

Mohammad S. Obaidat · Slawomir Koziel  
Janusz Kacprzyk · Leifur Leifsson  
Tuncer Ören *Editors*

# Simulation and Modeling Methodologies, Technologies and Applications

International Conference, SIMULTECH  
2013 Reykjavík, Iceland, July 29–31,  
2013 Revised Selected Papers

# **Advances in Intelligent Systems and Computing**

Volume 319

## **Series editor**

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland  
e-mail: [kacprzyk@ibspan.waw.pl](mailto:kacprzyk@ibspan.waw.pl)

## *About this Series*

The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually, all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, health care, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within “Advances in Intelligent Systems and Computing” are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and worldwide distribution. This permits a rapid and broad dissemination of research results.

## *Advisory Board*

### Chairman

Nikhil R. Pal, Indian Statistical Institute, Kolkata, India  
e-mail: nikhil@isical.ac.in

### Members

Rafael Bello, Universidad Central “Marta Abreu” de Las Villas, Santa Clara, Cuba  
e-mail: rbellop@uclv.edu.cu

Emilio S. Corchado, University of Salamanca, Salamanca, Spain  
e-mail: escorchado@usal.es

Hani Hagras, University of Essex, Colchester, UK  
e-mail: hani@essex.ac.uk

László T. Kóczy, Széchenyi István University, Győr, Hungary  
e-mail: koczy@sze.hu

Vladik Kreinovich, University of Texas at El Paso, El Paso, USA  
e-mail: vladik@utep.edu

Chin-Teng Lin, National Chiao Tung University, Hsinchu, Taiwan  
e-mail: ctklin@mail.nctu.edu.tw

Jie Lu, University of Technology, Sydney, Australia  
e-mail: Jie.Lu@uts.edu.au

Patricia Melin, Tijuana Institute of Technology, Tijuana, Mexico  
e-mail: epmelin@hafsamx.org

Nadia Nedjah, State University of Rio de Janeiro, Rio de Janeiro, Brazil  
e-mail: nadia@eng.uerj.br

Ngoc Thanh Nguyen, Wroclaw University of Technology, Wroclaw, Poland  
e-mail: Ngoc-Thanh.Nguyen@pwr.edu.pl

Jun Wang, The Chinese University of Hong Kong, Shatin, Hong Kong  
e-mail: jwang@mae.cuhk.edu.hk

More information about this series at <http://www.springer.com/series/11156>

Mohammad S. Obaidat  
Slawomir Koziel · Janusz Kacprzyk  
Leifur Leifsson · Tuncer Ören  
Editors

# Simulation and Modeling Methodologies, Technologies and Applications

International Conference, SIMULTECH 2013  
Reykjavík, Iceland, July 29–31, 2013 Revised  
Selected Papers

 Springer



*Editors*

Mohammad S. Obaidat  
Department of Computer Science  
Monmouth University  
West Long Branch, NJ  
USA

Janusz Kacprzyk  
Systems Research Institute  
Polish Academy of Sciences  
Warsaw  
Poland

Slawomir Koziel  
Leifur Leifsson  
School of Science and Engineering  
Reykjavik University  
Reykjavik  
Iceland

Tuncer Ören  
Faculty of Engineering  
University of Ottawa  
Ottawa, ON  
Canada

ISSN 2194-5357

ISBN 978-3-319-11456-9

DOI 10.1007/978-3-319-11457-6

ISSN 2194-5365 (electronic)

ISBN 978-3-319-11457-6 (eBook)

Library of Congress Control Number: 2014953111

Springer Cham Heidelberg New York Dordrecht London

© Springer International Publishing Switzerland 2015

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](http://www.springer.com))

# Preface

This book includes extended and revised versions of a set of selected papers from the 3rd International Conference on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH 2013) which was coorganized by the Reykjavík University (RU) and sponsored by the Institute for Systems and Technologies of Information, Control and Communication (INSTICC). SIMULTECH 2013 was held in cooperation with the ACM SIGSIM—Special Interest Group (SIG) on Simulation and Modeling (SIM), Movimento Italiano Modellazione e Simulazione (MIMOS) and AIS Special Interest Group on Modeling and Simulation (AIS SIGMAS) and technically co-sponsored by the Society for Modeling and Simulation International (SCS), Liophant Simulation, Simulation Team, and International Federation for Information Processing (IFIP).

This conference brings together researchers, engineers, applied mathematicians, and practitioners interested in the advances and applications in the field of system simulation. We believe the papers here published, demonstrate new and innovative solutions, and highlight technical problems that are challenging and worthwhile.

SIMULTECH 2013 received 142 paper submissions from 39 countries in all continents. A double-blind paper review was performed by the International Program Committee members, all of them recognized in at least one of the main conference topic areas. After the reviewing process, 35 papers were selected to be published as full papers and 58 papers were selected as short papers. The full paper acceptance ratio was thus 25 %, and the total oral paper acceptance ratio was less than 41 %.

The papers included in this book were selected from those with the best reviews taking also into account the quality of their presentation at the conference, assessed by the session chairs. Therefore, we hope that you find the papers included in this book interesting, and we trust they may represent a helpful reference.

We wish to thank all those who supported and helped to organize the conference. On behalf of the conference Organizing Committee, we would like to thank the authors, whose work mostly contributed to a very successful conference and the members of the Program Committee, whose expertise and diligence were instrumental to ensure the quality of final contributions. We also wish to thank all the

members of the Organizing Committee whose work and commitment were invaluable. Thanks also are due to the organizations that technically cosponsored the conference. Last but not least, we would like to thank INSTICC for sponsoring and organizing the conference.

December 2013

Mohammad S. Obaidat  
Slawomir Koziel  
Janusz Kacprzyk  
Leifur Leifsson  
Tuncer Ören

# Organization

## Conference Co-chairs

Mohammad S. Obaidat, Monmouth University, USA  
Slawomir Koziel, Reykjavík University, Iceland

## Program Co-chairs

Tuncer Ören, University of Ottawa, Canada  
Janusz Kacprzyk, Systems Research Institute—Polish Academy of Sciences,  
Poland  
Leifur Leifsson, Reykjavík University, Iceland

## Organizing Committee

Marina Carvalho, INSTICC, Portugal  
Helder Coelhas, INSTICC, Portugal  
Bruno Encarnação, INSTICC, Portugal  
Ana Guerreiro, INSTICC, Portugal  
André Lista, INSTICC, Portugal  
Andreia Moita, INSTICC, Portugal  
Raquel Pedrosa, INSTICC, Portugal  
Vitor Pedrosa, INSTICC, Portugal  
Cláudia Pinto, INSTICC, Portugal  
Ana Ramalho, INSTICC, Portugal  
Susana Ribeiro, INSTICC, Portugal  
Sara Santiago, INSTICC, Portugal

André Santos, INSTICC, Portugal  
Fábio Santos, INSTICC, Portugal  
Mara Silva, INSTICC, Portugal  
José Varela, INSTICC, Portugal  
Pedro Varela, INSTICC, Portugal

## **Program Committee**

Magdiel Ablan, Venezuela  
Erika Ábrahám, Germany  
Marco Aldinucci, Italy  
Mikulas Alexik, Slovak Republic  
Manuel Alfonseca, Spain  
Musab AlTurki, Saudi Arabia  
Jan Awrejcewicz, Poland  
Gianfranco Balbo, Italy  
Simonetta Balsamo, Italy  
Juan Antonio Barcelo, Spain  
Isaac Barjis, USA  
Fernando Barros, Portugal  
Lucian Bentea, Norway  
Louis Birta, Canada  
Wolfgang Borutzky, Germany  
Ipek Bozkurt, USA  
Marian Bubak, Poland  
Juan C. Burguillo-Rial, Spain  
Christian Callegari, Italy  
Jesus Carretero, Spain  
Federico Cecconi, Italy  
François E. Cellier, Switzerland  
Srinivas Chakravarthy, USA  
Ioannis Chatzigiannakis, Greece  
Bo Chen, UK  
Chun-Hung Chen, USA  
Dan Chen, China  
E. Jack Chen, USA  
Jiangzhuo Chen, USA  
Jean-Yves Choley, France  
Bruno Ciciani, Italy  
Silvano Cincotti, Italy  
Christophe Combastel, France  
Douglas Creighton, Australia  
Andrea D'Ambrogio, Italy

Saber Darmoul, Saudi Arabia  
Gabriella Dellino, Italy  
Atakan Dogan, Turkey  
Werner Dubitzky, UK  
Bruce Edmonds, UK  
Stephan Eidenbenz, USA  
Sabeur Elkosantini, Tunisia  
Roland Ewald, Germany  
Denis Filatov, Mexico  
Paul Fishwick, USA  
Ian Flood, USA  
Claudia Frydman, France  
José Manuel Galán, Spain  
Charlotte Gerritsen, The Netherlands  
Norbert Giambiasi, France  
Daniele Gianni, The Netherlands  
John (Yannis) Goulermas, UK  
Jan Tommy Gravdahl, Norway  
Murat Gunal, Turkey  
Mykola Gusti, Austria  
Zhi Han, USA  
Scott Y. Harmon, USA  
Monika Heiner, Germany  
Herbert Hoeger, Venezuela  
Brian Hollocks, UK  
Xiaolin Hu, USA  
Eric Innocenti, France  
Luis Izquierdo, Spain  
Segismundo Samuel Izquierdo, Spain  
Mats Jägstam, Sweden  
András Jávör, Hungary  
Tania Jiménez, France  
Björn Johansson, Sweden  
Cara Kahl, Germany  
Rihard Karba, Slovenia  
Korina Katsaliaki, Greece  
Peter Kemper, USA  
William Knottenbelt, UK  
Juš Kocijan, Slovenia  
Petia Koprinkova-Hristova, Bulgaria  
Jirí Kunovský, Czech Republic  
Witold Kwasnicki, Poland  
Stephanie Jane Lackey, USA  
Timo Lainema, Finland  
Béla Lakatos, Hungary

Kennard Lavers, USA  
Sanghyun Lee, USA  
Leifur Leifsson, Iceland  
Johannes Lüthi, Austria  
Emilio Jiménez Macías, Spain  
Carla Martin-Villalba, Spain  
Radek Matušu, Czech Republic  
Adel Mhamdi, Germany  
Qi Mi, USA  
Federico Milani, Italy  
Gabriele Milani, Italy  
Jairo R. Montoya-Torres, Colombia  
Il-Chul Moon, Korea, Republic of  
Àngela Nebot, Spain  
Manuel Noguera, Spain  
Michael J. North, USA  
James J. Nutaro, USA  
Peter Csaba Ölveczky, Norway  
Stephan Onggo, UK  
C. Michael Overstreet, USA  
Ioannis Paraskevopoulos, UK  
George Pavlidis, Greece  
Ana Peleteiro, Spain  
Petr Peringer, Czech Republic  
L. Felipe Perrone, USA  
Arne Petermann, Germany  
H. Pierreval, France  
Gary Polhill, UK  
Marta Posada, Spain  
Francesco Quaglia, Italy  
Jacinto A. Dávila Quintero, Venezuela  
Urvashi Rathod, India  
Manuel Resinas, Spain  
M.R. Riazi, Kuwait  
José Risco-Martín, Spain  
Andrea Emilio Rizzoli, Switzerland  
Paolo Romano, Portugal  
Rosaldo Rossetti, Portugal  
Ella E. Roubtsova, The Netherlands  
Willem Hermanus le Roux, South Africa  
Jerzy W. Rozenblit, USA  
Jean-François Santucci, France  
Rune Schlanbusch, Norway  
Florence Sèdes, France  
Philippe Serré, France

Jaime Sichman, Brazil  
Jaroslav Sklenar, Malta  
Jefrey Smith, USA  
Yuri Sotskov, Belarus  
James C. Spall, USA  
Giovanni Stea, Italy  
Steffen Straßburger, Germany  
Nary Subramanian, USA  
Samarth Swarup, USA  
Claudia Szabo, Australia  
Antuela A. Tako, UK  
Elena Tànfani, Italy  
Pietro Terna, Italy  
Andreas Tolk, USA  
Klaus G. Troitzsch, Germany  
Emmanuel Tseklevs, UK  
Kay Tucc, Venezuela  
Bruno Tuffin, France  
Alfonso Urquia, Spain  
Mayerlin Uzcategui, Venezuela  
Timo Vepsäläinen, Finland  
Manuel Villen-Altamirano, Spain  
Natalie van der Wal, The Netherlands  
Frank Werner, Germany  
Philip A. Wilsey, USA  
Kuan Yew Wong, Malaysia  
Yiping Yao, China  
Nong Ye, USA  
Gregory Zacharewicz, France  
František Zboril, Czech Republic  
Durk Jouke van der Zee, The Netherlands  
Suiping Zhou, UK  
Armin Zimmermann, Germany  
Leon Zlajpah, Slovenia  
Konstantinos Zografos, Greece  
Paolo Zuliani, UK

## **Auxiliary Reviewers**

Scott Bourne, Australia  
Andrea Bracciali, UK  
Diego Didona, Portugal  
Graçaliz Pereira Dimuro, Brazil



Pierangelo Disanzo, Italy  
Fei Liu, China  
Luis Gustavo Nardin, Brazil  
Alessandro Pellegrini, Italy  
Christian Rohr, Germany  
Martin Schwarick, Germany  
Irfan Uddin, Italy

## **Invited Speakers**

François E. Cellier, ETH Zürich, Switzerland  
Alexander Smirnov, SPIIRAS, Russian Academy of Sciences, Russian Federation  
Norbert Giambiasi, Aix-Marseille Université, France

# Contents

## Part I Invited Paper

<b>Context-Aware Decision Support in Dynamic Environments: Theoretical and Technological Foundations</b> . . . . .	3
Alexander Smirnov, Tatiana Levashova, Nikolay Shilov and Alexey Kashevnik	

## Part II Papers

<b>Creating Hybrid Simulation Systems Using a Flexible Meta Data Approach</b> . . . . .	23
Juergen Rossmann, Michael Schluse and Ralf Waspe	
<b>Simulation of Real-Time Multiprocessor Scheduling Using DES</b> . . . . .	37
Maxime Chéramy, Anne-Marie Déplanche and Pierre-Emmanuel Hladik	
<b>Epidemics and Their Implications in Urban Environments: A Case Study on a National Scope</b> . . . . .	55
Robert Elsässer, Adrian Ogierman and Michael Meier	
<b>Hybrid Multilinear Modeling and Applications</b> . . . . .	71
Georg Pangalos, Annika Eichler and Gerwald Lichtenberg	
<b>Modeling Interdependent Socio-technical Networks: The Smart Grid—An Agent-Based Modeling Approach</b> . . . . .	87
Daniël Worm, David Langley and Julianna Becker	

**A Heuristic Bidding Price Decision Algorithm Based on Cost Estimation Accuracy Under Limited Engineering Man-Hours in EPC Projects.** . . . . . 101  
 Nobuaki Ishii, Yuichi Takano and Masaaki Muraki

**Intelligent Agents for Human Behavior Modeling as Support to Operations.** . . . . . 119  
 Agostino G. Bruzzone, Marina Massei, Simonluca Poggi, Christian Bartolucci and Angelo Ferrando

**Alternative to Multifractal Analysis of Scalable Random Variables Applied to Measured and Estimated Soil Properties at an Arizona Field Site.** . . . . . 133  
 Alberto Guadagnini, Shlomo P. Neuman, Marcel G. Schaap and Monica Riva

**Adaptive Neuro Fuzzy Inference System Used to Build Models with Uncertain Data: Study Case for Rainfed Maize in the State of Puebla (Mexico)** . . . . . 145  
 Anáis Vermonden, Carlos Gay-García and Iván Paz-Ortiz

**Separation of Carbon Dioxide from Synthesis Gas Containing Steam by Pressure Swing Adsorption at Mid-high Temperature** . . . . . 157  
 Cheng-tung Chou, Yu-Hau Shih, Yu-Jie Huang and Hong-sung Yang

**Fuzzy Climate Scenarios for Temperature Indicate that Things Could Be Worse Than Previously Thought** . . . . . 171  
 Carlos Gay García and Oscar Sánchez Meneses

**Efficient Design of Inline E-Plane Waveguide Extracted Pole Filters Through Enhanced Equivalent Circuits and Space Mapping.** . . . . . 185  
 Oleksandr Glubokov, Slawomir Koziel and Leifur Leifsson

**Decomposition and Space Mapping for Reduced-Cost Modeling of Waveguide Filters.** . . . . . 199  
 Slawomir Koziel, Stanislav Ogurtsov and Leifur Leifsson

**Quasi-Monte Carlo and RBF Metamodeling for Quantile Estimation in River Bed Morphodynamics** . . . . . 211  
 Tanja Clees, Igor Nikitin, Lialia Nikitina and Sabine Pott

**Multi-objective Optimization by Using Modified PSO  
Algorithm for Axial Flow Pump Impeller . . . . . 223**  
H.S. Park and Fu-qing Miao

**A Public Health Model for Simulating Policy Interventions  
to Reduce Nonmedical Opioid Use . . . . . 239**  
Alexandra Nielsen, Wayne Wakeland and Amanuel Zimam

**Supervisory Fuzzy Cognitive Map Structure for Triage Assessment  
and Decision Support in the Emergency Department. . . . . 255**  
Voula C. Georgopoulos and Chrysostomos D. Stylios

**Author Index . . . . . 271**

**Part I**  
**Invited Paper**

# Context-Aware Decision Support in Dynamic Environments: Theoretical and Technological Foundations

Alexander Smirnov, Tatiana Levashova, Nikolay Shilov  
and Alexey Kashevnik

**Abstract** The paper addresses the issue of context-aware operational decision support in dynamic environments. The context model specifies conceptual knowledge describing the situation and problems to be solved in this situation. This model comprises knowledge captured from an application ontology, which is formalized by a set of constraints. The context aware decision support system (DSS) developed within the research has a service-oriented architecture. The Web-services constituting the architecture provide the DSS with the contextualized information from information resources, solve problems specified in the context, and participate in decision making. A decision making model overstepping the limits of the three-phase Simon's model is offered. The paper proposes a set of technologies that can be used to implement the ideas behind the research. An application of these ideas is illustrated by an example of usage of the developed DSS for planning fire response actions.

**Keywords** Context-aware decision support · Context model · Decision making model · Service-oriented architecture · Emergency response · Ridesharing

---

A. Smirnov (✉) · T. Levashova · N. Shilov · A. Kashevnik  
St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences,  
St. Petersburg, Russia  
e-mail: smir@iias.spb.su

T. Levashova  
e-mail: Tatiana.levashova@iias.spb.su

N. Shilov  
e-mail: nick@iias.spb.su

A. Kashevnik  
e-mail: alexey@iias.spb.su

A. Smirnov  
ITMO University, St. Petersburg, Russia

## 1 Introduction

Operational decision support is required in situations happened in dynamic, rapidly changing, and often unpredictable distributed environments. Such situations can be characterized by highly decentralized, up-to-date data sets coming from various resources. The goals of context-aware support to operational decision making are to timely provide the decision makers with up-to-date information, to assess the relevance of information and knowledge to the decision, and to gain insight in seeking and evaluating possible decision alternatives.

The present research addresses theoretical and technological foundations of context-aware operational decision support in dynamic environments. The theoretical foundations are built around ontologies. The ontologies are a widely accepted means for context information modeling. They provide efficient facilities to represent application knowledge and to enable the resources of the dynamic environments to be context-aware and interoperable. The proposed fundamentals are supported by advanced intelligent technologies with their application to the Web.

The decision making model used in the research follows the three-phase model proposed by Simon [14]. The used model oversteps the limits of the three-phase model towards automatic search for an efficient workable decision and communications on the decision implementation.

The ideas behind the research are incorporated into a context-aware decision support system (DSS). The developed DSS has service-oriented architecture. Such architecture facilitates the interactions of service components and the integration of new ones [1, 12, 19]. The system is intended to support decisions on involvement of independent parties in the joint actions according to the current situation and scheduling these actions.

The rest of the paper is structured as follows. Section 2 proposes theoretical foundations to be followed when building context-aware DSSs. Section 3 presents technologies supporting the proposed theoretical foundations and service-oriented architecture of the developed DSS. Application of the presented ideas to the emergency management domain is illustrated in Sect. 4. Main research results are summarized in the conclusion.

