13. Konno H, Yamazaki H (1991) Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market. *Manage Sci* 37:519–531
14. Konno H, Shirakawa H, Yamazaki H (1993) A mean-absolute deviation-skewness portfolio optimization model. *Ann Oper Res* 45:205–220
15. Leon RT, Liern V, Vercher E (2002) Validity of infeasible portfolio selection problems: fuzzy approach. *Eur J Oper Res* 139:178–189
16. Lintner BJ (1965) Valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets. *Rev Econ Stat* 47:13–37
17. Liu B (2002) *Theory and practice of uncertain programming*. Physica Verlag, Heidelberg
18. Lobo MS (2000) *Robust and convex optimization with applications in finance*. Doctor thesis of the department of Electrical engineering and the committee on graduate studies, Stanford University, Stanford
19. Luenberger DG (1997) *Investment science*, Oxford University Press, Oxford
20. Markowitz HM (1952) Portfolio selection. *J Financ* 7(1):77–91
21. Mamdani EH (1974) Application of fuzzy algorithms for control of simple dynamic plant. *Proc IEE* 121(12):1585–1588
22. Mossin J (1966) Equilibrium in capital asset markets. *Econometrica* 34(4):768–783
23. Rockafellar RT, Uryasev S (2000) Optimization of conditional value-at-risk. *J Risk* 2(3):1–21
24. Ross S (1976) The arbitrage theory of capital asset pricing. *J Econ Theory* 13(3):341–360
25. Sharpe WF (1964) Capital asset prices: A theory of market equivalent under conditions of risk. *J Financ* 19(3):425–442

26. Takagi T, Sugeno M (1985) Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans Syst Man Cybern SMC-15*(1), pp 116–132
27. Tanaka H, Guo P (1999) Portfolio selection based on upper and lower exponential possibility distributions. *Eur J Oper Res* 114:115–126
28. Tanaka H, Guo P, Turksen IB (2000) Portfolio selection based on fuzzy probabilities and possibility distributions. *Fuzzy Sets Syst* 111:387–397
29. Vercher E, Bermúdez JD, Segura JV (2007) Fuzzy portfolio optimization under downside risk measures. *Fuzzy Sets Syst* 158:769–782
30. Watada J (1997) Fuzzy portfolio selection and its applications to decision making. *Tatra Mountains Math Pub* 13:219–248

Hierarchical Multiobjective Stochastic Linear Programming Problems Through a Fractile Optimization Model Using Reference Membership Intervals

Hitoshi Yano

Abstract In this paper, we focus on hierarchical multiobjective stochastic linear programming problems (HMOP) where multiple decision makers in a hierarchical organization have their own multiple objective linear functions together with common linear constraints. In order to deal with HMOP, a fractile optimization model is applied. By considering the conflict between permissible probability levels and the corresponding objective functions in such a model, it is assumed that each of the decision makers has fuzzy goals for not only permissible probability levels but also the corresponding objective functions, and such fuzzy goals can be quantified by eliciting the membership functions. Through the fuzzy decision, such membership functions are integrated. In the integrated membership space, the extended Pareto optimality concept is introduced. The interactive algorithm to obtain a satisfactory solution from among a Pareto optimal solution set is proposed on the basis of linear programming technique, in which the hierarchical decision structure is reflected by the decision power and the proper balance between permissible objective levels and the corresponding probability function is attained.

Keywords Hierarchical multiobjective stochastic linear programming · Decision power · A fractile optimization model · Interactive decision making · Pareto optimality · Fuzzy decision

H. Yano (✉)
Graduate School of Humanities and Social Sciences,
Nagoya City University, Nagoya 467-8501, Japan
e-mail: yano@hum.nagoya-cu.ac.jp

1 Introduction

The decision makers in practical hierarchical decision making situations often encounter two kinds of decision making processes, one is well known as a multi-level programming process and the other is the interactive decision making process [3]. The Stackelberg games are well-known as multilevel programming problems with multiple decision makers, in which the decision maker in each level makes his/her decision independently in order to optimize his/her own objective function [1, 10]. On the other hand, the interactive decision making process can be found in large scale hierarchical organizations such as multi-hierarchical companies, in which the decision maker in each level makes his/her decision through the interaction between the decision makers and the lower level decision makers submit their own decision and then such decision is modified by the upper level decision makers with considerations of the overall benefits [3]. In order to deal with such an interactive decision making process, Lai [2], Shih et al. [9] and Lee et al. [3] introduced concepts of memberships of optimalities and degrees of decision powers and proposed fuzzy approaches to obtain a satisfactory solution. As a natural extension of their approaches, Yano [11] proposed a fuzzy approach for hierarchical multiobjective linear programming problems. On the other hand, in the actual decision making situations, the decision makers often encounter difficulties to deal with vague information or uncertain data. Sakawa et al. [5–7] formulated multiobjective stochastic linear programming problems through a probability maximization model and a fractile optimization model, and proposed interactive algorithm to obtain a satisfactory solution from among a Pareto optimal solution set. Using a probability maximization model or a fractile optimization model, it is required for the decision maker to specify parameters called permissible objective levels or permissible probability levels in advance. However, it seems to be very difficult to specify such values in advance. In order to cope with such difficulties, Yano et al. [14] proposed fuzzy approaches to multiobjective stochastic linear programming problems, where the decision maker has fuzzy goals for permissible objective levels and permissible probability levels, and such fuzzy goals are quantified by eliciting the membership functions. Unfortunately, in the proposed method, it is assumed that the decision maker adopts the fuzzy decision [4] to obtain the satisfactory solution.

In this paper, we focus on hierarchical multiobjective stochastic linear programming problems [12], and propose an interactive algorithm to obtain a satisfactory solution from among a Pareto optimal solution set. In the proposed method, by considering the conflict between permissible probability levels and the corresponding objective functions in a fractile optimization model, the corresponding membership functions are integrated through the fuzzy decision. In the integrated membership space, the extended Pareto optimal concept is introduced. In Sect. 2, hierarchical multiobjective programming problems through a fractile optimization model is formulated. In Sect. 3, an interactive algorithm based on linear programming technique is proposed to obtain a satisfactory solution.

2 Hierarchical Multiobjective Stochastic Linear Programming Problems Through a Fractile Optimization Model

We consider the following hierarchical multiobjective stochastic linear programming problem (HMOP), where each of the decision makers (DM r , $r = 1, \dots, q$) has his/her own multiple objective linear functions together with common linear constraints, and random variable coefficients are involved in each objective function.

[HMOP1]

first level decision maker: DM1

$$\min_{x \in X} \bar{z}_1(x) = (\bar{z}_{11}(x), \dots, \bar{z}_{1k_1}(x))$$

.....

q-th level decision maker: DMq

$$\min_{x \in X} \bar{z}_q(x) = (\bar{z}_{q1}(x), \dots, \bar{z}_{qk_q}(x))$$

where $x = (x_1, x_2, \dots, x_n)^T$ is n -dimensional decision column vector whose elements $x_i, i = 1, \dots, n$ are nonnegative, X is a linear constraint set with respect to x . Each objective function of DM r is defined by $\bar{z}_{r\ell}(x) = \bar{c}_{r\ell}x + \bar{\alpha}_{r\ell}$, $\bar{c}_{r\ell} = c_{r\ell}^1 + \bar{t}_{r\ell}c_{r\ell}^2$, $\bar{\alpha}_{r\ell} = \alpha_{r\ell}^1 + \bar{t}_{r\ell}\alpha_{r\ell}^2$, where $\bar{c}_{r\ell}, \ell = 1, \dots, k_r$ are n dimensional random variable row vectors, $\bar{\alpha}_{r\ell}, \ell = 1, \dots, k_r$ are random variables, and $\bar{t}_{r\ell}$ is a random variable whose cumulative distribution function $T_{r\ell}(\cdot)$ is assumed to be strictly monotone increasing and continuous.

Similar to the formulations of multilevel linear programming problems proposed by Lee et al. [3], it is assumed that the upper level decision makers make their decisions with consideration of the overall benefits for the hierarchical organization, although they can take priority for their objective functions over the lower level decision makers.

If we adopt a fractile optimization model [6] for HMOP1, we can convert HMOP1 to the following multiobjective programming problem.

[HMOP2 (\hat{p})]

first level decision maker: DM1

$$\min_{x \in X, f_{1\ell} \in R^1, \ell=1, \dots, k_1} (f_{11}, \dots, f_{1k_1})$$

.....

q-th level decision maker: DMq

$$\min_{x \in X, f_{q\ell} \in R^1, \ell=1, \dots, k_q} (f_{q1}, \dots, f_{qk_q})$$

subject to

$$p_{r\ell}(x, f_{r\ell}) \geq \hat{p}_{r\ell}, r = 1, \dots, q, \ell = 1, \dots, k_r \tag{1}$$

where $\hat{p}_r = (\hat{p}_{r1}, \dots, \hat{p}_{rk_r})$, $r = 1, \dots, q$, $\hat{p} = (\hat{p}_1, \dots, \hat{p}_q)$ are vectors of permissible probability levels which are specified by the decision maker in his/her subjective manner.

In HMOP2 (\hat{p}), the constraint (1) can be transformed into the following form.

$$\begin{aligned} \hat{p}_{r\ell} &\leq \hat{p}_{r\ell}(x, f_{r\ell}) = T_{r\ell} \left(\frac{f_{r\ell} - (c_{r\ell}^1 x + \alpha_{r\ell}^1)}{c_{r\ell}^2 x + \alpha_{r\ell}^2} \right) \\ \Leftrightarrow f_{r\ell} &\geq T_{r\ell}^{-1}(\hat{p}_{r\ell}) \cdot (c_{r\ell}^2 x + \alpha_{r\ell}^2) + (c_{r\ell}^1 x + \alpha_{r\ell}^1) \end{aligned}$$

Let us define the right-hand side of the above inequality as follows.

$$f_{r\ell}(x, \hat{p}_{r\ell} \stackrel{\text{def}}{=} T_{r\ell}^{-1}(\hat{p}_{r\ell}) \cdot (c_{r\ell}^2 x + \alpha_{r\ell}^2) + (c_{r\ell}^1 x + \alpha_{r\ell}^1) \tag{2}$$

Then, HMOP2 (\hat{p}) can be equivalently reduced to the following simple form.

[HMOP3 (\hat{p})]

first level decision maker: DM1

$$\min_{x \in X} (f_{11}(x, \hat{p}_{11}), \dots, f_{1k_1}(x, \hat{p}_{1k_1}))$$

.....

q-th level decision maker: DMq

$$\min_{x \in X} (f_{q1}(x, \hat{p}_{q1}), \dots, f_{qk_q}(x, \hat{p}_{qk_q}))$$

In order to deal with HMOP3 (\hat{p}), the decision maker must specify permissible probability levels \hat{p} in advance. However, in general, the decision maker seems to prefer not only the less value of the objective function $f_{r\ell}(x, \hat{p}_{r\ell})$ but also the larger value of the permissible probability level $\hat{p}_{r\ell}$. From such a point of view, we consider the following multiobjective programming problem which can be regarded as a natural extension of HMOP3 (\hat{p}).

[HMOP4]

first level decision maker: DM1

$$\min_{x \in X} (f_{11}(x, \hat{p}_{11}), \dots, f_{1k_1}(x, \hat{p}_{1k_1}), -\hat{p}_{11}, \dots, -\hat{p}_{1k_1})$$

.....

q-th level decision maker: DMq

$$\min_{x \in X} (f_{q1}(x, \hat{p}_{q1}), \dots, f_{qk_q}(x, \hat{p}_{qk_q}), -\hat{p}_{q1}, \dots, -\hat{p}_{qk_q})$$

Considering the imprecise nature of the decision maker's judgment, we assume that the decision maker has a fuzzy goal for each objective function in HMOP4. Such a fuzzy goal can be quantified by eliciting the corresponding membership

function. Let us denote a membership function of an objective function $f_{r\ell}(x, \hat{p}_{r\ell})$ as $\mu_{\hat{f}_{r\ell}}(f_{r\ell}(x, \hat{p}_{r\ell}))$, and a membership function of a permissible probability level $\hat{p}_{r\ell}$ as $\mu_{\hat{p}_{r\ell}}(\hat{p}_{r\ell})$ respectively.

Then, HMOP4 can be transformed as the following problem.

[HMOP5]

first level decision maker: DM1

$$\max_{x \in X, \hat{p}_1 \in (0,1)^{k_1}} \left(\mu_{\hat{f}_{11}}(f_{11}(x, \hat{p}_{11})), \dots, \mu_{\hat{f}_{1k_1}}(f_{1k_1}(x, \hat{p}_{1k_1})), \mu_{\hat{p}_{11}}(\hat{p}_{11}), \dots, \mu_{\hat{p}_{1k_1}}(\hat{p}_{1k_1}) \right),$$

.....

q-th level decision maker: DMq

$$\max_{x \in X, \hat{p}_q \in (0,1)^{k_q}} \left(\mu_{\hat{f}_{q1}}(f_{q1}(x, \hat{p}_{q1})), \dots, \mu_{\hat{f}_{qk_q}}(f_{qk_q}(x, \hat{p}_{qk_q})), \mu_{\hat{p}_{q1}}(\hat{p}_{q1}), \dots, \mu_{\hat{p}_{qk_q}}(\hat{p}_{qk_q}) \right)$$

Throughout this section, we make the following assumptions with respect to the membership functions $\mu_{f_{r\ell}}(f_{r\ell}(x, \hat{p}_{r\ell}))$, $\mu_{\hat{p}_{r\ell}}(\hat{p}_{r\ell})$, $r = 1, \dots, q, \ell = 1, \dots, k_r$.

Assumption 1

$\mu_{\hat{p}_{r\ell}}(\hat{p}_{r\ell})$, $r = 1, \dots, q, \ell = 1, \dots, k_r$ are strictly increasing and continuous with respect to $\hat{p}_{r\ell}$, which is defined on the interval $[p_{r\ell\min}, p_{r\ell\max}] \subset (0, 1)$, where $\mu_{\hat{p}_{r\ell}}(\hat{p}_{r\ell}) = 0$ if $\hat{p}_{r\ell} \leq p_{r\ell\min}$, and $\mu_{\hat{p}_{r\ell}}(\hat{p}_{r\ell}) = 1$ if $\hat{p}_{r\ell} \geq p_{r\ell\max}$.

Assumption 2

$\mu_{f_{r\ell}}(f_{r\ell}(x, \hat{p}_{r\ell}))$, $r = 1, \dots, q, \ell = 1, \dots, k_r$ are strictly decreasing and continuous with respect to $f_{r\ell}(x, \hat{p}_{r\ell})$, which is defined on the interval $[f_{r\ell\min}, f_{r\ell\max}] \in R^1$, where $\mu_{f_{r\ell}}(f_{r\ell}(x, \hat{p}_{r\ell})) = 0$ if $f_{r\ell}(x, \hat{p}_{r\ell}) \geq f_{r\ell\max}$, and $\mu_{f_{r\ell}}(f_{r\ell}(x, \hat{p}_{r\ell})) = 1$ if $f_{r\ell}(x, \hat{p}_{r\ell}) \leq f_{r\ell\min}$.

In order to determine these membership functions appropriately, let us assume that the decision maker sets $p_{r\ell\min}$, $p_{r\ell\max}$, $f_{r\ell\min}$ and $f_{r\ell\max}$ as follows.

At first, the decision maker specifies the parameters $p_{r\ell\min}$ and $p_{r\ell\max}$ in his/her subjective manner, where $p_{r\ell\min}$ is an acceptable minimum value and $p_{r\ell\max}$ is a sufficiently satisfactory minimum value, and sets the intervals $p_{r\ell} = [p_{r\ell\min}, p_{r\ell\max}]$, $r = 1, \dots, q, \ell = 1, \dots, k_r$. Corresponding to the interval $p_{r\ell}$, $f_{r\ell\min}$ can be obtained by solving the following linear programming problem.

$$f_{r\ell\min} \stackrel{\text{def}}{=} \min_{x \in X} f_{r\ell}(x, p_{r\ell\min}) \tag{3}$$

In order to obtain $f_{r\ell\max}$, we first solve the following linear programming problems, $\min_{x \in X} f_{r\ell}(x, p_{r\ell\max})$, $r = 1, \dots, q, \ell = 1, \dots, k_r$. Let $x_{r\ell}$, $r = 1, \dots, q, \ell = 1, \dots, k_r$ be the above optimal solution. Using the optimal solutions $x_{r\ell}$, $f_{r\ell\max}$ can be obtained as follows.

$$f_{r\ell\max} \stackrel{\text{def}}{=} \max_{s=1, \dots, q, t=1, \dots, k_s, s \neq r, t \neq \ell} f_{r\ell}(x_{st}, p_{r\ell\max}) \tag{4}$$

It should be noted here that, from (2), $\mu_{\tilde{f}_{r\ell}}(f_{r\ell}(x, \hat{p}_{r\ell}))$ and $\mu_{\tilde{p}_{r\ell}}(\hat{p}_{r\ell})$ are conflict with respect to $\hat{p}_{r\ell}$ for any $x \in X$. Here, let us assume that the decision maker adopts the fuzzy decision [4] in order to integrate both the membership functions $\mu_{\tilde{f}_{r\ell}}(f_{r\ell}(x, \hat{p}_{r\ell}))$ and $\mu_{\tilde{p}_{r\ell}}(\hat{p}_{r\ell})$. Then, the integrated membership function can be defined as follows.

$$\mu_{D_{f_{r\ell}}}(x, \hat{p}_{r\ell}) \stackrel{\text{def}}{=} \min \left\{ \mu_{\tilde{p}_{r\ell}}(\hat{p}_{r\ell}), \mu_{\tilde{f}_{r\ell}}(f_{r\ell}(x, \hat{p}_{r\ell})) \right\} \tag{5}$$

Using the membership functions, $\mu_{D_{f_{r\ell}}}(x, \hat{p}_{r\ell})$, HMOP5 can be transformed into the following form.

[HMOP6]

first level decision maker: DMI

$$\max_{x \in X, \hat{p}_{1\ell} \in (0,1), \ell=1, \dots, k_1} \left(\mu_{D_{f_{11}}}(x, \hat{p}_{11}), \dots, \mu_{D_{f_{1k_1}}}(x, \hat{p}_{1k_1}) \right)$$

.....

q-th level decision maker: DMq

$$\max_{x \in X, \hat{p}_{q\ell} \in (0,1), \ell=1, \dots, k_q} \left(\mu_{D_{f_{q1}}}(x, \hat{p}_{q1}), \dots, \mu_{D_{f_{qk_q}}}(x, \hat{p}_{qk_q}) \right)$$

In order to deal with HMOP6, we introduce a D_f -Pareto optimal solution concept.

Definition 1

$x^* \in X, \hat{p}_{r\ell}^* \in (0, 1), r = 1, \dots, q, \ell = 1, \dots, k_r$ is said to be a D_f -Pareto optimal solution to HMOP6, if and only if there does not exist another $x \in X, \hat{p}_{r\ell} \in (0, 1), r = 1, \dots, q, \ell = 1, \dots, k_r$ such that $\mu_{D_{f_{r\ell}}}(x, \hat{p}_{r\ell}) \geq \mu_{D_{f_{r\ell}}}(x^*, \hat{p}_{r\ell}^*) = 1, \dots, q, \ell = 1, \dots, k_r$, with strict inequality holding for at least one r and ℓ .

In order to generate a candidate of the satisfactory solution from among a D_f -Pareto optimal solution set, it has been suggested to ask the decision makers to specify their reference levels of achievement of the membership functions, called the reference membership values [4]. However, considering the imprecise nature of the decision makers' judgments, it seems to be more appropriate to obtain fuzzy-valued assessments of the reference membership values such as "it should be between $[\mu_{r\ell}^L, \mu_{r\ell}^R]$ for $\mu_{D_{f_{r\ell}}}(x, \hat{p}_{r\ell})$ " called reference membership intervals [8], where $\mu_{r\ell}^L < \mu_{r\ell}^R$, and

$$\mu_r^L = (\mu_{r1}^L, \mu_{r2}^L, \dots, \mu_{rk_r}^L), r = 1, \dots, q, \tag{6}$$

$$\mu_r^R = (\mu_{r1}^R, \mu_{r2}^R, \dots, \mu_{rk_r}^R), r = 1, \dots, q, \tag{7}$$

$$\mu^L = (\mu_1^L, \mu_2^L, \dots, \mu_q^L), \tag{8}$$

$$\mu^R = (\mu_1^R, \mu_2^R, \dots, \mu_q^R). \tag{9}$$

Once the reference membership intervals are specified, the corresponding D_f -Pareto optimal solution, which is, in a sense, close to their requirement, is obtained by solving the following minmax problem.

[MINMAX1(μ^L, μ^R)]

$$\min_{x \in X, \hat{p}_{r\ell} \in (0,1), r=1, \dots, q, \ell=1, \dots, k_r, \lambda \in \Lambda_1} \lambda \tag{10}$$

subject to

$$\mu_{r\ell}^R - \mu_{\hat{p}_{r\ell}}^{\bar{r}}(f_{r\ell}(x, \hat{p}_{r\ell})) \leq \lambda/d_{r\ell}, \tag{11}$$

$$\begin{aligned} \mu_{r\ell}^R - \mu_{\hat{p}_{r\ell}}^{\bar{r}}(\hat{p}_{r\ell}) &\leq \lambda/d_{r\ell}, \\ r = 1, \dots, q, \ell = 1, \dots, k_r \end{aligned} \tag{12}$$

where

$$d_{r\ell} \stackrel{\text{def}}{=} \left(\frac{1}{\mu_{r\ell}^R - \mu_{r\ell}^L} \right) / \left(\sum_{\ell=1}^{k_r} \frac{1}{\mu_{r\ell}^R - \mu_{r\ell}^L} \right), \tag{13}$$

$$\begin{aligned} \Lambda_1 \stackrel{\text{def}}{=} [&\max_{r=1, \dots, q, \ell=1, \dots, k_r} d_{r\ell}(\mu_{r\ell}^R - 1) + \delta, \\ &\min_{r=1, \dots, q, \ell=1, \dots, k_r} d_{r\ell}\mu_{r\ell}^R - \delta], \end{aligned} \tag{14}$$

and δ is a sufficiently small positive constant. As a special case of the above definition (13) where $\mu_{r\ell}^L = \mu_{r\ell}^R, \ell = 1, \dots, k_r$, let us define $d_{r\ell} = 1/k_r, \ell = 1, \dots, k_r$.

It should be noted here that, in general, the optimal solution of MINMAX1(μ^L, μ^R) does not reflect the hierarchical structure between q decision makers where the upper level decision maker can take priority for his/her distribution functions over the lower level decision makers. In order to cope with such a hierarchical preference structure between q decision makers, we introduce the concept of the decision power [2] $w = (w_1, \dots, w_q)$, where the r -th level decision maker (DM r) can specify the decision power w_{r+1} in his/her subjective manner and the last decision maker (DM q) has no decision power. In order to reflect the hierarchical preference structure between multiple decision makers, the decision powers $w = (w_1, \dots, w_q)$ have to satisfy the following inequality conditions.

$$w_1 = 1 \geq w_2 \geq \dots \geq w_{q-1} \geq w_q > 0 \tag{15}$$

Then, the corresponding modified MINMAX1(μ^L, μ^R) is reformulated as follows.

[MINMAX2 (μ^L, μ^R, w)]

$$\min_{x \in X, \hat{p}_{r\ell} \in (0,1), r=1, \dots, q, \ell=1, \dots, k_r, \lambda \in \Lambda_2} \lambda \tag{16}$$

subject to

$$\mu_{r\ell}^R - \mu_{\hat{f}_{r\ell}}(f_{r\ell}(x, \hat{p}_{r\ell})) \leq \lambda / (d_{r\ell} \cdot w_r), \quad (17)$$

$$\begin{aligned} \mu_{r\ell}^R - \mu_{\hat{p}_{r\ell}}(\hat{p}_{r\ell}) &\leq \lambda / (d_{r\ell} \cdot w_r), \\ r &= 1, \dots, q, \ell = 1, \dots, k_r \end{aligned} \quad (18)$$

where

$$\Lambda_2 \stackrel{\text{def}}{=} \left[\max_{r=1, \dots, q, \ell=1, \dots, k_r} d_{r\ell} \cdot w_r \cdot (\mu_{r\ell}^R - 1) + \delta, \min_{r=1, \dots, q, \ell=1, \dots, k_r} d_{r\ell} \cdot w_r \cdot \mu_{r\ell}^R - \delta \right], \quad (19)$$

and δ is a sufficiently small positive constant.

Because of $c_{r\ell}^2 x + \alpha_{r\ell}^2 > 0$, the constraint (17) can be transformed as follows.

$$\hat{p}_{r\ell} \leq T_{r\ell} \left(\frac{\mu_{\hat{f}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda / (d_{r\ell} \cdot w_r)) - (c_{r\ell}^1 x + \alpha_{r\ell}^1)}{c_{r\ell}^2 x + \alpha_{r\ell}^2} \right) \quad (20)$$

where $\mu_{\hat{f}_{r\ell}}^{-1}(\cdot)$ is an inverse function of $\mu_{\hat{f}_{r\ell}}(\cdot)$. From the constraints (18), it holds that $\hat{p}_{r\ell} \geq \mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda / (d_{r\ell} \cdot w_r))$, where $\mu_{\hat{p}_{r\ell}}^{-1}(\cdot)$ is an inverse function of $\mu_{\hat{p}_{r\ell}}(\cdot)$. Therefore, the constraint (20) can be reduced to the following inequality where a permissible probability level $\hat{p}_{r\ell}$ is disappeared.

$$\begin{aligned} &\mu_{\hat{f}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda / (d_{r\ell} \cdot w_r)) - (c_{r\ell}^1 x + \alpha_{r\ell}^1) \\ &\geq T_{r\ell}^{-1}(\mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda / (d_{r\ell} \cdot w_r))) \cdot (c_{r\ell}^2 x + \alpha_{r\ell}^2) \end{aligned} \quad (21)$$

Then, MINMAX2 (μ^L, μ^R, w) can be equivalently reduced to the following problem.

[MINMAX3 (μ^L, μ^R, w)]

$$\min_{x \in X, \lambda \in \Lambda_2} \lambda \quad (22)$$

subject to

$$\begin{aligned} &\mu_{\hat{f}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda / (d_{r\ell} \cdot w_r)) - (c_{r\ell}^1 x + \alpha_{r\ell}^1) \\ &\geq T_{r\ell}^{-1}(\mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda / (d_{r\ell} \cdot w_r))) \cdot (c_{r\ell}^2 x + \alpha_{r\ell}^2), r = 1, \dots, q, \ell = 1, \dots, k_r \end{aligned} \quad (23)$$

It should be noted here that an optimal solution (x^*, λ^*) of MINMAX3(μ^L, μ^R, w) can be obtained by combined use of the bisection method with respect to λ and the first-phase of the two-phase simplex method of linear programming.

The relationship between the optimal solution (x^*, λ^*) of $\text{MINMAX3}(\mu^L, \mu^R, w)$ and D_f -Pareto optimal solutions can be characterized by the following theorem.

Theorem 1

If $x^* \in X, \lambda^* \in \Lambda_2$ is a unique optimal solution of $\text{MINMAX3}(\mu^L, \mu^R, w)$ then $x^* \in X, \hat{p}_{r\ell}^* = \mu_{\hat{p}_{r\ell}^*}^{-1}(\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r)), r = 1, \dots, q, \ell = 1, \dots, k_r$ is a D_f -Pareto optimal solution.

Proof

From (23), it holds that $\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r) \leq \mu_{\hat{p}_{r\ell}}(f_{r\ell}(x^*, \mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r))))), r = 1, \dots, q, \ell = 1, \dots, k_r$. Assume that $x^* \in X, \mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r)), r = 1, \dots, q, \ell = 1, \dots, k_r$ is not a D_f -Pareto optimal solution. Then, there exist $x \in X, \hat{p}_{r\ell}, r = 1, \dots, q, \ell = 1, \dots, k_r$ such that $\mu_{D_{f_{r\ell}}}(x, \hat{p}_{r\ell}) = \min \{ \mu_{\hat{p}_{r\ell}}(\hat{p}_{r\ell}), \mu_{\hat{p}_{r\ell}}(f_{r\ell}(x, \hat{p}_{r\ell})) \} \geq \mu_{D_{f_{r\ell}}}(x^*, \mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r))) = \mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r), r = 1, \dots, q, \ell = 1, \dots, k_r$ strict inequality holding for at least one r and ℓ . Then it holds that

$$\mu_{\hat{p}_{r\ell}}(\hat{p}_{r\ell}) \geq \mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r), \tag{24}$$

$$\mu_{\hat{p}_{r\ell}}(f_{r\ell}(x, \hat{p}_{r\ell})) \geq \mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r), \tag{25}$$

$r = 1, \dots, q, \ell = 1, \dots, k_r$. From the definition (2), the inequalities (24) and (25) can be transformed into the inequalities, $\hat{p}_{r\ell} \geq \mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r)), \hat{p}_{r\ell} \leq T_{r\ell} \left(\frac{\mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r)) - (c_{r\ell}^1 x + \alpha_{r\ell}^1)}{c_{r\ell}^2 x + \alpha_{r\ell}^2} \right)$. This means that there exists some $x \in X$ such that $\mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r)) - (c_{r\ell}^1 x + \alpha_{r\ell}^1) \geq T_{r\ell}^{-1}(\mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r))) \cdot (c_{r\ell}^2 x + \alpha_{r\ell}^2), r = 1, \dots, q, \ell = 1, \dots, k_r$ which contradicts the fact that $x^* \in X, \lambda^* \in \Lambda_2$ is a unique optimal solution to $\text{MINMAX3}(\mu^L, \mu^R, w)$.

3 An Interactive Algorithm

In this section, we propose an interactive algorithm to obtain a satisfactory solution of the decision makers from among D_f -Pareto optimal solution set. Unfortunately, it is not guaranteed that the optimal solution (x^*, λ^*) of $\text{MINMAX3}(\mu^L, \mu^R, w)$ is D_f -Pareto optimal, if (x^*, λ^*) is not unique. In order to guarantee D_f -Pareto optimality, we first assume that $\sum_{r=1}^q k_r$ constraints (23) of $\text{MINMAX3}(\mu^L, \mu^R, w)$ are active at the optimal solution (x^*, λ^*) . If one of the constraints of (23) is inactive, i.e.,

$$\begin{aligned} & \mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r)) - (c_{r\ell}^1 x^* + \alpha_{r\ell}^1) \\ & > T_{r\ell}^{-1}(\mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r))) \cdot (c_{r\ell}^2 x^* + \alpha_{r\ell}^2), \end{aligned} \tag{26}$$

we can convert the inactive constraint (26) into the active one by applying the bisection method, where $G_{r\ell}(\mu_{r\ell}^R) = \mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r)) - f_{r\ell}(x^*, \mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r)))$.

[The bisection method for the inactive constraint]

Step 1. Set $q_{r\ell}^L \leftarrow \lambda^*/(d_{r\ell} \cdot w_r)$, $q_{r\ell}^R \leftarrow \lambda^*/(d_{r\ell} \cdot w_r) + 1$.

Step 2. Set $q_{r\ell} \leftarrow (q_{r\ell}^R + q_{r\ell}^L)/2$.

Step 3. If $G_{r\ell}(q_{r\ell}) > 0$ then $q_{r\ell}^L \leftarrow q_{r\ell}$ and go to Step 2, else if $G_{r\ell}(q_{r\ell}) < 0$ then $q_{r\ell}^R \leftarrow q_{r\ell}$ and go to Step 2, else if $G_{r\ell}(q_{r\ell}) = 0$, then update the reference membership value as $\mu_{r\ell}^R \leftarrow q_{r\ell}$ and stop.

For the optimal solution (x^*, λ^*) of MINMAX3(μ^L, μ^R, w), where the active conditions of the constraints (23) are satisfied, we solve the D_f -Pareto optimality test problem formulated as follows.

[Test problem for D_f -Pareto optimality]

$$\max_{x \in X, \epsilon_{r\ell} \geq 0, r=1, \dots, q, \ell=1, \dots, k_r} w = \sum_{r=1}^q \sum_{\ell=1}^{k_r} \epsilon_{r\ell} \quad (27)$$

subject to

$$\begin{aligned} & T_{r\ell}^{-1}(\mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda^*/(d_{r\ell} r))) \cdot (c_{r\ell}^2 x + \alpha_{r\ell}^2) + (c_{r\ell}^1 x + \alpha_{r\ell}^1) + \epsilon_{r\ell} \\ & = T_{r\ell}^{-1}(\mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r))) \cdot (c_{r\ell}^2 x^* + \alpha_{r\ell}^2) \\ & + (c_{r\ell}^1 x^* + \alpha_{r\ell}^1), r = 1, \dots, q, \ell = 1, \dots, k_r \end{aligned} \quad (28)$$

For the optimal solution of the above test problem, the following theorem holds.

Theorem 2

Let $\tilde{x} \in X, \tilde{\epsilon}_{r\ell} \geq 0, r = 1, \dots, q, \ell = 1, \dots, k_r$ be an optimal solution of the test problem (27–28). If $w = 0, x^* \in X, \mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r)), r = 1, \dots, q, \ell = 1, \dots, k_r$ is a D_f -Pareto optimal solution.

Proof

From the active conditions of the constraints (23), it holds that $\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r) = \mu_{\hat{p}_{r\ell}}(f_{r\ell}(x^*, \mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r))))$, $r = 1, \dots, q, \ell = 1, \dots, k_r$. If $x^* \in X, \mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r)), r = 1, \dots, q, \ell = 1, \dots, k_r$ is not a D_f -Pareto optimal solution, there exists some $x \in X, \hat{p}_{r\ell}, r = 1, \dots, q, \ell = 1, \dots, k_r$ such that $\mu_{D_{f_{r\ell}}}(x, \hat{p}_{r\ell}) = \min \{ \mu_{\hat{p}_{r\ell}}(\hat{p}_{r\ell}), \mu_{\hat{p}_{r\ell}}(f_{r\ell}(x, \hat{p}_{r\ell})) \} \geq \mu_{D_{f_{r\ell}}}(x^*, \mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r))) = \mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r), r = 1, \dots, q, \ell = 1, \dots, k_r$ with strict inequality holding for at least one r and ℓ . This means that the following inequalities hold.

$$\mu_{\hat{p}_{r\ell}}(\hat{p}_{r\ell}) \geq \mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r), \quad (29)$$

$$\mu_{\hat{p}_{r\ell}}(f_{r\ell}(x, \hat{p}_{r\ell})) \geq \mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r), \quad (30)$$

$r = 1, \dots, q, \ell = 1, \dots, k_r$. This means that there exists some $x \in X, \hat{p}_{r\ell}, r = 1, \dots, q, \ell = 1, \dots, k_r$ such that $\mu_{r\ell}^{-1}(\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r)) \geq (c_{r\ell}^1 x + \alpha_{r\ell}^1) + T_{r\ell}^{-1}(\mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r))) \cdot (c_{r\ell}^2 x + \alpha_{r\ell}^2)$. Because of the active conditions of the constraints (23), it holds that $T_{r\ell}^{-1}(\mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r))) \cdot (c_{r\ell}^2 x^* + \alpha_{r\ell}^2) + (c_{r\ell}^1 x^* + \alpha_{r\ell}^1) \geq T_{r\ell}^{-1}(\mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r))) \cdot (c_{r\ell}^2 x + \alpha_{r\ell}^2) + (c_{r\ell}^1 x + \alpha_{r\ell}^1), r = 1, \dots, q, \ell = 1, \dots, k_r$, with strict inequality holding for at least one r and ℓ . This contradicts the fact that $w = 0$.

Now, following the above discussions, we can present the interactive algorithm in order to derive a satisfactory solution from among a D_f -Pareto optimal solution set.

[An interactive algorithm]

Step 1: Each of the decision makers (DM $r, r = 1, \dots, q$) sets his/her membership functions $\mu_{\hat{p}_{r\ell}}^{\pm}(\hat{p}_{r\ell}), \ell = 1, \dots, k_r$ in his/her subjective manner.

Step 2: Corresponding to the membership functions $\mu_{\hat{p}_{r\ell}}(\hat{p}_{r\ell}), \ell = 1, \dots, k_r$, each of the decision makers (DM $r, r = 1, \dots, q$) sets his/her membership functions $\mu_{f_{r\ell}}(f_{r\ell}(x, \hat{p}_{r\ell})), \ell = 1, \dots, k_r$.

Step 3: Set the initial decision powers as $w_r = 1, r = 1, \dots, q$.

Step 4: Each of the decision makers (DM $r, r = 1, \dots, q$) sets his/her initial reference membership intervals $[\mu_{r\ell}^L, \mu_{r\ell}^R], \ell = 1, \dots, k_r$ in his/her subjective manner, where $\mu_{r\ell}^R = 1, \ell = 1, \dots, k_r$.

Step 5: Solve MINMAX3(μ^L, μ^R, w) by combined use of the bisection method and the first-phase of the two-phase simplex method of linear programming. If the active condition of the constraints (23) is not satisfied at the optimal solution (x^*, λ^*), then the bisection method with respect to the value of the right-hand-side of the reference membership value $[\mu_{r\ell}^L, \mu_{r\ell}^R]$ is applied, and D_f -Pareto optimality test problem is solved.

Step 6: If each of the decision makers (DM $r, r = 1, \dots, q$) is satisfied with the current values of the D_f -Pareto optimal solution $\mu_{D_{f_{r\ell}}}(x^*, \hat{p}_{r\ell}^*), \ell = 1, \dots, k_r$, where $\hat{p}_{r\ell}^* = \mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda^*)/(d_{r\ell} \cdot w_r)$, then stop. Otherwise, let the s -th level decision maker (DM s) be the uppermost of the decision makers who are not satisfied with the current values. Considering the current values of his/her membership functions, DM s updates his/her decision power w_{s+1} and/or his/her reference membership intervals $[\mu_{s\ell}^L, \mu_{s\ell}^R], \ell = 1, \dots, k_s$ according to the following two rules, and return to Step 5.

Rule 1 w_{s+1} must be set as $w_{s+1} \leq w_s$. After updating w_{s+1} , if $w_{s+1} \leq w_t, s + 2 \leq t \leq q, w_t$ is replaced by w_{s+1} ($w_t \leftarrow w_{s+1}$).

Rule 2 Before updating DM s 's reference membership intervals $[\mu_{s\ell}^L, \mu_{s\ell}^R], \ell = 1, \dots, k_s$, the other decision makers' reference membership intervals are fixed as the current values $(\mu_{r\ell}^L, \mu_{r\ell}^R \leftarrow \mu_{D_{f_{r\ell}}}(x^*, \hat{p}_{r\ell}^*), r = 1, \dots, q, r \neq s, \ell = 1, \dots, K_r)$.

4 A Numerical Example

In order to demonstrate the proposed method for HMOP1, we consider the following hierarchical two-objective stochastic linear programming problem under three hypothetical decision makers.

[HMOP1]

first level decision maker: DM1

$$\min \bar{Z}_{11}(x) = (c_{11}^1 + \bar{t}_{11}c_{11}^2)x + (\alpha_{11}^1 + \bar{t}_{11}\alpha_{11}^2)$$

$$\min \bar{Z}_{12}(x) = (c_{12}^1 + \bar{t}_{12}c_{12}^2)x + (\alpha_{12}^1 + \bar{t}_{12}\alpha_{12}^2)$$

second level decision maker: DM2

$$\min \bar{Z}_{21}(x) = (c_{21}^1 + \bar{t}_{21}c_{21}^2)x + (\alpha_{21}^1 + \bar{t}_{21}\alpha_{21}^2)$$

$$\min \bar{Z}_{22}(x) = (c_{22}^1 + \bar{t}_{22}c_{22}^2)x + (\alpha_{22}^1 + \bar{t}_{22}\alpha_{22}^2)$$

third level decision maker: DM3

$$\min \bar{Z}_{31}(x) = (c_{31}^1 + \bar{t}_{31}c_{31}^2)x + (\alpha_{31}^1 + \bar{t}_{31}\alpha_{31}^2)$$

$$\min \bar{Z}_{32}(x) = (c_{32}^1 + \bar{t}_{32}c_{32}^2)x + (\alpha_{32}^1 + \bar{t}_{32}\alpha_{32}^2)$$

subject to $x \in X \stackrel{\text{def}}{=} \{x \in R^{10} | a_i x \leq b_i, i = 1, \dots, 7, x \geq 0\}$

In HMOP1, $x = (x_1, x_2, \dots, x_{10})^T$ is the decision column vector, $a_i, i = 1, \dots, 7, c_{r\ell}^1, c_{r\ell}^2, r = 1, 2, 3, \ell = 1, 2, \alpha_{r\ell}^1, \alpha_{r\ell}^2, b_i, i = 1, \dots, 7$ are the constant coefficients which are same to [13]. $t_{r\ell}(\omega), r = 1, 2, 3, \ell = 1, 2$ are Gaussian random variables defined as $\bar{t}_{11} \sim N(4, 2^2), \bar{t}_{12} \sim N(3, 3^2), \bar{t}_{21} \sim N(3, 1^2), \bar{t}_{22} \sim N(3, 2^2), \bar{t}_{31} \sim N(3, 2^2), \bar{t}_{32} \sim N(3, 3^2)$.

According to the proposed interactive algorithm, we assume that the hypothetical decision makers (DM1, DM2 and DM3) set their membership functions as follows (Step 1 and 2).

$$\mu_{\bar{p}_{r\ell}}(\hat{p}_{r\ell}) = \frac{p_{r\ell\min} - \hat{p}_{r\ell}}{p_{r\ell\max} - p_{r\ell\min}}$$

$$\mu_{\bar{f}_{r\ell}}(f_{r\ell}(x, \hat{p}_{r\ell})) = \frac{f_{r\ell}(x, \hat{p}_{r\ell}) - f_{r\ell\max}}{f_{r\ell\min} - f_{r\ell\max}}$$

where

$$\begin{aligned}
 [p_{11\min}, p_{11\max}] &= [0.023, 0.959], [p_{12\min}, p_{12\max}] = [0.015, 0.993], \\
 [p_{21\min}, p_{21\max}] &= [0.001, 0.999], [p_{22\min}, p_{22\max}] = [0.259, 0.995], \\
 [p_{31\min}, p_{31\max}] &= [0.136, 0.859], [p_{32\min}, p_{32\max}] = [0.001, 0.987], \\
 [p_{11\min}, p_{11\max}] &= [2000, 2200], [p_{12\min}, p_{12\max}] = [400, 700], \\
 [p_{21\min}, p_{21\max}] &= [800, 1000], [p_{22\min}, p_{22\max}] = [650, 800], \\
 [p_{31\min}, p_{31\max}] &= [-1050, -950], [p_{32\min}, p_{32\max}] = [-200, 50].
 \end{aligned}$$

At Step 3, set the initial decision powers as $w_r = 1, r = 1, 2, 3$. At Step 4, set the initial reference membership intervals as $[\mu_{s\ell}^L, \mu_{s\ell}^R] = [0.8, 1], r = 1, 2, 3, \ell = 1, 2$. At Step 5, solve MINMAX3(μ^L, μ^R, w) by combined use of the bisection method with respect to λ and the first-phase of the two-phase simplex method of linear programming. The corresponding D_f -Pareto optimal solution is obtained as

$$\mu_{D_{r\ell}}(x^*, \mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r))) = 0.5452, r = 1, 2, 3, \ell = 1, 2.$$

For the optimal solution (x^*, λ^*) , DMI updates his/her decision power as $w_2 = 0.8$ in order to improve his/her own membership functions at the expense of the membership functions of the lower level decision makers (Step 6), and go back to Step 5. Then, the corresponding optimal solution is obtained as

$$\mu_{D_{f_{1\ell}}}(x^*, \mu_{\hat{p}_{r\ell}}^{-1}(\hat{\mu}_{1\ell} - \lambda^*/(d_{1\ell} \cdot w_1))) = 0.6059, \ell = 1, 2,$$

$$\mu_{D_{r\ell}}(x^*, \mu_{\hat{p}_{r\ell}}^{-1}(\mu_{r\ell}^R - \lambda^*/(d_{r\ell} \cdot w_r))) = 0.5074, r = 2, 3, \ell = 1, 2.$$

For this optimal solution, DMI is satisfied with the current values of the membership functions, but DM2 is not satisfied with the current values. Therefore, DM2 updates his/her decision power as $w_3 = 0.75$ and the optimal solution is obtained as

$$\mu_{D_{f_{1\ell}}}(x^*, \mu_{\hat{p}_{1\ell}}^{-1}(\hat{\mu}_{1\ell} - \lambda^*/(d_{1\ell} \cdot w_1))) = 0.6220, \ell = 1, 2,$$

$$\mu_{D_{f_{2\ell}}}(x^*, \mu_{\hat{p}_{2\ell}}^{-1}(\hat{\mu}_{2\ell} - \lambda^*/(d_{2\ell} \cdot w_2))) = 0.5275, \ell = 1, 2,$$

$$\mu_{D_{f_{3\ell}}}(x^*, \mu_{\hat{p}_{3\ell}}^{-1}(\hat{\mu}_{3\ell} - \lambda^*/(d_{3\ell} \cdot w_3))) = 0.4960, \ell = 1, 2,$$

Since the decision makers (DMI and DM2) are satisfied with current values of the membership functions, but the third level decision maker (DM3) is not satisfied with the current values. Therefore, DM3 updates his/her reference membership intervals as $[\mu_{31}^L, \mu_{31}^R] = [0.52, 0.53], [\mu_{32}^L, \mu_{32}^R] = [0.48, 0.49]$, in order to improve $\mu_{D_{p_{31}}}(x, f_{31})$ at the expense of $\mu_{D_{p_{32}}}(x, f_{32})$. Then, since the decision makers (DMI, DM2, and DM3) are satisfied with the following optimal solution, stop the interactive processes.

$$\mu_{D_{f_{1\ell}}}(x^*, \mu_{\hat{p}_{1\ell}}^{-1}(\hat{\mu}_{1\ell} - \lambda^*/(d_{1\ell} \cdot w_1))) = 0.6206, \ell = 1, 2,$$

$$\mu_{D_{f_{2\ell}}}(x^*, \mu_{\hat{p}_{2\ell}}^{-1}(\hat{\mu}_{2\ell} - \lambda^*/(d_{2\ell} \cdot w_2))) = 0.5257, \ell = 1, 2,$$

$$\mu_{D_{f_{31}}}(x^*, \mu_{\hat{p}_{31}}^{-1}(\hat{\mu}_{31} - \lambda^*/(d_{31} \cdot w_3))) = 0.5281,$$

$$\mu_{D_{f_{32}}}(x^*, \mu_{\hat{p}_{32}}^{-1}(\hat{\mu}_{32} - \lambda^*/(d_{32} \cdot w_3))) = 0.4881.$$

5 Conclusion

In this paper, we have proposed an interactive decision making method for hierarchical multiobjective stochastic linear programming problems to obtain a satisfactory solution from among a Pareto optimal solution set. In the proposed method, by considering the conflict between permissible objective levels and permissible probability levels, the corresponding membership functions are integrated through the fuzzy decision. In the integrated membership space, the candidate of a satisfactory solution is obtained from among Pareto optimal solution set by updating the reference membership intervals and/or the decision powers. In our proposed method, it is expected to obtain the satisfactory solution, in which the proper balance between permissible objective values and permissible probability levels are attained.

References

1. Anandalingam G (1988) A mathematical programming model of decentralized multi-Level systems. *J Oper Res Soc* 39:1021–1033
2. Lai Y-J (1996) Hierarchical optimization : a satisfactory solution. *Fuzzy Sets Syst* 77:321–335
3. Lee ES, Shih H (2001) *Fuzzy and multi-level decision making*. Springer, London
4. Sakawa M (1993) *Fuzzy sets and interactive multiobjective optimization*. Plenum Press, New York
5. Sakawa M, Nishizaki I, Katagiri H (2011) *Fuzzy stochastic multiobjective programming*. Springer, New York
6. Sakawa M, Katagiri H, Kato K (2001) An interactive fuzzy satisficing method for multiobjective stochastic programming problems using fractile optimization model. In: *The 10th IEEE international conference on fuzzy systems 3*, Melbourne, pp 25–31
7. Sakawa M, Kato K, Katagiri H (2004) An Interactive fuzzy satisficing method for multiobjective linear programming problems with random variable Coefficients through a probability maximization model. *Fuzzy Sets Syst* 146:205–220
8. Sakawa M, Yano H (1985) Interactive fuzzy decision-making for multi-objective nonlinear programming using reference membership intervals. *Int J Man Mach Stud* 23:407–421
9. Shih H, Lai Y-J, Lee ES (1996) Fuzzy approach for multi-level programming problems. *Comput Oper Res* 23:73–91

10. Wen U-P, Hsu S-T (1991) Linear bi-level programming problems—a review. *J Oper Res Soc* 42:125–133
11. Yano H (2010) A fuzzy approach for hierarchical multiobjective linear programming problems. Lecture notes in engineering and computer science. In: Proceedings of the international multiconference of engineers and computer scientists 2010, IMECS 2010, 17–19 March 2010, Hong Kong, pp 2098–2103
12. Yano H (2012) Interactive fuzzy decision making for hierarchical multiobjective stochastic linear programming problems. Lecture Notes in Engineering and Computer Science. In: Proceedings of the international multiconference of engineers and computer scientists 2012, IMECS 2012, 14–16 March, Hong Kong, pp 1623–1628
13. Yano H (2012b) Hierarchical multiobjective stochastic linear programming problems considering both probability maximization and fractile optimization. *IAENG Int J Appl Math* 42(2):91–98
14. Yano H, Matsui K (2011) Two fuzzy Approaches for multiobjective stochastic programming and multiobjective fuzzy random linear programming through a probability maximization model. *IAENG Int J Comput Sci* 38(3):234–241

Subsidy, Revenue and Tragedies in Heterogeneous Communities

Mbuyu Sumbwanyambe and Andre L. Nel

Abstract Developing countries have embarked on the promotion of “ICT access for all” through subsidized Information and Communication Technologies (ICTs), especially in underserved areas of such countries. The main aim of the “ICT access for all” is to extend the communication services to the large areas of underserved regions through subsidized communication services. In some instances, subsidies may lead to high ICT penetration and high resource utilization while in some instances unsubsidized services may lead to low utilization of resources and low ICT penetration, which may eventually lead to market failure and destroy market efficiency. With explicitly defined objectives, regarding subsidy policy, however, developing countries always fall short on the implementation of such subsidy policy due to economic reasons and unrealistic subsidy driven pricing models. In this paper we investigate the impact of subsidy driven pricing model on resource utilization and revenue maximization in a developing country. We try to find a middle ground between promoting “ICT access for all” (given a subsidy and diverse income variations between the groups) and resource utilization in a network.

Keywords Heterogeneous communities · ICTs · Pricing developing country · Revenue · Subsidy · Tragedy of the commons · Tragedy of the anti-commons

M. Sumbwanyambe (✉) · A. L. Nel
Department of Mechanical Engineering Science, University of Johannesburg,
Auckland Park, P.O. Box 524, Johannesburg 2006, South Africa
e-mail: sumbwam@gmail.com

A. L. Nel
e-mail: andren@uj.ac.za

1 Introduction

In developing countries the amount and characteristics of consumers of ICT services affect revenue, Quality of Service (QoS) and have substantial financial and business model implications for providers and consumers alike [1]. On the whole, competitive environments with numerous service providers are more likely to offer lower costs, high QoS, and are more inclined to create and adjust to a dynamic communication markets. On the other hand, monopolistic service providers, who are common in developing countries, are generally, inclined to serving large urban areas, and are reluctant to enter smaller and rural markets where there is little or no possibility of deriving a profit. To that effect, governments of developing countries have required and persuaded such monopolistic suppliers to serve rural areas through a combination of service requirements and subsidies [2–4], but have had limited success in creating competitive rural markets.

In much of the developing world, therefore, rural access remains limited in scope, regularity, and quality. New entrants typically give precedence to the profitable urban markets, and may enter rural areas and underserved regions with infrastructure that is insufficient for effective service provision. This observation has led to the exclusion of underserved areas or rural communities from the urban areas in terms of ICT access [5].

The exclusion of rural communities and underserved areas from ICT resources in developing countries perpetuates the problems associated with socio-economic deprivation (specifically poverty). Thus, most rural and underserved areas in developing countries lack access to basic ICT services, creating a gap between the information “haves” and “have nots” [4, 6, 7].

The gap between the information “haves” and “have nots” creates a group of customers whose preferences and socio-economic status are very diverse. This heterogeneity can arise from economical imbalances due to poverty (income disparity) or from different geographic locations or both [8]. To effectively rectify the existing inequalities in access to ICT services, governments of developing countries use different economic mechanisms, commonly funded by taxes [2].

The current practice in such countries has been to promote information access by means of tiered subsidized pricing scheme [2, 9, 10]. However, in countries where technical expertise is still in its embryonic stages, subsidized pricing schemes have proved to be very difficult to implement, and, in certain cases, it is unclear whether the user’s reservation price and sensitivity towards price are taken into consideration [1, 3]. Subsidy driven pricing mechanisms have, often times, resulted in controversies and have on certain occasion led to resource tragedies in such countries.

In a heterogeneous society (especially in developing countries), subsidized resource use can lead to the tragedy of the (anti) commons [11]. If too large a subsidy is given, consumers will demand more than is available creating the tragedy of the commons [12]. In order to prevent the tragedy of the commons, service providers and government policy makers have to cut the subsidy level

which will eventually reduce the number of consumers. The decrease in the number of consumers (especially rural consumers) may then lead to the tragedy of the anti-commons [11], leaving a gap in government policy and objectives of promoting social and economic growth. In such an eventuality, the government will have to lower the prices yet again, through a subsidy, in order fulfill their objectives, creating a see-saw effect.

In order to prevent this tragedy of the commons and tragedy of the anti-commons in telecom markets, a correct subsidy enhanced price structure must be developed for such a market. The use of the correct subsidy level to promote ICT access in underserved areas can be seen as a balanced two tier-pricing scheme (taking into consideration customer sensitivities and price thresholds). Furthermore, it can be seen as a means of preventing the tragedy of the (anti) commons and at the same time promoting social and economic growth. By setting up the right subsidy (an equilibrium subsidy) mechanism, to induce a desirable number of users, the service providers and government can achieve the optimal utility. In this paper, we study the impact of subsidies on the network resource usage by the information “have nots” and propose a subsidy driven pricing model that will mitigate the tragedy of the (anti) commons. We study all the possible outcomes of the subsidy on the information “have nots” and find a fair solution that will prevent the underusage or overusage of network resources by the subsidized group.

2 Why Should We Price the Internet in a Heterogeneous Community?

Pricing of ICTs, especially the internet, is of great interest to many organizations and researchers [1, 8, 13] because it presents a pricing optimization problem with multiple players. The players involved in the pricing of internet services include the service providers, the users and the government.

Currently, the internet is shared by a number of heterogeneous users who have different values, sensitivities, usage patterns and demands in the way they use the internet. When internet service providers set up a price for the internet service, users choose the amount of demand based on their own preferences as well as the prices set by ISPs. Therefore, the choice of pricing strategy is an optimal problem for ISP subject to users’ utility maximization and vice versa.

Typically, pricing of internet connectivity in a heterogeneous community is a complex matter that should involve appropriate and well executed economic decisions among the government, ISPs and the consumers. Even if it is government responsibility to provide free internet access to its underprivileged populace of the country, however, such a provision may result in what is called the “tragedy of the commons”. Especially, when there is no restriction on the number of internet subscribers, users may face no marginal usage costs, resulting in the internet being treated much like a public good [12]. In most of the other public

goods especially those that are not depleted with use; that is the usage of it does not affect the utility or payoff you get from it, government intervention or control is not necessary. However, this is clearly not the case with the internet and we should therefore expect it to exhibit problems that face other finite resources.

Overpricing of internet may result in what is known as the tragedy of anti-commons [1, 11], a problem that occurs when the prices of public good are overpriced and therefore no consumer is willing to pay for the services. The result is that such public goods are underutilized.

Sumbwanyambe and Nel [1, 8] have shown that pricing is a very effective tool in controlling the usage of resources in societies which are complicated like heterogeneous societies. Pricing mechanisms provide a resourceful method to ensure QoS guarantees in service provision and to regulate network resource usage. Additionally, pricing is very significant because it determines the revenue and financial viability of a venture and sometimes defines consumers' time cost of accessing or using a particular facility, affects the allocation of economic resources especially in dynamic markets and sometimes defines the ICT subsidies received by the rural people [1].

In a system such as the one found in developing countries, where heterogeneous users often compete for a same set of limited resources, pricing can, either on one hand, be used to provide resources for a certain group and denying others, while in certain circumstances pricing can be used to evenly distribute resources in communities with an observable income disparity among the users or consumers.

Thus far, researchers and other organization have proposed pricing models for a heterogeneous society or community. In fact, such pricing strategies have been proposed by Sumbwanyambe, Nel and Clarke [1], where a pricing model that will mitigate the tragedy of the commons and the free rider's problems in heterogeneous communities through differential pricing of telecommunication services is developed.

3 Pricing Model for Heterogeneous Communities

We consider a case of a developing country with heterogeneous consumers and an observable income disparity or geographical disadvantage; the information "haves" N_2 and "have-nots" N_1 . In heterogeneous societies, consumers have different service requirements and sensitivities towards price. In such societies, service providers can choose a suitable price based on consumers' price acceptance, sensitivities and threshold values and adjust their price accordingly so as to obtain an optimal price at which a both types of consumers will be more willing to pay for the service. Such pricing models can be learned through a market price adaptive research which estimates the threshold values of users and sensitivities in such an environment.

Within this current perspective, we assume that the N_j consumers are offered a subsidy so as to promote social and economic growth in underserved areas, while

the N_2 users pay the price that is set by the services provider. We take p_1 and p_2 as the price that is payable to the service provider by the heterogeneous consumers N_1 and N_2 , respectively, where $p_1 = \beta p_2$, and β is a subsidy factor which is in the range of $0 < \beta < 1$ is in the range of $(0, p_\infty)$ while c and r is the cost and the rate of providing network services respectively. The following definition is required.

Definition 1 The threshold or reserve price, p_{th1} and p_{th2} is the price at which a customer from group N_1 and N_2 is indifferent between subscribing to the current network or opting out of the current network.

Assuming that p_{th1} and p_{th2} is the reservation price, the consumer's utility function can thus be written as $u(p_{th1,2}, p_{1,2})$. If $p_{th1,2} \cdot p_{1,2} > 0$ the consumer has positive utility, if $p_{th1,2} \cdot p_{1,2} < 0$ then the consumer has a negative utility. For notational simplicity we assume that consumers attempt to maximize their utility subject to the reservation price of p_{th1} for N_1 and p_{th2} for N_2 , where $p_{th1} < p_{th2} < p_\infty$.

4 Tragedies and Resource Utilization in Heterogeneous Communities

Depending on the subsidy level, the total number of information “have nots” will increase or decrease leading to the tragedy of the commons or tragedy of the anti-commons respectively. We define the following:

Definition 2 The tragedy of the commons is a dilemma arising from the situation in which consumers, acting independently and rationally consulting their own self-interest will ultimately deplete a shared limited resource, even when it is clear that it is not in anyone's long-term interest for this to happen [12].

Definition 3 The tragedy of the anti-commons, describes a coordinated breakdown in a game where the existence of high prices for a public good or numerous holders of certain rights hinder the achievement of more socially desirable outcome [11].

In developing countries, it is obvious that pricing structures that are currently in use affect network usage and are not helping in promoting social and economical growth [2]. To that effect, subsidy driven pricing has been seen as a valuable tool in promoting social and economic growth in underserved regions. From the users' point of view, price affects their behavior in terms of network usage, which is correlated to their perception about the price for such a service. Therefore, choosing the right price and subsidy by the service provider and government, is of great importance in maximizing revenue and preventing resource tragedies. It is therefore imperative that government policy makers and service providers find a middle ground between “ICT access for all” and resource overuse or underuse given a subsidy β and users' reservation price i.e. p_{th1} and p_{th2} .

In heterogeneous communities the reservation price is associated with a number of factors. Primarily, any heterogeneous user is assumed to maximize his or her utility subject to a predetermined level of his or her income or reservation price [1]. Some of the probable factors associated with users' reservation price in heterogeneous communities are what is perceived as positive or negative utility experienced by the consumer. For example, for people who live on less than USD 2 a day [6], pricing the internet services beyond such an amount, i.e. USD 2, will constitute what is known as a negative utility towards the customer or consumer. To remove such a negative utility, a subsidy has to be paid by the government to the ISP. However, such a subsidy may have an adverse effect on resource usage, number of users and revenue maximization in a network.

To understand the impact of subsidies on user behavior and revenue maximization in developing countries, especially on the information "have nots", one has to understand how subsidies affect the number of users subscribing to an ISP. The number of users in this case determines the revenue and the number of packets sent in any network. For example, Figs. 1 and 2 show a relationship between packets and price given a government subsidy factor $\beta = 0.1$, $\beta = 1$, sensitivity $\alpha = 0.1$, $\lambda = 20$, $p_{th1} = 40$ and $p_{th2} = 100$. Analyses of the graphs show that the number of packets seems to fluctuate with an increasing or decreasing subsidy factor β . For instance, for $\beta = 0.1$, as in Fig. 1, the number of packets increases significantly due to an increase in the number of users, while the opposite is true for no subsidy i.e. $\beta = 1$. Obviously, there is a relationship between price and the number of users that will enable us to make some generalized observations about the entire information "have nots" population. We list these observations as below:

- a. There exists a finite price $p_1 > p_{th1}$ where a customer will not subscribe to any ISP.
- b. Information "have nots" are rational, they will prefer a subsidized price or a lower price to a higher price.
- c. The number of information "have nots" who will accept a given price in any heterogeneous communities of developing countries does not increase with price.

Given the above motivation, we can therefore define the number of packets for the N_1 and N_2 in a network as a function of price p_1 and p_2 and the reservation price p_{th1} and p_{th2} respectively. Jagannathan and Almeroth [2] and Sumbwanyambe [1] investigate user behavior upon price change using a demand function model. We adopt this model with some modifications to reflect our price setting and the number of users. When the price tends to zero the number of users increases significantly and when the price is very high the number of users decreases significantly. The following equations are given for the number of N_1 and N_2 users and number of packets in a network accordingly:

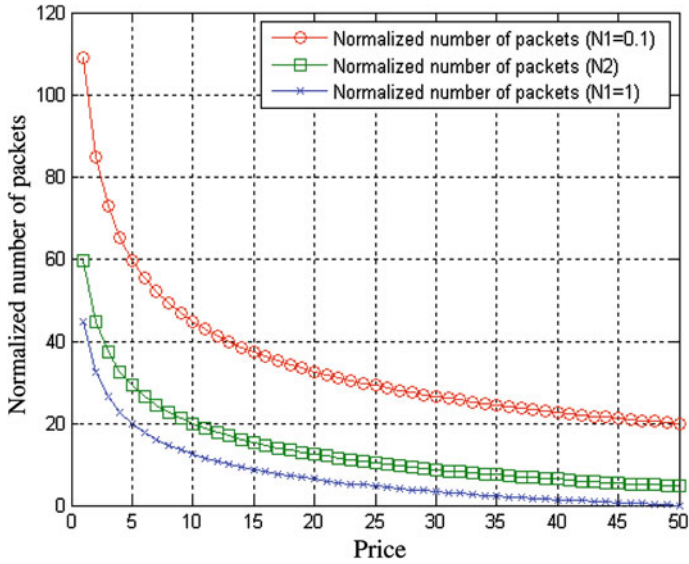


Fig. 1 Number of packets versus the price at different subsidy levels

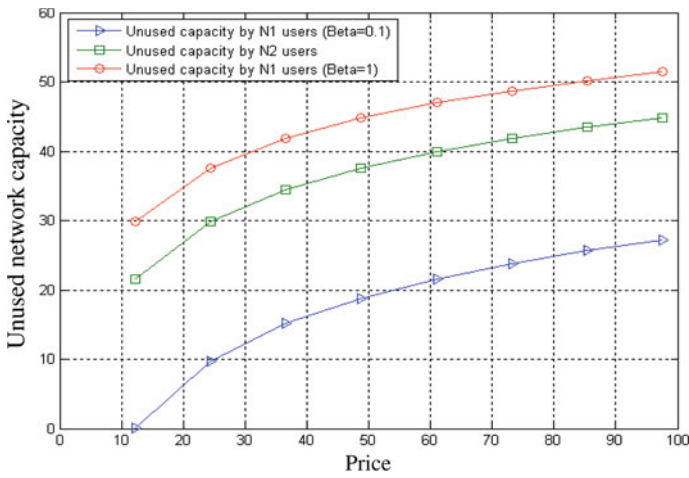


Fig. 2 Unused network capacity versus the price at different subsidy levels (Network capacity = 45 for N1 and N2)

$$N_1 = \left(\frac{\beta p_2}{p_{th1}} \right)^{-\alpha} - 1 \tag{1}$$

$$N_2 = \left(\frac{p_2}{p_{th2}} \right)^{-\alpha} - 1 \tag{2}$$

Combining Eqs. (1) and (2) and multiplying with λ will give us the total number of packets in a network:

$$Total_{packets} = \lambda \left(\frac{\beta p_2}{p_{th1}} \right)^{-\alpha} - 1 + \lambda \left(\frac{p_2}{p_{th2}} \right)^{-\alpha} - 1 \quad (3)$$

where $0 < \alpha \leq 1$ is the probabilistic response of consumers towards price, and it increases with an increasing value of α . Actually, α can be described as the perceived utility of consumers given a certain price. From Eq. (3) we can define the unused capacity for the network by subtracting number of packets from the total network capacity.

5 Subsidies, Tragedies and Revenue Maximization in Heterogeneous Societies

When a consumer uses shared resources in a network, that consumer should pay a corresponding price. The dynamics of correct subsidy implementation in a heterogeneous community could affect the individual user's (in this case the information "have nots") decision in such a way that tragedy of the commons and tragedy of the anti-commons can be avoided. Given the preferences of consumers for a given price p_2 and a cost c , the profit function of the service provider can be written and defined as follows:

$$\Pi = \lambda \left(\left(\frac{\beta p_2}{p_{th1}} \right)^{-\alpha} - 1 \right) (p_2 - c) + \lambda \left(\left(\frac{p_2}{p_{th2}} \right)^{-\alpha} - 1 \right) (p_2 - c) \quad (4)$$

Taking the partial derivatives of Eq. (4), we obtain an expression of Π as follows:

$$\frac{\partial \Pi}{\partial p_2} = \frac{\lambda}{\left(\frac{\beta p_2}{p_{th1}} \right)^{-\alpha} - 1} + \frac{\lambda}{\left(\frac{p_2}{p_{th1}} \right)^{-\alpha} - 1} + \frac{\lambda \alpha (c - p_2)}{p_{th2} \left(\frac{p_2}{p_{th2}} \right)^{\alpha+1}} + \frac{\lambda \alpha \beta (c - p_2)}{p_{th1} \left(\frac{\beta p_2}{p_{th1}} \right)^{\alpha+1}} \quad (5)$$

Further differentiation of Eq. (5) provides us with:

$$\begin{aligned} \frac{\partial^2 \Pi}{\partial p_2^2} = & \left(\frac{2\lambda\alpha}{p_{th2} \left(\frac{p_2}{p_{th2}} \right)^{\alpha} + 1} \right) - \left(\frac{2\lambda\alpha\beta}{p_{th1} \left(\frac{\beta p_2}{p_{th1}} \right)^{\alpha} + 1} \right) - \left(\frac{\lambda\alpha(c-p_2)(\alpha+1)}{p_{th2}^2 \left(\frac{p_2}{p_{th2}} \right)^{\alpha+2}} \right) \\ & - \left(\frac{\lambda\alpha\beta^2(c-p_2)(\alpha+1)}{p_{th1}^2 \left(\frac{\beta p_2}{p_{th1}} \right)^{\alpha+2}} \right) \end{aligned} \quad (6)$$

A close analysis of Eqs. (4), (5) and (6) shows that there is an optimal solution (price) to Eq. (4) that provides the service provider with the maximum revenue or profit. Thus, from Eqs. (4), (5) and (6) the following theorem can be proposed:

Theorem 1 *Suppose the profit function Π is twice differentiable and strictly concave in the region $0 < p_2 < p_{th2}$, then there exists an optimal price p_2^* that maximizes the service provider’s revenue due to unsubsidized users.*

Proof From Eqs. (4), (5) and (6) we see that $\frac{\partial \Pi}{\partial p_2} = 0$ has a maximum point since $\frac{\partial^2 \Pi}{\partial p_2^2} < 0$, meaning that within the interval of $0 < p_0 < p_{th2}$, the profit function is positive for all the values of p_2 in $c < p_{1,2} < p_{th1,2}$. Therefore, it is clear that the profit function, Π , is strictly concave in the region of $c < p_{1,2} < p_{th1,2}$. Moreover, it can be inferred from Eq. (4) that if $\frac{\partial^2 \Pi}{\partial p_2^2} < 0$, Π contains at least one maximization point or fixed point (see Kakutani fixed point theorem), at which $\frac{\partial \Pi}{\partial p_2} = 0$ within the range of $c < p_2 < p_{th2}$. Thus the optimal price can be found by equating the first derivative to zero i.e. $\frac{\partial \Pi}{\partial p_2} = 0$ and solving for p_2^* .

Depending on the price set by the ISP, and within the range of $0 \leq p_1 \leq p_{th1}$, the number of information “have-nots” in such a network can be determined by the value of the subsidy β set by the government. Actually, β also determines the maximum profit that the ISP can receive under no network capacity constraints i.e. the assumption of infinite available bandwidth. From theorem (1), we provide the following lemma to support our claim:

Lemma 1 *In any developing country the maximum number of information “have nots” (N_1) subscribing to an ISP is determined by the subsidy factor β and is a maximum when $\beta \approx 0$ i.e. $p_1 \approx 0$*

Proof Consider Eqs. 1 and 2 as follows:

$$N_{total} = \left(\frac{\beta p_2}{p_{th1}}\right)^{-\alpha} - 1 + \left(\frac{p_2}{p_{th2}}\right)^{-\alpha} - 1$$

By differentiating Eq. 3 with respect to β we obtain:

$$\frac{\partial N}{\partial \beta} = \frac{p_2 \alpha (c - p_2)}{p_{th1} \left(\frac{\beta p_2}{p_{th1}}\right)^{\alpha+1}} \tag{7}$$

From $\frac{\partial N}{\partial \beta} = 0$, we can solve for β^* by equating 7 to zero:

$$\frac{\partial N}{\partial \beta} = \left(\frac{p_2 \alpha (c - p_2)}{p_{th1} p_2^{\alpha+1}}\right)^{\frac{1}{\alpha+1}} \tag{8}$$

It is clear from the *Lemma 1* that if $\beta^* \approx 0$ then the tragedy of the commons is the most likely outcome. Increasing the value of β such that $\beta^* \approx 1$, will reduce the number of information “have nots” subscribing to the ISP. Obviously, a solution to Eq. (8) exists that maximizes the revenue of a service provider without creating the tragedy of the (anti) commons within the range $0 < \beta < 1$. Since the solution to the equation above is mathematically intractable, we use Matlab to approximate the value of β which is the solution to Eq. (8). Note that this solution is dependent on the value of the threshold values for the two groups and it varies with varying threshold and sensitivity (α) towards price.

6 Simulation Results

In this section we provide the simulation results on the consequences of subsidy on the number of users and revenue maximization. We use Matlab to approximate our solutions.

6.1 Effect of Subsidy on Revenue and Price

In Figs. 3, 4 and 5, we evaluate the effect of subsidy on revenue maximization, given any price p_2 and p_{th1} and p_{th2} at a constant sensitivity α Fig. 3, for example, shows that for a subsidy factor of $\beta = 1$ (in this case we assume that the government does not subsidize the information “have nots”), the revenue generated from the information “have nots” is, actually, dependant on the maximum price p_{th1} that they are willing to pay. This phenomenon, especially in underserved areas of developing countries, may sometimes result in the tragedy of anti-commons since very few information “have nots” will be willing to subscribe to the ISP. For all intents and purposes, it is worthwhile to note that the information “haves nots” are contributing very little towards the ISP’s overall revenue since no subsidy is being given by the government towards the ISP (see Fig. 3).

Consider again that the government subsidizes the information “have nots” as shown in Fig. 4. Figure 4 shows that too much subsidy has a drastic consequence on revenue maximization and the number of information “have nots” subscribing to an ISP. Since decreasing the subsidy factor β , i.e. $\beta = 0.1$, increases the number of information “have nots”, it is obvious that the maximum revenue obtained by the ISP will be highest when the government pays for all the “information have nots” ($\beta \approx 0$) as shown in Fig. 4.

In actual fact, Fig. 4 represents an outcome that will objectively fulfill government policies of universal access for all, as it allows for more information “have nots” to subscribe to the ISP at a subsidized price. However, such a situation would definitely promote the tragedy of the commons and may lead to market failure in once stable markets. Thus, Fig. 4 highlights the implications of

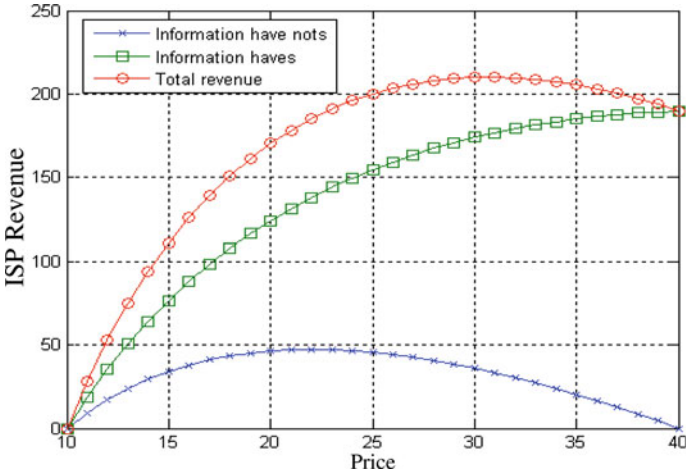


Fig. 3 Revenue versus price at $\alpha = 0.3$, $p_{th1} = 40$, $p_{th2} = 100$, $\beta = 1$, $c = 10$

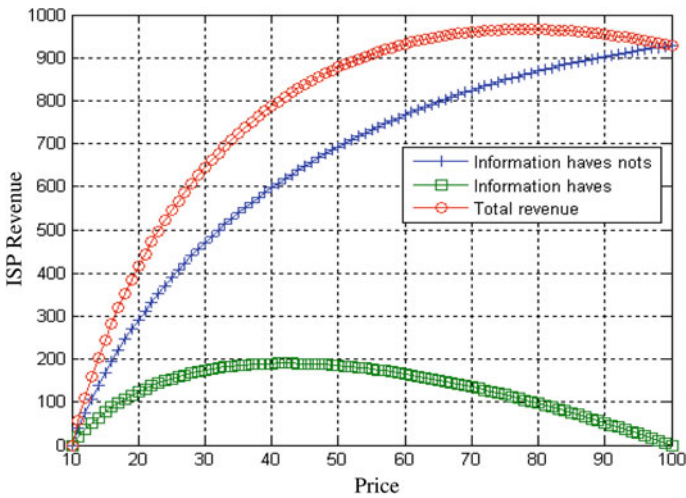


Fig. 4 Revenue versus price at $\alpha = 0.3$, $p_{th1} = 40$, $p_{th2} = 100$, $\beta = 0.1$, $c = 10$

heavy subsidy towards revenue maximization in a heterogeneous society, and shows that under heavy government subsidy there will be a skewed information access between the information “haves” and the information “have nots” due to unequal subsidy distribution.

By equating (consider Eq. 4) the revenue generated by the information “have nots” to the “information haves” we can solve for equilibrated value of β . Figure 5, actually, represents an equilibrium between the information “haves” and “have nots” as the revenue curve of the N_1 and N_2 follow the same trajectory. Both users (as depicted in Fig. 5) will not subscribe to the ISP at the price of 100 units

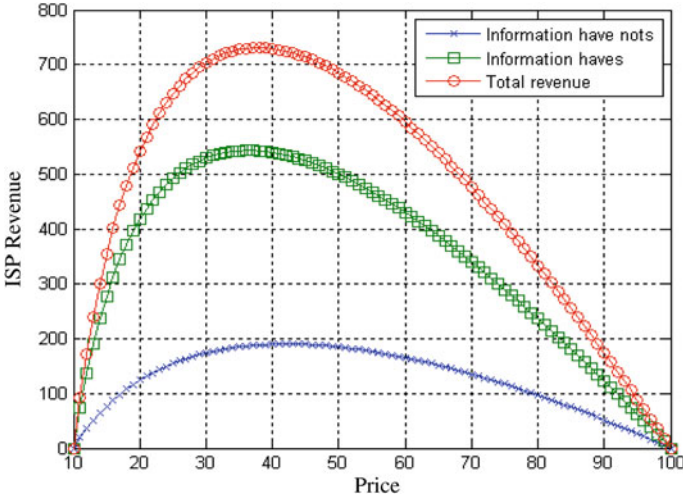


Fig. 5 Revenue vs price at $\alpha = 0.3$ for N_1 , $\alpha = 0.7$ for N_2 and $p_{th1} = 40$, $p_{th2} = 100$, $\beta = 0.4$, $c = 10$

preventing free riding. Contrary to Figs. 3 and 4, Fig. 5 represents a desirable outcome between the information “haves” and “have nots” as it represents a balanced price versus revenue curve trajectory. Generally, Fig. 5, in point of fact, may prevent the tragedy of the commons and tragedy of the anti-commons in heterogeneous society as no member of a group is disadvantage due to skewed pricing.

6.2 Revenue, Subsidy and Consumer Sensitivity Towards Price

This section presents results on how β and the sensitivity α affects revenue maximization given the two groups of people. We adopt the following values for our simulation: $p_{th1} = 40$, $\lambda = 2$, $c = 10$ and $p_{th2} = 100$.

In this subsidy driven pricing model, understanding the properties of subsidy factor β and sensitivity α is useful to ascertain revenue maximization. The major question to be asked in this pricing model is: how does customer sensitivity α and subsidy factor β affect the revenue obtained by the service provider given the threshold price of the consumer p_{th1} and p_{th2} . As observed from Figs. 6, 7 and 8, revenue maximization is affected by the value of the subsidy factor and sensitivity of consumers towards price. The figures show that by adjusting the value of α and β , the government and the ISP can attain a required number of users so as to prevent the tragedy of the commons and the tragedy of anti-commons. As a matter of analysis, Figs. 6, 7, 8 expresses the revenue as a function of β and α that determines and optimizes net revenue over the ISP’s services horizon.