## 2 Theoretical Foundations

Decision support in the dynamic environments has to take into account constant environmental changes. In the present research, resources of the environment provide information of any changes to the DSS. These resources are referred to as *information resources*. The information resources perform the needed computations and solve problems, as well. The collection of information resources comprises various kinds of sensors, electronic devices, databases, services, etc. Besides

information resources, the research distinguishes one more type of resources that is *acting resources*. These resources include people and/or organizations that can be involved in the joint actions.

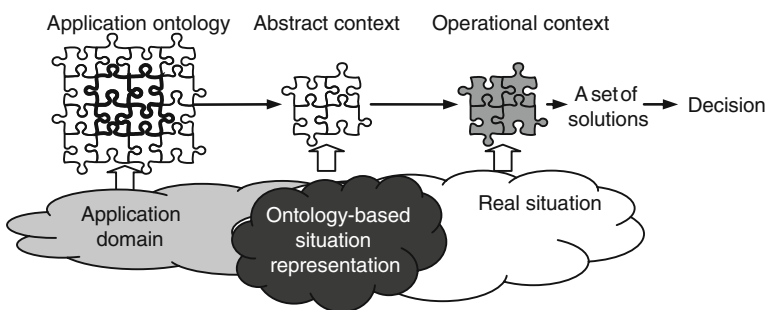
The research follows the knowledge-based methodology to building DSSs. The idea behind the research is to represent the application knowledge by means of constraints. This knowledge is described using two independent sorts of reusable components: domain ontology and task ontology. The domain ontology represents conceptual knowledge about the application domain. The task ontology describes problems occurring in the application domain and methods for achieving solutions to these problems (problem-solving methods). The both components make up the application ontology, which is represented as a set of constraints. This ontology specifies non-instantiated knowledge.

The resources' representations are supposed to be compatible with the ontology representation. The application ontology and the resources' representations are aligned. The alignment indicates what information resource(s) instantiates the given property of the given object specified in the ontology.

In the research, context model serves to represent the knowledge about a decision situation (the settings in which decisions occur and the problems requiring solutions). Context is suggested being modeled at two levels: abstract and operational. These levels are represented by abstract and operational contexts, respectively (Fig. 1).

*Abstract context* is an ontology-based model integrating information and knowledge relevant to the current decision situation. The DSS's user (the decision maker) in his/her request to the DSS indicates the type of the current situation or smart sensors provide this type to the system. The relevant information and knowledge are extracted from the application ontology. As the two components make up this ontology, the abstract context specifies domain knowledge describing the current situation and problems to be solved in this situation.

The abstract context reduces the amount of knowledge represented in the application ontology to the knowledge relevant to the decision situation. In the application ontology this knowledge is related to the resources via the alignment, therefore the abstract context allows the set of resources to be reduced to the



**Fig. 1** Context-aware decision support



resources needed to instantiate knowledge specified in the abstract context. The reduced set of resources is referred to as contextual resources.

*Operational context* is an instantiation of the domain constituent of the abstract context with data provided by the contextual resources. This context reflects any changes in environmental information, in this way it is a near real-time picture of the current situation. The operational context embeds the constraint-based specifications of the problems to be solved. Those input parameters of the problems, which correspond to properties of the objects specified in the domain constituent, are instantiated.

The embedded in the operational context problems are processed as a constraint satisfaction problem in its enumeration form. As a result, a set of feasible alternative ‘satisfactory’ solutions in the current situation is produced. Each solution is a plan of joint actions for the acting resources in the current situation. Decision making is regarded as a choice between the alternatives.

If one or more efficiency criteria are applied to the set of feasible solutions an efficient solution can be found. The efficient solution is considered as the workable decision. The acting resources included in the efficient plan communicate with the DSS in the person of the decision maker on acceptance/rejection of this plan, i.e. on the plan implementation.

In order to enable capturing, monitoring, and analysis of the implemented decisions and their effects the abstract and operational contexts with references to the respective decisions are retained in an archive. As a result, the DSS is provided with reusable models of decision situations. These models, for instance, are used to reveal user preferences based on the analysis of the operational contexts in conjunction with the implemented decisions.

Search for a ‘satisfactory’ decision is the main principle of the decision making model proposed by Simon [14]. The main conclusions from the Simon’s investigation of decision making process can be formulated as follows: the efforts of decision makers to evaluate consequences of the possible alternatives and dependence of the decision makers on the multiple factors influencing their choice should be minimized. Simon proposed a ‘satisfactory’ decision as a result of decision making. That is a decision that is neither efficient nor optimal, but the decision that satisfies all the stakeholders interested in it.

The presented here constrained-based approach enables to express the multiple influencing factors (e.g., the preferences of the stakeholders interested in the decision, intervals of the resources’ availabilities, the resources’ costs, etc.) by means of constraints. The constraints formalizing the multiple factors along with the constraints specified in the operational context are processed as a constraint satisfaction problem. A set of feasible (satisfactory) plans is the result of problem solving. At that, these plans do not depend on decision makers’ attentions, information they have, or stress. Moreover, the decision makers are saved from information overload. The decision maker can choose any plan from the set or take advantage of some predefined efficiency criteria.

The Simon’s model specifies decision making consisting of “intelligence”, “design”, and “choice” phases. Table 1 shows the correspondences between the

**Table 1** Three-phase model

Phase	Phase content	Steps of Simon's model	Steps of constraint-based approach
Intelligence	Finding, identifying, and formulating the problem or situation that calls for a decision	Fixing goals	Abstract context creation
		Setting goals	Operational context producing
Design	Search for possible decisions	Designing possible alternatives	Constraint-based generation of feasible alternative satisfactory solutions
Choice	Evaluation of alternatives and choosing one of them	Choice of a satisfactory decision	Choice of an efficient satisfactory decision
Implementation	Putting the decision into action	–	Communications of the actors with the DSS for their actions

steps of the constrained-based approach and the phases of the Simon's model. The proposed approach exceeds the bounds of the Simon's model proposing two more steps: search for an efficient satisfactory decision and communications on the implementation of this decision.

In the following Section, technologies supporting the proposed theoretical foundations are presented.

### 3 Technological Foundations

According to the theoretical foundations, the application ontology should be represented by means of constraints. Knowledge specified in such a way can be treated as a constraint satisfaction problem. The formalism of object-oriented constraint networks (OOCN) [15] is proposed to be used for the ontology representation. This formalism supports object-oriented knowledge representation (class-attribute-range modeling) and allows specification of the following sets of constraints: (1) *taxonomical* (“is-a”) relationships, (2) *hierarchical* (“part-of”) relationships, (3) *class cardinality* restrictions, (4) *class compatibility* relationships, (5) *associative* relationships, and (6) *functional* relations.

DSSs intended for the dynamic environments become more efficient when they take advantage of Web service technologies. Hence, it is proposed to implement the DSS as a service-oriented system. Web services can support and simplify the exchange of context information; they enable service-oriented systems to utilize various types of context information to adapt their behaviors and operations to dynamic changes. Service-oriented architecture can support integration and consolidation of activities of independent actors.

In the research, capabilities of various resources (resource functionalities) and their delivery constraints are made available through the Web-services [21]. These capabilities and constraints are captured by a service profile [8]. A profile describes the functional and non-functional service semantics. The functional service semantics is described in terms of the input and output parameters of the service. The non-functional service semantics is described with respect to service's cost model, availability, competence, and weight.

In the ontology-based service-oriented systems the idea of the alignment of the resource representations against an ontology is implemented using methods of semantic matching [8]. In the present research the Web-services' descriptions in WSDL [20] are aligned against the application ontology. As the result, the WSDL-descriptions are complemented with appropriate SA-WSDL [13] annotations. The alignment is based on semantic matching between the application ontology and the WSDL-descriptions. An approach combining three classes of matching methods—combined, linguistic, and contextual—is used for the matching [16]. The alignment enables Web-services to be interoperable and facilitates Web-service composition.

The problem of relevant knowledge determination in the ontology-based systems is treated as slicing operation. The purpose of this operation is to extract pieces of knowledge from one or more ontologies, which considered as relevant to the user request. The implementation of the operation depends on the ontology representation formalism (see e.g., [3, 18]).

In the present research, the slicing operation is based on the determination in the application ontology knowledge semantically similar to the user request. The operation captures knowledge related to the semantically similar knowledge based on attribute inheritance and constraints processing rules [11]. The captured knowledge is extracted and integrated into the abstract context. In terms of OOCN the abstract context is a set of constraints formalizing (1) the domain knowledge to be instantiated to produce the picture of the current situation and (2) the situation's problems in general form. According to the alignment associations, the set of Web-services is reduced to the contextual ones.

In the part of the abstract context instantiation by the resources, the research follows the idea to integrate these resources based on the service composition [2, 7, 8]. Web-service composition is the act of taking several semantically annotated component services, and bundling them together to meet the needs of a given customer [9].

In the present research the customer needs are formalized in the abstract context. This context specifies an abstract workflow of the required composite Web-service. Contextual Web-services participate in the composition. They communicate in terms of their inputs/outputs to create a service execution sequence. If alternative services available a set of sequences is created. A specific alternative is chosen based on the principles of maximum functionality, maximum access interval, and minimum service weight [17].

The contextual Web-services produce an operational context. The used knowledge representation by means of the OOCN-formalism is compatible with representations supported by constraint solvers. Therefore, the operational context can be

processed as a constraint satisfaction problem using such solvers. Some of these solvers provide mechanisms to search for optimal or efficient solutions, e.g., ILOG [5].

The acting resources communicate on the decision implementation using wire- or wireless Internet-accessible devices.

Table 2 summarizes the technologies supporting the proposed here concept.

The service-oriented architecture of the DSS comprises three groups of services (Fig. 2). The first group is made up of core services responsible for the registration of the Web-services in the service register and producing the real-world model of

**Table 2** Technological framework

Objectives in the theoretical foundations	Techniques	Technology	Result in terms of OOCN
Application ontology building	Ontology building from scratch, integration of existing ontologies	Ontology engineering, ontology management	OOCN with non-instantiated variables
Resource representation	Service-based representation	Semantic Web-services	OOCN with non-instantiated variables
Overcoming Web-services heterogeneity	Alignment of ontology and service descriptions	Semantic matching	OOCN with associative (alignment) constraints
Abstract context creation	Ontology slicing	Ontology management	OOCN with non-instantiated variables
Determination of contextual resources	Ontology slicing	Ontology management	OOCN with associative (alignment) constraints
Operational context producing	Service communications	Web-service composition, context management, constraint satisfaction	OOCN with partly instantiated variables
Generation of alternative action plans	Solving of constraint satisfaction problem	Constraint satisfaction	OOCN with fully instantiated variables, a set of feasible solutions
Choice of a specific plan	Optimization	Constraint programming	An efficient solution
Plan implementation	Service communications	Mobile applications, collective decision making	The efficient solution
Context reusability	Context archiving	Context management	OOCN with partly instantiated variables
Revealing user preferences	Context-based decision archiving	Profiling, decision mining	A set of user constraints
DSS implementation	Service-oriented architecture	Web-services	–

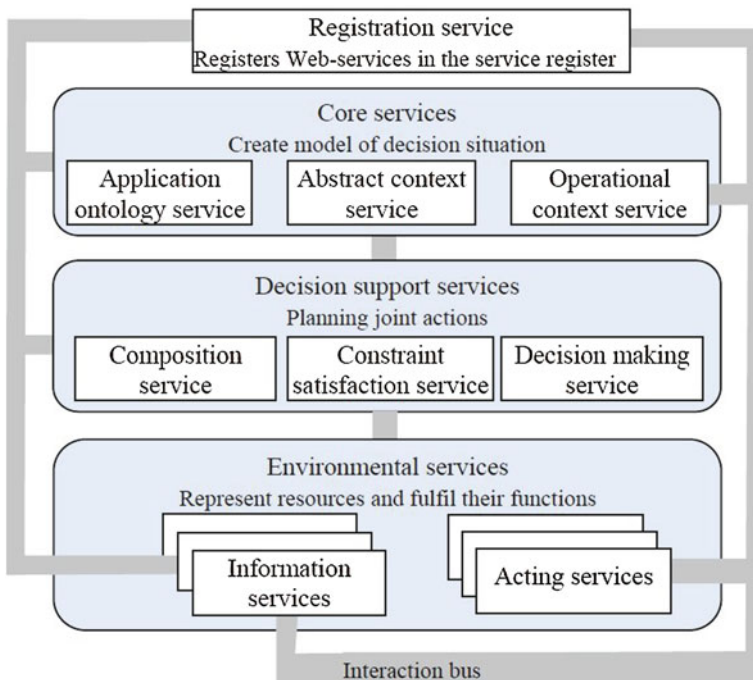
the decision situation, i.e. the creation of the abstract and operational contexts. Services belonging to this group are as follows:

- *registration service* registers the Web-services in the service register;
- *application ontology service* provides access to the application ontology;
- *abstract context service* creates, stores, maintains, and reuses the abstract contexts;
- *operational context service* produces the operational contexts.

Web-services comprising the second group are responsible for the generation of alter-native plans for actions and the selection of an efficient plan. This group contains:

- *composition service* coordinates the service composition process;
- *constraint satisfaction service* generates sets of feasible plans for actions;
- *decision making service* selects efficient plan(s) from the sets of feasible plans, coordinates the communications between the acting resources and the plan implementation.

The third group comprises sets of services responsible for the representation of the resources and implementation of their functions. This group includes:



**Fig. 2** Service-oriented architecture of DSS

- *information services* provide data stored in the resources' profiles and implementation functions of the information resources;
- *acting services* provide data stored in the profiles of the acting resources; represent roles played by people or organizations; communicate on the plan implementation.

### 4 Fire Response

The DSS to support decisions on planning joint actions of independent parties has been developed based on the presented here foundations. The system is used to support decisions in emergency scenarios. Emergencies are a good example of ever-changing situations.

This Section illustrates the system usage in the simulated fire scenario. The purposes of the DSS in the fire situation are to produce a fire response plan for emergency responders, to offer an evacuation plan for potential victims, and to support the communications on decision implementation between the parties involved in the plans. Independent parties participating in the implementation of the fire response plans organize the fire response community (Fig. 3).

Figure 4 presents the abstract context created to model the fire situation. This context is created from the application ontology of the emergency management

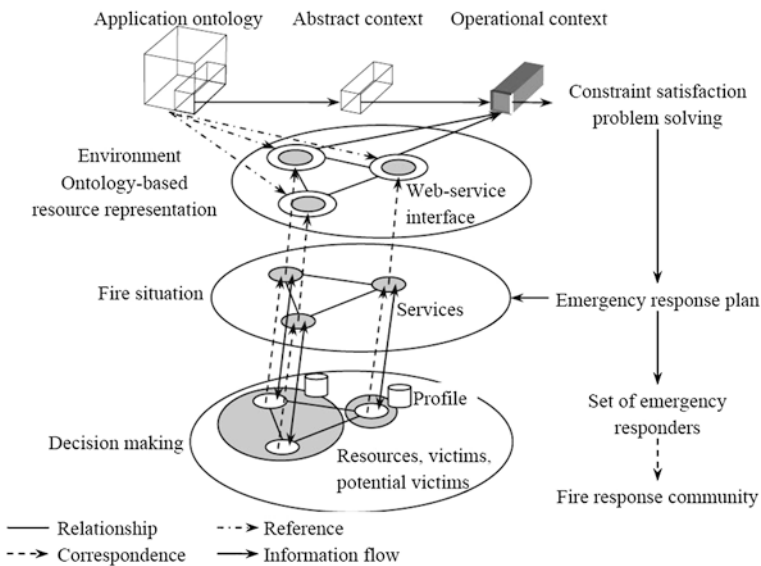


Fig. 3 Organization of fire response community

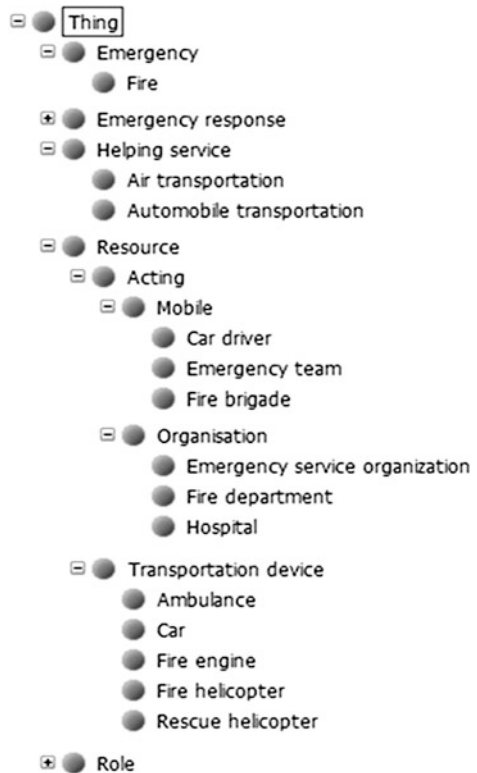
domain. The ontology has 7 taxonomy levels, contains more than 600 classes, 160 class attributes, and 120 constraints of different types. The abstract context has 4 taxonomy levels, contains 19 bottom-level classes to be instantiated, 38 class attributes, and around 30 constraints.

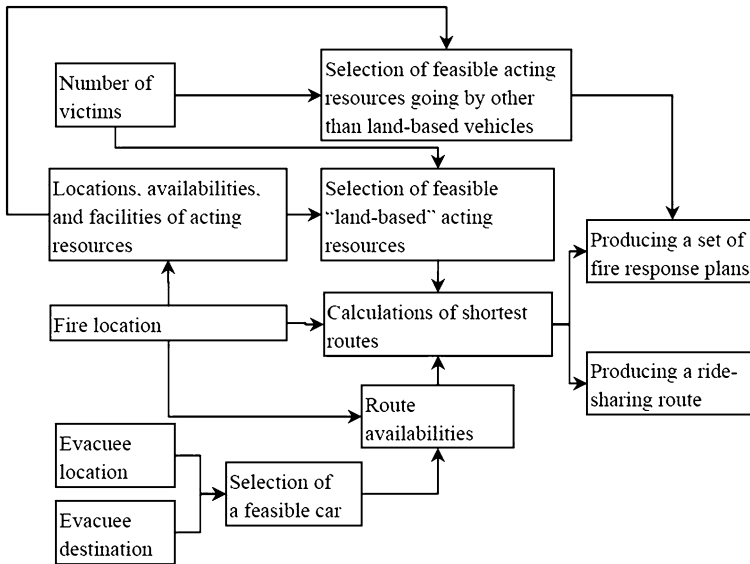
The abstract context represents the following kinds of acting resources needed in the response actions: fire brigades, emergency teams, hospitals, and car drivers. As well, this context represents kinds of transportation devices the mobile acting resources can go by. The context specifies (not shown by the taxonomy) that the emergency teams can go by ambulances and rescue helicopters, the fire brigades can go by fire engines and fire helicopters, and the car drivers go by cars.

Task knowledge is hidden in the class “emergency response”. This class specifies the following problems:

- select feasible hospitals, emergency teams, fire brigades, and car drivers;
- determine feasible transportation routes for ambulances and fire engines depending on the transportation network and traffic situation;
- calculate the shortest routes for transportation of the emergency teams by ambulances, fire brigades by fire engines, and evacuees by cars;

**Fig. 4** Fire situation: abstract context (a piece)





**Fig. 5** Execution sequence of information web-services

- produce a set of feasible response plans for emergency teams, fire brigades, and hospitals;
- produce a set of feasible ridesharing routes.

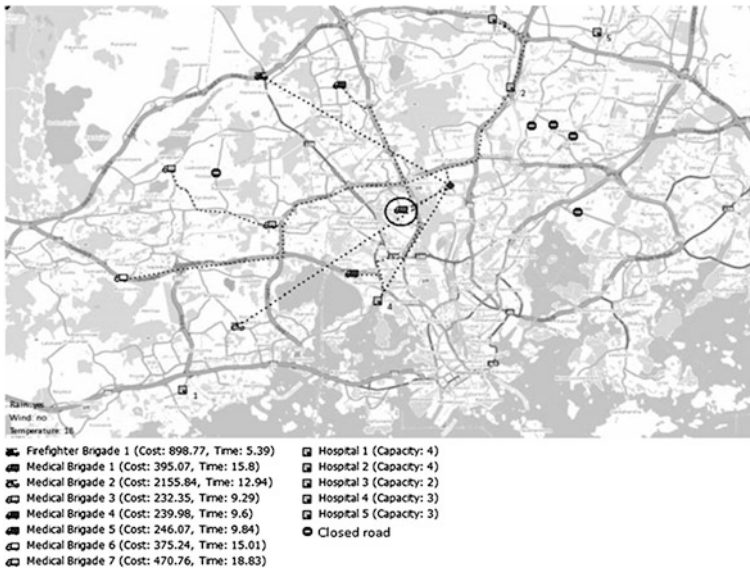
The execution sequence of Web-services composed to instantiate the abstract context and solve the specified problems is shown in Fig. 5.