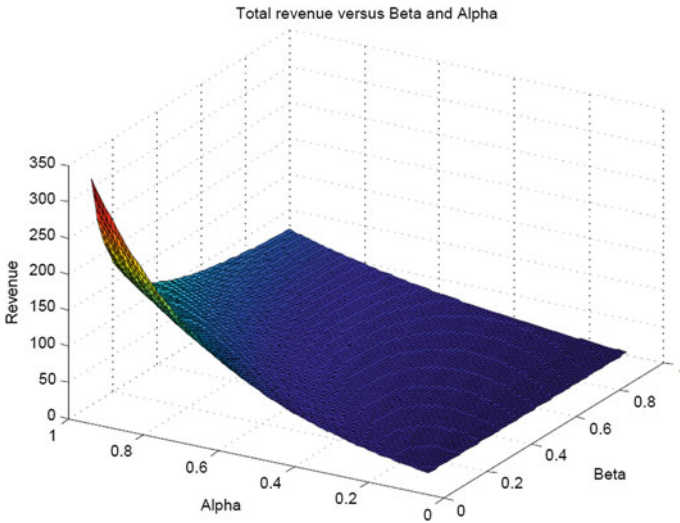


Fig. 6 Revenue versus beta and alpha when $p2 = 15$

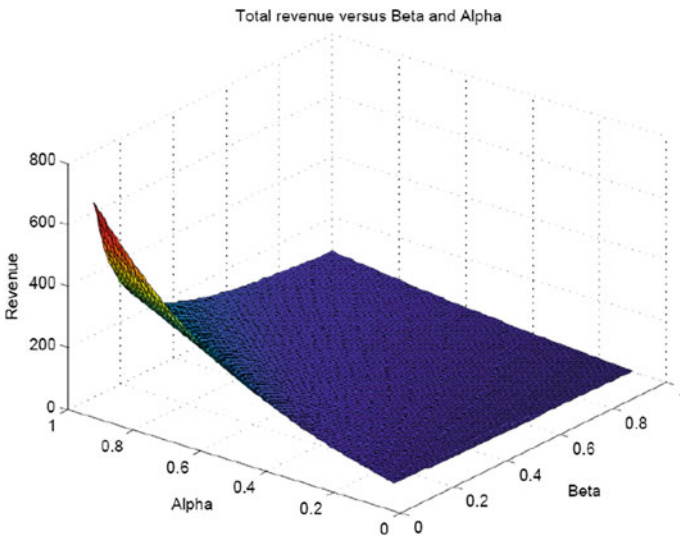


Fig. 7 Revenue versus beta and alpha when $p2 = 40$

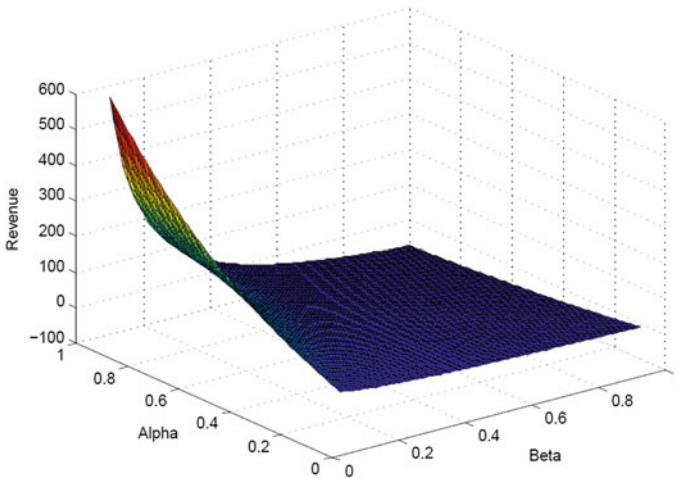


Fig. 8 Revenue versus beta and alpha when $p_2 = 90$

7 Conclusion

We have analyzed the effect of subsidy on the increase and decrease of users in a network which has a direct effect on the underuse and overuse of resources in heterogeneous communities. We have shown that if implemented properly, subsidies can prevent the tragedy of the (anti) commons and could be the solution to promoting social and economical growth in developing countries. In a matter of fact, developing countries must consider the tradeoff between increasing or decreasing subsidy and its effect on the underuse or overuse of resources. Our subsidy driven pricing model addresses the importance of a balanced subsidy pricing scheme with a view of achieving social and economical growth while enhancing the usage of network resources efficiently. Using this framework, we have shown that a government subsidy in a developing country can result in the increase or decrease of information “have nots” leading to the tragedy of (anti) commons. We have also shown that a correct subsidy, given customer sensitivities, can promote desired revenue and that the revenue is a concave function of price. We have derived the optimal revenue given the price, subsidy factor β and α as shown in Figs. 6, 7 and 8. In summary we have shown that an over-subsidized or under-subsidized group of users in a heterogeneous society can lead to either the tragedy of commons or anti-commons accordingly.

Acknowledgments This work was partially supported by the University of Johannesburg commonwealth bursary fund.

References

1. Sumbwanyambe M, Nel AL (2012) Subsidy and revenue maximization in developing countries. In: International multiconference of engineers and computer scientist, IMECS 2012. Lecture notes in engineering and computer science. Hong Kong, 14–16 March, pp 1568–1573
2. Jagannathan S, Almeroth KC (2002) Prices issues in delivering E-content on-demand. *ACM Sigecom Exch* 3(2):19–26
3. Levin SL (2010) Universal service and targeted support in a competitive telecommunication environment. *Telecommun Policy J* 34:92–97
4. Mariscal J (2005) Digital divide in a developing country. *Telecommun Policy J* 29(5–6):406–428
5. Garbacz C, Thompson HG (2005) Universal telecommunication service: a world perspective. *Inf Econ Policy* 17:495–512
6. Sumbwanyambe M, Nel AL, Clarke WA (2011) Challenges and proposed solutions towards telecentre sustainability: a Southern Africa case study. In: Proceedings of IST-Africa 2011 Gaborone Botswana 2011
7. TIBA (2012) South Africa’s universal service agency to offer big players subsidy to take voice and data to most rural areas”, Feb 2012. <http://www.balancingact-africa.com/news/en/issue-no-398/telecoms/south-africa-s-unive/en>
8. M. Sumbwanyambe, A. L Nel, W.A. Clarke “Optimal pricing of telecomm services in a developing country: A game theoretical approach” in proceedings of AFRICON 2011 pages 1–6 Livingstone, Zambia
9. Garbacz C, Thompson HG (2007) Demand for telecommunication services in developing countries. *Telecommun Policy* 31:276–289
10. Ramos B, Saeed K, Pavlov O (2010) The impact of universal service obligations and international cross-subsidies on the dispersion of telephone services in developing countries. *Socio-Econ Plan Sci* 44:57–72
11. Heller MA (1997) The tragedy of the anti-commons: property in the transition from Marx to markets. In: Proceedings of Harvard Law Review, vol 3
12. Harding G (1968) The tragedy of the commons. *J Sci* 168:1246–1248
13. Mandjes M (2003) Pricing strategies under heterogeneous service requirements. In: Proceedings of 22nd IEEE Computer and Communications (INFOCOM)

A Game Theoretical Analysis for the Quantity Discount Problem with Weibull Ameliorating Items

Hidefumi Kawakatsu, Toshimichi Homma and Kiyoshi Sawada

Abstract This paper presents a model for determining optimal all-unit quantity discount strategies in a channel of one seller (poultry farmer) and one buyer (retailer). The poultry farmer's stock increases due to the growth of items. In contrast, the retailer purchases items as fresh chicken meat from the poultry farmer, the inventory level of the retailer is therefore depleted due to the combined effects of its demand and deterioration. The poultry farmer attempts to increase her profit by controlling the retailer's ordering schedule through a quantity discount strategy. We formulate the above problem as a Stackelberg game between the poultry farmer and the retailer to analyze the existence of the poultry farmer's optimal quantity discount pricing policy, which maximizes her total profit per unit of time. Numerical examples are presented to illustrate the theoretical underpinnings of the proposed formulation.

Keywords Ameliorating items · Deteriorating items · Quantity discounts · Stackelberg game · Total profit · Weibull distribution

H. Kawakatsu (✉)

Department of Economics and Information Science, Onomichi City University,
1600-2 Hisayamada-cho, Onomichi 722-8506, Japan
e-mail: kawakatsu@onomichi-u.ac.jp

T. Homma

Faculty of Business Administration of Policy Studies, Osaka University of Economics,
2-2-8 Osumi, Higashiyodogawa-ku, Osaka, Japan
e-mail: t-homma@osaka-ue.ac.jp

K. Sawada

Department of Policy Studies, University of Marketing and Distribution Sciences,
3-1 Gakuen-Nishimachi, Nishi-ku, Kobe 651-2188, Japan
e-mail: sawada@umds.ac.jp

1 Introduction

Retailers who deal in fresh food-stuffs are particular about farms that produce high-quality products since their consumers today are concerned about food safety. For this reason, the number of retailers who directly deal with the poultry farmers has been increasing in Japan. The poultry farmers are expected to frequently ship the products in order to provide the fresh items for the retailers. In this study, we consider the case where the poultry farmer uses a quantity discount strategy to control the retailer's ordering schedule. Many researchers have considered the seller's quantity discount decision. By offering a discounted price to induce the buyer to order in larger quantities, the seller can increase her/his profit through reductions in her/his total transaction cost associated with ordering, shipment, and inventorying. Monahan [1] was one of the early authors who formulated the transaction between the seller and the buyer, and proposed a method for determining an optimal all-unit quantity discount policy with a fixed demand (see also Rosenblatt and Lee [2] and Data and Srikanth [3]). Lee and Rosenblatt [4] generalized Monahan's model to obtain the "exact" discount rate offered by the seller, and to relax the implicit assumption of a lot-for-lot policy adopted by the seller. Parlar and Wang [5] proposed a model using a game theoretical approach to analyze the quantity discount problem as a perfect information game. For more work, see also Sarmah et al. [6]. These models assumed that both the seller's and the buyer's inventory policies can be described by classical economic order quantity (EOQ) models. The classical EOQ model is a cost-minimization inventory model with a constant demand rate. It is one of the most successful models in all the inventory theories due to its simplicity and easiness.

In this study, we discuss the quantity discount problem between the poultry farmer and the retailer for ameliorating items under circumstances where the poultry farmer deals in the broiler. The ameliorating items include the fast growing animals such as the broiler in the poultry farm [7–9]. The poultry farmer purchases chicks from an upper-leveled supplier and then feeds and raises them until they grow up to be fowls. The retailer purchases (raw) chicken meat from the poultry farmer. The stock of the poultry farmer increases due to the growth of the items, in contrast, the inventory level of the retailer is depleted due to the combined effect of its demand and deterioration. The poultry farmer is interested in increasing her/his profit by controlling the retailer's order quantity through the quantity discount strategy. The retailer attempts to maximize her/his profit considering the poultry farmer's proposal. Our previous work [10] has formulated the above problem as a Stackelberg game between the poultry farmer and retailer to analyze the existence of the poultry farmer's optimal quantity discount pricing policy, which maximizes her/his total profit per unit of time. Our previous study showed how to determine the poultry farmer's optimal quantity discount policy, but did not clarify the existence of her/his optimal order quantity, which maximizes her/his total profit.

In this study, we reveal additional properties associated with the poultry farmer's optimal order quantity per lot. Numerical examples are presented to illustrate the theoretical underpinnings of the proposed formulation.

2 Notation and Assumptions

The poultry farmer uses a quantity discount strategy in order to improve her/his profit. The poultry farmer proposes, for the retailer, an order quantity per lot along with the corresponding discounted selling price, which induces the retailer to alter her/his replenishment policy.

We consider the two options throughout the present study as follows:

Option V_1 : The retailer does not adopt the quantity discount proposed by the poultry farmer. When the retailer chooses this option, she/he purchases the products from the poultry farmer at an initial price in the absence of the discount, and she/he determines her/himself an optimal order quantity, which maximizes her/his own total profit per unit of time.

Option V_2 : The retailer accepts the quantity discount proposed by the poultry farmer.

The main notations used in this paper are listed below:

Q_i	the retailer's order quantity per lot under Option V_i ($i = 1,2$).
S_i	the poultry farmer's order quantity per lot under Option V_i .
T_i	the length of the retailer's order cycle under Option V_i ($T_1 \leq T_2$).
h_s	the poultry farmer's inventory holding cost per item and unit of time (including the cost of amelioration).
h_b	the retailer's inventory holding cost per item and unit of time.
a_s, a_b	the poultry farmer's and the retailer's ordering costs per lot, respectively.
c_s	the poultry farmer's unit acquisition cost (unit purchasing cost from the upper-leveled manufacturer).
p_s	the poultry farmer's initial unit selling price, i.e., the retailer's unit acquisition cost in the absence of the discount.
p_b	the retailer's unit selling price, i.e., unit purchasing price for her/his customers.
y	the discount rate for the selling price proposed by the poultry farmer, i.e., the poultry farmer offers a unit discounted price of $(1-y)p_s$ ($0 \leq y < 1$).
θ_b	the deterioration rate of the retailer's inventory.
μ	the constant demand rate of the product.

The assumptions in this study are as follows:

- (1) The poultry farmer's inventory increases due to the growth of the items during the time period $[0, T_{\max}]$.

- (2) The retailer's inventory level is continuously depleted due to the combined effects of its demand and deterioration. The retailer's inventory level $I_b(t)$ at time t can be expressed by the following differential equation [10]:

$$dI_b(t)/dt = -\theta_b I_b(t) - \mu. \quad (1)$$

- (3) The rate of replenishment is infinite and the delivery is instantaneous.
 (4) Backlogging and shortage are not allowed.
 (5) The quantity of the item can be treated as continuous for simplicity.
 (6) Both the poultry farmer and the retailer are rational and use only pure strategies.
 (7) The length of the poultry farmer's order cycle is given by $N_i T_i$ ($N_i = 1, 2, \dots$) under Option V_i ($i = 1, 2$), where N_i is a positive integer. This is because the poultry farmer can possibly improve her/his total profit by increasing the length of her/his order cycle from T_i to $N_i T_i$.
 (8) The instantaneous rate of amelioration of the on-hand inventory at time t is denoted by $r(t)$ which obeys the Weibull distribution [7–9] i.e.,

$$r(t) = \frac{g(t)}{1 - F(t)} = \alpha \beta \cdot t^{\beta-1} (\alpha > 0, \beta > 0), \quad (2)$$

where $F(t)$ is the distribution function of Weibull distribution.

- (9) The poultry farmer's stock increases due to the growth of the items. Her/his inventory level, $I_s(t)$, at time t can be expressed by the following differential equation:

$$dI_s(t)/dt = r(t)I_s(t). \quad (3)$$

3 Retailer's Total Profit and Optimal Response

In our previous work [10], we have formulated the retailer's total profits per unit of time under Options V_1 and V_2 available to the retailer, and then revealed the retailer's optimal response. For this reason, in the following we briefly summarize the results associated with the retailer's profits under these two options and her/his optimal response.

3.1 Retailer's Total Profit

If the retailer chooses Option V_1 , her/his order quantity per lot and her/his unit acquisition cost are, respectively, given by $Q_1 = Q(T_1)$ and p_s , where p_s is the unit initial price in the absence of the discount. In this case, she/he determines her/himself the optimal order quantity $Q_1 = Q_1^*$, which maximize her/his total profit per unit of time.

Under assumption (1), Kawakatsu et al. [10] has formulated the retailer's total profit per unit of time, which is given by

$$\begin{aligned} \pi_1(T_1) &= \frac{p_b \mu T_1 - p_s Q(T_1) - h_b \int_0^{T_1} I_b(t) dt - a_b}{T_1} \\ &= \rho(p_b \theta_b + h_b) - \frac{(p_s + h_b/\theta_b)Q(T_1) + a_b}{T_1}, \end{aligned} \tag{4}$$

where

$$\rho = \mu/\theta_b, \tag{5}$$

$$Q(T_1) = \rho(e^{\theta_b T_1} - 1). \tag{6}$$

We showed that there exists a unique finite $T_1 = T_1^*$ (>0), which maximizes $\pi_1(T_1)$ in Eq. (4) [10].

The optimal order quantity is therefore given by

$$Q(T_1) = \rho(e^{\theta_b T_1^*} - 1). \tag{7}$$

The total profit per unit of time becomes

$$\pi_1^* = \rho \theta_b [(p_b + h_b/\theta_b) - (p_s + h_b/\theta_b)e^{\theta_b T_1^*}]. \tag{8}$$

In contrast, if the retailer chooses Option V_2 , the order quantity and unit discounted selling price are, respectively, given by $Q_2 = Q(T_2) = \rho(e^{\theta_b T_2} - 1)$ and $(1-y)p_s$.

The retailer's total profit per unit of time can therefore be expressed by

$$\begin{aligned} \pi_2(T_2, y) &= \rho(p_b \theta_b + h_b) \\ &\quad - \frac{[(1-y)p_s + h_b/\theta_b]Q(T_2) + a_b}{T_2}. \end{aligned} \tag{9}$$

3.2 Retailer's Optimal Response

The retailer prefers Option V_1 over Option V_2 if $\pi_1^* > \pi_2(T_2, y)$, but when $\pi_1^* < \pi_2(T_2, y)$, she/he prefers V_2 to V_1 . The retailer is indifferent between the two options if $\pi_1^* = \pi_2(T_2, y)$, which is equivalent to

$$y = \frac{1}{p_s} \cdot \frac{(p_s + h_b/\theta_b)[Q_2(T_2) - \rho \theta_b T_2 e^{\theta_b T_1^*}] + a_b}{Q_2(T_2)}. \tag{10}$$

Let us denote, by $\psi(T_2)$, the right-hand side of Eq. (10). It can easily be shown from Eq. (10) that $\psi(T_2)$ is increasing in $T_2 (\geq T_1^*)$.

4 Poultry Farmer's Total Profit

This section formulates the poultry farmer's total profit per unit of time, which depends on the retailer's decision. Figure 1 shows the poultry farmer's transition of inventory level in the case of $N_i = 4$ under Option V_i ($i = 1, 2$). The length of the poultry farmer's order cycle is divided into N_i shipment cycles as described in assumption (7), where N_i is also a decision variable for the poultry farmer. The poultry farmer ships the goods to the retailer at time jT_i ($j = 0, \dots, N_i - 1$). The poultry farmer's inventory level increases due to the growth of the items during the interval $[jT_i, (j + 1)T_i)$, but at time jT_i her/his inventory level drops from $z_j(jT_i)$ to $(z_j(jT_i) - Q_i)$ because of the shipment to the retailer, where $z_j(jT_i)$ denotes the poultry farmer's inventory level at the end of j th shipment cycle. Obviously, we have $S_i = z_0(jT_i)$.

4.1 Total Profit Under Option V_1

If the retailer chooses Option V_1 , her/his order quantity per lot and unit acquisition cost are given by $Q_1 = Q(T_1)$ and p_s , respectively. The poultry farmer's order quantity per lot can be expressed as $S_1 = S(N_1, T_1)$. Under assumption (9), for a given N_1 , the poultry farmer's total profit per unit of time under Option V_1 can be given by the following Eq. [10]:

$$\begin{aligned} P_1(N_1, T_1) &= \frac{p_s N_1 Q(T_1) - c_s S(N_1, T_1) - h_s \sum_{j=1}^{N_1-1} \int_{(j-1)T_1}^{jT_1} I_s^{(j)}(t) dt - a_s}{N_1 T_1} \\ &= \frac{p_s Q(T_1)}{T_1} \\ &\quad - \frac{Q(T_1) \left[c_s + \sum_{j=1}^{N_1-1} e^{-\alpha(jT_1)^\beta} \left(c_s + h_s \int_0^{jT_1} e^{\alpha t^\beta} dt \right) \right] + a_s}{N_1 T_1}, \end{aligned} \quad (11)$$

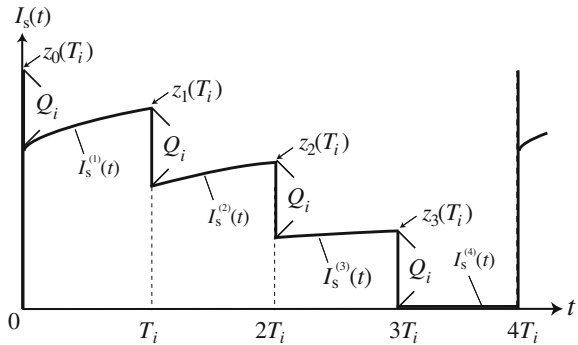
where

$$I_s^{(j)}(t) = z_j(T_1) e^{-\alpha[(jT_1)^\beta - t^\beta]}, \quad (12)$$

$$z_j(T_1) = Q(T_1) e^{\alpha(jT_1)^\beta} \sum_{k=j}^{N_1-1} e^{-\alpha(kT_1)^\beta}, \quad (13)$$

$$S(N_1, T_1) = Q(T_1) \sum_{j=0}^{N_1-1} e^{-\alpha(jT_1)^\beta}. \quad (14)$$

Fig. 1 Transition of inventory level ($N_i = 4$)



4.2 Total Profit Under Option V_2

When the retailer chooses Option V_2 , she/he purchases $Q_2 = Q(T_2)$ units of the product at the unit discounted selling price $(1 - y)p_s$. In this case, the poultry farmer's order quantity per lot under Option V_2 is expressed as $S_2 = S(N_2, T_2)$, accordingly the poultry farmer's total profit per unit of time under Option V_2 is given by

$$\begin{aligned}
 P_2(N_2, T_2, y) &= \frac{1}{N_2 T_2} \cdot \left[(1 - y)p_s N_2 Q(T_2) \right. \\
 &\quad \left. - c_s S(N_2, T_2) - h_s \sum_{j=1}^{N_2-1} \int_{(j-1)T_2}^{jT_2} I_s^{(j)}(t) dt - a_s \right] \\
 &= \frac{(1 - y)p_s Q(T_2)}{T_2} \\
 &\quad - \frac{Q(T_2) \left[c_s + \sum_{j=1}^{N_2-1} e^{-\alpha(jT_2)^\beta} (c_s + h_s \int_0^{jT_2} e^{\alpha t^\beta} dt) \right] + a_s}{N_2 T_2},
 \end{aligned} \tag{15}$$

where

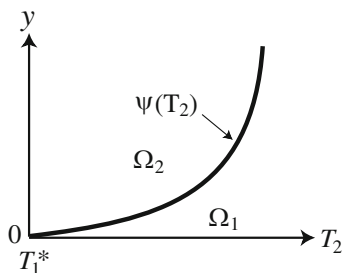
$$Q(T_2) = \rho(e^{\theta_b T_2} - 1), \tag{16}$$

$$S(N_2, T_2) = Q(T_2) \sum_{j=0}^{N_2-1} e^{-\alpha(jT_2)^\beta}. \tag{17}$$

5 Poultry Farmer's Optimal Policy

The poultry farmer's optimal values for T_2 and y can be obtained by maximizing her/his total profit per unit of time considering the retailer's optimal response, which was discussed in Sect. 3.2.

Fig. 2 Characterization of retailer's optimal responses



Henceforth, let $\Omega_i (i = 1, 2)$ be defined by

$$\Omega_1 = \{(T_2, y) | y \leq \psi(T_2)\},$$

$$\Omega_2 = \{(T_2, y) | y \geq \psi(T_2)\}.$$

Figure 2 depicts the region of $\Omega_i (i = 1, 2)$ on the (T_2, y) plane.

5.1 Under Option V_1

If $(T_2, y) \in \Omega_1 \setminus \Omega_2$ in Fig. 2, the retailer will naturally select Option V_1 . In this case, the poultry farmer can maximize her/his total profit per unit of time independently of T_2 and y on the condition of $(T_2, y) \in \Omega_1 \setminus \Omega_2$. Hence, for a given N_1 , the poultry farmer's locally maximum total profit per unit of time in $\Omega_1 \setminus \Omega_2$ becomes

$$P_1^* = \max_{N_1 \in N} P_1(N_1, T_1^*). \tag{18}$$

where N signifies the set of positive integers.

5.2 Under Option V_2

On the other hand, if $(T_2, y) \in \Omega_2 \setminus \Omega_1$, the retailer's optimal response is to choose Option V_2 . Then the poultry farmer's locally maximum total profit per unit of time in $\Omega_2 \setminus \Omega_1$ is given by

$$P_2^* = \max_{N_2 \in N} \hat{P}_2(N_2), \tag{19}$$

where

$$\hat{P}_2(N_2) = \max_{(T_2, y) \in \Omega_2 \setminus \Omega_1} P_2(N_2, T_2, y). \tag{20}$$

More precisely, we should use “sup” instead of “max” in Eq. (20).

For a given N_2 , we show below the existence of the poultry farmer’s optimal quantity discount pricing policy $(T_2, y) = (T_2^*, y^*)$ which attains Eq. (20). It is easily proven that $P_2(N_2, T_2, y)$ in Eq. (15) is strictly decreasing in y , and consequently the poultry farmer can attain $\hat{P}_2(N_2)$ in Eq. (20) by letting $y \rightarrow \psi(T_2) + 0$. By letting $y = \psi(T_2)$ in Eq. (15), the total profit per unit of time on $y = \psi(T_2)$ becomes

$$P_2(N_2, T_2) = (p_s + h_b/\theta_b)\rho\theta_b e^{\theta_b T_1^*} - \frac{(h_b/\theta_b)Q(T_2) + a_b}{T_2} - \frac{Q(T_2) \left[c_s + \sum_{j=1}^{N_2-1} j^\beta e^{-(jT_2)^\beta} \left(c_s + h_s \int_0^{jT_2} e^{\alpha t^\beta} dt \right) \right] + a_s}{N_2 T_2}, \tag{21}$$

Let us here define

$$L(T_2) = [\rho\theta_b T_2 e^{\theta_b T_2} - Q_2(T_2)] \times \left[\left(\frac{N_2 h_b}{\theta_b} + c_s \right) + \sum_{j=1}^{N_2-1} e^{-\alpha(jT_2)^\beta} \left(c_s + h_s \int_0^{jT_2} e^{\alpha t^\beta} dt \right) \right] + Q(T_2) T_2 \left[h_s \frac{N_2(N_2 - 1)}{2} - r(T_2) \sum_{j=1}^{N_2-1} j^\beta e^{-(jT_2)^\beta} \left(c_s + h_s \int_0^{jT_2} e^{\alpha t^\beta} dt \right) \right], \tag{22}$$

then we can summarize the results of analysis in relation to the optimal quantity discount policy which attains $\hat{P}_2(N_2)$ in Eq. (20) when N_2 is fixed to a suitable value [10].

1. $N_2 = 1$:

In this case, there exists a unique finite $T_0 (> T_1^*)$ which maximizes $P_2(N_2, T_2)$ in Eq. (21), and therefore (T_2^*, y^*) is given by

$$(T_2^*, y^*) \rightarrow (\tilde{T}_2, \psi(\tilde{T}_2)), \tag{23}$$

where

$$\tilde{T}_2 = \begin{cases} T_o, & T_o \leq T_{\max}/N_2, \\ T_{\max}/N_2, & T_o > T_{\max}/N_2. \end{cases} \tag{24}$$

The poultry farmer’s total profit then becomes

$$\hat{P}_2(N_2) = \rho\theta_b [(p_s + h_b/\theta_b)e^{\theta_b T_1^*} - (c_s + h_b/\theta_b)e^{\theta_b T_2^*}]. \tag{25}$$

2. $N_2 \geq 1$:

Let us define $T_2 = \tilde{T}_2 (> T_1^*)$ as the unique solution (if it exists) to

$$L(T_2) = (a_b N_2 + a_s). \tag{26}$$

In this case, the optimal quantity discount pricing policy is given by Eq. (23).

5.3 Under Options V_1 and V_2

In the case of $(T_2, y) \in \Omega_1 \cap \Omega_2$, the retailer is indifferent between Option V_1 and V_2 . For this reason, this study confines itself to a situation where the poultry farmer does not use a discount pricing policy $(T_2, y) \in \Omega_1 \cap \Omega_2$.

5.4 Optimal Value for N_2

For a given T_i , we here derive a lower bound and an upper bound for the optimal value of $N_i = N_i^* (N_i = 1, 2, 3, \dots, \lfloor T_{\max}/T_i \rfloor)$, which maximizes $P_1(N_1, T_1)$ in Eq. (11) and $P_2(N_2, T_2, y)$ in Eq. (15). It should be reminded that the poultry farmer's order quantity per lot is the function of N_i and T_i .

Let us here define

$$N_{\max} = \lfloor T_{\max}/T_i \rfloor, \tag{27}$$

$$A_j = e^{-\alpha(jT_i)^\beta} \left(c_s + h_s \int_0^{jT_i} e^{\alpha t^\beta} dt \right) (j = 1, 2, 3, \dots, N_{\max}), \tag{28}$$

$$L_i^{(L)}(N_i) = N_i A_{N_i} - \sum_{j=1}^{N_i-1} A_j, \tag{29}$$

$$L_i^{(U)}(N_i) = N_i A_{N_i-1} - \sum_{j=1}^{N_i-1} A_j. \tag{30}$$

We focus on the case where A_{j+1} is greater than A_j , or A_{j+1} is less than or equal to A_j ($j = 1, 2, 3, \dots, N_{\max} - 1$). In this case, by using a difference method, we can here summarize the results of analysis in relation to a lower bound $N_i = N_i^{(L)}$ and an upper bound $N_i = N_i^{(U)}$ for N_i^* .

1. Lower bound $N_i^{(L)}$ for N_i^* :

- $L_i^{(L)}(1) \geq c_s + a_s/Q(T_i)$ and $A_j < A_{j+1}$:

In this subcase, we have

$$N_i^{(L)} = 1. \tag{31}$$

- $L_i^{(L)}(1) < c_s + a_s/Q(T_i) < L_i^{(L)}(N_{\max})$ and $A_j < A_{j+1}$:

There exists a unique finite $N_i^{(L)}$ which is the solution to

$$L_i^{(L)}(N_i) = c_s + a_s/Q(T_i). \tag{32}$$

- $\left\{ L_i^{(L)}(N_{\max}) \leq c_s + a_s/Q(T_i) \text{ and } A_j < A_{j+1} \right\}$ or $\left\{ A_j \geq A_{j+1} \right\}$:

In this subcase, we have

$$N_i^{(L)} = N_{\max}. \tag{33}$$

2. Upper bound $N_i^{(U)}$ for N_i^* :

- $L_i^{(U)}(N_{\max}) > c_s + a_s/Q(T_i)$ and $A_j < A_{j+1}$:

There exists a unique finite $N_i^{(U)}$ which is the solution to

$$L_i^{(U)}(N_i) = c_s + a_s/Q(T_i). \tag{34}$$

- $\left\{ L_i^{(U)}(N_{\max}) \leq c_s + a_s/Q(T_i) \text{ and } A_j < A_{j+1} \right\}$ or $\left\{ A_j \geq A_{j+1} \right\}$:

In this subcase, we have

$$N_i^{(U)} = N_{\max}. \tag{35}$$

The above results indicate that the optimal N_i^* satisfies the following inequality:

$$(1 \leq) N_i^{(L)} \leq N_i^* < N_i^{(U)} (\leq N_{\max}). \tag{36}$$

6 Numerical Examples

Table 1 reveals the results of sensitively analysis in reference to N_1^* , Q_1^* , $p_1 (= p_s)$, S_1^* , P_1^* , Q_2^* , N_2^* , $p_2^* (= (1-y^*) p_s)$, S_2^* , P_2^* for $(p_s, p_b, a_s, a_b, h_s, h_b, \alpha, \beta, \theta_b, \mu, T_{\max}) = (100, 200, 1000, 1200, 20, 1, 0.8, 0.8, 0.015, 5, 30)$ when $c_s = 35, 40, 45,$ and 50.

Table 1 Sensitivity analysis

a_s	N_1^*	Q_1^*	p_1	$S_1^* (N_1^*)$	P_1^*
(a) Under Option V_1					
35	1	71.64	100	71.64	281.75
40	2	71.64	100	71.79	265.82
45	2	71.64	100	71.79	251.99
50	2	71.64	100	71.79	238.17
a_s	N_2^*	Q_2^*	p_2^*	$S_2^* (N_2^*)$	P_2^*
(b) Under Option V_2					
35	1	127.12	95.43	127.12	310.24
40	1	123.95	95.81	123.95	280.79
45	2	84.11	99.63	84.19	254.73
50	2	84.11	99.63	84.19	240.70

In Table 1a, we can observe that Q_1^* takes a constant value ($Q_1^* = 71.64$). Under Option V_1 , the retailer does not adopt the quantity discount offered by the poultry farmer. The poultry farmer cannot therefore control the retailer’s ordering schedule, which signifies that Q_1^* is independent of c_s . Table 1a also shows that both N_1^* and S_1^* increase when c_s increases from 35 to 40 (more precisely, at the moment when c_s increases from 35.761 to 35.762). Increasing the poultry farmer’s lot size results in increasing the time period of feeding and raising the items. This signifies that, when her/his lot size becomes larger, the poultry farmer’s stock more increases due to the growth of the items. Under Option V_1 , when c_s takes larger values, the poultry farmer should increase her/his stock by means of increasing her/his own lot size since the length of the retailer’s order cycle cannot be changed.

Table 1b indicates that, under Option V_2 , Q_2^* is greater than Q_1^* (compare with Table 1a). Under Option V_2 , the retailer accepts the quantity discount proposed by the poultry farmer. The poultry farmer’s lot size can therefore be increased by stimulating the retailer to alter her/his order quantity per lot through the quantity discount strategy. We can also notice in Table 1 that we have $P_1^* < P_2^*$. This indicates that using the quantity discount strategy can increase the poultry farmer’s total profit per unit of time.

7 Conclusion and Future Work

In this study, we have discussed a quantity discount problem between a poultry farmer and a retailer for ameliorating items.

These items include the fast growing animals such as broiler in poultry farm. The poultry farmer purchases chicks from an upper-leveled supplier and then feeds and raises them until they grow up to be fowls. The retailer purchases (raw) chicken meat from the poultry farmer. The stock of the poultry farmer increases

due to the growth of the items, in contrast, the inventory level of the retailer is depleted due to the combined effect of its demand and deterioration. The poultry farmer is interested in increasing her/his profit by controlling the retailer's order quantity through the quantity discount strategy. The retailer attempts to maximize her/his profit considering the poultry farmer's proposal. We have formulated the above problem as a Stackelberg game between the poultry farmer and the retailer to show the existence of the poultry farmer's optimal quantity discount policy that maximizes her/his total profit per unit of time in the same manner as our previous work [10]. Our previous study mainly showed the existence of the optimal quantity discount pricing policy. In this study, we have derived the additional properties associated with the poultry farmer's optimal order quantity per lot. It should be pointed out that our results are obtained under the situation where the inventory holding cost is independent of the value of the item. The relaxation of such a restriction is an interesting extension.

References

1. Monahan JP (1984) A quantity discount pricing model to increase vendor's profit. *Manag Sci* 30(6):720–726
2. Rosenblatt MJ, Lee HL (1985) Improving pricing profitability with quantity discounts under fixed demand. *IIE Trans* 17(4):338–395
3. Data M, Srikanth KN (1987) A generalized quantity discount pricing model to increase vendor's profit. *Manag Sci* 33(10):1247–1252
4. Lee HL, Rosenblatt MJ (1986) A generalized quantity discount pricing model to increase vendor's profit. *Manag Sci* 32(9):1177–1185
5. Parlar M, Wang Q (1995) A game theoretical analysis of the quantity discount problem with perfect and incomplete information about the buyer's cost structure. *RAIRO/Oper Res* 29(4):415–439
6. Sarmah SP, Acharya D, Goyal SK (2006) Buyer vendor coordination models in supply chain management. *Eur J Oper Res* 175(1):1–15
7. Hwang HS (1997) A study on an inventory model for items with weibull ameliorating. *Comput Ind Eng* 33:701–704
8. Mondal B, Bhunia AK (2003) An inventory system of ameliorating items for price dependent demand rate. *Comput Ind Eng* 45:443–456
9. Chou SY, Chouhuang WT (2008) An analytic solution approach for the economic order quantity model with weibull ameliorating items. *Math Comput Model* 48:1868–1874
10. Kawakatsu H, Homma T, Sawada K (2012) An optimal quantity discounting pricing policy for ameliorating items, Lecture notes in engineering and computer science. In: *Proceedings of the international multiconference of engineers and computer scientists 2(IMECS2012)* Hong Kong, 14–16 March, pp 1549–1554

Simulation Study of an Integrated Reverse Logistics in Fuzzy Environment

Debabrata Das and Pankaj Dutta

Abstract Along with forward supply chain organization needs to consider the impact of reverse logistics due to its economic advantage, social awareness and strict legislations. In this work, we develop a system dynamics framework to analyze the long-term behavior of a multi-echelon integrated forward-reverse supply chain with fuzzy demand, satisfaction rate and collection rate. The uncertainty associated with satisfaction of customers and collection of used product has been quantified using fuzzy possibility measures. In the proposed model, it is assumed that the customer can exchange their old used product with a fresh new product in a primary market or a relatively better refurbished product in a secondary market at a discounted price. From the simulation study, it is observed that the inclusion of product exchange policy reduce the order variation and bullwhip effect at both retailer and distributor level. Finally, sensitivity analysis is performed to examine the impact of various parameters, namely; satisfaction rate, collection percentage, refurbishing percentage, inventory cover time and inventory adjustment time on recovery process and bullwhip effect.

Keywords Bullwhip effect · Fuzzy parameters · Possibility measures · Reverse supply chain · Simulation · System dynamics

D. Das · P. Dutta (✉)
SJM School of Management, Indian Institute of Technology Bombay,
Mumbai 400076, India
e-mail: pdutta@iitb.ac.in

1 Introduction

A large number of successful companies, especially in emerging economies, focus on forward supply chain but experience a lack of control over their reverse logistics (RL) process. According to Blackburn et al. [1], most reverse supply chains are organized to carry out five processes—product acquisition, reverse logistics, inspection and disposition, remanufacturing or refurbishing, and creation of alternative markets for the recovered products. Refurbished products are often offered as a substitute to the original products in a secondary market to the customers those are attracted by the brand, but do not wish to pay the price of a new product (e.g. electronics and automobile) [2]. In the present study, we address the benefits of employing product exchange (PE) policy both in primary as well as in secondary markets to increase the collection rate of used products and consequently selling of refurbished products.

Fleischmann et al. [3] provided a review of the quantitative models for reverse logistics in which they reported that most of the papers in the area of integrated reverse logistics are confined to single issues while comprehensive approaches are rare as variety of factors are involved in a general framework and the complexity of their interdependencies. System dynamics (SD) is a powerful methodology for obtaining the insights of these kinds of problems having dynamic complexity; but there are very few literatures which modeled the integrated aspects of forward and reverse supply chain using SD. Spengler and Schroter [4] modeled an integrated production and recovery system for supplying spare parts using SD to evaluate various strategies. Georgiadis and Vlachos [5] developed a SD model to evaluate the effect of environmental issues on long-term decision making in collection and remanufacturing activities.

As the quality and quantity of used products return to the collection points are uncertain in the reverse channel, the systematic distortion is inevitable and bullwhip effect may occur at retailer, distributor and manufacturer level. There are very few literatures that consider order variations and bullwhip effect in an integrated reverse supply chain ([6, 7]), but it is yet to receive attention in the context of SD framework ([8, 9]).

In this chapter, we propose a SD framework for an integrated reverse logistics network by incorporating fuzzy demand, fuzzy collection rate and fuzzy satisfaction rate with three way recovery (TWR) options, namely; product refurbishing, component reuse & refurbishing, and raw material recovery. The uncertainty associated with acquisition and collections of used product have been quantified using fuzzy possibility measures. We simulate the order variation at retailer and distributor levels and compare the bullwhip effect of the integrated forward-reverse logistics with and without PE policy. Also, sensitivity analysis is performed to examine the impact of various parameters on recovery process and bullwhip effect.

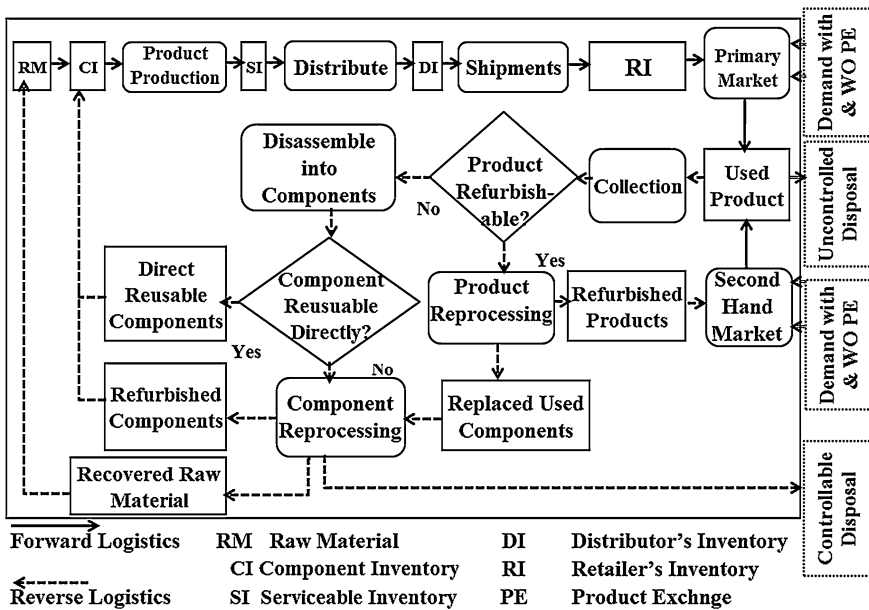


Fig. 1 Flow Chart of the integrated forward-reverse supply chain

2 Problem Description

In this work, we focus on an integrated forward-reverse supply chain (see Fig. 1). Due to the high complexity; we divide the whole system into two parts:

2.1 Forward Supply Chain with Product Exchange

The forward supply chain comprises three echelons: producer, distributor and retailer. There are two kinds of demand which is generally being seen in developing countries, namely;—demand in “primary market” for fresh new products and demand in “secondary market” for refurbished products. For example, in India, due to the increasing standards of living, the concept of product exchange and secondary market are getting popularity, especially for automobile and electronics products. In this model we categorize the demand as “demand with exchange” and “demand without exchange” both in primary as well as in secondary market. The customer can exchange their old used product with a fresh new product in a primary market or a relatively better refurbished product in a secondary market at a discounted price which will effectively increase the product collection rate, market share and satisfy the customer’s need.

2.2 Reverse Supply Chain with Three Way Recovery

In the simulation study, it is assumed that there is no constraint on the capacity of collection, inspection, sorting and restoring. After the initial inspection, if the collected products are accepted for refurbishing, then with some reprocessing, the *refurbished products* can be sold in the secondary market. If the products are not in a condition to refurbish, then it is disassembled into various components. During the process of product refurbishing, if new replacement is required for some components, then the old components are sent to reprocessing center for further recovery. In this model, we assume that the derived components can have three categories: one is *direct reusable components* that can be directly used without any further processing to increase the inventory of component in the forward channel; the second one is the *refurbished component* which require some reprocessing before adding it to the component inventory in the forward channel; the rest of the components can be used either to *recover raw material* which effectively increase the raw materials inventory in forward channel or can be sent directly for controllable disposal as shown in Fig. 1.

3 Fuzzy Set Theory

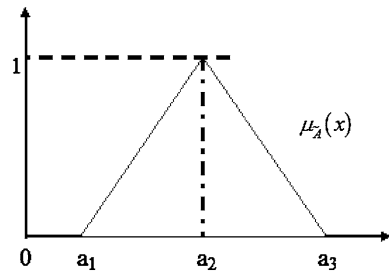
In real life supply chain system, due to the lack of historical data and/or insufficient information, prediction of actual demand from end users (both in primary and secondary markets) and collections of returned/used products are highly uncertain. It has been argued in a large body of recent literature that fuzzy sets theory could provide an appropriate framework for dealing with uncertainties in areas where intuition and subjective judgment play an important role [10].

3.1 Triangular Fuzzy Number

Triangular fuzzy number (TFN) $\tilde{a} = (a_1, a_2, a_3)$ (see Fig. 2) is the fuzzy number with the membership function $\mu_{\tilde{A}}(x)$, a continuous mapping: $\mu_{\tilde{A}}(x) : R \rightarrow [0, 1]$ such that

$$\mu_{\tilde{A}}(x) = \begin{cases} 0 & \text{for } -\infty < x < a_1 \\ \frac{x-a_1}{a_2-a_1} & \text{for } a_1 \leq x \leq a_2 \\ \frac{a_3-x}{a_3-a_2} & \text{for } a_2 \leq x \leq a_3 \\ 0 & \text{for } a_3 < x < \infty \end{cases}$$

Fig. 2 Membership function of TFN \tilde{a}



3.2 Possibility Measures

There are several representations of fuzzy constraints. Here we use possibility measure concept in which fuzzy numbers are interpreted by the degree of uncertainty. Possibility (optimistic sense) means the maximum chance (at least) to be selected by the decision maker (DM). Analogous to chance constrained programming with stochastic parameters, in a fuzzy environment; it is assumed that some constraints may be satisfied with a least possibility, η ($0 \leq \eta \leq 1$) (i.e. here the ‘chance’ is represented by the ‘possibility’) [11]. Therefore, if a DM desires to impose the resource constraint in possibility sense, he/she is optimistic about his/her decisions.

According to Dubois and Prade [11], if \tilde{A} and \tilde{B} be two fuzzy numbers with membership function $\mu_{\tilde{A}}(x)$ and $\mu_{\tilde{B}}(x)$ respectively, then

$Pos(\tilde{A} * \tilde{B}) = \{sup(min(\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(y)), x, y \in \mathfrak{R}, x * y)\}$, where the abbreviation *Pos* represents possibility and $*$ is any of the relations $<, >, =, \leq, \geq$.

Let $\tilde{a} = (a_1, a_2, a_3)$ be a triangular fuzzy number and b is a crisp number then as in Maity and Maiti [12] the following lemma holds:

$$Pos(\tilde{a} \geq b) \geq \eta \text{ if and only if } \frac{a_3 - b}{a_3 - a_2} \geq \eta \tag{1}$$

4 System Dynamics Model

System dynamics is a modeling and simulation methodology for framing, understanding, and discussing complex issues and problems. The structure of a SD model contains stock (state), flow (rate) and auxiliary/constant variables. Stock variables are the accumulations (e.g. inventories) within the system. The flow variables represent the flows in the system (e.g. refurbishing rate) from one stock to another. The mathematical formulation consists of a system of differential equations, which is numerically solved via simulation. Nowadays, high-level graphical simulation programs (viz. Vensim, i-think, Powersim etc.) support the analysis and study of these systems. Here, we choose Vensim (version: windows 5.10 e) as a tool to simulate the model.

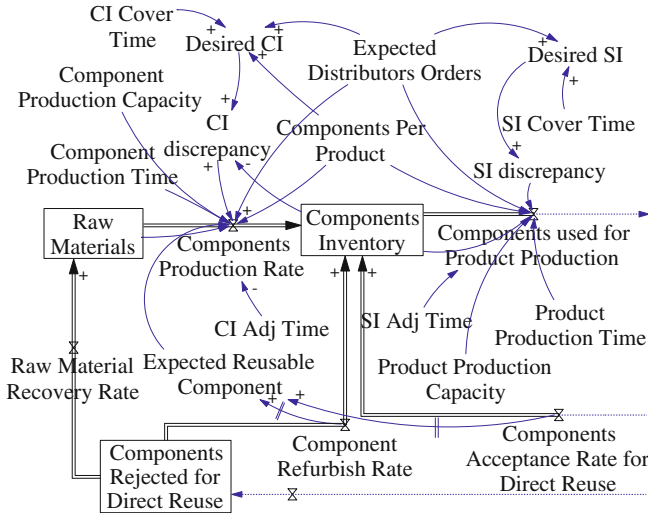


Fig. 3 Stock-flow diagram for component production and recovery

4.1 Stock-Flow Diagram

Because of the high complexity of the integrated model, whole system has been divided in four sub sections. Figures 3, 4 and 5 depicts the stock-flow diagram of the various parts of SD model in which the stock variables are represented by the symbol “□” and the flow variables by “X”.

4.1.1 Forward Supply Chain

The forward supply chain begins from the upper left corner of Fig. 3. *Raw Materials* are furnished by external suppliers and recycling the used products (*Raw Material Recovery Rate*) from the reverse channel. *Components Production Rate* depletes raw materials and increase *Components Inventory*. The equations related to component production rate are following:

$$\text{Components Production Rate} = \max(\min(\min(\text{Raw Materials}/\text{Component Production Time} (\text{Expected Distributors Orders} * \text{Components Per Product} - \text{Expected Reusable Component} + \text{CI discrepancy}/\text{CI Adj Time})), \text{Component Production Capacity}), 0)$$

$$\text{CI discrepancy} = \text{MAX}(\text{Desired CI} - \text{Components Inventory}, 0)$$

$$\text{Components Inventory} = \text{INTEG}((\text{Components Production Rate} + \text{Component Refurbish Rate} + \text{Components Acceptance Rate for Direct Reuse}) - \text{Components used for Product Production}, 2400)$$

The remanufacturing process supplements the production process. Producer’s requirement for components is satisfied with a mix of new components produced

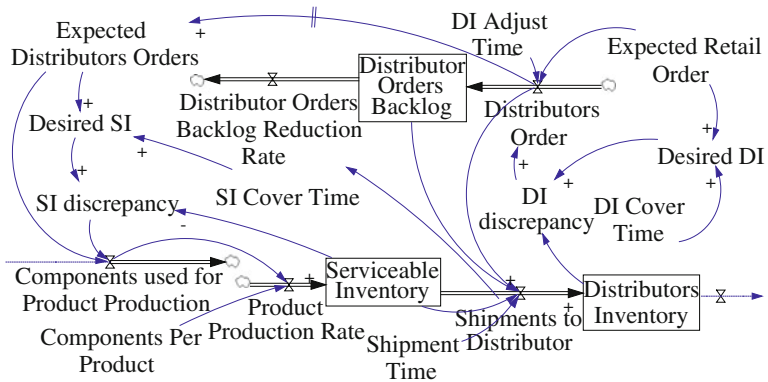


Fig. 4 Stock-flow diagram for product production and distribution

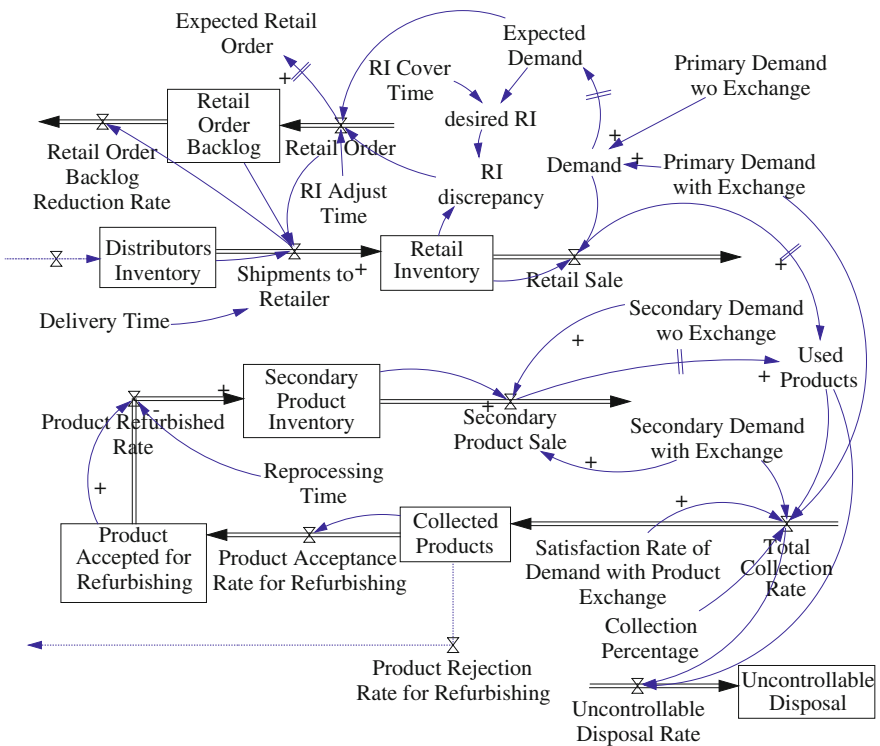


Fig. 5 Stock-flow diagram for demand, product exchange, satisfaction, collection and product refurbishment

by firm, and reusable/refurbished components derived from used products. The equations related to product production rate are following:

Table 1 Demand at various markets with and without PE

Market	Policy	Demand (products/week)
Primary market	Without PE policy	Fuzzy (400,500,600)
	With PE policy	Fuzzy (180,200,220)
Secondary market	Without PE policy	Fuzzy (115,150,185)
	With PE policy	Fuzzy (85,100,115)

Product Production Rate = Components used for Product Production/Components Per Product

SI discrepancy = max (Desired SI – Serviceable Inventory, 0)

Serviceable Inventory = INTEG (Product Production Rate – Shipments to Distributor, 800)

Product production rate depletes Component Inventory and increase Serviceable Inventory. Shipments to Distributor deplete Serviceable Inventory and increase Distributors Inventory. In the same way, products delivered from the upper stream increase the inventory of retailer, which can satisfy the demand of end-users. The equations related to distributor’s inventory, transportation and order are presented below:

Distributors Inventory = INTEG (Shipments to Distributor – Shipments to Retailer, 800)

Shipments to Distributor = IF THEN ELSE (Serviceable Inventory – Distributors Order – Distributor Orders Backlog >=0, Distributors Order + Distributor Orders Backlog, Serviceable Inventory)/Shipment Time

Distributors Order = Expected Retail Order +DI discrepancy/DI Adjust Time

4.1.2 Demand with Product Exchange Policy

Demand in primary as well as in secondary market has been categorized as “demand with exchange” and “demand without exchange”. The product exchange policy has been described in details in “Problem Description” Sect. 2.1. In this chapter, we assume that both primary and secondary demand is lost if it is not satisfied in the current period. Although, *Distributor Orders Backlog* and *Retailer Orders Backlog* are satisfied in a future period. In the proposed SD model, we assume that all the demands are triangular fuzzy number which is shown in Table 1.

4.1.3 Reverse Supply Chain

The recycling process which we incorporate into our SD framework consists of collection, product recovery, and component and material recovery. Sold products after their uses turn into used products. Then, Used Products are either uncontrollably disposed (*Uncontrollable Disposal*) or collected for reuse (*Collected Products*). The equation related to collection rate is following:

*Total Collection Rate = (Secondary demand with exchange + primary demand with exchange)*Satisfaction Rate of Demand with Product Exchange + used products*Collection Percentage*

Out of the *Product Accepted for Refurbishing*, refurbished products (*Product Refurbished Rate*) are sold to the secondary market and the components that are replaced (*Component Replacement Rate*) during product refurbishing by new components are processed further for raw material recovery and component refurbishing.

Product Accepted for Refurbishing = INTEG (Product Acceptance Rate for Refurbishing – Product Refurbished Rate, 0)

In the model, it is assumed that the disassembled components can have three categories: one is direct reusable components (*Components Accepted for Direct Reuse*) that can be directly used to increase the *Components Inventory* in the forward channel; the second is the part of *Components Rejected for Direct Reuse* which requires further reprocessing. After reprocessing, the *Refurbished Components* can be used to increase the *Components Inventory* in the forward channel. The third is rejected components that does not survive the first two screening levels but can be used either for raw material recovery (*Recovered Raw Material*) to increase the *Raw Materials* inventory in the forward channel or sent directly for *Controllable Disposal*.

4.1.4 Possibility Constraints on Collection and Satisfaction Rate

In the proposed SD framework, the uncertainty issues associated with *collection of used products* and *satisfaction of customers willing to exchange their used products for a fresh new/refurbished product* have been quantified using the possibility constraint programming approach.

Collection Rate

It is always expected that the DM would like to maintain a predefined threshold value of collection rate in every period to increase profitability in remanufacturing and to satisfy the legislations requirements. In the proposed SD model, it is assumed that the collection rate (CR) is a triangular fuzzy distribution, $\widetilde{CR} = (a_1, a_2, a_3)$ with a constant deviation of 0.20 from the central value i.e. $a_2 - a_1 = a_3 - a_2 = 0.2$ and that the CR should be more than or equal to 50 % of used product with at least 95 % probability. But, as CR is a fuzzy number, the following possibility constraint has to be satisfied to fulfill DM's requirement:

$$Pos(\widetilde{CR} \geq 0.5) \geq 0.95.$$

Now, from lemma (1) of section (3), it is clear that

$$\frac{a_3 - 0.5}{a_3 - a_2} \geq 0.95 \text{ i.e. } \frac{a_3 - 0.5}{0.2} \geq 0.95 \text{ i.e. } a_3 \geq 0.69. \text{ So, } a_2 \geq 0.49 \text{ and } a_1 \geq 0.29.$$

Hence, we can say with 95 % possibility that the collection percentage will be greater than or equal to 50 % if $\widetilde{CR} = (0.29, 0.49, 0.69)$.

Satisfaction Rate of Demand with Product Exchange Policy

The incorporation of PE policy plays an important role in the process of integrated forward-reverse supply chain as discussed in earlier sections. Therefore, it is always expected that the DM would like satisfy almost all the customers who are interested to exchange their used products to buy a fresh new product from primary market or refurbished products from secondary market. In the proposed model, it is assumed that the satisfaction rate (SR) is a triangular fuzzy distribution $\widetilde{SR} = (c_1, c_2, c_3)$ with a constant deviation of 0.10 from the central value i.e. $c_2 - c_1 = c_3 - c_2 = 0.1$ and that the SR should be than or equal to 85 % of used product in that period with at least 95 % probability. But, as SR is a fuzzy number, the following possibility constraint has to be satisfied to fulfill DM's requirement: $Pos(\widetilde{SR} \geq 0.85) \geq 0.95$

Now, from lemma (1) of section (3), it is clear that $\frac{c_3 - 0.85}{c_3 - c_2} \geq 0.95$ i.e. $\frac{c_3 - 0.85}{0.1} \geq 0.95$ i.e. $c_3 \geq 0.945$. So, $c_2 \geq 0.845$ and $c_1 \geq 0.745$.

Hence, we can say with 95 % possibility that collection percentage will be more than or equal to 85 % if $\widetilde{SR} = (0.745, 0.845, 0.945)$.

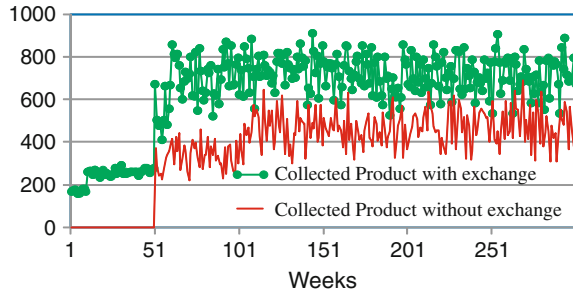
5 Result and Discussion

Before analyzing the performance of the integrated system, we set the important parameters as follows: *components per product* = 3, *component production time* = 1.2 weeks, *product production time* = 2 weeks, *shipment time* from producer to distributor = 1.5 weeks, *delivery time* from distributor to retailer = 1.5 weeks, *cycle life of product* = 50 weeks. We assume that 80 % of collected products are accepted for refurbishing after initial inspection, 15 % of the components get replaced by the new ones from the product which accepted for refurbishing, 65 % of the components are reusable immediately from the collected products which rejected for refurbishing. The length of the time horizon is 300 weeks for the simulation.

5.1 Effect of Product Exchange Policy

We analyze the behavior of the integrated forward-reverse supply chain over the time periods using SD simulation. The incorporation of PE policy in the primary as well as in the secondary market increases the collection rate which is shown in Fig. 6.

Fig. 6 Comparison of collection rates with and without PE policy



5.2 Effect of Three Way Recovery Policy

Table 2 shows the rate at which used products, components and raw materials get recovered. Refurbished products are sold in secondary market. Components which are refurbished or accepted for direct reuse are supplied to the manufacturer to increase the component inventory in forward channel and helps in reducing the fresh production of components. Recovered raw materials are used in the forward process to produce the new components, thus reducing the use of new raw materials.

5.3 Bullwhip Effects and Order Variations

We simulate the demand and order variation of the integrated forward-reverse supply chain. Figure 7 shows the variance of actual demand at retailer and order placed by retailer and distributor. We compute the bullwhip effect of the systems using the following formula given by Chen et al. [13]: $\text{Bullwhip} = \text{Var}(\text{Order Rate}) / \text{Var}(\text{Demand})$ and make a comparison of bullwhip effect at retailer and distributor level *with* and *without* PE policy in Table 3. It is observed that the bullwhip effects of the retailer and distributor *without* PE is bigger than that of *with* PE in the integrated RL. So the results indicate that remanufacturing with PE policy in the integrated reverse supply chain can reduce the bullwhip effect.

5.4 Sensitivity Analysis

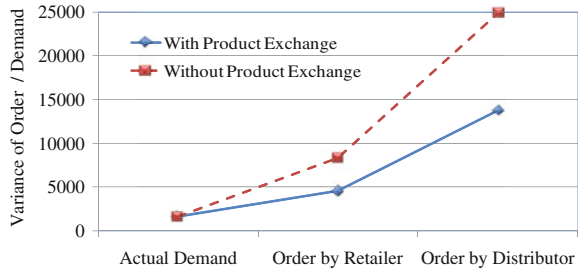
5.4.1 Collection and Satisfaction Rate

Total collection rate depends upon the collection (*Collection Percentage*) of used products directly from the end-user plus the rate (*Satisfaction Rate of Demand with Product Exchange*) at which the used products get collected from the customers through the PE policy in primary as well as secondary market.

Table 2 Effect of three way (product, component and raw material) recovery

Three way recovery	Refurbished product	Component		Raw material
		Direct reusable	Refurbished	
Recovery rate (Units Per Week)	502	243	203	87

Fig. 7 Variance of actual demand and order placed by retailer and distributor



From Table 4 it is observed that, with the increment of collection percentage, refurbished products per week increases. Again, if we can increase the satisfaction rate of the customers who are willing to exchange their used product, then the bullwhip effect reduces at both retailer as well as distributor level (Table 5). This result strengthens our hypothesis that the incorporation of product exchange policy plays an important role in RL.

5.4.2 Refurbish Percentage

Product Refurbishable Percentage assumes how much percentage of collected products is accepted for refurbishing after initial inspection. From Fig. 8, it is clear that if we increase the refurbishable percentage, the average refurbished products per week increases which help to satisfy the demand in the secondary market but the component and raw material recovery rate decreases.

If we increase the product refurbish percentage (PRP), Table 6 shows that the bullwhip effect increases at retailer and distributor level but with the increment of component refurbishment percentage (CRP) and direct reusable component percentage (DRCP), bullwhip effect at both level reduces. The reason behind this phenomenon is that, all the refurbished products are assumed to be sold in secondary market; only refurbished components and recovered raw materials supplement the component and raw material inventory of the forward channel to reduce the inventory discrepancy.

5.4.3 Inventory Cover and Adjustment Time

Inventory cover time determines the safety stock for the inventory. *Inventory adjustment time* represents how quickly a firm tries to correct the discrepancy

Table 3 Order variation and bullwhip effect at various markets with and without PE

	Retailer		Distributor	
	With PE	Without PE	With PE	Without PE
Order variance	4,555	8,331	13,749	24,787
Bullwhip	2.72	4.98	8.21	14.81

Table 4 Sensitivity analysis of collection percentage on product refurbishment

Collection percentage (Refurbished products/week)	30 %	40 %	50 %	60 %	70 %	80 %
	378	440	502	563	625	686

Table 5 Sensitivity analysis of satisfaction rate on bullwhip effect

Satisfaction rate	20 %	40 %	50 %	70 %	80 %	90 %
Bullwhip effect at retailer	4.17	3.29	3.08	2.88	2.82	2.78
Bullwhip effect at distributor	12.25	9.78	9.21	8.67	8.51	8.40

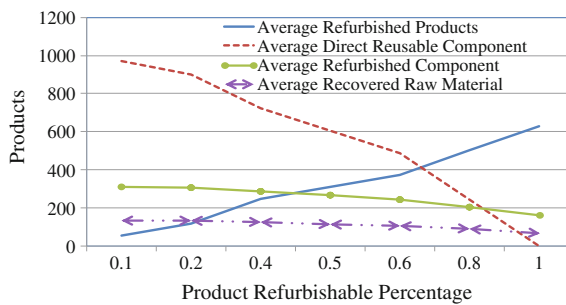


Fig. 8 Sensitivity analysis of refurbish percentage

Table 6 Sensitivity analysis of product, component refurbishing percentage on bullwhip

PRP/CRP/ DRCP	Bullwhip effect due to PRP		Bullwhip effect due to CRP		Bullwhip effect due to DRCP	
	Retailer	Distributor	Retailer	Distributor	Retailer	Distributor
0 %	2.57	7.18	3.13	9.35	2.92	8.84
20 %	2.60	7.38	2.97	8.93	2.88	8.73
40 %	2.66	7.70	2.88	8.69	2.84	8.59
60 %	2.72	8.06	2.82	8.51	2.81	8.47
80 %	2.80	8.45	2.78	8.40	2.78	8.38
100 %	3.60	10.70	2.77	8.34	2.76	8.31

between desired serviceable inventory and actual serviceable inventory. From Table 7, it is clear that bullwhip effect increases both at retailer and distributor level as the retail inventory cover time (RICT) increases; but the bullwhip effect

Table 7 Sensitivity analysis of inventory cover and adjustment time on bullwhip effect

Cover/Adjustment time (week)		0.5	1	1.5	2	2.5	3
Bullwhip effect due to RICT	Retailer	1.00	1.23	2.80	6.30	12.45	21.43
	Distributor	1.66	2.79	8.45	19.88	39.34	67.65
Bullwhip effect due to RIAT	Retailer	24.71	7.32	3.99	2.80	2.24	1.92
	Distributor	79.10	23.35	12.43	8.45	6.52	5.41
Bullwhip effect due to DICT	Retailer	2.89	2.89	2.80	2.78	2.78	2.78
	Distributor	2.91	3.56	8.45	17.51	33.99	59.22
Bullwhip effect due to DIAT	Retailer	2.78	2.79	2.79	2.80	2.80	2.81
	Distributor	45.73	17.38	11.07	8.45	7.06	6.22

decreases at both levels as the retail inventory adjustment time (RIAT) increases. The main reason is that if a firm adjusts the discrepancy between desired serviceable inventory and actual serviceable inventory very quickly, then the variations in order increases. Again, it can be seen from Table 7 that the changes of cover time and adjustment time in distributor level has almost no impact in determining the bullwhip effect at retailer level. But the bullwhip effect increases at distributor level with the increment of distributor inventory cover time (DICT) and decreases with the increment of distributor inventory adjustment time (DIAT).

6 Conclusion and Future Work

In this chapter, we have proposed a SD framework to analyze long-term behavior of a multi-echelon forward-reverse supply chain with fuzzy demand, collection rate and satisfaction rate under possibility constraints by incorporating various recycling activities. The behavior analysis of the developed model indicated that collection rate increases by incorporating PE facility in both primary and secondary market. The simulation results showed that inclusion of PE facility can reduce the order variance and bullwhip effect both at retailer and distributor levels of the integrated reverse logistics. Sensitivity analysis is performed to examine the impact of various parameters on the order variance and bullwhip effect.

Acknowledgments This work was supported partly by IRCC, IIT Bombay under Grant 08IRCC037.

References

1. Blackburn J, Guide V, Souza G, Van Wassenhove L (2004) Reverse supply chain for commercial returns. *Calif Manage Rev* 46(2):6–22
2. Ferrer G, Swaminathan JM (2010) Managing new and differentiated remanufactured products. *Eur J Oper Res* 203:370–379
3. Fleischmann M, Dekker R, van der Laan E, van Numen J, van Wassenhove L, Ruwaard J (1997) Quantitative models for reverse logistics: A review. *Eur J Oper Res* 103:1–17

4. Spengler T, Schroter M (2003) Strategic management of spare parts in closed-loop supply chains: a system dynamics approach. *Interfaces* 33(6):7–17
5. Georgiadis P, Vlachos D (2004) The effect of environmental parameters on product recovery. *Eur J Oper Res* 157(2):449–464
6. Zhou L, Disney S (2006) Bullwhip and inventory variance in a closed loop supply chain. *OR Spectrum* 28:127–149
7. Pati RK, Vrat P, Kumar P (2010) Quantifying bullwhip effect in a closed loop supply chain. *OP SEARCH* 47(4):231–253
8. Qingli D, Hao S, Hui Z (2008) Simulation of remanufacturing in reverse supply chain based on system dynamics. In: IEEE, service systems and service management, 2008 international conference, Melbourne, pp 1–6
9. Das D, Dutta P (2012) A system dynamics framework for an integrated forward-reverse supply chain with fuzzy demand and fuzzy collection rate under possibility constraints. In: Proceedings of the international multiconference of engineers and computer scientists 2012, IMECS 2012, Hong Kong, 14–16 March 2012. Lecture notes in engineering and computer science pp 1592–1597
10. Zimmermann HJ (1996) Fuzzy set theory and its applications, 2nd edn, Kluwer Academic Publisher, Boston
11. Dubois D, Prade H (1983) Ranking fuzzy numbers in the setting of possibility theory. *Inf Sci* 30:183–224
12. Maity K, Maiti M (2007) Possibility and necessity constraints and their defuzzification—a multi-item production-inventory scenario via optimal control theory. *Eur J Oper Res* 177:882–896
13. Chen F, Drezner Z, Ryan JK, Simchi-Levi D (2000) Quantifying the Bullwhip effect in a simple supply chain: the impact of forecasting, lead times, and information. *Manage Sci* 46(3):436–443

A Hybrid-Heuristics Algorithm for k -Minimum Spanning Tree Problems

Hideki Katagiri and Qingqiang Guo

Abstract A combinatorial optimization problem, namely k -Minimum Spanning Tree Problem (KMSTP), is to find a subtree with exactly k edges in an undirected graph G , such that the sum of edges' weights is minimal. This chapter provides a Hybrid algorithm using Memetic Algorithm (MA) as a diversification strategy for Tabu Search (TS) to solve KMSTPs. The genetic operator in the proposed MA is based on dynamic programming, which efficiently finds the optimal subtree in a given tree. The experimental results show that the proposed algorithm is superior to several existing algorithms in terms of solution accuracy and that the algorithm updates some best known solutions that were found by existing algorithms.

Keywords Combinatorial optimization · Hybrid algorithm · k -Minimum spanning tree · Memetic algorithm · Tabu search

1 Introduction

A k -minimum spanning tree problem (KMSTP) or a k -cardinality tree problem is one of combinatorial optimization problems. Let $G = (V, E)$ be an undirected graph, which is made up by connecting vertices V and edges E . Each edge e is

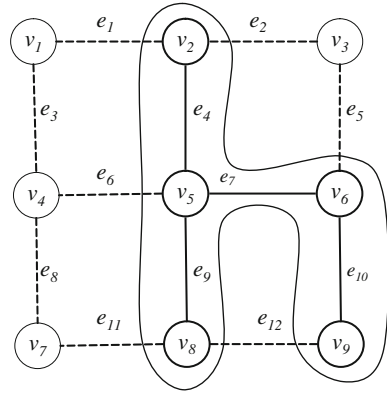
H. Katagiri (✉)

Department of System Cybernetics, Graduate School of Engineering, Hiroshima University,
4-1, Kagamiyama 1-chome, Higashi-Hiroshima 739-8527, Japan
e-mail: katagiri-h@hiroshima-u.ac.jp

Q. Guo

Graduate School of Engineering, Hiroshima University, 4-1, Kagamiyama 1-chome,
Higashi-Hiroshima 739-8527, Japan
e-mail: guoqingqiang@hiroshima-u.ac.jp

Fig. 1 A k -spanning tree
 ($k = 4, \{v_2, v_5, v_6, v_8, v_9\},$
 $\{e_4, e_7, e_9, e_{10}\}$)



attached with a nonnegative value w_e , called weight. The goal is to find a subtree with exactly k ($k \leq |V| - 1$) edges, so that the sum of the weights of edges, denoted by $f(T_k)$, is minimal. The problem is mathematically formulated as follows:

$$\begin{aligned} &\text{minimize } f(T_k) = \sum_{e \in E(T_k)} w_e \\ &\text{subject to } T_k \in S_{T_k} \end{aligned}$$

where $E(T_k)$ is the edges set of tree T_k , and S_{T_k} is a set containing all feasible solutions with k edges in graph G . Like other existing data structures of trees, we represent the solution of the KMSTP as an ordered list of vertices and a list of edges. A simple example of T_k is shown in Fig. 1.

KMSTP was firstly introduced by Hamacher et al. [1]. It has been broadly applied to real-world decision making problems, such as oil-field leasing [2] and image processing [3]. KMSTP was proved to be an NP-hard problem by previous studies [1, 4, 5]. Incidentally, this problem can be polynomially solved in two cases. One is that there are only two distinct weights in a graph [5], and the other is that a graph is given as a tree [6].

Many algorithms have been proposed in order to solve KMSTP. The first exact algorithm was presented by Fischetti et al. [4], in which KMSTP was formulated into an integer linear program (ILP) with generalized circle elimination constraints. Recently, Quintao et al. [7] have proposed two integer programming formulations, Multiflow Formulation and a formulation based on the Miller–Tucker–Zemlin constraints, for solving KMSTP.

Besides the exact solution methods, various types of heuristics and metaheuristic methods have been also proposed. Heuristics based on greedy strategy and Dynamic Programming (DP) were introduced by Ehrgott et al. [8]. Moreover, a heuristics-based on Variable Neighborhood Decomposition Search (VNDS), which has a good performance for problems of small size, was presented by Urosevic et al. [9]. Blum [10] proposed an improved dynamic programming

approach; after a minimum spanning tree for a given graph is obtained, DP is applied in order to obtain an optimal subtree with k edges from the spanning tree. This algorithm has been proved to be efficient even for problems of large size. Concerning metaheuristics, Tabu Search (TS), Evolutionary Computation (EC) and Ant Colony Optimization (ACO), for solving KMSTP, were introduced in [11]. It showed that the performances of the three metaheuristics depend on the characteristics of the tackled instances, such as the numbers of vertices and edges, degree, and cardinality. Recently, a hybrid algorithm based on TS and ACO was constructed by Katagiri et al. [12]. Their experimental results using benchmark instances demonstrate that the hybrid algorithm provides a better performance with solution accuracy over existing algorithms.

However, existing approaches may not be effective in some cases, especially for the problems with large size graphs. That is because the problem becomes complicated exponentially as the size of graph increases. In this chapter, by combining MA and TS for solving the KMSTP, we propose a hybrid algorithm, called HMATS. Numerical results show that the proposed algorithm is competitive to classical algorithms from the viewpoint of solution accuracy and updates some of the best known results.

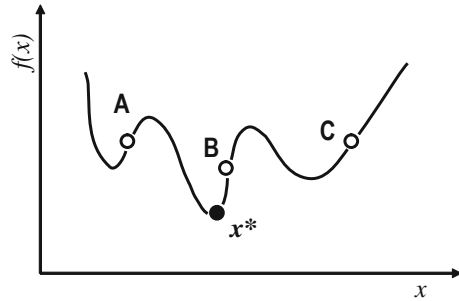
The remainder of this article is organized as follows. Section 2 introduces preliminary background of this paper, including Local search, TS and MA. We describe the proposed hybrid algorithm in Sect. 3. Our new experimental results and analysis are described in Sect. 4. Finally, Sect. 5 concludes this chapter and discusses future studies.

2 Back Ground

2.1 Local Search

A local search is often conducted via some move operators. To be explicit, it translates a solution s to a new one s' ($s' \in N(s)$), where $N(s)$ is a set of neighborhood solutions. In other words, each of solutions is improved by replacing a current solution with a new one. If a movement is carried out once a new solution with a smaller objective function value is found, we call such a strategy the *first improvement*. If a movement is not carried out until a solution with the smallest objective function value from its neighborhood solutions is found, we call such a strategy the *best improvement*. In general, although the solution can be smoothly improved at first, it may not reach the best solution in most cases. The local search may repeatedly search some areas that had been searched, which may cause a cycle among particular solutions in the worst case. In other words, the search may easily fall into local optimum if the local search is simply used.

Fig. 2 A minimization problem



2.2 Tabu Search

Tabu search (TS), firstly proposed by Glover et al. [13, 14], is one of the most well-known metaheuristics for solving combinatorial optimization problems. TS can be considered as an extended local search on the basis of a variety of iterative local search strategies for discrete optimization. The most important characteristic of TS is that it uses a concept of memory to control movements via a dynamic list of forbidden moves. To be more specific, the specific search space will be “tabu” (prohibited) to visit for a while. This mechanism as well as the dynamic use of memory allows TS to intensify or diversify its search procedure in order to escape from local optima. Many experiments showed that TS can lead to a significant improvement in terms of solution quality. TS has also been proved to be effective in solving KMSTPs [11].

2.3 Memetic Algorithm

As a rising field of Evolutionary Computation, Memetic Algorithm (MA) was first introduced by Moscato in 1989 [15]. Traditional Evolutionary Computations (e.g. Genetic Algorithm) have been applied widely to solve optimization problems because of their good search abilities. However, they may not be an efficient method for solving some problems which contain many local optima. For example, for a minimization problem shown in Fig. 2, it may take a number of iterations (generations) to reach the best solution x^* by Evolutionary Computation. On the other hand, it is easy for local search to find the best solution if the search starts from solution B . Usually solution B can be easily generated by evolutionary computation operators (e.g. crossover or mutation).

3 Proposed Algorithm

For the problem with a large graph, the length of tabu list may become very long in order to enlarge its search area. Accordingly, the computational cost would be expended rapidly as the length increases. Then, MA with evolutionary computation operators is expected to help TS overcome this shortcoming. In the proposed hybrid algorithm, MA is used as a diversification strategy for TS.

3.1 Initial Solution

Let e_i denote an edge, one of its vertices belongs to k -spanning tree T_k and the other one does not. The set of all the e_i for a k -spanning tree T_k is denoted by $E_{NH}(T_k)$, which means a neighbor edges set for T_k . To generate a k -spanning tree, a vertex is randomly selected as the initial vertex for T_k . Then, edges will be added to T_k one by one until a k -spanning tree is constructed. Each edge e to be added is selected as $e := \operatorname{argmin} \{w_e | e \in E_{NH}(T_k)\}$ with probability p , or it is selected randomly from $E_{NH}(T_k)$ with probability $1 - p$. In an extreme case, k -spanning tree is constructed in a greedy strategy if the probability $p = 1$, or it is randomly constructed if the probability $p = 0$. This procedure is expected to attain a good balance between the goodness of initial solutions and their diversity if a proper probability p is set. In this research, we determine the probability $p = 0.85$ due to preliminary numeral experiments.

3.2 Tabu Search

This section devotes describing the details on TS in the proposed algorithm. The TS used in this paper is similar to the method in [16, 17], in which aspiration criterion and dynamic length of tabu lists are used. To begin with, the flowchart of the proposed Hybrid algorithm, which is based on MA and TS, called HMATS, is shown in Fig. 3.

In Fig. 3, T_k^{cur} denotes the current k -spanning tree, and T_k^{lb} means the best solution that was found at the present moment in the local search. The details of other parameters are the same as those in [17].

3.2.1 Local Search

In local search, the current subtree is replaced with a new one, by exchanging one of its vertices with a vertex not contained in it. Let $V(T_k)$ be a set of vertices in k -spanning tree T_k . Then, we define the neighbourhood vertex set of T_k as follows:

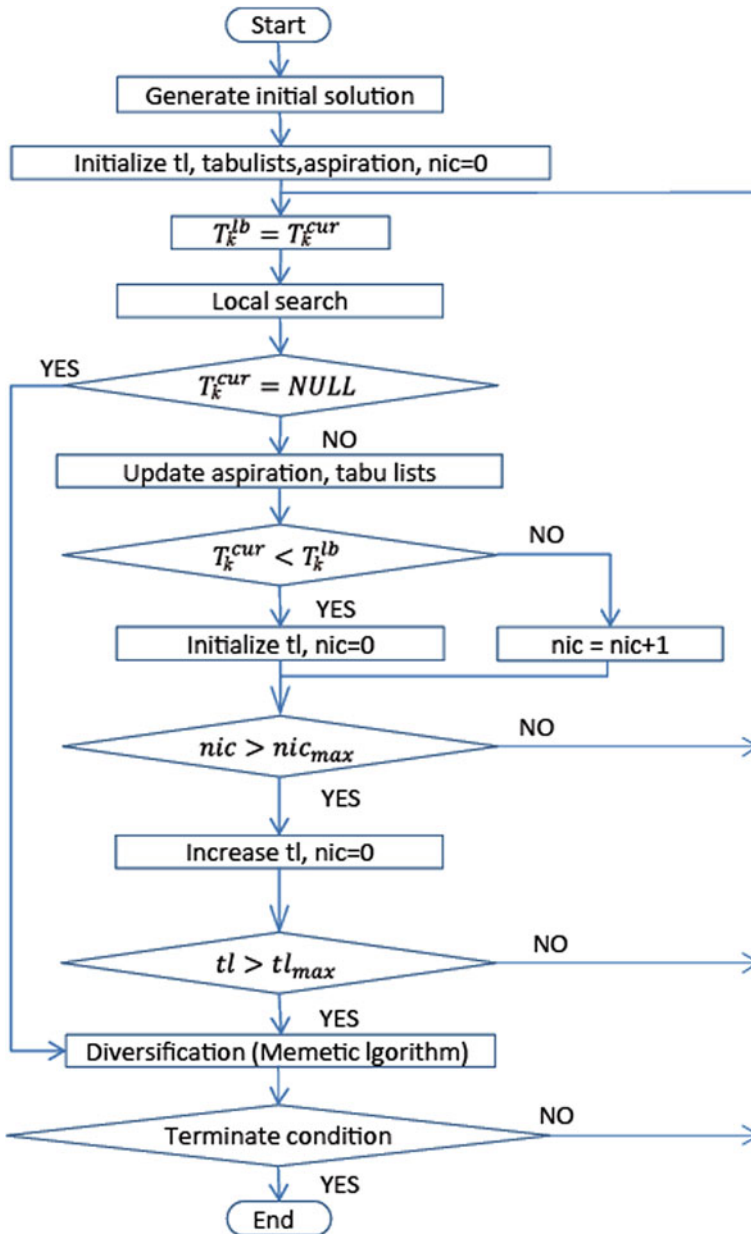


Fig. 3 Flowchart of HMATS

$$V_{NH}(T_k) := \{v|(v, v') \in E, v \notin V(T_k), v' \in V(T_k)\}.$$

Let T_k^{NH} be a new k -spanning tree obtained by adding one vertex $v_{add} \in V_{NH}(T_k)$ to T_k and deleting one vertex $v_{del} \in V(T_k)$. The set of neighbourhood k -spanning tree, denoted by $NH(T_k)$, contains all the possible k -spanning trees T_k^{NH} in G .