In the simulated scenario it is supposed that 9 injured people have to be transported to hospitals. In the territory adjacent to the fire place 7 available fire brigades, 8 emergency teams, 5 hospitals having free capacities for 4, 4, 2, 3, and 3 patients are found; 6 fire trucks and 1 fire helicopter are allocated to the fire brigades, 7 ambulances and 1 rescue helicopter are allocated to the emergency teams; 1 fire brigade is calculated to be required to extinguish the fire. The plan for actions designed for the emergency teams supposes that one vehicle can house one injured person.

The set of feasible plans for actions is generated for the criteria of minimal time and cost of transportation of all the victims to hospitals, and minimal number of mobile emergency responders involved in the response actions. The set of plans comprises 4 plans. The efficient plan is selected based on the key indicator of minimal time of victim transportations.

Figure 6 presents the operational context and the efficient plan for actions for the emergency teams, fire brigades, and hospitals, i.e. for professional emergency responders. Such a plan is a set of emergency responders with transportation routes for the mobile responders, required helping services, and schedules for the responders' activities. In Fig. 6 the big dot denotes the fire location; the dotted lines





**Fig. 6** Plan for actions for professional emergency responders

designate the routes proposed for the transportations of the emergency teams and fire brigades. One ambulance (encircled in the figure) and the rescue helicopter go from the fire location to hospitals twice. The estimated time of the operation of transportations of all the victims to hospitals is 1 h. 25 min.

The interaction sequence of the architectural Web-services for producing the efficient plan for the professional emergency responders is shown in Fig. 7.

The efficient plan is presented to the emergency responders for their decision making on the plan implementation. They can access the operational context using any Internet-accessible devices (notebooks, PDAs, mobile phones, etc.). Figure 8 shows part of the plan displayed on the Tablet PC of the leader of an emergency team going by ambulance. He/she can accept or reject this plan (a special option is provided for this). The option of rejection is provided for due to the rapidly changing emergency situations—something may happen between the moment when the plan is selected and the time when the emergency responders receive this plan.

The procedure of making decisions by the professional emergency responders is as follows (Fig. 9). If the plan is approved by all the responders, this plan is supposed to be the plan for actions. Otherwise, either this plan is adjusted (so that the potential participant who refused to act according to the plan does not appear in the adjusted plan) or another set of plans is produced.

The plan adjustment is a redistribution of the actions among emergency responders that are contained in the set of feasible plans. If such a distribution does not lead to a considerable loss of time (particularly, the estimated time of the

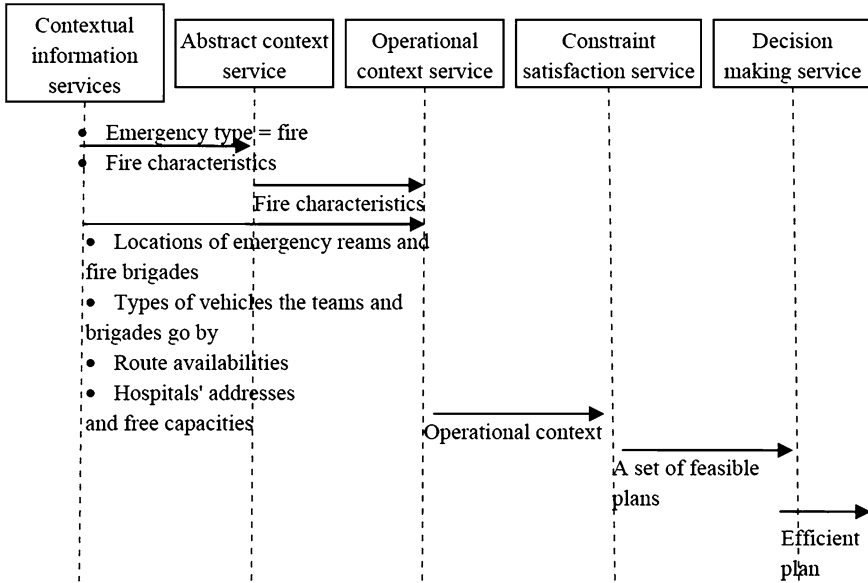


Fig. 7 Service interactions for planning actions of professional emergency responders

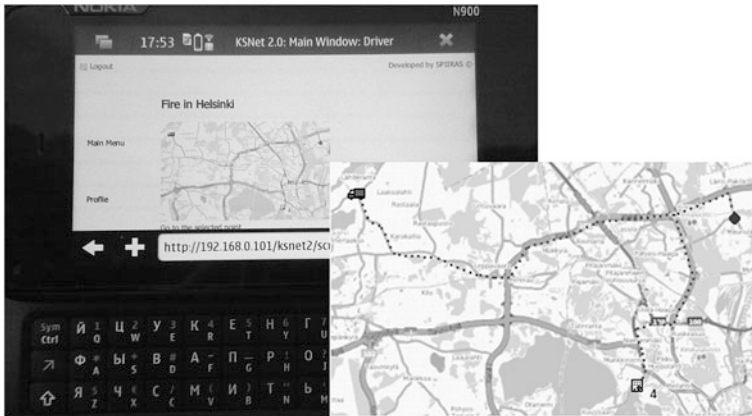
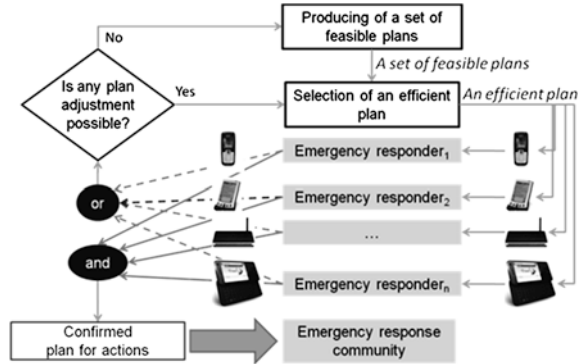


Fig. 8 Plan for actions for an emergency team

transportation of the injured people to hospitals does not exceed “The Golden Hour” [10]) then the adjusted plan is submitted to the renewed set of emergency responders for approval. If a distribution is not possible or leads to loss of response time a new set of plans is produced, from which a new efficient plan is selected and submitted to approval. Simultaneously with planning the joint actions for professional emergency responders, the evacuation activities are planned. Features provided by the ridesharing technology are used to organize the evacuation of potential

**Fig. 9** Decision making by professional emergency responders



victims. Potential victims here are people who have been out of danger so far or have got themselves out of the dangerous area. Success of the evacuation operation heavily depends on the availability of volunteers as drivers of the passing cars.

Persons who need to be evacuated invoke a ridesharing service that is responsible for the evacuation. Clients of this service are supposed to be installed on the Internet-accessible devices of car drivers and other people involved in the fire situation. The persons enter the locations they would like to be conveyed. The ridesharing service determines the persons' locations and searches for cars going to or by the same or close destinations that the persons would like to be. It searches the cars among the vehicles passing the persons' locations. This service reads information about the destinations that the car drivers are going to from the navigators that the drivers use or from the drivers' profiles. The profiles store periodic routes of the drivers and cars' properties as the number of passenger seats, the availabilities of baby car seats, etc. (Fig. 10).

Based on the information about the locations and destinations of the person and the found cars a set of feasible routes for the person transportations is generated. An efficient route is determined based on the criterion of minimum transportation time.

The ridesharing service sends appropriate signals to the drivers included in the ridesharing routes and displays on the drivers' devices the routes each driver is selected for. The points where the driver is expected to pick up the passenger(s) is indicated in the routes. The ways the passengers have to walk to these points are routed for them as well. Besides the routes, the passengers are informed of the model, color, and license plate number of the car intended for their transportation. The persons that cannot be evacuated by passing cars are informed that they can be evacuated by taxi. If they agree, the ridesharing service makes orders for taxi.

Generally speaking, the destinations for the evacuated people do not matter. In actual usage the evacuees can just run the ridesharing service and it will search for passing cars.

Decision making on an evacuation plan is in making agreement between the driver and the evacuee to go according to the scheduled ridesharing route (Fig. 11). In case, when there is no agreement between a driver and an evacuee, another car

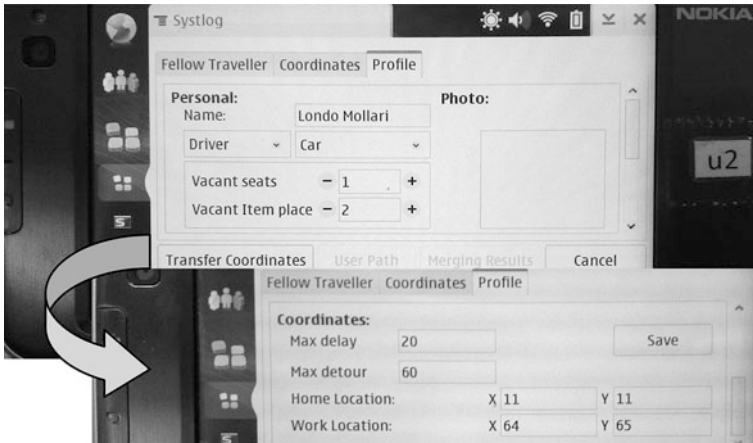


Fig. 10 Driver’s profile

for evacuation of this passenger is sought for. At that, the confirmed routes are not revised.

In the considered scenario results of evacuation using the ridesharing technology are as follows: 26 persons desire to be evacuated from the scene of fire; 22 persons have been driven directly to the destinations by 16 cars whereas for 4 persons no cars have been found. Examples of ways routed for a driver and a passenger are given in Figs. 12 and 13. The encircled car in the figures shows the location where the driver is offered to pick up the passenger.

The interaction sequence of Web-services for producing a ridesharing route is shown in Fig. 14. The boxes “Evacuee” and “Car driver” mean Web-services representing the evacuees and car drivers, respectively.

The Smart-M3 platform [4] was used in the execution of the fire response scenario. Tablet PC Nokia N810 (Maemo4 OS), smart phone N900 (Maemo5 OS),

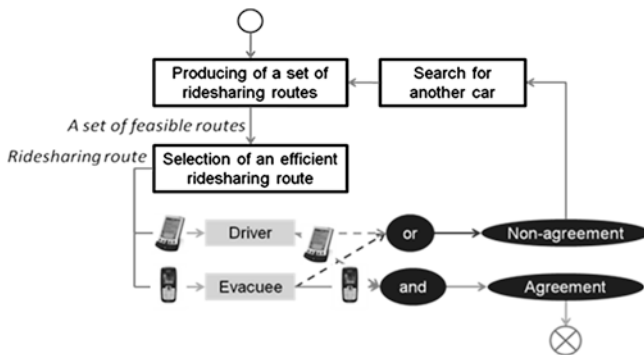


Fig. 11 Decision making by car drivers and evacuees



Fig. 12 Ridesharing route: driver’s view



Fig. 13 Ridesharing route: passenger’s view

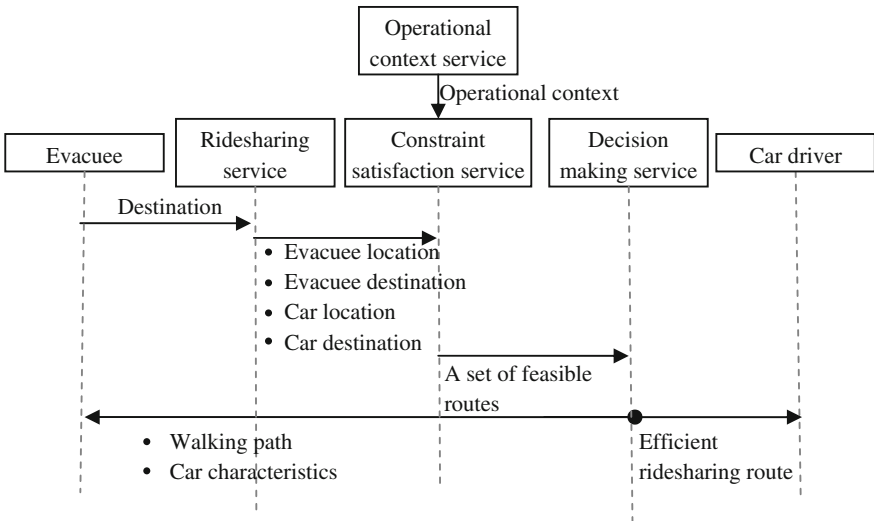


Fig. 14 Service interactions for evacuation planning

and different mobile phones served as the user devices. Personal PCs based on Pentium IV processors and running under Ubuntu 10.04 and Windows XP were used for hosting other services.

## 5 Conclusions

Theoretical and technological foundations for building context-aware DSSs intended for the dynamic environments are proposed. The theoretical foundations offer a concept of context-aware DSS. According to this concept the DSS uses ontology-based context model formalized as a set of constraints. The technological foundations propose a set of technologies enabling Web-based implementation of the theoretical ideas.

The constraint-based context representation allows the DSS to process the problems specified in the context along with other constraints possibly influencing the decision maker's choice as a constraint satisfaction problem. As the result, the DSS provides the decision maker with a workable satisfactory decision. Such a decision is the main principle of the Simon's decision making model. The presented approach exceeds the bounds of this model towards the automatic search for an efficient workable decision and the actors' communications on the decision implementation. It can be concluded that the constraint-based approach enables to minimize the efforts of decision makers to evaluate consequences of the possible alternatives and dependence of the decision makers on the multiple factors influencing their choice.

The main ideas behind the research are illustrated via the support of decisions on planning fire response actions. The response actions are considered comprising two simultaneous scenarios: (1) the response from the professional emergency responders, which is purposed on transportation of the injured people to hospitals and extinguishing the fire; and (2) the response from the volunteers agreeing to evacuate the potential victims. The problem of planning actions of the professional emergency responders is treated as a dynamic logistic problem. The problem of planning volunteers' actions is processed as a ridesharing problem. The constraint satisfaction technology is used for problem solving. This technology enables decision support systems to automatically provide decision makers with efficient 'satisfactory' solutions.

**Acknowledgments** The present research was partly supported by the projects funded through grants 12-07-00298, 13-07-00336, 13-07-12095, 13-07-13159, 14-07-00345, 14-07-00427 (the Russian Foundation for Basic Research), the Project 213 (the research program "Information, control, and intelligent technologies & systems" of the Russian Academy of Sciences (RAS)), the Project 2.2 (the Nano- & Information Technologies Branch of RAS), and grant 074-U01 (the Government of the Russian Federation).

## References

1. Alonso G, Casati F, Kuno HA, Machiraju V (2004) Web services—concepts, architectures and applications. Springer, Heidelberg Berlin
2. Bangemann T, Diedrich C, Riedl M et al (2009) Integration of automation devices in web service supporting systems, preprints of the 30th IFAC workshop on real-time programming and 4th international workshop on real-time software, pp 161–166

3. Chaudhri VK, Lowrance JD, Stickel ME et al (2000) Ontology construction toolkit. Technical note ontology, AI Center. Report, 2000. SRI Project no. 1633
4. Honkola J, Laine H, Brown R, Tyrkko O (2010) Smart-M3 information sharing platform. In: Proceedings of IEEE symposium computers and communications, IEEE computer society, pp 1041–1046. doi:<http://ieeecomputersociety.org/10.1109/ISCC.2010.5546642>
5. IBM ILOG, Web (2011) URL: <http://www-01.ibm.com/software/websphere/ilog/>. Accessed 27 Feb 2014
6. Jammes F, Mensch A, Smit H (2005) Service-oriented device communications using the devices profile for web services, proceedings of the 3rd international workshop on middleware for pervasive and ad-hoc computing (MPAC'05), ACM. <http://doi.acm.org/10.1145/1101480.1101496>
7. Karnouskos S (2011) Realising next-generation web service-driven industrial systems. Int J Adv Manuf Technol. doi:[10.1007/s00170-011-3612-z](https://doi.org/10.1007/s00170-011-3612-z) (Springer)
8. Klusch M (2008) Semantic web service description. In: Helin H, Schuldt H (eds) Intelligent service coordination in the semantic web. Birkhaeuser Verlag, Springer, pp 41–67
9. Klusch M (2008) Semantic web service coordination. In: Helin H, Schuldt H (eds) Intelligent service coordination in the semantic web. Birkhaeuser Verlag, Springer, pp 69–108
10. Lerner EB, Moscati RM (2001) The golden hour: scientific fact or medical “urban legend?”. Acad Emerg Med 8(7):758–760
11. Levashova T (2008) Context model in intelligent decision support systems, decision support. In: Petrovsky A (ed) Proceeding of institute of systems analysis of the Russian academy of sciences, Moscow, URSS 35, pp 33–42 (in Russian)
12. Papazoglou MP, van den Heuvel W-J (2007) Service oriented architectures: approaches, technologies and research issues. VLDB J 16(3):389–415
13. Semantic Annotations for WSDL and XML Schema, W3C Recommendation, Web (2007) URL: <http://www.w3.org/TR/sawSDL/>. Accessed 27 Feb 2014
14. Simon HA (1987) Making management decisions: the role of intuition and emotion. Acad Manage Exec 1:57–64
15. Smirnov A, Pashkin M, Chilov N, Levashova T, Haritatos F (2003) Knowledge source network configuration approach to knowledge logistics. Int J Gen Syst 32(3):251–269
16. Smirnov A, Kashevnik A, Shilov N, Balandin S, Oliver I, Boldyrev S (2010) On-the-Fly ontology matching in smart spaces: a multi-model approach, smart spaces and next generation wired/wireless networking. In: Balandin S, Dunaytsev R, Koucheryavy Y (eds) Proceedings of the 3rd conference ruSMART 2010 and the 10th international conference NEW2AN 2010, St. Petersburg, Russia, 2010, Springer, LNCS 6294, pp 72–83
17. Smirnov A, Levashova T, Shilov N, Kashevnik A (2010) Hybrid technology for self-organization of resources of pervasive environment for operational decision support. Int J Artif Intell Tools 19(2):211–229. doi:<http://dx.doi.org/10.1142/S0218213010000121> (World Scientific Publishing Company)
18. Swartout B, Patil R, Knight K, Russ T (1996) Toward distributed use of large-scale ontologies, tenth knowledge acquisition for knowledge-based systems workshop (KAW'96), Banff, Canada, 1996, Web. URL: [http://ksi.cpsc.ualgary.ca/KAW/KAW96/swartout/Banff\\_96\\_final\\_2.html](http://ksi.cpsc.ualgary.ca/KAW/KAW96/swartout/Banff_96_final_2.html). Accessed 27 Feb 2014
19. Web Services Architecture, W3C Working Group Note, Web (2004) URL: <http://www.w3.org/TR/ws-arch/>. Accessed 27 Feb 2014
20. Web Services Description Language (WSDL), W3C Note, Web (2001) URL: <http://www.w3.org/TR/wsdl>. Accessed 27 Feb 2014
21. Web Services on Devices, Microsoft Corp. (2007) Web, URL: <http://msdn.microsoft.com/en-us/library/bb756908.aspx>. Accessed 27 Feb 2014

# **Part II**

## **Papers**



# Creating Hybrid Simulation Systems Using a Flexible Meta Data Approach

Juergen Rossmann, Michael Schluse and Ralf Waspe

**Abstract** Our goal was to realize a truly hybrid simulation system, which allows the simultaneous use of discreet event simulation and continuous 3D-simulation on a unified database. The key component is an active real-time simulation database, which is an object-oriented, self-reflecting graph database, with a powerful meta-information system. We achieve this by using State Oriented Modeling, which combines the ideas of object-oriented Petri-nets and supervisory control (using discreet event simulation as a control component). The object-oriented Petri-nets are formally described in the State Oriented Modeling Language, which is itself an extension scheme of the simulation database.