In order to obtain a better solution, we employ the best improvement in the local search. The k -spanning tree T_k^{NHbest} which has the smallest objective function value in the neighbourhood k -spanning tree set of T_k is selected as follows:

$$T_k^{NHbest} := argmin\{f(T_k^{NH})|T_k^{NH} \in NH(T_k)\}.$$

3.2.2 Parameters

The core procedure of TS is to forbid some movements to be employed based on memory. In the proposed algorithm, the “tabu” (forbiddance) is applied to edges that have added to or deleted from the k -spanning tree recently, and tabu lists are used as a memory to record the edges that are forbidden to add or delete. In the proposed algorithm, we use two tabu lists, namely, InList and Outlist: InList and OutList are adopted to keep the records of removed edges and added edges, respectively. Tabu tenure, which is generally dependent on the length of tabu lists, is a period for which it forbids edges in the tabu lists from adding or deleting. The lengths of tabu lists are dynamically changed in the proposed algorithm, which helps the local search implement intensification and diversification strategies. If the best solution in this iteration has not been updated for nic_{max} movements, the length of tabu lists will be increased by tl_{inc} . The search stops if the length of either of tabu lists is larger than tl_{max} .

3.2.3 Aspiration Criterion

The “tabu” mechanism, which forbids some of the movements to be employed, helps the algorithm avoid falling into local optima. However, this mechanism may also forbid movements to employed even if the movements are so excellent that it reaches the best solution. In order to avoid such a situation, a procedure called aspiration criterion is used in the proposed algorithm. That is, if $\gamma_e > f(T_k^{NH})$ is satisfied, the movement is acceptable even if e is included in InList or OutList. Parameters γ_e called aspiration criterion levels are given to all of edges and are initially set as:

$$\gamma_e = \begin{cases} f(T_k^{cur}), & e \in E(T_k^{cur}) \\ \infty, & e \notin E(T_k^{cur}) \end{cases}$$

For each explored solution T_k , γ_e is updated as $\gamma_e := f(T_k)$ for each $e \in E(T_k)$.

3.3 Memetic Algorithm

When it is judged that a solution is not improved any more by TS, MA is applied in the proposed algorithm as a diversification strategy. This procedure is also helpful to make the algorithm escape from local optimum by evolutionary computation operators. The proposed MA is shown as follows:

Memetic Algorithm

- Step 1 $P := \text{Initialize } (P)$
 while the stop criterion is not satisfied do
 Step 2 $P' := \text{Genetic Operations } (P)$
 Step 3 $P' := \text{Updating Population } (P')$
 Step 4 $P' := \text{Replace } (P \cup P')$
 Step 5 $P'' := \text{Tabu Search based Local Search } (P')$
 Step 6 $P'' := \text{Updating Population } (P'')$
 Step 7 $P := \text{Replace } (P \cup P'')$
 endwhile
 Step 8 Return (P)

In the above algorithm, P means the current population of individuals, P' the population of the individuals generated from the Genetic Operations, and P'' the population of the individuals improved by TS based Local Search.

3.3.1 Generating Initial Population

Firstly, the size of P is determined. Obviously the bigger the size is, the more new individuals is expected to be generated in general. However, for the sake of the computing cost for large size problems, we select the size of P as 4 in this study. The population P includes one individual obtained by TS and the individuals generated using the procedure described in Sect. 3.1. Additionally, in order to make sure that the structures of all the individuals are not the same, we replace the reduplicate individuals with new k -spanning trees.

3.3.2 Genetic Operation

In this paper, crossover, usually adopted in Evolutionary Computation, is used as the Genetic Operation. We use crossover operator to enlarge the explored domain, so that the algorithm can escape from local optima easily. The crossover operator is completed by two procedures.

Firstly, each individual except T_k itself in P is considered to be a cross partner T_k^C for T_k . More concretely, if T_k and its cross partner T_k^C have at least one common vertex, then a vertex set $V(G^C)$ is defined as: $V(G^C) = V(T_k) \cup V(T_k^C)$.

Otherwise, edges and vertices should be added to T_k until at least one common vertex is found, by the procedure described in Sect. 3.1 with probability $(1 - p)$. A spanning tree T^{SP} , which contains all the vertices of $V(G^C)$, is constructed using the procedure introduced in Sect. 3.1.

Then, a DP-based algorithm for obtaining k -minimum spanning tree, originally introduced in [10], is applied to T^{SP} for finding out the best k -spanning tree. Since the DP is efficient, the crossover operator will help us obtain a good solution in a short time.

3.3.3 Updating Population

Since there are so many improvements for individuals under the operators above, some of the individuals may be the same in offspring populations. In order to avoid searching redundancy, an updating operator is applied to the offspring populations. Concretely, through the updating operator, the repeated individuals is replaced with k -spanning trees constructed by the procedure introduced in Sect. 3.1 with probability $(1 - p)$.

3.3.4 Tabu Search-Based Local Search

After Genetic Operation, each offspring will be further improved by local search. Now we introduce TS-based local search in order to conduct local search effectively. Consequently, the local search here used is the same as the one described in Sect. 3.2, except for the length of tabu lists.

3.3.5 Stopping Criteria of Memetic Algorithm

We define a generation as an idle generation if the objective function value is not improved in that generation. If idle generation occurs continuously for several times, the MA stops. In order to have a complete searching, the MA will not end until the number of idle generations is twice as much as the largest idle generations happened before. The stopping criteria function is shown as below:

$$i := \operatorname{argmax}\{i_s, 2 * i_{max}\},$$

where i is the number of continuously occurred idle generations, i_s is the smallest number of idle generations which is determined in advance, and i_{max} is the largest idle generation happened before.

3.4 Centralization Strategy

Centralization strategy is also concerned in the proposed algorithm. Realizing the fact that the best solution may exist near to the space where MA just searched, the algorithm is restarted after the best solution generated from MA is regarded as the initial solution for TS.

3.5 Stopping Criterion

The stopping criterion of the proposed algorithm is the same as that described in Sect. 3.3.5. The proposed algorithm ends when the objective function value is no longer improved for a previously specified number of iterations.

4 Experimental Study

4.1 Comparing HMATS to Existing Algorithms

In this section, computational experiments using well-known benchmark instances are conducted in order to evaluate the effectiveness of the proposed algorithm (HMATS). To be more explicit, the proposed algorithm is compared with three state-of-the-art existing algorithms described in [17]. One is a Hybrid algorithm (Hybr.K) based on the TS and ACO, originally proposed by Katagiri et al. [12]. The other two algorithms are metaheuristics based on tabu search algorithm (TSB) and ant colony optimization (ACOB), both of which are introduced by Blum et al. [11].

HMATS is coded in C language and compiled with C-Compiler: Microsoft Visual C++ 7.1. Computation environment is CPU: Pentium 4 3.06 GHz, RAM: 1 GB, OS: Microsoft Windows XP. The parameter settings as well as the source code used for all the experiments in this study were the same as those provided by Katagiri et al. and Blum et al. All of the 75 benchmark instances were downloaded from the web page [18]. The prepared benchmark instances are classified into small graphs ($|E| \leq 2000$) and large graphs ($|E| > 2000$), where $|E|$ indicates the number of edges in graph G . For small graphs, the Limit Time is set to be 300 (s), and experiments are executed for 30 independent runs. For large graphs, the Limit Time is set to be 6000 (s), and experiments are executed for ten independent runs.

Tables 1, 2, 3, and 4 show parts of the results of experiments for several benchmark instances. $|V|$ indicates the number of vertices in a graph, and k denotes the cardinality of KMSTP. BKS means the Best Known Solutions which have been obtained by Blum et al. through their experiments. The descriptions “Best”, “Mean” and “Worst” in the tables mean the best, average, and the worst objective function values, respectively. Results highlighted in bold mean that the corresponding

Table 1 |Vl: 400 |El: 800

G	k:160 BKS:3062				k:240 BKS:5224			
	HMATS	Hybr.K	TSB	ACOB	HMATS	Hybr.K	TSB	ACOB
Best	3062	3067	3070	3063	5224	5230	5238	5228
Mean	3067.1	3069.2	3071.8	3065.6	5226.5	5235.3	5247.2	5240.1
Worst	3087	3072	3076	3069	5241	5238	5257	5261

Table 2 |Vl:1000 |El: 1250

G	k:200 BKS:3456				k:600 BKS:15916			
	HMATS	Hybr.K	TSB	ACOB	HMATS	Hybr.K	TSB	ACOB
Best	3456	3553	3588	3460	15917	16041	16132	16186
Mean	3505.2	3600.1	3608.3	3471.3	15955.8	16084.9	16162.3	16285.7
Worst	3569	3644	3636	3487	15998	16111	16218	16384

Table 3 |Vl:2500 |El: 4900

G	k:500 BKS:8150				k:1000 BKS:17437			
	HMATS	Hybr.K	TSB	ACOB	HMATS	Hybr.K	TSB	ACOB
Best	*8146	8621	8722	8406	17508	18343	18374	17973
Mean	8259.4	8673.9	8824.5	8481	17810.2	18663.8	18465.2	18155.5
Worst	8356	8811	8874	8605	18242	19031	18624	18358

Table 4 |Vl:2500 |El: 4900

G	k:1500 BKS:28683				k:2000 BKS:43627			
	HMATS	Hybr.K	TSB	ACOB	HMATS	Hybr.K	TSB	ACOB
Best	*28549	29771	29878	30410	*43558	44101	43955	45520
Mean	28748	29938	30097.4	30518.2	43597.7	44411.3	44059.7	45697.7
Worst	29353	30061	30404	30722	43646	44876	44151	45940

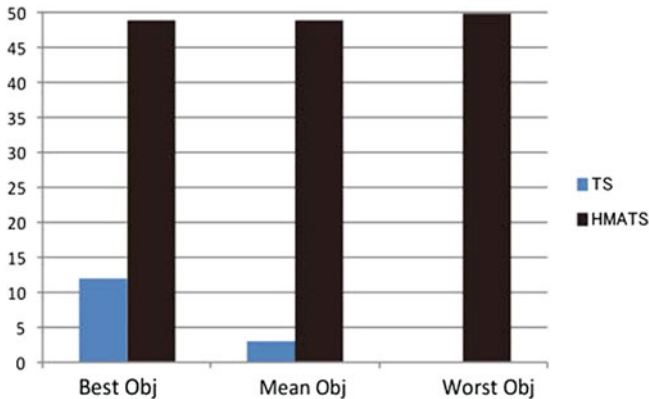
algorithm beats others. The values marked by * denote that the best known solutions are updated by that algorithm.

Tables 1 and 2 show the results for the instances with small graphs. The performance of HMATS is better than existing algorithms in terms of “Best value”. Tables 3 and 4 show the results for the instances with large graphs. It is concluded from the results of “Best value”, “Mean value” and “Worst value” that the performance of HMATS is superior to the comparative algorithms. This is due to the effect of the diversification strategy based on MA, which enlarges the search area through genetic operators.

Table 5 summarizes the number of instances which the corresponding algorithm outperforms or performs as much as other comparative algorithms. For example, the description “Best” indicates the number of instances in which the corresponding algorithm reaches the best results among four algorithms. It can be

Table 5 Performances for each algorithms

Objective function value	HMATS	Hybr.K	TSB	ACOB
Best	75 (instances)	30	15	18
Mean	59	27	4	11
Worst	47	35	5	13
BKS	47	30	15	17

**Fig. 4** Comparing results between TS and HMATS

found that HMATS reached the best solutions in 75 instances, which is significantly better than the other algorithms. In addition, HMATS reaches the best known solutions in 47 instances. Moreover, HMATS is robust because it outperforms the other three algorithms in terms of not only mean values but also worst values. Thus we can conclude that the proposed method is superior in terms of precision for various types of graphs.

4.2 Analysis of the Effectiveness of the Memetic Algorithm

The experimental results shown in the previous subsection suggests that HMATS outperforms three existing algorithms. In order to investigate the effect of MA, which is considered to play an important role in the proposed algorithm, we compare HMATS to TS. The programming language C is used as the programming language, and the program was compiled with C Compiler: Microsoft Visual C++ 2010 Express. Both HMATS and TS employ the same parameters and the same values of the parameters, and all the experiments are conducted for ten independent runs on the same computer. Similar to the former experiments, we apply two algorithms to 50 benchmark problems downloaded from [18].

The experimental results are shown in Fig. 4. The vertical axis provides the number of instances that an algorithm reaches a better objective function value out of 50 instances. Best objective function values (“Best Obj”) shows that MA improves TS in reaching a better solution. Mean objective function values (headed “Mean Obj”) and the worst objective function values (headed “Worst Obj”) show that MA plays a very important role in the robustness of HTSMA. Thus it is shown that the solution accuracy of the proposed algorithm is obviously improved by incorporating MA into TS.

5 Conclusion

In this chapter, we have proposed a hybrid algorithm based on MA and TS for solving KMSTP. The experiments using existing benchmark instances have shown that the proposed algorithm outperforms other existing algorithms and has stronger robustness. It has been observed through additional experiments that a proper combination of MA and TS is efficient for solving k -spanning tree problems. As one of future works, we will conduct experiments with much bigger size of benchmark problems in order to clearly show the effectiveness of the proposed algorithm.

References

1. Hamacher HW, Jorsten K, Maffioli F (1991) Weighted k -cardinality trees. Technical report, Politecnico di Milano, Dipartimento di Elettronica, Italy
2. Hamacher HW, Jorsten K (1993) Optimal relinquishment according to the Norwegian petrol law: a combinatorial optimization approach. Technical report, no. 7/93, Norwegian School of Economics and Business Administration, Bergen, Norway
3. Ma B, Hero A, Gorman J, Michel O (2000), Image registration with minimum spanning tree algorithm. In: IEEE international conference on image processing
4. Fischetti M, Hamacher HW, Jorsten K, Maffioli F (1994) Weighted k -cardinality trees: complexity and polyhedral structure. *Networks* 24:11–21
5. Ravi D, Sundaram R, Marathe MV, Rosenkrantz DJ, Ravi SS (1996) Spanning trees-short or small. *SIAM J Discrete Math* 9(2):178–200
6. Maffioli F (1991) Finding a best subtree of a tree. Technical report, Politecnico di Milano, Dipartimento di Elettronica e Informazione
7. Quintaoa FP, da Cunha AS, Mateus GR, Lucena A (2010) The k -cardinality tree problem: reformulations and Lagrangian relaxation. *Discrete Appl Math* 158:1305–1314
8. Ehrgott M, Freitag J, Hamacher HW, MaLoli F (1997) Heuristics for the k -cardinality tree and subgraph problem. *Asia-Pacific J Oper Res* 14(1):87–114
9. Urosevic D, Brimberg J, Mladenovic N (2004) Variable neighbourhood decomposition search for the edge weighted k -cardinality tree problem. *Comput Oper Res* 31:1205–1213
10. Blum C (2007) Revisiting dynamic programming for finding optimal subtrees in trees. *Eur J Oper Res* 177:102–115
11. Blum C, Blesa M (2005) New metaheuristic approaches for the edge-weighted k -cardinality tree problem. *Comput Oper Res* 32:1355–1377

12. Katagiri H, Hayashida T, Nishizaki I, Guo Q (2012) A hybrid algorithm based on tabu search and ant colony optimization for k-minimum spanning tree problems. *Expert Syst Appl* 39(5):5681–5686
13. Glover F (1986) Future paths for integer programming and links to artificial intelligence. *Comput Oper Res* 5:533–549
14. Glover F, Laguna M (1997) *Tabu search*. Kluwer Academic Publishers, Dordrecht
15. Moscato P (1989) *On evolution, search, optimization, genetic algorithms and martial arts: toward memetic algorithms*. Caltech concurrent computation program, CalTech, Pasadena, CA, Rep 826
16. Katagiri H, Hayashida T, Nishizaki I, Ishimatsu J (2010) An approximate solution method based on tabu search for k-minimum spanning tree problems. *Int J Knowl Eng Soft Data Paradigms* 2(3):263–274
17. Guo Q, Katagiri H, Hayashida T, Nishizaki I (2012) A hybrid algorithm based on memetic algorithm and tabu search for k-minimum spanning tree problems. In: *Lecture notes in engineering and computer science: proceedings of the international multiconference of engineers and computer scientists 2012, IMECS 2012, 14–16 March, 2012, Hong Kong*, pp 1611–1616
18. A library for the edge-weighted k-cardinality tree problem (2003). <http://iridia.ulb.ac.be/~cblum/kctlib/>. Accessed 20 June 2012

An Optimal Model of Adding Relation Between the Top and a Member of a Linking Pin Organization Structure with K Subordinates

Kiyoshi Sawada, Hidefumi Kawakatsu and Takashi Mitsuishi

Abstract This study considers the addition of relation to an organization structure such that the communication of information between every member in the organization becomes the most efficient. This paper proposes a model of adding relation to a linking pin organization structure where every pair of siblings in a complete K -ary tree of height H is adjacent. When a new edge between the root and a node with a depth N is added, an optimal depth N^* is obtained by minimizing the total distance which is the sum of lengths of shortest paths between every pair of all nodes.

Keywords Complete K -ary tree · Linking pin · Organization structure · Pyramid organization · Shortest path · Total distance

K. Sawada (✉)

Department of Policy Studies, University of Marketing and Distribution Sciences,
Kobe 651-2188, Japan
e-mail: Kiyoshi_Sawada@red.umds.ac.jp

H. Kawakatsu

Department of Economics and Information Science,
Onomichi University, Onomichi 722-8506, Japan
e-mail: kawakatsu@onomichi-u.ac.jp

T. Mitsuishi

Department of Commerce, University of Marketing and Distribution Sciences,
Kobe 651-2188, Japan
e-mail: mitsuishi@umds.ac.jp

1 Introduction

A pyramid organization is a hierarchical structure based on the principle of unity of command [1] that every member except the top in the organization should have a single immediate superior. On the other hand an organization characterized by System 4 [2] has a structure in which relations between members of the same section are added to the pyramid organization structure. Members of middle layers of System 4 which are both members of the upper units and chiefs of the lower units are called linking pins, and this type of organization is called a linking pin organization. In the linking pin organization there exist relations between each superior and his direct subordinates and those between members which have the same immediate superior.

The linking pin organization structure can be expressed as a structure where every pair of siblings which are nodes which have the same parent in a rooted tree is adjacent, if we let nodes and edges in the structure correspond to members and relations between members in the organization respectively. Then the linking pin organization structure is characterized by the number of subordinates of each member, that is, the number of children of each node and the number of levels in the organization, that is, the height of the rooted tree, and so on [3, 4]. Moreover, the path between a pair of nodes in the structure is equivalent to the route of communication of information between a pair of members in the organization, and adding edges to the structure is equivalent to forming additional relations other than those between each superior and his direct subordinates and between members which have the same direct subordinate.

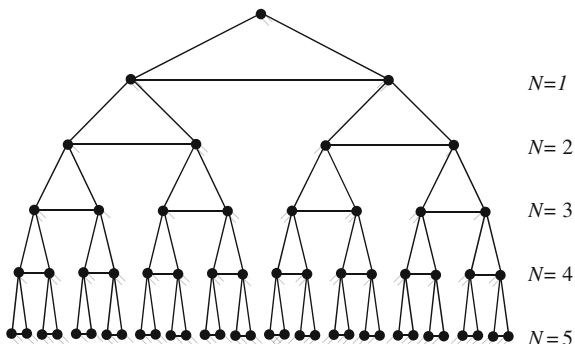
The purpose of our study is to obtain an optimal set of additional relations to the linking pin organization such that the communication of information between every member in the organization becomes the most efficient. This means that we obtain a set of additional edges to the structure minimizing the sum of lengths of shortest paths between every pair of all nodes [5].

We have obtained an optimal depth for each of the following two models of adding relations in the same level to a complete K -ary linking pin structure of height H ($H = 2, 3, \dots$) where every pair of siblings in a complete K -ary tree of height H is adjacent: (i) a model of adding an edge between two nodes with the same depth and (ii) a model of adding edges between every pair of nodes with the same depth [6]. A complete K -ary tree is a rooted tree in which all leaves have the same depth and all internal nodes have K ($K = 2, 3, \dots$) children [7]. Figure 1 shows an example of a complete K -ary linking pin structure of $K = 2$ and $H = 5$. In Fig. 1 the value of N expresses the depth of each node.

This paper proposes a model of adding an edge between the root and a node with a depth N ($N = 2, 3, \dots, H$) in a complete K -ary linking pin structure of height H ($H = 2, 3, \dots$) [8]. This model corresponds to the formation of an additional relation between the top and one member of an organization.

If $l_{i,j}$ ($= l_{j,i}$) denotes the distance, which is the number of edges in the shortest path from a node v_i to a node v_j ($i, j = 1, 2, \dots, (K^{H+1} - 1)/(K - 1)$) in the

Fig. 1 An example of a complete K -ary linking pin structure of $K = 2$ and $H = 5$



complete K -ary linking pin structure of height H , then $\sum_{i < j} l_{i,j}$ is the total distance. Furthermore, if $l'_{i,j}$ denotes the distance from v_i to v_j after adding an edge in this model, $l_{i,j} - l'_{i,j}$ is called the shortening distance between v_i and v_j , and $\sum_{i < j} (l_{i,j} - l'_{i,j})$ is called the *total shortening distance*. Minimizing the total distance is equivalent to maximizing the total shortening distance. When an edge between the root and a node with a depth N is added to the complete K -ary linking pin structure of height H , an optimal depth N^* is obtained by maximizing the total shortening distance.

In Sect. 2 we formulate the total shortening distance of the above model. In Sect. 3 we show an optimal depth N^* which maximizes the total shortening distance and in Sect. 4 we illustrate an optimal depth N^* with numerical examples.

2 Formulation of Total Shortening Distance

This section formulates the total shortening distance when an edge between the root and a node with a depth N ($N = 2, 3, \dots, H$) is added to a complete K -ary linking pin structure of height H ($H = 2, 3, \dots$).

Let v_N denote the node of depth N which gets adjacent to the root. The set of descendants of v_N is denoted by V_1 . The set of ancestors of parent of v_N is denoted by V_2 . (Note that every node is a descendant and an ancestor of itself [7].) Let V_3 denote the set obtained by removing V_1 and V_2 from the set of all nodes of the complete K -ary linking pin structure.

The sum of shortening distances between every pair of nodes in V_1 and nodes in V_2 is given by

$$A_H(N) = M(H - N) \sum_{i=1}^{\lfloor \frac{N}{2} \rfloor} (N - 2i + 1), \tag{1}$$

where $M(h)$ denotes the number of nodes of a complete K -ary tree of height h ($h = 0, 1, 2, \dots$), and $\lfloor x \rfloor$ denotes the maximum integer which is equal to or less than x . The sum of shortening distances between every pair of nodes in V_2 is given by

$$B(N) = \sum_{i=1}^{\lfloor \frac{N}{2} \rfloor - 1} \sum_{j=1}^{\lfloor \frac{N}{2} \rfloor - i} (N - 2i - 2j + 1), \quad (2)$$

where we define $\sum_{i=1}^0 \cdot = 0$. The sum of shortening distances between every pair of nodes in V_1 and nodes in V_3 is given by

$$C_H(N) = M(H - N) \sum_{i=1}^{\lfloor \frac{N-1}{2} \rfloor} (K - 1)M(H - i)(N - 2i). \quad (3)$$

The sum of shortening distances between every pair of nodes in V_2 and nodes in V_3 is given by

$$\begin{aligned} D_H(N) = & \sum_{i=1}^{\lfloor \frac{N-1}{2} \rfloor - 1} (K - 1)M(H - i) \sum_{j=1}^{\lfloor \frac{N-1}{2} \rfloor - i} (N - 2i - 2j) \\ & + \sum_{i=1}^{\lfloor \frac{N-1}{2} \rfloor} (K - 1)M(H - N + i - 1) \sum_{j=1}^{\lfloor \frac{N-1}{2} \rfloor - i + 1} (N - 2i - 2j + 2), \end{aligned} \quad (4)$$

where we define $\sum_{i=1}^{-1} \cdot = 0$. The sum of shortening distances between every pair of nodes in V_3 is given by

$$E_H(N) = \sum_{i=1}^{\lfloor \frac{N}{2} \rfloor - 1} (K - 1)M(H - N + i - 1) \sum_{j=1}^{\lfloor \frac{N}{2} \rfloor - i} (K - 1)M(H - j)(N - 2i - 2j + 1). \quad (5)$$

From the above equations, the total shortening distance $S_H(N)$ is given by

$$S_H(N) = A_H(N) + B(N) + C_H(N) + D_H(N) + E_H(N). \quad (6)$$

3 An Optimal Depth

This section obtains an optimal depth N^* which maximizes the total shortening distance $S_H(N)$.

Let us classify $S_H(N)$ into two cases of $N = 2L$ where $L = 1, 2, \dots, \lfloor H/2 \rfloor$ and $N = 2L + 1$ where $L = 1, 2, \dots, \lfloor (H - 1)/2 \rfloor$. Since the number of nodes of a complete K -ary tree of height h is

$$M(h) = \frac{K^{h+1} - 1}{K - 1}, \quad (7)$$

$S_H(2L)$ and $S_H(2L + 1)$ become

$$\begin{aligned}
 S_H(2L) &= \frac{K^{H-2L+1} - 1}{K - 1} \sum_{i=1}^L (2L - 2i + 1) + \sum_{i=1}^{L-1} \sum_{j=1}^{L-i} (2L - 2i - 2j + 1) \\
 &+ \frac{K^{H-2L+1} - 1}{K - 1} \sum_{i=1}^{L-1} (K^{H-i+1} - 1)(2L - 2i) \\
 &+ \sum_{i=1}^{L-2} (K^{H-i+1} - 1) \sum_{j=1}^{L-i-1} (2L - 2i - 2j) \\
 &+ \sum_{i=1}^{L-1} (K^{H-2L+i} - 1) \sum_{j=1}^{L-i} (2L - 2i - 2j + 2) \\
 &+ \sum_{i=1}^{L-1} (K^{H-2L+i} - 1) \sum_{j=1}^{L-i} (K^{H-j+1} - 1)(2L - 2i - 2j + 1) \\
 &= \frac{1}{(K - 1)^3} \left\{ K^{2H-3L+3} - K^{2H-2L+3} - K^{2H-2L+2} + K^{2H-L+2} - K^{H-2L+2} \right. \\
 &\quad + K^{H-2L+1} - K^{H-L+3} - K^{H-L+1} - (L - 1)K^{H+3} + K^{H+2} \\
 &\quad \left. + LK^{H+1} - L(K - 1)^2 \right\},
 \end{aligned} \tag{8}$$

and

$$\begin{aligned}
 S_H(2L + 1) &= \frac{K^{H-2L} - 1}{K - 1} \sum_{i=1}^L (2L - 2i + 2) + \sum_{i=1}^{L-1} \sum_{j=1}^{L-i} (2L - 2i - 2j + 2) \\
 &+ \frac{K^{H-2L} - 1}{K - 1} \sum_{i=1}^L (K^{H-i+1} - 1)(2L - 2i + 1) \\
 &+ \sum_{i=1}^{L-1} (K^{H-i+1} - 1) \sum_{j=1}^{L-i} (2L - 2i - 2j + 1) \\
 &+ \sum_{i=1}^L (K^{H-2L+i-1} - 1) \sum_{j=1}^{L-i+1} (2L - 2i - 2j + 3) \\
 &+ \sum_{i=1}^{L-1} (K^{H-2L+i-1} - 1) \sum_{j=1}^{L-i} (K^{H-j+1} - 1)(2L - 2i - 2j + 2) \\
 &= \frac{1}{(K - 1)^3} \left\{ \frac{K - 1}{K + 1} K^{2H-3L+2} + K^{2H-3L+1} - K^{2H-2L+2} - K^{2H-2L+1} \right. \\
 &\quad + \frac{2}{K + 1} K^{2H-L+2} - K^{H-2L+2} + K^{H-2L+1} - K^{H-L+3} + K^{H-L+2} \\
 &\quad - 2K^{H-L+1} - (L - 1)K^{H+3} + (L + 1)K^{H+1} \\
 &\quad \left. - L(K - 1)^3 - L(K - 1)^2 \right\},
 \end{aligned} \tag{9}$$

respectively.

Lemma 1 (i) If $L = 1$, then $S_H(2L) < S_H(2L + 1)$.

(ii) If $L \geq 2$, then $S_H(2L) > S_H(2L + 1)$.

Proof

(i) If $L = 1$, then

$$S_H(2L) - S_H(2L + 1) = \frac{K^{2H-2}}{K-1} \left(-1 + \frac{1}{K^{H-2}} + \frac{K-1}{K^{2H-2}} \right) < 0. \quad (10)$$

(ii) If $L \geq 2$, then

$$\begin{aligned} & S_H(2L) - S_H(2L + 1) \\ &= \frac{1}{(K-1)^2} \left\{ \frac{K^{2H-L+2}}{K+1} \left(1 - \frac{K^2 + 2K + 1}{K^{L+1}} + \frac{K^2 + K + 1}{K^{2L+1}} \right) \right. \\ & \quad \left. + K^{H+1} \left(1 - \frac{1}{K^L} \right) + L(K-1)^2 \right\} \\ & > 0, \end{aligned} \quad (11)$$

where $L = 2, 3, \dots, \lfloor (H-1)/2 \rfloor$.

Q.E.D.

Lemma 2 If $L \geq 2$, then $L^* = 2$ maximizes $S_H(2L)$.

Proof

If $L \geq 2$, then $L^* = 2$ maximizes $S_H(2L)$ since

$$\begin{aligned} & S_H(2L) - S_H(2L + 2) \\ &= \frac{1}{(K-1)^2} \left\{ K^{2H-L+1} \left(1 - \frac{K^2 + 2K + 1}{K^{L+1}} + \frac{K^2 + K + 1}{K^{2L+1}} \right) \right. \\ & \quad \left. + K^{H+2} \left(1 + \frac{1}{K} - \frac{K^2 + 1}{K^{L+2}} - \frac{K^2 - 1}{K^{2L+3}} \right) + K - 1 \right\} \\ & > 0, \end{aligned} \quad (12)$$

where $L = 2, 3, \dots, \lfloor H/2 \rfloor - 1$.

Q.E.D.

Lemma 3 (i) If $H = 4$, then $S_H(3) > S_H(4)$.

(ii) If $H \geq 5$, then $S_H(3) < S_H(4)$.

Proof

(i) If $H = 4$, then

$$S_H(3) - S_H(4) = K^2 - 2 > 0. \quad (13)$$

(ii) If $H \geq 5$, then

$$S_H(3) - S_H(4) = \frac{K^{2H-3}}{K-1} \left(-1 + \frac{K^4 + K^2 - K - 1}{K^H} - \frac{K-2}{K^{2H-3}} \right) < 0. \quad (14)$$

Q.E.D.

Table 1 Optimal depth N^* and the total shortening distance $S_H(N^*)$

H	$K = 2$		$K = 3$		$K = 4$	
	N^*	$S_H(N^*)$	N^*	$S_H(N^*)$	N^*	$S_H(N^*)$
2	2	1	2	1	2	1
3	3	10	3	30	3	68
4	3	54	3	336	3	1300
5	4	298	4	3905	4	25626
6	4	1366	4	37961	4	430186
7	4	5806	4	350081	4	6963626
8	4	23902	4	3176009	4	111740586
9	4	96958	4	28659905	4	1789139626
10	4	390526	4	258166601	4	28631394986

Theorem 4 (i) If $H = 2$, then $N^* = 2$ maximizes $S_H(N)$.

(ii) If $H = 3$ or $H = 4$, then $N^* = 3$ maximizes $S_H(N)$.

(iii) If $H \geq 5$, then $N^* = 4$ maximizes $S_H(N)$.

Proof

(i) If $H = 2$, then $N^* = 2$ trivially.

(ii) If $H = 3$, then $N^* = 3$ since $S_H(2) < S_H(3)$ from (i) of Lemma 1. If $H = 4$, then $N^* = 3$ since $S_H(2) < S_H(3)$ from (i) of Lemma 1 and $S_H(3) > S_H(4)$ from (i) of Lemma 3.

(iii) If $H \geq 5$, then $N^* = 4$ since

(a) $S_H(2) < S_H(3)$ from (i) of Lemma 1,

(b) $S_H(3) < S_H(4)$ from (ii) of Lemma 3,

(c) If $L \geq 2$, then $S_H(2L) > S_H(2L + 1)$ from (ii) of Lemma 1, and

(d) If $L \geq 2$, then $L^* = 2$ maximizes $S_H(2L)$ from (ii) of Lemma 2.

Q.E.D.

4 Numerical Examples

Table 1 shows numerical examples of the optimal depth N^* and the total shortening distance $S_H(N^*)$ in the case of $K = 2, 3, 4$ and $H = 2, 3, \dots, 10$.

Table 1 reveals that $N^* = 2$ maximizes $S_H(N)$ in the case of $H = 2$, $N^* = 3$ maximizes $S_H(N)$ in the case of $H = 3, 4$, and $N^* = 4$ maximizes $S_H(N)$ in the case of $H = 5, 6, \dots, 10$, irrespective of K .

5 Conclusions

This study considered the addition of relation to a linking pin organization structure such that the communication of information between every member in the organization becomes the most efficient. For a model of adding an edge between the root and a node with a depth N to a complete K -ary linking pin structure of height H where every pair of siblings in a complete K -ary tree of height H is adjacent, we obtained an optimal depth N^* which maximizes the total shortening distance. Theorem 4 shows that the most efficient manner of adding relation between the top and a subordinate is to add the relation to a subordinate of the second, the third or the fourth level depending on the number of levels in the organization structure. Furthermore, we illustrate an optimal depth N^* and the total shortening distance $S_H(N^*)$ with numerical examples.

References

1. Koontz H, O'Donnell C, Weihrich H (1980) *Management*, 7th edn. McGraw-Hill, New York
2. Likert R, Likert JG (1976) *New ways of managing conflict*. McGraw-Hill, New York
3. Robbins SP (2003) *Essentials of organizational behavior*, 7th edn. Prentice Hall, Upper Saddle River
4. Takahara Y, Mesarovic M (2003) *Organization structure: cybernetic systems foundation*. Kluwer Academic/Plenum Publishers, New York
5. Sawada K, Wilson R (2006) Models of adding relations to an organization structure of a complete K -ary tree. *Eur J Oper Res* 174:1491–1500
6. Sawada K (2008) Adding relations in the same level of a linking pin type organization structure. *IAENG Int J Appl Math* 38:20–25
7. Cormen TH, Leiserson CE, Rivest RL, Stein C (2001) *Introduction to algorithms*, 2nd edn. MIT Press, Cambridge
8. Sawada K, Kawakatsu H, Mitsuishi T (2012) A model of adding relation between the top and a member of a linking pin organization structure with K subordinates. In: *Proceedings of the international multiconference of engineers and computer scientists 2012, IMECS 2012. Lecture notes in engineering and computer science*. Hong Kong, 14–16 Mar 2012, pp 1598–1601

An Agri-food Supply Chain Model to Empower Farmers for Supplying Deteriorated Product to Modern Retailer

Wahyudi Sutopo, Muh. Hisjam and Yuniaristanto

Abstract Small-scale vegetables farmers have to deal with marketing and low selling price problems although they produce good quality of vegetables. The low prices are mostly because of the market information asymmetry. An agri-food supply chain (ASC) model that involves the corporate social responsibility (CSR) activities to empower the farmers as qualified supplier for modern retail (MR) is proposed to overcome the problems. The CSR activities are designed to enhance business skills and improve the quality of vegetables distribution system. For farmers, quality improvement of vegetable is expected to increase supply of vegetable to MR, whereas for MR to increase selling price and lengthen deterioration time of vegetable thus reducing the risk of unsold vegetable to consumer. Multi-objective optimization programming is developed to determine the amount supply, level of farmers training skills, quality improvement target, and the CSR total cost. The results show that both farmers and MR gain benefit from CSR activities although MR must allocate some amount of budget as CSR cost.

Keywords Agri-food supply chain · Corporate social responsibility · Deterioration time · Multi-objectives optimization · Risk of unsold vegetable · Supplier of modern retailer

W. Sutopo (✉) · Muh. Hisjam · Yuniaristanto
Laboratory of Logistics and Business System, Department of Industrial Engineering,
Faculty of Engineering, Sebelas Maret University, Jl. Ir. Sutami 36A,
Kentingan, Surakarta 57126, Indonesia
e-mail: sutopo@uns.ac.id

Muh. Hisjam
e-mail: mhisjam@yahoo.com

Yuniaristanto
e-mail: yuniaristanto@gmail.com

1 Introduction

Small-scale vegetables farmers who live around Mt. Merapi in Boyolali regency, Indonesia, produce good quality of vegetables but they persistently face with the marketing problems. As a result, they are forced to sell their commodities at a very low price to the customers [1, 13]. They may get better profit from their harvest if their group or cooperative could sell to modern retailers directly. To become supplier for modern retailers, the Farmer Group and/or Cooperative (FGC) not only have to deal with market information asymmetry in managing small-scale farmers so they lost in competition from large-scale competitors, but also have to meet the relevant provisions of modern retail on product specifications, delivery terms, and internal business requirements [2, 3].

The case described in the previous paragraph can be seen as the integration of key business processes from the integrated system in agri-food supply chain (ASC) that consists of three main entities namely farmers, the FGC, and the modern retailers (MR), and also the customers as end users. The ASC is created by the organizations responsible for producing, processing, distribution, process, and marketing the commodities to the final consumers [4]. Moreover, corporations that do business in Indonesia must be accountable to put into practice the environmental and social responsibility [5, 6]. As a consequence, modern retailers must commit to take part in sustainable economic development, in order to improve the quality of life and environment, which will be beneficial for the company itself, the local community, and society in general. Thus, implementing the Corporate Social Responsibility (CSR) programs in the integrated system of ASC could solve the problem described in the first paragraph.

Several researches had been conducted to improve supply chain coordination [7, 8], to make business contracts [9, 10], and to understand the effect of ASC improvements [2, 11–15]. Reference [13] studies the effects of CSR activities which consist of farmers' skill level enhancement and vegetable's quality improvement to benefit both of farmers and MR. In [14], several options of CSR activities are studied using goal programming technique. Deterioration time of vegetable and risk faced by MR are included in [15].

Nowadays firms are able to promote and implement the responsible business practices throughout its supply chain [16, 17]. Furthermore, many researchers have tried to develop the implementation of CSR program in supply chains. Firms engage in CSR activities as a way to enhance firm-supplier relationship [18, 19], to generate customer loyalty and manage their risk [20, 21], and to propose the guidance of CRS implementation in supply chain responsibility [22–24].

However, few of them propose a suitable model to solve the problem by considering the internal and external problems of the small-scale vegetables farmers. Furthermore, all existing models do not consider the particular weaknesses of the small-scale farmers in accessing capital, adopting new technology, and managing internal business. This chapter is an extension of [15] by providing comprehensive results and analysis. By allocating an amount of money to fund

CSR activities, farmers can provide high quality vegetables to MR. As a result, deterioration time of sold vegetables can be lengthen hence the risk faced by MR can be reduced.

2 Problem Formulation

ASC network which is considered as relevant system of the problem is depicted in Fig. 1. The members of FGC consist of several farmers which plant several types of vegetables. Each farmer delivers it to the FGC, and then the FGC sells the vegetable to modern retailers (MR). However due to restrictive quality specification imposed by modern retailers, the FGC must conduct strict quality inspection before it can be sold to modern retailers. As a consequence, FGC can only sell high quality vegetables to MR. On the other hand; MR needs adequate supply of vegetables to conduct their business. FGC can also sell their products directly to end customer. However, this option offers less attractive revenue.

In order to increase supply of vegetables, MR proposes CSR activities conducted by the division of human resource development (HRD) of the modern retailers as well as CSR benefits for empowering farmer. The CSR activities are designed to enhance business skills of the farmers and to improve the quality of vegetables distribution system. The objective of the CSR is not only to maximize the profit of farmers, but also to maximize CSR benefits for the modern retailers. The decisions concern on the amount of supply, level of farmers training skills, quality improvement target, and the CSR total cost. By allocating an amount of money to fund CSR activities, farmers can provide high quality vegetables to MR. As a result, deterioration time of sold vegetables can be lengthen hence the risk of vegetable decay faced by MR can be reduced.

From the description above, this paper tries to answer the following questions: given an amount of CSR fund by MR how much must be allocated to business skills improvement of the farmers, how much must be allocated to improve the quality of vegetables distribution system, and what is the effect of quality improvement to deterioration time of vegetables and the risk faced by MR.

3 Model Formulation

The ASC model is formulated as mix integer linear programming (MILP). The following notations are used to develop the proposed model.

Index

- $t \in T$ period set
- $i \in I$ farmer set
- $j \in J$ cooperative group set

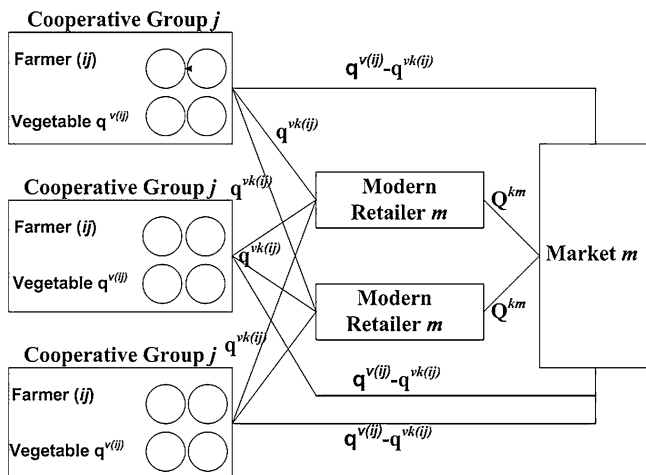


Fig. 1 ASC network of relevant system (Source Sutopo et al. [15])

- $k \in K$ modern retailer set
- $m \in M$ consumer market set
- $v \in V$ vegetable set

Variables and parameters

- $q_t^{v(ij)}$ the quantity of the vegetables produced by farmer i in cooperative group j
- $q_t^{vk(ij)}$ the quantity of the vegetables transacted by retailer k from farmer i in cooperative group j at period t
- p_t^{ymk} price of vegetable v from retailer k to market m as the effect of CSR
- $p_t^{vm(ij)}$ price of vegetables v transacted by market m from farmer i in cooperative group j
- $p_t^{vk(ij)}$ price of vegetable v transacted by modern retailer k from farmer i in cooperative group j
- p_t price of vegetable v from retailer k to market m
- $c_t^{v(ij)}$ production cost of farmer i in cooperative group j
- $h_t^{(ij)}$ training fund received by farmer i in cooperative group j
- $g_t^{v(ij)}$ quality improvement fund received by farmer i in cooperative group j
- $\varpi^{(ij)}$ initial skill level of farmer i in cooperative group j
- ϑ_t^k CSR cost faced by modern retailer k
- ϕ maximum skill level of farmers business skill
- p_t normal selling price transacted by consumer from modern retailer
- r_t risk faced by modern retailer
- α skill level factor of farmers in quality improvement
- y_0^{ij} initial quality of vegetable produced by farmer i at cooperative group j

- $g(y)$ function of vegetable selling price as CSR effects
- $f(y)$ function of vegetable quality as CSR effects
- $h(\tau)$ function of risk faced by modern retailer
- Q_t^{km} the quantity of the vegetables transacted between retailer k and each demand market m at time t
- $F_t^{(ij)}$ training taken by farmer i at cooperative groups j in period t
- $\psi_t^{v(ij)}$ quality improvement percentage of vegetable v farmer i at cooperative group j .

The FGC consists of several farmers who inhabit the area nearby the cooperative and/or group, and a farmer cannot be a member of more than one cooperative and/or group. The FCG sell the vegetables to a local modern retailer (MR) with better price than to traditional market (TM). However, they must select the vegetables based on their quality because of the modern retailer quality requirement. Therefore, not all vegetables produced by farmers can be sold to the modern retailer. The relationship between the quantity of the vegetables that produced by farmers and that can be sold to modern retailer can be expressed by (1). Modern retailer sells the vegetables acquired from the cooperatives groups to consumer market. The products flow transacted by consumer is defined as (2). Equation (2) express the sum of all vegetables sold to customer market less than or equal to the sum of all vegetables bought from all cooperative groups.

$$\sum_{i \in I} q_t^{v(ij)} \geq \sum_{k \in K} q_t^{vk(ij)} \tag{1}$$

$$\sum_{m \in M} Q_t^{ymk} \geq \sum_{k \in K} q_t^{vk(ij)} \tag{2}$$

The proposed model assumes that by improving quality of vegetable, the followings are occurred: farmers can increase their volume of vegetables sold to MR, MR obtained high-quality vegetables from farmers thus improving its selling price to customer and reducing the risk. The relationship between the quantities of vegetable that can be sold to modern retailer with the quality of vegetable produced by farmers is depicted in Fig. 2.

For instance, if the quality of vegetable produced by farmers is less than fifty percent, modern retailer will not buy it from farmers. As the quality of vegetable increases, modern retailer will buy some amount of vegetable. Suppose that the quality of vegetable is eighty percent then modern retailer will buy seventy percent of the vegetable produced by farmers, and the remaining vegetable is sold to traditional market. This relationship can be written as

$$q_t^{vk(ij)} = f(y) q_t^{v(ij)} \tag{3}$$

As part of CSR commitments, modern retailer assists farmers to improve vegetable quality. By improving quality of the vegetable, farmers can sell their

Fig. 2 Quantity of vegetables sold to MR as a function of quality

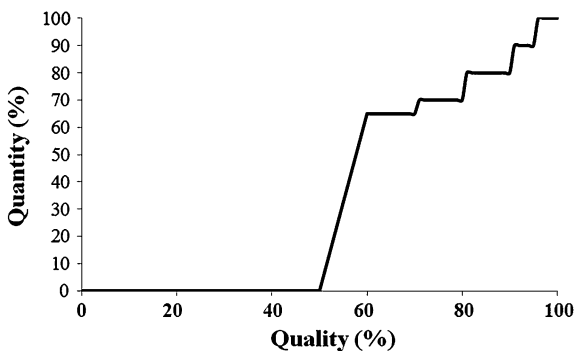
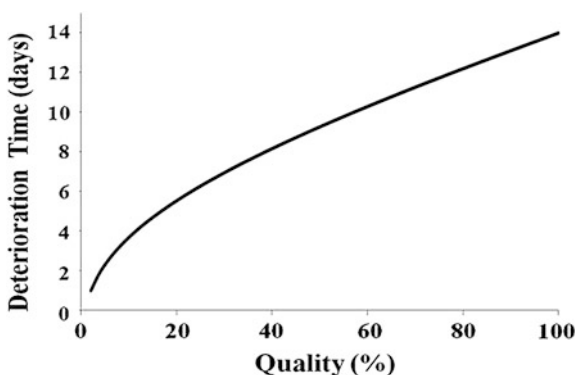


Fig. 3 Deterioration time as a function of vegetable quality



product to modern retailer with higher price than sell it to traditional market. Hence for farmers, CSR will benefit them with additional revenue. As for modern retailer, quality improvement can make modern retailer increase the selling price to consumer. Thus, both farmers and modern retailer will get advantage by CSR activities. Impact of quality improvement to selling price is expressed as

$$p_t^{vmk} = g(y) p_t \tag{4}$$

Not only benefit does modern retailer get, but also it has to face the risk to conduct CSR activities. The risk caused by unsold vegetable that suffers from deterioration. In order to clarify this concept, Fig. 3 depicts the relationship between vegetable quality and deterioration time. This relationship is expressed as

$$y = a \cdot \exp[-c/(\tau + b)] \tag{5}$$

where y denotes quality of vegetable, τ denotes deterioration time of vegetable, whereas a , b , and c are parameters [10].

For instance, modern retailer buys vegetable with category of seventy percent which will deteriorate in 12 days. After 12 days, vegetable becomes obsolete hence it can not be sold to consumer. Thus, if and all parameters in (5) are known,

the risk faced by modern retailer is a function of deterioration time and can be calculated by the expression below

$$r_t = h(\tau) \tag{6}$$

The ASC model can be formulated as constrained MILP. It consists of two objective functions that must be minimized, i.e. MR profit and farmers' profit. Equation (7) consists of two parts. The first three terms belong to farmers' objectives and the remaining terms belong to MR objective. The first term in (7) represents the revenue of farmers from selling vegetables to modern retailer. The second term represents the production cost, whereas the last term represents revenue from selling to traditional market. The fourth and remaining terms of (7) belong to MR objective. The fourth term represents revenue, the fifth represents the purchasing cost, the total CSR cost, and the last term represents the risk cost.

Objective function

$$\begin{aligned} \text{Max} \quad & \sum_{t \in T} \sum_{v \in V} \sum_{k \in K} \sum_{j \in J} \sum_{i \in I} p_t^{vk(ij)} q_t^{vk(ij)} - \sum_{t \in T} \sum_{v \in V} \sum_{j \in J} \sum_{i \in I} c_t^{v(ij)} q_t^{v(ij)} \\ & + \sum_{t \in T} \sum_{v \in V} \sum_{m \in M} \sum_{j \in J} \sum_{i \in I} p_t^{vm(ij)} \left(q_t^{v(ij)} - q_t^{vk(ij)} \right) + \sum_{t \in T} \sum_{v \in V} \sum_{k \in K} \sum_{m \in M} p_t^{vmk} Q_t^{km} \\ & - \sum_{t \in T} \sum_{v \in V} \sum_{k \in K} \sum_{j \in J} \sum_{i \in I} p_t^{vk(ij)} q_t^{vk(ij)} - \sum_{t \in T} \vartheta_t^k - \sum_{t \in T} p_t^{vk(ij)} q_t^{vk(ij)} r_t \end{aligned} \tag{7}$$

Constraints

$$\sum_{t \in T} \vartheta_t^k = \sum_{t \in T} \sum_{v \in V} \sum_{j \in J} \sum_{i \in I} \psi_t^{v(ij)} g_t^{v(ij)} + \sum_{t \in T} \sum_{j \in J} \sum_{i \in I} F_t^{(ij)} h_t^{(ij)} \tag{8}$$

$$\sum_{i \in I} q_t^{v(ij)} \geq \sum_{k \in K} \left(f(y) + \alpha F_t^{(ij)} \right) q_t^{vk(ij)}, \forall t, j, v \tag{9}$$

$$\sum_{m \in M} Q_t^{vmk} \leq \sum_{k \in K} q_t^{vk(ij)}, \forall t, j, v \tag{10}$$

$$\sum_{t \in T} \vartheta_t^k \leq CSR \tag{11}$$

$$\sum_{t \in T} \sum_{j \in J} \sum_{i \in I} \varpi^{(ij)} + F_t^{(ij)} \leq \phi \tag{12}$$

$$f(y) + \alpha F_t^{(ij)} \leq 1, \forall t \tag{13}$$

$$\psi_t^{v(ij)} = f(y) - y_0^{ij} \tag{14}$$

$$F_t^{(ij)} \in \mathbb{Z}_+, \alpha_t^{(ij)} \geq 0, \psi_t^{v(ij)} \geq 0, \forall i, j, t \tag{15}$$

The total CSR cost that the modern retailer have to deal with is expressed in (8), which states that the total CSR cost is equal to the sum of the vegetables quality improvement cost and the farmers skill enhancement cost. The first term of the right hand side of (8) expresses the cost for improving vegetables which can be obtained by multiplying the quality improvement percentage and the associated improvement cost. The vegetables flows transacted by modern retailer from farmers balance is defined in (9). The vegetables flows transacted by consumer market from modern retailer are expressed in (10). It stated that the sum of the vegetables bought by all consumer markets must not exceed the quantity bought by modern retailer. Modern retailer spends budget for CSR activities. The amount of the budget is limited to the amount of the CSR budget authorized by modern retailer owner as in (11). Equation (12) states that the training level taken by farmer added with the current level must not exceed the maximum skill level determined by the modern retailer. Quality improvement and farmer's skill enhancement effects to quantity of vegetable sold by farmer to modern retailer must not exceed 100 % as expressed in (13). Equation (14) described the quality improvement balance. Finally, the last equation is utilized to enforce non-negativity and integrity for decision variables.

4 Results and Discussion

In this computational study, we investigate the impact of the changes in parameters in the multi period ASC supply chain model on optimum vegetables flow, training level taken by farmers, quality improvement percentage, and risk faced by MR. The algorithm used to solve the MILP formulation was branch and bound [25]. We use IBM® ILOG® CPLEX Academic version to solve the MILP formulation.

In order to illustrate the proposed model, agri-food supply chain in Boyolali regency is taken as case study. The supply chain comprises three cooperative groups $j, j = 1, 2, 3$; 1 modern retailer $k, k = 1$; 1 vegetable $v, v = 1$; 1 consumer market $m, m = 1$; and 2 periods $t, t = 1, 2$. The numbers of farmers associated with the cooperative groups are 3, 2, and 4 respectively. Table 1 represents the cooperative groups and their members, along with the skill of each farmer. The initial skill level owned by each farmer has been investigated by [15], which was conducted by field survey and interview with cooperative groups, farmer, and MR. MR determine what kinds of skill must be possessed by each farmer for each level as qualified supplier. For instance, farmer A which is member of Cooperative Group 1 was assessed by [15] initially in Level 1. MR requires that each farmer must have skill level 4 to become a supplier, i.e. all the vegetables produced by farmer are acceptable to be purchased by MR.

The parameters values for (4) are $a = 10.8159$, $b = 50$, and $c = 7$. These values are estimated based on direct observation on the real system and interviews with farmers, cooperative groups, and modern retailer. Farmers in Boyolali mostly produce cabbage and carrot. For this computational study, cabbage is taken as

Table 1 Farmer management skill data

Farmer	Cooperative group	Initial skill level	Required skill level	Training need
A	1	1	4	3
B	1	2	4	2
C	1	3	4	1
D	2	1	4	3
E	2	3	4	1
F	3	2	4	2
G	3	2	4	2
H	3	1	4	3
I	3	2	4	2

Table 2 Quantity and selling price as function of quality

Vegetable quality (y) (%)	Quantity function f(y) (%)	Price function g(y)
0 to <60	0 to <60	1.00
–60 to <75	–60 to <75	1.06
75 to <90	75 to <90	1.09
90–100	90–100	1.12

object. Initial quality of vegetable is estimated and averaged 60 %, i.e. currently all farmers can only sell 60 % of their production to MR and the remaining is sold to traditional market.

Table 2 represents the effect of quality to selling price of MR and quantity sold by farmers to MR. Currently, MR assess’ the quality of vegetable sold by farmers in Boyolali regency is only about 50–60 %. This value is estimated based on the amount of vegetables that currently accepted by MR. In order to increase the quantity of vegetable sold to MR becomes 65 %, the quality of vegetable produced by farmers must be about at level 60–70 %. For farmers, the benefit of this quality improvement is additional revenue since additional quantity is sold to MR which gives better price than sold to traditional market. For MR, the benefit is additional revenue expected from selling higher quality vegetable with higher price.

Aside from selling price increase, MR also expects that quality improvement would also reduce the risk that vegetable can not be sold to customer due to deterioration, by increasing shelf life of vegetable. Table 3 presents risk function faced by modern retailer as a function of deterioration time. Suppose by doing quality improvement, vegetable will deteriorate after 12 days. Then the risk faced by modern retailer that it is unable to get gain from vegetable sold is 0.2.

The effects of CSR to both farmers and MR are investigated. First, we examined the effects of CSR to farmers in Fig. 4. The profit of farmer tends to increase as the CSR budget increases. This can be explained by (3) that as the quality of vegetable increases, so does the quantity sold to MR. As a consequence, farmer gain more revenue from selling to MR which offers higher revenue than selling to traditional market. The slope of farmer profit is increase sharply until CSR budget

Table 3 Shelf life of vegetable and risk function

Shelf life (τ)	Risk function $h(\tau)$
1 to <9	0.17
9 to <11	0.13
11 to <13	0.10
13–14	0.05

in amount of 200,000, after CSR budget exceeds 200,000 the slope is also increase but not as big as former. Overall, Fig. 4 gives us an indication that CSR activities will bring benefit to farmer. Based on this result, we develop a proposition that supports CSR effect for farmers.

Proposition 1 *Farmers will get benefit from CSR activities which give more revenue by selling more vegetable to MR that gives higher price.*

Proof The proof is trivial. Suppose the amount of vegetable produced by farmer i in cooperative j for any period t is denoted by $q^{v(i)}$ and its value is the sum of the amount vegetable sold to MR $q^{vk(i)}$ and to traditional market $q^{vm(i)}$. Off course by (3) CSR budget will increase the quality of vegetable thus increase the number of $q^{vk(i)}$. Since the price offered by MR is higher than that of traditional market, the revenue of farmers will also increase. \square

Figure 5 describes the effect of CSR activities to profit and revenue of MR. The dash-line plot represents the revenue of MR because of various allocated CSR budget. The plot is similar to effect of CSR to profit of farmers. It is not surprising since MR is expected to gain more revenue by selling additional amount of vegetable as in (3) with higher selling price as in (4). However, this is not entirely true for MR profit. The slope of MR profit increase sharply until the amount of CSR budget is about 200,000, however after that value is exceeded the profit slowly goes downward or the slope is negative, i.e. the profit is decreased as the amount of CSR budget increase. The reason behind the decrease of profit is CSR budget that must be spent by MR to improve quality is greater than of additional revenue from quality improvement of vegetable. Furthermore, as the amount of vegetable purchased also takes into account for cost incurred.

Proposition 2 *MR will get benefit from CSR activities, however as the CSR budget, which can be considered fixed cost incurred, is increased the profit might be reached to the same level where no CSR activities conducted, i.e. no benefit gained.*

Proof MR objective function in (7) expressed the profit as the revenue subtracted by operational cost, CSR cost, and risk cost. Suppose that all vegetable purchased by MR is sold to customer, i.e. $Q^{km} = q^{vk(i)} = q$. Then benefit from CSR can be expressed as $(p^{vmk} - p^{vk(i)} - rp^{vk(i)})q - \vartheta$. We can see that the expression negatively correlated with CSR cost, the bigger CSR cost is the smaller profit. \square

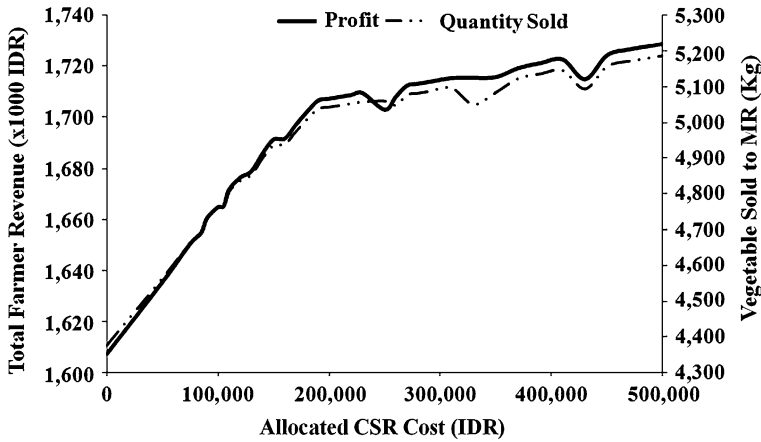


Fig. 4 CSR effect on profit of farmers

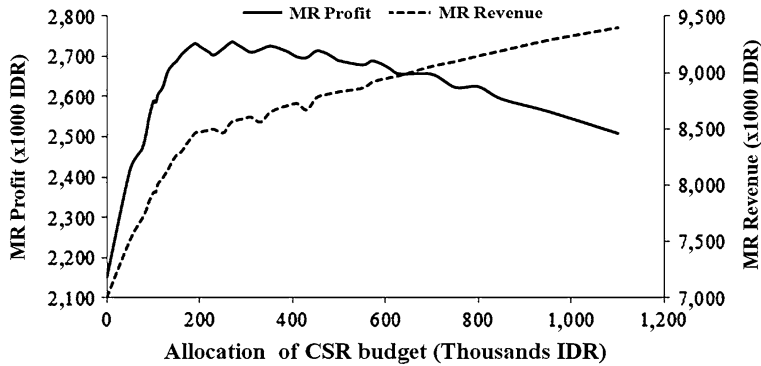


Fig. 5 CSR effect on profit and revenue of MR

The effect of CSR activities to risk faced by MR is depicted in Fig. 6. Risk cost of MR is increase until optimum budget allocation. However, the value is decreased until its value exceeds optimum allocation. We suggest that this fact support the proposed model that before optimum allocation, CSR budget is allocated mostly to quality improvement. Until it exceeds the optimum allocation, CSR budget is allocated mostly to improve deterioration time. Interesting result is found when we divide risk cost with quantity purchased. The plot line is somewhat decrease as the allocated budget increase. Hence we suggest that the effectiveness of CSR activities to risk faced by MR caused by the deterioration time is measured by this ratio.

The proposed model can be used by decision makers as follows:

1. Determine the optimum CSR budget allocation.
2. Determine the CSR effects to vegetable quality.

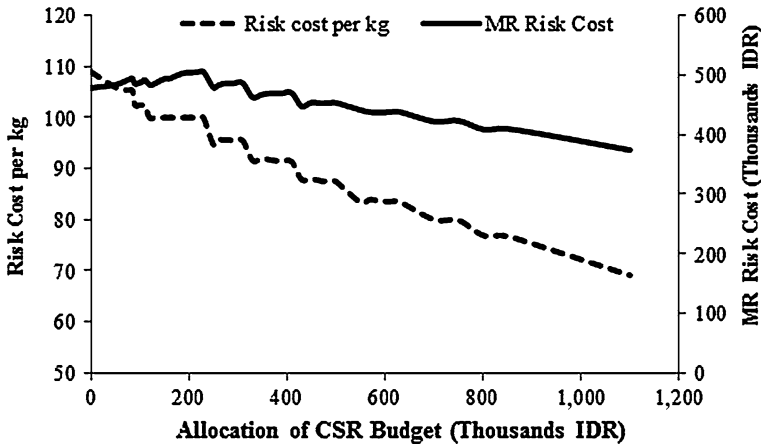


Fig. 6 CSR effect on MR’s risk caused by deterioration time

3. Determine what kind of program that must be consisted in CSR activities.

We believe that optimum CSR budget allocation and vegetable quality has been presented clearly in the previous section. For the CSR program, the work of [15] can be used as reference. The various programs proposed by [15] include product knowledge and packaging system, accounting system and quality system, procurement, ordering, contracting, and etc. Our difficulties in implementing the proposed model include how to determine quality function, the deterioration time parameters, and the risk function. We suggest that readers take careful attention to them.

5 Conclusion and Future Work

In our proposed model, CSR activities which consist of farmers’ skill enhancement and vegetable’s quality improvement give benefit not only to farmers but also MR. Based on numerical result, we develop proposition to support the benefit of CSR activities both for farmers and MR. It has been showed that farmer will gain benefit from this CSR schema, however for MR there is a CSR budget allocation that gives optimum benefit. The benefit of CSR for MR somehow is reduced when the allocation of CSR budget exceeds this value. The effect of CSR activities is also shown by presenting risk cost per unit. The numerical results show that risk cost per unit is lowered by CSR activities by lengthening the deterioration time of vegetable.

This paper has certain limitations that should be overcome in order to empower farmers in accessing capital and adopting new technology. This proposed model

should be extended for enhance the capabilities of the farmers accessing market and adopting technology. However, further research is required to extend uncertainty factors such as price, demand, and supply.

Acknowledgments The research is supported by the Directorate General of Higher Education (DGHE), Ministry of National Education Republic of Indonesia in HIBAH BERSAING Research Program (DIPA UNS, contract No. 02338/UN27.16/PN/2012). Thanks to the Farmers Groups and the Modern Retail Corporation which have been partners in the research project.

References

1. Hastuti EY (2007) The influence of agribusiness system applied to vegetables farmers income improvement in Boyolali in regency. Master Thesis, Diponegoro University, Indonesia
2. Austin JE (2007) Supporting buyer-supplier relationships. Agency for International Development
3. Arumugam N, Fatimah MA, Chew EFC, Zainalabidin M (2010) Supply chain analysis of fresh fruits and vegetables. *Agric Econ Czech* 56(9):435–442
4. Ahumada OJ, Villalobos JR (2009) Application of planning models in the agri-food supply chain: a review. *Eur J Oper Res* 195(1):1–20
5. Law Number 40 of 2007 the Republic of Indonesia concerning on limited liability companies [online]. <http://www.djpp.depkmham.go.id>
6. Ardana IK (2008) Business and social responsibility. *Buletin Studi Ekonomi* 13(1):32–39
7. Barbarosolu G (2000) An integrated supplier-buyer model for improving supply chain coordination. *Prod Planning Control* 11(8):732–741
8. Esmaeili M, Zeephongsekul P (2010) Seller-buyer models of SCM with asymmetric information structure. *Int J Prod Econ* 123(1):146–154
9. Blanco AM, Masini G, Petracci N, Bandoni JA (2005) Operations management of a packaging plant in the fruit industry. *J Food Eng* 70(1):299–307
10. Sutopo W, Nur Bahagia S (2008) An inventory model for deteriorating commodity under stock dependent selling rate. *Proceedings of the 9th Asia Pacific Industrial Engineering and Management Society* 2008, 3–5 Dec, Bali, Indonesia, pp 1152–1159
11. Narrod C, Roy D, Okello J, Avendaño B, Rich K, Thorat A (2009) Public-private partnerships and collective action in high value fruit and vegetable supply chains. *Food Policy* 34(1):8–15
12. Chaddad F, Fischer FC, Hartmann M (2010) Lessons learned: recommendations for future research on agri-food chain relationships. In: Fischer C, Hartmann M (eds) *Agri-food chain relationships*. CAB International Press, Oxford, pp 267–279
13. Sutopo W, Hisjam M, Yuniaristanto (2011) An agri-food supply chain model for empowering farmers as supplier of modern retailer using corporate social responsibility. *Proceedings of the 6th NIEC 2011*, 3–5 Dec, Surabaya, Indonesia, pp 216–224
14. Sutopo W, Hisjam M, Yuniaristanto (2011) An agri-food supply chain model for cultivating the capabilities of farmers accessing market using corporate social responsibility program. *Proceedings of World Academic Science, Engineering, and Technology (WASET) 2011*, Venice, Italy, pp 2440–2444
15. Sutopo W, Hisjam M, Yuniaristanto (2012) An agri-food supply chain model to empower farmers as supplier for modern retailer using corporate social responsibility activities on deteriorated product. *Proceedings of the international multi conference of engineers and computer scientists, IMECS 2012*, Hong Kong, China, pp 1423–1427
16. Maloni MJ, Brown ME (2006) Corporate social responsibility in the supply chain: an application in the food industry. *J Bus Ethics* 68:35–52

17. Reynolds E (2009) Reset: corporate social responsibility in the global electronics supply chain [Online]. <http://goodelectronics.org>
18. Cruz JM (2008) Dynamics of SC networks with CSR through integrated environmental decision-making. *EJOR* 184(1):1005–1031
19. Cruz JM, Wakolbinger T (2008) Multiperiod effects of CSR on supply chain networks, transaction costs, emissions, and risk. *Int J Prod Econ* 116(1):61–74
20. Mark-Herbert C, von Schantz C (2007) Communicating CSR—brand management. *EJBO* 12(2):4–11
21. Amaeshi K, Osuji O, Nnodim P (2008) CSR in supply chains of global brands: a boundary less responsibility? *J Bus Ethics* 81(1):223–234
22. Hohnen P (2007) Corporate social responsibility an implementation guide for business [Online]. http://www.iisd.org/pdf/2007/csr_guide.pdf
23. Castka P, Balzarova M (2008) ISO 26000 and supply chains—on the diffusion of the social responsibility standard. *Int J Prod Econ* 111(1):274–286
24. Goering GE (2012) Corporate social responsibility and marketing channel coordination. *Res Econ* 66:142–148
25. Rao SS (2009) *Engineering optimization: theory and practice*, 4th edn. Wiley, Hoboken, pp 287–393

Fuzzy Logic Approach to Inventory Lot-Sizing Problem Under Uncertain Environment

Busaba Phruksarphanrat and Thipbodee Tanthatemee

Abstract In this research fuzzy logic approach is applied to solve the problem of inventory lot-sizing under uncertain demand and supply. Conventional stochastic inventory lot-sizing model and existing dynamic models determine only uncertain demand. However, unavailability of supply can happen in many types of manufacturing system. It causes the problems of shortage and overstock. In the proposed Fuzzy Inventory Control (FIC) system, both demand and availability of supply are considered and described by linguistic terms. Then, the developed fuzzy rules are used to extract the fuzzy order quantity and the fuzzy reorder point continuously. The model is more effective than the conventional approach due to adjustment of both order quantity and reorder point. A time step simulation is used to analyze the results of different types of inventory lot-sizing model. Inventory costs of the proposed fuzzy inventory system are compared with existing models based on historical data of a case study factory. It found that FIC system can obtain extremely lower cost than the conventional stochastic and dynamic lot-sizing models.

Keywords Fuzzy logic · Lot-sizing · Uncertainty · Demand and supply · Inventory control · Dynamic model

B. Phruksarphanrat (✉)

ISO-RU, Department of Industrial Engineering, Faculty of Engineering, Thammasat University, Pathumthani 12120, Thailand
e-mail: lbusaba@enr.tu.ac.th

T. Tanthatemee

The Industrial Statistics and Operational Research Unit (ISO-RU), Department of Industrial Engineering, Faculty of Engineering, Thammasat University, Pathumthani 12120, Thailand
e-mail: cheerjuice_junea@hotmail.com

1 Introduction

Inventory decisions can increase both risk and benefit throughout the supply chain [1]. A shortage or not enough supply can disrupt manufacturing plan or inventory management, overstock on inventory level also barrier to management. The most important task of inventory management is making a trade-off between the minimization of the total cost and the maximization of the customer satisfaction. In real cases, these objectives are very difficult to satisfy regarding to the great number of factors involved and unpredictable events such as uncertainties of demand and supply. It is necessary to apply a suitable control system and lot-sizing policy for each type of product [2].

Harris (1913) introduced the Economic Order Quantity (EOQ) for solving a static inventory lot-sizing problem, which assumes deterministic static demand [3]. The objective is to minimize the sum of ordering and inventory holding costs. The stochastic EOQ model was also presented to remedy the problem of uncertain demand [2]. Dynamic lot-sizing model establishes the amount and timing of replenishment of an inventoriable item subject to time-varying demand for realistic applications [4]. It can be solved by simple rules, optimization models and heuristic models such as the Wagner-Whitin (WW) algorithm, Silver-Meal (SM) method, Part Period Balancing (PPB), Lot for Lot (L4L) etc. [4–10]. Nevertheless, supply is assumed to be available. However, in reality supply may not exist when it is needed because of random capacity of suppliers, unpredictable events or seasonal factors. Meanwhile, some of uncertainties within the inventory system cannot be considered appropriately using concepts of probability theory, fuzzy set theory has been used in modeling of inventory systems since 1980s. Fuzzy set theory, originally introduced by Zadeh [11], provides a framework for considering parameters that are vague or unclearly defined or whose values are imprecise or determined based on the subjective beliefs of individuals. Some researches applied fuzzy set and fuzzy number in uncertain demand, order quantity or lead time [6–8]. Roy and Maiti [9] solved the classical EOQ problem with a fuzzy goal and fuzzy inventory costs using a fuzzy non-linear programming method. Uncertain supply was examined by use of Markov chain [12] and a mathematical approach [13]. Both methods determine time to wait before the next order. However, these methods are complicated and difficult for practitioners.

Another approach to simplify a complicated system is to use Fuzzy Logic Control (FLC). FLC has been applied in many applications in industries such as machine control, scheduling and system controls including inventory control [14–16]. Timer and Demirli [17] presented a fuzzy simulation model using FLC of a single item inventory system with variable demand to evaluate the EOQ under uncertain lead time. Dynamic inventory control models under demand, supplier yield and lead time uncertainties have also been presented [2, 10, 13]. However, few of them consider both uncertain quantities of demand and supply. They determine uncertain supply by considering timing to reorder with the same quantity of EOQ. Nevertheless, varying amount of reorder point has not been in concern. In case of unavailability of supply,

increasing of reorder point when the material is obtainable is necessary to protect shortage. The comparison of EOQ and fuzzy logic system under uncertain demand and supply has been proposed [18]. However, the comparison with a dynamic inventory lot-sizing model has not been discussed.

In this research, Fuzzy Inventory Control (FIC) is used to treat the uncertainty regarding demand and supply in inventory lot-sizing problem. MATLAB’s fuzzy toolbox is employed to represent the continuous inventory control system. The demand, supply, order quantity and reorder point are described by linguistic terms. The main objective is to evaluate the order quantity and reorder point in each period taking into account the demand and supply uncertainties. The proposed model is compared with the stochastic EOQ and dynamic models to show its effectiveness.

2 Inventory Lot-Sizing System

The purpose of inventory management is to determine the amount of inventory to keep in stock -how much to order and when to replenish. Inventory should be enough to meet customer demand and also be cost effective [10].

2.1 Static Inventory Lot-Sizing Models

Economic order quantity (EOQ), a static lot-sizing model, is one of the most fundamental of all inventory models. The importance of the model is that it is still one of the most widely used inventory model in industry, and served as a basis for more sophisticated inventory models. EOQ model is extended to the stochastic EOQ model to cure the problem of uncertain demand [2], which is used when the uncertainties are treated as random and are handled by applying probability theory. Assuming that, the demand is represented by a normal distribution. How much to order can be determined from the stochastic EOQ model that can be calculated by the following equation [2].

$$Q^* = EOQ = \sqrt{\frac{2AH\bar{d}(h + s)}{h \times s}}, \tag{1}$$

where

- A* Ordering cost per time.
- h* Holding cost per unit per period.
- s* Shortage cost per unit per period.
- \bar{d} Average weekly demand.
- H* Total length of the planning horizon (number of weeks).

The determinant of when to order in a continuous inventory system is the reorder point, the inventory level at which a new order is placed. If demand is uncertain we must add safety stock into the reorder point, the reorder point and the safety stock can be computed by [19]

$$R = \bar{d}L + SS, \quad (2)$$

$$SS = z\sigma_d\sqrt{L}, \quad (3)$$

where

R Unit of reorder point.

SS Safety stock.

L Lead time.

σ_d The standard deviation of weekly demand.

z The number of standard deviation corresponding to the service level probability.

2.2 Dynamic Inventory Lot-Sizing Models

Dynamic inventory lot-sizing models can be classified into simple rule methods, heuristic methods and optimization methods. In this research a heuristic method, called Silver-Meal (SM) and an optimization method called, Wagner-Whitin (WW) algorithm are used to compare with the proposed method. SM was proposed in 1973 [20]. It tries to achieve the minimum average cost per period for the m period span. The future demand for the next n periods is given as

$$(D_1, D_2, \dots, D_n). \quad (4)$$

Let $K(m)$ be average variable cost per period if the order covers m periods. The general form of $K(m)$ is

$$K(m) = \frac{1}{m}(A + hD_2 + 2hD_3 + \dots + (m-1)hD_m). \quad (5)$$

Compute $K(m)$, $m = 1, 2, \dots, m$, and stop when $K(m+1) > K(m)$ i.e., the period in which the average cost per period starts increasing. We order in the period 1 a quantity to meet the demand of the next m periods, i.e.,

$$Q_1 = D_1 + D_2 + \dots + D_m. \quad (6)$$

WW algorithm [20] is the optimal model for finding the optimal order quantity, Q_i for a dynamic inventory lot-sizing model. The objective is to minimize the variable inventory cost, ordering cost and holding cost over the planning horizon. The optimal order quantities satisfy

$$Q_i = \sum_{k=1}^j D_k \text{ for some } j \geq i. \tag{7}$$

Q_i is the number of units, ordered in a period i to cover demand through period j , with the next order placed at period $j + 1$. Let $K_{t,l}$ be the cost to place an order to cover the demand in period $t, t + 1, \dots, l$, assuming zero inventory at the beginning of period t and zero inventory at the end of period l . Mathematically, this cost is

$$K_{t,l} = A + h \left(\sum_{j=t+1}^l (j - t) D_j \right), t = 1, 2, \dots, n; l = t + 1, t + 2, \dots, n. \tag{8}$$

The equation for the minimum can be found recursively, Let K_l^* denotes this minimum in period l , and it is given by

$$K_l^* = \min_{t=1,2,\dots,l} \{K_{l-1}^* + K_t^*\}, l = 1, 2, \dots, N. \tag{9}$$

K_0^* is defined as zero, and the least-cost solution value is given by K_N^* .

3 Fuzzy Inventory Control Model

In the proposed Fuzzy Inventory Control (FIC) model, there are three components; fuzzy inputs, fuzzy outputs and fuzzy rules. A fuzzy logic toolbox of MATLAB is used to create the FIC model for calculating order quantity and reorder point in any time period. Each element of FIC is shown in fuzzy inference system editor (FIS editor) as shown in Fig. 1. Two fuzzy input variables are demand and availability of supply. Two output variables are ordered quantity and reorder point. These variables are represented by linguistic variables.

3.1 Fuzzy Inputs

Fuzzy inputs are demand and availability of supply, described by membership functions, μ_D and μ_S , respectively. Fuzzy demand and fuzzy availability of supply were defined based on observation and testing of historical data. Both of them are assumed to be represented by 3 linguistic values; low, medium, high as shown in Figs. 2 and 3.