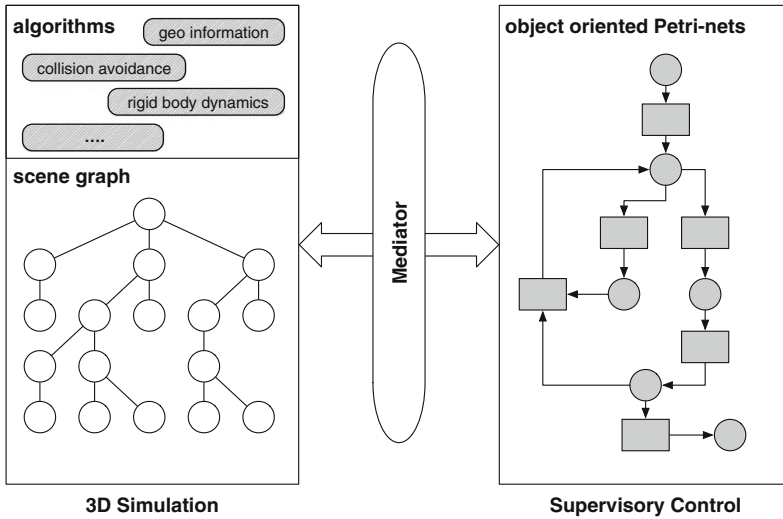
**Keywords** Hybrid simulation · Meta data system · Supervisory control · State oriented modeling language

## 1 Introduction

State Oriented Modeling [15] combines the ideas of supervisory control introduced by [11] and object-oriented Petri-nets (OPN) [3]. It has already been used for a large variety of different applications in the field of simulation (e.g. to simulate robot programs as described in [2]), but also for the real-time control of physical systems using simulation technology as described in [14]. To realize these applications, Supervisory Control provides the methods necessary to link the control algorithms with simulations or physical devices. To implement the controllers, Petri-nets are known to be able to map almost all of the most important state oriented description languages and even modern programming paradigms, to model complex scenarios.

---

J. Rossmann · M. Schluse · R. Waspe (✉)  
Institute for Man-Machine Interaction, RWTH Aachen University, Aachen, Germany  
e-mail: waspe@mimi.rwth-aachen.de



**Fig. 1** A scene graph based simulation system with an exterior supervisory control

Figure 1 illustrates the integration of Supervisory Control concepts, Petri-Nets and 3D simulation used so far. Here the simulation system and the supervisory control are separate entities, working on disjoint sets of data. Communication between these two sets has to be established via a mediator.

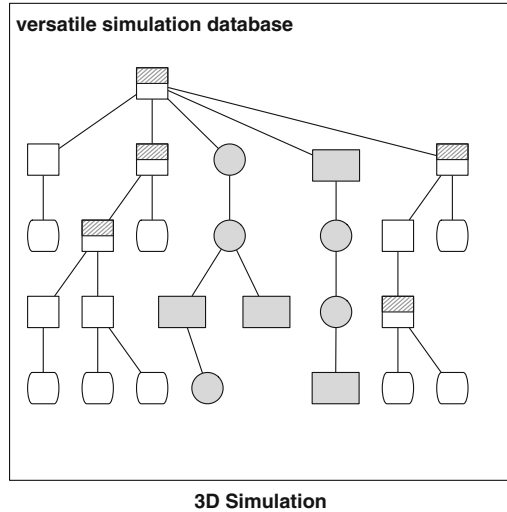
With this concept as a starting point, our goal was to realize a truly hybrid simulation system, which allows the simultaneous use of discrete event simulation and continuous 3D simulation on a unified database. It is then up to the simulation developer how to use these paradigms *in parallel* to realize convincing simulation applications in a wide range of application areas from “classical” simulation applications (driving simulators, virtual production, etc.), to new application areas like user interface design or Virtual Testbeds providing simulation-based development frameworks for complex systems, a key technology in the emerging field of eRobotics [12].

This contribution will detail the progress we have made in making State Oriented Modeling not only an add-on to a 3D simulation system, but incorporating the principles directly into our real time simulation system database.

The key component is an active real-time simulation database, which is an object-oriented, self-reflecting graph database. To reach the integration goal outlined above, the database has to fulfill the following demands:

- The database must support the integration of data (e.g. 3D simulation data) and algorithms (e.g. Petri Nets) in one single—now *active simulation*—database, supporting interface definition and providing means for state oriented as well as event based communication.

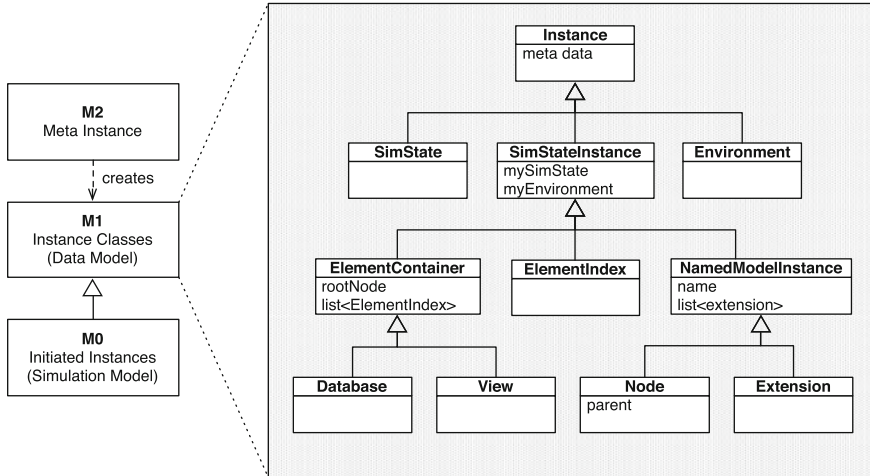
**Fig. 2** A simulation system based on an active database with integrated supervisory control



- The database must be able to flexibly adopt new data schemata (e.g. for the representation of various kinds of OPN) for its internal database, without additional alteration to the core programming.
- For the (real-time) simulation performance it is important how the data can be accessed and manipulated by the simulation algorithms (e.g. to implement even complex controllers using OPN). Ideally, database management itself should be time efficient, thus leaving computing power available to the simulation routines.
- In addition to this, it must be possible to easily add new simulation algorithms or enhance existing methods, while guaranteeing stability and performance of the overall system.
- The database itself must be independent from the type of simulation, to be able to incorporate quasi continuous as well as discrete event simulation paradigms into one single integrated simulation framework as depicted in Fig. 2.

## 2 The Real-Time Simulation Database

To fulfill the requirements mentioned in the introduction, and to eliminate unnecessary dependencies and provide a sustainable basis for various and diverse simulation applications, we developed a new architecture for 3D simulation systems, which is based on a small (micro-) kernel. This kernel is the *Versatile Simulation Database* (VSD), a real-time database containing all the data and algorithms needed for simulation applications. Fully implemented in C++, it provides the central building blocks for data management, meta-information,



**Fig. 3** The meta model hierarchy and the core database class hierarchy

communication, persistence and user interface. The design of the simulation system as shown in on the left of Fig. 3 is inspired by the Object Management Group (OMG) meta model hierarchy [8].

## 2.1 Meta Information

The uppermost layer (labeled M2 in Fig. 3) is the meta-information system. It is essential for the flexibility, as well as the developer and end user friendliness of the database and the simulation system. The meta-information system is the basis for persistence, user interface, parallel and distributed simulation, scripting and communication. It mainly consists of the following classes:

- **MetaTypeVal.** Describes all data types that can be used as values (e.g. int, double, string, simple structs, enumerations, flags).
- **MetaProperty.** Describes a property (see Sect. 2.3) with its getter and setter functions, its data type and a number of additional flags. These flags describe the behavior of the property as exposed to the user (editable, savable, etc.) as well as the properties ability to be used in parallel and distributed simulation.
- **MetaMethod.** Describes a method (member function) of an instance.
- **MetaInstance.** Describes an instance including its class hierarchy. Each non-abstract meta-instance is able to create corresponding instances. Each meta-instance holds a list of the corresponding meta-methods and meta-properties, and furthermore provides a central entry point for executing member functions.

In addition to “build-in” classes, it is also possible to generate meta-instances with the corresponding meta-properties and meta-methods during runtime (for example for object oriented scripting or new data models). Such “run-time meta-instances” are treated in exactly the same way as the build in meta-instances, with no performance overhead in the data management.

## 2.2 Instances

The middle layer (labeled M1 in Fig. 3) describes the data model of the simulation. In order to be able to retain semantic information and integrate data and algorithms into one single database, the VSD data model is an object oriented graph database [6], whose structure is detailed in this section. A simplified class hierarchy of the VSD core is shown on the right of Fig. 3.

All nodes in the graph database, the database itself and even the simulation environment are derived from a single base class called “Instance”. This base class provides mechanisms for inter-instance communication, as well as access to the meta-information system, which allows introspection of class hierarchy, properties and methods (see Sect. 2.1).

The simulation model (labeled M0 in Fig. 3) is an instantiation of the data model.

## 2.3 Properties

Derived from the instance class is the “SimStateInstance”. Besides providing a reference to its simulation state (see Sect. 2.4) it may contain so called “Properties” and encapsulates the access to them. Properties are standardized getter and setter functions that encapsulate the data itself. All data in the simulation system is stored as properties. Properties can encapsulate any single value or value containers (lists, vectors, etc.), whose data types are known to the meta-information system. Properties can also hold references or lists of references. References come in two different varieties, composite aggregation (with reference counting as described in [9]), and shared aggregation. All parent child relations within the database are implemented as composite aggregation references. Shared aggregation references do not change the reference counter of the instance, but are informed if the instance gets deleted.

## 2.4 Database Structure

As shown on the left of Fig. 3 all nodes in the graph database, as well as the database itself are derived from the instance base class.

- **Environment and Simulation States.** The complete simulation is described by an “Environment”, which contains at least one “Simulation State” (SimState).

Simulation states provide mechanisms for copying content from one state to another and are thus the basis for data partitioning in distributed or parallel simulation. For this, a simulation state can keep a list of all transactions, which can then be used to apply all bundled state changes to another simulation state on the same or on other computers. For example, when streaming data from an external database a separate thread with its own simulation state does handle the database interface. When the data has been loaded it will be transferred into the main simulation state.

- **Container and Element Index.** The database graph itself is kept by a “Container” class, a collection of graph nodes. A special container is the “Database” class, which acts as the spanning tree of the database. Other containers can be constructed, offering a different “View” onto the database, by rearranging all nodes or a subset thereof in different order. An example for this is the spatial view, which shows nodes in their spatial arrangements and gets updated when objects are grabbed or moved by other objects.

Furthermore the database offers convenience access to specific instance types via the “ElementIndex” class, which for every meta-instance provides a list of all instances of that type. By this mechanism it is possible to view the graph database in a traditional table based manner without performance restrictions.

- **NamedModelInstances.** This derivation from SimStateInstance provides a name property, as well as a list of extensions.
- **Nodes.** Most commonly used is the “Node” class, which adds a child reference list property to the “NamedModelInstances” class.
- **Extensions.** “Extensions” are used to add data and functionality to a variety of nodes. Extensions can not only be attached to nodes, but also to other extensions.

## 2.5 Active Database

As mentioned above the VSD is not only a static data container but also contains the simulation algorithms itself. The environment, as well as all containers and element indexes actively inform interested listeners about new instance creation or deletion, as well as property modifications. Furthermore each instance sends a signal when one of its properties has been changed. Thus interested parts of the simulation systems can be informed about state changes in the simulation, eliminating the need to continuously poll the database content. With the active messaging system and the availability of element index lists we have minimized the need for data polling and tree traversal. By creating derived classes (like “Actuator”, “Sensor”, “Robot”, etc.) from the Node or Extension base class, the simulation algorithms (actuator control, sensor simulation, robot controller, etc.) itself are integrated into VSD. Simulation algorithms that need a complete overview over the simulation state (like rigid body

simulations) are integrated on the database level but still manage their data on the node and extensions level as illustrated before.

That’s why we call the VSD *active*. To achieve a complete decoupling of the different system components with well defined interfaces (introspectable using the meta-information system), methods are provided for event and state based communication. These methods can be used to let the components exchange information as defined by the algorithm developer or the simulation expert.

### 2.6 3D Simulation

As already stated above, even the 3D simulation capabilities are an extension of the core database. An excerpt of class hierarchy for these classes is shown in Fig. 4.

The data for 3D simulation may be interpreted by a collision detection system, a kinematic animation system, the physics simulation, the renderer or other application specific simulation algorithms. The most important classes are:

- **3D Node.** A node which contains a frame property describing its position and orientation relative to the parent node.
- **Hull Node.** A 3D node that holds a reference to geometry. It also has a reference list of map nodes and an associated material node.
- **Geometries.** This instance holds all data necessary to describe a geometry. Properties include vertex, facet and texture coordinate lists. It can be referenced (shared) from many hulls.

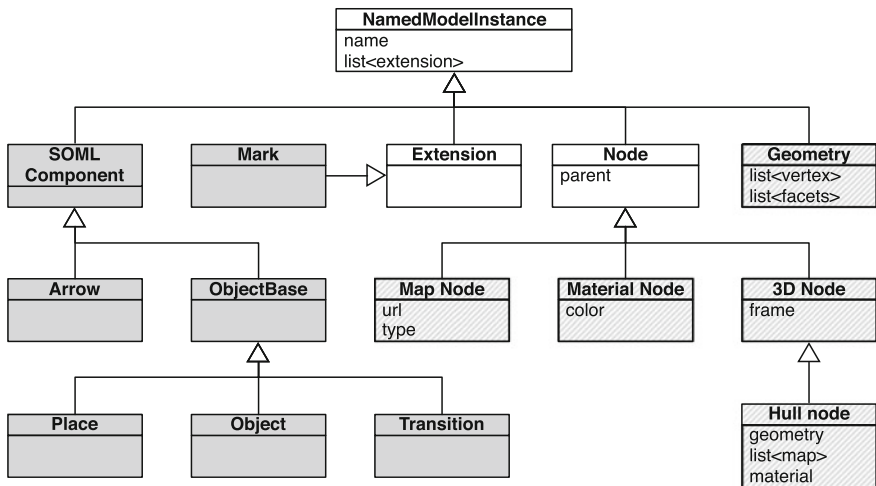


Fig. 4 Add-ons to the core database: components for 3d simulation (light gray) and state oriented modeling (dark gray)

- **Maps and Materials.** A map is a node with an url to a texture. Furthermore the function of the texture is given (for example: color map, normal map or diffuse map). A material node describes physical properties of a geometry, like color or electric conductivity. Maps and material are not only used for rendering. For example it is possible to attach a radar reflection intensity map to a hull node, in order to enable a more realistic radar simulation.

## 2.7 Simulation System

For real-world applications the database must be extended by new data schemes and simulation algorithms like 3D simulation (VSD3D, see Sect. 2.6) or State Oriented Modeling (VSDNet, see Sect. 3). Further functionality like rendering, data processing, file loading, hardware interfaces or simulation scheduling, is handled by plugins, which may also add database enhancements like kinematics, dynamics (detailed in [7]), process simulation or GIS (Geo Information Systems, see [10]).

## 3 State Oriented Modeling Language

In this section we introduce the features of the “State Oriented Modeling language” (SOML++) and show how object oriented Petri-nets are integrated into our simulation system. As mentioned before, the goal was to make State Oriented Modeling and Petri-nets an integral part of the simulation system.

### 3.1 Supervisory Control

This basic concept to integrate Petri-nets with 3D simulation systems is based on the Supervisory Control approach. The idea of Supervisory Control describes a technique to regard a control component (Supervisor) as a discrete-event simulation (DES), with its own state space and transitions caused by events that reflect changes at observed outputs of the controlled system (Plant). Figure 5 shows the basic structure of Supervisory Control: selected, re-fed state transitions  $\sigma$  of the plant generate events, which trigger further state transitions in the supervisor. Based on those state transitions, the supervisor can react, in order to adjust the plant with control commands  $\gamma$ .

With respect to the focus of this contribution, the integration of Petri-Nets and 3D Simulation, Petri Nets are used to implement the control component (Supervisor) as a DES supervising and controlling the 3D simulation (Plant).





**Fig. 5** The basic structure of supervisory control, where one DES (supervisor) commands another DES (plant)

### 3.2 Language Features

SOML++ is a language that describes objects that can contain Petri-nets (see [3]). In an object-oriented fashion these objects can be derived from other objects and can encapsulate data. In the global name-space all the language constructs described below are allowed.

- **Object-classes.** An object-class must be instantiated as an object in order to be used. It is possible to subclass any object-class, regardless whether it has been defined in the SOML++ code or is a build-in class like VSD3D::Node.
- **Objects.** An object can be created from scratch or it can be derived from any object-class. Objects can be constructed with arguments, which will get passed to the constructor function.

An object or object-class may contain any number of further objects or object-classes. Additional language elements and the building blocks of Petri-nets for use within objects or object-classes are:

- **Properties.** A value of all data types, that are known to the meta-information system of the simulation. Properties can also hold references to SOML++ objects.
- **Functions.** A block of code to be executed. Like in C++, a function has any number of arguments and a defined return type. Functions can be called from other functions or transitions anywhere within the SOML++ script.
- **Places.** A place as defined for a traditional Petri-net. Places and transitions are special objects, thus it is possible to define properties or functions within them.
- **Transitions.** Transitions may contain conditions and actions. If the conditions are met a marker may pass the transition and the actions are executed. Both, conditions and actions, are defined by user defined code which may in turn call other functions.
- **Arrows.** Link places to transitions and transitions to places.
- **Start-place.** Declares the object containing this statement to be a token, which moves through the Petri-net as defined by places, arrows and transitions.

This short summary outlines only a few language elements. In addition to this, the Petri-net implementation of State Oriented Modeling provides arrow conditions, different arrow types (normal, inhibitor, communication), Petri-net substitution and invocation hierarchy with arguments and return values, to name only the most important features.

### 3.3 *Interpreting and Executing SOML++ Code*

For the representation of the SOML++ code within the simulation database, a new data schema “VSDNet” as described in Fig. 4 has been developed, providing classes for all SOML++ language elements.

When a piece of SOML++ code is loaded by the simulation system, equivalent VSDNet “SourceComponent”-instances are created for classes, objects, arrows, etc. After that, the source representation in the simulation database is traversed and meta-instances (see Sect. 2.1) are generated for each object-class and each object. Properties are mirrored as meta-properties and functions as meta-methods. Afterwards new instances (see Sect. 2.2) are created from this meta-data and added to the database. Contained places and transitions are sub nodes of the object nodes, arrows are modeled as sub nodes of the originating place and transition nodes.

Since every SOML++ object has now become a “normal” database node, net functions can use all the functionality provided by the database or the meta-information system, respectively. Net functions can interact with the rest of the database by creating new instances (using not only those defined in the SOML++ code, but all classes known to the simulation system), obtain references to other database nodes and call functions of these nodes.

Of course other database nodes can interact in the same way with the net objects, for example signals are sent when a net object is created or a property within this object has changed. There is no additional interface layer between the core database and the Petri-nets—the Petri-nets are an integral part of the simulation database.

### 3.4 *Realizing Supervisory Control*

At this point, all components necessary for realizing Supervisory Control with State Oriented Modeling technique in 3D simulation systems are available. Using the same concepts, Petri-nets can observe any simulation state. State changes cause transitions to “fire” (representing events  $\sigma$ ) which produce control commands  $\gamma$ .

## 4 Applications

The hybrid simulation approach presented in this contribution greatly simplifies the realization of new 3D simulation applications. Nearly all the applications realized so far benefit this approach. They use quasi continuous simulation technology to simulate kinematics, physics, actuators, sensors, various processes, etc. and State Oriented Modeling for supervisor and controller implementation, interfaces, user interaction, and supervisory control of the overall simulation. In this section we will

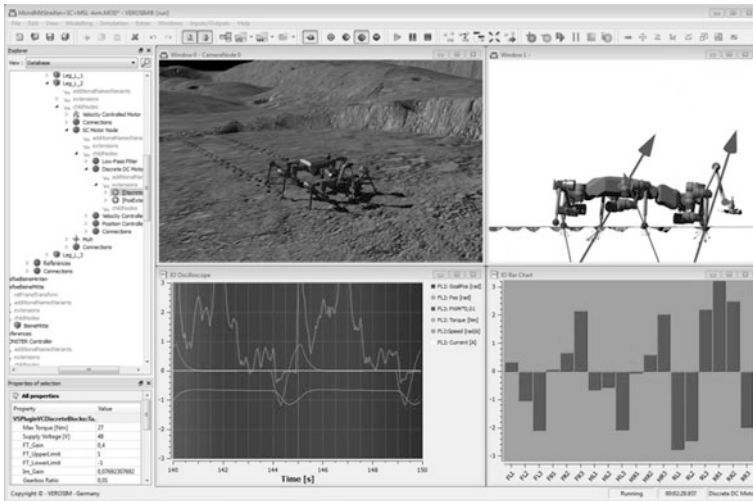


Fig. 6 A virtual testbed. (spaceclimber model—copyright DFKI bremen)

focus on two application areas illustrating the application range of the concept presented above.<sup>1</sup>

### 4.1 Virtual Testbeds

Using the concepts illustrated above we are now able to simulate complex systems with all relevant system components and their interdependencies. The result is a comprehensive development and testing environment based on simulation technology, a Virtual Testbed (see Fig. 6).

The Virtual Testbed concept is a key technology in the emerging field of eRobotics, because Virtual Testbeds can act as a central focal point in multi-disciplinary development projects. For this reason, the first application areas of Virtual Testbeds are in the field of robotics, i.e. for the development of exploration robots, production plants or other complex systems.

### 4.2 Simulation Based Control

The use of supervisory control and state oriented modeling is not confined to the virtual world. The very same concept can also be applied to control real hardware

<sup>1</sup> More examples can be found at <http://www.youtube.com/user/VEROSIMSimulations>, all of which are based on the VSD and most of them use supervisory control.

with the same software. We are using simulation technology to directly control physical systems, which we call “Simulation-Based Control” [13]. This way, the same simulation and algorithms which were prototyped in a Virtual Testbed, control the actual hardware afterwards. An example is the control system of the multi-robot workcell consisting of two redundant 8-axis robots (linear axis plus a 7-axis robot).

For robot control the database of our simulation system is extended with new types of node extensions, able to model and control kinematic chains and kinematic trees. The extension supports rotary and prismatic joints, as well as universal joints and joints directly defined via their Denavit-Hartenberg parameters.

The control concept of the multi robot system is based on the Intelligent Robot Control System (IRCS) structure, developed and introduced in the 1990s by [5]. The IRCS addresses the main aspects of multi robot control by breaking up given tasks into smaller, manageable pieces in a “divide and conquer” fashion, delegating control over several layers of abstraction and responsibility. Figure 7 shows a simplified structure for the robot controller.

Here, the 3D simulation control (named “Multi Robot Control”) acts as a coordinator of the vendor specific robot control units by implementing the simulation-based control concept. To communicate with the physical devices, the Ethernet based Fast-Research-Interface (FRI) is used for the KUKA Light-Weight-Robots [4], while a Profibus-Interface is used for the linear axes. The user interaction via a “User Interface” and the real-time robot coordination “Multi Robot Control” is performed on different computers, both running the same simulation system, though with different configurations on different operating systems (Windows and QNX). Both simulations use the same model, which is kept in sync between the computers by distribution methods provided by the core database.

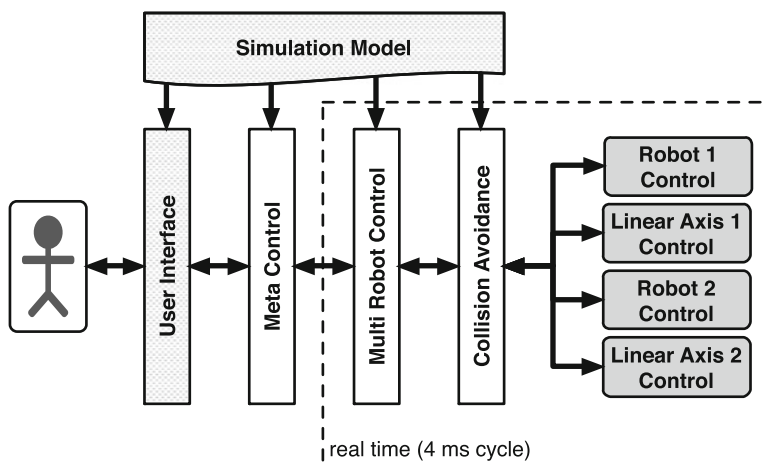


Fig. 7 The intelligent robot control system

The “Meta Control” layer (action generation and distributed planning using algorithms from the field of artificial intelligence), as well as the “Multi Robot Control” layer are Supervisory Controls implemented in SOML++.

## 5 Conclusions and Future Work

In this contribution we presented a new structure for an object oriented graph database for versatile 3D simulation systems. Due to their meta-information management, such systems can adapt to new data schemes even at run-time of the simulation, without the need for further programming. This approach allows us to integrate Petri-net objects as modeled in the State Oriented Modeling language. These Petri-net objects become an integral part of the simulation database and have full access to the 3D simulation data and algorithms, which enables supervisory control of quasi continuous simulation applications using discrete event simulations. The result is a hybrid simulation system which has proven its applicability in large variety of applications, “classical” simulation applications like driving or production simulators, but also new fields of applications like GUI modeling or Virtual Testbeds.

Although performance of the interpreted SOML++ code segments never was a problem so far, we plan to introduce a compiler, which will transfer the generated meta-instances into native C++ code. In addition to this we plan the integration of further Petri-net approaches like hybrid and continuous petri nets [1] to widen the methodical base of the overall concept.

## References

1. Alla H, David R (1998) Continuous and hybrid petri nets. *J Circuits Syst Comput* 08:159–188
2. Baldini F, Bucci G, Vicario E (2005) A tool set for modeling and simulation of robotic workcells. In: 2005 Workshop on techniques, methodologies and tools for performance evaluation of complex systems (FIRB-Perf 2005)
3. Bastide R (1995) Approaches in unifying petri nets and the object-oriented approach. In: Proceedings of the application and theory of petri nets (1995)
4. Bischoff R, Kurth J, Schreiber G, Koeppel R, Albu-Schäffer A, Beyer A, Eiberger O, Haddadin S, Stemmer A, Grunwald G, et al (2010) The kuka-dlr lightweight robot arm—a new reference platform for robotics research and manufacturing. In: International symposium on robotics (ISR2010)
5. Freund E, Rossmann J (1995) Systems approach to robotics and automation. In: Proceedings of 1995 IEEE international conference on robotics and automation
6. Gyssens M, Paredaens J, van den Bussche J, van Gucht D (1994) A graph-oriented object database model. *IEEE Trans Knowl Data Eng* 6:572–586
7. Jung T (2011) Methoden der Mehrkörperdynamiksimulation als Grundlage realitätsnaher Virtueller Welten. Ph.D. thesis, Institut ForMan-Machine Interaction, RWTH Aachen University, Germany

8. Kurtev I, van den Berg K (2005) Mistral: a language for model transformations in the mof meta-modeling architecture. In: European MDA workshops: foundations and applications, MDAFA 2003 and MDAFA 2004, Twente, The Netherlands, June 26–27, 2003 and Linköping, Sweden, June 10–11, 2004. Revised Selected Papers (2005)
9. Levanoni Y, Petrank E (2006) An on-the-fly reference-counting garbage collector for java. *ACM Trans Program Lang Syst* 28:1–69
10. Longley PA, Goodchild MF, Maguire DJ, Rhind DW (2005) *Geographical Information Systems and Science*. Wiley, Hoboken
11. Ramadge P, Wonham W (1984) Supervisory control of a class of discrete event processes. In: *Analysis and optimization of systems*. Springer, Berlin/Heidelberg
12. Rossmann J, Schluse M (2011) Virtual robotic testbeds: a foundation for e-robotics in space, in industry—and in the woods. In: *Developments in e-systems engineering (DeSE)*
13. Rossmann J, Schluse M, Schlette C, Waspe R (2012) Control by 3d simulation—a new robotics approach to control design in automation. In: *Intelligent Robotics and Applications*. Springer, Berlin, Heidelberg
14. Rossmann J, Schluse M, Waspe R (2008) 3D discrete event systems: an efficient way to model and supervise dynamic behavior in virtual environments. In: *ASME 2008 international design engineering technical conferences and computers and information in engineering conference (IDETC/CIE2008)*, vol 2008, pp 1503–1511 (ASME)
15. Schluse M (2002) *Zustandsorientierte Modellierung in Virtueller Realitt und Kollisionsvermeidung*. Ph.D. thesis, Institute of Robotics Research, University of Dortmund, Germany

# Simulation of Real-Time Multiprocessor Scheduling Using DES

Maxime Chéramy, Anne-Marie Déplanche  
and Pierre-Emmanuel Hladik

**Abstract** The evaluation of the numerous real-time scheduling algorithms is difficult without a real and complex implementation. Simulation allows to study the schedulers with more flexibility. This paper presents a simulation tool that uses a process-based discrete-event simulation engine. Compared to the other scheduling simulators, it is able to take into account the impact of the caches through statistical models and direct overheads such as context switches and scheduling decisions. The last Section shows how this tool can be used on concrete examples.

**Keywords** Real-time · Scheduling · Simulation · DES · Multiprocessor · Overheads · Cache

## 1 Introduction

The study of real-time scheduling has regained interest this last decade with the continuous introduction of multiprocessor architectures. Multiple approaches have been used to handle those architectures [10]. A first approach, called partitioning, consists of splitting the task set into subsets. Each of these subsets is allocated to a unique processor on which a uniprocessor scheduler is then run. In contrast, a second approach, called global scheduling, allows tasks to migrate from processor to processor. In that case, there is a single queue of ready tasks and a single

---

M. Chéramy (✉) · P.-E. Hladik  
LAAS, CNRS, 7 avenue du colonel Roche, 31400 Toulouse, France  
e-mail: maxime.cheramy@laas.fr

M. Chéramy · P.-E. Hladik  
Univ de Toulouse, INSA, LAAS, 31400 Toulouse, France

A.-M. Déplanche  
IRCCyN-Institut de Recherche en Communications et Cybernétique de Nantes, UMR CNRS 6597, LUNAM Université-Université de Nantes, 1 rue de la Noe, BP 92101, 44321 Nantes cedex 3, France

scheduler for all the processors. Finally, as a compromise that aims to alleviate limitations of partitioned (limited achievable processor utilization) and global (non-negligible overheads) algorithms, hybrid policies such as semi-partitioned and clustered scheduling have been proposed more recently [4].

By far the greatest focus on multiprocessor real-time scheduling has been put on algorithmic and theoretical issues. Indeed, for the various scheduling policies, a lot of attention has been paid to define analytical schedulability tests. However, those results rely on general and simple models of the considered software and hardware architectures quite far away from the practical ones. Such research must now address implementation concerns as well. Actually, multiprocessor architectures bring more complexity with shared caches and memory, new communication buses, inter-processor interrupts, etc. They also raise new implementation issues at the operating system level: which core should run the scheduler? what data should be locked? etc.

Thus, new scheduling policies that try to take benefits from the specificities of the hardware architecture (such as the caches) must be designed and tools for studying them must be made available. One way for this is to use a cycle-accurate simulator or even a real multiprocessor platform, and to execute real tasks. In that case, the results are very accurate, however it requires developing the scheduler in a low-level language and integrating it into an operating system. This work can potentially take a lot of time. Furthermore, the generation of various and realistic tasks for a massive evaluation is laborious.

In consequence, it is preferable to use an “intermediate-grained” simulator able to simulate with a certain level of accuracy the behavior of those (hardware and software) elements that act upon the performances of the system. Such a simulator allows fast prototyping and does not require a real implementation of the tasks nor the operating system. Moreover, extensive experiments can be easily conducted and various metrics be made available for analysis. Its intrinsic drawback is that it will never reflect exactly how a scheduler behaves in details on a real system but it should be enough to give good insights on general tendencies.

Our contribution is a simulation tool, called SimSo (“Simulation of Multiprocessor Scheduling with Overheads”), that is designed to be easy to use and able to take into account the specificities of the system, starting with LRU caches, context-save/load overheads and scheduling overheads. SimSo<sup>1</sup> is an open source tool, actively developed, designed to facilitate the study of the behavior of schedulers for a class of task systems and a given hardware architecture. For that, we propose to extend the Liu and Layland model [19] to bring enough information to characterize how the tasks access the memory. This allows us to use statistical models to calculate the cache miss rates and to deduce job execution times. Moreover, our simulator has been conceived as flexible as possible to be able to integrate other task and architecture models.

This paper is organized as follows. First, we explain our motivation in Sect. 2. The main principles of real-time scheduling are explained in Sect. 3. Then we describe how the simulator has been implemented in Sect. 4. Section 5 presents the scheduler

---

<sup>1</sup> Available at <http://homepages.laas.fr/mcheramy/simso/>.



component and Sect. 6 deals with the integration of the hardware models. We present the simulation software in Sect. 7 and we compare it to the existing similar tools in Sect. 8. To conclude, we summarize our contribution and present our future works in Sect. 9.

## 2 Motivation

Most of the multiprocessor real-time scheduling strategies have been designed without taking into account the presence of caches and their effects on the system behavior. Though, interferences on the cache of preempted and preempting tasks allocated to the same processor may cause additional delays [23]. In the same way, when a cache is shared by multiple processors, the execution of a task can have a significant impact on another task running on another processor. Furthermore, scheduling overheads and context switch overheads are often regarded as negligible. However, on a multiprocessor system, schedulers tend to generate more preemptions, more migrations and even more rescheduling points in order to achieve a high utilization of the processors [11].

Significant research effort has been focused on the problem of real-time multiprocessor scheduling since the late 1990s, in particular in the area of global scheduling. It led to a number of optimal algorithms (PFair and its variants PD and PD2, ERFair, SA, LLREF, LRE-TL, BF, RUN, U-EDF, etc.) that are very attractive because theoretically able to correctly schedule all feasible task sets without processing capacity unused [10]. However their practical use can be problematic due to the potentially excessive overheads they cause by frequent scheduling decisions, preemptions and migrations. Therefore, being able to take them into account helps in the predictability analysis of such real-time systems for which the first requirement is to meet time constraints. Moreover, reducing the overall execution time of the tasks can also bring significant benefits (for instance, better response times or less power consumption). Following these observations, recent research has emerged and new scheduling algorithms appeared which aim to reduce the overheads by bounding the amount of preemptions [4, 24]. Also, a few studies have shown that avoiding co-scheduling tasks that heavily use a shared cache can reduce the overall execution time [1, 14]. Finally, other researches focus on cache space isolation techniques to avoid cache contention on shared caches [5, 15].

Our primary objective is the comparison of those numerous scheduling policies and their associated variants. Currently, the only way to compare them is by far to try to put in relation the properties exhibited by their authors: computational complexity, number of scheduling points, utilization bound, number of task preemptions, number of task migrations. Such a task is quite intractable since evaluations have been made under separate conditions. Instead our intention is to make available a framework allowing to study as precisely as possible the performance of a scheduler and to establish relevant comparisons between different scheduling policies based on the same benchmarks. For instance, given a system correctly schedulable with multiple scheduling policies, we would like to pick the one that should be the most efficient (less overhead). For that, we would aim to identify general trends for classes

of tasks and hardware architectures. A typical result could be: scheduler A is better than B in most cases, except when the shared cache is too small given the characteristics of the tasks. We expect from these results to help the real-time community to better understand the cache effects on scheduling, and bring new ideas that could help to conceive schedulers which take benefits from the caches.

### 3 Context

In this part, we briefly present the context of real-time multiprocessor scheduling and its relevant models in order to facilitate the understanding of the following. This also precises a few assumptions made for the simulation.

A real-time application is composed of tasks, i.e. programs, to be run on a hardware architecture made of a limited number of processors. Real-time means that the computing of tasks has to meet time constraints (typically release times and deadlines). The scheduler is a software system component whose purpose is to decide at what time and on which processors tasks should execute.

*Tasks* The model most commonly used to describe the tasks is the Liu and Layland one [19]. In this model, a great abstraction is made since a task is simply viewed as a computation time. This means that its functional behavior is ignored as it will be discussed in Sect. 6. In our simulation, when the caches are taken into consideration, some extra parameters are necessary as explained in Sect. 6.2.

A task can be respectively periodic, sporadic, or aperiodic depending on its inter-activation delay, respectively constant, minimum, or unknown. A task activation gives rise to the release of a job (an instance of the task) that must complete before a given deadline date.

The tasks neither share memory nor communicate between each other but precedence relations between tasks may be specified so that the activation of an aperiodic task follows the end of another task.

*Processors* We consider symmetric multiprocessing hardware architectures (SMP) which are the most common multiprocessor design nowadays. In such architecture, the processors are identical and share a single main memory. Private and/or shared caches are associated to them in a hierarchical way. The modeling of the cache hierarchy as well as their time access costs are given in Sect. 6.2. Note that we focus our work on architectures with less than a few dozen processors. Thus, Network On Chip (NoC) architectures, which present an interest for many-core systems, are not considered.

*Scheduler* Among the various scheduling strategies, one distinguishes time- and event-triggered ones depending on the conditions in which the scheduler is invoked: a rescheduling has to be made either at specified instants, or when a job completes or a new one is released. In addition, the scheduler may be preemptive and decide to interrupt the execution of a job and to resume it later. In the same way, schedulers may allow tasks and their jobs to partly or freely migrate and execute on multiple processors.

## 4 Implementation

### 4.1 Discrete-Event Simulation

The core of our simulator is implemented using SimPy [27], a process-based discrete-event simulation library for Python. The advantage of a discrete-event simulation over a fixed-step one is that it is possible to handle short durations (such as a context-switch overhead) as well as long durations (such as a job execution) with the same computational cost. We have chosen SimPy because it can be easily embedded as part of a software, it is well-documented and easy to use.

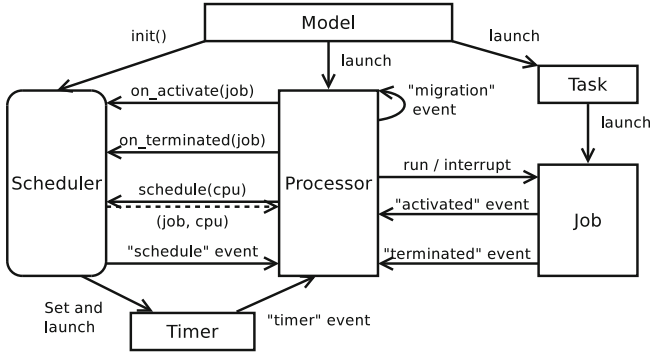
According to SimPy's vocabulary, a *Process* is an entity that can wait for a signal, a condition or a certain amount of time. When it is not waiting, a *Process* can execute code, send signals or wake up other processes. This *Process* state is called "active", opposed to "passive". A *Process* is activated by another *Process* or by the simulation main class itself.

The simulation unit is the processor cycle to allow a great precision. However, for user convenience, the attributes of the tasks, such as the period or the deadline, are defined in milliseconds (floating-point numbers) and converted in cycles using a parameter named *cycles\_per\_ms*.

### 4.2 Architecture

The main classes and their mutual interactions are represented in Fig. 1 and described below:

- *Model* is the simulation entry point, it will instantiate and launch the processors and the tasks as active *Processes*. It will also call the *init* method of the scheduler so that it can initialize its data structures and launch timers if needed.
- A *Task* handles the activations of its jobs. The activations are either periodic or triggered by an other task (aperiodic). Depending on a property of the task, the jobs that exceed their deadline can be aborted.
- A *Job* simulates, from a time-related aspect only, the execution of the task code. Its progression is computed by the execution time model (see Sect. 6). A signal is sent to its running processor when it is ready and when its execution is finished.
- A *Processor* is the central part and simulates the behavior of the operating system running on a physical processor. There is one *Processor* for each physical processor. It controls the state of the jobs (running or waiting) in accordance with the scheduler decisions. It also deals with the events: activation or end of a job, timer timeout, schedule request, etc. The attribute "running" of a processor points to the job that is running (if any). Figure 2 provides a very simplified diagram representing what a processor does. Similarly, as a real system, some actions can induce overheads (e.g. context switch or scheduling decision) and only affect the concerned processor.