The universe of discourse of demand input space was designed within the interval $[0, \max(D)]$, where $\max(D)$ is the maximum demand that had been ordered. Demand membership functions are based on these parameters $(0, \bar{d} - \sigma_d, \bar{d}, \bar{d} + \sigma_d, \max(D))$ as shown in Fig. 2. The parameters were designed according to the characteristics of a normal distribution of uncertain demand of the factory. Availability of supply was constructed based on real data within the interval $[0, \max(S)]$, where $\max(S)$ is the maximum availability of supply from the current suppliers of determining planning horizon. Membership functions are shown in

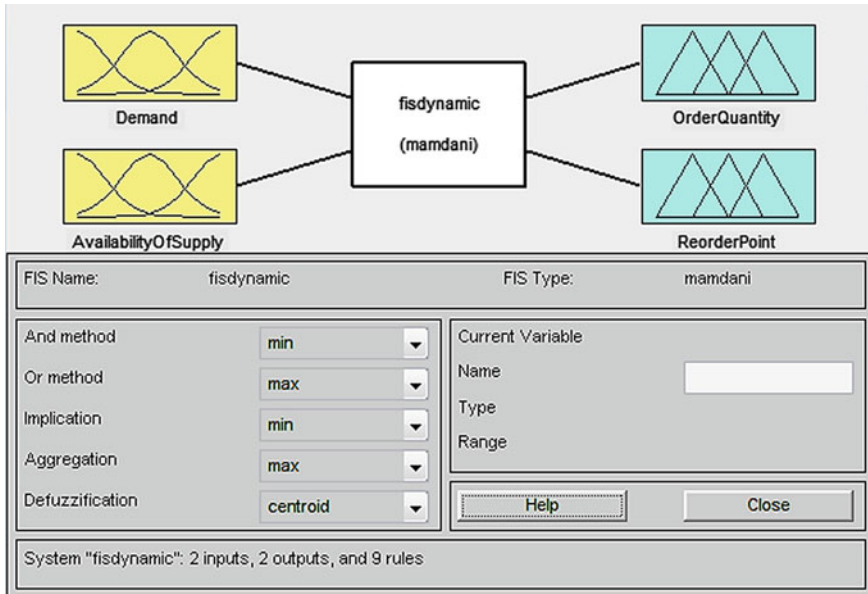


Fig. 1 Fuzzy inventory control model in an inference system editor

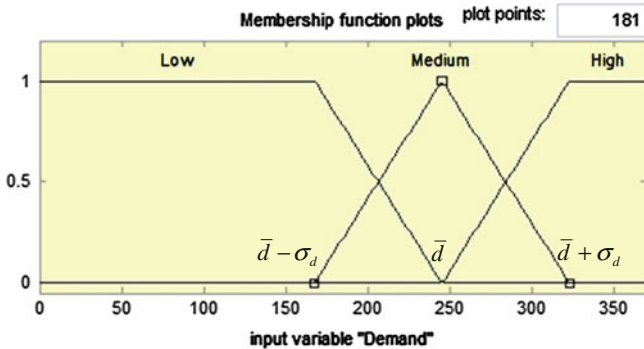


Fig. 2 Demand membership functions

Fig. 3. The parameters $(0, 0.25\max(S), 0.5\max(S), 0.75\max(S), \max(S))$ are used for supply linguistic values according to the normal distribution.

3.2 Fuzzy Outputs

Traditional inventory models use fixed value of order quantity and reorder point. However, in the actual situation, demand and availability of supply are uncertain.

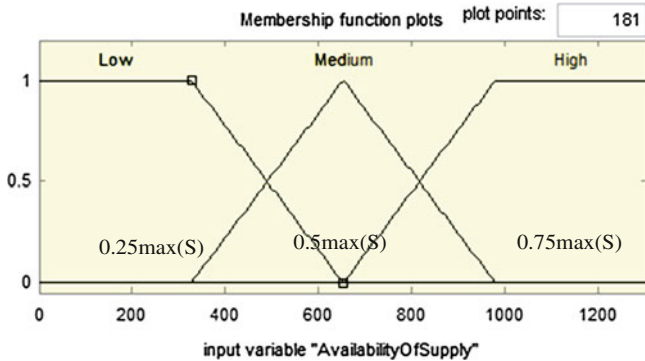


Fig. 3 Availability of supply membership functions

Fixed values of order quantity and reorder point are not appropriate because demands are lumpy and materials may not be available when they are strongly needed. Moreover, the situation of supply may tend to decrease. Safety stock should be carefully established. So, in the proposed model, two fuzzy outputs are constructed. They are fuzzy order quantity and fuzzy reorder point, described by membership functions, μ_Q and μ_R , respectively. Fuzzy order quantity is assumed to be represented by three linguistic values; low, medium, high as shown in Fig. 4. Reorder point is assumed to be represented by five linguistic values; very low, low, medium, high, very high as shown in Fig. 5. Linguistic values of order quantity are selected from availability of supply because in a real situation order quantity should be in the possible range of available supply so the universe of discourse for order quantity output is chosen from the interval $[0, \max(S)]$, where $\max(S)$ is the maximum availability of supply from historical data. The parameters that use to represent three linguistic values are low, medium and high $(0, 0.5\max(S) - R, 0.5\max(S), 0.5\max(S) + R, \max(S))$. The universe of discourse of the reorder point space is the set of real numbers within the interval $[0, 2R]$. Five Linguistic values of reorder point are $(0, R - SS, R, R + SS, 2R)$.

3.3 Fuzzy Rules

Fuzzy inference type of the proposed system is Mandani. The relationship between demand (x_1), availability of supply (x_2) and order quantity (y_1), reorder point (y_2) are described by the following rules:

- R1 IF (x_1 is 'Low') AND (x_2 is 'Low') THEN (y_1 is 'Medium') AND (y_2 is 'High') ELSE
- R2 IF (x_1 is 'Low') AND (x_2 is 'Medium') THEN (y_1 is 'Low') AND (y_2 is 'Medium') ELSE
- R3 IF (x_1 is 'Low') AND (x_2 is 'High') THEN (y_1 is 'Medium') AND (y_2 is 'Low') ELSE

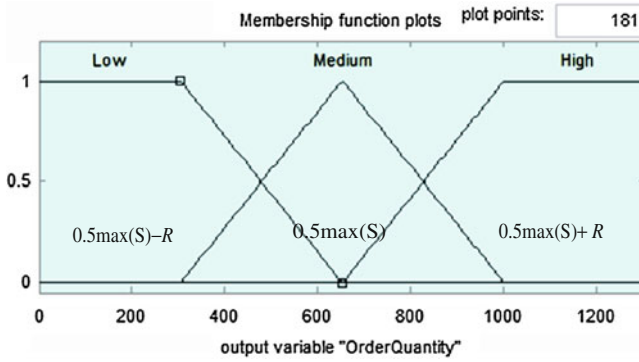


Fig. 4 Order quantity membership functions

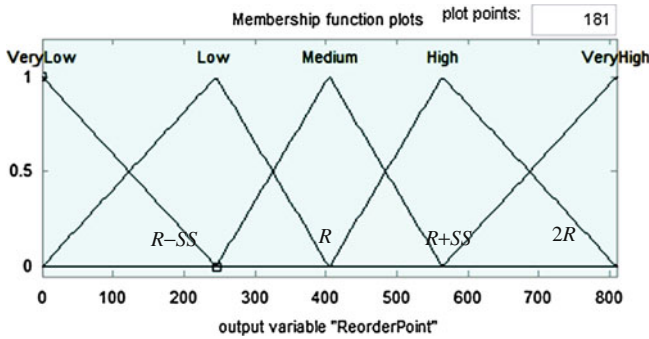


Fig. 5 Reorder point membership functions

- R4 IF (x_1 is 'Medium') AND (x_2 is 'Low') THEN (y_1 is 'Low') AND (y_2 is 'Very High') ELSE
- R5 IF (x_1 is 'Medium') AND (x_2 is 'Medium') THEN (y_1 is 'Medium') AND (y_2 is 'High') ELSE
- R6 IF (x_1 is 'Medium') AND (x_3 is 'High') THEN (y_1 is 'High') AND (y_2 is 'High') ELSE
- R7 IF (x_1 is 'High') AND (x_2 is 'Low') THEN (y_1 is 'Medium') AND (y_2 is 'Very High') ELSE
- R8 IF (x_1 is 'High') AND (x_2 is 'Medium') THEN (y_1 is 'High') AND (y_2 is 'High') ELSE
- R9 IF (x_1 is 'High') AND (x_2 is 'High') THEN (y_1 is 'High') AND (y_2 is 'High')

By taking the max–min compositional operation, the fuzzy reasoning of these rules yields fuzzy outputs. These outputs can be expressed as

$$\mu_{Q_0}(y_1) = (\mu_{D_1}(x_1) \wedge \mu_{S_1}(x_2)) \vee ..(\mu_{D_n}(x_1) \wedge \mu_{S_n}(x_2)), \tag{10}$$

$$\mu_{R_0}(y_2) = (\mu_{D_1}(x_1) \wedge \mu_{S_1}(x_2)) \vee ..(\mu_{D_n}(x_1) \wedge \mu_{S_n}(x_2)), \tag{11}$$

where \wedge is the minimum operation and \vee is the maximum operation. D_i , S_i , Q_i , and R_i are fuzzy subsets defined by the corresponding membership functions, i.e., μ_{D_i} , μ_{S_i} , μ_{Q_i} , μ_{R_i}

Finally, a defuzzification method, called the center-of-gravity method [12], is adopted here to transform the fuzzy inference output into a non-fuzzy value order quantity, y_{01} and non-fuzzy reorder point, y_{02} .

$$y_{01} = \frac{\sum y_1(\mu_{Q_0}(y_1))}{\sum \mu_{Q_0}(y_1)}, \tag{12}$$

$$y_{02} = \frac{\sum y_2(\mu_{R_0}(y_2))}{\sum \mu_{R_0}(y_2)}. \tag{13}$$

These operations can be done in the FIC model, which is implemented in Fuzzy Logic Tool Box of MATLAB.

4 Numerical Example

Historical inventory data of a furniture company have been investigated. The company is a Made-to-Order manufacturer. It is facing the problem of inventory control because both demand and supply are uncertain. The main materials of the company are woods, which are uncertain because the quantity of woods depends on environment, rainfall and sources of supply. Demand is uncertain and randomly fluctuated. Both of demand and availability of supply quantities can be represented by a normal distribution. Currently the high inventory level is used to protect shortage. Service level that the company wants to guarantee to customers is more than 97 %. However, shortage is still raised and the total inventory cost is also high.

So, the FIC model is proposed to reduce inventory levels and the total inventory cost. It is compared among the conventional stochastic EOQ model, the SM model and the WW model. Three levels of service level (97, 99 and 99.9 %) for the stochastic model, the SM model and the WW model are tested based on 97 % service level of the proposed FIC model. Five data sets that generate from the distribution of historical data were examined. The distribution of historical demand is the normal with mean 2,452 units and standard deviation 776 units. Availability of supply is also uncertain and can be represented by the normal distribution with mean 6,486 units and standard deviation 3,920 units. From Fig. 6, it can be seen that the availability of materials is extremely fluctuated comparing with demand, which may cause unavailability in some period. Holding cost and shortage per unit per period of the case study factory are 0.05 and 59 Baht, respectively.

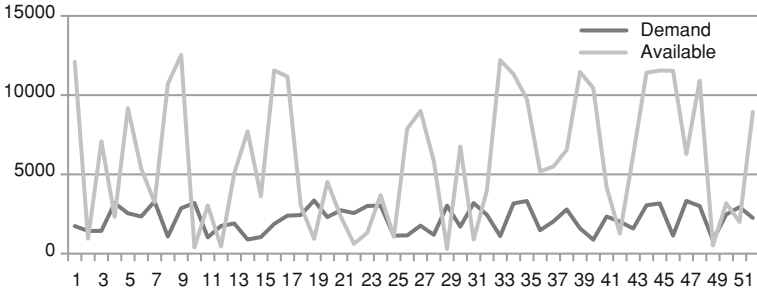


Fig. 6 Fluctuation of demand and supply in 52 weeks

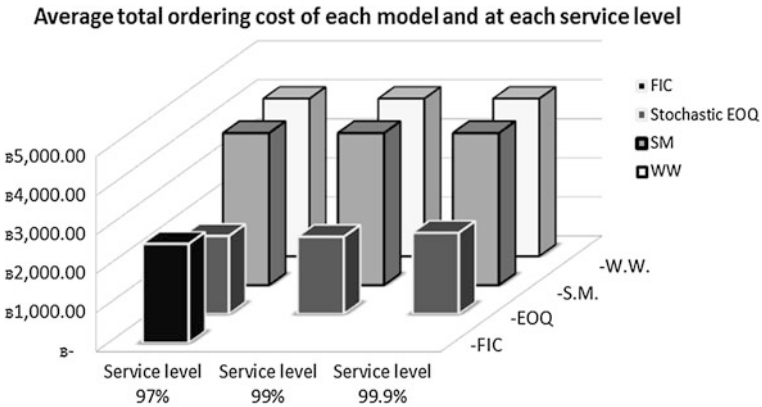


Fig. 7 A comparison of the average total ordering cost of 5 data sets of different models at service level 97, 99, 99.9 %

Order quantity, reorder point and safety stock of the stochastic EOQ model can be calculated using Eqs. (1)–(3). For the SM model and the WW model, Eqs. (4)–(9) are used. Then, the average holding cost, the average ordering cost, the shortage cost and the total cost of 52 weeks of all data sets can be generated and compared. They are represented in Figs. 7, 8, 9, and 10.

Figure 7 shown that the average ordering cost of the stochastic models for all service levels have the lowest cost due to less frequency of orders. Figure 8 shown that the average holding cost of the FIC model has the highest cost, whereas dynamic lot-sizing models have low average holding costs. It means that the FIC model has the highest average inventory level and dynamic lot-sizing models have low inventory levels. Figure 9 shown that the average shortage cost of the FIC model is the lowest because inventory level of the FIC model is high. The FIC model considers an availability of supply, so lacking of supply does not occur. Stochastic EOQ model, dynamic models (SM and WW) cannot handle the extreme fluctuated of availability of supply even higher service levels are considered so shortage cost is raised, although the FIC model has no shortage cost. This situation

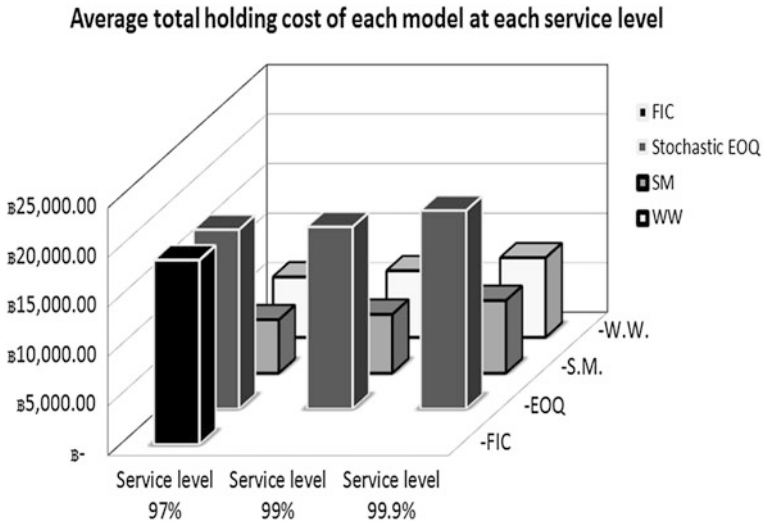


Fig. 8 A comparison of the average total holding cost of 5 data sets of different models at service level 97, 99, 99.9 %

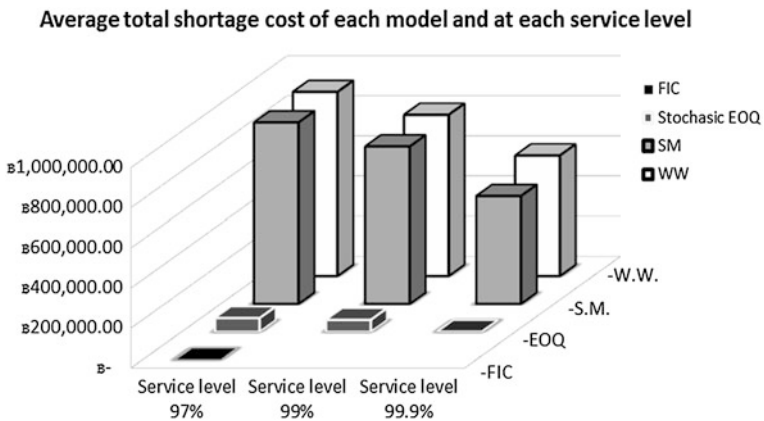


Fig. 9 A comparison of the average total shortage cost of 5 data sets of different models at service level 97, 99, 99.9 %

occurs because in the stochastic EOQ model and the dynamic models, availability of supply is assumed to be certain so it cannot react effectively when supply is unavailable. While the FIC model can control both reorder point and order quantity to the appropriate level when they are needed. So, FIC model is the most effective model when demand and supply are uncertain. Finally, the FIC model has the lowest total inventory cost as shown in Fig. 10.

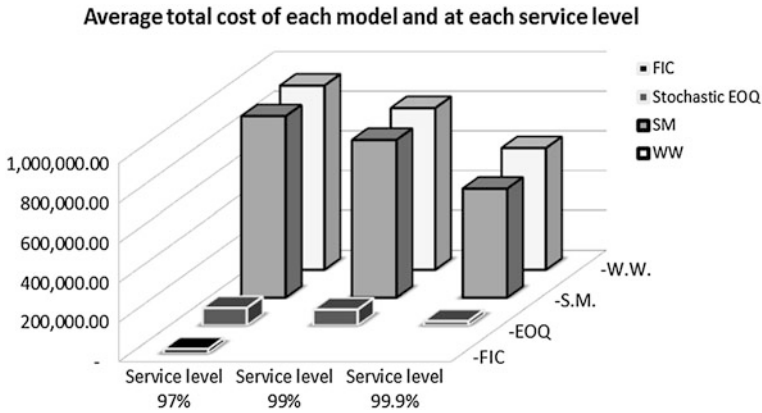


Fig. 10 A comparison of the average total cost of 5 data sets of different models at service level 97, 99, 99.9 %

Table 1 An average cost saving of the FIC model compared with the stochastic EOQ model at service level 97, 99, 99.9 % respectively

Data sets	EOQ 97 % (Baht)	EOQ 99 % (Baht)	EOQ 99.9 % (Baht)
1	45,634 (53 %)	66,010 (67 %)	22,301 (3 %)
2	71,142 (71 %)	71,142 (71 %)	23,517 (13 %)
3	144,175 (85 %)	144,175 (85 %)	21,524 (1 %)
4	56,207 (60 %)	56,692 (61 %)	22,989 (3 %)
5	123,828 (84 %)	60,899 (67 %)	20,105 (0.02 %)
Average	88,197 (76 %)	79,783 (73 %)	22,105 (4 %)

A cost saving of the FIC model is also shown in Tables 1, 2, and 3. These results confirm that the FIC model is better than the stochastic model, the SM model and the WW model when demand and supply are uncertain.

The FIC model can dramatically save inventory cost when compared with stochastic EOQ model, SM model and WW model at all service levels. If the service level is increased, the saving cost can increase because shortage is reduced due to higher safety stock. The higher the order quantity and reorder point, the better cost saving for the case study. The FIC model can save more money than the other models as shown in the results.

5 Conclusion

Fuzzy Inventory Control (FIC) model based on continuous inventory control system is proposed in this research for solving inventory lot-sizing problem under uncertain environment. Demand and availability of supply are inputs and order

Table 2 An average cost saving of the FIC model compared with the Silver-Meal (SM) model at service level 97, 99, 99.9 % respectively

Data sets	SM 97 % (Baht)	SM 99 % (Baht)	SM 99.9 % (Baht)
1	901,121 (98 %)	778,450 (97 %)	560,364 (96 %)
2	1,632,925 (99 %)	1,461,484 (99 %)	1,107,411 (98 %)
3	425,860 (95 %)	317,824 (93 %)	113,547 (81 %)
4	429,100 (95 %)	370,752 (94 %)	157,556 (86 %)
5	1,179,655 (98 %)	1,041,605 (98 %)	798,269 (97 %)
Average	913,732 (98 %)	794,023(97 %)	547,429 (96 %)

Table 3 An average cost saving of the FIC model compared with the Wagner-Whitin (WW) model at service level 97, 99, 99.9 % respectively

Data sets	WW 97 % (Baht)	WW 99 % (Baht)	WW 99.9 % (Baht)
1	785,788 (97 %)	688,154 (97 %)	533,067 (96 %)
2	1,633,195 (99 %)	1,461,764 (99 %)	1,148,972 (98 %)
3	525,276 (96 %)	417,779 (95 %)	214,352 (90 %)
4	146,767 (85 %)	98,403 (75 %)	32,641 (32 %)
5	1,548,352 (99 %)	1,404,423 (99 %)	1,130,367 (98 %)
Average	927,876 (98 %)	814,105 (97 %)	611,880 (97 %)

quantity and reorder point are outputs of the system. Linguistic values are used for both fuzzy inputs and outputs. Fuzzy Rules are constructed according to the historical experience. Five data sets generated from the distribution of the demand and supply of the case study factory are used to evaluate the FIC model comparing with the stochastic EOQ model, the SM model and the WW model at different service levels. The obtained results clearly show that FIC model is the most effective model when demand and supply are uncertain. The FIC model can extremely save the total inventory cost. Moreover, there is no shortage in any sets of test data when the FIC model is applied to control the inventory system. It means that the FIC model can provide customer satisfaction although demand and availability of supply are uncertain. The FIC model is more flexible than the other models because both order quantity and reorder point are reevaluated continuously. So, efficient solution can be obtained.

The FIC model allows a user to modify or readjust parameters when the situation changes. So, it is possible to extend learning methods for adjusting parameters to the FIC model.

Acknowledgments This work was supported by the Commission on Higher Education of Thailand and the Faculty of Engineering, Thammasat University, Thailand.

References

1. Bowersox DJ, Closs DJ, Cooper MB (2007) Supply chain logistics management. 2nd edn, International Edition, McGraw-Hill, New York
2. Kamal L, Sculfort J-L (2007) Fuzzy modeling of inventory control system in uncertain environment. In: International symposium on logistics and industrial informatics, 2007, pp 53–57
3. Robinson R, Narayanan A, Sahin F (2009) Coordinated deterministic dynamic lot-sizing problem: a review of models and algorithms. *Omega* 37:3–15
4. Simpson NC (2001) Questioning the relative virtues of dynamic lot sizing rules. *Comput Oper Res* 28:899–914
5. Rotshtein AP, Rakityanskaya AB (2006) Inventory control as an identification problem based on fuzzy logic. *Cybernet Syst Anal* 42(3):411–419
6. Park KS (1987) Fuzzy set theoretic interpretation of economic order quantity. *IEEE Trans Syst Man and Cybernet* 6 (17):1082–1084
7. Chen SH, Wang CC, Ramer A (1996) Backorder fuzzy inventory model under function principle. *Inf Sci* 95:71–79
8. Petrovic D, Petrovic R, Mirko V (1996) Fuzzy models for the newsboy problem. *Int J Prod Econ* 45:435–441
9. Roy TK, Maiti M (1997) A fuzzy EOQ model with demand dependent unit cost under limited storage capacity. *Eur J Oper Res* 99:425–432
10. Babai MZ, Dallery Y (2006) A dynamic inventory control policy under demand, yield and lead time uncertainties. In: Conf. Rec. 2006 IEEE international conference on service system and service management, pp 1439–1444
11. Zadeh LA (1965) Fuzzy sets. *Inf Control* 8:338–353
12. Roy TK, Maiti M (1998) Multi-objective inventory models of deteriorating items with some constraints in a fuzzy environment. *Comput Oper Res* 25(12):1085–1095
13. Parlar M, Perry D (1995) Analysis of a (Q,r,T) inventory policy with deterministic and random yields when future supply is uncertain. *Eur J Oper Res* 84:431–443
14. Wang C-H (2010) Some remarks on an optimal order quantity and reorder point when supply and demand are uncertain. *Comput Ind Eng* 58:809–813
15. Lin JL, Wang KS, Yan BH, Tang YS (2000) Optimization of the electrical discharge machining process based on the Taguchi method with fuzzy logics. *J Mater Proc Technol* 102:48–55
16. Petrovic D, Duenas A (2006) A fuzzy logic based production scheduling/rescheduling in the presence of uncertain disruptions. *Fuzzy Set Syst* 157:2273–2285
17. Yimer AD, Demirli K (2004) Fuzzy modeling and simulation of single item inventory system with variable demand. IEEE annual meeting of the North American Fuzzy Information Processing Society Banff, 2, Alberta, Canada, pp 985–989, 2004
18. Tanthatemee T, Phruksarphanrat B (2012) Fuzzy inventory control system for uncertain demand and supply. In: Proceedings of the international multicongress of engineers and computer scientists 2012, IMECS 2012, Hong Kong, 14–16 March 2012. Lecture notes in engineering and computer science, pp 1224–1229
19. Russell S, Taylor III W (2006) Operations management quality and competitiveness in a global environment. 5th edn, Wiley, New Jersey, pp 527–554
20. Sipper D, Bulfin RL (1998) Production planning, control and integration, McGraw-Hill International, Singapore

A Semantic Wiki to Support Knowledge Sharing in Innovation Activities

Inaya Lahoud, Davy Monticolo, Vincent Hilaire
and Samuel Gomes

Abstract We will present in this paper how to ensure the creation, the validation and the sharing of ideas by using a Semantic Wiki approach. We describe the system called Wiki-I which is used by engineers to allow them to formalize their ideas during the research solutions activities. Wiki-I is based on an ontology of the innovation domain which allows to structure the wiki pages and to store the knowledge posted by the engineers. In this paper, we will explain how Wiki-I ensures the reliability of the innovative ideas thanks to an idea of evaluation process. After explaining the interest of the use of semantic wikis in innovation management approach, we describe the architecture of Wiki-I with its semantic functionalities. At the end of the paper, we prove the effectiveness of Wiki-I with an ideas evaluation example in the case of students challenge for innovation.

Keywords Semantic wiki · Innovation · Ontology · Knowledge creation · Knowledge evaluation and sharing · SPARQL

I. Lahoud (✉) · V. Hilaire
SET Laboratory, University of Technology of Belfort-Montbeliard, Belfort, France
e-mail: Inaya.Lahoud@utbm.fr

V. Hilaire
e-mail: Vincent.Hilaire@utbm.fr

D. Monticolo
ERPI Laboratory, Polytechnic National Institute of Lorraine, Nancy, France
e-mail: Davy.Monticolo@ensgsi.inpl-nancy.fr

S. Gomes
M3M Laboratory, University of Technology of Belfort-Montbeliard, Belfort, France
e-mail: Samuel.Gomes@utbm.fr

1 Introduction

There are many definitions of the innovation. “Innovations are qualitatively new products or processes, which distinguish themselves significantly from previous ones” [9]. Innovation concerns R&D, product and service development. Armbruster in [1] distinguishes four different types of innovation: technical product innovations, non-technical service innovations, technical process innovations, and non-technical process innovations, these different innovations are organizational innovations.

Innovation is an approach to create value by solving problems. Creativity is a prerequisite for problem-solving and can be applied several times during a product development process. But creativity in an innovation process is not enough to develop a new innovating product or service [8]. The results of the creativity need to be stored [6] in order to ensure that the solution will use them in the future [5, 29]. In this paper we propose to organize, annotate and store the ideas resulting from the creativities activities by using a semantic Wiki.

Moreover in an innovation process, knowledge about the definition of the new product (problem), about the market, about the new technologies or the industrial processes and about the evolution of our cultures have to be managed to make innovation a learning and prioritization process [16]. [15] explains that the knowledge management is a key factor of the innovation process. In this paper we will describe how the Semantic Wiki using an innovation ontology [19] allow to develop the creativity of the engineers through idea cards and allows to build an innovative ideas base. This article is structured as follows: [Sect. 2](#) introduces the Semantic Wiki concept and describes its advantages for the innovation process; [Sect. 3](#) briefly describes the architecture of our Wiki; [Sect. 4](#) presents a simple application scenario to exploit knowledge and to represent it; [Sect. 5](#) presents the functionalities of the wiki; [Sect. 6](#) shows a study case which indicate the performance assessment of our wiki and finally [Sect. 7](#) concludes with some perspectives.

2 Semantic Wiki for Innovation Process

A wiki is a web site that allows collaborative distant creation of information and editing of hypertext content. Leuf and Cunningham [14] were the first to propose a web site where people could create, modify, transform and link pages from within their browser and in a very simple way. Wikis have become popular tools for collaboration on the web [17], and many active online communities employ wikis to publish or exchange information.

For most wikis, public or private, the primary goals are to organize the collected information and to share it. Wikis are usually viewed as tools to manage online

content in a quick and easy way, by editing some simple syntax known as wikitext [25]. Schaffert [23] enumerates the specifications of a wiki system:

- It allows the editing via a browser;
- It has a simplified wiki syntax i.e. simplified hypertext format usable by all the internet users;
- It manages a rollback mechanism i.e. it is able to versioned the changes in the content each time they are stored;
- Its access is unrestricted, everybody can write in the wiki;
- It manages the collaborative editing i.e. if someone create a article, everybody can extend this article;
- It proposes a strong linking, all the pages of the wiki are linked with each other using hyperlinks;
- It has a search function over the content of all pages stored;

It allows the uploading of different content like documents, images or videos. Taking into consideration all these properties, Wikis seems as a good candidates for collaborative knowledge engineering based on Web2.0 social networks [21]. Indeed new research works [23, 28] propose wikis to exchange knowledge. Knowledge can be seen as information with an added context and value that make it usable within that very context.

Knowledge can also be seen as what places someone in the position to perform a particular task by selecting, interpreting and evaluating information depending on the context [18, 27].

However a serious obstacle for the development of Semantic Web applications is the lack of formal ontologies and knowledge. Indeed, one of the main reasons of this is the rather high technical barrier for using Semantic Web [3] technologies that deters many domain experts from formalizing their knowledge.

On the other hand, wiki systems are becoming more and more popular as tools for content and information management. Much information is nowadays available in systems like Wikipedia [12, 13]. Unfortunately, this vast information is not accessible for machines. If a small amount of this information could be formalized to become knowledge, then, wiki systems could provide improved interfaces and advanced searching and navigation facilities.

Nevertheless, several analyses [4, 28] of traditional wikis as shown that they are not enough structured, and it's difficult to navigate and to find the relevant information. Besides, the wiki markup language (WikiML) used by most wiki engines makes internet users reluctant to contribute to the wiki. One solution to perform the ideas creation, evaluation and navigation inside wikis is to use technologies from the Semantic Web [2] to formalized information, content, structures and links in the wiki pages. These Wikis would take consideration of the semantic in their content management and become Semantic Wikis [7, 26]. "Semantic Wiki" systems aim to combine "traditional" wiki systems with Semantic Technology. "A semantic wiki is a system that allows collaborative authoring, editing and linking of pages, but also the authoring and adding semantics to the data on the wiki itself" [11].

The significance of semantic wiki is that it may contain machine-readable content and structure, which will improve the possibilities to browse, query, share and reuse the knowledge. Völkel [27] list a number of requirements for semantic wikis: usability, expressiveness, flexibility, scalability, interchange and compatibility.

The semantic wiki bears much potential in many application areas as Schaffert [23] lists a number of scenarios for semantic wiki:

- **Ontology engineering**, in which domain experts and knowledge engineers working together can be supported by a semantic wiki. The experts will have an easy way to enter knowledge, they can all collaborate on working on the ontology and the knowledge can be evolutionary formalized.
- **Knowledge management systems** must combine content creation and easy authoring with structuring knowledge in order to retrieve related information. Traditional wikis may be used for the first aspect, while semantic wikis adds support for the latter.
- **Educational environments** can be supported by semantic wiki by acting as a “learning content pool” where content can be created, edited and related to relevant learning material. Further, the structured and interlinked content in a semantic wiki can support self-directed learning.

In our work, we are more in the case of ontology engineering. Thus we propose to use a Semantic Innovation Wiki approach to propose a system, based on innovation ontology, to facilitate the creativity and to formalize ideas in the innovation activities by facilitating the knowledge sharing, updating and evaluation.

3 Wiki-I Architecture

In this section we detail the architecture of Wiki-I [20]. This architecture is defined by three layers (Fig. 1): Web Layer, Knowledge Persistent Layer and the Knowledge Base Layer. Each layer communicates with the others through a RDF flow in order to ease the knowledge diffusion.

3.1 *The Knowledge Base Layer*

The semantic Wiki is considered as a knowledge base. It is composed by the creation of wiki pages.

The Knowledge base is built with the RDF language i.e. it is formed by annotations describing the innovation ideas, the context where was captured the idea and their authors.

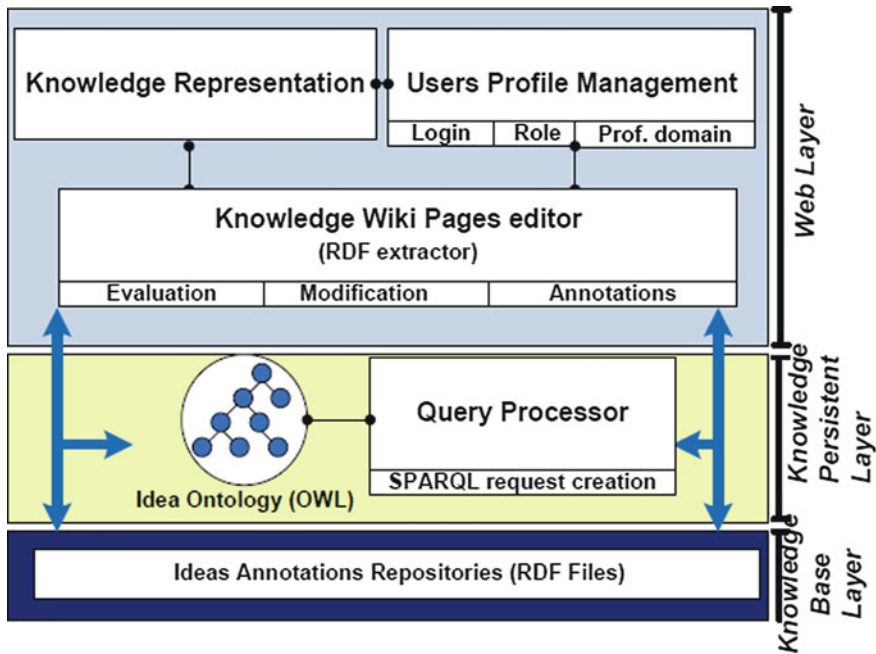


Fig. 1 Wiki-I architecture

3.2 The Knowledge Persistent Layer

The Knowledge Persistent Layer is based on the Idea ontology proposed by Riedl [22]. This ontology defines a vocabulary and a semantic of the knowledge used in innovation ideas. The Idea ontology is developed in OWL-DL.

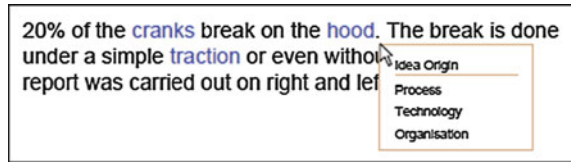
This language is based on Description Logics (hence the suffix DL). Description Logics are a decidable fragment of First Order Logic and are therefore amenable to automated reasoning. It is therefore possible to automatically compute the classification hierarchy and check for inconsistencies in an ontology that conforms to OWL-DL.

Consequently, the Idea Ontology provides an integrated conceptual model for sharing information related to an innovative idea. An OWL property is a binary relation that relates an OWL Class to another one, or to RDF literals or XML Schema datatypes.

For example, the “infoInput” property relates the Document class to the Activity class. Described by these formal, explicit and rich semantics, the domain concept of Activity, its properties and relationships with other concepts can be queried, reasoned or mapped to support the ideas sharing across the Semantic Wiki.

The Knowledge Persistent Layer is also composed by a Query Processor which allows formulating queries to exploit the knowledge based according to the

Fig. 2 Example of knowledge links from a term in Wiki-I



structure of the ontology. The Query Processor builds queries with the SPARQL language [24] in order to exploit the RDF files that compose the knowledge base.

3.3 The Web Layer

Wiki-I allows the use of semantic tags and navigating functionalities in the wikipages. We have seen that the Knowledge Persistent Layer is composed by the Idea ontology which defines a vocabulary and a semantic of the ideas created during the research activities. Thanks to the relations in the ontology, Wiki-I is able to automatically tag keywords in the wikipages. These tags provide to the users, not only a link to wikipages defining the term associated to the tag but also a links to the types of ideas origins associated to this term. The Fig. 2 shows three ideas origin links (Process, Technology, etc.) related to the term “hood”.

In addition Wiki-I has an Idea Wiki Pages editor. The editor proposes a structure of the article according to the concepts and sub concepts of the ontology. For example an article describing an idea is organizing with the tags “Idea Origin”, “What the Idea is useful for”, “Individuals or Structure involved in the development”, “Advantages”, “Limitation”, “Impact”, etc.

The Web Layer has also Users Profile Management module where users can create and refine their profile. To be a creator, user has to create a new profile. In this profile they can define the different roles they have in a project or their professional domains. According to this information, Wiki-I proposes to a user, when he is connected, a selection of wikipages created by other users and related to his profile.

3.4 Interface

Wiki-I uses a browser-based interface. A search page view is shown in Fig. 3. From keywords the users request the Innovative ideas base. The list of articles (wikipages) is generated in the same page. Each result corresponds to a wikipage and has an evaluation according to its maturity (number of stars describing the number of evaluations) and its percent of positive evaluation.

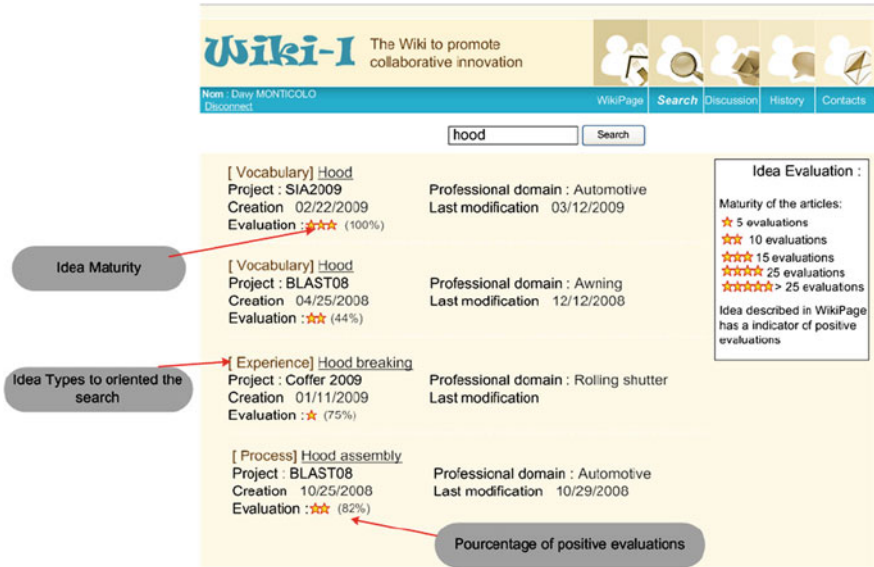


Fig. 3 Wiki-I interface

3.5 Collaborative Knowledge Evaluation

Inside Wiki-I, knowledge is subject to an evaluation process. This evaluation is done by the professional actors. An actor can modify or accept an article i.e. knowledge related to the wikipage (Fig. 4). When a user approves or modifies an article [10], he assigns a positive evaluation for this article.

Moreover, when he refuses, the article obtains a negative evaluation. Wiki-I allows to compute an indicator of knowledge maturity by positioning a percentage of positive evaluation and a number of stars.

Thus, knowledge which has just been created has one hundred percent of positive evaluation. Progressively, with each evaluation attributed by a user, the percentage can increase (resp. decrease) if the article obtains a positive (resp. negative) evaluation.

In addition, ideas that have a score lower than twenty-five percent of positive evaluation are deleted of the innovative idea base. The system is thus able to automatically delete ideas that have become obsolete or that do not reach a consensus inside the community of experts.

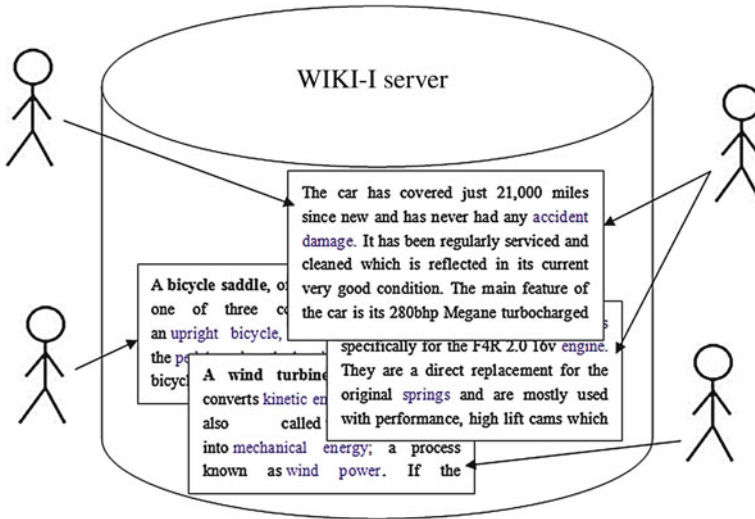


Fig. 4 A shared collection of articles that an actor can modify or accept

4 Exploiting Semantics and Knowledge Representation

4.1 Browsing

There are two types of Wikis users; the readers and the creators. The first one use the elements stored in the knowledge base to search pertinent information and the second one creates new wikipages.

The readers have access to the ideas. A reader uses a keyword to apply a research. The navigation in Wiki-I is made by clicking an idea which will take along the reader to the wikipage related to this idea. The other way to navigate in Wiki-I is to use a right click on a term which presents the knowledge types related to the term. Each associated idea leads to one or several wikipages. For example in the Fig. 3, we have an idea type “technology”, “organization” and “process” related to the term “Hood” and this type leads to four wikipages describing four different ideas implying a hood.

4.2 Querying

As shown in Fig. 1, Wiki-I has a semantic search engine for querying and reasoning on the knowledge base. This query processor used the Jena API. Jena allows loading ontological models in OWL or RDFS format and manages the SPARQL language. SPARQL is based on a boolean combination of triples that can be constrained by evaluable expressions. It is also processes datatyped RDF


```

PREFIX OntoDesign:http://www.ensgsi.fr/DM/OntoIdea.owl
SELECT xml ?Technology, ?Process, ?Organisation
WHERE
{
  ?Idea rdf:type OntoIdea:Name "Hood"
    Union
    ?IdeaOrigin rdf:type OntoIdea:origin "Technology"
}

```

Fig. 5 Example of request generated by the knowledge persistent layer

literals, optional properties, alternatives and the named graph scheme of SPARQL using a source statement. It returns an RDF/XML graph or an XML binding format. The bindings are available through an API. SPARQL provides the select, distinct, sort and an equivalent of limit statements.

The Knowledge Persistent Layer module allows building queries according to the keywords posted by the wiki readers. The readers can orient his requests on the ideas stored in the innovative ideas base. The Fig. 5 describes a classical query to search ideas associated to the keyword “Hood”.

The readers have the possibility to refine their requests according to the ideas and the origin of the ideas.

5 Wiki-I Functionalities

All the knowledge inside the wikipages of Wiki-I is annotated in RDF according to the idea ontology. Thus the ontology defines the inherent structure of the wiki. Moreover, the annotations facilitate the navigation between wikipages thanks to the links defined in the ontology. We describe in this section the advantages of Wiki-I.

5.1 Typing/Annotating of links

Wiki-I allows the annotation of links by giving them certain types defined in the idea ontology. Thus a link created by a user almost always carries meaning beyond mere navigation. Wiki-I manages annotations in its Web Layer. Each WikiPage is annotated as soon as a user (creator) defines the content related to an idea type.

5.2 *Context-Aware Presentation*

Wiki-I can change the way content is presented based on semantic annotations. This includes enriching pages by displaying semantically related pages in a separate link box, displaying informations that can be derived from the underlying knowledge base. Thus, a wikipage defining an idea related to a new technology is automatically associated to others wikipages corresponding to others ideas related to this technological innovation.

5.3 *Enhanced Navigation*

Ideas types facilitate annotated links and provide more information for navigation. Whereas a traditional wiki only allows following a link, Wiki-I offers additional information about the relation the link describes.

For example Wiki-I propose to the creator of the wikipages to define the semantic links with the relation defined in the ontology. For example a wikipage about an assembly process of a hood can have some links categorised by “has synonymous”, “has impact”, etc.

Such informations can be used to offer additional or more efficient navigations.

5.4 *Semantic Search*

Wiki-I allows a “semantic search” on the underlying RDF knowledge base. As described above, queries are expressed in the language SPARQL, a query language recently proposed as W3C recommendation for RDF querying. Using “semantic search”, users can ask queries like “retrieve all process innovation concerning a hood”.

The result of the query is a group of projects listed by the most important to the less important for the user. We have a specific classification approach to display projects (Fig. 6). In this approach we use the role of the user to list the projects by the most important to him. We apply also different type of filters on the list by using (1) the domains of interests of the user defined in his profile, (2) the projects he consults, (3) the projects he creates, and (4) the history of his searches in the system.

The Fig. 6 show an example of this mechanism where the user “Davy” has as role “designer” and as domains of interests “industry, mechanical, maintenance, aeronautical, nuclear power, hydraulic, automotive”; thus when he searches in the system the word “hood” for example the system of classification will sort the results by the mechanical domain because it appears in the center of interest of

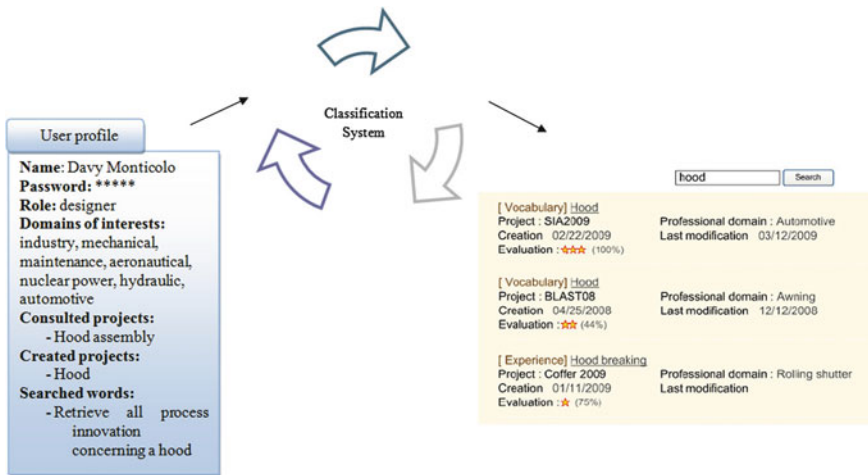


Fig. 6 Classification approach of projects resulting from query

the user and it will disadvantage the results of “hood” in the domains of agriculture, army, building, medicine, security, sports, etc.

6 Performance Assessment

6.1 Adhesion of the students

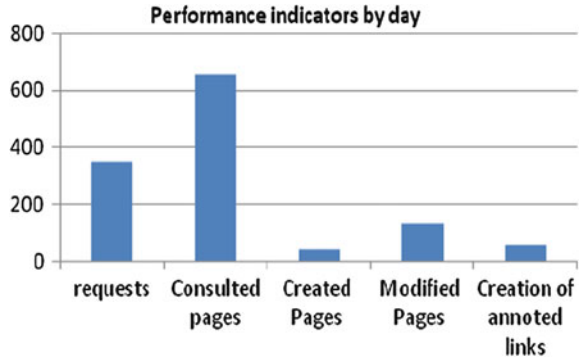
Since 2008, we organize in our university the innovation challenge “48 h to create Ideas”. In this challenge 500 students all over the world have to generate ideas to response to an industrial question like “what will be the glasses of the future” or “how to link generations with mobile phone”. After generating ideas the students use Wiki-I to store and to share their ideas.

Interviews of the students emphasize the two following advantages of the system Wiki-I:

- The capability to capitalize ideas and to annotate them in semi-automatic way and to make them reusable it in an easy way (thanks to a research based on the ideas types) inside Wiki-I;
- The possibility to make evolve Ideas in collaborative way, inside the wiki, by interacting with all the others students.

In the case of the innovation challenge in 2011 (Fig. 7), we have noted that during the 48 h, 876 ideas (articles) were automatically created in Wiki-I by the software agents. Among these 876 pages, 430 were evaluated and/or modified to enrich the root idea by students or professional actors.

Fig. 7 Performance indicators for Wiki-I



7 Perspectives and Conclusion

In this article, we have presented Wiki-I, a feature-rich semantic wiki allowing to create and evaluate new innovative ideas with pertinent links.

Wiki-I is currently used in several universities to support innovation challenge. The students, professors and professional actors use Wiki-I to share innovative ideas. Wiki-I seems to provide a good framework to evaluate and to create new Innovative ideas in an easy way.

We saw that nowadays semantic wiki helps users to organize and browse knowledge as well as allowing machine-to-machine integration.

Wiki-I is especially useful among groups of users or community which work collaboratively to create new ideas for innovation projects and maintain these ideas. It also offers a user friendly way to build and maintain ontologies and vocabularies.

Now, future directions for the evolution of Wiki-I might be to provide more support to Idea creation by using several ontologies. The system has to provide support for inferences or ontology import i.e. it has to allow users to import data from external ontologies and exploits schema data to provide editing support. Thus the system will be able to understand the knowledge about a new innovative product, process or technology and not only an idea.

We are currently working on how to merge several domain ontologies to propose more possibilities to create semantic links or to perform the semantic searches.

References

1. Armbrustera H, Bikfalvib A, Kinkela S, Laya G (2008) Organizational innovation: the challenge of measuring non-technical innovation in large-scale surveys. *The Int J Technovation* 28:644–657
2. Aumueller D, Auer S (2005) Towards a semantic wiki experience-desktop integration and interactivity in WikSAR. *Proceedings of the workshop on semantic desktop, Galway, Ireland, 2005*

3. Berners-Lee T, Hendler J, Lassila O (2001) The semantic web. *Scientific American*, pp 35–43, May 2001
4. Buffa M, Gandon F, Ereteo G, Sander P, Faron C (2008) SweetWiki: a semantic wiki. *J Web Semant* 6(1):84–97
5. Davis BM (2010) Creativity and innovation in business 2010 teaching the application of design thinking to business. *Int J Procedia Social Behav Sci* 2(4):6532–6538
6. Djaiz C, Monticolo D, Matta N (2008) Project memory decision making. *Int J e-Collaboration on Creativity, Innovation and e-Collaboration* 2(3):12–28
7. Fischer J, Gantner Z, Rendle S, Stritt M, Thieme LS (2006) Ideas and improvements for semantic wikis. *Lecture notes in computer science, the semantic web: research and applications* 4011:650–663
8. Gumusluoglu L, Ilsev A (2009) Transformational leadership, creativity, and organizational innovation. *J Bus Res* 62(4):461–473
9. Hauschildt J (2004) *Innovations management*, 3rd edn. Vahlen (Vahlens Handbücher der Wirtschafts- und Sozialwissenschaften), Munich
10. Ignat C-L, Oster G, Molli P, Cart M, Ferri_e J, Kermarrec A-M, Sutra P, Shapiro M, Benmou_ok L, Busca J-M, Guerraoui R (2007) A comparison of optimistic approaches to collaborative editing of wiki pages. *Proceedings of the international conference on collaborative computing: networking, applications and worksharing—collaboratecom 2007*, White Plains, IEEE Computer Society, New York, Nov 2007
11. Kousetti C (2008) *The wiki way to a semantic web*. University of Southampton, UK: School of Electronics and Computer Science, 2008
12. Krotzsch M, Vrandečić V, Volkel M, Haller H, Studer R (2007) Semantic Wikipedia. *J Web Semant: Sci Serv Agents World Wide Web* 5:251–261
13. Krotzsch M, Vrandečić D, Volkel M (2005) Wikipedia and the semantic web- the missing links, *WikiMania*
14. Leuf B, Cunningham W (2001) *The wiki way: quick collaboration on the web*. Addison-Wesley Longmann, Amsterdam
15. López-Nicolas C, Merono-Cerdan A (2011) Strategic knowledge management, innovation and performance. *Int J Inf Manage* 31(6):502–509
16. Liao S-H, Wu C (2010) System perspective of knowledge management, organizational learning, and organizational innovation. *Expert Syst Appl* 37(2):1096–1103
17. Majchrzac A, Wagner C, Yates D (2006) Corporate wiki users: results of a survey. *Proceedings of the ACM international symposium on wikis (Wikisym 2006)*, Odense, Denmark
18. Malone TW, Crowston K, Herman GA (2003) *Organizing business knowledge: the MIT process handbook*. The MIT Press, Cambridge
19. Monticolo D, Hilaire V, Koukam A, Gomes S (2007) *OntoDesign; A domain ontology for building and exploiting project memories in product design projects*. 2nd international conference in knowledge management in organizations, Lecce, Italia, Sept 2007
20. Monticolo D, Morel L, Boly V, lahoud I (2012) Wiki-I: a semantic wiki to support the ideas development and knowledge sharing in innovation activities. *Proceedings of the international multicongress of engineers and computer scientists 2012, IMECS 2012, Hong Kong, 14–16 March, 2012*. *Lecture notes in engineering and computer science* pp 1307–1311
21. Richards D (2009) A social software/Web2.0 approach to collaborative knowledge engineering. *Int J Inf Sci* 179:2515–2523
22. Riedl C, May N, Finzen J, Stathel S, Kaufman V, Krcmar H (2009) An idea ontology for innovation management. *Int J Semant Web Inf Syst* 5(4):1–18
23. Schaffert S (2006) *IkeWiki: a semantic wiki for collaborative knowledge management*, In: 1st international workshop on semantic technologies in collaborative applications, STICA, vol 6, 2006
24. Seaborne A, Prud'hommeaux E (2006) *SPARQL query language for RDF, 2006*. Technical Report <http://www.w3.org/TR/2006/CR-rdf-sparql-query-20060406/>, W3C

25. Singh AV, Wombacher A, Aberer K (2007) Personalized information access in a wiki using structured tagging. *On the move to meaningful internet systems: OTM 2007 Workshops, Lecture notes in computer science 4805*, Springer, ISSN 0302-9743 ISBN 978-3-540-76887-6, pp 427–436
26. Souzy A (2005) Building a semantic wiki. *IEEE Intel Syst* 20:87–91
27. Volkel M, Krtozsch M, Vrandecic D, Haller H, Studer R (2006) Semantic Wikipedia. In: *WWW '06: proceedings of the 15th international conference on world wide web*, ACM Press, New York, pp 585–594, 2006
28. Vrandecic D, Krötzsch M (2006) Reusing ontological background knowledge in semantic wikis. In: Max V, Sebastian S, Stefan D (eds) *Proceedings of the first workshop on semantic wikis—from wikis to semantics*, 2006
29. Yusuf S (2009) From creativity to innovation. *The Int J Technol Soc* 31(1):1–8

Overall Equipment Effectiveness Measurement: Weighted Approach Method and Fuzzy Expert System

M. Maran, G. Manikandan and K. Thiagarajan

Abstract This chapter is intended to enhance the original Overall Equipment Effectiveness (OEE). Calculating plant/equipment OEE can be very helpful for monitoring trends (such as whether a given plant is improving OEE over time) or as a rough measure of where a manufacturing plant lies in the OEE benchmarking spectrum. OEE concept in Total Productive Maintenance (TPM) implementation truly reduces manufacturing complexity into simple, intuitive presentation of information. The proposed approaches establish the functional technique for improvement of the effectiveness of production lines operating and dealing with uncertainty of the six major losses to OEE. The losses associated with production and limits for that losses are the major indexes of the production line performance, because it enables direct evaluation of production line output. The OEE is the process, which is acquired to specify an equivalent weight setting of every single element, even if, each concerning losses are totally different. Hence, the study proposes a weighted approach, to identify dissimilarity in weighting of each OEE element. Theoretical values for the losses improve the measurement of the OEE and Fuzzy analysis can help the decision makers to assess OEE for plant performances. Further, proposed fuzzy methodology can be used to reduce the indecisiveness. Therefore this technique introduces fuzzy theory for OEE computation and will also assist decision makers to evaluate uncertainty and imprecision. The

M. Maran (✉) · G. Manikandan
Department of Mechanical Engineering, Velammal College of Engineering and Technology,
Anna University, Madurai, Tamil Nadu, India
e-mail: gsmaran@yahoo.com

G. Manikandan
e-mail: manikandanguru@gmail.com

K. Thiagarajan
Department of Mathematics, Velammal College of Engineering and Technology,
Anna University, Madurai, Tamil Nadu, India
e-mail: vidhyamannan@yahoo.com

proposed concepts are used to find the OEE for the manufacturing plant, as well as to set the target for the plant and the area to focus for their improvement.

Keywords Availability · Decision making · Fuzzy analysis · OEE · Performance rate · Production losses · Quality rate · Signed rating · TPM effectiveness · Weighted rating

List of Acronyms

A	Availability
BD	Break-Down
JIPM	Japan Institute of Plant Maintenance
MI	Minor/idling stoppages
OEE	Overall Equipment Effectiveness
P	Performance rate
PE	Plant Effectiveness
Q	Quality rate
RL	Reject/rework Losses
RS	Reduced Speed
SA	Setup Adjustment
SL	Start-up Losses
TPM	Total Productive Maintenance

1 Introduction

In the highly competitive manufacturing environment prevailing at the global level, manufacturing plants of any country or region should be benchmarked and maintained at world-class level. Failing to do that, a plant may be exposed to various difficulties. The results of implementation of Total Productive Maintenance (TPM) program increased plant efficiency and productivity significantly by means of eliminating the major losses [1, 2–4], i.e., the total elimination of all six major losses, including breakdowns, equipment setup and adjustment losses, idling and minor stoppages, reduced speed, defects and rework, spills, process upset conditions, startup and yield losses.

1.1 Major Losses and Performance Evaluation

The Table 1 lists the Six Big Losses, and shows how they relate to the OEE Loss categories. World Class OEE is considered to be 85 % [5] or better. Clearly, there is room for improvement in most manufacturing plants. One of the major goals of

Table 1 Overall Equipment Effectiveness (OEE) and six major losses [2, 5]

OEE/loss category	Six big loss category	Event examples	Comment
Down time loss	Breakdowns	<ol style="list-style-type: none"> 1. Tooling failures 2. Unplanned maintenance 3. General breakdowns 4. Equipment failure 	There is flexibility on where to set the boundary between a breakdown (down time loss) and a small stop (speed loss)
	Set up and adjustments	<ol style="list-style-type: none"> 1. Setup/changeover 2. Material shortages 3. Operator shortages 4. Major adjustments 5. Warm-up time 	This loss is often addressed through setup time reduction programs
Speed loss	Small stops	<ol style="list-style-type: none"> 1. Obstructed product flow 2. Component jams 3. Miss- feeds 4. Sensor blocked 5. Delivery blocked 6. Cleaning/checking 	Typically only includes stops that are under 5 min and that do not require maintenance personnel
	Reduced speed	<ol style="list-style-type: none"> 1. Rough running 2. Under nameplate capacity 3. Under design capacity 4. Equipment wear 5. Operator inefficiency 	Anything that keeps the process from running at its theoretical maximum speed (also known as Ideal run rate or nameplate capacity)
Quality loss	Start-up rejects	<ol style="list-style-type: none"> 1. Product scrap 2. Product rework 3. In-process damage 4. In-process expiration 5. Incorrect assembly 	Rejects during warm-up, start-up and early production. May be due to improper setup or warm-up period
	Production rejects	<ol style="list-style-type: none"> 1. Product scrap 2. Product rework 3. In-process damage 4. In-process expiration 5. Incorrect assembly 	Rejects during steady-state production

TPM implementation and OEE computation is to eliminate or reduce the Six Big Losses—the most common causes of efficiency loss in manufacturing. Worldwide studies indicate that the average OEE rate in manufacturing plants is 60 % [5–7].

1.2 OEE and Its Significance

OEE is a “best practices” way to monitor and improve the manufacturing plants improvement and effectiveness of the manufacturing processes (i.e. machines, manufacturing cells, assembly lines). OEE is simple and practical. It takes the most common and important sources of manufacturing productivity loss, places them into three primary categories and distills them into metrics that provide an excellent gauge for measuring where and how to improve [8].

In general, OEE is calculated as the product of its three contributing factors:
Overall Equipment Effectiveness

$$\text{OEE} = A \times P \times Q,$$

where,

- A Availability of the machine. Availability is proportion of time, a machine is actually available to the scheduled time i.e., it should be available for work.
- P Performance Rate i.e., $P = RE \times SR$
- Q Quality rate is percentage of good parts out of total production, sometimes called yield.
- RE Rate Efficiency is actual average cycle time which is slower than design cycle time because of jams. Output is reduced because of jams.
- SR Speed Rate is actual cycle time which is slower than design cycle time; machine output is reduced because it is running at a reduced speed.

In other words, OEE can be expressed as follows,

$$\text{OEE} = \text{Utilization (A)} \times \text{Rate Loss (P)} \times \text{Yield (Q)}$$

OEE truly reduces complex production problems into simple, intuitive presentation of information. It helps to systematically improve the process with easy-to-obtain measurements.

OEE begins with Planned Production Time and scrutinizes efficiency and productivity losses that occur, with the goal of reducing or eliminating these losses. There are three general categories of loss to consider—Down Time Loss, Speed Loss and Quality Loss.

The proposed approaches give the mathematical approach for computing the Overall Equipment Effectiveness (OEE), i.e. plant effectiveness. OEE is essentially the ratio of Actual Productive Time to Planned Production Time.

The proposed approach assists decision makers to evaluate uncertainty and imprecision. Also to obtain improved OEE measurements as well as producing better production improvement plans and lean manufacturing implementation strategy. In general OEE (F) can be expressed as “n” number of factors namely $f_1, f_2 \dots f_n$.

$$F = f_1 * f_2 * \dots * f_n$$

In the above mentioned case, $n = 3$ where $f_1 = \text{Availability (A)}$, $f_2 = \text{Performance (P)}$, and $f_3 = \text{Quality (Q)}$.

OEE measurement is an effective way of analyzing the efficiency of a single machine or an integrated manufacturing system. OEE incorporates availability performance rate and quality rate and gives results. In other words, OEE addresses all losses caused by the equipment, not being available when needed due to breakdowns or set-up and adjustment losses, not running at the optimum rate due to reduced speed or idling and minor stoppage Losses and not producing first quality output due to defects and rework or start-up losses. A key objective of TPM is cost efficiency; maximizing Overall Equipment Effectiveness through the elimination or minimization of all these six losses by means of lucid approach.

The OEE is not that which gives one magic number; it gives three numbers other than OEE, which are all useful individually as the situation changes from day to day. In addition, it helps to visualize performance in simple terms—a very practical simplification. The OEE is probably the most important tool in continuous improvement program in manufacturing industry. Through the OEE analysis the operators can observe, where they lose most of the production (The six Major Losses as mentioned in above).

Significant improvement can be evident within a short period by means of eliminating the six major losses with result of enhanced maintenance activities and equipment management.

Increasingly, companies working in process manufacturing environments are discovering a surprisingly effective framework from which to tackle the continuous improvement challenge: the “Six Big Losses” approach. Developed in the 1970s by the Japan Institute of Plant Maintenance (JIPM), the Six Big Losses framework enables manufacturers to examine their efficiency problem with an unprecedented level of granularity According to JIPM, the OEE is based on three main aspects; each element concerns with different losses as shown in the Table 1.

2 OEE: Weighted Approach

2.1 OEE and its Computation Method

As described in World Class OEE, the OEE calculation is based on the three OEE Factors, Availability, Performance, and Quality. Here’s how each of these factors is calculated [1, 5].

2.1.1 Availability

Availability takes into account Down Time Loss, and is calculated as:

$$\text{Availability} = \frac{\text{Operating Time}}{\text{Planned Production Time}}$$

2.1.2 Performance

Performance takes into account Speed Loss, and is calculated as:

$$\text{Performance} = \text{Ideal Cycle Time}/(\text{Operating Time}/\text{Total Pieces})$$

Ideal Cycle Time is the minimum cycle time that your process can be expected to achieve in optimal circumstances. It is sometimes called Design Cycle Time, Theoretical Cycle Time or Nameplate Capacity. Since Run Rate is the reciprocal of Cycle Time, Performance can also be calculated as:

$$\text{Performance} = (\text{Total Pieces}/\text{Operating Time})/\text{Ideal Run Rate}$$

Performance is capped at 100 %, to ensure that if an error is made in specifying the Ideal Cycle Time or Ideal Run Rate the effect on OEE will be limited.

2.1.3 Quality

Quality takes into account Quality Loss, and is calculated as:

$$\text{Quality} = \text{Good Pieces}/\text{Total Pieces}$$

2.1.4 OEE

OEE takes into account all three OEE Factors, and is calculated as:

$$\text{OEE} = \text{Availability} \times \text{Performance} \times \text{Quality}$$

2.2 *Need of OEE: Weighted and Fuzzy Approaches*

Formerly, lots of research has been done to customize and fine tune the OEE computation Formula, i.e. $\text{OEE} = A \times P \times Q$ (specific to their plants/task). But there was no research was done based on the fixation of contributing factors and its effect on the OEE. So there was a chasm for real time values and targets fixed, that can be narrow down by appropriate method, for that this ideology will help the manufacturing industries to work on.

2.3 *Weighted Rating for A, P and Q*

In proposed system, OEE is considered as a signed graph, where each factors BD, SA, MI, RS, RL and SL are assigned with weight ratings based on the loss percentage. Signed weight [9, 10] (+, -, +-, --) is attached to each factors of

the graph based on the combination of above contributing factors for A, P, Q and OEE. Depending upon the BD, SA, MI, RS, RL and SL ratings A, P and Q are calculated. Similarly based on A, P and Q ratings OEE is estimated.

2.4 Existing Weighted Rating for A, P and Q

In a weighted graph, A1, A2, A3 and A4 are considered as the loss ratings, where $A1 < A2 < A3 < A4$. (i.e., A1 is the lowest loss level rating and A4 is the highest loss level rating) [10–13]. The ratings for the basic six parameters differ from one another based on user. In the subsequent ranking scheme, more weight-age is given for BD i.e., up to 10 % loss, where as other parameters have less weight-age of 2–3 %, [14] (Following set of data for model case plant limits values in %).

- | | |
|------------------------------|-----------------------------|
| 1. BD—Breakdown | 2. SA—Setup adjustment |
| a. A1 = 0 to \leq 1 | a. A1 = 0 to \leq 0.5 |
| b. A2 = $>$ 1 to \leq 5 | b. A2 = $>$ 0.5 to \leq 1 |
| c. A3 = $>$ 5 to \leq 10 | c. A3 = $>$ 1 to \leq 2 |
| d. A4 = $>$ 10 | d. A4 = $>$ 2 |
| 3. MI—Minor/Idling stoppages | 4. RS—Reduced speed |
| a. A1 = 0 to \leq 1 | a. A1 = 0 to \leq 1 |
| b. A2 = $>$ 1 to \leq 2 | b. A2 = $>$ 1 to \leq 2 |
| c. A3 = $>$ 2 to \leq 3 | c. A3 = $>$ 2 to \leq 3 |
| d. A4 = $>$ 3 | d. A4 = $>$ 3 |
| 5. RL—Reject / Rework losses | 6. SL—Start-up losses |
| a. A1 = 0 to \leq 0.5 | a. A1 = 0 to \leq 0.5 |
| b. A2 = $>$ 0.5 to \leq 1 | b. A2 = $>$ 0.5 to \leq 1 |
| c. A3 = $>$ 1 to \leq 2 | c. A3 = $>$ 1 to \leq 2 |
| d. A4 = $>$ 2 | d. A4 = $>$ 2 |

3 Program Dependence Graph

A Program Dependence Graph (PDG) is a suitable internal program representation for monolithic programs for the purpose of carrying out certain engineering operations such as scheming and computation of program metrics [13, 15] (Figs. 1, 2).

- In the following flow diagram, stage 1 calculation of availability with the inputs BD and SA with constrain limit shown in Fig. 3.
- In stage 2, calculation of performance with the inputs of MI and RS with constrain limit shown in the Fig. 4.
- In stage 3, calculation of quality with the inputs of RL and SL with constrain limit as in Fig. 5.
- Determination of OEE assigned as below $OEE = P_i \times Q_i \times R_i \times S_i \times T_i \times U_i$.

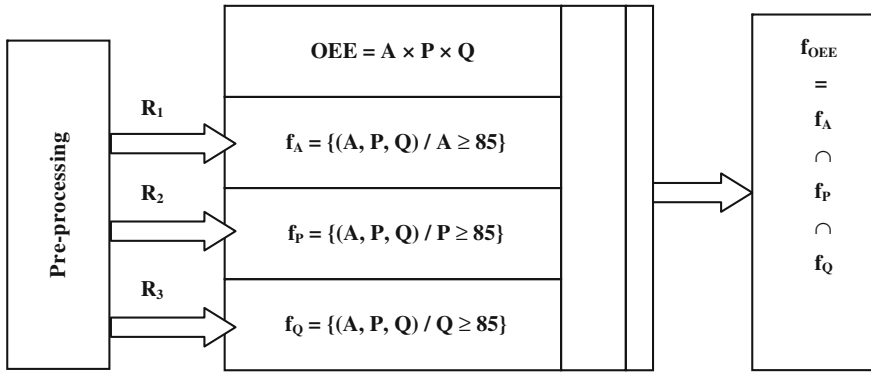


Fig. 1 PDG block diagram for OEE

- Stage 1: Calculation of Availability
- Stage 2: Calculation of Performance
- Stage 3: Calculation of Quality

4 Signed Approach for OEE

Signed weight [10] with following constrained users fixed methodology ++, +-, -, +, -- are fixed for best, better, good, worst ratings for A, P, Q and OEE is based on the combination of above contributing factors of A, P, Q and OEE. The BD, SA, MI, RS, RL and SL are assigned with signed ratings to allocate A, P and Q values. Similarly based on A, P and Q signed ratings OEE values are predicted [14] (Tables 2, 3, 4).

5 Fuzzyfication Algorithm for Current OEE

5.1 Programming Algorithm

For various A, P and Q values [1, 100], OEE computed and classified using following algorithm.

- Step 1: Start the process
- Step 2: Fix A = 0
- Step 3: Fix P = 0
- Step 4: Fix Q = 0
- Step 5: Calculate $OEE = A \times P \times Q$

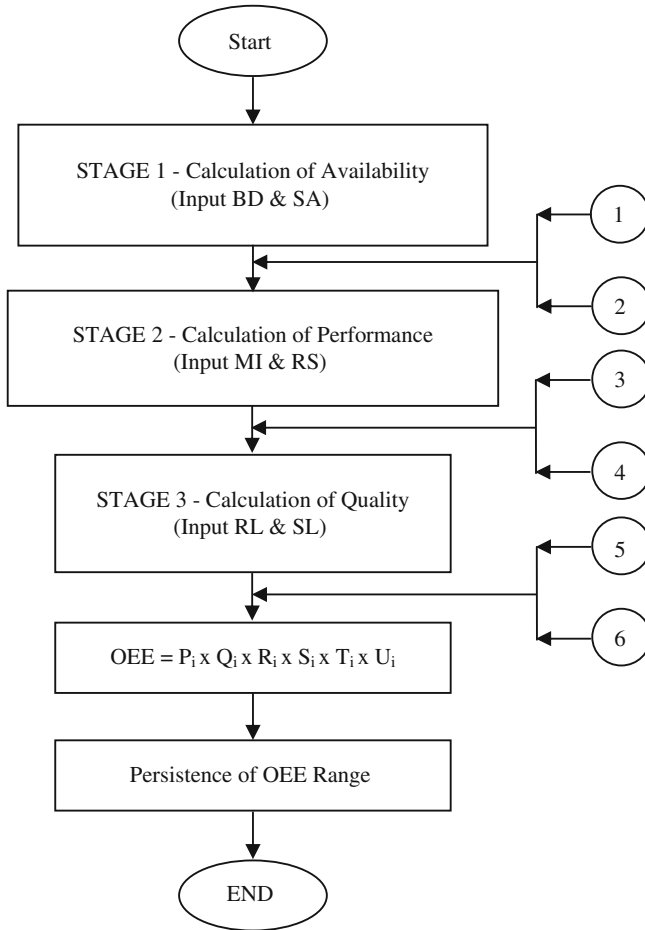


Fig. 2 Flow chart for current OEE

Step 6: Pre-processing

Step 7: Rotate Step 5 to Step 6 until $Q \leq 100$ with the step value of Q

Step 8: Rotate Step 4 to Step 7 until $P \leq 100$ with the step value of P

Step 9: Rotate Step 3 to Step 8 until $A \leq 100$ with the step value of A

Step 10: Stop the process

Note: In Pre-processing the constrains for A, P and Q are being fixed with plant required strategies. As stated in Sect. 2. The step values with respect to necessitate.

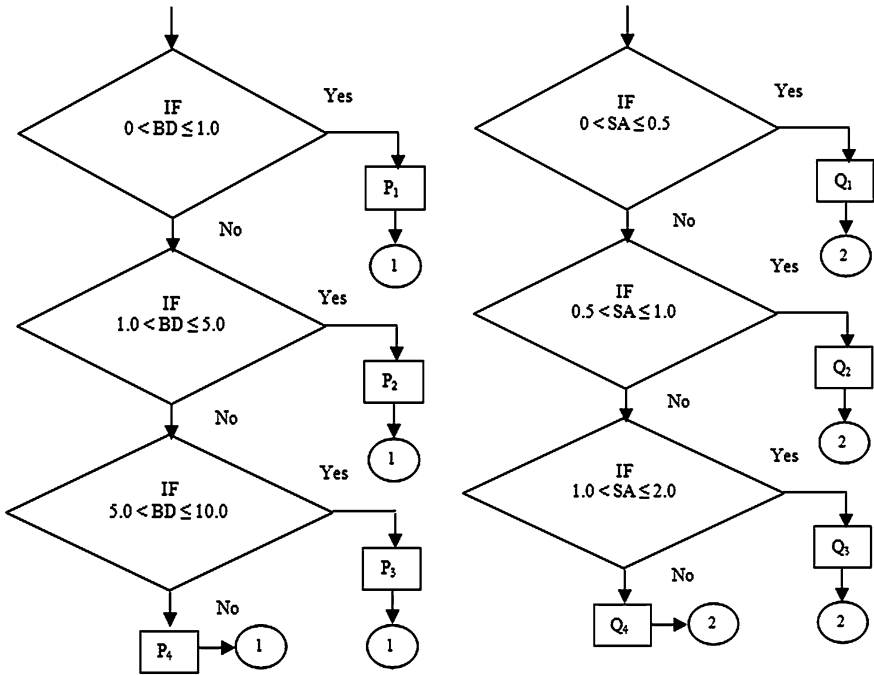


Fig. 3 Flow chart for availability calculation

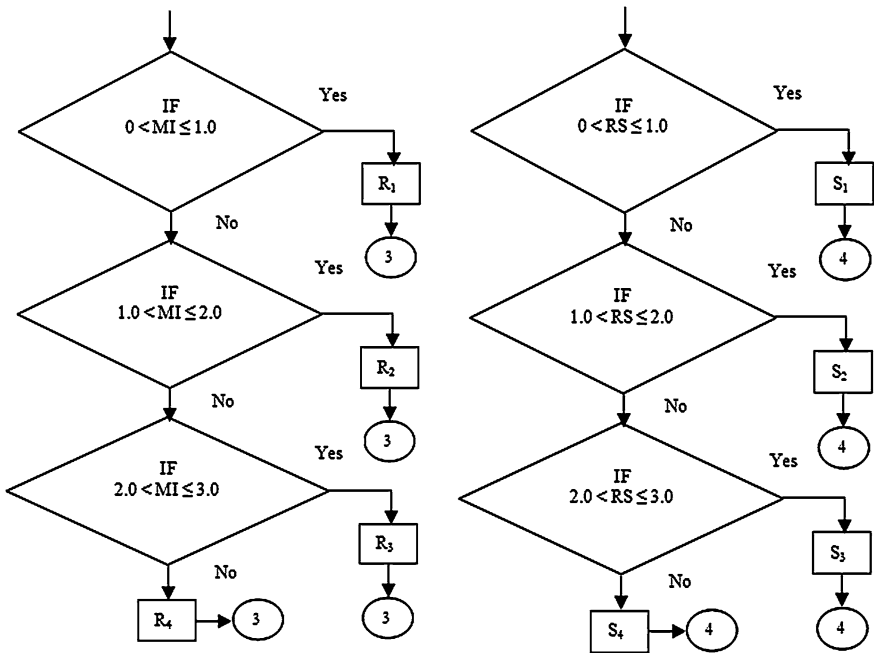


Fig. 4 Flow chart for performance calculation

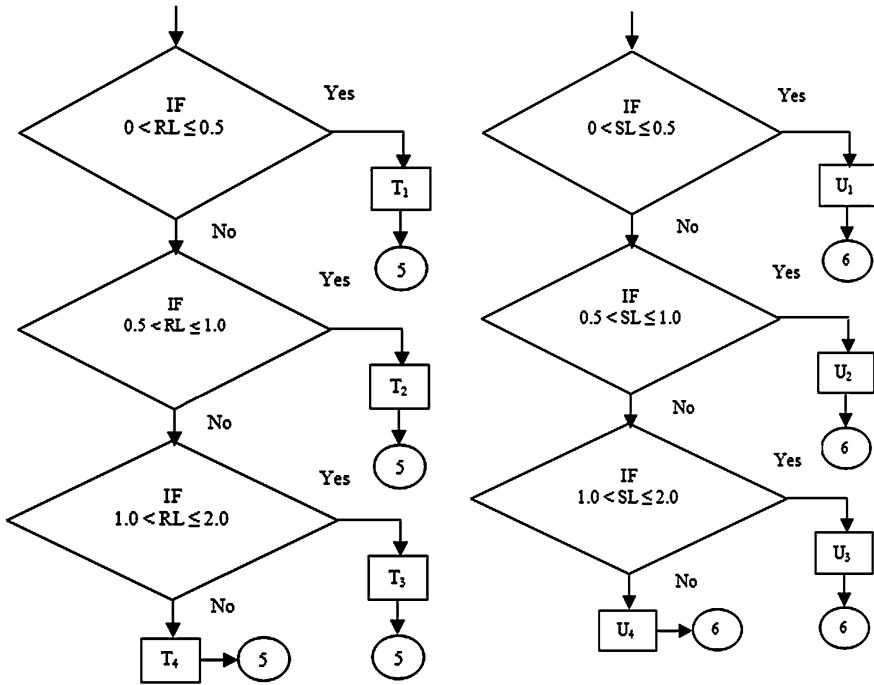


Fig. 5 Flow chart for quality calculation

Table 2 Availability rating based on breakdown and setup adjustment

Breakdown (BD)	Setup adjustment (SA)	Availability (A)
+	+	+
+	-	+
-	+	-
-	-	-

Table 3 Performance rating based on idling, minor stoppages and reduced speed

Idling and minor stoppages (IM)	Reduced speed (RS)	Performance (P)
+	+	+
+	-	+
-	+	-
-	-	-

5.2 Approximation Analysis for Current OEE

From the fuzzy analysis for the OEE in % using the formula [OEE = A × P × Q], even though the A, P, Q values are below the targeted

Table 4 Quality rating based on defect/rework and start-up losses

Defect/rework (DR)	Start-up losses (SL)	Quality (Q)
+	+	+
+	-	+
-	+	-
-	-	-

Table 5 OEE rating based on availability, performance and quality ratings

Availability (A)	Performance (P)	Quality (Q)	OEE
+	+	+	++
+	+	-	+ -
+	-	+	+ -
-	+	+	+ -
+	-	-	+ - *
-	+	-	- +
-	-	+	- +
-	-	-	--

*Availability (A) is having major impact on the OEE. In most circumstances, A (+) ranking gives better OEE, in spite of both P (-) and Q (-) ratings

Table 6 Experimental results

Availability	Performance	Quality	OEE
95	98	99	92.2
95	98	97	90.3
95	97	94	86.6
90	98	98	86.4
93	96	96	85.7*
88	98	94	81.1
87	91	92	72.8
86	90	90	69.7

* As availability (A) value is 93 %, i.e. + ratings give better OEE, even supposing P and Q have negative ratings. (As mentioned in Tables 5 and 6).

values Still OEE values or above the 85 %, the world class OEE value as target value. From the reading it's giving basic guidelines for setting the minimum criteria for A, P and Q values. The guiding principles are as follows;

5.2.1 Individual Constraint for Variables

$$A \geq 85; P \geq 85; Q \geq 85$$

5.2.2 Combine Constraint for Variables

$$(A + P + Q) \geq 284$$

So the target setting for a particular industry may use this as basic guidelines and set A, P and Q values based on their current performance results with benchmarked OEE.

For various A, P and Q values [1, 100], OEE computed and classified using following algorithm.

In a weighted graph, A1, A2, A3 and A4 are considered as the loss ratings, where $A1 < A2 < A3 < A4$. (i.e., A1 is the lowest loss level rating)

5.3 Boundaries for A, P and Q

For setting maximum (extreme) level values for A, P and Q for current bench mark OEE target is as follows.

5.3.1 Extreme Level

In this case one of the factors as minimum at benchmark setup and other two at maximum (100 %).

5.3.2 Average Level

In the average order case with equal rating at 95 % can be articulated.

6 Real Time Application

OEE data collection, analysis and reporting provide the principal basis for improving equipment effectiveness by eliminating the major equipment-related losses. The weighted approach for OEE calculation is presented and the analytical hierarchy process has been applied for setting the weight of all the factors namely BD, SA, MI, RS, RL and SL. The prospective of the proposed method can be applied for new as well as existing manufacturing plants. OEE data very quickly leads to root-cause identification and elimination of losses for plant performance improvement. Overall Equipment Effectiveness continues to gain acceptance as an effective method to measure production floor performance.

Capturing reliable production floor information is critical for producing reliable OEE focused progress of the plant or equipment. It will be optimal tool for plants

implementing TPM. This weighted approach and OEE results will be benchmarking tool to gauge manufacturing system, especially maintenance management system. OEE helps manufacturer to improve productivity and get better visibility of the operations and also allow management to take sound decisions.

7 Conclusion and Future Work

The proposed methodology through algorithmic approach to get computational work, the values of OEE categorized through fuzzy logic method on A, P and Q to get optimization.

No manual data collection or manual compilation for OEE calculations is the first step in improving both the accuracy of OEE reports as well as reducing the cost to produce the reports through neural network.

The utility of artificial neural network models lies in fact that, they can be used to infer for future purposes.

Acknowledgments The authors would like to thank Dr. Ponnammal Natarajan, Former Director of Research, Anna University, Chennai, India. Currently Advisor (Research and Development), Rajalakshmi Engineering College, for her intuitive ideas and fruitful discussions with respect to the paper's contribution and the Management of Velammal College of Engineering and Technology, Madurai and Dr. R. Vivekaanandhan, Assistant Professor, Department of English, Velammal College of Engineering and Technology, Madurai, for his constant encourage and support to complete the task.

References

1. Nakajima S (1988) Introduction to TPM. Productivity Press, Cambridge
2. Venkatesh J (2005) An introduction to total productive maintenance (TPM), The plant maintenance resource center, pp 2–3, 15–27
3. Wireman T (1990) World class maintenance management. Industrial Press, New York
4. Yamashina H (1995) Japanese manufacturing strategy and the role of total productive maintenance. *J Qual Maint Eng* 1(1):27–38
5. <http://www.oee.com/index.html>
6. Lungberg O (1998) Mearsument of overall equipment effectiveness as a basic for TPM activities. *Int J oper Prod Manag* 18(5):495–507
7. Lesshammar M (1999) Evaluation and improvement of manufacturing performance measurement systems—the role of OEE. *Int J Oper Prod Manag* 19(1):55–78
8. Dal B, Tugwell P, Greatbanks R (2000) Overall equipment effectiveness as a measure of operational improvement. *Int J Oper Prod Manag* 20(12):1488–1520
9. Yu B, Singh MP (2002) An evidential model for distributed reputation management, AAMAS'02, Bologna, Italy, 15–19 July 2002
10. Thiagarajan K, Raghunathan A, Natarajan P, Poonkuzhali G, Ranjan P (2009) Weighted graph approach for trust reputation management. *World Acad Sci Eng Technol* 32:830–836
11. Yu B, Singh MP (2000) Trust and reputation management in a small- world network. In: proceedings of the 4th international conference on multiagent systems, pp 449–450

12. Yu B, Singh MP (2000) A social mechanism of reputation management in electronic communities. In: proceedings of the 4th international workshop on cooperative information agents
13. Harary F (1969) Graph theory. Addison-Wesley, Reading
14. Maran M, Manikandan G, Thiagarajan K (2012) Overall Equipment effectiveness measurement by weighted approach method. In: Lecture notes in engineering and computer science: proceedings of the international multi conference of engineers and computer scientists 2012, IMECS 2012, 14–16 March, Hong Kong, pp 1280–1285
15. Srinivasan SP, Malliga P, Thiagarajan K (2008) An Algorithmic graphical approach for ground level logistics management of Jatropha seed distribution. *Ultra Sci Phys Sci* 20(3):605-612

A Supplier-Manufacturer Model for Procurement Plan in Export Oriented Furniture Industry with Sustainability Considerations

Muh. Hisjam, Achmad Habibie, Wahyudi Sutopo
and Kuncoro Harto Widodo

Abstract Due to limited supply of high quality of teak log as raw materials for export oriented furniture industry in Indonesia, procurement plan becomes a critical operation. The furniture industry (manufacturer) buys high quality teak log from Perum Perhutani (supplier), a state-owned forestry company that has responsibility not only to supply teak logs but also to keep the environmental and social role of the forest in Indonesia. The manufacturer should also consider sustainability due to the demands of the stakeholders. This study proposes a relationship model between supplier and manufacturer for procurement plan and formulated as Goal Programming to get the most favorable solution considering economical, environmental, and social aspect both of supplier and manufacturer. The results show that the model can be used to determine the decision variables considering sustainability aspects.

Keywords Export oriented furniture industry · Procurement plan · Goal programming · Supplier-manufacturer relationship model · Sustainability · Teak wood

M. Hisjam (✉) · A. Habibie · W. Sutopo
Logistics and Business System Laboratory, Department of Industrial Engineering,
Sebelas Maret University, Ir. Sutami Street 36A, Surakarta 57126, Indonesia
e-mail: hisjam@uns.ac.id

A. Habibie
e-mail: abiehay@gmail.com

W. Sutopo
e-mail: sutopo@uns.ac.id

K. H. Widodo
Department of Agro-industrial Technology, Gadjah Mada University,
Sosio Justisia Street, Bulaksumur, Yogyakarta 55281, Indonesia
e-mail: kuncorohw@yahoo.com

1 Introduction

Wooden furniture industry in Indonesia has been declining in its export at five recent years. From 2007 to 2011, wooden furniture export generally decreasing about 9 % per year for the export volume and about 6 % per year for the export value [1, 2]. The main problems are the limited supply of high quality teak wood and production inefficiency. The limited supply problem is related to the sustainability of teak forest that vulnerable due to illegal logging. While the production inefficiency impacts becomes more serious problem because export destination countries and foreign buyers impose strict requirement for furniture exporters to comply with several policy, i.e. environmentally benign and labor rights protection [3, 4].

Perum Perhutani (PP) is a state-owned forestry company that produce teak log as a raw material in furniture industry. PP plants, manages, harvests forest, and sells the teak log produced to furniture industry [4-7]. PP must supply teak log needed by furniture industry to fulfill customer demand and must conserve the forest to serve environmentally and socially. PP considers not only the recent demand but also future demand, so forest must be conserved [8]. As a state-owned company, PP must run Corporate Social Responsibility (CSR) [5, 9]. As a consequence, it will reduce the PP's profit but it will also give benefit to PP. CSR should motivate people around forest to help PP to keep the forest from illegal logging.

The study was held in CV. Valasindo Sentra Usaha (VSU), an export oriented furniture industry in Central Java, Indonesia. Teak log as raw material is bought from PP using a partnership contract. Relationship between PP and VSU is a relationship between supplier and manufacturer.

As a natural resources, for providing teak log needs long time because trees need to time to grow. Furniture industry should consider this in the strategic plan, so the growth of this industry should be supported by the availability of the raw material. Besides log is a commodity that related to forest that has environmental role, so the using of this resources should be wise, the minimum the better.

This will implies that VSU and other furniture industry face problem related to limited supply of high quality teak logs from PP. This condition makes the log price become more expensive. Of course, it will make furniture industry suffering because raw material cost contributes more than 60 % of total production cost [10]. Furniture industry must decide the kind and the quantity of logs so the purchasing cost can be minimized.

The raw material not only limited in quantity but also limited in time to be purchased. PP can sell the logs after harvested from forest. Harvesting begins each year around of March and completed around of October. With the limited time, furniture industry must make good schedule related in procurement plan.

Therefore, VSU should have good business relationship with PP to reduce the procurement cost of raw material. VSU considers minimizing scrap and labor productivity programs. Part of the scrap can be recycled to other products, another

part is used as fuel for drying operation and the rest is sold as bulky scrap. VSU must also prosecute the government regulation for employee safety and healthy, so VSU must provide Personal Protective Equipment (PPE) for VSU's employee as social aspect of sustainability.

The model is required to determine the new paradigm of business contract in three aspects i.e. economical, environmental, and social aspects for procurement plan in export oriented furniture industry with sustainability considerations [11, 12]. It is possible that between aspects considered are conflicted, so we propose a relationship model between supplier and manufacturer using goal programming technique [13].

The model is based on our previous work [4, 14]. In this book chapter, we conduct further analysis to the model that is sensitivity analysis to the model.

2 Problem Description

The problems concerned in this paper are related to three aspects i.e. economical, ecological, and social aspects. Analyzing business process between PP and VSU will help exploring the problem. Business process between PP and VSU is shown by Fig. 1.

In this model, we take a case study by PP as a supplier as well as VSU Company as a manufacturer. Therefore we consider economic variable, ecological variable, and social variable. So we develop six goals that can be categorized in three main objectives.

From the supplier side, three goals were developed. The first goal is to maximize PP's profit (economic variable). The second goal is to maximize conserved forest (ecological variable). The third goal is maximize PP's CSR (social variable).

Whilst from the manufacturer side, three other goals were developed. The fourth goal is to maximize profit (economic variable). The fifth goal is to maximize waste selling to minimize dispose waste (ecological variable). The last goal is to maximize PPE for VSU's employees (social variable).

Thus, in this paper we propose a relationship model between supplier (PP) and manufacturer (VSU) on teak log procurement with sustainability considerations. This model is expected give win-win solution for PP as a supplier and VSU and manufacturer.

3 Mathematical Modelling

Based on the above description, this paper consider three aspects and its mean the model will be developed is multi objectives. Goal Programming (GP) is a suitable tool for decision maker to analyze the achievement of the desired goals considering different and sometimes conflicting multiple objectives.

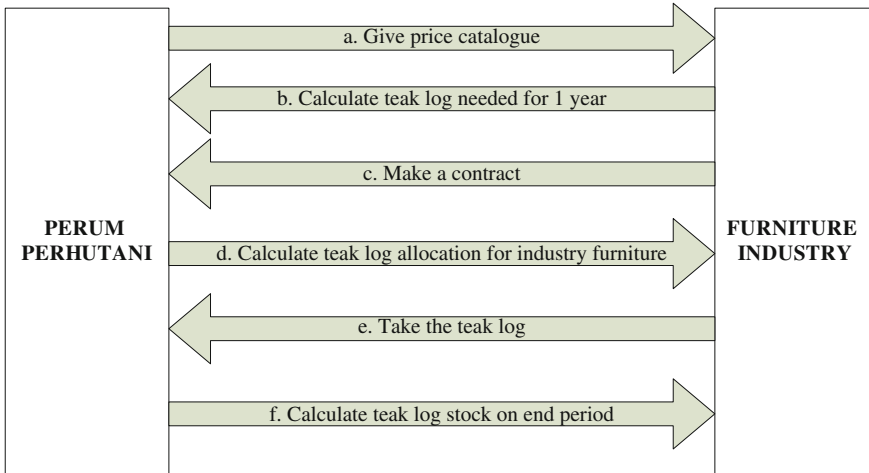


Fig. 1 The business process of PP—furniture industry relationship. Source [14]

3.1 Index and Notation

The notations in the formulation will be described. The notations of parameters and decision variables as follow [14].

Parameters:

- C_t^o Planting cost in period t (IDR/ha)
- C_t^p Maintenance cost in period t (IDR/ha)
- C_t^v Harvesting cost in period t (IDR/ha)
- C_t^{tk} Labor cost in period t (IDR/m³)
- C_t^b Overhead cost in period t (IDR/m³)
- C_{jt}^h Inventory holding cost for type j in period t (IDR/m³)
- P_{jt}^l Selling price of teak log in period t (IDR/m³)
- P_{kt}^v Selling price of furniture in period t (IDR/m³)
- P_t^s Selling price of waste in period t (IDR/m³)
- CSR_t CSR cost paid by Perhutani in period t (IDR)
- G_t APD cost in period t (IDR)
- L_t Waste sold by VSU in period t (IDR)
- TP_{Pt} Total profit of Perhutani in period t (IDR)
- $TP_{VSU t}$ Total profit VSU in period t (IDR)
- B Percentage of CSR cost
- γ_j Conversion value from log to furniture for type j
- α_j Conversion value from log to waste for type j
- n_i Negative deviation of function i
- p_i Positive deviation of function i
- b_i Desired value of function i

Decision variables:

- A_t Planted forest area (ha)
- B_t Conserved forest area (ha)
- C_t Harvested forest area (ha)
- q_{jt} Teak log sold by Perhutani type j in period t (m^3)
- Q_{jt} Teak log bought by VSU type j in period t (m^3)
- V_{kt} Furniture sold by VSU type k in period t (m^3)

3.2 Objective Function

There are three aspects considered in this paper and in this paper there are two entities, PP as a supplier and VSU as manufacturer. The six objective functions will cover all aspects from two entities [14].

The first goal (1) is maximization of PP's profit. The second goal (2) is maximization of furniture industry's profit. The third goal, maximization conserved forest was expressed in (3). The fourth goal (4), maximization waste selling from furniture production. Equations (5) and (6) state social goal. Equation (5) states the fifth goal, maximization CSR from PP. The last goal (6), maximization PPE for employee safety.

$$\sum_{t=1}^{12} TP_{Pt} = \sum_{t=1}^{12} \sum_{j=1}^3 P_{jt}^l q_{jt} - \sum_{t=1}^{12} C_t^o A_t - \sum_{t=1}^{12} C_t^p A_t - \sum_{t=1}^{12} C_t^v C_t - \sum_{t=1}^{12} CSR_t \quad (1)$$

$$\begin{aligned} \sum_{t=1}^{12} TP_{VSUt} &= \sum_{k=1}^2 \sum_{t=1}^{12} P_t^v V_{kt} - \sum_{t=1}^{12} \sum_{k=1}^2 C_t^{tk} V_{kt} - \sum_{t=1}^{12} \sum_{j=1}^3 P_{jt}^l Q_{jt} - \sum_{t=1}^{12} \sum_{j=1}^3 C_{jt}^h Q_{jt} \\ &- \sum_{t=1}^{12} \sum_{k=1}^2 C_t^b V_{kt} + \sum_{t=1}^{12} L_t - \sum_{t=1}^{12} G_t \end{aligned} \quad (2)$$

$$\sum_{t=1}^{12} B_t = \sum_{t=1}^{12} A_t - \sum_{t=1}^{12} C_t \quad (3)$$

$$\sum_{t=1}^{12} L_t = \sum_{t=1}^{12} \sum_{j=1}^3 P_t^s \alpha_j Q_{jt} \quad (4)$$

$$\sum_{t=1}^{12} CSR_t = \sum_{t=1}^{12} \beta TP_{Pt} \quad (5)$$

$$\sum_{t=1}^{12} G_t = \sum_{t=1}^{12} C_t^p k_t \tag{6}$$

3.3 Mathematical Formulation

The objective function is changed to soft constraint in Goal Programming (GP). The objective function is added with positive deviation, negative deviation, and desired value. The model then can be formulated as GP as follows [14].

$$\text{Min} \left(\sum_{i \in I} (n_i + p_i) \right) \tag{7}$$

Subject to:

$$\sum_{t=1}^{12} \sum_{j=1}^3 P_{jt}^l q_{jt} - \sum_{t=1}^{12} C_t^o A_t - \sum_{t=1}^{12} C_t^p A_t - \sum_{t=1}^{12} C_t^v C_t - \sum_{t=1}^{12} CSR_t + n_1 - p_1 = b_1 \tag{8}$$

$$\begin{aligned} \sum_{k=1}^2 \sum_{t=1}^{12} P_t^v V_{kt} - \sum_{k=1}^2 \sum_{t=1}^{12} C_t^{tk} V_{kt} - \sum_{t=1}^{12} \sum_{j=1}^3 P_{jt}^l Q_{jt} - \sum_{t=1}^{12} \sum_{j=1}^3 C_{jt}^h Q_{jt} - \sum_{k=1}^2 \sum_{t=1}^{12} C_t^b V_{kt} \\ + \sum_{t=1}^{12} L_t - \sum_{t=1}^{12} G_t + n_2 - p_2 = b_2 \end{aligned} \tag{9}$$

$$\sum_{t=1}^{12} A_t - \sum_{t=1}^{12} C_t + n_3 - p_3 = b_3 \tag{10}$$

$$\sum_{t=1}^{12} P_t^s \alpha_j Q_{jt} + n_4 - p_4 = b_4 \tag{11}$$

$$\sum_{t=1}^{12} \beta TP p_t + n_5 - p_5 = b_5 \tag{12}$$

$$\sum_{t=1}^{12} C_t^p k_t + n_6 - p_6 = b_6 \tag{13}$$

$$\sum_{t=1}^{12} \sum_{j=1}^3 Q_{j(t-1)} - \gamma_j Q_{jt} + Q_{jt} \leq 1000 \tag{14}$$

$$\sum_{t=1}^{12} \sum_{j=1}^3 Q_{jt} - \gamma_j Q_{jt} = 0 \quad (15)$$

$$\sum_{t=1}^{12} A_t + B_t - C_t \leq 446790 \quad (16)$$

where n_i and p_i are defined as preferential weight, negative deviational variable, and positive deviational of the goal, b_i denote the target level for each goal respectively. In this paper the number of goals is six. Some literatures defined (7) as the achievement function, which must be minimized to ensure that the solution is closely as possible to the desired goals [13].

4 Solution Method, Numerical Example, and Results

In this computational study, we analyze the impact of the changes in parameters in the supplier-manufacturer relationship model between PP and VSU considering several goals that must be achieved.

4.1 Solution Method

The first step to solve the GP formulation is determining initial target level for all goals. Initial target level is determined by decision maker. It will be compared with the result of GP formulation. The second step is applying simplex to minimize deviation variable, there are positive variable and negative variable. Then compare the result with initial target level and check them satisfied or not. If there are goals not satisfied yet, adjust the target level goal which the biggest deviation until all goals satisfied.

4.2 Numerical Example

To illustrate the capabilities of the proposed-model, we demonstrate a numerical example using data in Tables 1, 2, 3, 4, 5, 6 and 7.

Table 1 Parameter cost

Parameter	Quantity	Units
Planting cost	43,329,000.00	IDR/ha
Maintenance cost	160,500.00	IDR/ha
Harvesting cost	16,785,300.00	IDR/ha
Labor cost	2,230,400.00	IDR/m ³
Overhead cost	1,500,000.00	IDR/m ³

Source [14]

Table 2 Conversion value from furniture to log

Parameter	Quantity	Units
GF	AIII	5.2
GF	AII	20
Indoor	AIII	12.2
Indoor	AII	4.95

Source [14]

Table 3 Conversion value from log to scrap

Furniture type	Log type	Conversion value
GF	AIII	80.8 %
GF	AII	95 %
Indoor	AIII	91.8 %
Indoor	AII	79.8 %

Source [14]

4.3 Results

We run three scenarios to get the best result. The scenarios are scenario A to optimistic target level, scenario B to pessimistic target level, while scenario C to normal target level. The best scenario must satisfy all of goals. The results of scenario A, B, and C can be seen on Table 8.

To illustrate the capabilities of the model, we use numerical trial based on our observation in PP and furniture manufacturers in Central Java. Teak forest stand area in PP is estimated about 447,690 ha, and only 313,383 ha can be harvested. The remaining forest area is categorized as conserved forest. However, based on the government regulation, the minimum area of forest must cover at least 30 % of area within. As a consequence, decision makers set the target level of forest area that must be harvested at most 313,383 ha.

In optimistic scenario A, all goals set by decision makers are satisfied except G2. G2 refers to VSU's profit, it's mean that the target level for VSU's profit is too large. In order to do so, decision makers can adjust the target level set in scenario A.

Table 4 Furniture demand

Period	GF	Indoor	Units
1	0	19.63	m ³
2	8.74	15.44	m ³
3	7.39	8.74	m ³
4	0	12.30	m ³
5	4.68	0	m ³
6	4.52	3.22	m ³
7	9.68	14.71	m ³
8	8.93	22.25	m ³
9	0	9.77	m ³
10	0	0	m ³
11	16.56	13.98	m ³
12	0	7.43	m ³

Source [14]

Table 5 Furniture price

Furniture type	Price	Units
Indoor	27,750,000.00	IDR/m ³
GF	37,000,000.00	IDR/m ³

Source [14]

Table 6 Teak log price

Furniture type	Price	Holding cost	Units
AII	3,158,000.00	94,740.00	IDR/m ³
AIII	7,144,000.00	214,320.00	IDR/m ³

Source [14]

Table 7 PPE price

Level of PPE chosen	Price	Units
1 (a)	9,000.00	IDR
2 (a + b)	15,000.00	IDR
3 (a + b + c)	300,000.00	IDR
4 (a + b + c + d)	8,500.00	IDR
5 (a + b + c + d + e)	32,000.00	IDR

Note (a) Mask (b) Helmet (c) Safety shoes (d) Gloves (e) Ear-plugs

Source [14]

Scenario B is one alternate solution which its target level is lower than the scenario A. It can be seen in Table VIII that all target levels are lowered, except G3, G4, and G5. G3 refers to goal set by decision makers to comply with government regulation about forest area that must be conserved by PP. G4 refers to goal set by decision makers that VSU's revenue from waste selling is used for

Table 8 Results from scenarios

Scenario	Goals	Target level	Target level value	Achieved value	Satisfaction
A	G1	At least 30 % from PP's revenue	945,201,862.00	945,201,900.00	Yes
	G2	At least 30 % from VSU's revenue	1,732,601,140.00	1,225,865,000.00	No
	G3	At least 30 % from total forest	134,037	$\geq 134,037$	Yes
	G4	At least equal PPE cost	24,786,000.00	248,286,297.00	Yes
	G5	At most 2 % from PP's profit	18,904,037.00	18,904,000.00	Yes
	G6	At least level 5	24,786,000.00	24,786,000.00	Yes
B	G1	At least 10 % from PP's revenue	315,067,287.00	315,067,300.00	Yes
	G2	At least 10 % from VSU's revenue	577,533,700.00	577,533,700.00	Yes
	G3	At least 30 % from total forest	134,037	$\geq 134,037$	Yes
	G4	At least equal PPE cost	22,032,000.00	248,286,297.00	Yes
	G5	At most 2 % from PP's profit	6,301,345.00	6,301,345.00	Yes
	G6	At least level 3	22,032,000.00	22,302,000.00	Yes
C	G1	At least 20 % from PP's revenue	945,201,862.00	945,201,900.00	Yes
	G2	At least 30 % from VSU's revenue	1,155,067,000.00	1,155,067,000.00	Yes
	G3	At least 30 % from total forest	134,037	$\geq 134,037$	Yes
	G4	At least equal PPE cost	24,786,000.00	248,286,297.00	Yes
	G5	At most 2 % from PP's profit	18,904,037.00	18,904,000.00	Yes
	G6	At least level 5	24,786,000.00	24,786,000.00	Yes

Source [14]

buying PPE. G5 refers to goal set by decision makers to comply with government regulation about CSR. The result from scenario B is all goals are satisfied. However, the achieved value is lower than target value so the decision makers can improve the target level until the achieved value is approaching the target value and the goal is still satisfied.

Pareto efficient solution for GP formulation can be seen in scenario C. It can be seen that all goals are satisfied. It is the best scenario because if the target levels are increased, the goals will not be satisfied. From the result, the PP's profit is IDR 945,201,900.00 and the VSU's profit is IDR 1,155,067,000.00. The harvested forest is less than 30 % of total forest so it is not violate the government regulation.

The VSU's revenue from waste selling can be used for buying PPE. CSR cost that must be taken out by PP is IDR 6,301,345.00. Highest level of PPE can be bought by VSU for improving employee safety.

4.4 Sensitivity Analysis

In Fig. 2, economic and environmental factor of PP are presented as the function of planting cost. The first goal of the model which is the profit maximization of PP is kept in the fixed value, whereas the average of forest area in planning period is investigated for a given planting cost. The planting cost is allocated from the various proportion of revenue. The plot line shows that the bigger planting cost which allocated to plant teak log, the smaller average of forest area. The decline of forest area is caused by the economical is kept in the fixed value, as a result the average forest area is sacrificed to maintain the economic goal. However when environmental is set as the first priority goal, the average forest area is increase as the planting cost increase. The increase is almost linear to the increase of planting cost.

The trade-off between economical and social factor faced by VSU is presented in Fig. 3. The profit and total cost of VSU is plot against allocation of PPE per employee. Since PPE budget is proportional to number of employees, hence the plot is linear. It is not surprising that as the allocation of protective budget increase will reduce the profit because the budget is treated as fixed cost in the VSU objective. Hence as the budget increase will increase the total cost and in turn decreases the profit of VSU.

Proposition *In furniture supply chain that considers environmental and social factors aside of economical, there is trade-off that must be faced by decision makers.*

Proof It can be seen that from (1–6), the environmental and social aspects are the planting cost of PP and protective equipment cost of VSU. As a result, the profit of PP and VSU depend on them, the lower the cost the higher profit and the higher the cost the lower the profit.

Goal Programming (GP) presented above can be used by the decision makers in furniture industry to analyze many alternatives of goals that they desired. For example, if decision makers wish to consider the environmental or social aspects in their business, they can estimate the impacts of this consideration to the traditional and the main priority of business, i.e. economical aspects. As a result, they can choose the desired corresponding goals that suitable environmental regulation and social requirement while not sacrificing the business goal.

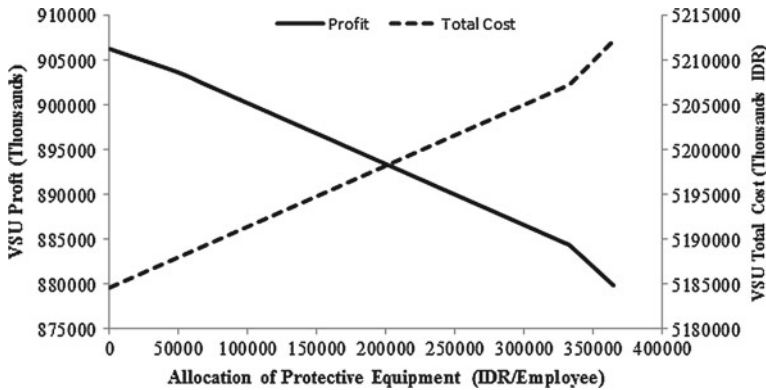


Fig. 2 Forest area as function of proportion of perum perhutani revenue

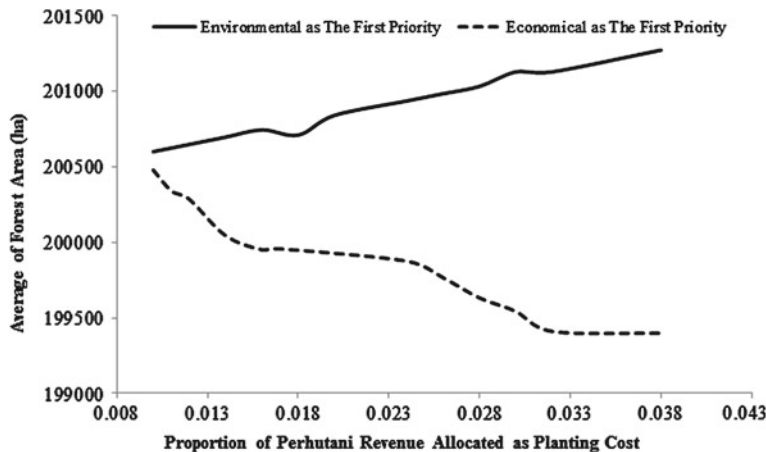


Fig. 3 Effect of PPE to profit and total cost of VSU

5 Conclusion and Future Work

The model is a relationship model between supplier and manufacturer with sustainability considerations. The model can determine the value of decision variables consist of planted forest area, conserved forest area, harvested forest area, teak log sold by supplier, teak log bought by manufacturer, and furniture sold by manufacturer. The proposed model can be used to determine the six goals for securing availability of teak log in export oriented furniture industry with sustainability considerations.

Business today requires not only economical aspect but also environmental and social aspects which can be conflicted. Trade-off cannot be avoided thus a set of decision that satisfied those objectives and goals must be made carefully by

decision makers. GP can be used to analyze the conflicting objectives in the furniture industry, hence the decision makers can make the suitable decision that they desired.

Further research can be conducted in adding more criteria of each aspects of sustainability. Another research topic is supplier selection with considering aspects of sustainability.

Acknowledgments Thanks to the Perum Perhutani Unit I Central Java, CV. Valasindo Sentra Usaha and ASMINDO Komda Solo Raya which have become partners in the research project. The research is supported by the Directorate General of Higher Education (DGHE), Ministry of National Education, Republic of Indonesia in HIBAH BERSAING Research Program (**contract No. 04/UN27.8/PN/2012**).

References

1. Ministry of Forestry (2010) Export and import of forest commodities, Year: 2007–2009. Ministry of Forestry, Republic of Indonesia
2. BPS—statistics of Indonesia (2012). Export Table by Commodity. (Online) Available. <http://www.bps.go.id/eng/exim-frame.php>. Accessed 4 June 2012
3. Hisjam M, Ota I, Guritno AD, Simon H, Tandjung SD (2010) Comparing the practices of forest product certification between perum perhutani and yusuhara forest owner's cooperative. Proceeding of the 3rd international seminar of Gajah Mada—Ehime network sustainable bio-resources for global welfare. Badung, Indonesia, pp 85–92, 7–8 Aug 2010
4. Hisjam M, Guritno AD, Simon H, Tandjung SD (2011) A framework for the development of sustainable supply chain management for business sustainability of export-oriented furniture industry in Indonesia (A case study of teak wooden furniture in Central Java Province). Proceedings of the 1st international conference on industrial engineering and service science 2011, IESS 2011, Solo, Indonesia, pp 285–290, 20–21 Sept 2011
5. Regulation Number 6 of 2007 the Republic of Indonesia concerning on Forest Governance, Development of Forest Management, and Utilization of Forest (Undang–Undang tentang Manajemen Pengembangan Hutan dan Fungsi Hutan No. 6 Tahun 2007), in Bahasa Indonesia
6. Regulation Number 19 of 2003 the Republic of Indonesia Concerning on State-Owned Enterprises (Undang–Undang tentang BUMN No. 19 Tahun 2003), in Bahasa Indonesia
7. Kurniawan B, Hisjam M, Sutopo W (2011) Integration of production and supply chain strategic planning for renewable resources under sustainability considerations: teakwood case study. In: Proceedings of IEEE, international conference on industrial engineering and engineering management, Singapore, pp 433–437, Dec 2011
8. Ginoga K, Wulan YC, Djaennudin D (2005) Carbon and its role in enhancing economic value of teak plantation in Saradan forest resort, East Java (Karbon dan Peranannya dalam Meningkatkan Kelayakan Usaha Hutan Tanaman Jati di KPH Saradan, Jawa Timur). Social Econ Res J 2(2):183–302 (in Bahasa Indonesia)
9. Regulation Number 9 of 2005 the Republic of Indonesia concerning on Small-Medium Enterprises (Undang–Undang tentang UKM No. 9 Tahun 2003), in Bahasa Indonesia
10. Sutopo W, Devi AOT, Hisjam M, Yuniaristanto S (2012) A model for procurement and inventory planning for export-oriented furniture industry in Indonesia: a case study. In: Proceedings of the international multiconference of engineers and computer scientists 2012, IMECS 2012, Hong Kong, 14–16 March 2012. Lecture notes in engineering and computer science pp 1214–1217

11. Zhou Z, Cheng S, Hua B (2000) Supply chain optimization of continuous process industries with sustainability considerations. *Comput Chem Eng* 24:1151–11158
12. Centikaya B, Cuthbertson R, Ewer G, Klass-Wissing T, Piotrowicz W, Tyssen C (2011) *Sustainable supply chain management : practical ideas for moving towards best practice*. Springer-Verlag, Heidelberg
13. Jonez D, Tamiz M (2010) *Practical goal programming*. International series in operations research and management science, Springer
14. Habibie A, Hisjam M, Sutopo W, Widodo KH (2012) A relationship model between supplier and manufacturer for securing availability of teak log in export oriented furniture industry with sustainability considerations. In: *Proceedings of the international multiconference of engineers and computer scientists 2012, IMECS 2012, Hong Kong, 14–16 March 2012*. Lecture notes in engineering and computer science, pp 1292–1297

Statistical Analysis of Friction Stir Weld Data

Esther T. Akinlabi and Stephen A. Akinlabi

Abstract This chapter discusses the results of statistical analysis conducted on the weld data obtained from friction stir welding of aluminium and copper. The welds were produced by varying the process parameters; the rotational speed was varied between 600 and 1200 rpm and the welding speed varied between 50 and 300 mm/min. The Statistica (version 9.0) statistical analysis software package was used to generate the scatter and surface plots relative to the experimental results obtained from the tensile testing and the FSW data. Regression analysis was also done on the weld data. It was found that the downward vertical force during the welding process has a significant effect on the Ultimate Tensile Strength of the weld and that strong relationships exist between the heat input into the welds and the measured electrical resistivities of the welds.

Keywords Analysis of scatter plot • Analysis of surface plot • Analysis of Variance • Dissimilar materials • Friction stir welding • Regression analysis

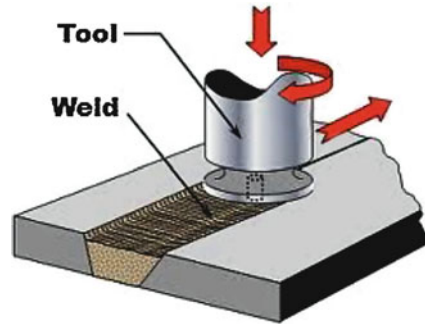
1 Introduction

Welding is a material joining process achieved by the application of heat, with or without pressure, and the addition of filler material in some cases. The applications of welding are varied and extensive from small-scale industry to large-scale industry

E. T. Akinlabi (✉) · S. A. Akinlabi
Department of Mechanical Engineering Science, University of Johannesburg,
P.O. Box 524, Auckland Park, South Africa
e-mail: etakinlabi@uj.ac.za

S. A. Akinlabi
e-mail: saakinlabi@uj.ac.za

Fig. 1 Schematic diagram of friction stir welding process [4]



and from small machines to large machineries. Welding is generally classified into two major groups—fusion welding and solid-state welding. Fusion welding process involves chemical bonding of the metal in the molten stage; and it may need a filler material, such as a consumable electrode or a spool of wire of the filler materials. Examples of the fusion welding processes include Metal Inert Gas welding (MIG), Tungsten Inert Gas welding (TIG) and Laser Beam Welding (LBW). The process may also need an inert ambience in order to avoid oxidation of the molten metal while solid-state welding is the process whereby coalescence is produced at temperatures below the melting point of the base metal without the use of any filler metal. Examples of solid-state welding processes include friction welding, Friction Stir Welding (FSW), ultrasonic welding, resistance welding, explosive welding and diffusion welding. Fewer defects are associated with solid-state welding due to the fact that the metals do not reach their melting temperatures [1, 2]. However, the base metals being joined retain their original properties, and the Heat Affected Zone (HAZ) is small when compared to the fusion welding techniques [1]. Friction Stir Welding is a variant of friction welding that produces a weld between two or more work pieces by the heating and plastic material displacement caused by a rapidly rotating tool that traverses the weld joint [3]. Figure 1 illustrates the process definitions for the tool and the work piece. The advancing side is on the right, where the tool rotation direction is the same as the tool travel direction (opposite the direction of metal flow), while the retreating side is on the left, where the tool rotation is opposite the tool travel direction (parallel to the direction of the metal flow).

The tool serves three primary functions, namely: the heating of the work piece, the movement of material to produce the joint and the containment of the hot metal beneath the tool shoulder. In FSW, the inter-relationship between the process parameters is complex; the two most important welding parameters being the tool rotational speed in a clockwise or anti-clockwise direction, and the tool traverse speed along the joint line [5]. The rotation of the tool during the welding process results in the stirring and mixing of material around the rotating pin which in turn affect the evolving properties of the weld. As such, understanding the relationship between the input process parameters and the resulting properties of the welds is important. Research studies based on material characterization to conclude on the process-property relationships [5–8] reported that the input process parameters are

found to exert significant effect on the resulting joint integrities of the welds. In attempting to further understand the process-property relationship in FSW, statistical analyses of the weld data have been conducted on similar joints of aluminium alloys [9–11]. Rajamanickam and Balusamy [9] conducted statistical analysis on the weld data obtained in FSW of 2014 aluminium alloy and concluded that the weld speed has the highest statistical influence on the mechanical properties of the welds produced. Also, Benyounis and Olabi [10] conducted a literature survey on optimization of different welding processes using statistical and numerical approaches and concluded that modeling; control of the process parameters and optimization of different welding processes can be achieved using different statistical tools. Further study was conducted by Rajakumar and Balasubramanian [11] on establishing relationships between mechanical properties of aluminium alloys and optimised friction stir welding process parameters, the optimal welding conditions to attain maximum strength for each alloy employed were identified using the Response Surface Methodology (RSM). Empirical relationships were established between the base metal mechanical properties of aluminium alloys and optimised FSW process parameters and the relationships established can be effectively used to predict the optimised FSW process parameters from the known base metal properties. The gap in the literature is that there is no report of statistical analysis on the weld data of dissimilar materials. This chapter therefore presents the statistical analysis on the weld data obtained from dissimilar friction stir welds of aluminium and copper in order to gain insight into the interaction between the process-properties of the resulting welds.

2 Theoretical Background to Statistical Analysis

The statistical analysis was done to evaluate the effects of a parameter on other results, and to establish if relationships exist amongst the parameters during the friction stir welding process. The Statistica (version 9.0) statistical analysis software package was used by the Unit for statistical support of Nelson Mandela Metropolitan University, Port Elizabeth South Africa to generate the scatter and surface plots relative to the experimental results obtained from the tensile testing and the FSW data. Regression analysis was also done on the weld data. Scatter plots are utilized to conduct a correlation analysis on the weld data. This method describes the direction and strength of a relationship between two variables. The correlation could be positive or negative [12]. Positive correlation is when an increase in values for one variable is associated with an increase in values of the other while Negative correlation is when an increase in values for one variable is associated with a decrease in the values of another variable.

The relationships that exist between the FSW process variables are explained using the strength of correlation, r value or the Pearson product-moment correlation coefficient, the p value, which indicates the statistical significance of the correlation. As a general guideline, a value of r ranging from 0.1 to 0.4 would be

classified as a weak correlation, while a value above 0.5 would be regarded as a strong correlation. Correlations close to 1.0 show a strong linear correlation and values close to zero indicate the absence of a linear relationship between the two variables. If the p value is less than 0.05, the corresponding correlation is statistically significant at the 5 % level. Multiple regression analysis was carried out on the data obtained from the FSW process and the results from the characterisation of the weld samples; in order to derive linear equations relating the dependent to the independent variables. The Statistical software was used to conduct regression analysis, Analysis of Variance, scatter plots and surface plots on the data.

In regression analysis, the regression line expresses the best prediction of the dependent variable on the independent variables. However, there is usually substantial variation of the observed points around the fitted regression line. The deviation of a particular point from the regression line (its predicted value) is called the residual value. R -square, also known as the coefficient of determination is commonly used in statistics to evaluate model fit. R -square is 1 minus the ratio of residual variability. When the variability of the residual values around the regression line relative to the overall variability is small, the predictions from the regression equation are good. In most cases, the ratio and R -square will fall somewhere between 0.0 and 1.0. The R -square value is an indicator of how well the model fits the data [13].

The adjusted R -square value is calculated by adjusting the R -square value for the number of independent variables. By reducing the number of independent variables, the adjusted R -square value will move closer to the unadjusted R -square. Usually, the degree to which two or more predictors (independent variables) explain the variation in the dependent variable is expressed in the correlation coefficient R , which is the square root of R -squared. The beta values are used to create the linear equations to predict the dependent variable (Y), when the independent variables (X_i) are specified. The generalized equation is given as [12, 13]:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 \beta_{3i} + \cdots \cdots + \beta_k \beta_{ki} + \varepsilon_i \quad (1)$$

The beta values that appear in red print in the Tables are statistically significant at the 5 % level, i.e. the p value is less than 0.05, and the corresponding predictor can be viewed as having a significant effect on the system. However, because the number of tests (sample size) is relatively small in this research study, twenty-seven parameter settings were employed; variables whose P values are bigger than 0.05, but still in the region of 0.05, could also be seen as having a noticeable effect.

The Analysis of Variance (ANOVA) is an important statistical tool for analyzing the effect of categorical factors on a response. It is a tool that separates the total variability observed within a data set. The ANOVA is also employed to determine the impact independent variables have on a dependent variable in a data.

The 2-dimensional scatter plot is a statistical technique used to visualize a relationship (correlation) between a dependent and an independent variable. Individual data points are represented in two-dimensional space, where the axes represent the variables [14].

Surface plots were also created from the weld data to aid visualization of the interrelationship that could exist between a dependent variable and two independent variables. In 3D- surface plots, the surface is defined by a smoothing technique or a defined mathematical expression fitted to the data (variables corresponding to sets of XYZ co-ordinates for subsets of data) [14].

3 Experimental Set-Up

Friction Stir Welds of 5754 Aluminium Alloy (AA) and C11000 Copper (Cu) in butt joint configurations were produced on $600 \times 120 \times 3.175$ mm thick sheets test coupons with an MTS Intelligent Stir Welding for Industry and Research Process Development System (I-STIR PDS) FSW platform at the Friction Processing Research Institute (FPRI) of Nelson Mandela Metropolitan University, Port Elizabeth South Africa. The I-STIR PDS is a robust self-contained system that is capable of welding ferrous and non-ferrous materials. The schematic diagram of the FSW platform is presented in Fig. 2. The plates were cleaned with Silicon Carbide paper to remove the oxide layer; and then they were cleaned with acetone prior to the welding procedure to remove grease, dirt and impurities.

To ensure the reproducibility of the welds, the tool holder and the tool were cooled to room temperature using compressed air after every weld. This cooling was done to prevent the tool from preheating the next weld. The deposited material on the tool pin and the shoulder was cleaned either by soaking in a solution of 20 g of Sodium Hydroxide (NaOH) and 100 ml of water for about four hours or by remachining to achieve the original tool geometry. The Cu sheet was placed at the advancing side (AS) and the AA at the retreating side (RS) during the welding process. Optimized tool displacement setting according to Akinlabi et al. [15] was used in which the tool pin was plunged in the AA and made to touch Cu during the welding process. Three different shoulder diameter tools—15, 18 and 25 mm with a constant tool pin diameter of 5 mm were used to produce the welds. The tools were machined from H13 tool steel and hardened to 52 HRC. The features of the tools were threaded pins with concave shoulders. A range of process parameter combination was employed to produce the welds. The rotational speeds of 600, 950 and 1200 rpm; and feed rates at 50, 150 and 300 mm/min were chosen to represent low, medium and high setting respectively. Other parameters like the tool tilt angle and the dwell time were kept constant at 2° and 2 s respectively. The tensile tests were conducted using a servo-hydraulic Instron 8801 tensile testing machine according to ASTM E8. An extension rate of 5 mm/min and a gauge length of 50 mm were used. The electrical resistances of the welds were measured using a Sign atone Four-Point probe meter with 1.6 mm probe spacing and the sample cross sectional area was 127 mm [2]. The output parameters i.e. the Force, Torque and the calculated heat input were taken from an average of the recorded values of the data obtained during the welding procedure from 25 to 135 mm (stable region) of the weld length. The output parameters are presented in Table 1.

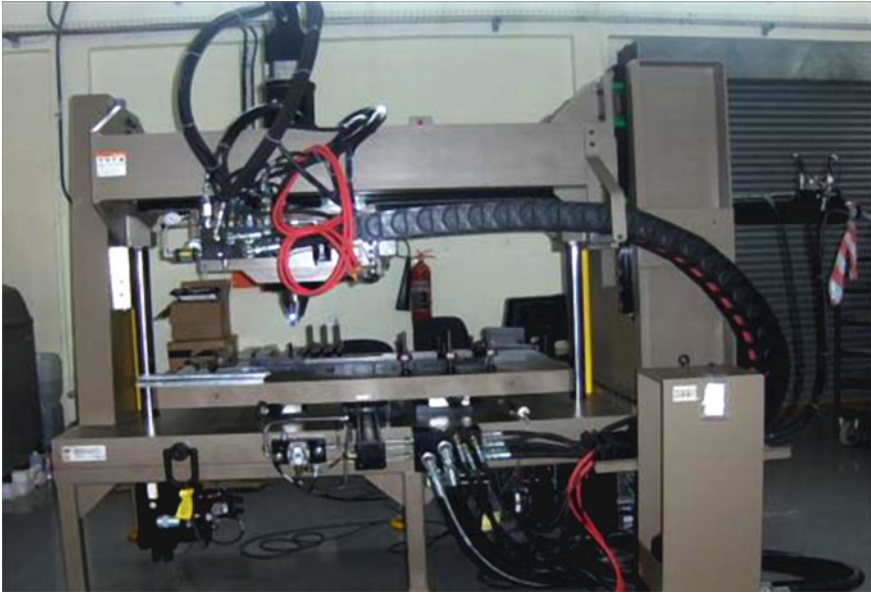


Fig. 2 The I-stir PDS FSW platform

4 Results and Discussion

In this section, we provide the results and discussion of the statistical analysis conducted on the weld data obtained from this research work.

4.1 Regression Analysis

Multiple regression analysis was conducted on the data obtained from the FSW process in order to derive linear equations relating the dependent to the independent variables. The equations derived from the multiple regression analysis are stated in Eqs. (2–6). The parameters are represented as follows: Torque—T, Feed rate—F, Spindle speed—S, and Interaction—I.

$$\text{Torque (Nm)} = 25.63624 - 0.01334 * S + 0.03330 * F - 0.00001 * I \quad (2)$$

$$\begin{aligned} \text{Heat input (KJ/mm)} = 1.13653501 - 0.00346301 \\ * F + 0.00033264 * S - 0.00000053 * I \end{aligned} \quad (3)$$

$$\text{UTS (MPa)} = 146.9909 + 0.275 * S + 0.0511 * F - 0.0001 * I \quad (4)$$

Table 1 FSW process parameters and the output data obtained

Weld No	Rotational speed (rpm)	Feed rate (mm/min)	Fx (kN)	Fy (kN)	Fz (kN)	Torque (kNm)	Q _{input} (J/mm)
S15_01	600	50	2.11	-1.17	10.14	16.50	1119.92
S15_02	600	150	2.60	-0.04	13.87	19.97	451.69
S15_03	600	300	3.02	-0.90	17.94	23.63	289.86
S15_04	950	50	1.63	-1.04	6.97	8.59	923.18
S15_05	950	150	2.02	-0.27	10.67	12.09	432.83
S15_06	950	30	2.63	-0.67	12.25	13.12	219.31
S15_07	1200	50	1.52	-0.76	7.45	8.01	1087.67
S15_08	1200	150	3.14	-0.65	11.59	12.67	524.30
S15_09	1200	300	3.54	-0.77	13.07	15.24	299.59
S18_01	600	50	2.87	-0.93	10.47	17.47	1185.71
S18_02	600	150	3.42	-0.60	14.27	20.96	474.02
S18_03	600	300	4.06	-0.09	18.53	26.20	296.37
S18_04	950	50	2.85	-0.72	11.56	12.80	1374.86
S18_05	950	150	3.12	-0.75	14.47	15.15	542.63
S18_06	950	30	3.47	-0.12	16.69	17.37	293.06
S18_07	1200	50	2.26	-0.18	10.60	12.95	1765
S18_08	1200	150	2.68	-0.23	12.51	15.12	683
S18_09	1200	300	3.16	-0.01	14.91	17.44	405
S25_01	600	50	3.26	-0.08	20.94	25.51	1731.02
S25_02	600	150	3.90	-0.64	22.49	26.12	590.92
S25_03	600	300	3.99	-0.75	24.28	28.91	326.91
S25_04	950	50	2.33	-0.09	12.14	14.74	1583.49
S25_05	950	150	4.43	0.12	26.08	29.26	934.11
S25_06	950	30	5.20	0.26	32.24	36.74	577.29
S25_07	1200	50	3.34	0.17	15.23	20.51	2067.58
S25_08	1200	150	4.54	0.48	21.00	33.02	950.15
S25_09	1200	300	5.91	-0.07	24.64	35.81	557.24

$$\text{Electrical resistivity } (\mu\Omega) = 0.10151956 - 0.00002725 * F + 0.00000081 * S - 0.0000001 * I \tag{5}$$

$$\text{UTS (MPa)} = 194.6103 + 10.8737 * F_x + 24.9234 * F_y - 3.0675 * F_z - 0.4985 * T \tag{6}$$

The linear equations outlined above can be used to predict independent variables (weld properties) when the dependent variables are known. It was observed that statistically, (F_y) and (F_z) could contribute significantly to changes in the UTS of welds. This can be explained further because the forces acting on the tool during the welding process dictate the forging force, the amount of heat input into the welds and the resulting weld defect, which can be related to the UTS of the welds.

Table 2 Univariate test of significance for UTS and FSW parameters

Effect	All groups Univariate tests of significance for UTS Sigma-restricted parameterization Effective hypothesis decomposition				
	SS	Degr. of Freedom	MS	F	P
Intercept	2045536	1	2045536	1837.593	0.000000
Spindle speed	2447	2	1223	1.099	0.338689
Feed rate	6727	2	3363	3.021	0.054951
Spindle speed*Feed rate	10113	4	2528	2.271	0.069819
Error	80148	72	1113		

4.2 Analysis of Variance (ANOVA)

The results of the analysis of variance of the data obtained for Ultimate Tensile Strength (UTS) of the weld samples are hereby presented in Table 2. Marked effects in first row are significant at $P < 0.05000$.

A noticeable effect observed where the p-value is not less than 0.05 but only marginally bigger is found in the effect of feed rate on the UTS of the welds produced in this research study. It can be interpreted that the feed rate influences the UTS of the welds produced. This can further be attributed to the fact that better weld consolidation is achieved at low feed rate due to the slow rotation of the tool which in turn influences the UTS of the weld produced.

4.3 Analysis of Scatter plots

The scatter plot of the results of the electrical resistivity and heat input to the welds is presented in Fig. 3.

The scatter plot of electrical resistivity versus heat input shows that a fairly strong relationship exists between them. It can be interpreted that the electrical resistivity increases as the heat input increases, but is limited at a certain point when the electrical resistivity becomes constant.

4.4 Analysis of Surface plots

Surface plots were created from the weld data to aid visualization of the interrelationship that could exist between a dependent variable and two independent variables [16]. The surface plot of the horizontal force, F_x against the spindle speed and the feed rate is presented in Fig. 4.

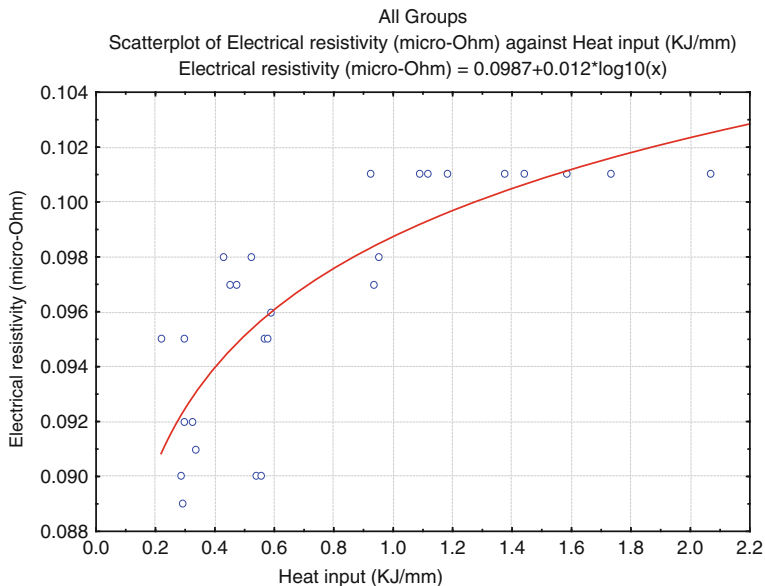
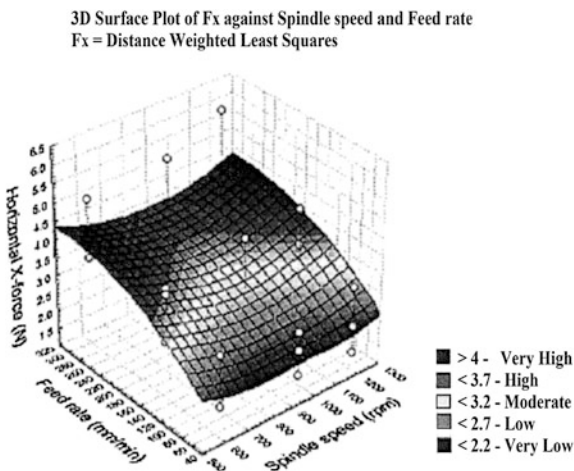


Fig. 3 Scatter plot of electrical resistivity versus heat input for all the welds

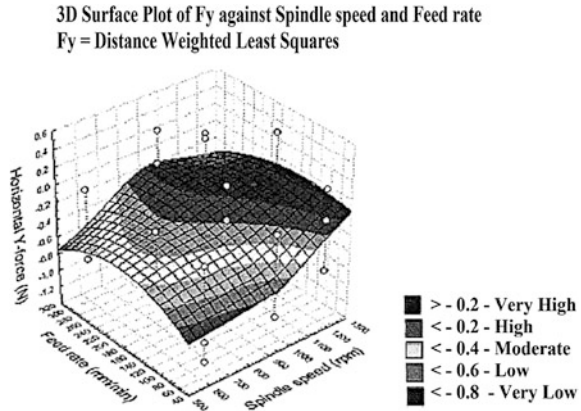
Fig. 4 Surface plot relating horizontal force (Fx), spindle speed and feed rate for all the welds



It was observed from Fig. 4 that the horizontal force (Fx) increases as the feed rate increases, while the spindle speed does not seem to have any significant effect on the horizontal force acting during the welding process. This is an important information for design purposes when considering the forces acting on the tool during the welding process.

The surface plot of the horizontal force (Fy) acting perpendicular to the (Fx) compared with the spindle speed and the feed rate is presented in Fig. 5.

Fig. 5 Surface plot relating horizontal force F_y , spindle speed and feed rate for all the welds



The surface plot revealed that the (F_y) increases as the spindle speed increases, while the feed rate does not seem to have much effect on the (F_y) acting on the tool. It stands to reason that side force (F_y) would increase slightly, because the material is pushed faster in the X and Y directions, as a result of increasing the spindle speed.

The surface plot relating vertical force (F_z), spindle speed and feed rate for all the welds is presented in Fig. 6.

From Fig. 6, it was observed that the downward vertical force F_z increases as the feed rate and spindle speed increase. This is expected because at high feed rates and high spindle speeds, the tool moves relatively fast; hence, less heat input is generated. As such, a high vertical force is practically required to ensure forging during the welding process.

The surface plot relating torque (T), spindle speed and feed rate for all the welds produced is presented in Fig. 7.

The trend observed in the torque values shown in the surface plot (Fig. 7) was that the torque increases as the feed rate increases, but it decreases as the spindle speed increases. The explanation given earlier on the relationship between vertical force, spindle speed and feed rate is also related to this case, since a linear relationship exists between vertical force and torque; that is to say, an increase in the downward vertical force gives an increase in torque values. Hence, it is revealed in this plot, that the feed rate plays a significant role in the resulting torque values compared with the spindle speed.

The surface plot relating heat input, spindle speed and feed rate for all the welds are presented in Fig. 8.

From the surface plot relating heat input to the process parameters (Fig. 8), it was observed that the heat input into the welds increases as the feed rate decreases, but not linearly. The spindle speed does not have a significant effect on the heat input. The explanation for this is based on the fact that the tool moves slowly at low feed rates; hence most of the heat generated is contained in the welds.

Fig. 6 Surface plot relating vertical force (F_z), spindle speed and feed rate for all the welds

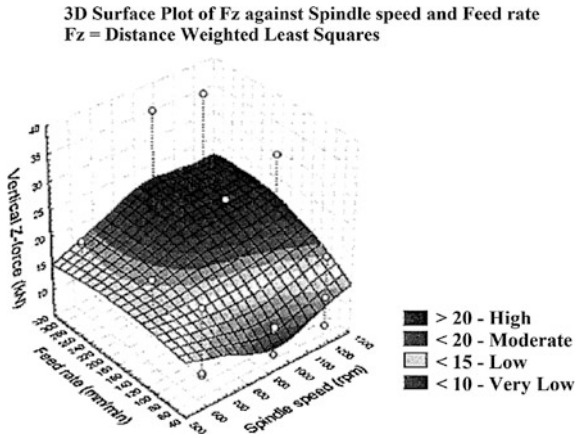


Fig. 7 Surface plot relating torque (T), spindle speed and feed rate for all the welds

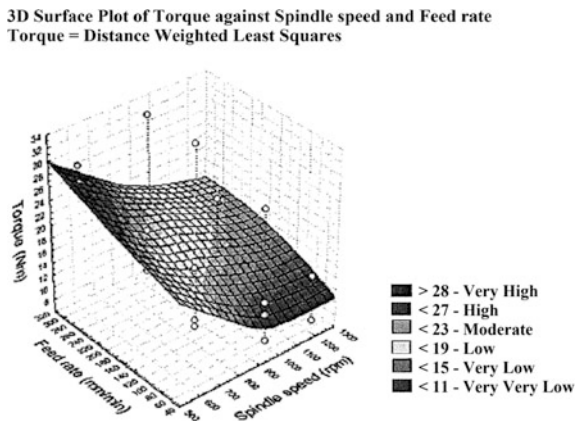


Fig. 8 Surface plot relating heat input, spindle speed and feed rate for all the welds

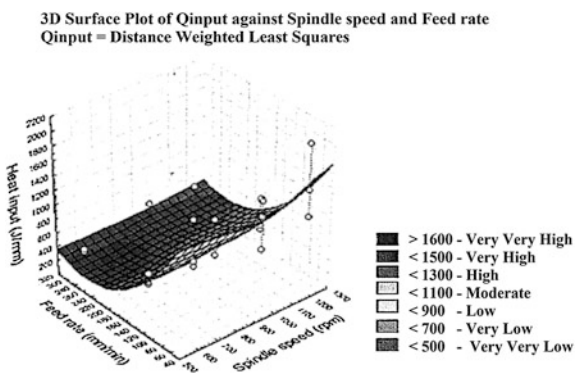


Fig. 9 Surface plot relating UTS, spindle speed and feed rate for all the welds

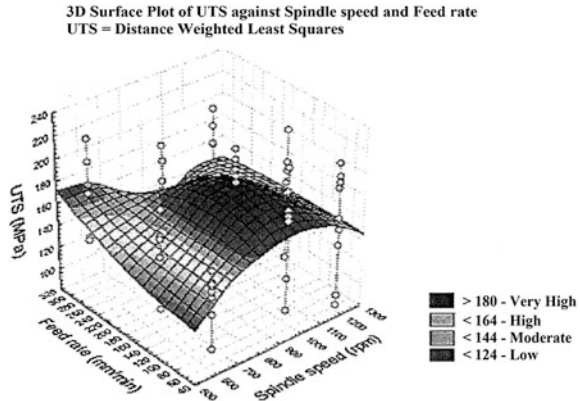
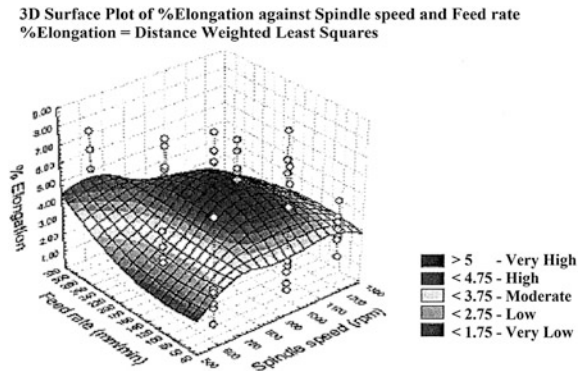


Fig. 10 Surface plot relating percentage elongation, spindle speed and feed rate for all the welds



The surface plot relating UTS, spindle speed and feed rate for all the welds is presented in Fig. 9.

The surface plot relating the UTS, spindle speed and feed rate of the entire weld data (Fig. 9) revealed that the UTS increases as the feed rate decreases, but decreases slightly at high spindle speeds. With respect to the entire weld matrix considered in this research, it can be said that the optimum weld setting with respect to the UTS, is 950 rpm and 150 mm/min based on the statistical analysis.

Figure 10 presents the surface plot relating percentage elongation, spindle speed and feed rate for all the welds.

The top darkest region of the surface plot relating the percentage elongation and the process parameters (Fig. 10) is considered statistically significant and looked similar to that of the UTS earlier discussed. Hence, the trends observed in both properties are similar.

5 Conclusion

The statistical analysis of the weld data obtained from dissimilar friction stir welds of aluminium and copper have been reported and discussed. Linear equations relating the dependent and independent variables in FSW process were achieved. The Analysis of Variance revealed that the downward vertical force, (F_z) has a significant effect on the UTS of the welds. There is also an indication of a strong relationship between the electrical resistivity and the heat input into the welds. Based on the statistical analysis, the optimal weld setting with respect to the UTS is the weld produced at 950 rpm and 150 mm/min. It can be concluded that the input process parameters in FSW play a very significant role in determining the joint integrity of the resulting weld.

Acknowledgments The authors wish to thank Dr T. Hua and Mr L. Von Wielligh for operating the FSW platform, Prof. A. Els-Botes for the opportunity to work in her research group, Mr G. C. Erasmus for assistance in the Metallurgy lab, Dr P. Jacques of the statistics department; all of Nelson Mandela Metropolitan University, Port Elizabeth, South Africa. The financial support of ESKOM for the Tertiary Education Support Program (TESP) award and the University Of Johannesburg research fund are also acknowledged.

References

1. O'Brien A, Guzman C (eds.) (2007) American welding society welding handbook welding processes part 2 vol 3 Ninth Edition Miami, American Welding Society
2. Akinlabi ET, Els-Botes A, McGrath PJ (2011) Analysis of process parameters and their effect on defect formation of dissimilar friction stir welds, in international friction processing seminar 2011 at Nelson Mandela Metropolitan University, Port Elizabeth, 31 Aug–1st Sept 2011
3. Thomas WM, Nicholas ED, Needham JC Murch MG, Templesmith P, Dawes CJ (1991) Improvements relating to friction welding international patent application, PCT/GB92/02203 (Patent) December 1991
4. Friction stir welding process. Retrieved from www.engineering.und.edu/research/aemc/fswprocess.phpo
5. Reynolds AP (2003) Friction stir welding of aluminium alloys. In: Totten GE, MacKenzie DS (eds.) Handbook of aluminium, volume 2 Alloy production and materials manufacturing. Marcel Dekker, New York pp 579–700
6. Lim S, Kim S, Lee C, Kim S (2004) Tensile behaviour of friction stir welded Al 6061-T651. Metall Mater Trans A 35:2829–2835
7. Abdollah-Zadeh A, Saeid T, Sazgari B (2008) Microstructural and mechanical properties of friction stir welded aluminium/copper lap Joints. J Alloy Compd 460:535–538
8. Akinlabi ET, Els-Botes A, McGrath PJ (2011) Effect of travel speed on joint properties of dissimilar metal friction stir welds. In: Proceedings of 2nd international conference on advances in engineering and technology (AET), Uganda. Jan 30–Feb 1
9. Rajamanickam N, Balusamy V (2008) Effects of process parameters on mechanical properties of friction stir welds using design of experiments. Indian J Eng Mater Sci 15:293–299
10. Benyounis KY, Olabi AG (2008) Optimization of different welding processes using statistical and numerical approaches—a reference guide. Adv Eng Softw 39:483–496

11. Rajakumar S, Balasubramanian V (2012) Establishing relationships between mechanical properties of aluminium alloys and optimised friction stir welding process parameters. *J Mater Des* 40:17–35
12. Greasley P (2008) *Quantitative data analysis using SPSS. An introduction for health and social science*. Open University Press, McGraw-Hill Education, pp 77–85
13. Ashenfelter O, Levine PB, Zimmerman DJ (2003) *Statistics and econometrics: Methods and applications*. Wiley, New York pp 158–165
14. Multiple Regressions. Available from: <http://www.statsoft.com/textbook/multiple-regression/> [Accessed on Aug 2010]
15. Akinlabi ET, Els-Botes A, Lombard H (2010) Effect of tool displacement on defect formation in friction stir welding of aluminium and copper. In: *Proceedings of the 8th international friction stir welding symposium, Hamburg, Germany, May*, pp 18–20
16. Akinlabi ET, Akinlabi SA (2012) Friction stir welding of dissimilar materials—statistical analysis of the weld data. In: *Lecture notes in engineering and computer science: proceedings of the international multi conference of engineers and computer scientists 2012, IMECS 2012, 14–16 March 2012, Hong Kong*, pp 1368–1373

Development and Evaluation of an Online Casebook for Teaching and Learning of Industrial Accidents

Ke Chen and Alan Hoi Shou Chan

Abstract The use of web based learning is becoming more common in educational institutes nowadays. An online casebook for improving the learning and teaching processes of industrial accident analysis was developed and evaluated in this project. The online casebook was developed with six main components: Analysis Model, Nature of Accident, Self-Checking, Glossary, Cost Estimation, and Industrial Statistic Analysis. Comprehensive ergonomics glossary, real life case studies, and some related web resources were provided in the courseware. It also contains self-checking for testing users' general qualitative concepts and terminology and further strengthens their understanding of the subjects. The effectiveness of the online courseware was evaluated with a questionnaire survey with a group of 92 students. Results showed that the online casebook allows students to acquire the knowledge and techniques of industrial accident analysis in a methodical manner. This courseware is expected to be useful for the students, occupational health and safety professional, and safety engineers in enhancing their knowledge on occupation health and safety. The courseware also could serve as a reference for companies, education institutions, and governments when designing web-based training course.

Keywords Industrial accident · Interactive training · Online casebook · System usability scale (SUS) · Teaching and learning · Web-based training

K. Chen (✉) · A. H. S. Chan
Department of Systems Engineering and Engineering Management,
City University of Hong Kong, Hong Kong
e-mail: kechen2@gapps.cityu.edu.hk

A. H. S. Chan
e-mail: alan.chan@cityu.edu.hk

1 Introduction

Occupational health and safety is a multidisciplinary discipline aiming at protection of the health of workers by preventing and controlling occupational diseases and accidents and by eliminating occupational factors and conditions hazardous to health and safety at work, and development and promotion of healthy and safe work, work environments and work organizations [1]. Industrial accidents can cause serious injury or wrongful death, and also can be extremely costly and have undesirable impacts on the morale, image, and productivity in an organization. In recent years, the Hong Kong government has shifted the emphasis from law enforcement to promoting safety management. As stated by the Commissioner for Labour [2], self-regulation in respect of reducing the risks at work by the proprietors and their workforce is the key to attaining long-term improvements in safety standards. However, it seems that the safety management among many organizations does not reach a satisfactory level since there are still many industrial accidents that cause serious injuries and deaths. It was reported by the Hong Kong Labour Department [3] that in 2010, the number of industrial accident in all industries stood at 14015, an increase of 3.1 % over 2009. There is definitely a need to make more careful planning and management of safety and health issues so as to improve safety and health management in all different industries. As accident investigation helps to find out the causes of any accident or incident for developing prompt arrangements and preventing future recurrence, the importance of knowledge and skills in industrial accident analysis should always be included and emphasized in engineering education. Obviously, identification of the root causes requires a higher level of knowledge, vision, and analytical power of the investigator. And usually industrial hazards could be attributed to the factors of work design, workplace design, engineering design, system design, product design (or procurement), maintenance program, personal physical and psychological factors, etc.

The International Labour Organization has identified that education and training are vital components for safe working environment [1]. With an understanding of the importance of ensuring occupational health and safety, City University of Hong Kong has provided Occupational Health and Safety Management as a core course for students in industrial engineering, manufacturing engineering, and industrial design programs. It is expected that students who take this course will be able to learn techniques that are used to prevent and investigate industrial accidents and to develop the skills required for accident prevention and analysis after taking the course. In recent years, the teaching style of this course has been increasingly changing from typical use of text based materials to an intensive use of computer aids like PowerPoint presentation. However, the use of computer aids for teaching is just not enough. A course conducted in classroom usually covers a considerable amount of material, but with no complementary time for discussion of real life case studies. Students may leave the classroom with a blurred

impression of techniques and some technical jargons. It is expected that using of web based learning kit performing in an interactive environment will improve the learning and teaching processes.

The Internet and Web can be an effective medium for the delivery of education and training because of its flexibility, timeliness, and breadth of access. The idea of using the Internet or Web based training for education delivery is not new in today's educational context. There is a tendency that education institutions are placing more course material online to supplement classroom training situations. Wyld and Eklund [4] pointed out that the Internet has begun to have a significant role in education and change the nature of teaching and learning, since it provides learners with a wide range of learning opportunities and experiences. Jain et al. [5] also highlighted that knowledge-based, generative, and intelligent instructional environments are just beginning to appear on the Internet, and Web was described as a widely accessible source, an information bank, and a method of transmitting knowledge to students.

Web-based learning is the delivery of and access to a coordinated collection of learning materials over an electronic medium using a Web server to deliver the materials, a Web browser to access them and the Transmission Control Protocol/Internet Protocol (TCP/IP) and HyperText Transfer Protocol (HTTP) protocols to mediate the exchange [6]. The advantages of Web-based Training (WBT) have been recognized in previous literatures. Hannum [7] stated that WBT has many different advantages and one of the advantages is the cross-platform compatibility of the software. By using WBT, only one version of a course would be needed to maintain and develop. A training organization does not have to maintain multiple versions of courses for different computer platforms and operation systems. Jolliffe et al. [6] pointed out that web-based learning allows for a learner-centered delivery strategy that can take into account the individual differences in learning. Another well-known advantage of WBT is that it provides an interactive learning environment. Previous studies have indicated that interactive learning style could facilitate the learning process. Steed [8] noted that the interactive nature of the presentation, combining multimedia (text, graphics, audio and video) in an interactive environment enables the development of courses that simulate real-life scenarios and provide immediate feedback. Learners retain more course material offered on Web than in classroom. McIntyre and Wolff [9] also indicated that one of the advantages of interactivity in a Web environment is the capability to engage by providing rapid, compelling interaction and feedback to the users. They also noted that interactive multimedia technology can help motivate learners by providing information in a form that is concrete and perceptually easy to process. The study of De Leeuwe et al. [10] found that animations are clearly preferred above static figures, and the learning-access-level of the learning materials are rated significantly higher when the learning materials are illustrated by animations. It is thus expected that the replacement of static materials with animated displays in the web teaching and learning could enrich the course materials to be more illustrative and interactive. It is also envisaged that the interactive web based training should provide alternative better means, and adds zest to the current seemingly monotonous teaching process. Therefore, in this project, an

online casebook is developed for improving the learning and teaching of Occupational Safety and Health Management.

The aim of this project was to produce a coherent and intriguing World Wide Web online casebook, which has genuine occupational accident cases for arousing students' learning interests and facilitating students' transfer of learning from knowledge in industrial safety to real-life situations in the health and safety profession. Three major objectives were identified for achieving the project aim. The first objective was the systematic sourcing and compilation of teaching materials, all the facts and information for selected real life cases for inclusion in the online casebook. The second objective was to develop the software aspect of this online casebook for delivering teaching materials and accident techniques in an interactive mode. The third objective was to evaluate the effectiveness of the online casebook.

2 Clients of the Project

In this project, the clients are basically the undergraduates and postgraduates who take the courses of work design and ergonomic workplace design in the City University of Hong Kong. The clients are expected to participate in the classroom lectures before using the courseware. This online casebook was developed with the aim of supplementing and enhancing their learning process outside the classrooms.

3 Methodology

Five processes were done to achieve the project aim. The five processes were review and selection, material compilation and preparation, production, project construction, and operation and maintenance.

3.1 Review and Selection

The pre-development stage was to intensively review related knowledge like web page design theory, computing language, computer graphics formats, modern training approaches, proper software, appropriate technologies for broadcasting, and concepts of occupational health and safety. The necessary skill, concepts and formats of each of these components were selected after the review.

Shin and Shin [11] stated that Synchronized Multimedia Integration Language (SMIL) offers an integrating format representing how various media objects are positioned spatially and combined temporally using a collection of XML elements and attributes for multimedia delivery over the Web. With the use of SMIL, the

web site was able to define and synchronize multimedia elements like video, sound, and still images for web presentation and interaction. Steele et al. [12] also noted that SMIL permits web page send multiple movies, still images, and sound separately in a well time-controlled and coordinated way. Each media object is accessed with a unique Uniform Resource Locator (URL), which means that presentations can be made of objects arriving from more than one place, and that objects can easily be reused in multiple presentations.

Considering the glossary's large database management, implementation, browsing and further development, MySQL and PHP (Hypertext Preprocessor) were chosen. According to Welling and Thomson [13], MySQL is a very fast, robust, relational database management system (RDBMS), which uses a standard database query language Structured Query Language (SQL). It would enable the user to efficiently store, search, sort, and retrieve data. MySQL has high performance and is easy to configure and learn. They also noted that MySQL server controls access to the data to ensure that multiple users can work with it concurrently, to provide fast access to it, and ensure that only authorized users can obtain access. Apart from the MySQL, Welling and Thomson [13] also noted that PHP is a server-side scripting language designed specifically for the Web. PHP code would be executed each time the page visit within an HyperText Markup Language (HTML) page, and it generates HTML or other output that the visitor preferring to browse. It was pointed out that PHP has many strengths including high performance, interfaces to many different database systems, and built-in libraries for many common Web tasks. Davis and Philips [14] pointed out that when working hand-in-hand, PHP and MySQL serve as the standard for the rapid development of dynamic and database-driven websites.

3.2 Preparation, Compilation and Collation of Teaching Materials and Case Studies Preparation

In this stage, a systematic preparation, compilation and collation of teaching materials were done. The three levels of quizzes and real life case studies were also developed. In the courseware, most of the interfaces were designed based on the materials found in the common occupation safety and health textbooks, websites, magazines, and newspapers.

3.3 Production

In this stage, the concepts, technologies, and programming skills selected in the first phase were applied on the materials of the online casebook. The computer-based training (CBT) and web-based training (WBT) guidelines were reviewed

and applied into the design. Sung and Ou [15] stated that WBT is a form of computer-based training, in which the training material resides on WWW pages. Web-based training (WBT) is an approach for distance learning in which CBT is transformed by the technologies and methodologies of the World Wide Web, the Internet and intranets [16]. In particular, the practical guide developed by Cook and Dupras [17] was taken as a reference for developing effective web-based learning. It is born in mind that effective webpages should be clear, concise, and consistent [18].

3.4 Construction

In this phase, the ideas and design from the preceding phase was transferred to the web pages with the JavaScript, HTML, Common Gateway Interface (CGI), MYSQL, and Flash.

Gillani and Relan [19] noted that JavaScript works quite effectively when it is embedded into HTML codes. Objects on the page, for example, buttons and texts, can respond to user action directly because the codes that run the document are downloaded into the client's computer rather than coming from the server's script. Schaller et al. [20] stated that Flash allows developers to go far beyond the text and image foundations of the Web to create sophisticated applications, full-fledged games, rich multimedia, and complex interactivity. The materials were then uploaded to the server for public browsing. The documents and guidelines for successors for further development of the web were also prepared.

3.5 Operation and Maintenance

Operation and maintenance are very important for developing informative learning websites. Operation refers to the frequently updating of the website while maintenance means the provision of supports. The maintenance of websites involves revising, editing existing web pages to keep them up to date, checking and correcting errors, and adding new web pages periodically. Without the maintenance, websites would become stagnant. Since a web page needs continuous updating and reviewing, the structure of the pages was designed for easy updating and information addition.

4 Results

The courseware can be accessed with a web browser on an Intranet or the World Wide Web (<http://personal.cityu.edu.hk/~meachan/IAA>). It incorporates with a

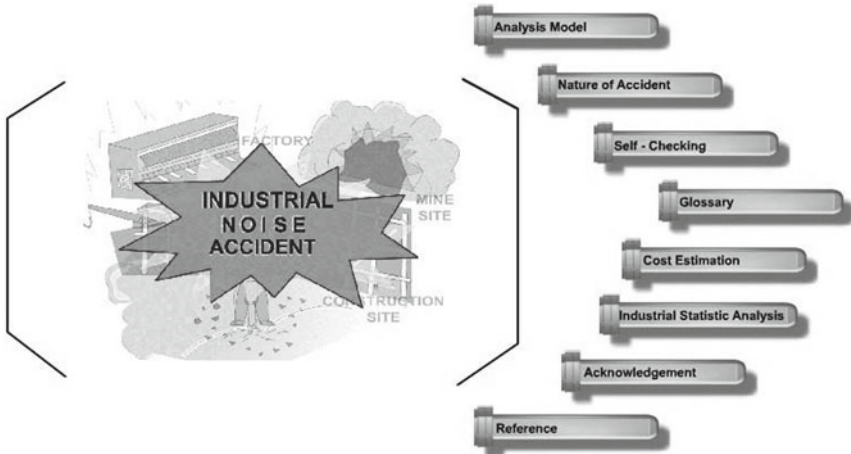


Fig. 1 A snapshot of an interface of the online casebook

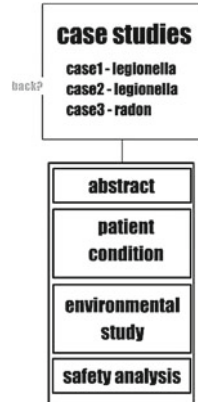
combination of text, sound, video, graphics, animation, and database. The materials and topics in the courseware are major health and safety issues in local industries. The systematic approach for preventing and analyzing industrial accidents was adopted for training in the courseware. The web page is developed with six main components, namely analysis model, nature of accident, self-checking, glossary, safety cost estimation, and industrial accident analysis (Fig. 1). The self-checking component allows users to acquire the knowledge, principles, and analysis assessment in a methodical manner. Besides, a comprehensive glossary containing all the related terms is compiled and appropriate web links with related sites are availed to the students. This online casebook is expected to allow students to learn more efficiently and understand better the subjects through series of self paced activities. It also enables students to evaluate themselves on the effectiveness of learning through series of three levels of quizzes. Through the study of the real life cases provided in the courseware, students are able to get familiar with more practical situations on industrial accident prevention and analysis.

The effectiveness of this online casebook is also evaluated by a questionnaire survey to users.

4.1 Analysis Model

Accident investigation is one of the important elements of a safety management system, since it helps to find out the causes of any accident or incident for developing prompt arrangements and preventing future recurrence. In this online

Fig. 2 A snapshot of an interface found in the topic of indoor air quality



courseware, six analysis models are provided for helping in the Industrial Accident Analysis. The analysis models are Pilot Error, Discrete Action, Petersen Accident-Incident Causation Model, Human Error Model, Heinrich's Domino Theory, and PROACT Method.

4.2 Nature of Accidents

In this component, eight major health and safety issues in local industries are included, which are Industrial Noise, Mechanical Accident, Indoor Air Quality (IAQ), Office Safety, Fire/Explosion, Electronic Accident, Confined Space and Chemical Accident, and Falls. Figure 2 is a snapshot of interface in this web site.

In the Industrial Noise, Mechanical and Electronic Accident components, common causes and safety precautions of these three components are discussed. In the IAQ component, subjects like About IAQ, How IAQ affects health, Ventilation, Recommended IAQ levels, and Organization about IAQ are developed. In the Office Safety component, information about Office workstation, Office equipment, and Work Posture are provided. In the Fire/Explosion component, subjects like Introduction of fire, and Safety information of Fire/Explosion are developed. In the Confined Space and Chemical Accident component, there are introduction of and information about Confined Space and Chemical Accident. In the Falls component, there are subjects like Introduction of fall, and Falls safety regulation. For each of the topics, case studies are provided. Khan [21] noted that case studies of real situations can be used in a Web-based course to engage learners in the problem-solving tasks in which they can be asked to identify and solve problems.