**Fig. 1** Interactions between main class instances. *Processor*, *Task*, *Job* and *Timer* are *Process* objects and can have multiple instances

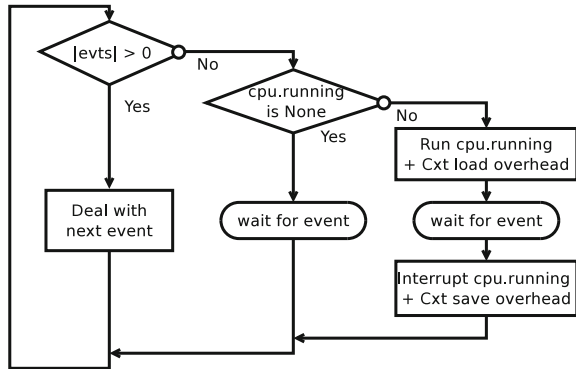
- A *Timer* allows the execution of a method after a delay, periodically or not. On a real system, this method would run on a physical processor, thereby inducing a context switch overhead if a job was running on the same processor. This behavior is reproduced by sending a “timer” event to the processor.
- The *Scheduler* is described in Sect. 5. Unlike the previous elements, the scheduler is not a *Process* object, all its methods except the *init* method are called by the *Processor* objects.

## 5 Scheduler Component

### 5.1 Scheduler Interface

In order to implement a scheduler, the user has to develop a class that inherits from the abstract *Scheduler* class. The scheduler interface is partly inspired by what can be found on real operating systems such as Linux but kept as simple as possible.

**Fig. 2** Simplified execution workflow of a *Processor*. The “terminated” event is a particular event that will not cause a context save overhead



This interface allowed us to develop partitioned, global and hybrid schedulers. The scheduler interface is shown in Fig. 1.

When the simulation is started, the *init* method is called. It is then possible to initialize data structures and set timers if required. When the scheduler needs to make a scheduling decision, it sends a “schedule” event to the processor that will execute the schedule method. This event is sent as a consequence of a job activation, termination or through a timer. This lets the possibility to write schedulers that are either time-driven, event-driven or both. The processor is in charge of applying the scheduling decision (which includes an inter-processor interrupt if needed).

## 5.2 Handling Various Kinds of Scheduling

In order to deal with multiprocessor scheduling, various strategies are possible: a global scheduler for all the processors, a scheduler for each processor, or even intermediate solutions. The support of any kind of scheduler is done at a user level.

Take a partitioned scheduling as an illustration of this, we define a “virtual” scheduler that will be instantiated by the simulation and called by the processors. This scheduler will then instantiate one uniprocessor scheduler for each processor and allocate each task to one scheduler. The links between the processors and the schedulers as well as the links between the tasks and the schedulers are saved. Thus, when a processor calls a method of the “virtual” scheduler, the latter retrieves the concerned scheduler and forwards the method call to it.

By generalizing this example by allocating one scheduler to any number of processors and by allowing a task to migrate from one scheduler to another, we see that any kind of scheduling is feasible. Thus, this approach has the advantage of being very flexible. Moreover, we provide a few examples in the user guide.

## 5.3 Lock

For global or hybrid strategies, some scheduler variables (such as the list of ready tasks) are shared between the processors. As a consequence, a protection mechanism can be required to avoid inconsistencies. Such protections form a bottleneck which induces extra overheads.

A mechanism of lock is provided by the simulator in order to reproduce these overheads. This lock is intended to prevent to run the scheduler at the same simulation time on two or more different processors. The developer of a scheduler can decide to deactivate the lock by overriding the *get\_lock* method.

## 5.4 Example

In this section, we present what a user could develop to simulate a global EDF. The source code is in Python.

A detailed explanation of this example is available in the documentation of the tool. Instead we would like to draw the reader's attention on the small number of lines required. An actual implementation of this policy in an operating system would require hundreds of lines.

```

from core import Scheduler

class G_EDF(Scheduler):
    """Global Earliest Deadline First"""
    def init(self):
        self.ready_list = []

    def on_activate(self, job):
        self.ready_list.append(job)
        # Send a "schedule" event to the processor.
        job.cpu.resched()

    def on_terminated(self, job):
        # Send a "schedule" event to the processor.
        job.cpu.resched()

    def schedule(self, cpu):
        decision = None # No change.

        if len(self.ready_list) > 0:
            # Get a free processor or the processor running the job with
            # the latest deadline.
            key = lambda x: (1 if not x.running else 0,
                            x.running.absolute_deadline if x.running else 0)
            cpu_min = max(self.processors, key=key)

            # Get the job with the highest priority within the ready list.
            job = min(self.ready_list, key=lambda x: x.absolute_deadline)
            # If the selected job has a higher priority than the one
            # running on the selected cpu:
            if (cpu_min.running is None or
                cpu_min.running.absolute_deadline > job.absolute_deadline):
                self.ready_list.remove(job)
                if cpu_min.running:
                    self.ready_list.append(cpu_min.running)
                # Schedule job on cpu_min.
                decision = (job, cpu_min)

        return decision

```

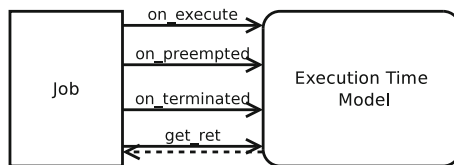


Fig. 3 Interface of any execution time model

## 6 Job Computation Time

### 6.1 A Generic Approach

Generally, scheduling simulation tools consider only the worst-case execution time (WCET) for the execution time of the jobs. Depending on the tool, the user may also have the possibility to configure the simulator to use the average-case execution time (ACET) or a random duration between the best-case execution time (BCET) and the WCET.

One of our objectives is to take into consideration the impact of the memory accesses on the computation time in order to be as accurate as possible. A significant difference with the classical approach is that the total execution time of a job can only be known when it finishes. Indeed, the execution time depends on the scheduling decisions (which tasks were executing on the other processors, was it preempted, etc.).

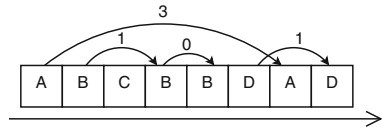
The components needed to compute the execution time are purposely isolated from the rest of the simulator and implement a generic interface to interact with the simulator. As shown in Fig. 3, the execution time model receives an event when the state of a job is changed. The job uses the *get\_ret* method to get a lower bound of its remaining execution time. While this duration is strictly positive, the job is not finished.

For example, a computation time model for an execution based on the WCET is trivial. The *get\_ret* method simply returns the WCET minus the duration already spent to run the job. The remaining methods of the interface have nothing to do because that duration is given and kept up-to-date by the job itself.

This design is sufficiently generic to easily swap the models used to compute the execution time of the jobs. Hence, alternative models could be developed to simulate a different hardware or to adjust the accuracy of the results.

### 6.2 Modeling Memory Behaviors

In this section, we briefly present how the impact of the caches is implemented as an execution time model as explained above.



**Fig. 4** Memory accesses sequence. A, B, C and D are cache lines and numbers indicate the stack distances

*Memory Behavior of a Task.* In order to characterize the memory behavior of the tasks, we extended the model of Liu-Layland with additional information. For each task  $\tau$ , the user must provide:

- Number of instructions: the average number of instructions executed by a job of  $\tau$ .
- Base CPI: the average number of cycles required to execute an instruction without considering the memory access penalties ( $base\_cpi_\tau$ ).
- Memory access rate:  $mix_\tau$  is defined as the proportion of instructions that access the memory among all.
- Stack distance profile (SDP): the distribution of the stack distances for all the memory accesses of a task  $\tau$  is the stack distance profile ( $sdp_\tau$ ), where a stack distance is by definition the number of unique cache lines accessed between two consecutive accesses to a same line [22]. An illustration of this distance is provided by Fig. 4. Such metric can be captured for both fully-associative and N-way caches [2, 9].

These information can be automatically generated or retrieved from a real application. The number of instructions, the memory access rate and the stack distance profile can be generated using various tools [2, 12, 18]. The base CPI requires a cycle accurate simulator. It is the computation time in cycles divided by the number of instructions.

*Cache Hierarchy* We consider hierarchical cache architectures. A list of caches (e.g. [L1, L2, L3]) can be associated to each processor. Caches can be shared between several processors while it respects the inclusive<sup>2</sup> property.

A cache is defined by a name, its associativity, its size, and the time needed to reach it (in cycles).

For now, only data caches with Least Recently Used (LRU) as replacement policy are considered, the generalization to instruction caches is left for future work. Hence, a few modifications in the cache description are likely to occur in order to make the distinctions between instruction caches, data caches and unified caches.

*Cache Models* Depending on which tasks are running in concurrency and the initial state of the caches, the *execution speed* of the jobs varies.

<sup>2</sup> Caches are inclusive if any data contained in a level of cache is contained on the upper level.



The goal of the cache models is to determine, on a given time interval, the average number of cycles per instructions (CPI) of a job, taking into consideration the impact of the various tasks on the caches. Using the CPI, it is then possible to determine the number of instructions executed by a job during that interval.

The duration returned by the *get\_ret* method (see Fig. 3) is simply the time required to execute the remaining number of instructions if the job was running alone on the system without any interruption.

Cache sharing induces two kinds of extra cache misses:

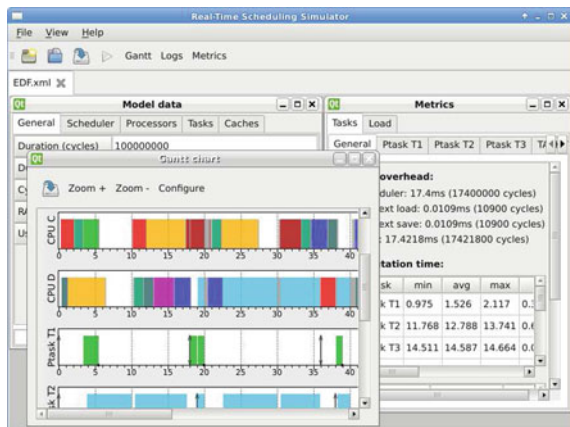
- Following a preemption: a job may have lost its cache affinity when another job is running on the same processor. Some of the evicted lines should then be reloaded.
- Shared between multiple processors: two or more tasks that are simultaneously running on different processors with a shared cache, do not tend to share this cache equally.

For the first case, we have taken the simplifying assumption that the cache filling follows an exponential distribution and we use the SDP of the task to determine the number of lines that should be reloaded. However, other models exist [20] and we are also currently working on better estimations of the cache loading using Markov chains.

For the second case, we used the FOA model [9] that has fast running times and gives reasonable results according to the authors. Obviously, other models [2, 9, 13] could be implemented as well.

The state of the caches (the number of lines for each task) is kept up-to-date at each change in the system (start, interruption or end of a job).

**Fig. 5** User interface of the SimSo showing the simulation of a global EDF



## 7 Simulation Tool

### 7.1 Features

*Open Source* The source code, the documentation and the examples are freely available at <http://homepages.laas.fr/mcheramy/simso/>.

*Configuration* The user interface of the simulator (Fig. 5) provides a straightforward graphical interface to load the scheduler class and to define the tasks, the processors, the caches and their hierarchy, and the various parameters for the simulation. The resulting configuration is saved into an XML file.

*Scriptable* The simulation can be run without the graphical user interface in order to automate a simulation campaign.

*Output* When the configuration is completed and checked, the user can launch the simulation and obtain a Gantt chart representing the result of the simulation or a textual equivalent representation. Some metrics are provided to the user such as the number of preemptions, migrations, the time spent in the scheduler, the computation time of each job, etc.

*Speed and Limitations* The simulation runs very fast, as an example, simulating a global EDF with 4 processors, 10 tasks, 2 levels of caches for a duration of 100 ms ( $10^8$  cycles) takes less than one second on an Intel Core i5. There are no technical limitations on the number of processors but the cache models implemented in the present version have not been validated by their authors for a large number of processors.

### 7.2 Use as an Educational Tool

The user interface of the simulator has been designed keeping in mind it could also be used for an educational purpose. This is the main reason why all the inputs can be set through a graphical user interface and the results displayed in a Gantt chart.

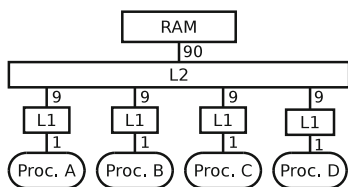
As shown in the previous example, it allows a fast prototyping of schedulers in Python. This language is easy to learn and yet very powerful and effective [25].

Regarding the simulation, it is also possible to consider WCETs as effective task computation times. Thus, the input is simpler and more usual.

This tool is already used by Master's students in a real-time systems course at INSA Toulouse. From an applicative real-time project, students have to model its various tasks and then use the simulator to understand how the processors are shared between the tasks using a fixed priority scheduler.

**Table 1** List of tasks (total utilization is 82.5 %)

	T1	T2	T3	T4	T5	T6	T7	T8
Period (ms)	20	20	15	15	10	10	10	10
WCET (ms)	11	9	7	5	2	6	4	3

**Fig. 6** Simulated hardware architecture. Numbers represent the access time in cycles ( $10^{-9}$  s)

### 7.3 Application Example

As a reminder, our primary goal is the study of scheduling policies. In the following, we present through a simple case study, how our tool could be used to better understand the real behavior of a system.

*Problem Description* In this example, we would like to compare the scheduling of a system using the Earliest Deadline First algorithm with a global (G-EDF) and a partitioned (P-EDF) strategy. G-EDF is a generalization of EDF for multiprocessor that uses a single ready task queue whereas P-EDF starts with a definitive allocation of the tasks on the processors and then runs multiple uniprocessor EDF schedulers for handling each processor.

*Input* The eight considered tasks are all periodic<sup>3</sup> and synchronous with the start of the simulation. Their SDP was taken from the MiBench benchmark [16], the first five tasks are making more accesses to the memory than the last three (according to their value of *mix*).

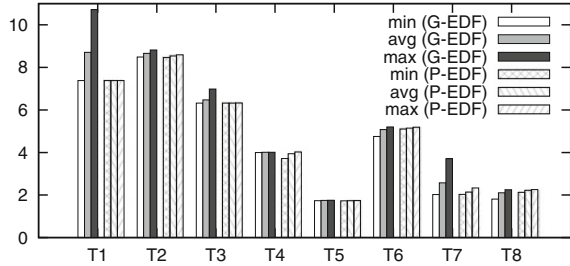
For the task partitioning phase (P-EDF only), a WCET for each task is mandatory. WCET values were chosen by bounding with a safety gap the experimental times given by the simulation. Table 1 synthesizes the period and WCET values of the tasks. Task partitioning was done using the First Fit algorithm. The result of the partitioning is: {T1, T2}, {T3, T4, T5}, {T6, T7}, {T8}.

The simulated hardware architecture, including four processors and a cache hierarchy, is summarized in Fig. 6. Each L1-cache is an LRU fully-associative cache of 2KiB (32 lines of 64 bytes). The L2 cache is an LRU fully-associative cache of 16KiB (256 lines of 64 bytes). The second level of cache is relatively small when compared to what can usually be found on modern architectures. This choice

<sup>3</sup> A periodic task releases a job every *period* time units.

**Table 2** Load for both schedulers

	Total load (%)	Payload (%)	System load (%)
G-EDF	72.2	68.0	4.2
P-EDF	68.3	65.1	3.2

**Fig. 7** Effective task computation times

is justified by the small memory footprint of the selected benchmarks and the will to show the impact of cache contention through this example.

The scheduling overhead is set to 0.1 ms (100.000 cycles) and a context save or load overhead to 0.0001 ms (100 cycles).

The duration of the simulation is 10 s.

*Observations* Table 2 shows the load using both strategies. The payload corresponds to the time spent executing the tasks (including cache overheads) and the system load is the time wasted in the system (scheduler and context-switch overheads). Because of the order of magnitude of a scheduling overhead compared to the overhead of a context save or load, the system load is mostly the time spent waiting for a scheduler decision. The number of scheduling decisions remains similar in both cases, however, the global lock required for G-EDF adds an additional overhead. We can assume that this gap in system load between the two strategies will increase with more processors. This is in accordance with the results of Bastoni et al. stating that G-EDF is not a viable choice for hard real-time systems with a large number of processors (24 in their study) [3].

The computation times of the jobs are shorter with P-EDF compared to G-EDF as shown in Fig. 7. In proportion to G-EDF, the payload for P-EDF is 4.2 % lower in this example. The first reason is that the tasks that use the more the memory are merged in the first two processors, partly avoiding co-scheduling. The second reason is the reduction of the number of preemptions (1501 for G-EDF against only 666 for P-EDF) and task migrations (3,500 against 0) which led to less cache reloading.

This result seems compatible with the work done by Fedorova et al. that shows in their case study an improvement of the system throughput up to 32 % for a non-real-time system [14].

Obviously, in order to confirm these results, we shall conduct larger studies. Because the duration of a simulation run is very short, it is then possible to run thousands of experiments with different configurations. Such complete studies are in progress but out of scope of this paper.

## 8 Related Work

Most of the work on real-time multiprocessor scheduling addresses the theory only. Davis and Burns give a good insight of the current state of the researches in their survey [10].

A first approach for considering real-world overheads in the study of such scheduling policies is to use a cycle-accurate simulator or a real system. There are two major simulator available. The first one, Gem5 is the merger of the M5 and GEMS simulators [6]. It simulates a full system with various CPU models and a flexible memory system that includes caches. The second one, Simics, is a commercial product able to simulate full-systems but it is not cycle-accurate [21].

LITMUS<sup>RT</sup> [7], developed at the University of North Carolina (UNC), offers a different approach. It is not a simulator but an extension of the Linux Kernel which provides an experimental platform for applied real-time research and that supports a large number of real-time multiprocessor schedulers.

With both kind of tools, a substantial investment in time is required to learn how to use them and to write some new scheduler components.

There are also several tools emerging from the academic community and dedicated to the simulation of real-time systems such as Cheddar [28], MAST [17], Storm [29] and others [8, 26]. Most of these tools are designed to validate, test and analyze systems. Storm is probably the most advanced tool focusing on the study of the scheduler itself. However it does not handle direct overheads such as context-switches or scheduling overheads. Nor does it handle the impact of caches.

## 9 Conclusions

This paper presents a simulator dedicated to the study of real-time scheduling. It was designed to be easy to use, fast and flexible. Our main contribution, when compared to the existing scheduling simulators, is the integration of overheads linked to the system (context-switching, scheduling decision) and the impact of the caches.

We have shown in this paper that it is possible to take the impact of the caches into consideration. However, the models we currently use could probably be replaced by better ones. This replacement can easily be done as explained in Sect. 6.

We are already thinking about new models but they have to be validated using cycle accurate simulators.

Once our cache models will be validated and integrated into the simulator, we will launch a large campaign of simulations. As a reminder, our long term goal is the classification of the numerous scheduling policies with practical considerations. We hope that it will also help the researchers to spot the weaknesses and the strengths of the various strategies. We would be pleased if our simulation tool could be the source of innovative ideas.

**Acknowledgments** The work presented in this paper was conducted under the research project RESPECTED (<http://anr-respected.laas.fr/>) which is supported by the French National Agency for Research (ANR), program ARPEGE.