4.3 Self-Checking

Self-test is an interactive component in each lesson provided to stimulate learners' thought and action, motivate them to learn and help them to know whether they understand the main concepts and ideas in the lesson [22]. In this online casebook, a Self-Checking component is developed and two major parts are involved—Analysis Model and Nature of Accidents. It attempts to test users' general qualitative concepts and terminology and further strengthens their understanding of the subjects.

In the Analysis Model component, three case studies are provided. For each case, a multiple choice test is provided. According to Davies and Marriott [23], multiple-choice questions have advantages in that they are relatively quick for students to answer, so a wide range of ideas could be tested. Immediate feedbacks are also given.

For the Nature of Accidents part, eight topics are provided—Industrial Noise, Mechanical Accident, Indoor Air Quality (IAQ), Office Safety, Fire/Explosion, Electronic Accident, Confined Space and Chemical Accident, and Falls. Users are required to answer questions after reading the related case studies for different topics. There are three catalogues in each topic. The first catalogue provides case description, and completed case analysis. It also provides a series of simple true/false statements with immediate feedback and score counting, and matching games with immediate feedback. For the second catalogue, multiple choice questions, shooting game or hooking game are set and arranged in order of increasing difficulty. Quinn [24] noted that making things interesting can make them more efficient and stated that e-learning stimulation games can engage learning. In the third catalogue, it presents the case background information, few analyses for the case, and the highest level of multiple choices, or fill in the blank. Users can check the suggested answer after completing the tasks. In this catalogue, students can learn through action and examine their ability of application of knowledge and principles in modern industrial accident investigation and analysis.

4.4 Accessories

The accessories component includes all the supplementary tools and learning aids for better understanding of the course materials. It includes the Glossary, Cost Estimation and Industrial Statistic Analysis. The Glossary is an online database containing more than 3,000 Occupational Health and Safety terms with brief explanations. In the Cost Estimation component, introduction and concept of cost estimation are provided. Different cost estimation models are also introduced. The advantages and limitation of each model are discussed. From this component, user can learn how to calculate the industrial accident cost through different models which are suitable for Hong Kong. For the Industrial Statistic Analysis component,

the reason of analysis of the industrial accident statistic is mentioned. The definitions and detail information of occupational disease, occupational and industrial accidents are also provided. Past analysis of industrial accident statistic in Hong Kong is provided as an example. The Industrial Statistic Analysis can serve as a reference for safety program planners to develop or to design a new safety training program for workers.

5 Educational Benefits

The effectiveness of this online courseware was evaluated by a questionnaire survey addressing the anticipated benefits. The questionnaire consisted of the ten questions of System Usability Scale (SUS) and four additional questions (Appendix). Each question is a statement and answers are given on a five-point scale from “Strongly Disagree” to “Strongly Agree”. The total score of the questionnaire is ranged from 10 to 100. In a study of Tullis and Stetson [25], five questionnaires for assessing the usability of a website were compared. It was found that the SUS questionnaire yielded among the most reliable results across sample sizes. Finstad [26] stated that SUS is a commonly used, freely distributed, and reliable questionnaire with ease of administration and scoring.

A group of 92 students who had attended the classroom lectures on the topics of Analysis Model and Nature of Accidents participated in the questionnaire survey. They were asked to try using the courseware for 90 min and answer the questions of the survey. Seventy users (76 %) completed and returned the questionnaires. The results revealed that an overall SUS score of 75.5 was obtained with the use of the courseware, indicating that an above the average satisfaction level was achieved by the users.

There were four additional questions addressing the educational benefits brought to the students. The average score of the additional question numbered 11 was 4.2 and indicated that students can understand and learn more effectively and efficiently the ergonomics issues of industrial accident analysis. The average score of the additional question numbered 12 was 3.9 and presented that students can learn the techniques of industrial accident analysis through a systematic and comprehensive task analysis procedure. The mean score of question numbered 14 was 3.0 and suggested that through action learning, students can learn in a flexible and interactive environment of the concepts and principles of industrial accident analysis with some structured real life cases.

Question 14 asked participants whether students thought they could evaluate the effectiveness of their learning process through the series of self-conducting checkpoints and quizzes. The equal to average score showed a neutral response and implied that there are still rooms for improving the design for allowing the participants to evaluate their learning effectiveness. In view of the increasing awareness of the importance of health and safety issues in Hong Kong, this online casebook is not only useful to those studying occupational health and safety but

also benefits all other related professionals who are responsible for preventing industrial accidents and protecting workers in their employment from risks. The development of this online casebook was acclaimed of supporting the University's mission of providing its students with quality higher education and simultaneously responding to local and regional needs.

6 Conclusion

In this project, an online casebook was developed successfully. Eight main areas of occupational health and safety are focused, namely Industrial Noise, Mechanical Accident, Indoor Air Quality (IAQ), Office Safety, Fire/Explosion, Electronic Accident, Confined Space and Chemical Accident, and Falls [27]. The results of questionnaire indicated that users achieved an above the average satisfaction level for the use of the courseware, and it also indicated that the courseware did bring certain educational benefits to the users. Given the educational needs and systematic approach mentioned above, the multimedia online casebook brings many educational benefits to the target students, except that there are rooms for the improvement in the self evaluation of learning effectiveness by the users. It is hoped that the online casebook would help students to have an in-depth learning of the techniques used for preventing and investigating industrial accidents. It is also hoped that engineers and occupational health and safety professionals could find this courseware helpful in providing them a range of information for accident analysis and helping them to acquire knowledge on occupational health and safety.

7 Recommendation

It is recommended that more topics of occupation health and safety could be provided in the courseware, for example, heat and temperature hazards, vibration and pressure hazards. Besides, it is recommended that more case studies could be given to the students for their self-learning. Moreover, it is suggested that communication tools, for example, discussion board, should be developed in the courseware to allow the discussion of any problems encountered in the online casebook or in the online materials among students. Past studies have emphasized the importance of social interaction among learners. Collaboration helps learners validate their learning experiences, and requires a level of articulation that promotes collective knowledge building and a deeper understanding of what is being studied [28]. Vonderwell [29] pointed out that social interaction among learners plays an important part in the learning process and can have a significant impact on learning outcomes.

Appendix

Evaluation of online casebook on industrial accident analysis

Please visit the above web site (<http://www.cityu.edu.hk/meem/ergoweb>) and try using the self-learning kit for 90 min before filling in this questionnaire. Put a tick in the appropriate box. After completion, please return to the mailbox of Dr Alan Chan.

	Strongly disagree				Strongly agree
	1	2	3	4	5
1. I thought I would like to use this courseware frequently					
2. I found the courseware unnecessarily complex					
3. I thought the courseware was easy to use					
4. I thought that I would need the support from others to use this courseware					
5. I found the various functions in this courseware were well integrated					
6. I thought there was too much inconsistency in this courseware system					
7. I would imagine that most people would learn to use this courseware very quickly					
8. I found the courseware very cumbersome to use					
9. I felt very confident using the courseware					
10. I needed to learn a lot of things before I could get going with this courseware					
11. I could understand and learn more effectively and efficiently the ergonomics issues of industrial accident analysis					
12. I could learn the techniques of industrial accident analysis through a systematic and comprehensive task analysis procedure					
13. Through action learning, I could learn in a flexible and interactive environment of the concepts and principles of industrial accident analysis with the structured real life cases					
14. I could evaluate the effectiveness of my learning process through the series of self-conducting checkpoints and quizzes					

References

1. International Labour Organization (2001) Fundamental principles of occupational safety and health. http://www.ilo.org/wcmsp5/groups/public/@dgreports/@dcomm/@publ/documents/publication/wcms_093550.pdf
2. The Government Information Centre (2000) Proposal to enhance the power of safety officers. <http://www.info.gov.hk/gia/general/200001/13/0113191.htm>
3. Hong Kong Labour Department Occupational Safety and Health Statistics Bulletin (2011) Issue no. 11. <http://www.labour.gov.hk/eng/osh/pdf/Bulletin2011.pdf>

4. Wyld S, Eklund J (1997) A case study of communication technology within the elementary school. *Aust J Educ Technol* 13(2):144–164
5. Jain A, Kensk K, Noble D (1998) An interactive web-based teaching tool for simplified 3D analysis of solar rhythms. *Automat Constr* 8(2):181–194
6. Jolliffe A, Ritter J, Stevens D (2001) *The online learning handbook: developing and using web-based learning*. Kogan Page, London
7. Hannum W (2001) Web-based training: advantages and limitations. In: Khan BH (ed) *Web-based training*, Educational Technology Publications, New Jersey, pp 13–20
8. Steed C (1999) *Web-based Training*. Gower, Aldershot
9. McIntyre D, Wolff F (1998) An experiment with WWW interactive learning in university education. *Comput Educ* 31(3):255–264
10. De Leeuwe MFJ, Portier SJ, Martens RL (1999) Using animations in an interactive learning environment: does it help? In: den Brinker BPLM, Beek PJ, Brand AN, Maarse FJ, Mulder LJM (eds) *Cognitive ergonomics, clinical assessment, and computer-assisted learning*, Swets and Zeitlinger, The Netherlands, pp 247–256
11. Shin D, Shin D (2002) Design and implementation of the synchronized multimedia integration language (SMIL) player. *IEEE Trans Consum Electron* 48(3):575–578
12. Steele R, Lubonski M, Ventsov Y, Lawrence E (2004) Accessing SMIL-based dynamically adaptable multimedia presentations from mobile devices. In: *Proceeding of the international conference on information technology: coding and computing (ITCC'04)*, Las Vegas, vol 2, pp 410–415
13. Welling L, Thomson L (2003) *PHP and MySQL Web development*, 2nd edn, Sams Publishing, Indiana
14. Davis ME, Philips JA (2006) *Learning PHP and MySQL*. O'Reilly Media, Inc., CA
15. Sung WT, Ou SC (2001) Integrating network CAD and VR into the design and development of web-based computer graphics learning materials. *Malays J Comput Sci* 14(2):47–63
16. Zhang R, Zhang B, Zhu J, Huang H (2008) Development of multi-video based virtual classroom and its application in english as second language learning. *2008 international symposium on computer science and computational technology*, vol 1, 175–179
17. Cook DA, Dupras DM (2004) A practical guide to developing effective web-based learning. *J Gen Intern Med* 19(6):698–707
18. Krug S (2000) *Don't make me think! a common sense approach to web usability*. New Riders, Indianapolis
19. Gillani BB, Relan A (1997) Incorporating interactivity and multimedia into web-based instruction. In: Khan BH (ed) *Web-based instruction*, Educational Technology Publications, Inc., New Jersey, pp 231–238
20. Schaller DT, Allison-Bunnell S, Chow A, Marty P, Heo M (2004) To flash or not to flash? Usability and user engagement of HTML versus flash, archives and museum informatics, *Selected proceedings of museums and the web*, pp 197–208
21. Khan BH (2001) *Web-based training*. Educational Technology Publication, Inc., New Jersey
22. Sadik A (2004) The design elements of web-based learning environments. *The Int J Instructional Technol Distance Learning* 1(8):27–42
23. Davies N, Marriott J (2010) Assessment and feedback in statistics. In: Bidgood P, Hunt N, Jolliffe F (eds) *Assessment methods in statistical education: an international perspective*, John Wiley & Sons Ltd, UK, pp 3–20
24. Quinn CN (2005) *Engaging learning: designing e-learning simulation games*. John Wiley & Sons, Inc., Pfeiffer
25. Tullis TS, Stetson JN (2004) A comparison of questionnaires for assessing website usability. In: *Proceedings of the 13th international usability professionals association (UPA) 2004 Conference*, Minneapolis, USA, June 2004
26. Finstad K (2006) The system usability scale and non-native english speakers. *J Usability Studies* 1(4):185–188
27. Chan AHS, Chan KL, Chen K (2012) An online casebook on industrial accident analysis. In: *proceedings of the international multicongference of engineers and computer scientists 2012*,

- IMECS 2012, Hong Kong, 14–16 March 2012. Lecture notes in engineering and computer science, pp 1434–1438
28. Grabinger RS, Dunlap JC (2000) Rich environments for active learning: a definition. In: Squires D, Conole G, Jacobs G (eds) *The changing face of learning technology*, University of Wales Press, Cardiff, 2000, pp 8–38
 29. Vonderwell S (2003) An examination of asynchronous communication experiences and perspectives of students in an online course: a case study. *Internet and Higher Education* 6(1):77–90

Comparison of Response to Visual and Auditory Digit Cues in Man–Machine System

Annie W. Y. Ng and Alan H. S. Chan

Abstract The purpose of this paper was to compare response time to visual and auditory digit cues in the context of man–machine system. The effects of age, gender, education level, time spent on computer in daily life, left/right finger, and choice alternative on response time were also examined. A total of 69 right-handed Chinese participants took part in the visual and auditory stimuli tests. The result showed that the auditory response time was significantly shorter than that for visual cues. For both visual and auditory cues, in general, the response time decreased with an increase of age up to the 21–30 years, and thereafter it increased gradually with an increase of age. Females were found to respond faster than males. The response of tertiary and secondary education groups was faster than that of primary education group. Besides, the longer the time spent on computer in daily life, the shorter was the response time. In addition, the right finger response time was shorter than the left finger response time. The response on single-choice task was the fastest, followed by two-choice task and then four- and eight-choice tasks. Implications of the results on the design of alerting cue and man–machine system were discussed. The findings of this study would act as a useful reference for engineers and designers to realize how the visual and auditory modality channels could interfere the users, so as to design a more user-friendly human–machine-interface.

Keywords Response time • Visual modality • Auditory modality • Digit cue • Man–machine interface • Human factors

A. W. Y. Ng (✉) • A. H. S. Chan
Department of Systems Engineering and Engineering Management,
City University of Hong Kong, Kowloon, Hong Kong
e-mail: anniewy.ng@hotmail.com

A. H. S. Chan
e-mail: alan.chan@cityu.edu.hk

1 Introduction

The human-machine-interfaces nowadays commonly utilize different kinds of visual and auditory stimuli for communicating with human operators. These stimuli can be found in the design of military communication system, driving vehicle systems, crosswalk walking system, smoke detector alarm, and many other industrial applications for provision of timely alert information [1–8]. Colors, numerals, signal words, pictures, and non-speech auditory tones are common cues used to guide users' attention to a man-machine system [9–14]. There are some studies addressed to the comparison of the effectiveness of visual and auditory cues in terms of subjective preference and response time [e.g. [11–13, 15]].

Response time is a measure of the processing rate of sensory stimuli by central nervous system and its execution in form of motor response [16]. It can be partitioned into reaction time and movement time [17]. Reaction time refers to the duration from the onset of a signal calling for a response until the beginning of the response, while movement time denotes the duration from the beginning of the response to its completion. Reaction time tasks can be distinguished according to the number of diverse stimuli in a task that need to be responded with a specific motor reaction. In case the number of stimuli is equal to one, this kind of reaction time task is called simple reaction time task; if higher than one, it is a choice reaction time task. In a simple reaction time task, only one particular stimulus can occur and the same response is always required, whereas there can be several different stimuli in the choice reaction time task which requires a particular response for each stimulus.

In this paper, we compared human response time to visual and auditory digit cues in the context of man-machine system in terms of some user characteristics and task factors. These factors included age, gender, education level, time spent on computer in daily life, left/right response finger, and number of choice alternative. The findings of this study provide a useful reference for engineers and designers to make best use of visual and auditory cues into human-machine systems more effectively and efficiently.

2 Method

Participants. A total of 69 right-handed Chinese participants (32 males and 37 females), aged between 11 and 60 years, voluntarily took part in this study. All of them were in good visual, auditory and physical condition at the time of the study.

Apparatus and stimuli. An application program prepared with Visual Basic 6.0 and a notebook computer were used to generate visual and auditory stimuli and to capture the participants' responses. The visual stimuli were digit numbers (1, 2, 3, 4, 6, 7, 8, or 9) presented at the centre of the 7" computer screen. The auditory stimuli were speech signals of the digit numbers (in English) emitted from the

built-in speaker of the computer. A USB number pad was placed at a convenient location at the front of the participants for entering the corresponding response number. An adjustable chair was provided for the participants to perform the study.

Procedure. The participants were asked about their demographics, average time spent on computer per day, and dominant hand. Then they were briefed with the objectives and procedure of the study. The distance between the participants and the computer was 600 mm.

Visual stimuli test. The visual stimuli test consisted of a simple, two-, four-, and eight-choice reaction time task. Each task was ended after 10 correct responses were made. The details of the tasks are as follows.

- (i) Simple reaction time task: For each trial, a number ‘8’ was presented at the centre of the computer screen. The participants responded by pressing the corresponding number i.e., 8 on the keypad with their middle finger as fast as possible.
- (ii) Choice reaction time tasks: In each trial, the stimulus was either number ‘2’ or ‘8’ in the two-choice reaction time task; number ‘2’, ‘4’, ‘6’, or ‘8’ in the four-choice reaction time task; and number ‘1’, ‘2’, ‘3’, ‘4’, ‘6’, ‘7’, ‘8’, or ‘9’ in the eight-choice reaction time task. Once the number appeared at the centre of the computer screen, the participants pressed the corresponding number key with their middle finger as soon as possible.

At the beginning of each reaction time task, the participants were told which number set would appear for them to get familiar with the response modes. Both left and right middle fingers were tested. The sequence of testing of reaction time tasks and left/right fingers was counterbalanced across participants to minimize order effects.

The number key ‘5’ served as resting, initiating, and response position. A stimulus would appear between 150 to 300 ms randomly after the number key ‘5’ was pressed and last until a response was captured by the computer.

Auditory stimuli test. The stimuli were presented in the auditory mode, and the procedure of the auditory test was the same as that of the visual test.

3 Results

A total of 5520 responses (69 participants \times 4 reaction time tasks \times 10 trials \times 2 test fingers) were recorded. The average response time for visual cues across all choice tasks and response fingers was 0.517 ms (S.D. = 0.181 ms). With regards to auditory cues, the average response time was 0.493 ms (S.D. = 0.178 ms). Time in response to visual cues was significantly longer than that to auditory cues (Mann–Whitney U test, $U = 1.4E + 07$, $p < 0.05$). Regressions of the response time data with respect to number of alternatives using the Hicks–Hyman model [18] were done. The results showed that the regressions were significant for both

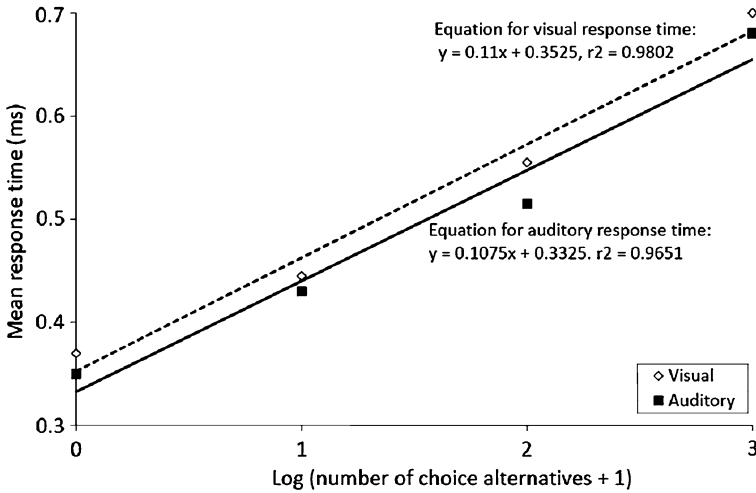


Fig. 1 Relations between response times and number of choice alternatives

the visual and auditory data (Fig. 1). This gave evidence for the variation of response time with respect to the logarithm of number of alternatives in reaction time tasks.

4 Response Time to Visual Cue: Effects of User Characteristics and Task Factors

Age. The results of ANOVA showed that age had significant effect on left/right finger response time for each choice task (ANOVA, p 's < 0.05). Post hoc analysis with Bonferroni method was used to analyse how age groups differed significantly from each other in terms of left/right finger response time.

Regarding the left finger response time, for *all choice tasks* the response time of 51–60 years was significantly longer than that of 11–20, 21–30, and 31–40 years. In the *single-choice task*, the response time of 31–40 years was significantly shorter than that of 41–50 years. In the *two-choice task*, the response time of 41–50 years was significantly shorter than 51–60 years. In the *four-choice task*, the response time of 41–50 years was significantly shorter than 51–60 years but significantly longer than 21–30 years and 31–40 years.

Regarding the right finger response time, for *all choice tasks*, the response time of 51–60 years was significantly slower than that of 11–20, 21–30, and 31–40 years. In the *single-choice task*, the response of 41–50 years was also significantly slower than that of 21–30 years. In the *two-choice task*, the response of 21–30 years was significantly faster than that of 11–20 and 31–40 years, while the response of 41–50 years was significantly slower than that of 11–20, 21–30, and 31–40 years.

In the *four-choice task*, the response time of 41–50 years was significantly faster than that of 51–60 years but slower than 21–30 years. In the *eight-choice tasks*, the response time of 51–60 years was significantly slower than that of 41–50 years.

Gender. There was significant difference between males and females in left/right finger response time. For the left finger response time, the response time of females was significantly shorter than that of males in single choice task ($t = -0.290$, $df = 688$, $p < 0.05$). No significant differences were found between females and males in the three choice reaction tasks.

For the right finger response time, the response time of female was significantly shorter than that of males in eight-choice task ($t = -2.393$, $df = 686$, $p < 0.05$). There were no significant differences between females and males in single, two- and four-choice tasks.

Education level. Education level had significant effect on right/left finger response time for each choice task (ANOVA, p 's < 0.05). Bonferroni post test was then conducted to determine which education levels differed significantly from each other.

For the left finger response time, tertiary education group had significantly shortest response times, followed by secondary education group and then primary education group in the single, two- and four-choice tasks. Under the eight-choice task, both secondary and tertiary education group responded significantly faster than primary education group, but no significant difference was found between secondary and tertiary education groups.

For the right finger response time, in the four-choice task, tertiary education group responded significantly faster than both primary and secondary education groups, while the response time of secondary education group was significantly shorter than primary education group. In the single, two- and eight-choice tasks, both secondary and tertiary education groups responded significantly faster than primary education group ($p < 0.05$), but no significant difference was revealed between secondary and tertiary education groups.

Time spent on computer in daily life. The time spent on computer in daily life had significant effect on response time for each choice task (ANOVA, p 's < 0.05). Bonferroni post test showed that, in general, the longer the time spent on computer in daily life, the shorter was the finger response time. However, for the left finger response time, there were no significant differences between 'less than 2 h' and '2–4 h' groups in the single choice task, between '2–4 h' and '4–6 h' groups in the four-choice task, and between '4–6 h' and 'more than 6 h' groups in the single, two- and four-choice tasks. For the right finger response time, there were no significant differences between '4–6 h' and 'more than 6 h' groups in the single, two-, four- and eight-choice tasks, and between '2–4 h', '4–6 h' and 'more than 6 h' groups in the single choice task.

Left/right response finger. The left finger response time was significantly longer than the right finger response time across all choice tasks (Independent samples t test, p 's < 0.05).

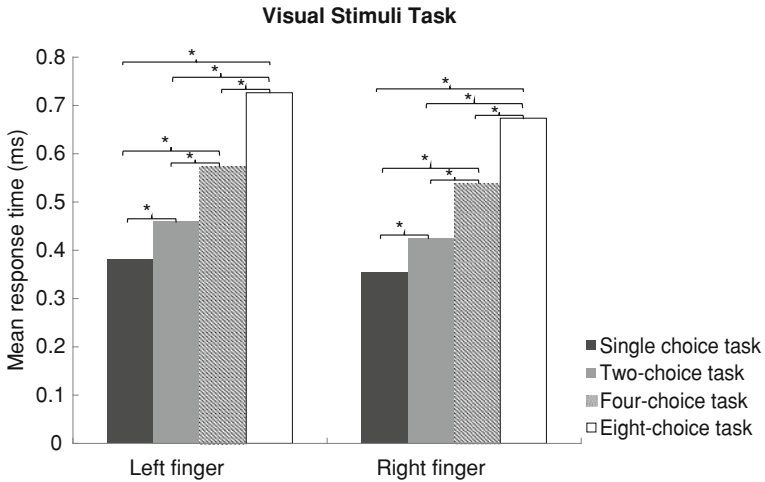


Fig. 2 Mean response time to visual stimuli in terms of choice alternative and left/right response finger factors for right-handers (*significant p value corrected with the Holm's sequential Bonferroni method)

Choice alternative. Choice alternative had significant effect on right/left finger response time (ANOVA, p 's < 0.05). The right/left finger response time of the single-choice task was significantly shorter, followed by two-choice task and then four- and eight-choice tasks (see Fig. 2).

5 Response Time to Auditory Cue: Effects of User Characteristics and Task Factors

Age. Age had significant effect on response time across all choice tasks (ANOVA, p 's < 0.05). Bonferroni post test was then conducted to determine which age groups differed significantly from each other.

For the left finger response time, the response of 51–60 years was significantly slower than that of 11–20, 21–30, and 31–40 years across all choice tasks. The response of 51–60 years was also significantly slower than that of 41–50 years in the single-, four-, and eight-choice tasks. In the *single choice task*, the response time of 41–50 years was significantly longer than that of 21–30 and 31–40 years, and the response time of 11–20 years was significantly longer than that of 21–30 and 31–40 years. In the *two- and four-choice task*, the response time of 21–30 years was also significantly shorter than 11–20, 31–40, and 41–50 years.

For the right finger response time, the response of 51–60 years was significantly slower than that of 11–20, 21–30, 31–40 years, and 41–50 years across *all choice tasks*. In the *single, two- and four-choice tasks*, the response time of 21–30 years was significantly shorter than that of 41–50 years. As compared to the response

time of 31–40 years, the response time of 21–30 years was significantly shorter in the two- and four-choice tasks but was marginally significantly longer in the eight-choice task.

Gender. Males responded significantly slower than females in two- and four-choice tasks with left/right finger (Independence samples t test, p 's < 0.05). No significant differences were found between males and females in single and eight-choice tasks.

Education level. Education level had significant effect on left and right finger response time across all choice tasks (ANOVA, p 's < 0.05). Bonferroni post test showed that the higher the education level, the shorter was the left/right finger response time in single, two- and four-choice tasks. In eight-choice task, the left/right finger response time of primary education group was significantly longer than that of secondary and tertiary education groups, whereas no significant difference was found between secondary and tertiary education groups.

Time spent on computer in daily life. Time spent on computer had significant effect on response time across all choice tasks (ANOVA, p 's < 0.05). In general, as the time spent on computer increased, the left/right finger response times decreased accordingly. However, Bonferroni post test indicated that regarding the left finger response time, no significant difference was found between '4–6 h' and 'more than 6 h' across all choice tasks. In the single-choice task, no significant difference between 'less than 2 h' and '2–4 h' was also found. In the four-choice task, there was also no significant difference between '2–4 h' and '4–6 h'. In the eight-choice task, no significant differences were also revealed between '2–4 h' and '4–6 h', and between '2–4 h' and 'more than 6 h'.

Regarding the right finger response time, in the single-choice task, there were no significant differences between '2–4 h' and '4–6 h', and between '4–6 h' and 'more than 6 h' groups. In the two-choice task, no significant differences were found between '2–4 h' and 'more than 6 h', and between '4–6 h' and 'more than 6 h'. In the four-choice task, no significant difference was also found between '4–6 h' and 'more than 6 h'. In the eight-choice task, no significant differences were revealed between '2–4 h' and '4–6 h' groups, between '2–4 h' and 'more than 6 h' groups, and between '4–6 h' and 'more than 6 h' groups.

Left/right response finger. The left finger response time was significantly longer than the right finger response time across all choice tasks (Independent samples t test, p 's < 0.05).

Choice alternative. Choice alternative had significant effect on response time (ANOVA, p 's < 0.05). The finger response time of the single-choice task was significantly shorter, followed by two-choice task and then four- and eight-choice tasks.

6 Discussion

This study examined human response to the visual and auditory digit cues in the context of man–machine system. The factors of age, gender, education level, left/right finger, time spent on computer in daily life, choice alternatives, and stimulus modality were found to be associated with response time significantly.

With regard to the age, in general, the response time decreased with an increase of age up to 21–30 years, and thereafter it increased with the age increase. Similar results have been reported by Ashoke et al. [19] with a group of subjects of 5–70 years of age. They found that the response was faster with an increase of age up to 21–25 years and then the response gradually slower with increase of age.

For the gender factor, females responded faster than males. This finding was similar to the research done by Han et al. [20], which revealed that females responded faster than males during the detection of threat cues in visual scenes.

Regarding the education level, the response time of tertiary and secondary education groups was shorter than that of primary education group. Previous studies indicated that the differences in reaction time tasks were due to processing time [21], and higher levels of education were associated with greater central executive efficiency and information processing speed [22, 23]. This implies that the higher the education level, the quicker will be the response. The study here confirmed this earlier hypothesis. Given the significant effects of age, gender and education level on response time, designers and engineers should take into account of users' demographic information into consideration when designing new human–machine-interfaces with the use of visual and auditory digit cues.

The right finger was shown to respond faster than the left finger of right-handed people in this study. Peters and Ivanoff [24] also found that right-handed people responded faster with their right-hand when using a right-handed computer mouse. Such results might be due to extensive right hand practice on keypad and mouse entries in the context of computer usage. This was in agreement with our findings that time spent on computer was positively associated with response time here. In addition, the preferred hand muscle strengths are generally stronger than the non-preferred ones [25], and a correlation was found to exist between muscle strength and reaction time [26]. The differences in left and right finger response time might also result from the muscle strength differences between the hands. Since people responded faster with their dominant hand, the man–machine systems should be designed in a way that facilitating users to use their dominant hand to make a prompt response to alerting cues.

Regarding the effect of choice alternative, the response on single-choice task was the fastest, followed by two-choice task and then four- and eight-choice tasks. Previous studies [16, 27] also found that the response on choice reaction time task was significantly longer than simple reaction time task. The possible explanation of such finding was that choice reaction task required not only stimulus perception and execution of the response but also decision-making processes, according to Kamitani et al. [27]. The results here indicated the smaller the number of digit cues

in a man–machine system, the faster the response time to the cue. Thus, the number of alerting cue alternatives in a system should be kept in minimum.

With respect to the sensory modality, time in response to the auditory cues was significantly shorter than the visual cues. The response time for visual cues was 5 % longer than that for auditory cues. This finding was similar to the studies done by Shelton and Kumar [28] and Solanki et al. [16]. Brebner and Welford [29] indicated that the variation of response across sensory modalities might be due to differences in the peripheral mechanisms such as some sensory systems are more sensitive than others. The temporal sensitivity of the skin is found to be very high which is close to that of the auditory system and larger than that of the visual system [30]. Sugiura et al. [31] revealed that in the simple reaction time tasks with visual and auditory cues, regions in the left primary motor area and the bilateral premotor cortex were activated; the distribution of premotor cortex subregions involved in sensorimotor association differed when the sensory cues were in different modalities. Based on the results here, it was suggested that auditory cues would be more effective for guiding respondents' attention to important information on displays than visual cues.

7 Limitations and Future Work

The human performance for visual and auditory digit cues in this study was determined based on the behavioral measure of response time. Subjective measure of the respondents towards these visual and auditory digit cues may also be assessed and considered in further study. The way to evaluate subjective measures can be referred to some previous studies [e.g. [31–33]]. In addition, there is an increasing use of multimodal coding and alerting systems. However, the nature of stimulus modalities here was considered individually. The combination of stimulus modalities effects should be considered in future study.

8 Conclusion

Response time to the auditory digit cues was shorter than the visual ones. Age, gender, education level, time spent on computer, choice alternative, and left/right finger were found to have significant effects on response time to both visual and auditory digit cues. Designers and engineers should take into account of these factors when designing human–machine-interfaces with the use of visual and auditory cues. In general, response time decreased with an increase of age up to the 21–30 years. After that, it increased gradually with increase of age. Females were found to respond faster than males. The response time of tertiary and secondary education groups was also faster than that of primary education group. Besides, the longer the time spent on computer in daily life, the shorter was the response time. The right finger response time was shorter than the left finger response time. In

addition, the response on single-choice task was the fastest, followed by two-choice task and then four- and eight-choice tasks. Implications of the results on the design of alerting cue and man-machine system were discussed. The findings of this study provided engineers and designers with useful information on how the visual and auditory modality channels could interfere the operators, so as to design a more user-friendly human-machine-interface in future.

Acknowledgments This work was supported in part by a grant from the Research Grant Council of the Hong Kong Special Administrative Region, China (Project No. CityU 110306). The authors would like to thank Chan Wing Kin for his help in the data collection process of the experiment.

References

1. Takahashi H, Honda H (2012) A study on the possibility of applying subliminal visual cue for guiding subject's attention. *J Adv Comput Intell Intell Inf* 16(1):96–107
2. Bruck D, Thomas I (2009) Smoke alarms for sleeping adults who are hard-of-hearing: comparison of auditory, visual, and tactile signals. *Ear Hear* 30(1):73–80
3. White TL, Kehring KL, Glumm MM (2009) Effects of unimodal and multimodal cues about threat locations on target acquisition and workload. *Mil Psychol* 21:497–512
4. Scott JJ, Gray R (2008) A comparison of tactile, visual, and auditory warnings for rear-end collision prevention in simulated driving. *Hum Factors* 50(2):264–275
5. Huang YC, Tsai CJ, Kuo JY, Wu FG (2011) The comparison of different sensory outputs on the driving overtake alarm system. In: Stephanidis C (ed) *Universal Access in HCI, Part III, HCI 2011, LNCS 6767*. Springer, Berlin, pp 290–297
6. Sanders AF (1975) The foreperiod effect revisited. *Q J Exp Psychol* 27:591–598
7. Krausman AS, Elliott LR, Redden ES, Petrov P (2005) Effects of visual, auditory, and tactile cues on army platoon leader decision making. In 10th international command and control research and technology symposium the future of C2
8. Baldwin CL, Eisert JL, Garcia A, Lewis B, Pratt SM, Gonzalez C (2012) Multimodal urgency coding: auditory, visual, and tactile parameters and their impact on perceived urgency. *Work* 41:3586–3591
9. Health and Safety Executive (2001) Review of wrong helideck landings, status lights and signalling lamps. Offshore technology report 2000/067. <http://www.hse.gov.uk/research/otopdf/2000/oto00067.pdf>
10. Hakkert AS, Gitelman V, Ben-Shabat E (2002) An evaluation of crosswalk warning systems: effects on pedestrian and vehicle behaviour. *Transp Res Part F* 5:275–292
11. Inao K, Akita T, Koga T (2009) Effects of various visual and auditory information upon reaction time of direction judgment—a study on cognition of environmental information from the viewpoint of cross modal effect. *J Asian Archit Build Eng* 8(1):95–102
12. Cheng SY, Hsu HT, Shu CM (2008) Effects of control button arrangements on human response to auditory and visual signals. *J Loss Prev Process Ind* 21:299–306
13. Smith CAP, Clegg BA, Heggstad ED, Hopp-Levine PJ (2009) Interruption management: a comparison of auditory and tactile cues for both alerting and orienting. *Int J Hum Comput Stud* 67:777–786
14. Ristic J, Wrright A, Kingstone A (2006) The number line effect reflects top-down control. *Psychon Bull Rev* 13(5):862–868
15. Ng AWY, Chan AHS (2012) Finger response times to visual, auditory and tactile modality stimuli. In: *Lecture notes in engineering and computer science: proceedings of the international multicference of engineers and computer scientists 2012, IMECS 2012, 14–16 March, Hong Kong*, pp 1449–1454

16. Solanki J, Joshi N, Shah C, Mehta HB, Gokhle PA (2012) A study of correlation between auditory and visual reaction time in healthy adults. *Int J Med Public Health* 2(2):36–39
17. Sanders MS, McCormick EJ (1993) *Human Factors in Engineering and Design*. McGraw-Hill, New York, pp 284–293
18. Hick WE (1952) On the rate of gain of information. *Q J Exp Psychol* 4:11–26
19. Ashoke B, Shikha D, Sudarsan B (2010) Reaction time with respect to the nature of stimulus and age of male subjects. *J Sport Health Res* 2(1):35–40
20. Han S, Gao X, Humphreys GW, Ge J (2008) Neural processing of threat cues in social environments. *Hum Brain Mapp* 29:945–957
21. Miller JO, Low K (2001) Motor processes in simple, go/no-go, and choice reaction time tasks: a psychophysiological analysis. *J Exp Psychol* 27(2):266–289
22. Bosma H, van Boxtel MPJ, Ponds RWHM, Houx PJH, Jolles J (2003) Education and age-related cognitive decline: the contribution of mental workload. *Educ Gerontol* 29:165–173
23. Tun PA, Lachman ME (2008) Age differences in reaction time and attention in a national telephone sample of adults: education, sex, and task complexity matter. *Dev Psychol* 44(5):1421–1429
24. Peters M, Ivanoff J (1999) Performance asymmetries in computer mouse control of right-handers, and left-handers with left- and right-handed mouse experience. *J Mot Behav* 31(1):86–94
25. Damon A, Stoudt HW, McFarland RA (1966) *The human body in equipment design*. Harvard University Press, Cambridge
26. Wojtys EM, Huston LJ (2000) Longitudinal effects of anterior cruciate ligament injury and patellar tendon autograft reconstruction on neuromuscular performance. *Am J Sports Med* 28(3):336–344
27. Kamitani T, Kuroiwa Y, Li M, Ikegami T, Matsubara S (2003) Relationship between cerebellar size and variation of reaction time during a visual cognitive task in normal subjects. *J Neurol* 250:1001–1003
28. Shelton J, Kumar GP (2010) Comparison between auditory and visual simple reaction times. *Neurosci Med* 1:30–32
29. Brebner JMT, Welford AT (1980) Introduction: an historical background sketch. In: Welford AT (ed) *Reaction times*. Academic Press, London, pp 1–23
30. van Erp JBF (2002) Guidelines for the use of vibro-tactile displays in human computer interaction. In: *Proceedings of eurohaptics 2002*, Edinburgh, pp 18–22
31. Sugiura M, Kawashima R, Takahashi T, Xiao R, Tsukiura T, Sato K, Kawano K, Iijima T, Fukuda H (2001) Different distribution of the activated areas in the dorsal premotor cortex during visual and auditory reaction-time tasks. *NeuroImage* 14:1168–1174
32. Chan AHS, Ng AWY (2009) Perceptions of implied hazard for visual and auditory alerting signals. *Saf Sci* 47:346–352
33. Masakura Y, Nagai M, Kumada T (2006) Effective visual cue for guiding peoples' attention to important information based on subjective and behavioral measures. In: *Proceedings of the first international workshop on Kansei*

Gender Differences of Human Response Under Vibration Condition

Hon Keung Yau

Abstract The purpose of this study is to investigate the gender differences of human response under different vibration frequencies. A total of nine frequencies (1.6, 2.0, 2.5, 3.15, 4.0, 5.0, 6.3, 8.0 and 10.0 Hz) were chosen as stimuli for testing in the experiment. In this experiment, four tests were conducted: Heart Rate Variety (HRV) Test, Posture test, Visual Reaction Test and Auditory Reaction Test. Ten males and ten females were invited to participate in the experiment. The major findings revealed that (i) there is difference in response of the feeling of discomfort for tensed upper body posture between males and females. However, (ii) there are no significant differences between males and females in visual and auditory tests. (iii) Vibration influenced our heart rate, it seems more serious on males than females as the heart rate of some female participants do not increase.

Keywords Auditory reaction test • Gender difference • Heart rate variety test • Posture test • Vibration • Visual reaction test

1 Introduction

Most people take the public transport to various destinations like schools and offices every day in Hong Kong. Common vibrations we can experience come mainly from the uneven road, changing of speed, sound produced when a sudden stop of the bus and train, etc., and these vibrations can cause damages to us. Different vibrations

H. K. Yau (✉)

Department of Systems Engineering and Engineering Management,
City University of Hong Kong, Kowloon Tong,
Kowloon, Hong Kong
e-mail: honkyau@cityu.edu.hk

can cause different levels of discomfort feeling on males and females. Some past studies have investigated how vibration created uncomfortable feeling [1] and affected working efficiency, safety and health [2]. However, few studies investigated the gender differences, if any, in the human response under different vibration frequencies in Hong Kong. Therefore, we fill this research gap in this study.

2 Literature Review

Human Response. Vibration interferes with people's working efficiency, safety and health [2]. It also causes discomfort, fatigue and physical pain [3]. Our health are affected if the strength of vibration is too large or the duration of vibration is too long [4]. Past research results showed that feeling of discomfort increases when the frequency increases [5]. Five hertz is the resonance frequency for whole-body vibration. It was also shown by them that the spine, inner organs and muscle are in resonance when vibration frequency reaches 8 Hz. The feeling of discomfort increases even higher when more than one organ of our body are in resonance [6, 7]. The thigh's stiffness can only affect little on the feeling of discomfort during vibration [7] while the tensed posture can increase our stiffness and decrease the feeling of discomfort [8]. In addition, the posture of tensed upper body increases the stiffness of body so as to increase the resonance frequency [9]. Due to the resonance of spine and inner organs, the uncomfortable index increases rapidly [6]. Besides, the uncomfortable feeling is much worse than the relax pose or even cause damage to our back if the vibration strength reaches the resonance frequency [3].

Visual Reaction. The normal visual reaction time is 0.19 s [10–12]. The eyes provide compensatory movement to protect the visual acuity for low frequencies between 1 and 2 Hz [13]. The vibration affects the visual ability at about 8–10 Hz and resonance at about 20 Hz [13].

Auditory Reaction. The auditory reaction time is faster because the signal in auditory modality arrives in the brain faster [10–12]. Arousal also affects the auditory reaction time [14].

Heart rate. It was reported that vibration influenced human physiological reactions such as heart rate, blood pressure and respiration [15, 16]. Heart rate variability (HRV) is one of the analyses of the interaction by subtraction of the heart beat-to-beat interval [17].

3 Methodology

3.1 Participants

Twenty people (10 males and 10 females) were invited to participate in the experiment. They were of average age 21.95 years, 168.1 cm tall and 59.1 kg weight.

3.2 Design of Experiment

To acquire the data, the following four tests were conducted:

- (i) Posture test
- (ii) Visual Reaction test
- (iii) Auditory Reaction test
- (iv) Heart Rate Variety (HRV) test

(i) Posture test

- Method of magnitude estimation

1. Try stimulus without upper body relax as a reference stimulus
2. Try the same stimulus with upper body tensed
3. Give uncomfortable index comparing to the reference stimulus

(ii-iii) Visual/Auditory Reaction Test

1. Participants did the reaction task for five times before the test
2. Did the task for five times again under vibration condition

(iv) Heart Rate Variety Test (HRV Test)

- The heart rate was collected three times
 1. Before the test
 2. After the second test
 3. 5 min after all test

3.3 Experiment Procedure

Before the experiment started, participants were asked to do the visual and auditory reaction test to determine whether human reaction was affected under different vibration conditions. A machine with lamp and timer was used for the visual reaction test. Participants were required to press the corresponding button to stop the timer when they saw the lamp which was on. For the auditory reaction test, same equipment was used also, when they sensed the “beep” sound, they needed to stop the timer by pressing the button on the machine panel. The reaction time of participants was recorded to compare with the performance under different vibration conditions. Then the heart rate of the participants was checked and the exposure test started. After the second random chosen test, the heart rate was checked at the same time. After all tests were completed, the final heart rate was checked again. Both left and right hands were also measured, and the higher one was recorded.

For the visual and auditory reaction tests, participants were required to do the test under different vibration frequencies again.

At the end, the heart rate was measured again 5 min after the test and a perceptual sensation questionnaire was used to ask to find out whether the vibration caused any symptoms to the participant for evaluating which frequency and posture caused harm to which body parts.

The test and frequency sequence was randomized and counterbalanced across trials but each pair of males and females was under the same sequence for the analysis of difference in males and females.

4 Results and Discussion

4.1 Posture Test

From the Table 1, we can see that the tensed posture did help for female before 5 Hz. But it seems not useful for male. The tensed posture also makes male participants feel uncomfortable at low frequencies, but the uncomfortable feeling was not very strong. Both male and female feel that the tensed posture make them feel more uncomfortable after 5 Hz. It is because the 5 Hz is the resonance frequency for whole-body resonance [6, 7]. In addition, the posture of tensed upper body will increase the stiffness of body so as to increase the resonance frequency [9]. At 8 and 10 Hz, the discomfort index increase rapidly, it should be due to the resonance of spine and inner organs [6]. When we are tensing our upper body, the loading of spine and our back will increase. If the vibration strength reaches the resonance frequency, the uncomfortable feeling will be much worse than the relax pose or even cause damage to our back [3]. From Table 1, we only found that 2, 3.15 and 5 Hz resulted in the significant difference in response to the increase in frequency. It indicates that the males and females have the same response at tensed posture to the increase in frequency. Also, the variation of response of females is larger than males in the same frequency before 5 Hz, but that male's variation of response becomes large at 6.3, 8 and 10 Hz become very large and even reaches 36.893 at 10 Hz. It indicates that the difference in male's strength of spine is much larger than female. For participants who have a healthy spine, their index should be much lower under this extreme condition.

4.2 Visual Reaction Test

From Table 2, we can observe that the visual reaction time is a bit lower than the normal response of 0.19 s reported by previous researchers [10–12]. We can see that the reaction time remains stationary from 1.6 to 5 Hz. At the beginning frequency up to 2 Hz, the eyes will provide compensatory movement to protect the visual acuity for low frequency at about 1–2 Hz [13]. The increase in reaction time

Table 1 Results of posture test

Frequency (Hz)	Index-combination	Index-male	Index-female
1.6	98.5	102	95
2	98.5	102	95
2.5	93	96	90
3.15	98.5	112	85
4	98.5	107	90
5	93.75	87	100.5
6.3	116.25	115	117.5
8	125.5	123	128
10	157.5	165	150

Table 2 Results of visual reaction test

Frequency (Hz)	Reaction time (male)	Reaction time (female)	Reaction time combination
Static	0.32912	0.32536	0.32724
1.5	0.33094	0.3271	0.32902
2	0.33528	0.33674	0.33601
2.5	0.34152	0.34222	0.34187
3.15	0.3454	0.35624	0.35082
4	0.34946	0.3689	0.35918
5	0.3692	0.38082	0.37501
6.3	0.36994	0.39698	0.38346
8	0.41138	0.43012	0.42075
10	0.46194	0.4738	0.46787

becomes obvious after 4 Hz indicating that the compensatory movement of the eyes is less useful after 4 Hz. After 6.3 Hz, the reaction time starts to increase more obviously. At 8 and 10 Hz, the slope becomes steeper. It is normal as the vibration will affect the visual ability at about 8–10 Hz and resonant at about 20 Hz [13]. The results also indicated that there was no significant difference in visual reaction time between males and females for all frequencies. Besides vibration frequency, one of the factors that may affect the result is the posture. During the test, participants need to put their hands on the testing machine without any backrest of armrest. The reaction time was also influenced by the posture.

4.3 Auditory Reaction Test

Table 3 shows that the auditory reaction time is faster as the signal in audiotroy will arrive in the brain faster. It has been proved by many researchers [10–12]. Auditory reaction test reflects a more accurate result for human’s reaction as candidate only needs to press a button when they hear a sound. But they need to make a choice for visual. So, the reaction time for auditory test should be lower.

Table 3 Results of auditory reaction test

Frequency (Hz)	Reaction time male	Reaction time female	Reaction time combination
Static	0.2051	0.20278	0.20394
1.6	0.21116	0.21014	0.21065
2	0.22124	0.22012	0.22068
2.5	0.22778	0.22956	0.22867
3.15	0.23346	0.23662	0.23504
4	0.24622	0.25232	0.24927
5	0.27116	0.26916	0.27016
6.3	0.2817	0.2881	0.2849
8	0.32102	0.33404	0.32753
10	0.4505	0.37902	0.41476

From Table 3, we can see that there is generally not much difference between males and females at auditory reaction time. But the variation of auditory reaction time for males is larger than males at the same frequency, that may be due to the arousal, since the variation of discomfort for males is larger than females, the larger variation of auditory reaction time for male may be due to the large variation of discomfort generated from the frequency. That may cause distraction to the participants because arousal will affect the auditory reaction time [14]. The result at 8 and 10 Hz may not be very accurate as the noise from machine was very large and some participants even cannot hear the signal clearly.

4.4 HRV Test

From Table 4, we can observe that the vibration influenced our heart rate which is consistent with the past studies that vibration influenced human physiological reactions such as heart rate [15, 16]. The correlation for both males and females are high. But it seems more serious on males than females as the heart rate of some female participants do not increase. It might be due to the sequence of the test. Apart from the testing sequence, the testing time might also influence not only the heart rate, but also the HRV. From the Table 4, we see that the variation of heart rate is quite large from 5 to -5 . So, it is not suitable for us to expose under vibration condition for long time. It is because some researchers have pointed out that our health will be affected if the strength is too large or the duration is too long [4]. Also, we cannot see any significant decrease in heart rate after the test. It may be due to the insufficient time for the participant to cool down, so some participant's heart rate keeps on a higher level than heart rate before the experiment.

Table 4 Results of HRV test

Candidate number	Original heart rate	Variability male	Variability female
Male 1	87	3	-2
Male 2	69	3	2
Male 3	78	1	-3
Male 4	79	2	-1
Male 5	75	4	1
Male 6	79	3	2
Male 7	77	2	-1
Male 8	83	1	2
Male 9	84	-1	-3
Male 10	79	2	1
Female 1	84	2	-2
Female 2	87	-1	-1
Female 3	101	-2	-2
Female 4	83	0	-1
Female 5	98	3	2
Female 6	89	-1	1
Female 7	89	1	2
Female 8	93	1	2
Female 9	98	-1	-1
Female 10	102	1	0

5 Conclusion

A series of human factors experiments concerning the gender difference of human response under different vibration conditions was conducted in this study. It was concluded that the tensed upper body posture was helpful for female before 5 Hz. The difference in response for tensed upper body posture between males and females was quite large. That may be due to 8 Hz which was the resonance frequency for spine and other inner organs. For visual and auditory reaction time, there was no significant difference of response between males and females. The reaction time varied according to the increase of frequency and it was more obvious after 5 Hz. For the heart, it was concluded that the heart rate increased during the vibration. However, it took more than 5 min for participants to cool down after exposure under vibration condition for more than 30 min. Finally, the exposure time for vibration should not be too long especially for people who do not have sufficient exercise. Their feeling of increase in heart rate was more serious because some participants told us that they felt the centrifugal force just like playing the roller coaster. Eight and ten Hz were the extreme frequencies in the test which was shown by previous researchers that those frequencies were the resonance frequency for spine and inner organs, we should prevent to expose our body under high frequencies because it is dangerous and may cause damage to our body.

References

1. Yau HK, Luk BL, Chan SS (2012) Evaluation of human response under vibration condition. Lecture notes in engineering and computer science, In: Proceedings of the international multiconference of engineers and computer scientists, IMECS 2102, 14–16 March 2012, Hong Kong, pp 1439–1442
2. Mcleod RW, Griffin MJ (1995) Mechanical vibration included interference with manual control performance. *Ergonomics* 38:1431–1444
3. Liu JZ, Kubo M, Aoki H (1995) A study on the difference of human sensation evaluation to whole body vibration in sitting and lying postures. *Appl Hum Sci* 14:219–226
4. Byung CM, Soon CC, Se JP (2002) Automatic responses of young passengers contingent to the speed and driving model of a vehicle. *Int J Ind Ergon* 29:187–198
5. Subashia GHMJ, Nawayseh N, Matsumoto Y, Griffin MJ (2008) Nonlinear subjective and dynamic responses of seated subjects exposed to horizontal whole-body vibration. *J Sound Vib* 321:416–434
6. Mester J, Spitznapefl P, Yue ZY (2003) Vibration loads: potential for strength and power development In: Komi P.V. (ed) *In strength and power in sport*. Blackwell Publishing Company, Oxford, pp. 488–501.
7. Nawayseh N, Griffin MJ (2002) Non-linear dual-axis biodynamic response to vertical whole-body vibration. *J Sound Vib* 268:503–523
8. Matsumoto Y, Griffin MJ (2002) Non-linear characteristics in the dynamic responses of seated subjects exposed to vertical whole-body vibration. *J Biomech Eng* 124:527–532
9. Huang Y, Griffin MJ (2006) Effect of voluntary periodic muscular activity on nonlinearity in the apparent mass of the seated human body during vertical random whole-body vibration. *J Sound Vib* 298:824–840
10. Brebner JT, Welford AT (1980) Introduction: an historical background sketch. In: Welford AT (ed) *Reaction times*. Academic Press, New York, pp 1–23
11. Von Fieandt K, Huhtala A, Kullberg P, Saarl K (1956) Personal tempo and phenomenal time at different age levels. The University of Helsinki Psychological Institute Report No. 2
12. Galton F (1899) On instruments for (1) testing perception of differences of tint and for (2) determining reaction time. *J Anthropol Inst* 19:27–29
13. Grether WF (1971) Vibration and human performance. *Hum Factors* 13(3):203–216
14. Welford AT (1977) Motor performance. In: Birren JE, Schaie KW (eds) *Handbook of the psychology of aging*. Van Nostrand Reinhold, New York, pp 450–496
15. Kubo M, Terauchi F, Aoki H (2001) An investigation into a synthetic vibration model for humans: an investigation into a mechanical vibration human model constructed according to the relations between the physical, psychological and physiological reactions of humans exposed to vibration. *Int J Ind Ergon* 27:219–232
16. Ullsperger P, Seidel H (1980) On auditory evoked potentials and heart rate in man during whole-body vibration. *Eur J Appl Physiol* 43:183–192
17. Kun J, Li Z, Chen M, Wang C, Qi S (2004) Effect of different vibration frequencies on heart rate variability and driving fatigue in healthy drivers. *Int Arch Occup Environ Health* 77:205–212

Extended Kalman Filter (EKF) with Sensitive Equation for the PEMFC Internal State Estimation

Jichen Liu, Guangji Ji and Su Zhou

Abstract State estimation is widely used in the field of process system engineering. There are several available technologies regarding this topic, e.g. Extended Kalman Filter (EKF). The EKF is the variant of the standard Kalman filter and is successfully applied on the nonlinear systems for state estimation. As well known, the PEMFC is a typical nonlinear system, and some of the internal states are obtained costly, and even cannot be measured directly. Hence, in order to obtain these internal states effectively via collecting measurable variables, the EKF is applied in this study. The goal of this paper is to demonstrate the implementation of the EKF based on a PEMFC model which is taken in a literature, in order to estimate the following internal states: concentrations of vapor and oxygen in cathode chamber, as well as cell temperature. The corresponding results show that the EKF can serve as a ‘software sensor’ for the control design or the supervision in the fuel cell system.

Keywords Case study · Extended Kalman filter · Implementation · PEMFC system · Sensitive equation · Simulation study · State estimation

J. Liu · G. Ji (✉)

School of Automotive Studies, Tongji University, Shanghai 201804, China

e-mail: jiguangji@126.com

J. Liu

e-mail: jichenliu2011@gmail.com

S. Zhou

School of Automotive Studies/Clean Energy Automotive Engineering Center/Sino-German

Postgraduate School, Tongji University, Shanghai 201804, China

e-mail: suzhou@tongji.edu.cn

1 Introduction

As well known, not all the states in a chemical process are easily obtained. The state estimation could be a costless solution to access these states via state estimation techniques. For the linear system, the state estimation can be done by the Kalman Filter (KF) because the probability distribution function (pdf) of the system state is propagated in an optimal way. However, the KF is not suitable for the nonlinear systems. An alternative solution is a variant of the standard KF, e.g. extended Kalman filter (EKF) or unscented Kalman filter (UKF). It is known that there exists no perfect solution that is superior to other techniques of state estimation. The main principle of choosing an estimation method is to trade off various aspects such as accuracy of estimation, difficulty of implementation, numerical robustness, and computational load [1].

The EKF with different applications encounter in several literatures. Simon [2] briefly introduced the mentioned different kinds of state estimation methods and their basic principles, respectively. By comparing these methods, the EKF is one of the most widely used and attractive methods owing to its relative simplicity and efficacy for nonlinear system. Kandepu et al. [3] depicted the principle of the EKF, and four nonlinear cases were introduced: the first one is the Van der Pol oscillator. The second is an estimation problem in an induction machine. The third is the state estimation in a reversible reaction. Finally, a hybrid solid oxide fuel cell (SOFC) system is introduced to evaluate the performance of the EKF. In the literature [4], an evaluation of the EKF was given by comparing to the moving horizon estimation (MHE). It turned out that, the computational load required to solve the MHE is greater than the EKF. For the differential algebraic equation (DAE) system, the EKF can be used through the proper modifications as well [5].

The combination of the EKF and the fuel cell system can be also found in several literatures. It is known that the fuel cell system is a strong nonlinear system. Some internal states in the system (e.g. cathode side relative humidity (RH)) would be detected costly by the delicate sensors. To reduce the cost of the system, those sensors can be removed by using the EKF. Some internal states (e.g. oxygen partial pressure and cell temperature) cannot be directly measured by sensors and could be estimated via the EKF as well. Recently, some studies concerning on the state estimation in the fuel cell systems have been carried out. Groetsch et al. [6] applied the EKF which is designed on the basis of the reduced model of a molten carbonate fuel cell (MCFC) system. The EKF was tested by both simulations and experiments. It is stated the temperature sensors in the stack as well as the expensive concentration measuring equipments can be removed. Kandepu et al. [7] used both UKF and EKF methods on the SOFC system to estimate the stack temperature. Soares and Kosanovich [8] used a nonlinear programming (NLP) formulation for both the parameter and state estimation in a PEMFC. Goerguen et al. [9] proposed a novel method for estimation of the inside humidity by exploiting its effect on cell resistive voltage drop. There are seldom

studies reporting the applications of EKF in the PEMFC system. Hence, the main motivation of this paper is to apply the EKF and describe its implementation in PEM fuel cell system with considerable clear manner.

The organization of this paper is as follows. In Sect. 2 the fundamental theory of Kalman filter and EKF is illustrated by using an example of a nonlinear system. In Sect. 3 the EKF is tested via a PEMFC system. The introduction of the fuel cell system model, implementation of the EKF technique and the simulation results are included. Finally, the conclusions are drawn in Sect. 4.

2 Fundamental Theory of the KF and the EKF

2.1 Principle of the KF

The KF is the minimum-variance state estimator, no matter with- or without Gaussian noise [3]. To illustrate the principle of the KF, the following linear discrete-time system is taken into account. The model equations are given in (1) and (2):

$$x_k = Fx_{k-1} + w_k \quad (1)$$

$$y_k = Hx_k + v_k \quad (2)$$

where k is the time step, x_k represents the state to be estimated at time k , y_k is the measurement at time k , w_k and v_k are the zero-mean process noise and measurement noise with co-variances Q and R , respectively, i.e. $E[w_k w_k^T] = Q$, $E[v_k v_k^T] = R$, $E[w_k] = E[v_k] = 0$, F and H are the state transition and measurement matrix, which represent the essential characteristics of the system. Figure 1 shows the framework of state estimation. It is assumed that the system can be observable fully, and the KF formulations for the states are given as follows [3]:

$$P_k^- = FP_{k-1}^+ F^T + Q \quad (3)$$

$$\hat{x}_k^- = F\hat{x}_{k-1}^+ \quad (4)$$

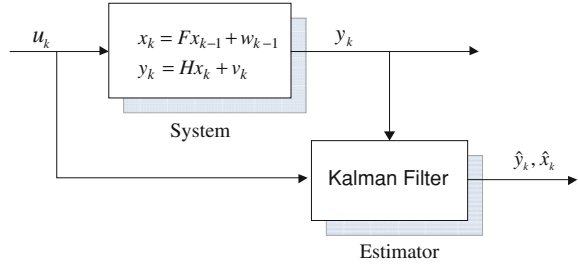
$$P_k^+ = (I - K_k H)P_k^- \quad (5)$$

$$\hat{x}_k^+ = \hat{x}_k^- + K_k(y_k - H\hat{x}_k^-) \quad (6)$$

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1} \quad (7)$$

where $k = 1, 2, \dots, I$ is the identity matrix, \hat{x}_k^- is the priori estimation of the state x_k , \hat{x}_k^+ is the posteriori estimation of the state, K_k is the Kalman gain, P_k^- is the covariance of the priori estimation error $x_k - \hat{x}_k^-$. P_k^+ is the covariance of the posteriori estimation error $x_k - \hat{x}_k^+$. The initialization of the KF is written as:

Fig. 1 Framework of state estimation with the KF



$$\hat{x}_0^+ = E(x_0) \quad (8)$$

$$P_0^+ = E[(x_0 - \hat{x}_0^+)(x_0 - \hat{x}_0^+)^T] \quad (9)$$

Generally there are two main steps in KF. The first step is called ‘prediction step’, as shown in (3) and (4). The second step is called ‘update step’, shown in (5), (6) and (7).

2.2 Introduction of the EKF Via an Exemplary Nonlinear Model

The KF is the unbiased estimation for the linear model [3]. For the nonlinear model, the expectation of the nonlinear function cannot be easily obtained. Under this situation, the EKF can be derived when the nonlinear model is linearized and the expectation can be approximated as the linear function of the estimated state. The main disadvantage of linearization lies in biased estimation and insufficient accuracy. However, for the state estimation in nonlinear systems, the EKF has several attractive features because it can be easily understood and implemented. In this study, the EKF is firstly illustrated with a discrete-time nonlinear dynamic system which is given as:

$$x_k = \psi(x_{k-1}) + w_{k-1} \quad (10)$$

$$y_k = h(x_k) + v_k \quad (11)$$

where process noise w_k and measurement noise v_k also abide by $E[w_k w_k^T] = Q$, $E[v_k v_k^T] = R$, $E[w_k] = E[v_k] = 0$. ψ and h represent the characteristics of the representative nonlinear system. The main differences between EKF and KF lie in the ‘prediction step’. The detailed formulations of the EKF can be found through (12)–(18).

$$\hat{x}_k^- = \psi(\hat{x}_{k-1}^+) \quad (12)$$

$$P_k^- = \frac{\partial \psi}{\partial x} \Big|_{\hat{x}_{k-1}^+} P_{k-1}^+ \frac{\partial \psi}{\partial x} \Big|_{\hat{x}_{k-1}^+}^T + Q \tag{13}$$

$$P_{xy} = P_k^- \cdot \frac{\partial h}{\partial x} \Big|_{\hat{x}_k^-} \tag{14}$$

$$P_{yy} = \frac{\partial h}{\partial x} \Big|_{\hat{x}_k^-} P_k^- \frac{\partial h}{\partial x} \Big|_{\hat{x}_k^-}^T + R \tag{15}$$

$$K_k = P_{xy} \cdot P_{yy}^{-1} \tag{16}$$

$$\hat{x}_k^+ = \hat{x}_k^- + K_k(y_k - \hat{y}_k^-) \tag{17}$$

$$P_k^+ = P_k^- - K_k P_{yy} K_k^T \tag{18}$$

Actually, a nonlinear system is usually represented by a continuous-time model rather than a discrete-time model because the continuous model can be obtained directly through the fundamental balances (e.g. mass balance, energy balance, etc.). Without losing generality, the mathematical formulation of a continuous-time nonlinear model can be written as:

$$\dot{x} = f(x, u) \tag{19}$$

In order to apply the EKF in discrete manner, Eq. (19) would be transferred into the discrete-time form in order to obtain $\frac{\partial \psi}{\partial x}$ in (13). There exist some available approaches to achieve such purpose, e.g. explicit or implicit Euler methods, and Runge–Kutta method. In this study, the discretization can be prevented by means of the sensitive equation which is defined as:

$$S(t) = \frac{\partial \hat{x}(t)}{\partial \hat{x}_{k-1}^+} \tag{20}$$

According to (20), the derivation of the sensitive variable, S, can be given as following:

$$\frac{dS(t)}{dt} = \frac{d}{dt} \left(\frac{\partial \hat{x}(t)}{\partial \hat{x}_{k-1}^+} \right) = \frac{\partial f(\hat{x})}{\partial \hat{x}_{k-1}^+} = \frac{\partial f(\hat{x})}{\partial \hat{x}(t)} \cdot \frac{\partial \hat{x}(t)}{\partial \hat{x}_{k-1}^+} = J \cdot S(t) \tag{21}$$

where the Jacobian matrix J is the partial differential of $f(\hat{x})$ to $\hat{x}(t)$. Combining (19) and (21) can lead to the extended differential equations, see (22):

$$\begin{cases} \dot{x} = f(x, u), & x \in R^n, f \in R^n \\ \dot{S} = J \cdot S, & J \in R^{n \times n} \end{cases} \tag{22}$$

The initial matrix for the sensitive equation equals to the unit matrix, which is given as:

$$S|_{k=0} = I \in \mathbf{R}^{n \times n} \quad (23)$$

Afterwards, the $\left. \frac{\partial \psi}{\partial x} \right|_{\hat{x}_{k-1}^+}$ in (13) can be written as:

$$\left. \frac{\partial \psi}{\partial x} \right|_{\hat{x}_{k-1}^+} = \frac{\partial \hat{x}(k)}{\partial \hat{x}_{k-1}^+} = S(t_k) \quad (24)$$

Hence, it is obviously that the introduction of a sensitive variable $S(t)$ is another way to calculate the $\left. \frac{\partial \psi}{\partial x} \right|_{\hat{x}_{k-1}^+}$.

To demonstrate the principle of the EKF with sensitive equation, a highly nonlinear Van der Pol oscillator is used. The model equations are given as follows:

$$\begin{aligned} \dot{x}_1 &= -x_2 + w_1 \\ \dot{x}_2 &= -0.2 \times (1 - x_1^2)x_2 + x_1 + w_2 \end{aligned} \quad (25)$$

The output of the nonlinear system is defined below:

$$\begin{aligned} x &= [x_1 \ x_2]^T \\ y &= [x_1 \ x_2]^T + v \end{aligned} \quad (26)$$

where covariance of process noise w and measurement noise v are both set as $10^{-3}I$. The initial value of x is chosen as $x_0 = [1.4 \ 0]^T$. The initial value of estimated states in the EKF is selected as $\hat{x}_0 = [0 \ 5]^T$. The elements in the Jacobian matrix used in the sensitive equation (see (22)) are given as following:

$$\begin{aligned} J(1, 1) &= 0 \\ J(1, 2) &= -1 \\ J(2, 1) &= 1 + 0.4x_2x_1 \\ J(2, 2) &= -0.2(1 - x_1^2) \end{aligned} \quad (27)$$

The corresponding simulations are carried out to verify the EKF in this case. As can be seen in Fig. 2, large errors at the beginning occur due to the selection of the initial values. After less than one second, the estimated states give the convergent behavior.

3 PEM Fuel Cell State Estimation by Means of the EKF

3.1 The Model of the Reference PEMFC

In order to use the EKF on the state estimation in a PEMFC system, a simple model should be considered. For this purpose, the study of Lauzze and Chmielewski [10]

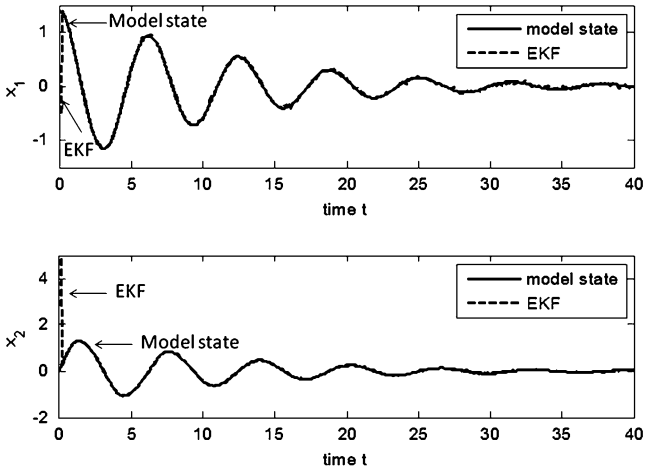


Fig. 2 Estimated states for Van der Pol oscillator using EKF

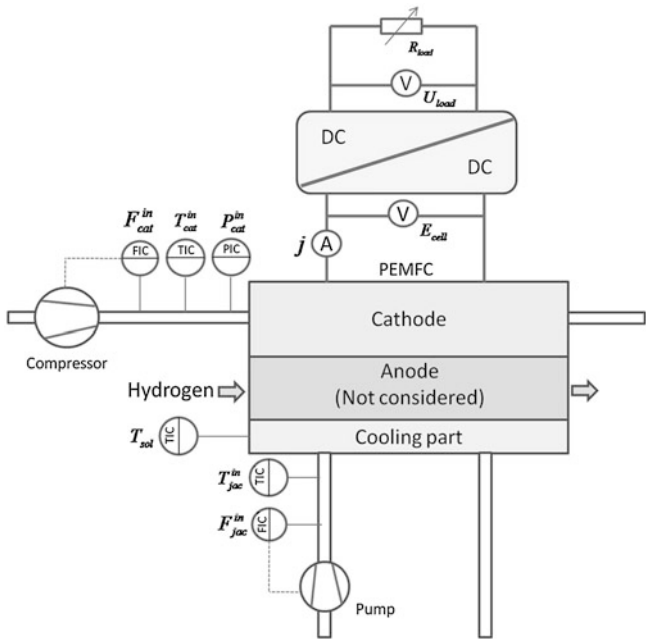


Fig. 3 P&ID of the selected PEM fuel cell system

is chosen. Figure 3 shows the piping and instrumentation diagram (P&ID) of the selected PEM fuel cell system considered in [10].

In the work of Lauzze and Chmielewski, the PEM fuel cell is treated as a Continuous Stirred Tank Reactor (CSTR) with lumped parameters. The following general assumptions are made: (1) The generated water in the fuel cell is in gas phase. (2) All gases obey the ideal gas law. (3) The temperature of all the solid materials and the coolant in the jacket is lumped. (4) The anode part is not considered in the model due to the assumption of a pure hydrogen feed. (5) The water transport in the membrane is neglected. (6) The dynamics of air stream and coolant delivery are neglected as well.

The formulated model bases on material and energy balances [10]. The material balance equations are developed for the oxygen, nitrogen and water vapor concentrations in the cathode gas chamber. The energy balance equations are developed for the cathode gas chamber, for the solid material and for cooling part to derive the temperature in each compartment. The electrochemical kinetics is modeled by using the Tafel equation. The oxygen transport from chamber to the reaction zone is modeled by Fickian diffusion equation. The oxygen diffusion coefficient is a function of relative humidity (RH) at the cathode side in order to consider flooding issues. As for the ohmic losses in the membrane, an empirical relationship between the proton conductivity, water content and temperature is employed from [11].

In the following, the model equations used in this work are briefly collected. The PEMFC model contains six dynamic states, i.e. the concentrations of oxygen, water vapor and nitrogen, the temperature of the gas stream in the cathode side, cell temperature and coolant temperature. The corresponding parameters in the model can be found in Table 1.

3.1.1 Cathode Mass Balances

$$V_{cat} \frac{dC_{O_2}}{dt} = F_{cat}^{in} C_{O_2}^{in} - F_{cat}^{out} C_{O_2} - \frac{1}{2} \gamma_{H_2O} A_{mem} \quad (28)$$

$$V_{cat} \frac{dC_{H_2O}}{dt} = F_{cat}^{in} C_{H_2O}^{in} - F_{cat}^{out} C_{H_2O} + \gamma_{H_2O} A_{mem} \quad (29)$$

$$V_{cat} \frac{dC_{N_2}}{dt} = F_{cat}^{in} C_{N_2}^{in} - F_{cat}^{out} C_{N_2} \quad (30)$$

The rate of the electrochemical reaction (generation of water per unit area of MEA) is expressed with the Faraday's law:

$$\gamma_{H_2O} = \frac{j}{nF} \quad (31)$$

where j is the current density, F is the Faraday's constant, n is number of electrons being dissociated. In Eqs. (28)–(30), the volumetric flow rate at the cathode outlet, F_{cat}^{out} , obeys a simple relationship:

Table 1 The parameters needed in the model

Parameter	Value	Unit
MEA effective area, A_{mem}	16e-4	m ²
Membrane thickness, t_{mem}	150e-6	m
Faraday's constant, F	96485	C/mol
Universe gas constant, R	8.314	J/(mol K)
Air density, ρ_{air}	1.23	kg/m ³
Air heat capacity, $c_{p,air}$	1003	J/(kg K)
Solid part density, ρ_{sol}	2500	kg/m ³
Solid part heat capacity, $c_{p,sol}$	100	J/(kg K)
Coolant density, ρ_{jac}	985	kg/m ³
Coolant heat capacity, $c_{p,jac}$	4200	J/(kg K)
Volume of cathode chamber, V_{cat}	2.7e-6	m ³
Volume of coolant jacket, V_{jac}	2.7e-8	m ³
Volume of solid part, V_{sol}	1.2e-4	m ³
Heat transfer coefficient for gas and solid part, UA_{cat}	0.042	W/K
Heat transfer coefficient for coolant and solid part, UA_{jac}	3.276	W/K
Heat transfer coefficient for ambient and solid part, UA_{eff}	0.01	W/K
H ₂ partial pressure, P_{H_2}	101325	Pa
Enthalpy of formation for water, $\Delta H_{f,H_2O}$	2.86e5	J/mol
Reference O ₂ concentration at standard condition, $C_{O_2}^0$	40.88	mol/m ³
Reference exchange current density, j_0^0	0.1	A/m ²
Charge transfer coefficient, α	0.5	-
Mass transfer coefficient, K_0	9.6e-3	m ² /s
Porosity coefficient, ϕ	0.025	-
Anti-flooding coefficient, F_0	1	-
Inlet coolant temperature, T_{jac}^{in}	298	K
Inlet cathode gas temperature, T_{cat}^{in}	343	K

$$CF_{cat}^{out} = CF_{cat}^{in} + \frac{1}{2}\gamma_{H_2O}A_{mem} \quad (32)$$

where C is the total concentration of gases in the cathode chamber.

3.1.2 Energy Balances

$$V_{cat} \frac{dT_{cat}}{dt} = F_{cat}^{in} T_{cat}^{in} - F_{cat}^{out} T_{cat} + \frac{UA_{cat}}{(\rho C_p)_{air}} (T_{sol} - T_{cat}) \quad (33)$$

$$\begin{aligned}
(\rho C_P)_{sol} V_{sol} \frac{dT_{sol}}{dt} &= UA_{cat}(T_{cat} - T_{sol}) \\
&+ UA_{jac}(T_{jac} - T_{sol}) \\
&+ UA_{eff}(T_{amb} - T_{sol}) \\
&+ Q_{gen} A_{mem}
\end{aligned} \tag{34}$$

$$(\rho C_P)_{jac} V_{jac} \frac{dT_{jac}}{dt} = (\rho C_P)_{jac} F_{jac}^{in} (T_{jac}^{in} - T_{jac}) + UA_{jac}(T_{sol} - T_{jac}) \tag{35}$$

As for the heat generation source term, Q_{gen} , it is the amount of heat generated by the exothermic reactions, and is expressed as follows:

$$Q_{gen} = (\Delta H_{f,H_2O}) \gamma_{H_2O} - P_e \tag{36}$$

$$P_e = jE_{cell} \tag{37}$$

where the $\Delta H_{f,H_2O}$ is the enthalpy of formation for water (lower heating value), P_e is the electrical power per unit MEA area.

3.1.3 Electrochemistry

The cell voltage is formulated as a combination of open-circuit voltage and several losses:

$$E_{cell} = E_{ner} - E_{act} - E_{mt} - E_{ohm} \tag{38}$$

The open-circuit voltage is obtained through the Nernst potential expression:

$$E_{ner} = E^\circ + \frac{RT_{sol}}{nF} \ln \frac{P_{H_2} \sqrt{P_{O_2}}}{P_{H_2O}} \tag{39}$$

The cathode activation overpotential, E_{act} , represents the driving force of the electrochemical reactions. The reaction kinetic is characterized by Tafel equation:

$$E_{act} = \frac{1}{\alpha} \frac{RT_{sol}}{nF} \ln \frac{j}{j_0} \tag{40}$$

where the exchange current density, j_0 , is calculated regarding to a reference exchange current density at a reference oxygen concentration:

$$j_0 = j_0^0 \left(\frac{C_{O_2}^{(s)}}{C_{O_2}^0} \right)^{0.5} \tag{41}$$

It is important to note that the Tafel equation is valid under the condition of $j > j_0$. If this does not hold, the over potential E_{act} equals to zero. In Eq. (41), the oxygen concentration at catalyst surface is calculated by the following:

$$KA_{mem}(C_{O_2} - C_{O_2}^{(s)}) = \frac{1}{2}\gamma_{H_2O}A_{mem} \quad (42)$$

where the mass coefficient K , is a function of RH in order to consider the flooding at cathode side and expressed as:

$$K = K_0 \left(1 - F_0 e^{\frac{RH-1}{\phi}}\right) \quad (43)$$

where K_0 is the normal, un-flooded mass transfer coefficient, ϕ and F_0 are newly defined porosity and anti-flooding coefficients, respectively. It is also assumed that there is no transport resistance for the water vapor moving from membrane surface toward the cathode chamber.

The third loss is caused by the mass transfer effects and defined as:

$$E_{mt} = -\frac{1}{2} \frac{RT_{sol}}{nF} \ln \left(\frac{C_{O_2}^{(s)}}{C_{O_2}} \right) \quad (44)$$

The last loss is related mainly to the ohmic loss in the membrane, and obeys:

$$E_{ohm} = j \cdot r_{mem}(C_{H_2O}, T_{sol}) \quad (45)$$

The ionic resistance of the membrane with respect to the membrane area is defined as $r_{mem} = t_{mem}/\sigma$, where the t_{mem} represents thickness of the membrane, and σ is the membrane conductivity which is defined as:

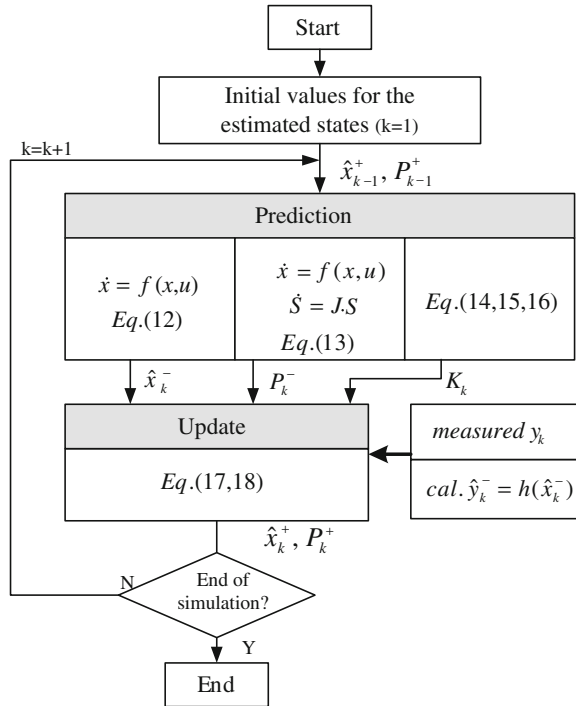
$$\sigma = (0.514\lambda - 0.326)e^{1268 \left(\frac{1}{303} - \frac{1}{T_{sol}} \right)} \quad (46)$$

where λ represents the water content in the membrane. In the work of Lauzze and Chmielewski, the water content in the membrane solely depends on RH at cathode chamber.

3.2 Implementation of the EKF

According to the selected model, the state vector x is $[C_{H_2O} \ C_{O_2} \ C_{N_2} \ T_{cat} \ T_{sol} \ T_{jac}]^T$, which represents vapor concentration, oxygen concentration, nitrogen concentration, the temperature of the gas stream in cathode side, cell temperature and coolant temperature, respectively. In the output vector $y = [u_{cell} \ T_{cat} \ T_{jac} \ p]^T$, the cell voltage, the temperature of the gas stream in cathode side, coolant temperature, and total gas pressure in cathode side are included. They are measured directly by the corresponding sensors. Depending on the measurable T_{cat} and p in output vector, the total concentration of the gas in cathode can be calculated through the ideal gas law. Hence, in order to access all the six states, a vector containing the states to be estimated $\hat{x} = [\hat{C}_{H_2O} \ \hat{C}_{O_2} \ \hat{T}_{sol}]^T$ is chosen.

Fig. 4 The flow chart for EKF algorithm implementation



In this study, the EKF is implemented in Matlab environment. The EKF algorithm is programmed subject to the flow chart in Fig. 4. Herein, the program starts with the given initial values for the states to be estimated. Afterwards, the pre-mentioned ‘prediction step’ is executed to calculate the \hat{x}_k^- by (12), P_k^- by (13) and Kalman gain K_k by (14)–(16). It is known that the value of $\frac{\partial \hat{y}}{\partial x}$ is critical for P_k^- and can be calculated by the sensitive equation. As discussed before, a Jacobian matrix of the system model is needed. In the next, the ‘update step’ is programmed to renovate the predicted states \hat{x}_k^- and covariance matrix P_k^- at the current sampling time. Finally, the updated \hat{x}_k^- and P_k^- (i.e. \hat{x}_k^+ and P_k^+) will be fed into the next prediction step as new initial values, if the end condition of the simulation is not fulfilled. Such procedure is executed iteratively in Matlab.

3.3 Simulation Results

In order to test the EKF for the state estimation in the PEMFC, the related simulation is carried out. In the simulation, the current density is applied as the system input, and its profile is given in Fig. 5. It contains two step-change at $t = 10$ s and $t = 20$ s, respectively.

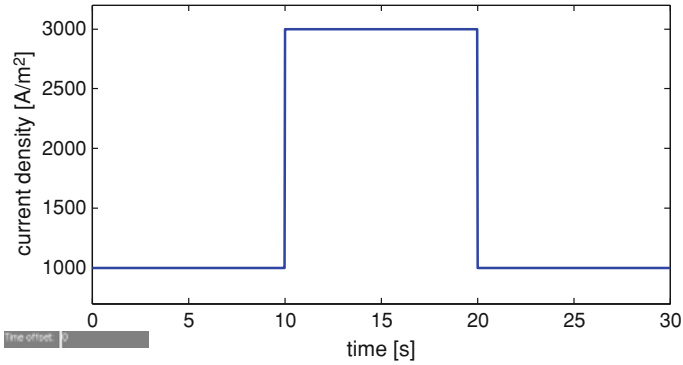


Fig. 5 Current density as the input of the PEM fuel cell system model

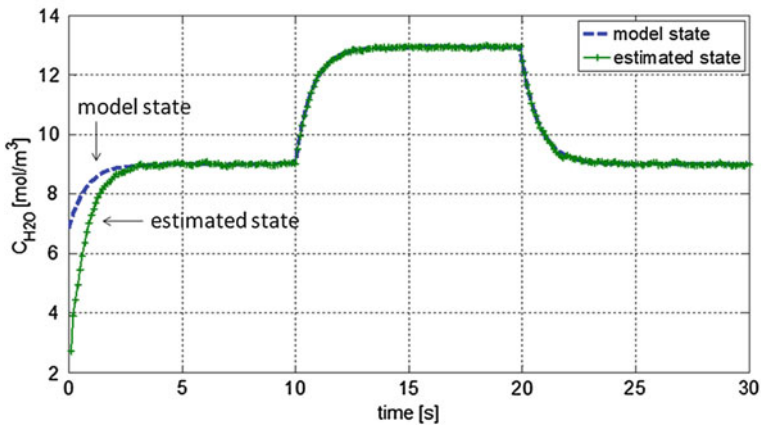


Fig. 6 Vapor concentration and its estimation

Figures 6 and 7 show the simulated results of vapor and oxygen concentrations and theirs estimations. The dashed line represents the profile of the concentration which is calculated via the model, the solid line is the estimated profile via the EKF approach. At the beginning, results of both model and EKF are different due to the initial values selection. After 5 s around, the results of EKF generally converge to the model results due to the EKF algorithm. Figure 8 shows the simulated results of cell temperature and its estimation. It is observed that the convergence time of cell temperature estimation is shorter than the other two states, around 1 s.

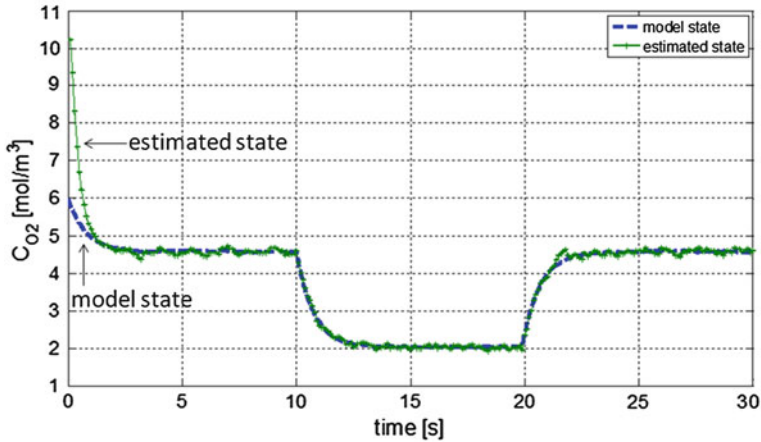


Fig. 7 Oxygen concentration and its estimation

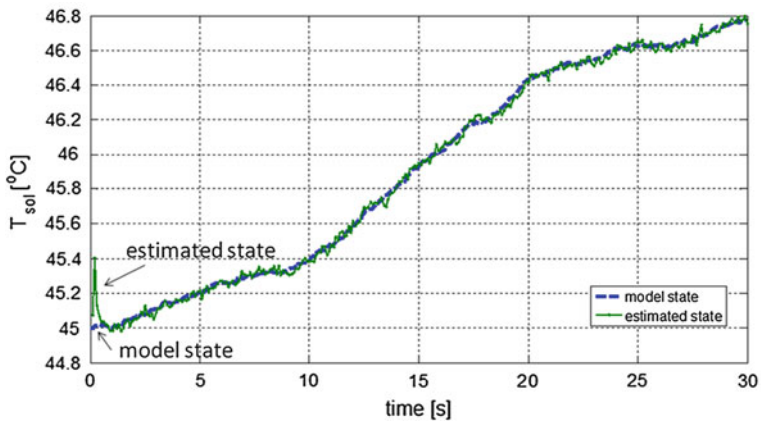


Fig. 8 Cell temperature and its estimation

4 Conclusion

In this paper, the principle of the EKF is introduced and illustrated by the example of the Van der Pol oscillator. The calculation of $\frac{\partial \psi}{\partial x} \Big|_{\hat{x}_{k-1}^+}$ is obtained by employing the sensitive equations. Afterwards, a PEM fuel cell system is chosen. The state estimations of three internal states are selected to be estimated via the EKF. During the implementation of the EKF, the sensitive equations are employed, resulting in no requirement of model discretization [12]. The simulation results show that the estimated states track the model results with a satisfactory manner.

The implementation of the EKF in the PEMFC in this paper can be used for the purpose of system control and monitoring. It can also be migrated to other nonlinear cases after the proper modifications.

References

1. Norgaard M, Poulsen NK, Ravn O (2000) New developments in state estimation for nonlinear systems. *Automatica* 36:1627–1638
2. Simon D (2009) Kalman filtering with state constraints: a survey of linear and nonlinear algorithms. *IET Control Theory Appl*
3. Kandepu R, Foss B, Imsland L (2007) Applying the unscented Kalman filter for nonlinear state estimation. *J Process Control* 18(7–8):753–768
4. Haseltine EL, Rawlings JB (2005) Critical evaluation of extended Kalman filtering and moving-horizon estimation. *Ind Eng Chem Res* 44:2451–2460
5. Mandela RK, Rengaswamy R, Narasimhan S, Sridhar LN (2010) Recursive state estimation techniques for nonlinear differential algebraic systems. *Chem Eng Sci* 65:4548–4556
6. Groetsch M, Mangold M, Sheng M, Kienle A (2006) State estimation of a molten carbonate fuel cell by an extended Kalman filter. In: 16th European symposium on computer aided process engineering and 9th international symposium on process systems engineering, pp 1161–1166
7. Kandepu R, Huang B, Imsland L, Foss B (2007) Comparative study of state estimation of fuel cell hybrid system using UKF and EKF. In: IEEE 2007 control and automation conference, pp 1162–1167
8. Soares GE, Kosanovich KA (1997) Parameter and state estimation of a proton-exchange membrane fuel cell using sequential quadratic programming. *Ind Eng Chem Res* 36:4264–4272
9. Goerguen H, Arcak M, Barbir F (2006) An algorithm for estimation of membrane water content in PEM fuel cells. *J Power Sour* 157:389–394
10. Lauzze KC, Chmielewski DJ (2006) Power control of a polymer electrolyte membrane fuel cell. *Ind Eng Chem Res* 45:4661–4670
11. Springer T, Zawodzinski T, Gottesfeld T (1991) Polymer electrolyte fuel cell model. *J Electrochem Soc* 138(8):2334–2342
12. Liu J, Ji G, Zhou S (2012) Applying and implementation of the extended Kalman filter (EKF) with sensitive equation: a PEMFC case study. In: Lecture notes in engineering and computer science: proceedings of the international multiconference of engineers and computer scientists 2012, IMECS 2012, 14–16 March, 2012, Hong Kong, pp 864–868

Global Localization and Position Tracking of Autonomous Transport Vehicles

Christof Röhrig, Christopher Kirsch, Julian Lategahn
and Marcel Müller

Abstract This chapter presents global localization and position tracking for a swarm of autonomous transport vehicles which transport Euro-bins in a distribution center or warehouse. Localization is realized by sensor fusion of range measurements obtained from an IEEE 802.15.4a network and laser range finders. The IEEE 802.15.4a network is used for communication as well as for global localization. Laser range finders are used to detect landmarks and to provide accurate positioning for docking maneuvers. Range measurements are fused in a Monte Carlo Particle Filter. The chapter presents the design of the global localization and position tracking algorithms. Experimental results are given to prove the effectiveness of the proposed methods.

Keywords Localization · IEEE 802.15.4a CSS · Autonomous transport vehicle · Mobile robot · Automated guided vehicle · AGV · Swarm intelligence

C. Röhrig (✉) · C. Kirsch · J. Lategahn · M. Müller
Intelligent Mobile Systems Lab, University of Applied Sciences and Arts in Dortmund,
Emil-Figge-Str. 42, 44227 Dortmund, Germany
e-mail: christof.roehrig@fh-dortmund.de

C. Kirsch
e-mail: christopher.kirsch@fh-dortmund.de

J. Lategahn
e-mail: julian.lategahn@fh-dortmund.de

M. Müller
e-mail: marcel.mueller@fh-dortmund.de

1 Introduction

Short production cycles and just-in-time inventory management require flexible material flow as well as usage of small transportation units. These demands can be met by using small autonomous transport vehicles which act as a swarm. Several companies have introduced small transport vehicles for logistic applications. Examples are “The Kiva Mobile Fulfillment System (MFS)”, the “Adept Courier” from Adept Technology, “RoboCourierTM” from swisslog and “ADAMTM (Autonomous Delivery and Manipulation)”. Inexpensive localization of transport vehicles is an important issue for many logistic applications and object of current research activities. The Kiva MFS uses bar codes on the floor that can be detected with a camera by the vehicles [17]. These bar codes specify the pathways and guarantee accurate localization. Drawbacks of this solution are the risk of polluting the bar codes and the need for predefined pathways which restrict the movements of the vehicles. The “Self-Driving Courier”, RoboCourierTM and ADAMTM are based on the same technology, which was developed by MobileRobots in Amherst USA. These transport vehicles use open path navigation with laser range finders to travel to their destination. Laser range finders can be used to track the position of an autonomous vehicle within a known environment using a predefined map, if the initial position is given, but it is difficult to find the initial position in a complex or dynamically changing environment without a priori information.

The chapter proposes an open path navigation, which is based on sensor fusion of range measurements using an IEEE 802.15.4a wireless network, measurements of laser range finders and dead-reckoning (odometry). The IEEE 802.15.4a wireless network is used for communication and global localization (without a priori information), the laser range finders and odometry are used for position tracking and improved local accuracy.

In the target application, a distribution center, transport vehicles transport bins with Euro footprint (600×400 mm) from a high bay racking to order picking stations and back to the racking. Order pickers collect the orders from Euro-bins and pack them into custom bins. This so called *Cellular Transport System* is based on the Multishuttle Move (MSM) technology [7]. MSM is a fusion of a conventional rack shuttle and a transport vehicle developed by Fraunhofer-Institute for Material Flow and Logistics (FhG IML). The vehicles are rail-guided while they are located in the racking system or the lift. The vehicles are able to leave the rail-system and to operate as transport vehicles with open path navigation. This scalable and flexible vehicle swarm concept is a compact, adaptable solution for high storage capacity and covers the entire performance spectrum of facility logistics with the maximum possible flexibility [7]. Since the vehicles navigate autonomously and act as a swarm, real-time communication and global localization is needed.

This chapter proposes the usage of an IEEE 802.15.4a Wireless Sensor Network (WSN) for communication as well as for global localization. A WSN consists of spatially distributed autonomous sensor nodes for data acquisition. The chapter

describes location tracking of transport vehicles using an Extended Kalman Filter and global localization using a Monte Carlo Particle Filter. Experimental results are given to prove the effectiveness of the proposed methods. This chapter extends the work presented in [13] to global localization using a Monte Carlo Particle Filter.

2 Related Work

Up to now several kinds of localization techniques have been developed for the use in wireless networks. A review of existing techniques is given in [16]. These techniques can be classified by: Connectivity, Received Signal Strength (RSS), Angle of Arrival (AoA), Time of Arrival (ToA) and Round-trip Time of Flight (RToF).

Connectivity information is available in all kinds of wireless networks. The accuracy of localization depends on the range of the used technology and the density of the beacons. In cellular networks Cell-ID is a simple localization method based on cell sector information.

RSS information can be used in most wireless technologies, since mobile devices are able to monitor the RSS as part of their standard operation. The distance between sender and receiver can be obtained with the Log Distance Path Loss Model described in [9]. Unfortunately, the propagation model is sensitive to disturbances such as reflection, diffraction and multi-path effects. The signal propagation depends on building dimensions, obstructions, partitioning materials and surrounding moving objects. Own measurements show, that these disturbances make the use of a propagation model for accurate localization in an indoor environment almost impossible [12]. A method to overcome this disadvantage is fingerprinting which is introduced in [3] and uses a radio map. Fingerprinting is divided in two phases: In the initial calibration phase, the radio map is built by moving around and storing RSS values at various predefined points of the environment. In the localization phase, the mobile device moves in the same environment and the position is estimated by comparing the current RSS values with the radio map.

ToA and RToF estimate the range to a sender by measuring the signal propagation delay. The Cricket localization system [10] developed at MIT utilizes a radio signal and an ultrasound signal for position estimation based on trilateration. TDoA of these two signals are measured in order to estimate the distance between two nodes. This technique can be used to estimate the position of a node in a WSN [8] or to track the position of a transport vehicle [1]. Ultra-Wideband (UWB) offers a high potential for range measurement using ToA, because the large bandwidth (> 500 MHz) provides a high ranging accuracy [6]. In [5] UWB range measurements are proposed for tracking a vehicle in a warehouse. The new WSN standard IEEE 802.15.4a specifies two optional signaling formats based on UWB and Chirp Spread Spectrum (CSS) with a precision ranging capability [14].

Nanotron Technologies distributes the nanoLOC TRX Transceiver with ranging capabilities using CSS as signaling format.

3 Location Tracking Using the Extended Kalman Filter

The Kalman Filter is an efficient recursive filter, which estimates the state of a dynamic system out of a series of incomplete and noisy measurements by minimizing the mean of the squared error. It is also shown to be an effective tool in applications for sensor fusion and localization [15].

The basic filter is well-established, if the state transition and the observation models are linear distributions. In the case, if the process to be estimated and/or the measurement relationship to the process is specified by a non-linear stochastic difference equation, the Extended Kalman Filter (EKF) can be applied. This filtering is based on linearizing a non-linear system model around the previous estimate using partial derivatives of the process and measurement function.

The Extended Kalman Filter is suitable to track the x- and y-position of a mobile system using measured distances to artificial landmarks (anchors). To estimate the initial position of a mobile system, at least three distances are necessary. Using trilateration the anchor distances r_i are calculated as follow:

$$r_i = \sqrt{(p_x - a_{x,i})^2 + (p_y - a_{y,i})^2}, \quad (1)$$

where $(a_{x,i}, a_{y,i})$ are the x- and y-positions of anchor i and (p_x, p_y) represents the x- and y-position of the mobile system to be located.

To gain the unknown initial position, Eq. (1) are solved for p_x and p_y , and are transformed in matrices:

$$\mathbf{H} \cdot \begin{pmatrix} p_x \\ p_y \end{pmatrix} = \mathbf{z} \quad \text{with } \mathbf{H} = \begin{pmatrix} 2 \cdot a_{x,1} - 2 \cdot a_{x,2} & 2 \cdot a_{y,1} - 2 \cdot a_{y,2} \\ \vdots & \vdots \\ 2 \cdot a_{x,1} - 2 \cdot a_{x,n} & 2 \cdot a_{y,1} - 2 \cdot a_{y,n} \end{pmatrix}, \quad (2)$$

$$\text{and } \mathbf{z} = \begin{pmatrix} r_2^2 - r_1^2 + a_{x,1}^2 - a_{x,2}^2 + a_{y,1}^2 - a_{y,2}^2 \\ \vdots \\ r_n^2 - r_1^2 + a_{x,1}^2 - a_{x,n}^2 + a_{y,1}^2 - a_{y,n}^2 \end{pmatrix},$$

where n is the overall number of anchor nodes. Equation 2 can be solved using the method of least squares:

$$\begin{pmatrix} \hat{p}_x \\ \hat{p}_y \end{pmatrix} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \cdot \mathbf{z} \quad (3)$$

For location tracking using EKF, Eq. (2) needs only to be solved for the initial position estimate $\hat{\mathbf{x}}_0$. The EKF addresses the general problem of estimating the interior process state of a time-discrete controlled process, that is governed by non-linear difference equations:

$$\tilde{\mathbf{x}}_{k+1} = \mathbf{f}(\hat{\mathbf{x}}_k, \mathbf{u}_k, \mathbf{w}_k), \tilde{\mathbf{y}}_{k+1} = \mathbf{h}(\tilde{\mathbf{x}}_{k+1}, \mathbf{v}_{k+1}). \quad (4)$$

The state vector contains the position of the mobile robot $\mathbf{x}_k = (p_x, p_y)^T$. The optional input control vector $\mathbf{u}_k = (v_x, v_y)^T$ contains the desired velocity of the vehicle. These values are set to zero, if the input is unknown. The observation vector \mathbf{y}_k represents the observations at the given system and defines the entry parameters of the filter, in this case the results of the range measurements. The process function \mathbf{f} relates the state at the previous time step k to the state at the next step $k + 1$. The measurement function \mathbf{h} acts as a connector between \mathbf{x}_k and \mathbf{y}_k . The notation $\tilde{\mathbf{x}}_k$ and $\tilde{\mathbf{y}}_k$ denotes the approximated *a priori* state and observation, $\hat{\mathbf{x}}_k$ typifies the *a posteriori* estimate of the previous step. Referring to the state estimation, the process is characterized with the stochastic random variables \mathbf{w}_k and \mathbf{v}_k representing the process and measurement noise. They are assumed to be independent, white and normal probably distributed with given covariance matrices \mathbf{Q}_k and \mathbf{R}_k . To estimate a process with non-linear relationships the equations in (4) must be linearized as follow:

$$\begin{aligned} \mathbf{x}_{k+1} &\approx \tilde{\mathbf{x}}_{k+1} + \mathbf{F}_{k+1} \cdot (\mathbf{x}_k - \hat{\mathbf{x}}_k) + \mathbf{W}_{k+1} \cdot \mathbf{w}_k \\ \mathbf{y}_{k+1} &\approx \tilde{\mathbf{y}}_{k+1} + \mathbf{C}_{k+1} \cdot (\mathbf{x}_{k+1} - \tilde{\mathbf{x}}_{k+1}) + \mathbf{V}_{k+1} \cdot \mathbf{v}_{k+1}, \end{aligned} \quad (5)$$

where \mathbf{F}_{k+1} , \mathbf{W}_{k+1} , \mathbf{C}_{k+1} and \mathbf{V}_{k+1} are Jacobian matrices with the partial derivatives:

$$\begin{aligned} \mathbf{F}_{k+1} &= \frac{\partial \mathbf{f}}{\partial \mathbf{x}}(\hat{\mathbf{x}}_k, \mathbf{u}_k, 0) & \mathbf{W}_{k+1} &= \frac{\partial \mathbf{f}}{\partial \mathbf{w}}(\hat{\mathbf{x}}_k, \mathbf{u}_k, 0) \\ \mathbf{C}_{k+1} &= \frac{\partial \mathbf{h}}{\partial \mathbf{x}}(\tilde{\mathbf{x}}_{k+1}, 0) & \mathbf{V}_{k+1} &= \frac{\partial \mathbf{h}}{\partial \mathbf{v}}(\tilde{\mathbf{x}}_{k+1}, 0). \end{aligned} \quad (6)$$

Because in the analyzed system the predictor equation contains a linear relationship, the process function \mathbf{f} can be expressed as a linear equation:

$$\mathbf{x}_{k+1} = \mathbf{F}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k + \mathbf{w}_k, \quad (7)$$

where the transition matrix \mathbf{F} and \mathbf{B} are defined as:

$$\mathbf{F} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} T & 0 \\ 0 & T \end{pmatrix}, \quad (8)$$

where T is the constant sampling time.

The observation vector \mathbf{y}_k contains the current measured distances:

$$\mathbf{y}_k = (r_1 \quad \cdots \quad r_n)^T. \quad (9)$$

The initial state estimate $\hat{\mathbf{x}}_0$ is calculated based on (2). For the subsequent estimation of the position $\mathbf{x} = (p_x, p_y)$ the functional values of the non-linear measurement function \mathbf{h} must be approached to the real position. The function \mathbf{h} comprises the trilateration Eq. (1) and calculates the approximated measurement $\tilde{\mathbf{y}}_{k+1}$ to correct the present estimation $\tilde{\mathbf{x}}_{k+1}$. The equation $\tilde{\mathbf{y}}_{k+1} = \mathbf{h}(\tilde{\mathbf{x}}_{k+1}, \mathbf{v}_{k+1})$ is given as:

$$\begin{pmatrix} \hat{r}_1 \\ \vdots \\ \hat{r}_n \end{pmatrix} = \begin{pmatrix} \sqrt{(\tilde{p}_x - a_{x,1})^2 + (\tilde{p}_y - a_{y,1})^2} \\ \vdots \\ \sqrt{(\tilde{p}_x - a_{x,n})^2 + (\tilde{p}_y - a_{y,n})^2} \end{pmatrix} + \mathbf{v}_{k+1}. \quad (10)$$

The related Jacobian matrix $\mathbf{C}_{k+1} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}}(\tilde{\mathbf{x}}_k, 0)$ describes the partial derivatives of \mathbf{h} with respect to \mathbf{x} :

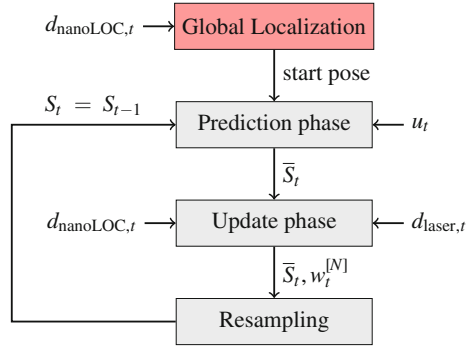
$$\mathbf{C}_{k+1} = \begin{pmatrix} \frac{\partial \hat{r}_1}{\partial p_x} & \frac{\partial \hat{r}_1}{\partial p_y} \\ \vdots & \vdots \\ \frac{\partial \hat{r}_n}{\partial p_x} & \frac{\partial \hat{r}_n}{\partial p_y} \end{pmatrix} \quad \text{with} \quad \begin{aligned} \frac{\partial \hat{r}_i}{\partial p_x} &= \frac{\tilde{p}_x - a_{x,i}}{\sqrt{(\tilde{p}_x - a_{x,i})^2 + (\tilde{p}_y - a_{y,i})^2}} \\ \frac{\partial \hat{r}_i}{\partial p_y} &= \frac{\tilde{p}_y - a_{y,i}}{\sqrt{(\tilde{p}_x - a_{x,i})^2 + (\tilde{p}_y - a_{y,i})^2}}. \end{aligned} \quad (11)$$

Given that \mathbf{h} contains non-linear difference equations the parameters r_i as well as the Jacobian matrix \mathbf{C}_{k+1} must be calculated newly for each estimation.

4 Global Localization Using the Monte Carlo Particle Filter

The KF and EKF rely on the assumption, that motion and sensor errors are Gaussian and that the estimated position can be modeled by using a Gaussian distribution. Because of this fact, KF and EKF can not handle position ambiguities. Another method which is based on the Bayesian filter is a Particle Filter (PF). A PF can handle position ambiguities and does not rely on the assumption that motion and sensor errors are Gaussian. Also PF can cope with multimodal distributions. In a PF, a set S of N samples is distributed in the environment or at known places. A sample s is defined by cartesian coordinates and an orientation. A widely used PF for transport vehicle localization is the Monte Carlo Particle Filter (MCPF), which is described in [4]. The estimated pose of a transport vehicle and its uncertainty about the correctness is represented by the samples. MCPF consists of two phases: The *prediction* phase and the *update* phase. Inside the prediction phase the motion information \mathbf{u}_t are applied on each sample s_{t-1}^i ($1 \leq i \leq N$). The prediction phase is also called *motion model*. The result of the motion model is a new

Fig. 1 MCPF flow chart



set of samples \bar{S}_t which represents the positions, where the transport vehicle could be after executing the movement u_t .

Inside the update phase, the set of distance measurements D_t is used to assign each sample with an importance factor w . The importance factor complies the probability $p(D_t | s_t^i, m)$, i.e. the probability of the distance measurements D_t at a point in the environment defined by sample s_t^i and by using the information from the map m . In m the positions of anchors and landmarks are stored. The result of the update phase—also called *measurement update*—is the set of samples \bar{S}_t of the prediction phase with the corresponding set of N importance weights w_t . Both sets together represent the current position likelihood of the transport vehicle. After the update phase, the resampling step follows. Inside the resampling step, samples with a low importance weight are removed and samples with a high importance factor are duplicated. The result of the resampling is the set S_t of N samples which represents the current position of the transport vehicle. In the next time step, the set S_t is used as S_{t-1} . There are two possibilities to extract the pose of the transport vehicle out of the sample set S_t : The first method is to use the weighted mean of all samples and the second method is to use the sample with the highest importance factor. MCPFs flow chart is drafted in Fig. 1. The MCPF has the advantages that it copes with global localization (no a priori information) and position tracking (given a priori information). The sensor fusion with some dependencies and special cases can be implemented easily. Generally the combination of sensor specific advantages and the compensation of sensor specific disadvantages is called sensor fusion.

In this work the MCPF uses distance measurements from the IEEE 802.15.4a WSN for global localization. The global localization task, which has to be done before the MCPF starts, is shown as a red block in Fig. 1. The probability density function of the IEEE 802.15.4a measurement error which is used to compute the importance factor, is presented in the next section.

4.1 IEEE 802.15.4a Measurement Model

In the update phase, the measurement model is used to calculate the importance factor w for each sample s . The measurement model is the probability density function $p(d_{\text{nanoLOC}}, k | s_k^i, m)$ which characterizes the measurement properties and error. The measurement set d_{nanoLOC}, k contains distance measurements to A anchors. The density function depends on sensors and environment. To estimate the density function for IEEE 802.15.4a distance measurements, line of sight measurements to four anchors are taken while a transport vehicle moves a straight path between them. While the vehicle moves, an accurate position was estimated by laser measurements to two walls. In Fig. 2, error histograms of measurements to four anchors are shown. The error is the difference between measured distance d_k^a and the Euclidean distance from vehicle position to anchor a .

The histograms show, that all measured distances are too large, the average error is 107 cm. The error depends on the position of the anchor and on the environment. The median and standard deviation of the error distributions are different but they all have a Gaussian structure. Owing to that fact, it is possible, to use a Gaussian distribution as IEEE 802.15.4a probability density function:

$$\mathcal{N}(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2} \frac{(x - \mu)^2}{\sigma^2}\right) \quad (12)$$

To calculate the importance weight of sample s_k^i , the Euclidean distance $d_k^{a,i*}$ between this sample and the anchor a is calculated as

$$d_k^{a,i*} = \sqrt{(x_i - x_a)^2 + (y_i - y_a)^2}, \quad (13)$$

where (x_a, y_a) is the position of anchor a and (x_i, y_i) are the Cartesian coordinates of sample s_k^i . The Euclidean distance and the measured distance d_k^a are used with an anchor specific constant d_c^a in a fixed Gaussian distribution:

$$p(d_{\text{nanoLOC}}^a, k | s_k^i, m) = \mathcal{N}(d_k^{a,i*} - (d_c^a - d_c^a), 0, \sigma^2) \quad (14)$$

where d_c^a is the median of the distance errors shown in the histograms. The advantages of this fixed Gaussian distribution are, that a normalization during the localization, to guarantee $\sum p = 1$, is not needed and that the domain can be restricted. This last advantage can be used to detect estimation failure. If a lot of samples are out of range, new samples can be drawn in the environment. This fact enables the MCPF to re-localize the transport vehicle.

The importance factor of a sample i is calculated with:

$$w_k^i = \prod_{a=1}^A p(d_{\text{nanoLOC}}^a, k | s_k^i) \cdot \prod_{l=1}^2 p(d_{\text{laser},k}^l | s_k^i) \quad (15)$$

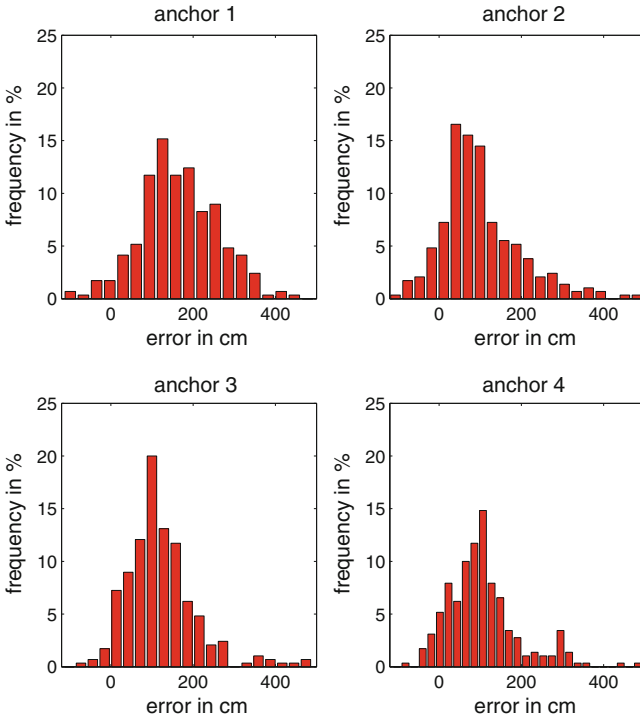


Fig. 2 Error histograms of IEEE 802.15.4a distance measurements to four anchors. The x -axis is the error in centimeter and the y -axis show their frequency in %

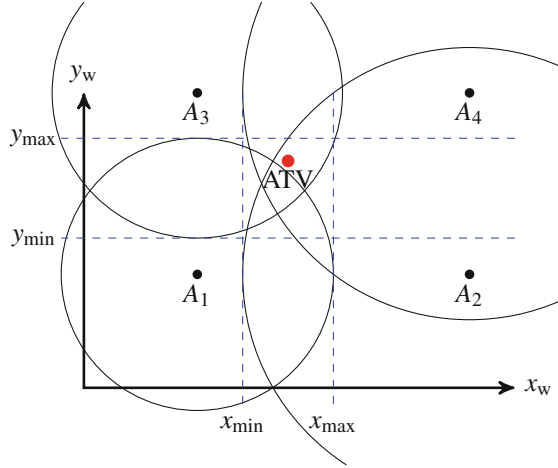
The importance factor w is the product of the probability of measurements to A anchors and to two landmarks. The probability $p(d_{\text{laser}}^l, k | s_k^i, m)$ is a fixed Gaussian with $\sigma = 28$ mm. The landmarks are equipped with reflectors, in order to allow easy detection by the laser range finders. If no landmarks are detected, the importance factor is equal to the probability of the distance measurements to A anchors.

The next section presents the global localization approach which uses distance measurements of the IEEE 802.15.4a WSN.

4.2 Anchorbox

For global localization range measurements of the IEEE 802.15.4a WSN are used to reduce the area in which particles are distributed. This method is based on a technique which was presented in [2]. Figure 3 shows an example of an Anchorbox which is computed by using range measurements to four anchors. The red dot is a vehicle which is equipped with a node.

Fig. 3 Anchorbox example by using range measurements to four anchor nodes with known positions



In the first step of the MCPF the particles are distributed in the area defined through calculation specifications 16 and 17, where (x_i, y_i) is the position of anchor i and \bar{d}_i is the average of I range measurements.

$$x_{\min} = \max_{i=1}^I (x_i - \bar{d}_i) \quad x_{\max} = \min_{i=1}^I (x_i + \bar{d}_i) \quad (16)$$

$$y_{\min} = \max_{i=1}^I (y_i - \bar{d}_i) \quad y_{\max} = \min_{i=1}^I (y_i + \bar{d}_i) \quad (17)$$

One disadvantage of the Anchorbox approach is that the orientation can not be estimated by IEEE 802.15.4a range measurements. To overcome this disadvantage more particles with a random orientation are distributed at the beginning of the algorithm. After the particles are distributed the vehicle drives 1 m in positive x vehicle direction. Because of this technique, the adapted MCPF is an active localization technique. Thereby, particles with an incorrect orientation remove themselves from the correct position and are getting a lower importance factor during the first measurement update. Another possibility to overcome this disadvantage is to compute the position through trilateration during the vehicle drives the path. The orientation can be estimated by computing the average orientation between these trilateration points. This approach is not used because these points vary significantly and the estimated orientation will be erroneous.

The advantage of the developed technique is a smaller particle cloud at the beginning of the algorithm. Because of this, the first position can be estimated faster. Another advantage is that this technique can also be used to solve the *kidnapped-robot problem*. To solve this problem, the Anchorbox can be computed in every MCPF cycle. Inside this computed Anchorbox a sample subset can be distributed to represent a much larger area where the vehicle can be. The advantage of this

Fig. 4 Wireless sensor node for anchors and mobile tags



approach is, that it does not rely on the odometry data, which is necessary to solve the kidnapped robot problem.

5 Implementation and Experimental Results

5.1 Hardware

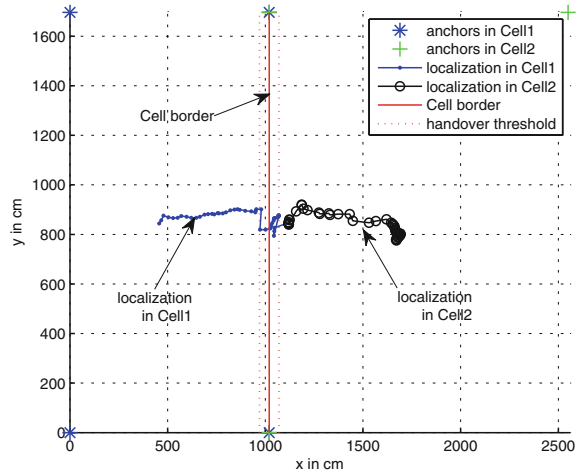
In order to fulfill the requirements of the target application, a wireless sensor board was developed that can be used as:

- Mobile node (tag) on a transport vehicle,
- Fixed anchor node,
- Master node with connection to the distributed system.

The board is designed around a STM32 micro-controller which includes an ARM Cortex-M3 core. The STM32 micro-controller provides interfaces and enough RAM and computational power to perform communication and location tracking using EKF in real-time. IEEE 802.15.4a radio is built with a nanoPAN 5375 module which supports up to 20 dBm output power and three frequency channels with 22 MHz bandwidth (Fig. 4).

The architecture of the wireless sensor board is modular, only necessary components are assembled. Master nodes are equipped with a Xport to connect to an Ethernet. Mobile nodes are equipped with an IMU (inertial measurement unit) which increases localization accuracy of the vehicles. Mobile nodes are connected via CAN-bus to the vehicle's PLC (programmable logic controller). Communication to the PLC is performed with CANopen protocol. As a fall back, the boards are equipped with a serial interface (RS-232).

Fig. 5 Wireless synchronisation



5.2 Position Tracking Using EKF

Several experiments have been conducted, to prove the localization accuracy. Figure 5 shows the result of a roaming experiment. The vehicle moves from Cell1 to Cell2 and performs a handover while crossing the boundary between the cells. The position tracking of the transport vehicle is estimated using the EKF as described in Sect. 3, the initial position is calculated with trilateration (Eq. (2)). The dots in Fig. 5 show the position tracked in Cell1, the circles show the position in Cell2. The position error near the border of the cells are caused by bad radio conditions in this area due to the directional antennas of the anchors. In this experiment, only range measurement using four anchors of each cell are used for tracking. The tracking error can be decreased, if odometry and laser range finders are included in the tracking algorithm [11].

5.3 Global Localization Using MCPF

To evaluate the proposed MCPF global localization, some experiments are conducted at the University of Applied Sciences and Arts in Dortmund with an omnidirectional transport vehicle, that is equipped with Mecanum wheels. The vehicle and it's motion model is described in [11]. It is equipped with two SICK S300 Professional laser range finders with a scanning angle of 270° . With both laser range finders, the vehicle gets a full 360° scan of the environment. The laser range finders provide a resolution $\Delta\alpha$ of 0.5° . A docking station for handing over bins serves as landmark. Two pillars of the docking station are equipped with reflectors, in order to allow easy detection by the laser range finders. The transport vehicle is also equipped with a IEEE 802.15.4a tag for ranging and communication

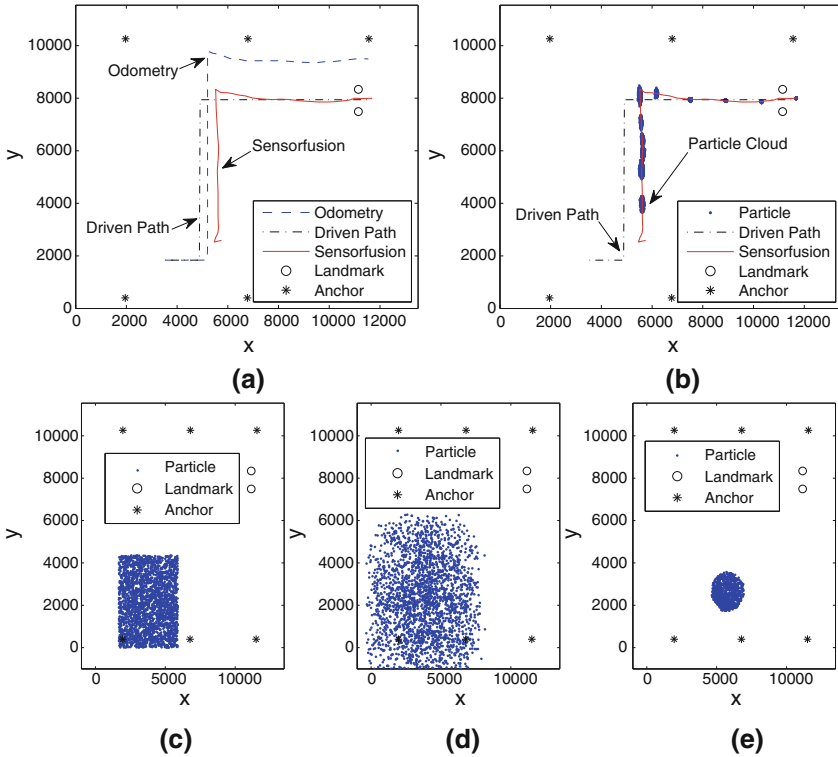


Fig. 6 Experimental results: **a** and **b** show results of global localization. **c–e** Show the process of global localization with the Anchorbox approach

purposes. At the margins of the environments six IEEE 802.15.4a anchors are placed. The Fig. 6a, b shows the first steps of the global localization and the estimated positions with a comparison to the driven path and the odometry data (dimensions in millimeter).

The transport vehicle is moved in manual mode from a starting point into the docking station which is shown by the black path in Fig. 6a, b. The transport vehicle is moved forwards first, then sideways and finally forwards into the docking station in the upper right corner, always with the same orientation $\theta = 0^\circ$. During the movement of the vehicle, all necessary sensor data for MCPF are stored. These values are odometry data, distance measurements to six IEEE 802.15.4a anchors and the laser range data. The first movement should simulate the first step of global localization if the transport vehicle acts in automatic mode with no a priori position information. The second and third movement represent the estimated path into the docking station, which guarantees a precise localization.

To perform the global localization first of all the Anchorbox is computed by using distance measurements to six anchors. Then 10000 samples with random

orientation are distributed inside the Anchorbox. Figure 6c shows the Anchorbox with the distributed sample set. The sample set after the first movement part is shown in Fig. 6d. It can be seen, that samples with an incorrect orientation have moved away from the real position. Subsequently the first importance factor is computed by using range measurements of the WSN. The resulting sample cloud of the resampling step is shown in Fig. 6e. The start pose is set to the weighted mean of the sample cloud.

After estimating the start position the MCPF segue from global positioning into position tracking and the sample set is reduced to 2000 samples. The estimated start position depends on the IEEE 802.15.4a measurements and on the set of samples with random orientation.

The resulting path of the global localization and position tracking is shown in Fig. 6a, b. Owing to an unequal floor contact, the vehicle has a large slippage when it moves sideways. Figure 6a shows odometry in blue and MCPF estimation using all sensor data in red. The sample clouds resulting from position tracking (after the global localization) are presented in Fig. 6b. Until the vehicle detects the landmark pair, the importance factors were computed by using the range measurements of the WSN. This results in bigger sample clouds and a higher position uncertainty, which can be seen in Fig. 6b. During the last movement the vehicle detects the landmark pair with the two laser range finders. Because of this, the resulting sample clouds are compressed and the uncertainty of the estimated position is reduced. The position estimated by using IEEE 802.15.4a measurements is good enough for planning the path and through using the sensor fusion a successful docking maneuver can be guaranteed.

6 Conclusions

In this chapter global localization of autonomous transport vehicles using IEEE 802.15.4a CSS and laser range finders was proposed. The IEEE 802.15.4a network is used for communication as well as for global localization. Laser range finders are used to detect landmarks and to provide accurate positioning for docking maneuvers. Range measurements are fused in a Monte Carlo Particle Filter. The presented experimental results presented in the chapter, show the effectiveness of the proposed methods.

Acknowledgments This work was supported by the Ministry of Innovation, Science and Research of the German State of North Rhine-Westphalia (FH-Extra, grant number 29 00 130 02/12) and the European Union Fonds for Regional Development (EFRE). Furthermore the project was financially supported by Nanotron Technologies GmbH in Berlin, Germany and the University of Applied Sciences and Arts in Dortmund (HIFF, project number 04 001 79).

References

1. Alriksson P, Rantzer A (2007) Experimental evaluation of a distributed kalman filter algorithm. In: Proceedings of the 46th IEEE conference on decision and control. New Orleans, pp 5499–5504
2. Baggio A, Langendon K (2006) Monte-Carlo localization for mobile wireless sensor networks. Technology report of Delft University (PDS: 2006–004)
3. Bahl P, Padmanabhan VN (2000) RADAR: an in-building RF-based user location and tracking system. Proceedings of the 19th annual joint conference of the IEEE computer and communications societies, vol 2. Tel Aviv, Israel, pp 775–784
4. Dellaert F, Fox D, Burgard W, Thrun S (1999) Monte Carlo localization for mobile robots. In: Proceedings of the IEEE international conference on robotics and automation (ICRA99)
5. Fernández-Madriral JA, Cruz E, González J, Galindo C, Blanco JL (2007) Application of UWB and GPS technologies for vehicle localization in combined indoor-outdoor environments. In: Proceedings of the international symposium on signal processing and its applications. Sharja, United Arab Emirates
6. Gezici S, Tian Z, Giannakis GB, Kobayashi H, Molisch AF, Poor HV, Sahinoglu Z (2005) Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks. *Signal Process Mag* 22(4):70–84
7. Kamagaew A, Stenzel J, Nettsträter A, Ten Hompel M (2011) Concept of cellular transport systems in facility logistics. In: Proceedings of the 5th international conference on automation, robots and applications (ICARA 2011)
8. Moore D, Leonard J, Rus D, Teller S (2004) Robust distributed network localization with noisy range measurements. Proceedings of the 2nd international conference on embedded networked sensor systems. Baltimore, USA, pp 50–61
9. Patwari N, Hero AO, Perkins M, Correal NS, O’Dea R (2003) Relative location estimation in wireless sensor networks. *IEEE Trans Signal Process* 51(8):2137–2148
10. Priyantha NB, Miu AKL, Balakrishnan H, Teller S (2001) The cricket compass for context-aware mobile applications. Proceedings of the 7th annual international conference on mobile computing and networking. Rome, Italy, pp 1–14
11. Röhrig C, Heß D, Kirsch C, Künemund F (2010) Localization of an omnidirectional transport robot using IEEE 802.15.4a ranging and laser range finder. In: Proceedings of the 2010 IEEE/RSJ international conference on intelligent robots and systems (IROS 2010). Taipei, Taiwan, pp 3798–3803
12. Röhrig C, Künemund F (2007) Estimation of position and orientation of mobile systems in a wireless LAN. In: Proceedings of the 46th IEEE conference on decision and control. New Orleans, USA, pp 4932–4937
13. Röhrig C, Lategahn J, Müller M, Telle L (2012) Global localization for a swarm of autonomous transport vehicles using IEEE 802.15.4a CSS. In: Lecture notes in engineering and computer science: proceedings of the international multicongference of engineers and computer scientists, (IMECS 2012). Hong Kong, pp 828–833
14. Sahinoglu Z, Gezici S (2006) Ranging in the IEEE 802.15.4a standard. In: Proceedings of the IEEE annual wireless and microwave technology conference (WAMICON ’06). Clearwater, Florida, USA, pp 1–5
15. Thrun S, Burgard W, Fox D (2005) Probabilistic robotics. MIT press, Cambridge
16. Vossiek M, Wiebking L, Gulden P, Wiegardt J, Hoffmann C, Heide P (2003) Wireless local positioning. *Microw. Mag* 4(4):77–86
17. Wurman PR, D’Andrea R, Mountz M (2008) Coordinating hundreds of cooperative, autonomous vehicles in warehouses. *AI Mag.* 29(1):9–19

An Intelligent Flow Measurement Technique by Venturi Flowmeter Using Optimized ANN

Santhosh Krishnan Venkata and Binoy Krishna Roy

Abstract An intelligent flow measurement technique by venturi flow meter using an optimized Artificial Neural Network (ANN) is presented. The objectives of the present work are (i) to extend the linearity range of measurement to 100 % of the full scale, (ii) to make the measurement technique adaptive of variation in (a) venturi diameter ratio, (b) discharge coefficient, (c) liquid density, and (d) liquid temperature, and (iii) to achieve objectives (i) and (ii) by using an optimized neural network.

Keywords Artificial neural networks • Flow measurement • Non linear estimation • Optimization • Sensor modeling • Venturi

1 Introduction

Flow control is a very important process in any industry, since control of many parameters in industry is achieved by control of flow. For example, temperature control is done by controlling the flow of steam; pressure control in pneumatic or hydraulic driven instruments is done by flow control of air or liquid, respectively etc. An important step for control of flow is the accurate measurement of flow. In other cases, inaccurate flow measurements or failure to take measurements can cause serious (or even disastrous) results. Liquid flow measuring instruments mostly derive its principle from Bernoulli laws. The flow rate is determined inferentially by measuring the liquid's velocity or the change in kinetic energy. Velocity depends

S. K. Venkata · B. K. Roy (✉)

Department of Electrical Engineering, National Institute of Technology, Silchar, India
e-mail: bkr_nits@yahoo.co.in

S. K. Venkata

e-mail: kv.santhu@gmail.com

on the pressure difference that is forcing the liquid to flow through a pipe or a conduit. Since the pipe's cross-sectional area is known, the average velocity is measure of the flow rate. Many sensors are used for this purpose. Venturi finds a very wide application because of its high sensitivity and ruggedness. However, the problem of offset, high non-linear response characteristics, dependence of output on the ratio between venturi and pipe diameter, liquid density, discharge coefficient, and temperature of liquid have restricted its use and further impose some difficulties. Several techniques, which are tedious and time consuming have been suggested to overcome the difficulties faced due to the nonlinear response characteristics of the venturi. Most of these techniques have discussed one or other method of linearization of the nonlinear response characteristic within certain range of input scale. Thus, it is not possible to use the full range of input scale for measurement of flow. Datasheet of venturi flowmeter suggests that only 10–60 % of input range is used for measurement in practice. Further, the process of calibration needs to be repeated whenever the diameter ratio and/or discharge coefficient and/or liquid under measurement is changed. The output of venturi is dependent on flow rate as well as temperature of the liquid. Hence, the problem of nonlinear response characteristics of a venturi further aggravates the situation when there is change in liquid temperature.

To overcome the above difficulties, an intelligent flow measurement technique is presented here using optimized neural network model. An optimized ANN is considered by comparing five various schemes, algorithms and number hidden layers based on minimum mean square error (MSE) and Regression close to 1. This optimum network is suitably used in the measurement to extend linearity range of the system and makes the output adaptive of variations in ratio of diameter between venturi and pipe, liquid density, discharge coefficient, and temperature.

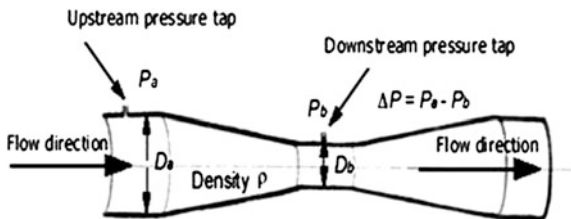
Literature survey suggests that in [1], calibration of orifice is discussed. In [2], measurement of flow for different areas of venturi nozzle is discussed. In [3], Calibration of flow meter is done with the help of microcontrollers. In [4, 5], a simulation model of venturi flow meter is used for measurement of flow rate is discussed. In [6–10] linearization of venturi is discussed using neural network algorithms. In [11], different flow measurements are discussed. In [12], linearization of venturi flow meter is discussed using mathematical computations. In [13], linearization of venturi flow meter using neural network and making the measurement independent of liquid density and temperature is discussed.

Since a venturi flow meter is used for intelligent flow measurement presented here, a mathematical model of venturi flow meter is discussed in the next section.

2 Venturi Flow Meter

A venturi nozzle (shown in Fig. 1) is a device used for measuring the volumetric flow rate. It uses the Bernoulli's principle which gives a relationship between the pressure and the velocity of the fluid. When the velocity increases, the pressure decreases and vice versa. A venturi nozzle is a tapered structure. It is usually placed in a pipe in

Fig. 1 Venturi nozzle



which fluid flows. When the fluid reaches the venturi nozzle, the fluid is forced to converge to go through the small hole. The point of maximum convergence actually occurs shortly downstream of the venturi nozzle, which is called vena-contracta point. As it flows so, the velocity and the pressure changes. Beyond the vena-contracta, the fluid expands and the velocity and pressure changes once again. The volumetric and mass flow rates can be obtained from Bernoulli’s equation by measuring the difference in fluid pressure between the normal pipe section and at the vena-contracta [14, 15]. The flow rate (Q) is given in Eq. 1.

$$Q = C_d \cdot A_b \cdot \sqrt{\frac{1}{(1 - \beta^4)}} \sqrt{\frac{2\Delta P}{\rho}} \tag{1}$$

where:

- C_d —Discharge coefficient
- A_b —Area of the flowmeter cross section
- β —Ratio of D_b to D_a
- P_a —Pressure at study flow
- P_b —Pressure at vena-contracta
- ρ —Density of liquid

The flow rate given in Eq. 1 varies with temperature due to variation of liquid density. The effect of temperature on the density [16, 17] is shown by

$$\rho_t = \frac{[\rho_{t_0}/(1 + \alpha(t - t_0))]}{[1 - \frac{P_t - P_{t_0}}{k}]} \tag{2}$$

where:

- ρ_t —specific density of liquid at temperature ‘t °C’
- ρ_{t_0} —specific density of liquid at temperature ‘t₀ °C’
- P_t —pressure at temperature ‘t °C’
- P_{t_0} —pressure at temperature ‘t₀ °C’
- k —Bulk modulus of liquid
- α —temperature coefficient of liquid

Again, the discharge coefficient is a function β , ρ , and Reynolds number (R_D) of the liquid. It is given in Eq. 3

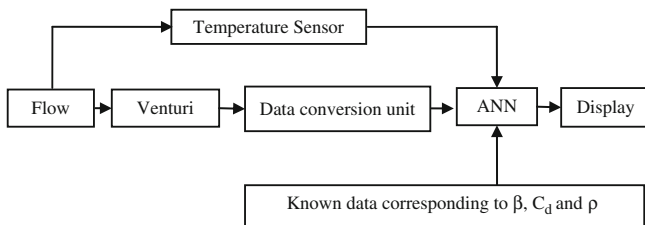


Fig. 2 Block diagram of the proposed flow measurement technique

$$C_d = 0.5959 + 0.0312\beta^{2.1} - 0.184\beta^8 + 0.0029\beta^{2.5} \left(\frac{\rho 10^6}{R_D} \right) \tag{3}$$

The block diagram of the flow measurement technique discussed here is given in the next section. Such representation is helpful in visualizing one of the possible solutions to overcome the difficulties discussed in Sect. 1.

3 Block Diagram

The block diagram representation of the flow measurement technique is shown in Fig. 2.

The Data Conversion Unit (DCU) consists of the pressure to voltage converter. The DMD 331 is a differential pressure transmitter for industrial applications. It is based on a piezo-resistive stainless steel sensor which can be pressurized on both sides with fluids or gases compatible with IEC 60770. The compact design allows an integration of the DMD 331 in machines and other applications with limited space. The DMD 331 calculates the difference between the pressure on the positive and the negative side and converts it into a proportional electrical signal within a range of 0–5 V [18].

4 Characteristics of Venturi with DCU

In this section, characteristic of venturi is simulated to understand the difficulties associated with the available measuring scheme. For this purpose, simulation is carried out with three different ratios of diameter between venturi and pipe is considered. These are $\beta = 0.3, 0.6,$ and 0.9 . Three different specific densities as $\rho = 500, 1000,$ and 1500 kg/m^3 are chosen. Three different discharge coefficients as $C_d = 0.85, 0.9,$ and 0.95 are chosen. Three different temperatures, like $t = 50,$ and $100,$ and $150 \text{ }^\circ\text{C}$ are used to find the output differential pressure (ΔP) of venturi nozzle with respect to various values of input flow considering a particular diameter ratio between venturi and pipe, discharge coefficient, liquid density, and temperature. These output pressure are used as inputs of data conversion circuit and output voltages are generated. The simulation results are shown in Figs. 3, 4, 5 and 6.

Fig. 3 Input flow versus output voltage of DCU for variation of flow rate and temperature for diameter ratio of 0.3, liquid density of 500 kg/m³ and discharge coefficient of 0.95

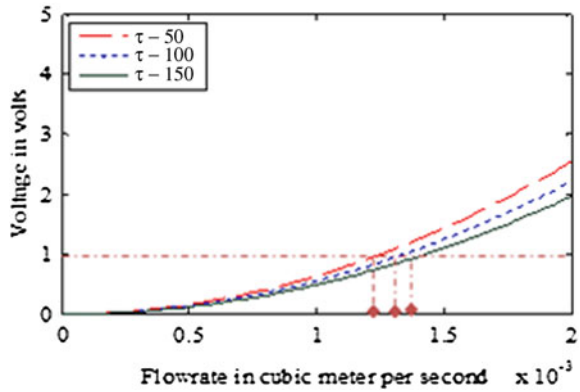


Fig. 4 Input flow versus output voltage of DCU for variation of flow rate and diameter ratio for temperature of 25 °C, discharge coefficient of 0.95 and liquid density of 500 kg/m³

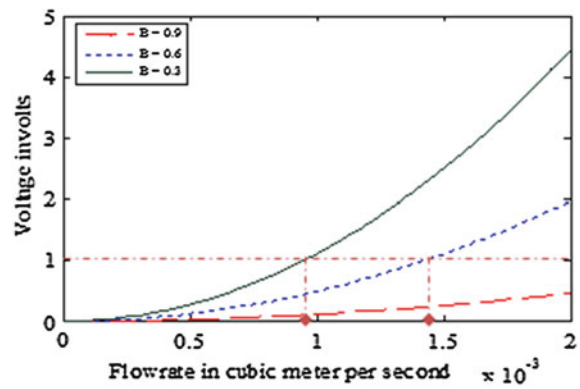
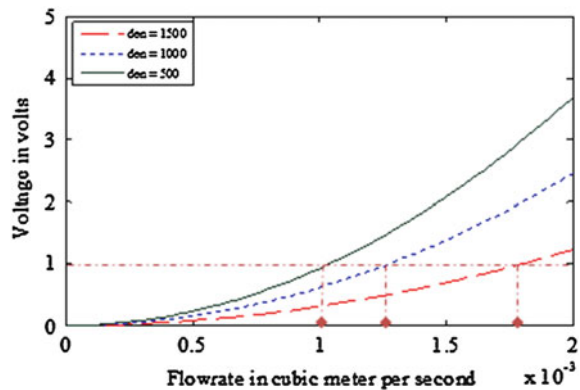
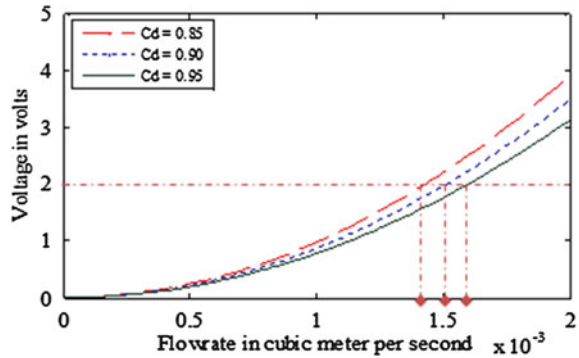


Fig. 5 Input flow versus output voltage of DCU for variation of flow rate and liquid density for diameter ratio of 0.3, discharge coefficient of 0.85 and temperature of 25 °C



Figures 3, 4, 5, and 6 show the nonlinear variation of voltage at the output of DCU with the change in flow rate, discharge coefficient, venturi to pipe diameter ratio, liquid density, and temperature.

Fig. 6 Input flow versus output voltage of DCU for variation of flow rate and discharge coefficient for liquid density of 500 kg/m^3 , diameter ratio of 0.3 and temperature of 25°C



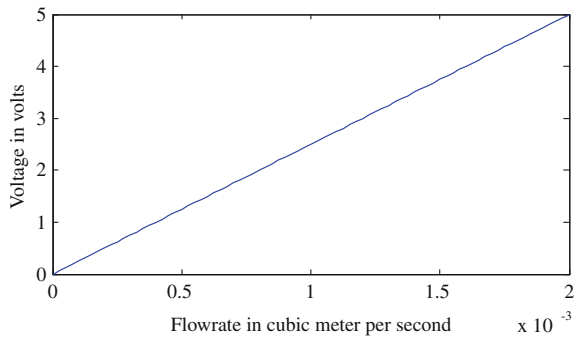
It has been observed from the above graphs that the output from data converter unit is non linear. The datasheet suggests that 10–60 % of full scale is only used as input range considering linearity. The output voltage of DCU also varies with the change in discharge coefficient, venturi to pipe diameter ratio, liquid density, discharge coefficient, and temperature. These are the reasons which have made the user to go for calibration techniques using some circuits. Whenever the required calibration is done for a particular set of values for ρ , β , C_d , and t , it does not work properly/accurately for any change in these parameters without repeated calibration. For example, the output voltage of DCU with variation in flow considering $\beta = 0.3$, $C_d = 0.85$, $t = 25^\circ\text{C}$, and $\rho = 1000 \text{ kg/m}^3$ is shown in Fig. 5. Let, calibration is done for this setup. This calibration will not give correct reading whenever there is variation in any of these parameters. Say, a new liquid having ρ different from 1000 kg/m^3 is used for measurement of its flow without making any change in the calibration circuit. It will read an error as can be seen in Figs. 3, 4, 5, and 6. These conventional techniques have drawbacks that its time consuming and need to be calibrated every time discharge coefficient, venturi to pipe diameter, liquid density, and temperatures are changed in the system. Such measurement is not desirable in industries. So, improvement in the measurement technique by making it adaptive to variations in discharge coefficients, venturi to pipe diameter ratios, liquid densities and temperatures is essential.

A detail of the intelligent design technique is discussed in next section.

5 A Solution Using ANN

In this chapter an attempt is made to design a technique incorporating intelligence to produce full scale linear output and to make the system adaptive of variation in discharge coefficients, venturi to pipe diameter ratios, liquid densities, and temperatures using the concept of an optimized ANN.

The drawbacks discussed in the earlier section are overcome by adding an optimized ANN model in place of conventional data converter unit. This model is designed using the neural network toolbox of MATLAB. Available schemes and algorithms in MATLAB nntool box are used.

Fig. 7 Target graph

The first step in developing a neural network is to create a database. The output voltage of data conversion unit for a particular change in discharge coefficient, venturi to pipe diameter ratio, liquid density and temperature is stored as row of input data matrix. Various such combinations of input flow rate, discharge coefficient, venturi to pipe diameter ratio, liquid density, temperature and their corresponding voltages at the output of data conversion unit are used to form the other rows of input data matrix. The output matrix is the target matrix consisting of data having a linear relation with the flow rate and adaptive of variation in discharge coefficient, venturi to pipe diameter ratio, liquid density, and temperature as shown in Fig. 7.

The process of finding the weights to achieve the desire output is called training. The optimized ANN is found by considering different schemes and algorithms with variation in the number of hidden layer as shown in Table 1. Mean Squared Error (MSE) is the average squared difference between outputs and targets. Lower values of MSE are better. Zero means no error. Regression (R) measure the correlation between output and target. Regression equal to 1 means a close relationship and 0 means a random relationship. The optimized ANN problem is set to find the structure of neural network model having minimum number of hidden layers subject to minimization of performance index as MSE and R and searching among some available schemes and algorithms in MATLAB ANN toolbox.

Five different schemes and algorithms are used to find the optimized ANN. These are Gauss-newton Algorithm (GNA) [19, 20], Levenberg–marquardt algorithm (LMA) [19, 21], Artificial Bee Colony (ABC) [22, 23], Back Propagation neural network (BP) trained by Ant Colony Optimization (ACO) [24–26] and Radial Basis Function (RBF) trained by ACO [25–29]. Training of ANN is first done assuming only one hidden layer. MSE and R values are noted. Hidden layer is increased to two and training is repeated. This process is continued up to four hidden layers. In all cases MSE and R are noted and shown in Table 1. Mesh of MSE's and R's corresponding to different algorithms and hidden layers are shown in Figs. 9 and 10. From Table 2, Figs. 8 and 9 its very clear that RBF trained by ACO yields most accurate results with less number of hidden layers. RBF trained by ACO with 2 layers is considered as the most optimized ANN for desired accuracy of result as shown in Fig. 10. Summary of the optimized ANN model is shown in Table 2.

Table 1 Comparison of number of hidden layers with R and MSE

N	SA	AL1	AL2	AL3	AL4	AL5
H	PM					
L						
1	R	0.625	0.685	0.758	0.876878	0.9124
	MSE	1.01E-1	2.25E-1	8.25E-1	6.87E-2	5.42E-3
2	R	0.785	0.778	0.865	0.93215	0.9926
	MSE	9.84E-2	7.85E-2	5.365E-2	3.76E-3	5.27E-4
3	R	0.886	0.911	0.925	0.99991	0.9994
	MSE	8.68E-3	5.68E-3	1.28E-3	3.12E-6	7.25E-7
4	R	0.9825	0.991	0.9958	0.99999	0.999998
	MSE	1.85E-4	9.98E-5	6.28E-5	7.33E-8	5.258E-9
5	R	0.99925	0.99991	0.99998	0.999993	0.9999999
	MSE	8.86E-6	1.84E-6	5.25E-7	6.32E-9	1.854E-10

NHL Number of hidden layers
SA Scheme and algorithm
PM Performance measure
AL1 Linear design using GNA
AL2 Linear design using LMA
AL3 Linear design Using ABC
AL4 BP using ACO
AL5 RBF using ACO

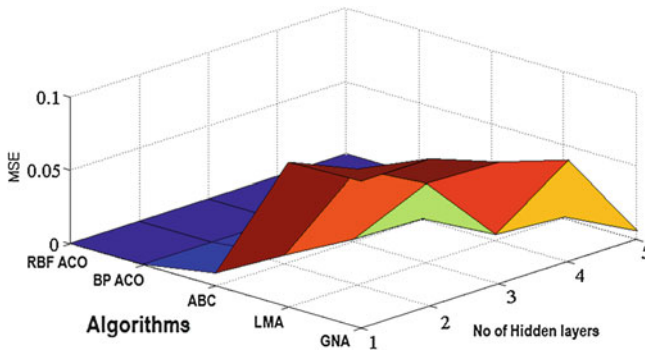


Fig. 8 Mesh showing the MSE corresponding to different ANN models

6 Results and Conclusion

The proposed ANN is trained, validated and tested with the simulated data. It is then subjected to various test inputs corresponding to different discharge coefficients, venturi to pipe diameter, liquid, density, discharge coefficient, and temperatures of liquid, all within the specified ranges. For testing purposes, the range of flow rate is considered from 0.0 to 2.0 m³/s, range of β is 0.3–0.9, range of liquid density is 500–1500 kg/m³, range of discharge coefficient is 0.85–0.95, and range of temperature is 50–150 °C. The outputs corresponding to sampled test

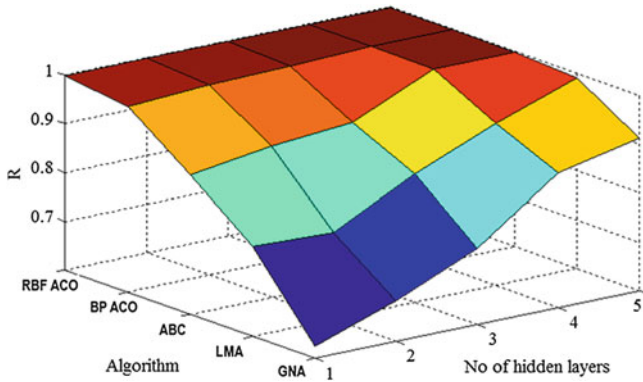


Fig. 9 Mesh showing the R corresponding to different ANN models

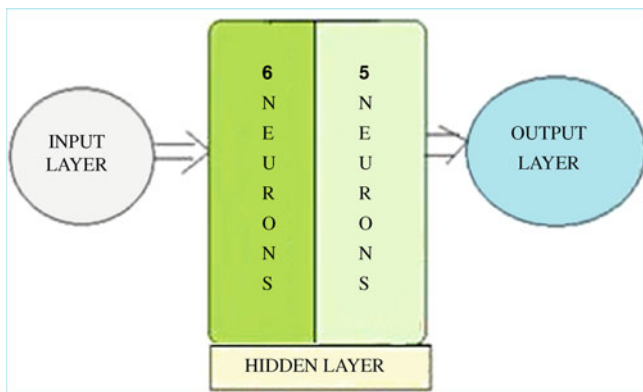


Fig. 10 Structure of optimized neural network model

inputs are tabulated in Table 3. The input output result is plotted and is shown in Fig. 11. The output graph matches with the target graph which is shown in Fig. 8.

It is evident from Fig. 11 and Table 3 that the proposed flow measurement technique has gained intelligence in addition to increase in the linearity range. The output is made adaptive of variation in discharge coefficient, venturi to pipe ratio, liquid density, discharge coefficient, and liquid temperatures. Thus, if the liquid is changed by another liquid having different density and/or if the discharge coefficient is changed and/or if the venturi to pipe diameter ratio is changed and/or if the liquid temperature is changed, then also calibration process need not be repeated. All these have been achieved by using an optimized ANN. Five different schemes and algorithms are used for training, validation and testing. RBF using ACO is found to have achieved the desired MSE and Regression close to 1 with only two hidden layers. Hence, RBF using ACO is used as an optimized ANN.

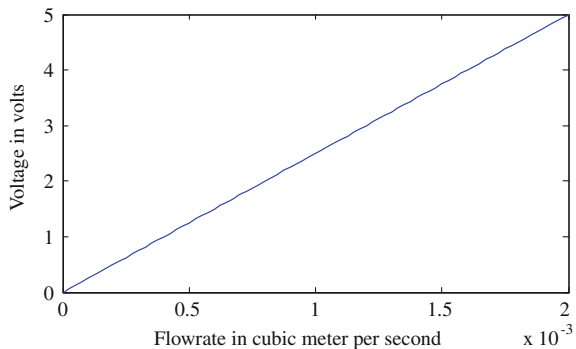
Table 2 Summary the optimized network model

Parameters of the optimized neural networks model						
Database	Training base		84			
	Validation base		18			
	Test base		18			
No of neurons in	1st layer		8			
	2nd layer		6			
Transfer function of	1st layer		Tansig			
	2nd layer		Tansig			
	Output layer		Linear			
Input	Flow (m ³ /s)	β	C_d	Den (kg/m ³)	Temp (°C)	
	Min	0	0.3	0.85	500	50
	Max	0.002	0.9	0.95	1500	150

Table 3 Result of measurement technique using optimized ANN

Actual flow rate in m ³ /s	β	ρ in Kg/m ³	C_d	t in °C	Data conversion o/p in V	ANN o/p in V	Measured flow rate in m ³ /s
0.0005	0.30	500	0.85	125	0.1017	1.25	0.0005
0.0005	0.35	550	0.92	50	0.0716	1.25	0.0005
0.0005	0.40	600	0.90	65	0.0973	1.25	0.0005
0.0010	0.45	650	0.91	75	0.4643	2.5	0.0010
0.0010	0.50	700	0.87	62	0.6097	2.5	0.0010
0.0010	0.55	750	0.95	90	0.5072	2.5	0.0010
0.0015	0.60	800	0.94	100	1.4063	3.75	0.0015
0.0015	0.65	850	0.89	120	2.2876	3.75	0.0015
0.0015	0.70	900	0.90	105	2.5167	3.75	0.0015
0.0020	0.75	950	0.91	130	4.6911	5.0	0.0020
0.0020	0.80	1000	0.95	85	4.5268	5.0	0.0020
0.0020	0.85	1050	0.88	90	4.8951	5.0	0.0020

Fig. 11 Response of proposed system



Available reported works have discussed different techniques for calibration of flow measurement, but these are not adaptive of variations in venturi to pipe diameter ratios, discharge coefficients, liquid densities, discharge coefficient,

and temperatures of liquid. Hence, repeated calibration is required for any change of these parameters. Some time, the calibration circuit may itself be replaced which is a time consuming and tedious procedure. Further, some reported works have not utilized the full scale of measurement. In comparison to these, the proposed intelligent flow measurement technique achieves linear input output characteristics for full input range and makes the output adaptive of variations in venturi to pipe diameter ratio, discharge coefficient, liquid density, discharge coefficient, and temperature. So, proposed technique avoids repeated calibration whenever there is any change in these parameters.

References

1. Lawley JE (1982) Orifice meter calibration'. In: 57th international school of hydrocarbon measurement, document ID: 16233
2. Yanagihara S, Mochizukia O, Satoa K, Saitob K (1999) Variable area venturi type exhaust gas flow meter'. *JSAE Rev, Science direct* 20(2):265–267
3. Zhang JQ, Yan Y (2001) A self validating differential pressure flow sensor'. In: Proceedings of IEEE conference on instrumentation and measurement technology, Budapest
4. Na MG, Lee YJ, Hwang IJ (2005) A smart software sensor for feedwater flow measurement monitoring. *IEEE Trans Nucl Sci* 52(6):3026–3034
5. Xu L, Zhou W, Li X, Wang M (2011) Wet-gas flow modeling for the straight section of throat-extended venturi meter. *IEEE Trans Instrum Meas* 60(6):2080–2087
6. Xu L, Li H, Tang S, Tan C, Hu B (2008) Wet gas metering using a venturi-meter and neural networks. In: Proceedings of IEEE conference on instrumentation and measurement technology, Canada
7. Pereira JMD, Postolache O, Girao PMBS (2009) PDF-based progressive polynomial calibration method for smart sensors linearization. *IEEE Trans Instrum Meas* 58(9):3245–3252
8. Yang HY, Lee SH, Na MG (2009) monitoring and uncertainty analysis of feedwater flow rate using data-based modeling method. *IEEE Trans Nucl Sci* 56(4):2426–2433
9. Fang L, Cao S, Li J, Cheng H (2009) Venturi wet gas measurement based on homogenous and Chisholm model. In: Proceedings of international conference on computer and automation engineering, Thailand
10. Xu L, Zhou W, Li X, Tang S (2011) Wet gas metering using a revised venturi meter and soft-computing approximation techniques. *IEEE Trans Instrum Meas* 60(3):946–2087
11. Pereira M (2009) Flow meters: Part 1. *IEEE Mag Instrum Meas* 12(1):18–26
12. Xu L, Tang S (2009) Wet gas metering using a venturi-meter and support vector machines. In: Proceedings of IEEE conference on instrumentation and measurement technology, Singapore
13. Santhosh KV, Roy BK (2012) An intelligent flow measuring technique using venturi. Lecture notes in engineering and computer science: proceeding of international multiconference of engineers and computer scientists, IMECS 2012, 14–16 March 2012, Hong Kong, pp 902–907
14. Doebelin EO (2003) Measurement systems—application and design, 5th edn. Tata McGraw Hill publishing company, New York
15. Liptak BG (2003) Instrument engineers' handbook: process measurement and analysis, 4th edn. CRC Press, Boca Raton
16. Plaza RJ (2006) Sink or swim: the effects of temperature on liquid density and buoyancy'. California state science fair
17. Density of fluids—changing pressure and temperature, the engineering toolbox (2004)
18. Sensorone DMD331 datasheets 2008

19. Björck A (1996) Numerical methods for least squares problems. SIAM Publications, Philadelphia, ISBN 0-89871-360-9
20. Roger F (1987) Practical methods of optimization, 2nd ed. Wiley, New York, ISBN 978-0-471-91547-8
21. Dias FM, Antunes A, Vieira J, Mota AM (2004) Implementing the levenberg-marquardt algorithm on-line: a sliding window approach with early stopping. In: International conference on proceedings of IFAC, USA
22. Karaboga D (2005) An idea based on honey bee swarm for numerical optimization. Technical report-tr06, Erciyes University, engineering faculty, computer engineering department
23. Venkata Rao R (2006) Multi-objective optimization of multi-pass milling process parameters using artificial bee colony algorithm. Artificial intelligence in manufacturing. Nova Science Publishers, USA
24. Li J-B, Chung Y-K (2005) A novel back propagation neural network training algorithm designed by an ant colony optimization. IEEE/PES transmission and distribution conference & exhibition: Asia and Pacific Dalian, China
25. Bianchi L, Gambardella LM, Dorigo M (2002) An ant colony optimization approach to the probabilistic travelling salesman problem. In: Proceedings of PPSN-VII, seventh inter 17 national conference on parallel problem solving from nature. Springer, Berlin
26. Russell S, Norvig P (2009) Artificial intelligence a modern approach, 3rd edn. Prentice Hall, New York
27. Park J, Sandberg JW (1991) Universal approach using radial basis function network. Neural Comput 3:246–257
28. Poggio T, Girosi F (1990) Networks for approximation and learning. Proc IEEE 78(9): 1484–1487
29. Paul Yee V, Haykin S (2001) Regularized radial basis function networks: theory and applications. Wiley, New York

Efficient Video Streaming Over Wireless Mesh Networks

Kwok Tung Sze, King Man Ho and Kwok Tung Lo

Abstract In this paper, a video-on-demand (VoD) system based on Home-to-Home Online (H2O) for wireless mesh network (WMN) is proposed and studied. Unlike H2O which focuses on the minimization of system storage, we also take the bandwidth usage into account in order to improve system performance. The VoD system mainly consists of three components: a video block replication scheme, a video block discovery scheme, and several enhancements on video block transmission. The enhancements are named Video Block Broadcasting, Peer-to-Peer, and Time-Shifting. A simulation model is built to evaluate the performance of the proposed VoD system against each enhancement in terms of blocking probability. The results of simulations show the potential and limitations of each enhancement.

Keywords Peer-to-peer (P2P) · Video-on-demand (VoD) · Wireless mesh network (WMN)

1 Introduction

Wireless mesh networks (WMNs) [1] is emerging as a key technology for next generation wireless networks. A typical WMN composes of mesh routers and mesh clients. The mesh routers form the backbone of the network that provides the

K. T. Sze · K. M. Ho · K. T. Lo (✉)
Department of Electronic and Information Engineering,
The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
e-mail: enktlo@inet.polyu.edu.hk

K. T. Sze
e-mail: Kwok.Tung.SZE@connect.polyu.hk

K. M. Ho
e-mail: enkmho@hotmail.com

network access to the mesh clients, and have minimum mobility. The mesh clients are mobile devices such as laptop and PDA, which connects to the mesh routers for network access. Currently, WMNs are undergoing rapid development progress and inspiring numerous applications such as video-on-demand (VoD) service. VoD over WMN means that the subscribers can playback any favorite video anytime once their mobile device connects to the WMN. With the increasing popularity of mobile devices, it can be expected that providing VoD service over WMN shall have a bright future.

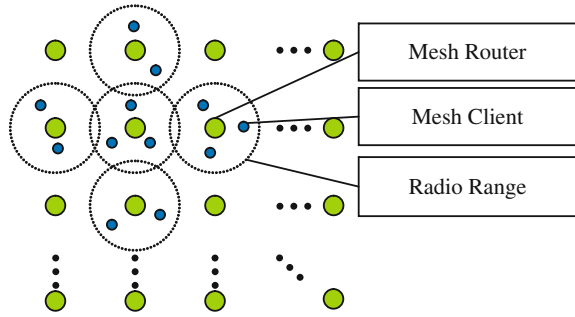
In order to provide quality VoD services over WMN, many research works have done. For example, several VoD schemes based on caching and cooperative clients are proposed in [2]. A work on supporting VoD over WiMax mesh network is presented in [3]. In [4], it introduces a video placement scheme called Home-to-Home Online (H2O) which distributes video data among the H2O devices. This scheme aims to save the storage of a distributed VoD system, and can also apply to the VoD service over WMN. Ding et al. [5] modifies H2O by taking the behaviors of the users into account and develops a new replication scheme called Home-to-Home Video Clip Replication (H2-VIP). In [6], challenge on selecting the best path is addressed. Several enhancements for Ad Hoc On-Demand Distance Vector (AODV) routing protocol are proposed to improve the video quality over WMN. Performance of the peer-to-peer video applications against user mobility is investigated in [7]. The experiment results indicate that the applications using frequent short-lived connection for video transmission has advantage in WMN. A distributed multicast technique, Dynamic Stream Merging (DSM), is proposed in [8]. DSM creates the multicast topologies by dynamically merging the server streams among mesh nodes that no video server is involved.

In this paper, we propose a VoD system [9, 10] over WMN based on H2O. The system mainly consists of three components: a video block replication scheme, a video block discovery scheme, and several enhancements on video block transmission. The replication scheme is based on [4] that equally divides a video into blocks and evenly distributes the copies of those blocks to the routers in the network. The required copies of each video block decreases with its timing order, i.e., fewer copies are needed for video blocks in later part of the video. Each router only stores a portion of video. The discovery scheme is to search the video blocks from the network.