## References

1. Anderson J, Calandrino J, Devi U (2006) Real-time scheduling on multicore platforms. In: Proceedings of the 12th IEEE real-time and embedded technology and applications symposium (RTAS)
2. Babka V, Libiř P, Martinec T, Tůma P (2012) On the accuracy of cache sharing models. In: Proceedings of the third joint WOSP/SIPEW international conference on performance engineering (ICPE)
3. Bastoni A, Brandenburg B, Anderson J (2010) An empirical comparison of global, partitioned, and clustered multiprocessor EDF schedulers. In: Proceedings of the IEEE 31st real-time systems symposium (RTSS)
4. Bastoni A, Brandenburg B, Anderson J (2011) Is semi-partitioned scheduling practical? In: Proceedings of the 23rd euromicro conference on real-time systems (ECRTS)
5. Berna B, Puaat I (2012) PDPA: period driven task and cache partitioning algorithm for multicore systems. In: Proceedings of the 20th international conference on real-time and network systems (RTNS)
6. Binkert N, Beckmann B, Black G, Reinhardt SK, Saidi A, Basu A, Hestness J, Hower DR, Krishna T, Sardashti S, Sen R, Sewell K, Shoaib M, Vaish N, Hill MD, Wood DA (2011) The gem5 simulator. SIGARCH computer architecture news
7. Calandrino JM, Leontyev H, Block A, Devi UC, Anderson JH (2006) Litmus<sup>RT</sup>: a testbed for empirically comparing real-time multiprocessor schedulers. In: Proceedings of the 27th IEEE international real-time systems symposium (RTSS)
8. Chandarli Y, Fauberteau F, Masson D, Midonnet S, Qamhieh M (2012) Yartiss: a tool to visualize, test, compare and evaluate real-time scheduling algorithms. In: 3rd international workshop on analysis tools and methodologies for embedded and real-time systems (WATERS)
9. Chandra D, Guo F, Kim S, Solihin Y (2005) Predicting inter-thread cache contention on a chip multi-processor architecture. In: Proceedings of the 11th international symposium on high-performance computer architecture (HPCA)
10. Davis RI, Burns A (2011) A survey of hard real-time scheduling for multiprocessor systems. ACM Comput Surv 43(4)
11. Devi U, Anderson J (2005) Tardiness bounds under global EDF scheduling on a multiprocessor. In: Proceedings of the 26th IEEE real-time systems symposium (RTSS)
12. Eklov D, Hagersten E (2010) StatStack: efficient modeling of LRU caches. In: Proceedings of the IEEE international symposium on performance analysis of systems software (ISPASS)

13. Eklov D, Black-Schaffer D, Hagersten E (2011) Fast modeling of shared caches in multicore systems. In: Proceedings of the 6th international conference on high performance and embedded architectures and compilers (HiPEAC)
14. Fedorova A, Seltzer M, Smith M (2006) Cache-fair thread scheduling for multicore processors. Technical report TR-17-06, Division of Engineering and Applied Sciences, Harvard University
15. Guan N, Stigge M, Yi W, Yu G (2009) Cache-aware scheduling and analysis for multicores. In: Proceedings of the 7th ACM international conference on embedded software (EMSOFT)
16. Guthaus M, Ringenberg J, Ernst D, Austin T, Mudge T, Brown R (2001) Mibench: a free, commercially representative embedded benchmark suite. In: Proceedings of the IEEE international workshop on workload characterization (WWC-4)
17. Harbour MG, Garca JGG, Gutierrez JCP, Moyano JMD (2001) Mast: modeling and analysis suite for real time applications. In: Proceedings of the 13th Euromicro conference on real-time systems (ECRTS)
18. Hoste K, Eeckhout L (2007) Microarchitecture-independent workload characterization. *IEEE Micro* 27(3)
19. Liu CL, Layland J (1973) Scheduling algorithms for multiprogramming in a hard-real-time environment. *J ACM* 20
20. Liu F, Guo F, Solihin Y, Kim S, Eker A (2008) Characterizing and modeling the behavior of context switch misses. In: Proceedings of the 17th international conference on parallel architectures and compilation techniques (PACT)
21. Magnusson P, Christensson M, Eskilson J, Forsgren D, Hallberg G, Hogberg J, Larsson F, Moestedt A, Werner B (2002) Simics: a full system simulation platform. *Computer* 35(2)
22. Mattson R, Gecsei J, Slutz D, Traiger I (1970) Evaluation techniques for storage hierarchies. *IBM Syst J* 9(2)
23. Mogul JC, Borg A (1991) The effect of context switches on cache performance. *SIGOPS Oper Syst Rev* 25(4):299–313
24. Nelissen G, Funk S, Goossens J (2012) Reducing preemptions and migrations in EKG. In: IEEE 18th international conference on embedded and real-time computing systems and applications (RTCSA)
25. Radenski A (2006) “python first”: a lab-based digital introduction to computer science. In: Proceedings of the 11th annual SIGCSE conference on innovation and technology in computer science education (ITICSE)
26. Rodriguez-Cayetano M (2011) Design and development of a cpu scheduler simulator for educational purposes using sdl. In: System analysis and modeling: about models. Springer, Berlin, Heidelberg
27. SimPy Developer Team (2012) <http://simpy.sourceforge.net/>
28. Singhoff F, Legrand J, Nana L, Marcé L (2004) Cheddar: a flexible real time scheduling framework. *Ada Lett XXIV*(4)
29. Urunuela R, Déplanche AM, Trinquet Y (2010) Storm a simulation tool for real-time multiprocessor scheduling evaluation. In: Proceedings of the emerging technologies and factory automation (ETFA)

# Epidemics and Their Implications in Urban Environments: A Case Study on a National Scope

Robert Elsässer, Adrian Ogierman and Michael Meier

**Abstract** In times where urbanization becomes more important every day, epidemic outbreaks may be devastating. Powerful forecasting and analysis tools are of high importance for both, small and large scale examinations. Such tools provide valuable insight on different levels and help to establish and improve embankment mechanisms. Here, we present an agent-based algorithmic framework to simulate the spread of epidemic diseases on a national scope. Based on the population structure of Germany, we investigate parameters such as the impact of the number of agents, representing the population, on the quality of the simulation and evaluate them using real world data provided by the Robert Koch Institute [4, 22]. Furthermore, we empirically analyze the effects of certain non-pharmaceutical countermeasures as applied in the USA against the Influenza Pandemic in 1918–1919 [18]. Our simulation and evaluation tool partially relies on the probabilistic movement model presented in [8]. Our empirical tests show that the amount of agents in use may be crucial. Depending on the existing knowledge about the considered epidemic, this parameter alone may have a huge impact on the accuracy of the achieved simulation results. However, with the right choice of parameters—some of them being obtained from real world observations [10]—one can efficiently approximate the course of a disease in real world.

**Keywords** Epidemic algorithms · Power law distribution · Disease spreading · Public health

---

Partially supported by the Austrian Science Fund (FWF) under contract P25214 and by DFG project SCHE 1592/2-1. A preliminary version of this paper was published in the Proceeding of SIMULTECH 2013 [9].

---

R. Elsässer

Department of Computer Sciences, University of Salzburg, 5020 Salzburg, Austria

A. Ogierman (✉) · M. Meier

Department for Computer Science, University of Paderborn, 33102 Paderborn, Germany  
e-mail: adriano@upb.de



# 1 Introduction

In order to improve our chances to control an epidemic outbreak, we need proper models which describe the spread of a disease. Institutes, governments, and scientists all over the world work intensively on forecasting systems in order to be well prepared if an unknown disease appears.

In recent years a huge amount of theoretical and experimental study has been conducted on this topic. While theoretical analysis provides important and sometimes even counter intuitive insights into the behavior of an epidemic (e.g. [6, 8]), in an experimental study one can take many different settings and parameters [16, 17] into account. These usually cannot be considered simultaneously in a mathematical framework. A specific topology, for example, may have its own attributes that are completely different in other topological settings.

The goal of this paper is to present and empirically analyze a dynamic model for the spread of epidemics in an extended manner. One of our objectives is to find the right parameters, which lead to realistic settings. Therefore, we investigate a general simulation environment, in which the different parameters can easily be adjusted to real world observations. A second objective is to evaluate similarities between countermeasure approaches in our model and the real world. We use empirical data for the comparison. Our tool is agent-based, i.e., the individuals (or groups of such) are modeled by agents interacting with each other. The environment approximates the geography of Germany, in which agents may travel between cities. Within a city the agents interact according to the probabilistic model presented in [8] in a distributed manner. For a detailed description of the algorithmic framework see Sect. 2. In order to describe the problems and the related work, we often utilize the style and wording of [8].

## 1.1 Related Work

There is plenty of work considering epidemiological processes in different scenarios and on various networks. The simplest model of mathematical disease spreading is the so called SIR model (see e.g. [14, 20]). The population is divided into three categories: susceptible (S), i.e., all individuals which do not have the disease yet but can become infected, infective (I), i.e., the individuals which have the disease and can infect others, and recovered (R), i.e., all individuals that recovered and have permanent immunity (or have been removed from the system). Most papers model the spread of epidemics using a differential equation based on the assumption that any susceptible individual has uniform probability  $\beta$  to become infected from any infective individual. Furthermore, any infected player recovers at some stochastically constant rate  $\gamma$ .

This traditional (fully mixed) model can easily be generalized to a network. It has been observed that the corresponding process can be modeled by bond percolation on the underlying graph [13, 19]. Interestingly, for certain graphs with a

power law degree distribution, there is no constant threshold for the epidemic outbreak as long as the power law exponent is less than 3 [20] (which is the case in most real world networks, e.g. [1, 3, 11, 21]). If the network is embedded into a low dimensional space or has high transitivity, then there might exist a non-zero threshold for certain types of correlations between vertices [19]. However, none of the papers above considered the dynamic movement of individuals, which seems to be the main source of the spread of diseases in urban areas [10].

In [10] the physical contact patterns are modeled by a dynamic bipartite graph, which results from movement of individuals between specific locations. The graph is partitioned into two parts. The first part contains the people who carry out their daily activities moving between different locations. The other part represents the various locations in a certain city. There is an edge between two nodes, if the corresponding individual visits a certain location at a given time. Obviously, the graph changes dynamically at every time step.

In [7, 10] the authors analyzed the corresponding network for Portland, Oregon. According to their study, the degrees of the nodes describing different locations follow a power law distribution with exponent around 2.8.<sup>1</sup> For many epidemics, transmission occurs between individuals being simultaneously at the same place, and then people's movement is mainly responsible for the spread of the disease.

The authors of [8] considered a dynamic epidemic process in a certain (idealistic) urban environment incorporating the idea of attractiveness based distributed locations. The epidemic is spread among  $n$  agents, which move from one location to another. In each step, an agent is assigned to a location with probability proportional to its attractiveness. The attractiveness of the locations follow a power law distribution [10]. If two agents meet at some location, then a possible infection may be transmitted from an infected agent to an uninfected one. The authors obtained two results. First, if the exponent of the power law distribution is between 2 and 3, then at least a small (but still polynomial) number of agents remains uninfected and the epidemic is stopped after a logarithmic number of rounds. Secondly, if the power law exponent is increased to some large constant, then the epidemic will only affect a polylogarithmic number of agents and the disease is stopped after  $(\log \log n)^{O(1)}$  steps. In this case each agent is allowed to spread the disease for a number of time steps, which is bounded by some large constant.

In addition to the theoretical papers described above, plenty of simulation work exists. Two of the most popular directions are the so called agent-based and structured meta-population-based approach, respectively (cf. [2, 15]). Both models have their advantages and weaknesses. The main idea of the meta-population approach is to model whole regions, e.g. georeferenced census areas around airport hubs [5], and connect them by a mobility network. Then, within these regions the spread of epidemics is analyzed by using the well known mean field theory. The agent-based approach models individuals with agents in order to simulate their behavior. In this

---

<sup>1</sup> In [10] the degree represents the number of individuals visiting these places over a time period of 24 h.

context, the agents may be defined very precisely, including e.g. race, gender, educational level, etc. [16, 17], and thus provide a huge amount of detailed data conditioned on the agents setting. Furthermore, these kinds of models are also able to integrate different locations like schools, theaters, and so on. Thus, an agent may or may not be infected depending on his own choices and the ones made by agents in his vicinity. The main issue of the agent-based approach is the huge amount of computational capacity needed to simulate huge cities, continents or even the world itself [2]. This limitation can be attenuated by reducing the number of agents, which then entails a decreasing accuracy of the simulation. In the meta-population approach the simulation costs are lower, sacrificing accuracy and some kind of noncollectable data.

A specific field of application of such simulations is the investigation of the impact of (non-)pharmaceutical countermeasures on the behavior of epidemics. Germann et al. [12] investigated the spread of a pandemic strain of the influenza virus through the U.S. population. They used publicly available 2000 U.S. Census data to identify seven so-called mixing groups, in which each individual may interact with any other member. Each class of mixing group is characterized by its own set of age-dependent probabilities for person-to-person transmission of the disease. They considered different combinations of socially targeted antiviral prophylaxis, dynamic mass vaccination, closure of schools and social distancing as countermeasures in use, and simulated them with different basic reproductive numbers  $R_0$ . It turned out that specific combinations of the countermeasures have a different influence on the spreading process. For example, with  $R_0 = 1.6$  social distancing and travel restrictions did not really seem to help, while vaccination limited the number of new symptomatic cases per 10,000 persons from  $\sim 100$  to  $\sim 1$ . With  $R_0 = 2.1$ , such a significant impact could only be achieved with the combination of vaccination, school closure, social distancing and travel restrictions.

## 1.2 Our Results

The results of this paper are two-fold. First, we show that by increasing the number of agents we are able to significantly improve the accuracy of our results in the scenarios we have tested. This is due to different phenomena which are only visible if the amount of agents in use is large enough. For example, if the number of agents exceeds a certain value, then the epidemic manages to keep a specific (low) amount of infected individuals over a long time period. Furthermore, the number of agents has to be above some threshold to allow the epidemic to enter some specific areas/cities in the environment we used. Obviously, a certain amount of agents is also needed to avoid significant fluctuations in our results.

The second main result of this paper is that by setting the parameters properly, one can approximate the effect of some non-pharmaceutical countermeasures, that are usually adopted if an epidemic outbreak occurs. This observation is supported by the empirical study of [18]. Interestingly, the right choices of parameters in our experiments seem to be in line with previous observations in the real world (e.g. the

right power law exponent seems to be in the range of 2.6–2.9, cf. [10]). To analyze the effect of the countermeasures mentioned above, we integrate the corresponding mechanisms on a smaller scale, and then verify their impact on a larger scale too.

## 2 Theoretical Models and Algorithmic Framework

In the following we provide an empirical analysis on a small as well as on a large scale. Hereby, the cities are chosen from a list in descending order of their population size. It is intuitively clear that large (and thus attractive) cities play a major role for the epidemic pace since a higher population density entails a potentially higher infection probability. Excluding such hotspots would of course slow down the infection spread. The problem is, this could only be achieved by putting the whole city into quarantine. However, isolating an entire city is not a trivial task. For example, people living in the suburbs but working in the city might not be willing to risk their job by obeying the quarantine. Therefore, we consider such strategies only on a smaller scale.

In our model the agents may not only move between locations within a city but between cities as well. Furthermore, due to simplicity, the agents are not categorized (i.e., they do not provide further properties like gender etc.). Since we are not interested in the evaluation of such details. In the following, we briefly introduce the model. Due to readability we present our model with respect to the following four categories: (1) The environment on a large scale (inter-city movement), (2) The environment on a small scale (intra-city movement), (3) The epidemic model, and (4) The countermeasure model.

### 2.1 The Environment on a Large Scale

Let  $G(d) = (V, E)$  be a complete graph with  $m$  nodes  $V := \{c_1, \dots, c_m\}$  and parameter  $d := \{d_{c_1}, \dots, d_{c_m}\}$ , whereas  $c_i = G_{c_i}$  (see below) represents the corresponding city  $c_i$ . Let further  $p_{c_i}$  be the population size of  $c_i$  and  $p := \{p_{c_1}, \dots, p_{c_m}\}$ . Then the attractiveness of  $c_i$  is given by  $d_{c_i} = \frac{p_{c_i}}{\sum_{1 \leq i \leq m} p_{c_i}}$ . Consequently,  $G$  represents the topology, which contains all cities, on a large scale. In other words, we model the inter-city movement using the complete graph  $G(d)$ . In this graph, each  $c_i \in V$  corresponds to a city of Germany. However, depending on the size, not every city of Germany is represented by a node in  $V$ .<sup>2</sup> The population is represented by

---

<sup>2</sup> The amount of overall agents in use ( $n$ ) determines how many cities are represented by  $V$ . Therefore we sort the list of all cities of Germany in descending order of their population size. Then, starting from the top, we include the currently considered city  $c_i$  to  $V$  if and only if the assigned amount of agents to said city is at least 1. The latter amount is given by  $n \cdot d_{c_i}$ .

$n = \sum_{1 \leq i \leq m} nd_{c_i}$  agents, with  $nd_{c_i}$  being the number of agents assigned to  $c_i$ . Note that said number is proportional to the cities real world population. Furthermore, each city  $c_i \in V$  is assigned an attractiveness  $d_{c_i}$  proportional to its population size (w.r.t. the whole population). Let  $A_{i,s,t}$  be the event that agent  $i$  travels from city  $s$  to  $t$ . Let further  $p'$  be the probability that an agent decides to travel at all, and let  $dist(s, t)$  be the Euclidean distance between cities  $s$  and  $t$ . Then the probability that event  $A_{i,s,t}$  occurs is given by

$$Pr(A_{i,s,t}) = p' \cdot \frac{d_t \cdot dist^{-1}(s, t)}{\sum_{(s,j) \in E} d_j \cdot dist^{-1}(s, j)}.$$

Thus, the probability for an agent entering a specific city depends on the distance to said city, its population size as well as the current position of the considered agent.

## 2.2 The Environment on a Small Scale

Let  $G_{c_i}(d(v)) = \{V_{c_i}, E_{c_i}\}$  be a complete graph with  $m_{c_i} = \lceil \kappa nd_{c_i} \rceil$  nodes (also called *cells*), with  $d(c_i) := \{d_{v_1}, \dots, d_{v_{m_{c_i}}}\}$ . Here,  $\kappa > 0$  is a constant, which will be specified in the upcoming experiments. Further, note that  $\kappa$  does not affect the amount of agents but the amount of cells only. Then the attractiveness of cell  $v_i \in V_{c_i}$  within a city  $c_i$  is given by  $d_{v_i}$ . Said attractiveness is chosen randomly with probability proportional to  $1/\beta^\alpha$  for a value  $\beta > 1$ , where  $\alpha > 2$  is a constant depending on the simulation run. In other words, each  $c_i \in V$ , representing city  $c_i$ , is a clique of cells on its own, thus incorporating intra-city movement into our model. The cells represent locations within a city an agent can visit. Each cell may contain agents (individuals), depending on the cells so-called *attractiveness*. If an agent decides to stay within its current city, said agent moves to a randomly chosen cell according to the distribution of the attractiveness' among the cells. This also holds for the first cell an agent is accommodated in after entering a city.

## 2.3 Epidemic Model

We use three different states to model the distributed spreading process. These states partition the set of agents into three groups;  $\mathcal{I}(j)$  contains the infected agents in step  $j$ ,  $\mathcal{U}(j)$  contains the uninfected (susceptible) agents in step  $j$ , and  $\mathcal{R}(j)$  contains the resistant agents in step  $j$ . Whenever it is clear from the context, we simply write  $\mathcal{I}$ ,  $\mathcal{U}$ , and  $\mathcal{R}$ , respectively. If an uninfected agent  $i$  visits a cell (within

a city) which also contains agents of  $\mathcal{I}(j)$ , then  $i$  becomes infected with probability  $1 - (1 - \gamma)^{|\mathcal{I}'(j)|}$ , where  $\mathcal{I}'(j)$  represents the set of infected agents accommodated in the same cell as  $i$ . We refer to the concrete value of  $\gamma$  in the upcoming simulations.

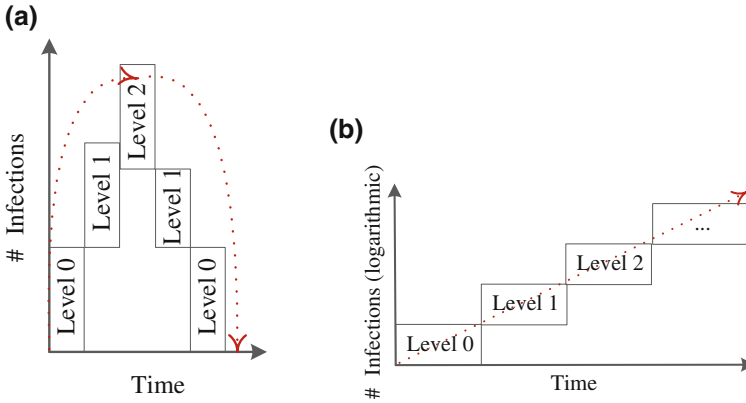
## 2.4 Countermeasure Models

Our countermeasure models mainly take advantage of the parameters  $\alpha$  and  $\kappa$ . That is, high values of these two parameters imply a high level of countermeasures and vice versa. With countermeasures applied, individuals avoid places with a large number of persons more often, waive needless tours, and are more careful when meeting other people. While  $\alpha$  is mostly responsible for a decreasing number of visitors within a cell, and thus for the avoidance of crowded areas for example,  $\kappa$  determines the total space available for all individuals. As pointed out in [18], a single countermeasure alone is most likely not sufficient to stop an epidemic. Therefore, we assume a combination of them to be in place, which then would be able to sufficiently influence the parameters  $\alpha$  and  $\kappa$ . Note, our countermeasure models apply to each city  $c_i \in V$  individually.