Although the replication scheme can save system storage, the bottleneck is resulted in the mesh routers holding the sparse video blocks. To tackle this issue, we consider different transmission strategies and introduce three possible enhancements named Video Block Broadcasting, Peer-to-Peer, and Time-Shifting. These enhancements aim to better utilize the routers' bandwidth so that more clients can be served by the routers concurrently. A simulation model is built to evaluate the performance of system against each enhancement in terms of blocking probability.

The main focus of this paper is to study how the proposed enhancements can improve the performance of the proposed VoD system in WMN. The rest of this chapter is organized as follows: Sect. 2 describes the basic elements of the

Fig. 1 Topology used in simulation



proposed VoD system. Section 3 introduces the enhancements. Section 4 presents the simulation results that evaluate the performance of system against each enhancement in terms of blocking probability. Section 5 is the conclusion.

2 Proposed System Architecture

In this section, we first describe the network topology used in the proposed system. Then, we show how the video blocks are distributed to the mesh routers, and how a mesh router gathers the missing video blocks from the network.

2.1 Network Topology

The topology of the proposed framework is shown in Fig. 1. It consists of N mesh routers and some mesh clients. The mesh routers are evenly distributed in a square area and the location of each router is fixed during the simulation. Each mesh router can only communicate with the neighboring routers in cardinal directions. It means a mesh router normally has four neighboring routers, i.e., at the top, bottom, left and right, except those routers on the edge of the topology. Moreover, every router holds a portion of video data in its storage. The mesh clients are mobile devices which connect to its nearest router and request for the VoD service.

2.2 Video Block Replication Scheme

Instead of deploying dedicated VoD servers in a mesh network, the video data is distributed to the storage of each mesh router. Moreover, in order to reduce the storage required, the idea of H2O replication scheme [4] is adopted. Briefly, suppose a video has a length of L seconds. The video is first divided into X blocks equally as

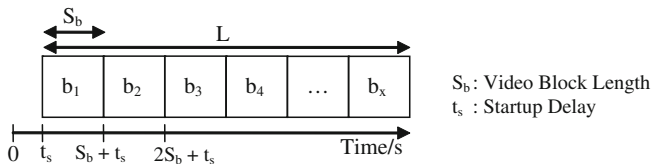


Fig. 2 Video block placement

shown in Fig. 2. The first video block, b_1 , is copied to every router to reduce the startup delay, t_s . The remaining blocks are distributed to the routers which ensure each router can acquire the n th video block before the time $(n-1)S_b + t_s$ if a mesh client requests a VoD service at time 0.

In the proposed system, the n th video block is copied to routers every $n-1$ hop(s). For example, the first video block will be copied to every mesh router; the second video block will be copied to routers every 1 hop and so on. As a result, fewer copies are needed for a video block in higher order.

The total number of mesh routers within $n-1$ hop(s) in the grid topology is equal to $2(n-1)^2 + 2(n-1) + 1$ as illustrated in Fig. 3. For example, there are five mesh routers, which include the source mesh router, in the distance of 1 hop. Therefore, the minimum number of replicas of n th video block is equal to $r_n = \left\lceil \frac{N}{2(n-1)^2 + 2(n-1) + 1} \right\rceil$, where N is the total number of mesh routers in the grid topology. Then, the minimum storage required of whole VoD system is equal to $\sum_{n=1}^X r_n$, where X is total number of video blocks.

2.3 Video Block Discovery Mechanism

Suppose a video is equally divided into blocks; these blocks are properly scattered over the storage of mesh routers. When a mesh client connects to its nearest mesh router, the client sends a request for the video to the router. Once the router receives the request, the router checks its local storage, generates request messages for those missing video blocks, and sends those request messages to its neighbors. These request messages should at least contain five elements which are Sequence Number (SEQ), Video Block ID, Video Block Arrival Time, Time to Live (TTL), and Timeout. The TTL limits the number of hops that a request message can be forwarded. The SEQ prevents request message from looping in the network.

When a neighboring router receives a new request message, the router first reserves a channel. If there is not channel available, the router discards the request message. Otherwise, the router checks the existence of the requesting video block in its local storage. If the requesting video block is not found, the router forwards the request message to all of its neighbors except the sender of the request message. If it is found, a response message will be returned along the requesting path.

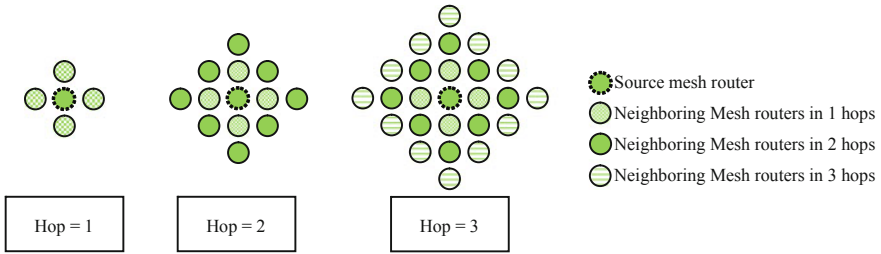


Fig. 3 Total number of mesh router within N hops

After all, when the requesting mesh router receives the response messages, it sends an acknowledgement message to one of the sources mesh routers which has the fewest hop counts. The requesting stage is completed when the source mesh router receives the acknowledgement message. Then, the transmission is started according to the scheduled video block arrival time.

Besides, all the corresponding reserved channels will be released if a request message is timeout and does not have any acknowledgement message. In the current setting, a mesh client terminates the VoD service if any of the requesting video blocks is not found in the requesting stage. As a result, all the reserved channels for this client are released also.

Figure 4 shows the request mechanism of a video block as an example. At first, a client connects to its nearest mesh router and sends a request for the video to the connected router. Then, the router searches its local storage and sends the requests of all the missing video blocks to neighboring routers. Suppose one of the missing video blocks is distributed to the routers every 3 hops. The requesting mesh router sends a request, which TTL is equal to 3, for this video block to all neighboring routers. This request message keeps flowing on the mesh network until there is not available channel of current intermediate router, or the TTL is reached, or the request message reaches the sources mesh router. After that, if the request message is successfully forwarded to the sources mesh router, it sends a response message back to the requesting mesh router along the path. Finally, the requesting mesh router sends the acknowledgement to one of the sources mesh router to confirm the transmission time.

As mentioned in Sect. 2.2, the number of copies of video block in the network is decreasing with the order of video block, i.e., video block in higher order has fewer numbers of copies in the network. Therefore, as shown in the example above, it can be expected that the degree of difficulty on accessing a video block shall be generally proportional to its order. As a result, the routers which hold video block in high order shall become the bottleneck of the system. To tackle this issue, we consider different transmission strategies in order to better utilize the router’s bandwidth and improve the accessibility of video blocks.

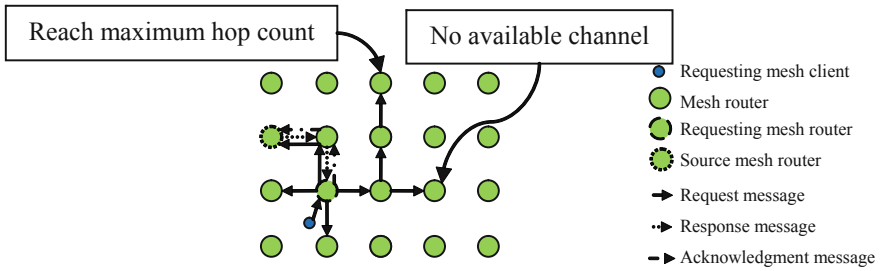


Fig. 4 Block discovery mechanism

3 System Enhancements

In order to improve the services performance, three enhancements are proposed and evaluated separately. In the following, video block broadcasting based on broadcast transmission is first described. Then, Peer-to-Peer and Time-Shifting based on unicast transmission are presented.

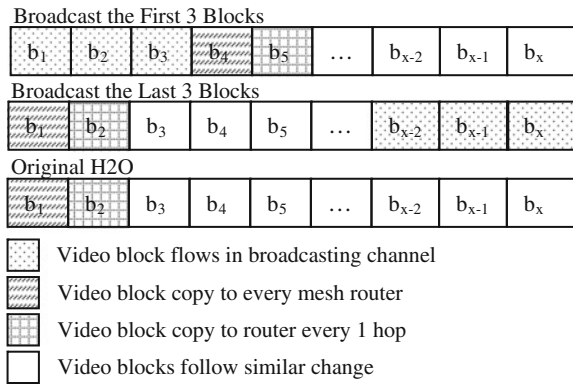
3.1 Video Block Broadcasting

Video Block Broadcasting assumes each mesh router reserves channel(s) to forward video block(s) from source mesh router to all other mesh routers actively and continually. For example, suppose the first three video blocks are selected to broadcast. At first, each of these three video blocks is placed to a mesh router separately, and three channels are reserved in all mesh routers. After that, source mesh routers broadcast the video block to their neighbors. Once the neighbors receive the data, they will also broadcast to their neighbors, and so on. It makes these three video blocks available at anytime anywhere by sacrificing three free channels.

Moreover, as illustrated in Fig. 5, the video block placement is modified accordingly when the Video Block Broadcasting is used. For example, suppose a video is divided into 15 blocks. When the first three video blocks are selected to broadcast, video block b_4 is copied to every router, video block b_5 is copied to the router every 1 hop, and so on. The distribution of the rarest video block, b_{15} , is changed from every fourteen hops to eleven hops. No additional channel is needed when the client is downloading the first three video blocks. Channels are only reserved to transmit the video blocks started from b_4 to b_{15} . When the last three video blocks are selected to broadcast, the rarest video block is changed from b_{15} to b_{12} which is also distributed every eleven hops. However, additional channels, which transmit the beginning video blocks, are needed immediately once the client requests for the VoD service.

In current design, the number of channels reserved in each mesh router is equal to the number of broadcasting video blocks. It is also assumed that the

Fig. 5 Changes in video block placement



total number of channels in each mesh router is much smaller than total number of video blocks. Therefore, the system can only broadcast a portion of video blocks.

3.2 Peer-to-Peer

Peer-to-Peer assumes that each connected mesh client stores a number of latest received video blocks in its buffer and acts as a source mesh client that uploads video block based on request. With this enhancement, the router first searches the connected clients' buffer before flooding the request of the missing video block to neighboring routers. For example, as shown in Fig. 6, when an intermediate mesh router receives a request message from neighboring router, the intermediate router checks its storage for the requesting video block as usual. Assume it cannot be found, the intermediate router then broadcasts the request message to all connected mesh clients. If the client(s) possesses the requesting video block, an acknowledgement message is sent back to the intermediate router. Then, the intermediate router returns a response message along the requesting path. However, if no client possesses the requesting video block, the intermediate router forwards the request message to neighboring routers as described in Sect. 2.3. And finally, the requesting mesh router send an acknowledgment message to the sources mesh router, or client, if response message is received.

In current design, it is assumed that the source mesh client will stay in the VoD system before the end of uploading.

3.3 Time-Shifting

In Time-Shifting, requests which arrive closely are grouped together and delivered at the same time.

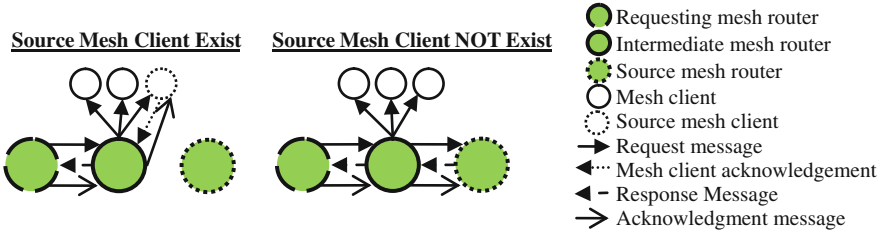


Fig. 6 Video delivery mechanism with peer-to-peer

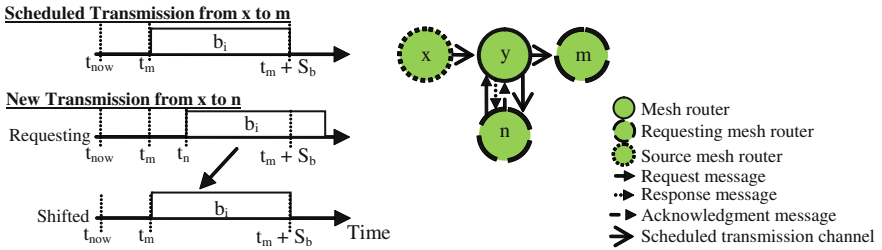


Fig. 7 Video delivery mechanism with time-shifting

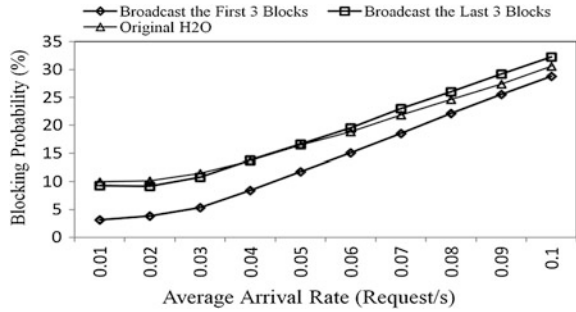
As illustrated in Fig. 7, suppose there is a scheduled transmission for video block b_i from router x to router m starting from time t_m to $t_m + S_b$. As an intermediate router y , it stores this information in a table named shifting table. When router y accepts another request from router n , router y first searches the records in the shifting table. If the requesting video block is also b_i and the requesting transmission time, t_n , falls between the scheduled transmission time t_m and $t_m + S_b$, router y returns a response message with updated transmission time which is shifted from t_n to t_m .

However, if the requesting video block is not found in the shifting table or the requesting transmission time does not match, router y acts according to the description in Sect. 2.3. It means router y then search its local storage. If the requesting video block is found, a response message is returned along the path. If the video block is not found, router y forwards the request to its neighboring routers, and so on.

4 Simulation Result

In order to evaluate the proposed VoD system and enhancements, a simulation model is built based on C programming language. It assumes that there are N mesh routers formed a grid topology. Each mesh router can only communicate with the neighboring routers in cardinal direction, and has a number of channels which

Fig. 8 Blocking probability against varies arrival rates



transmission rate is equal to the playback rate of the video as described in Sect. 2.1. The video is divided into blocks equally and distributed among the storage of mesh routers according to the method described in Sects. 2.2 and 3.1. Each client asks for the video after it connects to its nearest mesh router. The request pattern is modeled as the Poisson Process. The mechanism of video block discovery follows the description in Sects. 2.3, 3.2 and 3.3. Connected clients watch the video if it is confirmed in requesting stage that all the video blocks are available in the desired transmission time. Otherwise, clients quit the VoD service. We call these clients are blocked.

4.1 Video Block Broadcasting

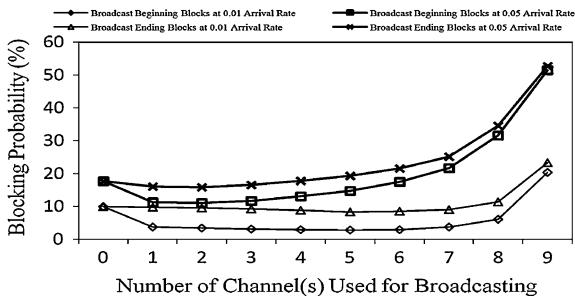
In this set of simulations, there are 400 mesh routers in total. Each mesh router equips 10 channels for transmission. The length of video is 2 h. The video is divided into 15 blocks equally and distributed according to the method described in Sects. 2.2 and 3.1.

It is expected that the system performance will be improved by adopting the Video Block Broadcasting since it improves the availability of the rare video blocks by sacrificing free channels. However, the improvement should be decreased with the increase of arrival rate. It is because the bottleneck shall be shifted from the routers which hold the rare video blocks to the total number of free channels available in the VoD system.

Figure 8 shows the blocking probability against varies average arrival rates in three different scenarios which are broadcasting the first three video blocks, broadcasting the last three video blocks, and the original H2O without broadcasting. Result suggests that

1. Broadcasting the first three video blocks has lower blocking probability than broadcasting the last three video blocks.
2. The use of Video Block Broadcasting does give lower blocking probability but the improvement does decrease with the increase of arrival rate and finally deteriorate the system performance.

Fig. 9 Blocking probability against varies number of channels used for broadcasting in two different arrival rates



In order to study the point 1 further, another simulation is carried out and the result is in Fig. 9. It shows the blocking probability against varies number of broadcasting channels in two different arrival rates. It should be noted that the number of broadcasting video blocks is equal to the number of channels used for broadcasting, and the total number of channels in each mesh router remains 10.

As shown in Fig. 9, broadcasting beginning blocks always give lower blocking probability than the one broadcasting ending blocks under the same arrival rate and number of broadcasting channel. The result suggests that broadcasting beginning blocks seems better than the one broadcasting ending blocks in current VoD system.

Figure 10 is the result of another simulation which is to study point 2. Figure 10 shows the blocking probability against varies number of channels equipped in each mesh router. In this simulation, half of channels in each mesh router are reserved for broadcasting and the rest are used as usual.

Figure 10 shows that the Video Block Broadcasting can improve the blocking probability when mesh routers have greater number of channels. For example, an improvement in blocking probability is observed when total number of transmission channels reaches 8 at 0.01 arrival rates and 12 at 0.05 arrival rates.

In conclusion, in order to obtain better system performance with Video Block Broadcasting, the following conditions are required:

1. It is broadcasting the video blocks in lower order.
2. There are a substantial number of channels equipped in mesh routers.

4.2 Peer-to-Peer and Time-Shifting

In this set of simulations, there are 400 mesh routers in total. The length of video is 2 h. The video is divided into 30 blocks equally and distributed according to the method described in Sect. 2.2.

Figure 11 shows the blocking probability against varies arrival rates. In general, the blocking probability increases with the increase of arrival rate but decreases with the increase of number of channels. This result will be used to compare the simulation results of Peer-to-Peer and Time-Shifting in the following section.

Fig. 10 Blocking probability against varies video channels equipped in mesh router in two different arrival rates

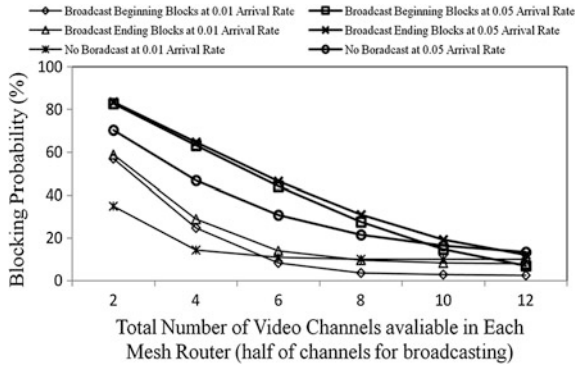
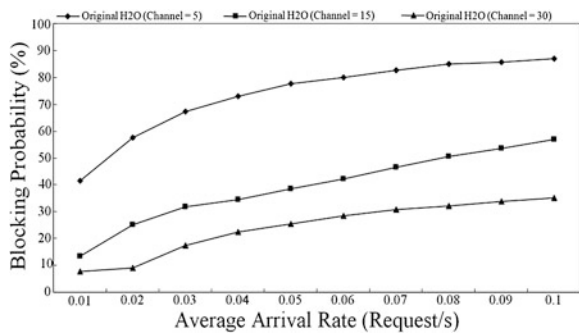


Fig. 11 Blocking probability against varies arrival rates



4.2.1 Peer-to-Peer

It is expected that the use of Peer-to-Peer should improve the system performance and the extent of improvement shall be proportional to the number of video blocks stored in mesh clients.

Figure 12 shows that the blocking probability against arrival rates in various buffer sizes. Compare with Fig. 11,

1. The improvement in blocking probability is obvious when the size of buffer is greater than one.
2. The extent of improvement in blocking probability decreases with the increase of buffer size.

Figure 13 shows blocking probability against various buffer sizes in different number of channels and arrival rates. This result further indicates that the improvement on blocking probability is negligible when the buffer size increases beyond certain limit. For example, when the number of channels and arrival rate is equal to 5 and 0.01 respectively, the improvement on blocking probability is negligible if the buffer size is greater than 4.

Fig. 12 Blocking probability against arrival rates in various buffer size of the mesh client

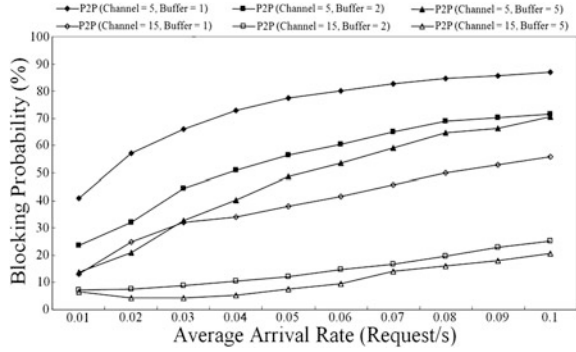


Fig. 13 Blocking probability against buffer size in various arrival rates

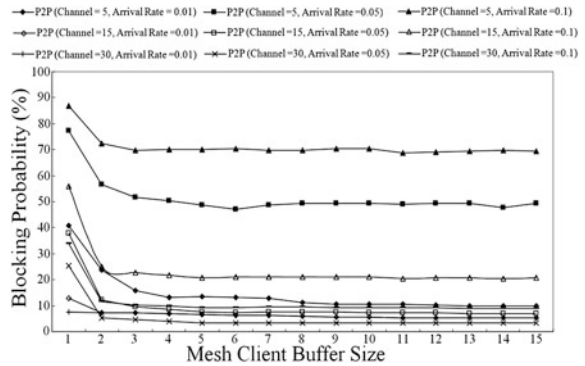
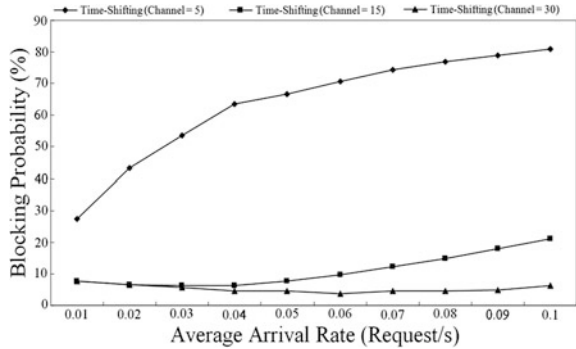


Fig. 14 Blocking probability against arrival rate



4.2.2 Time-Shifting

It is expected the use of Time-Shifting should improve the system performance as it effectively reduce the loading of upstream router by putting requests of same video block in a group and delivering at the same time.

Figure 14 shows the blocking probability against arrival rate in different number of channels. Compare with Fig. 11, obvious improvement is observed. The improvement increases with the number of channels in mesh router and arrival rate.

5 Conclusion and Future Work

In this paper, a VoD system with various enhancements over WMN is proposed and studied. The major components of the VoD system are a video block replication scheme, a video block discovery scheme, and several enhancements. The video block replication scheme adopts the idea from [4] in order to reduce the storage required. Briefly, this scheme divides a video into blocks equally, and distributes the n th block to every router $n-1$ hop(s); the necessary copies of each video block gradually decreases with its timing order, and all the mesh routers only hold a portion of video. The video block discovery scheme is to acquire all the missing video blocks from the mesh network.

To tackle the bottleneck introduced by H2O, we consider different transmission strategies and propose three enhancements, which are named Video Block Broadcasting, Peer-to-Peer, and Time-Shifting. According to the results of simulations, Peer-to-Peer and Time-Shifting are suitable for all situations while Video Block Broadcasting can only apply to the VoD System which already has substantial number of free channels.

Currently, the simulation model is being migrated to ns-3 [11], which is a discrete-event network simulator for research and education, in Ubuntu 10.04 LTS [12], which is one of the popular Linux distributions. It is expected the use of ns-3 can reduce the effort on coding especially debugging and have greater degree of flexibility as well as retain fast execution speed.

References

1. Akyildiz IF, Wang X, Wang W (2005) Wireless mesh networks: a survey. *Comput Netw* 47(4):445–487
2. Zhu Y, Zeng W, Liu H, Guo Y, Mathur S (2008) Supporting video streaming services in infrastructure wireless mesh networks: architecture and protocols. In: *IEEE international conference on communications 2008, ICC '08, 19–23 May, Beijing*, pp 1850–1855
3. Xie F, Hua KA, Jiang N (2007) Achieving true video-on-demand service in multi-hop wimax mesh networks. In: *32nd IEEE conference on local computer networks 2007, LCN 2007, 15–18 October, Dublin*, pp 287–294
4. Ghandeharizadeh S, Krishnamachari B, Song S (2004) Placement of continuous media in wireless peer-to-peer networks. *IEEE Trans Multimedia* 6(2):335–342
5. Ding JW, Wang WT, Wang CF (2008) An efficient data replication scheme for peer-to-peer video streaming over wireless-mesh community networks. In: *International conference on intelligent information hiding and multimedia signal processing 2008, IHHMSP '08, 15–17 August, Harbin*, pp 767–770

6. Zhu Y, Liu H, Wu M, Li D, Mathur S (2007) Implementation experience of a prototype for video streaming over wireless mesh networks. In: 4th IEEE consumer communications and networking conference 2007, CCNC 2007, Las Vegas, pp 23–28
7. Moraes IM, Campista MEM, Costa LHMK, Duarte OCMB, Duarte JL, Passos DG, de Albuquerque CVN, Rubinstein MG (2008) On the impact of user mobility on peer-to-peer video streaming. *IEEE Wirel Commun* 15(6):54–62
8. Hua KA, Xie F (2010) A dynamic stream merging technique for video-on-demand services over wireless mesh access networks. In: 7th annual IEEE communications society conference on sensor mesh and ad hoc communications and networks 2010, SECON 2010, 21–25 June, Boston, pp 1–9
9. Sze KT, Ho KM, Lo KT (2011) Supporting video-on-demand services over wireless mesh network. In: International conference on computers, communications, control and automation 2011, CCCA 2011, 20–21 February, Hong Kong, pp 357–360
10. Sze KT, Ho KM, Lo KT (2012) Performance evaluation of a video broadcasting system over wireless mesh network. *Lecture notes in engineering and computer science*. In: Proceedings of the international multiconference of engineers and computer scientists 2012, IMECS 2012, 14–16 March, Hong Kong, pp 386–390
11. ns-3 Website [Online] <http://www.nsnam.org/>
12. Ubuntu Website [Online] <http://www.ubuntu.com/>

Subject Index

Page numbers followed by ‘*f*’ indicate figures and ‘*t*’ indicate tables respectively.

A

- AA. *See* Aluminium Alloy
- AAT. *See* Available arrival time
- ACO. *See* Ant Colony Optimization
- Actual demand and order, 162*f*
- Ad Hoc On-Demand Distance Vector routing protocol (AODV routing protocol), 354
- ADAM. *See* Autonomous Delivery and Manipulation
- Advancing side (AS), 265
- Agri-food supply chain network (ASC network), 190, 192
 - constraints, 195
 - CSR effect
 - budget allocation, 199, 200
 - on farmers profit, 197, 199, 199*f*
 - on MR’s risk, 199, 200, 200*f*
 - on profit and MR revenue, 198, 199, 199*f*
 - CSR programs, 190
 - farmer management skill data, 196, 197, 197*t*
 - objective functions, 195
 - problem formulation, 191
 - deterioration time, 194, 194*f*
 - FGC, 191, 193
 - model formulation, 191–193
 - vegetables quantity sold to MR, 194, 194*f*
 - quantity and selling price, as quality function, 197, 197*t*
 - of relevant system, 192*f*
 - total CSR cost, 196
 - vegetable shelf life and risk function, 197, 198, 198*t*
- Aircraft batches, 71
 - sequence change and adjusted time of arrival, 73, 73*f*
 - two heavy and one medium aircraft, 72, 72*f*
 - two medium and one heavy aircraft, 73, 73*f*
- Aircraft combination, in longer batches, 74
 - two heavy and one medium aircraft, 74, 75
 - two medium and one heavy aircraft, 75, 76
- Aluminium Alloy (AA), 265
- Ameliorating items, 138
- Ameliorating items, Weibull, 138–140
 - poultry farmer’s optimal policy, 143–147
 - poultry farmer’s total profit, 142–143
 - retailer’s optimal response, 141–142
 - retailer’s total profit, 140–141
- Analysis model, 281–282
- Analysis of Variance (ANOVA), 264, 268, 273
- Anchorbox, 333, 334, 335, 337–338
 - example, 334*f*
- Angle of Arrival (AoA), 327
- ANOVA. *See* Analysis of Variance
- Ant Colony Optimization (ACO), 169
- Anti-commons tragedy, 125
- AoA. *See* Angle of Arrival
- AODV routing protocol. *See* Ad Hoc On-Demand Distance Vector routing protocol
- APT. *See* Arbitrage Pricing Theory
- Arbitrage Pricing Theory (APT), 92
- Arrival aircraft sequencing
 - aircraft fuel burn, 66
 - problem formulation, 70
 - arrival time, 70, 70*f*
 - Boeing 737, 69
 - Boeing 747, 69

A (*cont.*)

- first served sequence, 72, 72f
- minimum aircraft separation, 69
- MTOW, 69
- wake turbulence, 69
- wake vortex, 69
- terminal area and single aircraft trajectory optimization
 - assumptions, 67f
 - coordinates and waypoint altitudes, 67
 - fuel burn vs. descent time, 68, 68f
 - ICAO separation minima, 69, 69t
 - optimal descent trajectories, 67
 - quadratic function, 68
 - SQP method and ETA, 68
 - waypoint coordinates, 68, 68t
- terminal area assumptions, 67
- AS. *See* Advancing side
- ASC network. *See* Agri-food supply chain network
- Auditory reaction, 302
- Auditory reaction test, 303, 305–306
- Auditory stimuli test, 291
- Autonomous Delivery and Manipulation (ADAM), 326
- Autonomous transport vehicles
 - ADAM, 326
 - connectivity information, 327
 - CSS, 327–328
 - global localization using MCPF, 336
 - anchorbox, 337–338
 - experimental results, 337, 337f
 - laser range finders, 336
 - transport vehicle, 336, 337
 - global localization using MCPF
 - anchorbox, 333–335
 - IEEE 802.15.4a measurement error, 331
 - IEEE 802.15.4a measurement model, 332–333
 - KF and EKF, 330
 - measurement update, 331
 - motion model, 330, 331
 - odometry data, 335
 - prediction phase and update phase, 330
 - hardware
 - with CANopen protocol, 335
 - IMU, 335
 - vehicle's PLC, 335
 - wireless sensor node, 335f
 - wireless synchronisation, 336f
 - IEEE 802.15.4a wireless network, 326f
 - location tracking, using EKF, 328–330
 - MSM technology, 326

- position tracking using EKF, 336
- techniques, 327
- ToA and RToF, 327
- WSN, 326–327
- Availability, 235
 - flow chart for, 240
 - rating based on, 241
- Available arrival time (AAT), 70

B

- Back Propagation neural network (BP neural network), 347
- Backtracking, 37
- Bays, 51
- Best Obj. *See* Best objective function values
- Best objective function values (Best Obj), 179
- Blacklisting, 37
- Boundary conditions
 - freestream, 5, 6
 - inflow, 5
 - outflow, 7
 - spanwise, 7
 - wall, 6, 7
- Boundary layer turbulence transition, 1
 - fundamental waves, 2
 - K-type transition, 2
 - linear wave interactions, 2
 - N-type transition, 2
 - nonlinear wave interactions, 2
 - numerical method, 9
 - secondary instability, 2
 - spectral method, 7
 - Fourier series approximation, 8
 - governing equations, 8
 - with symmetric flow, 8
 - TS waves, 1–2
- BP neural network. *See* Back Propagation neural network
- Browser-based interface, 222–223
- Browsing, 224
- Bullwhip effects
 - and order variations, 161
 - at markets with and without PE, 158t, 163
 - retailer and distributor, 162

C

- Calibration, 342, 346, 350, 351
- Capital Asset Pricing Model (CAPM), 92
- CAPM-based robust random fuzzy mean-variance model, 92
- Cathode mass balances, 316, 317

- CBLS algorithm. *See* Constraint Based Local Search algorithm
 CBT. *See* Computer-based training
 CCD schemes. *See* Combined Compact Difference schemes
 Cell temperature estimation, [321f](#)
 Cell voltage, [318](#)
 Cellular transport system, [326](#)
 CGI. *See* Common Gateway Interface
 Check-by-voting method, [34](#), [37–38](#)
 effect of each function of
 $\varepsilon_{acc} = 0.05$, [45f](#)
 $\varepsilon_{acc} = 0.01$, [46f](#)
 Chirp Spread Spectrum (CSS), [327](#)
 Choice reaction time tasks, [291](#)
 Classical economic order quantity model (Classical EOQ model), [138](#)
 Classical EOQ model. *See* Classical economic order quantity model
 Collection rate (CR), [159](#)
 Colluding attack, [35](#), [38](#)
 colluding methods, [38–39](#)
 example of, [35f](#)
 proposed model, [40](#)
 sabotage-tolerance effect, [39–40](#)
 Colluding rate, error rate and throughput for
 $\varepsilon_{acc} = 0.05$, [42f](#)
 $\varepsilon_{acc} = 0.01$, [43f](#)
 Combinatorial game, [82](#)
 Combined Compact Difference schemes (CCD schemes), [9](#)
 Common Gateway Interface (CGI), [280](#)
 Commons tragedy, [125](#)
 Complete K-ary linking pin structure, [182](#), [183f](#)
 Component refurbishment percentage (CRP), [162](#)
 Computation method, [235](#), [236](#)
 Computer-based training (CBT), [279](#)
 Constraint Based Local Search algorithm (CBLS algorithm), [50](#), [55](#)
 Constraint Programming model (CP model), [50](#)
 Container shipping, [49–50](#)
 Containers, [50](#)
 Continuous Stirred Tank Reactor (CSTR), [316](#)
 Conversion value
 from furniture to log, [254t](#)
 from log to scrap, [254t](#)
 Convex quadratic programming problem, [98](#)
 Corporate Social Responsibility (CSR), [190](#), [200](#)
 CP model. *See* Constraint Programming model
 CR. *See* Collection rate
 Creativity, [218](#)
 Creators, [224](#)
 CRP. *See* Component refurbishment percentage
 CSR. *See* Corporate Social Responsibility
 CSS. *See* Chirp Spread Spectrum
 CSTR. *See* Continuous Stirred Tank Reactor
- D**
 DAE. *See* Differential algebraic equation
 Data Conversion Unit (DCU), [344](#)
 input flow versus output voltage of
 DCU. *See* Data Conversion Unit
 Decision maker (DM), [155](#), [199–200](#)
 Goal Programming for, [249](#), [257](#)
 HMOP1, [107–116](#)
 HMOP2, [107–108](#)
 HMOP3, [108](#)
 HMOP4, [108–109](#)
 HMOP5, [109–110](#)
 HMOP6, [110](#)
 interactive decision making process, [106](#)
 multi-level programming process, [106](#)
 Delta evaluation, [59](#)
 DIAT. *See* Distributor inventory adjustment time
 DICT. *See* Distributor inventory cover time
 Differential algebraic equation (DAE), [310](#)
 Direct reusable component, [154](#), [159](#)
 Direct reusable component percentage (DRCP), [162](#)
 Distributor inventory adjustment time (DIAT), [164](#)
 Distributor inventory cover time (DICT), [164](#)
 Disturbance velocities, transverse
 profiles of, [13f](#)
 DM. *See* Decision maker
 Domineering game, [81](#)
 Downstream amplification of waves, [13f](#)
 Downstream amplification rates, [12f](#)
 DP. *See* Dynamic Programming
 Drawing strategy (ds), [82](#)
 DRCP. *See* Direct reusable component percentage
 ds. *See* Drawing strategy
 DSM. *See* Dynamic Stream Merging
 Dynamic inventory lot-sizing models, [206–207](#)
 Dynamic lot-sizing model, [204](#)
 Dynamic Programming (DP), [167](#), [168](#)
 Dynamic Stream Merging (DSM), [354](#)

E

EC. *See* Evolutionary Computation
 Economic Order Quantity model (EOQ model), 204, 205
 Eigenvalues, stability range of, 12*f*
 EKF. *See* Extended Kalman filter
 EKF algorithm implementation, flow chart for, 320*f*
 Electrochemical kinetics, 316, 318
 cell voltage, 318
 electrochemical reactions, 318
 ionic resistance, 319
 mass transfer effects, 319
 ohmic loss, 319
 open-circuit voltage, 318
 Electrochemical reactions, 318
 Energy balance equations, 317
 EOQ model. *See* Economic Order Quantity model
 ETA. *See* Expected time of arrival
 Evolutionary Computation (EC), 169
 Expected time of arrival (ETA), 68
 Extended Kalman filter (EKF), 310, 312, 328
 continuous-time model, 313
 discrete-time nonlinear dynamic system, 312
 discretization, 313
 implementation
 in Matlab environment, 320
 in output vector, 319
 from nonlinear model, 312
 output, 314
 PEMFC state estimation, 314
 cathode mass balances, 316, 317
 CSTR, 316
 dynamic states, 316
 electrochemical kinetics, 318–319
 energy balance equations, 317–318
 P&ID, 315*f*
 parameters, 316, 317
 principle, 314
 sensitive equation, 313–314
 simulation results, 320–321

F
 Farmer Group and/or Cooperative (FGC), 190
 FGC. *See* Farmer Group and/or Cooperative
 FhG IML. *See* Fraunhofer-Institute for Material Flow and Logistics
 FIC model. *See* Fuzzy inventory control model
 Fickian diffusion equation, 316
 FLC. *See* Fuzzy Logic Control

Flow control, 341
 accurate and inaccurate measurement, 341
 flow domain, 3, 4
 schematic of, 4*f*
 flow measurement technique, 344
 flow rate, 341, 343
 Forward supply chain, 156–158
 component production rate, 156
 for demand, product exchange, satisfaction and collection, 157*f*
 collection rates with and without PE policy, 161*t*
 demand at markets, 158
 with and without PE, 158*t*
 with product exchange, 153
 for product production and distribution, 157*f*
 product production rate, 157–158
 for product refurbishment, 157
 remanufacturing process supplements, 156
 Fractile optimization model, 106, 107
 Fraunhofer-Institute for Material Flow and Logistics (FhG IML), 326
 Freestream boundary conditions, 5–6
 Friction stir weld data
 ANOVA, 268
 experimental set-up, 265
 regression analysis
 FSW process parameters, 267*t*
 linear equations, 267
 multiple, 266
 using parameters, 266
 UTS and FSW parameters, 268*t*
 scatter plots analysis, 268
 electrical resistivity vs. heat input, 269
 schematic diagram of, 262*f*
 statistical analysis
 ANOVA, 264
 FSW process variables, 263–264
 R-square, 264
 regression analysis, 263
 Statistica, 263
 statistical software, 264
 surface plots, 265
 2-dimensional scatter plot, 264
 surface plots analysis, 268, 269
 heat input, 270, 271
 horizontal force, 269–271
 percentage elongation, 272
 top darkest region, 272
 torque, 270, 271
 UTS, 272
 vertical force, 270

- Friction Stir Welding (FSW), 262
 - FSW. *See* Friction Stir Welding
 - Fundamental waves, 2
 - Fusion welding process, 262
 - Fuzzy inputs, 207
 - demand and availability, 207
 - demand membership functions, 207, 208f
 - parameters, 208
 - supply membership functions, 209, 209f
 - Fuzzy inventory control model (FIC model), 203, 205, 207
 - comparison with
 - SM model, 215f
 - WW model, 215f
 - fuzzy inputs, 207
 - demand and availability, 207
 - demand membership functions parameters, 208
 - supply membership functions, 209, 209f
 - fuzzy outputs, 208, 209
 - fuzzy rules, 209, 210–211
 - inference system editor, 208f
 - Fuzzy logic approach
 - FIC system, 203, 205, 213
 - FLC, 204
 - fuzzy inventory control model, 207
 - fuzzy inputs, 207, 208
 - fuzzy outputs, 208, 209
 - fuzzy rules, 209, 210–211
 - inventory lot-sizing system
 - continuous inventory system, 206
 - dynamic inventory lot-sizing models, 206–207
 - static inventory lot-sizing models, 205
 - numerical example, 211
 - cost saving FIC model with SM model, 215r
 - cost saving FIC model with WW model, 215r
 - fluctuation of demand and supply holding cost, 213f
 - inventory data, 211
 - order quantity and reorder point, 214
 - ordering cost, 212f
 - shortage cost, 213f
 - SM models, 211
 - stochastic EOQ model, 214
 - stochastic models, 212
 - total cost, 214f
 - WW models, 211
 - PPB, 204
 - Fuzzy Logic Control (FLC), 204
 - Fuzzy outputs, 208, 209
 - Fuzzy reasoning method, 93
 - Fuzzy rules, 209, 210–211
 - Fuzzy set theory, 154
 - membership function, 155
 - possibility measures, 155
 - triangular fuzzy number, 154
 - Fuzzyfication algorithm. *See also* Overall equipment effectiveness (OEE)
 - approximation analysis, 241, 242
 - average level, 243
 - boundaries, 243
 - combine constraint for variables, 242
 - extreme level, 243
 - flow chart, 239
 - individual constraint for variables, 242
 - pre-processing steps, 239
 - programming algorithm, 238
- G**
- Gender differences of human response, 302
 - auditory reaction test, 305–306, 306
 - results of, 306r
 - HRV test, 306, 307
 - results of, 307r
 - literature review, 302
 - methodology
 - experiment design, 303
 - experiment procedure, 303, 304
 - participants, 302
 - posture test, 304, 305
 - results of, 305r
 - under vibration condition, 301, 302
 - visual reaction test, 304, 305
 - results of, 305r
 - Global localization
 - anchorbox, 333, 334, 335, 337–338
 - experimental results, 337
 - IEEE 802.15.4a measurement error, 331
 - IEEE 802.15.4a measurement model, 332–333
 - KF and EKF, 330
 - laser range finders, 336
 - measurement update, 331
 - motion model, 330, 331
 - odometry data, 335
 - prediction phase, 330
 - transport vehicle, 336, 337
 - update phase, 330
 - GNA. *See* Gauss-Newton Algorithm
 - Goal Programming tool (GP tool), 249
 - GP tool. *See* Goal Programming tool
 - GPU. *See* Graphics processing units
 - Graphics processing units (GPU), 18

G (*cont.*)

- Green's function method, 18
 - algorithm, flowchart for, 19*f*
 - benchmarks, 28
 - charge-based devices, 18
 - CPU-based algorithms, 24*r*
 - implementation, 21, 25
 - GPU-memory addressing, 28, 29
 - LVGPU, 27
 - multi-GPU setup, 30
 - multiple scenarios, 29, 30
 - NVIDIA, 27, 28
 - pipelined block-tridiagonal solver, 27
 - RDMA, 30
 - simulation steps, 29*r*
 - single precision elements data transfers, 25*f*
 - mathematical and physics background
 - algebra equation, 20–21
 - Green's function algorithm flowchart, 19
 - Hamiltonian matrix, 19, 20
 - through matrix inversion, 20
 - in mesoscopic structures, 18
 - Schrödinger equation, 19
 - transmission probability, 21
 - matrix inversion, 23
 - NEGF method, 18
 - optimizations, 21, 22
 - block-tridiagonal solver, 24, 25–27, 25*r*
 - linear algebra functions, 22
 - sparse matrices, 24
- Guass-Newton Algorithm (GNA), 347

H

- H2-VIP. *See* Home-to-Home Video Clip Replication
- H2O. *See* Home-to-Home Online
- Hatch cover, 51
- HAZ. *See* Heat Affected Zone
- Headed Mean Obj. *See* Mean objective function values
- Headed Worst Obj. *See* Worst objective function values
- Heart rate, 302
- Heart rate variety test (HRV test), 303, 306–307
- Heat Affected Zone (HAZ), 262
- Heterogeneous communities
 - pricing model, 124–125
 - resource utilization, 126, 128–130
 - revenue maximization, 126, 128–130
 - tragedies, 128–130
 - anti-commons tragedy, 125
 - commons tragedy, 125
 - reservation price, 126
- Hicks-Hyman model, 291, 292
 - relations, 292
- Hierarchical multiobjective stochastic linear programming problems (HMOP), 107
 - through fractile optimization model, 106, 107
 - HMOP1, 107
 - HMOP2, 107, 108
 - HMOP3, 108
 - HMOP4, 108, 109
 - HMOP5, 109, 110
 - HMOP6, 110, 111
 - interactive algorithm, 113, 114
 - bisection method, 114
 - rules, 115
 - steps, 115
 - test problem, 114
 - using theorem, 114, 115
 - minmax problems, 111–113
 - numerical example, 116–118
 - using reference membership intervals, 110
 - theorem, 113
- High-cubes, 51
- High-order numerical methods, 9
- HMATS, 169, 171
 - flowchart of, 172*f*
 - and TS, comparing results, 178*f*
- HMOP. *See* Hierarchical multiobjective stochastic linear programming problems
- Home-to-Home Online (H2O), 354
- Home-to-Home Video Clip Replication (H2-VIP), 354
- Horizontal winning strategy, 88
- HRV test. *See* Heart rate variety test
- HTTP. *See* HyperText Transfer Protocol
- Human-machine-interfaces, 290. *See also*
 - Gender differences of human response
 - auditory cue, response time to age, 294
 - choice alternative, 295
 - computer in daily life, time spent on, 295
 - education level, 295
 - gender, 295
 - left/right response finger, 295
 - in single, two- and four-choice tasks, 294–295
 - in single-choice task, 294
 - in two- and four-choice tasks, 294
 - decision-making processes, 296, 297

- using Hicks-Hyman model, 291, 292
 - method, 290, 291
 - reaction time, 290
 - response time, 290
 - sensory modality, 297
 - stimulus modalities, 297
 - visual and auditory digit cues, 296
 - visual cue, response time to
 - age, 292
 - choice alternative, 292*f*, 294*f*
 - computer in daily life,
 - time spent on, 293
 - education level, 293
 - in four-choice task, 293
 - gender, 293
 - left/right response finger, 293
 - mean response time, 294
 - right finger response time, 293
 - in single-choice task, 292
 - in two-choice task, 292
- HyperText Transfer Protocol (HTTP), 277
- I**
- I-STIR PDS. *See* Intelligent Stir Welding for Industry and Research Process Development System
 - I-STIR PDS FSW platform, 266*f*
 - IAQ. *See* Indoor air quality
 - ICTs. *See* Information and Communication Technologies
 - Ideal Cycle Time, 236
 - IEEE 802.15.4a CSS, 338
 - IEEE 802.15.4a measurement model, 332, 333
 - IEEE 802.15.4a wireless network, 326
 - ILP. *See* Integer linear program
 - IMU. *See* Inertial measurement unit
 - Indoor air quality (IAQ), 283
 - interface found in, 282*f*
 - Industrial accidents, 276
 - Inertial measurement unit (IMU), 335
 - Inflow boundary conditions, 5
 - Information and Communication Technologies (ICTs), 122
 - lack access, 122
 - monopolistic service providers, 122
 - pricing, 123
 - anti-commons tragedy, 124
 - internet connectivity, 123
 - internet services, 123
 - mechanisms, 124
 - models, 124
 - service providers, 123
 - subsidy driven pricing mechanisms
 - Innovations, 218
 - Semantic Wiki, 218, 219
 - application areas, 220
 - significance, 220
 - specifications, 219
 - web applications, 219
 - WikiML, 219
 - types, 218
 - Integer linear program (ILP), 168
 - Integer Programming model (IP model), 52
 - Integrated forward-reverse supply chain
 - flow chart, 153*f*
 - fuzzy set theory, 154
 - possibility measures, 155
 - triangular fuzzy number, 154, 155
 - problem description
 - forward supply chain, 153
 - reverse supply chain, 154
 - result and discussion
 - bullwhip effects and order variations, 161, 163*r*
 - product exchange policy effect, 160
 - three way recovery policy effect, 161, 161*f*
 - sensitivity analysis
 - collection percentage, 161, 162, 163*r*
 - inventory cover and adjustment time on bullwhip, 162, 164*t*
 - product, component refurbishing percentage on bullwhip, 163*r*
 - refurbish percentage, 162, 163*f*
 - satisfaction rate, 161, 162, 163*r*
 - system dynamics model, 155
 - with and without PE policy, 161
 - demand with product exchange policy, 158
 - forward supply chain, 156–158
 - possibility constraints, 159–160
 - reverse supply chain, 158–159
 - stock-flow diagram, 156
 - Intelligent Stir Welding for Industry and Research Process Development System (I-STIR PDS), 265, 266
 - Interactive decision making process, 106
 - International Labour Organization, 276
 - Internet and Web in education, 277
 - Internet based learning. *See* Web based learning
 - Inventory adjustment time, 162
 - Inventory cover time, 162
 - DIAT, 164
 - DICT, 164
 - RIAT, 164
 - RICT, 163

I (*cont.*)

- Inventory decisions, 204
- Inventory level, transition of, 143^f
- Inventory lot-sizing system
 - continuous inventory system, 206
 - dynamic inventory lot-sizing models, 206, 207
 - static inventory lot-sizing models, 205
- Ionic resistance, 319
- IP model. *See* Integer Programming model

J

- Jacobian matrix, 313, 314, 320, 329, 330
- Just-in-time inventory management, 326

K

- K*-ary tree, 182
- K*-cardinality tree problem. *See* *K*-minimum spanning tree problem (KMSTP)
- K*-minimum spanning tree problem (KMSTP), 167, 168
 - experimental study
 - comparing HMATS to existing algorithms, 176–178, 178^t
 - MA effectiveness analysis, 178–179
 - heuristics based approaches, 168, 169
 - proposed algorithm
- aspiration criterion, 173
- centralization strategy, 176
- generating initial population, 174
- genetic operation, 174–175
- initial solution, 171
- local search, 169, 171, 173
- MA, 170, 174
- parameters, 173
- stopping criteria, 175
- stopping criterion, 176
- TS, 171
- TS-based local search, 175
- updating population, 175
 - real-world decision making problems, 168
 - TS, 169, 170
- k*-spanning tree, 168^f
- K*-type transition, 2
- Kalman Filter state estimation
 - (KF state estimation), 310
 - framework, 312^f
 - principle, 311
- Karush-Kuhn-Tucker condition
 - (KKT condition), 98
- KF state estimation. *See* Kalman Filter state estimation

- Kidnapped-robot problem, 334
- KKT condition. *See* Karush-Kuhn-Tucker condition
- KMSTP. *See* *K*-minimum spanning tree problem
- Knowledge Base Layer, 220
- Knowledge Persistent Layer, 221–222
 - request generated by, example of, 225^f

L

- L4L. *See* Lot for Lot
- LabVIEW GPU Analysis Toolkit (LVGPU), 27
- Lagrange function, 98
- Laser Beam Welding (LBW), 262
- LBW. *See* Laser Beam Welding
- Levenberg-Marquardt algorithm (LMA), 347
- Linear algebra functions, 22
- Linear shipping vessel, 50
 - bays arrangement, 51
 - below-deck part, 51, 52
 - containers, 50, 51
 - discharge ports, 54
 - IMO containers, 51, 52
 - IP model, 52–53
 - low-level constraints, 52
 - minimizing overstay, 52
 - objectives, 51
 - on-deck part, 51, 52
 - OOG containers, 52
 - reefer containers, 51, 52
 - vessel stability, 51–52
- Linking pin organization, 182
 - K*-ary linking pin structure, 182, 183
 - structure, 182
 - total shortening distance, 183
 - and optimal distance, 187^t
- Liquid flow measuring instruments, 341
- LMA. *See* Levenberg–Marquardt algorithm
- Logic based model, 55
 - computational results
 - algorithm analysis, 60, 61^t
 - execution time, 61, 61^f, 62
 - instances, 60
 - incremental computations, 59
 - objectives, 57
 - search states, 57
 - feasibility phase, 58, 59
 - optimality phase, 58, 59
 - placement heuristic, 58
 - slot planning problem, 56
 - stacking rules, 56
 - swapping containers, 57
 - violations, 57

Losing strategy (ls), 82
 Lot for Lot (L4L), 204
 LVGPU. *See* LabVIEW GPU Analysis Toolkit

M

M-first voting method, 34, 37
 MA. *See* Memetic Algorithm
 Mandani, 209
 Man-machine interfaces. *See* Human-machine-interfaces
 Markowitz model, 92
 MASMT. *See* Multicore Analysis and Sparse Matrix Toolkit
 Mass transfer effects, 319
 Master nodes, 335
 Material balance equations, 316
 Math Kernel Library function (MKL function), 22, 28
 MATLAB neural network toolbox (MATLAB nntool box), 346
 Matrix inversion, visualization of different algorithms for, 23f
 Maximum take-off weight (MTOW), 69
 MCFC system. *See* Molten carbonate fuel cell system
 MCPF. *See* Monte Carlo Particle Filter
 Mean objective function values (Headed Mean Obj), 179
 Mean square error (MSE), 342, 347
 Mean-absolute deviation model, 92
 Mean-variance model, 92
 Memetic Algorithm (MA), 170
 Metal Inert Gas welding (MIG), 262
 MFS. *See* The Kiva Mobile Fulfillment System
 MHE. *See* Moving horizon estimation
 Micro-scaled wave interactions, 9
 MIG. *See* Metal Inert Gas welding
 MILP. *See* Mix integer linear programming
 “Mind the Size” rule, 71, 72
 revised, 76r
 Minmax problems, 111–113
 Minimization problem, 170f
 Mix integer linear programming (MILP), 191
 MKL function. *See* Math Kernel Library function
 Mobile nodes, 335
 Modern portfolio theory, 92
 Modern retailers (MR), 190
 Molten carbonate fuel cell system (MCFC system), 310
 Monte Carlo particle filter (MCPF), 330
 flow chart, 331f
 Monte-Carlo simulations, 73, 74

 for “Mind the Size” rule, 74f
 Motion model, 330, 331
 Moving horizon estimation (MHE), 310
 MR. *See* Modern retailers
 MSE. *See* Mean square error
 MSM technology. *See* Multishuttle Move technology
 MTOW. *See* Maximum take-off weight
 Multicore Analysis and Sparse Matrix Toolkit (MASMT), 22
 Multiflow formulation, 168
 Multishuttle Move technology (MSM technology), 326
 Multi-GPU setup, visualization of, 30f

N

$n \times 11$ board game, 85
 horizontal winning strategy, 88
 number of moves, 87–88
 3×11 sub-rectangles, 86, 87
 vertical winning strategy, 86
 N-type transition, 2
 Nature of accidents, 282
 NEGF method. *See* Non-equilibrium Green’s function method
 NLP formulation. *See* Nonlinear programming formulation
 Non-equilibrium Green’s function method (NEGF method), 18
 Non-uniform grid
 computational efficiency, 9
 numerical method
 high-order CCD schemes, 10
 Lagrange polynomial interpolation, 10, 11
 stability, 11
 validation, 11, 12, 13
 sub-harmonic resonance, 13, 14
 downstream amplification, 13, 14
 downstream amplification rates, 12, 12f
 Fourier modes, 13
 transverse profile, 14
 in wall-normal y-direction, 10
 Nonlinear fractional programming problem, 101
 Nonlinear programming formulation (NLP formulation), 310

O

Occupational health and safety, 276
 Odometry data, 335
 OEE. *See* Overall equipment effectiveness

O (*cont.*)

- Ohmic loss, 319
- On-deck part, 51
- Online casebook, interface of, 281*f*
- OOG containers. *See* Out-of-gauge containers
- Open-circuit voltage, 318
- Optimal objective value, 99–101
- Optimality phase, 58
- Order quantity membership functions, 210*f*
- Order variations, 161
 - at markets with and without PE, 163, 163*t*
 - retailer and distributor, 162
- Organizational innovations, 218
- Orr-Sommerfeld equation (OS equation), 2, 11
- OS equation. *See* Orr-Sommerfeld equation
- Out-of-gauge containers (OOG containers), 51
- Outflow boundary condition, 7
- Overall equipment effectiveness (OEE), 231, 234
 - availability, 234, 235
 - availability calculation, flow chart for, 240*f*
 - availability rating, 241*t*
 - flow chart, current, 239*f*
 - losses and performance evaluation, 232
 - OEE computation, 233
 - and six major losses, 233*t*
 - measurement, 235
 - OEE rating, 242*t*
 - PDG, 237, 238
 - block diagram of, 238*f*
 - performance calculation, flow chart for, 240*f*
 - performance rating, 241*t*
 - process manufacturing environments, 235
 - quality qualification, flow chart for, 241*f*
 - quality rating, 242*t*
 - real time application
 - analysis and reporting data collection, 243
 - production floor information, 243, 244
 - signed approach, 238
 - weighted approach method
 - availability, 235
 - computation method, 235
 - existing weighted rating, 237
 - performance, 236
 - quality, 236
 - weighted and fuzzy approaches, 236
 - weighted rating, 236, 237
- Oxygen concentration estimation, 321*f*

P

- P&ID. *See* Piping and instrumentation diagram
- P2P. *See* Peer-to-Peer
- Pallet-wide containers, 51
- Part Period Balancing (PPB), 204
- Particle Filter (PF), 330
- PDG. *See* Program Dependence Graph
- PE policy. *See* Product exchange policy
- Peer-to-Peer (P2P), 359
 - blocking probabilities, 362, 363, 364
 - against arrival rates, 361*f*, 364*f*
 - against buffer size of arrival rates, 364*f*
 - against buffer size of mesh counts, 364*f*
 - against channels, 362*f*
 - against video channels, 363*f*
 - video delivery mechanism with, 360
- PEM fuel cell state estimation (PEMFC state estimation), 314
- PEM fuel cell system model, current density as input for, 321*f*
- PEMFC state estimation. *See* PEM fuel cell state estimation
- Personal Protective Equipment (PPE), 249
- Perum Perhutani (PP), 248
 - and furniture industry relationship, 250*f*
 - PP revenue, forest area as function of proportion of, 258*f*
 - PP-VSU, 249, 250
- PF. *See* Particle Filter
- Piping and instrumentation diagram (P&ID), 315
- Placement heuristics, 54, 58
- PLC. *See* Programmable logic controller
- Poisson's equations, 5
- Portfolio selection problem, 92
 - APT, 95
 - asset pricing theory, 92
 - CAPM, 95
 - formulation with random fuzzy returns, 96
 - convex quadratic programming problem, 98
 - future return, 97–98
 - KKT condition, 98–99
 - Lagrange function, 98
 - maximize total future returns, 97
 - maximizing total variance, 97
 - nonlinear fractional programming problem, 101
 - optimal objective value, 99–101

- safety-first models, 97
 - standard mean-variance portfolio model, 98
 - future returns, 92
 - fuzzy mean values, 93
 - fuzzy reasoning method, 93
 - fuzzy sets, 92, 93
 - investment, 92
 - portfolio models, 92
 - random fuzzy variables, 94
 - single input type fuzzy reasoning method, 95
 - fuzzy inference, 95, 96
 - K rule modules, 95–96
 - maximum values, 96
 - minimum values, 96
 - Possibility constraint programming approach, 159–160
 - Possibility measures, 155
 - Posture test, 303, 304, 305
 - PP. *See* Perum Perhutani
 - PPB. *See* Part Period Balancing
 - PPE. *See* Personal Protective Equipment
 - Pressure control, 341
 - Process innovations, 218
 - Product exchange policy (PE policy), 152
 - Product exchange policy effect, 160
 - Product refurbish percentage (PRP), 162
 - Program Dependence Graph (PDG), 237, 238
 - Programmable logic controller (PLC), 335
 - PRP. *See* Product refurbish percentage
 - Pyramid organization, 182
- Q**
- QoS. *See* Quality of Service
 - Quality, 236
 - flow chart for, 240
 - rating based on, 242
 - Quality of Service (QoS), 122
 - Quantity discount strategy
 - notation
 - and assumptions, 139–140
 - options for, 139
 - poultry farmer, 139
 - numerical examples, 147–148
 - poultry farmer
 - optimal policy, 143, 144–147
 - total profit, 142–143
 - quantity discount problem, 138
 - retailer's optimal response, 141–142
 - retailer's total profit, 140–141
 - sensitivity analysis, 148
 - Quay cranes, 49, 50
 - Querying, 224–225
 - classification approach of, 227f
- R**
- R-square, 264
 - Random fuzzy mean-variance model, 93
 - Random fuzzy returns, 96
 - convex quadratic programming problem, 98
 - future return, 97–98
 - KKT condition, 98–99
 - Lagrange function, 98
 - maximize total future returns, 97
 - maximizing total variance, 97
 - nonlinear fractional programming problem, 101
 - optimal objective value, 99–101
 - safety-first models, 97
 - standard mean-variance portfolio model, 98
 - RDF literals, 221
 - RDMA. *See* Remote direct memory access
 - Readers, 224
 - Received Signal Strength (RSS), 327
 - Reference membership intervals, 110
 - Refurbish percentage, 162
 - Regression analysis. *See also* Sensitivity analysis
 - FSW process parameters, 267
 - linear equations, 267
 - multiple, 266
 - using parameters, 266
 - UTS and FSW parameters, 268
 - Relative humidity (RH), 310
 - Remote direct memory access (RDMA), 30
 - Reorder point membership functions, 210f
 - Reserve price, 125
 - Response Surface Methodology (RSM), 263
 - Retail inventory adjustment time (RIAT), 164
 - Retail inventory cover time (RICT), 163
 - Retailer's optimal response, 141–142
 - Retailer's total profit, 140–141
 - Retreating side (RS), 265
 - Reverse logistics process (RL process), 152
 - Reverse supply chain, 154, 158–159
 - RH. *See* Relative humidity
 - RIAT. *See* Retail inventory adjustment time
 - RICT. *See* Retail inventory cover time
 - RK schemes. *See* Runge–Kutta schemes
 - RL process. *See* Reverse logistics process
 - Round-trip Time of Flight (RTof), 327

R (*cont.*)

- RS. *See* Retreating side
- RSM. *See* Response Surface Methodology
- RSS. *See* Received Signal Strength
- RTof. *See* Round-trip Time of Flight
- Runge-Kutta schemes (RK schemes), 9

S

- Sabotage model, 36
- Sabotage rate s , error rate for, 44*f*
- Sabotage-tolerance methods, 34, 36
 - check-by-voting method, 37–38
 - M -first voting method, 37
- Safety-first model, 92, 97
- Satisfaction rate (SR), 160
- Scatter plots analysis, 268
 - electrical resistivity vs. heat input, 269
- Scenario simulations. *See also* Arrival aircraft sequencing
 - with available arrival time constraints
 - aircraft combination in longer batches, 74
 - fuel savings, 77, 77*f*, 78
 - “Mind the Size” revised, 76
 - simple swap, 77, 78*f*
 - two heavy and one medium aircraft, 74, 75
 - two medium and one heavy aircraft, 75, 76
 - without available arrival time constraints
 - aircraft batches, 71
 - “Mind the Size” rule, 71, 72
 - Monte-Carlo simulations, 73, 74
- SD. *See* System dynamics
- Self-checking component, 281, 283
- Sensitivity analysis, 148
 - collection and satisfaction rate, 161, 162
 - of collection percentage on product refurbishment, 163
 - of inventory cover and adjustment time, 164
 - inventory cover time, 162
 - DIAT, 164
 - DICT, 164
 - RIAT, 164
 - RICT, 163
 - of product, component refurbishing percentage, 163
 - of refurbish percentage, 162, 163
 - of satisfaction rate on bullwhip effect, 163
- SEQ. *See* Sequence Number
- Sequence Number (SEQ), 356
- Sequencing problem formulation, 69–70
- Sequential quadratic programming method (SQP method), 68
- Silver-Meal method (SM method), 204
- Simple reaction time task, 290, 291
- Simulation, topology used in, 355*f*
- Single aircraft trajectory optimization
 - coordinates and waypoint altitudes, 67
 - fuel burn vs. descent time, 68
 - ICAO separation minima, 69
 - optimal descent trajectories, 67
 - quadratic function, 68
 - SQP method and ETA, 68
 - waypoint coordinates, 68
- Single input type fuzzy reasoning method, 95, 96
- SM method. *See* Silver-Meal method
- SMIL. *See* Synchronized Multimedia Integration Language
- SOFC. *See* Solid oxide fuel cell
- Solid oxide fuel cell (SOFC), 310
- Solid-state welding process, 262
- Spanwise boundary condition, 7
- Sparse matrices, 24
- Spectral method, 7
- SQP method. *See* Sequential quadratic programming method
- SR. *See* Satisfaction rate
- Stacks, 51
- Standard mean-variance portfolio model, 98
- Static inventory lot-sizing models, 205
 - continuous inventory system, 206
 - EOQ model, 204, 205
- Statistical analysis. *See also* Regression analysis, 263
 - ANOVA, 264
 - FSW process variables, 263, 264
 - R-square, 264
 - regression analysis, 263
 - Statistica, 263
 - statistical software, 264
 - surface plots, 265
 - 2-dimensional scatter plot, 264
- Stock-flow diagram, 156
 - demand with product exchange policy, 158
 - forward supply chain, 156–158
 - component production rate, 156
 - for demand, product exchange, satisfaction and collection, 157
 - demand at markets, 158
 - for product production and distribution, 157
 - product production rate, 157, 158
 - for product refurbishment, 157

- remanufacturing process supplements, 156
- possibility constraints, 159
 - collection rate, 159, 160
 - satisfaction rate of demand, 160
- reverse supply chain, 158, 159
- Stowage plan, 50
 - linear shipping vessel, 50
 - bays arrangement, 51, 51f
 - below-deck part, 51, 52
 - containers, 50, 51
 - discharge ports, 54
 - IMO containers, 51, 52
 - IP model, 52, 53
 - low-level constraints, 52
 - into master and slot planning, 51f
 - minimizing overstowage, 52
 - objectives, 51
 - on-deck part, 51, 52
 - OOG containers, 52
 - reefer containers, 51, 52
 - vessel stability, 51, 52
 - multi-phase approaches, 54
 - single-phase approaches, 54
 - solution approach, 54
 - CBLS, 55
 - logic based model, 55–57
 - 2-phase hierarchical decomposition, 50, 51
 - CBLS algorithm, 50
 - CP model, 50
 - multi-port master planning phase, 50, 51
 - slot planning phase, 50, 51
- Sub-harmonic resonance, 13, 14
- Subsidy driven pricing model, 132
- Subsidy levels. *See also* Heterogeneous communities, 126, 127–130
 - effect
 - revenue vs. beta and alpha, 133f, 134
 - revenue vs. price, 130–132, 131f
 - subsidy driven pricing model, 132
 - number of packets vs. price, 126, 127, 127f
 - total packets, 128
 - unused network capacity vs. price, 126, 127
- Supplier-manufacturer model
 - mathematical formulation, 252–253
 - mathematical modeling
 - GP tool, 249
 - notations, 250
 - numerical example
 - conversion values, 254
 - furniture demand, 255t
 - furniture price, 255t
 - parameter cost, 254
 - PPE price, 255t
 - teak log price, 255t
- objective function, 251–252
- problem description
 - goal, 249
 - PP-VSU, 249, 250
- results, 256
 - goals satisfaction, 255, 256
 - GP formulation Pareto efficient solution, 256–257
 - target level, 254
- sensitivity analysis
 - planting cost, 257, 258
 - VSU, 257, 258
- solution method, 253
- Surface plots analysis, 265, 268, 269
 - heat input, 270, 271
 - horizontal force, 269–271
 - percentage elongation, 272
 - top darkest region, 272
 - torque, 270, 271
 - UTS, 272
 - vertical force, 270
- SUS. *See* System Usability Scale
- Swapping, 59
 - containers, 57
- Synchronized games, 82
 - draw, 82
 - drawing strategy, 83
 - horizontal wins, 84
 - $n \times 11$ board, 85, 85f
 - horizontal winning strategy, 88
 - number of moves, 87–88
 - 3×11 sub-rectangles, 86, 86t, 87
 - vertical winning strategy, 86
 - possible outcomes, 82, 83
 - vertical and horizontal, 82
 - vertical wins, 84
 - winning strategy, 83
- Synchronized Multimedia Integration Language (SMIL), 278
- Synchronized triomineering, 82
 - examples of, 83–84
 - possible outcomes in, 93t
 - vertical strategy on the $n \times 11$ board of, 85f
- System dynamics (SD), 152
- System dynamics model, 155
 - demand with product exchange policy, 158
 - forward supply chain, 156–158
 - possibility constraints, 159–160
 - reverse supply chain, 158–159
 - stock-flow diagram, 156

S (*cont.*)

System Usability Scale (SUS), 284

T

Tabu Search (TS), 169, 170

and HMATS, comparing results, 178*f*

Tafel equation, 316, 318

TCP/IP. *See* Transmission Control Protocol/Internet Protocol

Temperature control, 341

Temporal discretization, 11

Test set characteristics, slot planning instances, 60*r*

TFN. *See* Triangular fuzzy number

The Kiva Mobile Fulfillment System (MFS), 326

Three dimensionality. *See* Boundary layer turbulence transition

flow domain, 3, 4

formation, 3

governing equation, 4

non-dimensional variables, 4

Poisson's equations, 5

total flow velocity and vorticity, 4

vorticity transport equation, 4–5

ZPG boundary layer flow, 5

Three way recovery option (TWR option), 152 effect of three way recovery, 161, 162

Threshold. *See* Reserve price

TIG. *See* Tungsten Inert Gas welding

Time of Arrival (ToA), 327

first come, first served sequence in, 72*f*

Time to Live (TTL), 356

Time-Shifting, 359, 360, 364

blocking probability, 362, 363, 364, 365

video delivery mechanism with, 360

TM. *See* Traditional market

ToA. *See* Time of Arrival

Tollmien-Schilliching waves (TS waves), 1

Total Productive Maintenance program (TPM program), 231, 232

Total shortening distance, 183

formulation, 183

between every pair of nodes, 183–184

numerical examples, 187

optimal depth, 184–187

TPM program. *See* Total Productive Maintenance program

Traditional market (TM), 193

Transmission Control Protocol/Internet Protocol (TCP/IP), 277

Triangular fuzzy number (TFN), 154, 155 membership function of, 155*f*

Triomineering game, 82

TS. *See* Tabu Search

TS waves. *See* Tollmien-Schilliching waves

TTL. *See* Time to Live

Tungsten Inert Gas welding (TIG), 262

Two heavy and one medium aircraft, 74, 75, 75*f*

Two medium and one heavy aircraft, 75, 75*f*, 76

TWR option. *See* Three way recovery option

U

Ultimate Tensile Strength (UTS), 268

Ultra-Wideband (UWB), 327

Unscented Kalman filter (UKF). *See* Extended Kalman filter (EKF)

UTS. *See* Ultimate Tensile Strength

UWB. *See* Ultra-Wideband

V

Valasindo Sentra Usaha (VSU), 248

effect of PPE total cost of, 258*f*

Van der Pol oscillator, estimated states for, 315*f*

Vapor concentration estimation, 321*f*

Variable Neighborhood Decomposition Search (VNDS), 168

VC system. *See* Volunteer computing systems

Vena-contracta point, 343

Venturi flow meter, 342

characteristics with DCU, 344

input flow vs. output voltage, 344, 345, 346

output pressure, 344

output temperature, 346

output voltage, 346

flow measurement technique, 344

block diagram of, 344*f*

flow rate, 343

results, 348

calibration, 351

optimized ANN, 350*r*

optimized network model, 350*r*

neural network model structure, 349*f*, 350

system response, 350, 350*f*

solution using ANN, 346

MATLAB nntoolbox, 346

MSE, 348, 348*f*

output matrix, 347

regression measure, 347

target graph, 347*f*

training, 347
 temperature effect, on density, 343
 vena-contracta point, 343
 Venturi nozzle, 342, 343f
 Vertical winning strategy, 86
 Vibrations, 301, 302
 Video block broadcasting, 358
 blocking probabilities, 361, 362
 changes in video block placement, 358, 359
 simulation results, 361, 362
 Video block discovery mechanism, 356, 357–358, 358f
 Video block placement, 356f
 changes in, 359f
 Video block replication scheme, 355
 total number of mesh router, 357f
 video block placement, 356
 Video delivery mechanism
 with peer-to-peer, 360f
 with time shifting, 360f
 Video-on-demand system (VoD system), 354
 Violations, 57
 Visual reaction, 302
 Visual reaction test, 303, 304–305
 Visual stimuli test, 291
 VNDS. *See* Variable Neighborhood Decomposition Search
 VoD system. *See* Video-on-demand system
 Volunteer computing systems (VC system), 34
 colluding attack, 38
 accidental colluding method, 39
 colluding method, 38
 direct colluding method, 39
 example, 39
 indirect colluding method, 39
 proposed model, 40
 sabotage-tolerance effect, 39–40
 against colluding attacks, 35
 M-first voting and check-by-voting, 34
 model, 35, 35f
 sabotage model, 36
 sabotage-tolerance methods, 34, 36
 check-by-voting method, 37–38
 M-first voting method, 37
 simulation
 backtracking method, 45, 46
 in check-by-voting, 42, 43
 check-by-voting, effect of, 44, 46
 colluding rate c calculation, 41
 condition, 41
 error rate and throughput, 42, 43
 evaluation method, 40, 41
 function and methods, 45

 parameters, 41r
 re-evaluations and re-calculations, 46
 in sabotage rate, 44
 sabotage rate s calculation, 44
 VC model
 sabotage model, 36
 work-pool-based master-worker model, 35–36
 work-pool-based master-worker model, 35, 36
 worker-to-worker communication, 35
 Vorticity Transport Equation, 3, 4–5
 VSU. *See* Valasindo Sentra Usaha

W

Wagner-Whitin model (WW model), 215
 Wall boundary condition, 6–7
 Wave amplitude, transverse
 profiles of, 14f
 Waypoint coordinates, 68
 WBT. *See* Web-based Training; Web-based training
 Web based learning, 277
 Web page, 281
 Web-based Training (WBT), 277
 advantages, 277
 aim, 278
 clients, 278
 educational benefits, 284–285
 learning materials, 277
 methodology
 construction, 280
 operation and maintenance, 280
 preparation, compilation and collation, 279
 production, 279–280
 review and selection, 278–279
 results
 accessories component, 283–284
 analysis model, 281–282
 nature of accidents, 282
 self-checking component, 281, 283
 snapshot of interface, 281, 282
 Web-based training (WBT), 279
 Weibull distribution, 140
 Weighted approach method
 computation method, 235
 availability, 235
 OEE, 236
 performance, 236
 quality, 236
 existing weighted rating, 237
 weighted and fuzzy approaches, 236

W (*cont.*)

- weighted rating, 236–237
- Welding, 261
 - electrical resistivity versus heat input, scatter plot of, 269*f*
 - fusion welding process, 262
 - primary functions, 262
 - property relationships, 262–263
 - solid-state welding process, 262
 - surface plot relating
 - heat input, spindle speed and feed rate, 271*f*
 - horizontal force (F_x), spindle speed and feed rate, 269*f*
 - horizontal force F_y, spindle speed and feed rate, 270*f*
 - percentage elongation, spindle speed and feed rate for, 272*f*
 - torque (T), spindle speed and feed rate, 271*f*
 - vertical force (F_z), spindle speed and feed rate, 271*f*
 - UTS, spindle speed and feed rate, 272*f*
- Wiki markup language (WikiML), 219
- Wiki system, 218
 - application areas, 220
 - educational environments, 220
 - knowledge management systems, 220
 - ontology engineering, 220
 - web applications, 219
 - semantic wiki, 218, 219
 - significance, 220
 - specifications, 219
 - WikiML, 219
- Wiki-I, performance indicators for, 228*f*
- Wiki-I architecture, 220, 221*f*
 - browser-based interface, 222–223
 - exploiting semantics and knowledge representation
 - browsing, 224
 - querying, 224–225
 - functionalities
 - context-aware presentation, 226
 - enhanced navigation, 226
 - semantic search, 226–227
 - typing/annotating of links, 225
 - knowledge evaluation, 223
 - knowledge links, 222*f*
 - layers, 220, 221
 - Knowledge Base Layer, 220
 - Knowledge Persistent Layer, 221–222
 - performance assessment, 228
 - adhesion of students, 227
 - advantages, 227
 - indicators, 228
- Wiki-I interface, 223*f*
- Wiki-I server, 224*f*
- WikiML. *See* Wiki markup language
- Wikis users, 224
- Wikitext, 218–219
- Winning strategy (ws), 82
- Wireless mesh network (WMN), 353–354
 - simulation results, 360–363
 - system architecture
 - network topology, 355
 - video block discovery mechanism, 356, 357–358
 - video block replication scheme, 355, 356
 - system enhancements
 - peer-to-peer, 359, 360
 - time-shifting, 359, 360
 - video block broadcasting, 358–359
- Wireless Sensor Network (WSN), 326
- Wireless sensor node, 335*f*
- Wireless synchronisation, 336*f*
- WMN. *See* Wireless mesh network
- Wooden furniture industry, 248
- Work-pool-based master-worker model, 35–36
- Worst objective function values (headed Worst Obj), 179
- ws. *See* Winning strategy
- WSN. *See* Wireless Sensor Network
- WW model. *See* Wagner-Whitin model

X

- XML binding format, 225
- XML Schema datatypes, 221

Z

- Zero pressure gradient boundary layer flow (ZPG boundary layer flow), 5

Author Index

A

Akinlabi, Esther T., [261](#)
Akinlabi, Stephen A., [261](#)
Amin, Vivek, [17](#)
Andreeva-Mori, Adriana, [65](#)

C

Chan, Alan Hoi Shou, [275](#), [289](#)
Chen, J. C., [1](#)
Chen, Ke, [275](#)
Chen, Weijia, [1](#)
Cincotti, Alessandro, [81](#)

D

Das, Debabrata, [151](#)
Dutta, Pankaj, [151](#)

F

Fukushi, Masaru, [33](#)
Funabiki, Nobuo, [33](#)

G

Gomes, Samuel, [217](#)
Guo, Qingqiang, [167](#)

H

Habibie, Achmad, [247](#)
Hasuike, Takashi, [91](#)

Hilaire, Vincent, [217](#)
Hisjam, Muh, [189](#), [247](#)
Ho, King Man, [353](#)
Homma, Toshimichi, [137](#)

I

Itoh, Eri, [65](#)

J

Jacob, Jan, [17](#)
Jensen, Rune Møller, [49](#)
Ji, Guangji, [309](#)

K

Katagiri, Hideki, [91](#), [167](#)
Kawakatsu, Hidefumi, [137](#), [181](#)
Kirsch, Christopher, [325](#)
Krause-Kyora, Bodo, [17](#)

L

Lahoud, Inaya, [217](#)
Lategahn, Julian, [325](#)
Liu, Jichen, [309](#)
Lo, Kwok Tung, [353](#)

M

Manikandan, G., [231](#)
Maran, M., [231](#)

M (*cont.*)

Mitsuishi, Takashi, [181](#)
Monticolo, Davy, [217](#)
Müller, Marcel, [325](#)

N

Nakanish, Toru, [33](#)
Nel, Andre L., [121](#)
Ng, Annie W. Y., [289](#)

P

Pacino, Dario, [49](#)
Phruksarphanrat, Busaba, [203](#)

R

Röhrig, Christof, [325](#)
Roy, Binoy Krishna, [341](#)
Ruan, Qing, [17](#)

S

Sawada, Kiyoshi, [137](#), [181](#)
Schmidt, Darren, [17](#)
Sinova, Jairo, [17](#)
Sumbwanyambe, Mbuyu, [121](#)
Sutopo, Wahyudi, [189](#), [247](#)

Suzuki, Shinji, [65](#)
Sze, Kwok Tung, [353](#)

T

Tanthatemee, Thipbodee, [203](#)
Thiagarajan, K., [231](#)
Tsuda, Hiroshi, [91](#)

V

Venkata, Santhosh Krishnan, [341](#)

W

Watanabe, Kan, [33](#)
Wenzel, Lothar, [17](#)
Widodo, Kuncoro Harto, [247](#)

Y

Yano, Hitoshi, [105](#)
Yau, Hon Keung, [301](#)
Yuniaristanto, [189](#)

Z

Zhou, Su, [309](#)