We use two different types of countermeasure-models: a (multi-tier) *level based approach* considering the amount of infected agents in the current step, and a *ratio based approach* considering the amount of newly infected agents in the current step compared to the one in the step before. In the following we use  $\alpha_0$  and  $\kappa_0$  as initial values for  $\alpha$  and  $\kappa$ , respectively.

In the *level based model* we have one or more levels in which a certain pair of parameters  $\alpha$  and  $\kappa$  is used. Let  $LM_m$  stand for the level based model with  $m$  levels  $L = \{l_1, \dots, l_m\} \cup l_0$ . Further, let  $T = \{l_0^d, l_0^u, l_1^d, l_1^u, \dots, l_m^d, l_m^u\}$  be the set of transition points for all levels, i.e.,  $l_i^d$  defines the transition point from level  $i$  to  $i - 1$  whereas  $l_i^u$  defines the transition point from level  $i$  to  $i + 1$ . Note that,  $l_i^u = l_{i+1}^d$  does not necessarily hold. Additionally,  $\alpha_0, \alpha_1, \dots, \alpha_m$  and  $\kappa_0, \kappa_1, \dots, \kappa_m$  define the parameters  $\alpha$  and  $\kappa$ , which are applied in the corresponding levels  $l_0, \dots, l_m$ . Figure 1a depicts an example situation.

In contrast, the *ratio based model RM* uses a non static approach. Let the set of newly infected nodes of a city  $c_i$  in step  $j$  be denoted by  $\mathcal{I}_{c_i}^*(j)$ . Furthermore, let  $\alpha_j$  and  $\kappa_j$  denote the corresponding parameters used in step  $j$ . If  $\frac{|\mathcal{I}_{c_i}^*(j)|}{|\mathcal{I}_{c_i}^*(j-1)|} \geq a$ , for some constant  $a$ , then we set  $\alpha_{j+1} = \alpha_j + z_1$  and  $\kappa_{j+1} = \kappa_j + z_2$ , where  $z_1, z_2$  are some small constants which will be specified later. Consequently, we set  $\alpha_{j+1} = \max\{\alpha_0, \alpha_j - z_1\}$  and  $\kappa_{j+1} = \max\{\kappa_0, \kappa_j - z_2\}$  whenever  $\frac{|\mathcal{I}_{c_i}^*(j)|}{|\mathcal{I}_{c_i}^*(j-1)|} \leq 1/a$ . The values applied in the various models are specified in Sect. 3.3. Figure 1b depicts an example situation.



**Fig. 1** Basic examples for the two countermeasure models. **a** An example demonstrating a possible configuration of  $LM_2$ . The *broken line* depicts the amount of total infections in the city while the scope of activity of the countermeasure levels is represented by the *rectangles*. **b** An example demonstrating a possible configuration of  $RM$ . The *broken line* depicts the amount of total infections in the city (on a logarithmic scale) while the scope of activity of the countermeasure levels is represented by the *rectangles*

### 3 Experimental Analysis

The environment in our simulations approximates the geography of Germany utilizing 10 million agents and more. Note that, the obtained results remain similar utilizing up to 100 million agents. Depending on the number of agents, our simulations incorporate several hundred cities as visitable areas spread all over the country. Only cities with an initial agent amount of at least one are included in the simulation. Each city is assumed to be reachable from any other city. However, an agent may travel at most 1,000 km within a single round. Each round represents a whole day in the real world. Consequently, an agent moving from one city to another has to wait until its destination is reached before it can interact with other agents at said destination.

#### 3.1 Simulations

In the upcoming sections we present and evaluate our results with the focus on: (1) the impact of the number of agents on the characteristics of the simulated epidemic compared to real world data, and (2) the impact of non-pharmaceutical countermeasures on the behavior of the epidemic (e.g. social distancing, school closures, and isolation [18]).

Furthermore, we also analyze our parameter settings. Although this is only a short part, our settings seem to coincide with the real world observations of [10],

and thus provide an additional valuable insight. Note that the figures presented in this section show values based on the real world population size and not on the number of agents.

### 3.2 Relevance of the Chosen Parameters

Based on real world observations (e.g. [10]), we chose  $\alpha = 2.8$  and  $\kappa = 1$  as a starting point for a series of simulations concerning  $\alpha$  and  $\kappa$ , respectively. Each plot represents the average value of 50 different simulation runs for each parameter constellation utilizing 10 million agents. The parameter notation is as follows:  $\gamma$  is the probability for an agent  $v \in \mathcal{U}$  to become infected independently by each  $w \in \mathcal{I}$  occupying the same cell at the same time,  $\xi$  is the amount of rounds an agent  $v \in \mathcal{I}$  is infectious, thus being able to infect others,  $City_{init}$  is the initial amount of cities the infection is being placed in,  $Agent_{init}$  is the amount of initially infected agents which are placed in  $City_{init}$  different cities,  $\alpha$  is the power law exponent used to individually compute the attractiveness of the cells within each city, and  $\kappa$  is a multiplicative factor to increase/decrease the amount of cells proportional to the initially assigned amount of agents.

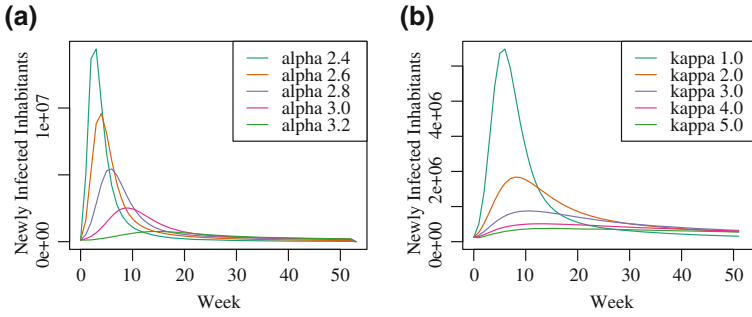
Figure 2a and b show the impact of  $\alpha$  and  $\kappa$  on the behavior of the epidemic, and compare the results to the characteristics of a typical Influenza case reported by the RKI.<sup>3</sup> To increase the readability, we omit to add the RKI-plot as the 6th one. Instead we refer the reader to Fig. 3. Among all five  $\alpha$ -values  $\alpha = 2.8$ , which has also been obtained in [10] in a different context, represents the best tradeoff between curve similarity and amount of infections. All remaining parameters were set to identical values as in Case 1 (cf. Sect. 3.3). For  $\kappa$  a similar phenomenon can be observed. With increasing  $\kappa$  (including fractional values), the characteristics of the curve (i.e., the amount of infected individuals at the peak vs. total number of infections and total duration of the outbreak) become less and less accurate. Even if we increase the value to 2, the obtained curve does not follow the characteristics of the real world observations reported by the RKI [see footnote 3] anymore.

In terms of the probability of infection  $\gamma$  we simply chose a reasonable value low enough to model an Influenza epidemic, but high enough to provoke an outbreak. This seemed reasonable due to the (at least to our knowledge) lack of concrete values deduced from real world observations.

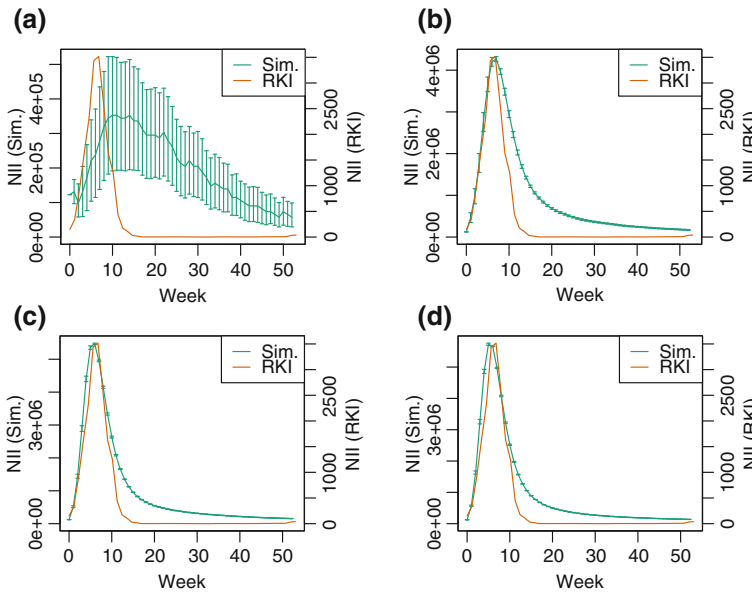
---

<sup>3</sup> The Robert Koch Institute (RKI) is the central federal institution in Germany responsible for disease control and prevention and is therefore the central federal reference institution for both, applied and response-orientated research. (Source [http://www.rki.de/EN/Home/homepage\\_node.html](http://www.rki.de/EN/Home/homepage_node.html)).





**Fig. 2** A composition of multiple simulation runs concerning varying  $\alpha$  (a) and  $\kappa$  (b) only. All other parameters are identical to Case 1 (cf. Sect. 3.3). Each result represents the average of 50 different simulation runs with 10 million agents for the topology of Germany. **a**  $2.4 \leq \alpha \leq 3.2$ . **b**  $1 \leq \kappa \leq 5$



**Fig. 3** Simulation results for Case 1a) (green) in comparison to real world data (red) provided by the RKI for a varying amount of agents. The abbreviation *NII* stands for *Newly Infected Inhabitants*. All results up to and including 10 million agents represent the average values of 50 different simulation runs, whereas the result for 100 million agents averages over 20 different simulation runs. The reliability of the averaged value is indicated by the corresponding confidence interval. **a** 10.000 agents. **b** 500.000 agents. **c** 10.000.000 agents. **d** 100.000.000 agents

### 3.3 Case 1—Number of Agents

Before we present the results for this case, we first introduce the relevant sources for comparison. Here we compare our results to real world data provided by the RKI [22] SurvStat system for the year 2007. The parameter values were taken from reference data provided by the RKI [4] where possible, or set to reasonable ones otherwise (cf. Sect. 3.2 for more details).

#### 3.3.1 RKI: Basis of Comparison

In the following we compare our results to the real world data provided by the RKI [see footnote 3]. For this purpose we use two different data sources: the official report of the Influenza epidemiology of Germany for 2010/2011 [4] and an online database containing obligatory reports called SurvStat [22].

*Relevance.* The data for the SurvStat database and the report of 2010/2011 itself were obtained from more than 1 % of all primary care doctors spread all over Germany. This indicates the significance according to international standards. Unfortunately, there are some drawbacks resulting from the type of data ascertainment itself. Note that not every infected person consults a doctor, which implies that the data of the SurvStat system contains only the serious courses of the disease. Nonetheless, these sources provide a valuable tool to obtain insight into the spread and persistence of the Influenza virus in Germany. Further, due to the data's significance, it is possible to estimate the number of infections within certain areas as well. Since the spread of infections in the real world is influenced by factors like seasonal fluctuations or the virus' aggressiveness, the number of total infections may significantly differ from year to year, cf. [22] for different years. However, the course of the curve usually does not change. Consequently, we do not focus on absolute values in our simulations, but on the *characteristics* of our results. These characteristics remain, up to some scaling factor, identical over the whole data set provided by the RKI.

#### 3.3.2 Case 1

The parameters for this subcase are as follows. We set  $\gamma = 7\%$ ,  $\xi$  to 5 days, the amount of initially infected cities  $City_{init}$  to 1 (namely Berlin), and the amount of initially infected agents  $Agent_{init}$  to 0.0015 % of the overall agents used for these simulations. Furthermore,  $\alpha = 2.8$  and  $\kappa = 1$ .

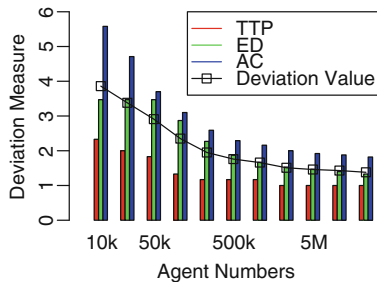
Figure 3 shows the results for this case. Here the green curve represents the real-world data provided by the RKI for the year 2007 while the red curve represents our simulation results. Note that both curves vary significantly in terms of absolute numbers. However, this is not our focus here. Due to the level of abstraction in our model and since the RKI data only contains reported cases (see above), the absolute

numbers do not coincide. Additionally, as stated above, the data provided by the RKI also differs significantly (in terms of absolute numbers but not the disease characteristics) from year to year (cf. [22]). Therefore, we focus on the course of the disease and the resulting characteristics of the plotted curves.

It is easy to see that the more agents are used, beginning from Fig. 3a up to 3d, the more the curve characteristics converge. Moreover, the accuracy of each simulation run increases as well (cf. the confidence intervals in Fig. 3). With at least 500,000 agents in use, both curves become similar.

To obtain a more formal evaluation, we define three measures, which are used to compare our results to the data provided by the RKI. These are: the time to peak (TTP), the epidemic duration (ED), and the area of the curve (AC). The time to peak describes the week with the maximum amount of newly infected agents of the corresponding curve. The area of the curve is simply the summation of the area between the origin and the endpoint EP (defined by the epidemic duration). Finally, the endpoint of the epidemic duration is the week in which only a minor amount of new infections occur, and no significant new infections are observed anymore. Minor infections are defined to start in a round  $i$  and last till the last round  $j$  of the simulation while for all rounds  $i \leq i' \leq j$  it holds that the amount of newly infected agents does not exceed 9 % of the maximum value.

In the following we consider the relative values of these measurements compared to the RKI data. That is, the numbers represent the ratio between the value obtained in our simulations and the value provided by the RKI. For example, a value of 4.00 for TTP in the case of 10,000 agents implies that the time to peak in our case divided by the time to peak in the real world data is 4. Using the individual deviation measurements, we define a global deviation value by the formula  $\frac{1}{3} \cdot \text{TTP} + \frac{1}{3} \cdot \text{AC} + \frac{1}{3} \cdot \text{ED}$ . For simplicity we consider each individual measurement uniformly weighted. The results, which confirm our previous observations, are shown in Fig. 4. All obtained results and previous statements imply a fragile balance between the accuracy, the parameter setting, and the amount of agents in use.



**Fig. 4** A visual representation of the data computed using the deviation measurements for the experiments conducted in Case 1

### 3.3.3 Case 2—Non-pharmaceutical Countermeasures

Now we extend our analysis to incorporate non-pharmaceutical countermeasures such as school closures and social distancing. Here we stick to a fictional epidemic simply because it simplifies the presentation, i.e., due to the increase of  $\gamma$  to 12%, a faster spread is achieved and the impact of the used countermeasures is amplified. Similar to Case 1, we set  $\zeta$  to 5 days, the amount of initially infected cities  $City_{init}$  to 1 (namely Berlin), and the amount of initially infected agents  $Agent_{init}$  to 0.00075% of the overall agents (to compensate the higher  $\gamma$  in the beginning). All relevant parameters regarding the countermeasure models can be found in the original paper [9].

We assume that non-pharmaceutical countermeasures basically affect the parameters  $\alpha$  and  $\kappa$ , since the individuals will most likely avoid places with a large number of persons, waive needless tours, and be more careful when meeting other people. For obvious reasons, we cannot compare our simulation results to current real-world data containing results for different epidemics with varying (or no) countermeasures in use. Therefore, we use the work of Markel et al. [18] for this purpose. Especially Fig. 5 is of particular interest. There, the direct correlation between establishing countermeasures and a decreasing amount of new infections (and vice versa) is well highlighted. We observed an identical effect in our simulations (cf. Fig. 6). Note that different combinations of countermeasures used in [18] entail different kinds of impacts on the death rates. In contrast, we do not focus on specific combinations but on sufficient ones to actually achieve an immediate impact on the epidemic.

As already described in Sect. 2, we use two different countermeasure approaches: the level based ( $LM_1, LM_2$  and  $LM_3$ ), and the ratio based ( $RM$ ), respectively. Both use different mechanisms and parameters to achieve the embankment of the epidemic. Recall that all transition points in the level based model are chosen w.r.t.

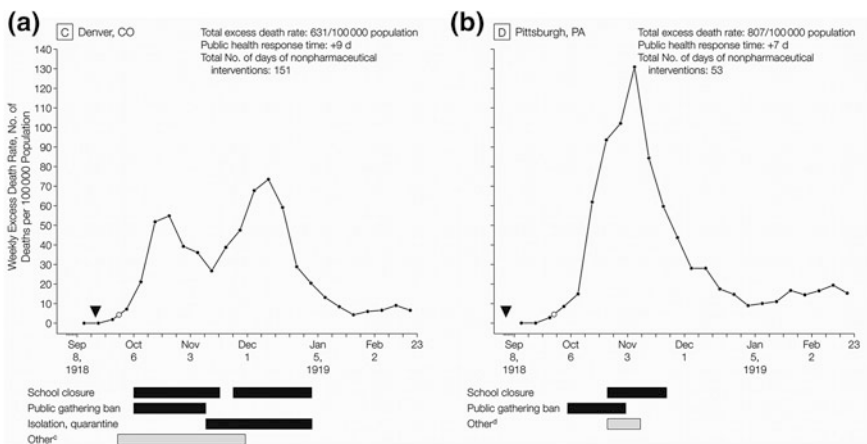
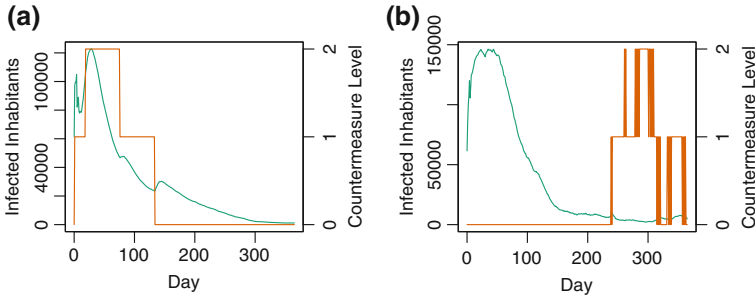
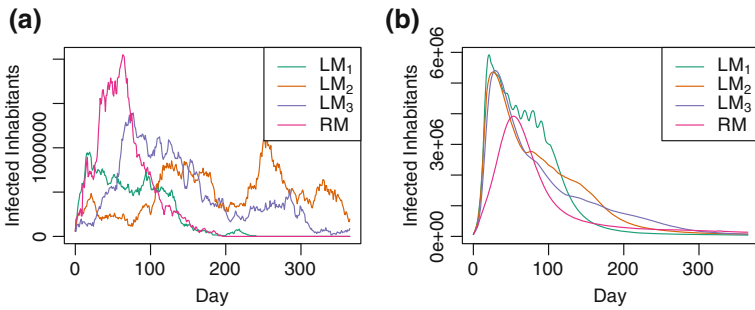


Fig. 5 Weekly excess death rates from September 8, 1918, through February 22, 1919 [18, Fig. 3]



**Fig. 6** Example results for BER for Case 2. The *green curve* represents the amount of infected inhabitants in the corresponding city while the *red curve* indicates the activated countermeasure level with respect to the countermeasure model in use. The number of agents refers to the total amount of agents used for the simulation. **a** Model  $LM_3$ , 100 million agents. **b** Model  $RM$ , 10 million agent



**Fig. 7** Comparison of different countermeasure models on a large scale, i.e., each value represents the situation for the corresponding countermeasure model in all cities combined. **a** 10.000 agents. **b** 100.000.000 agents

the ratio between the amount of currently infected individuals and the population size of the city. Figure 7 compares all models to each other on the national level.

Although the level based approach is completely different compared to the ratio based approach, the achieved results are similar. However, the overall increase of  $\alpha$  and  $\kappa$  by the ratio based approach may be noticeably higher, especially if a large number of agents is used. That is, while all  $LM$ -models use  $\alpha \leq 3.3$ , the  $RM$ -model goes above 4. This implies that the  $LM$ -models are more cost efficient, since both  $\alpha$  and  $\kappa$  are kept lower and therefore less effort is needed to achieve and maintain said values.

Additionally, to be able to compare our results to the findings in [18], we examine the following city in more detail: Berlin (BER) with a population size of  $\approx 3.5$  million. Figure 6 represents a composition of some interesting results. Note

that the green curve in these figures represents the countermeasure level for the *LM* models at the corresponding time, and indicates the number of times  $z_1, z_2$  have increased  $\alpha, \kappa$  in the *RM* model. Recall that, in the *RM*-model level  $j$  implies  $\alpha_i = \alpha_0 + \sum_{k=1}^j z_1$  and  $\kappa_i = \kappa_0 + \sum_{k=1}^j z_2$  for a step  $i$ .

Our results confirm the impact of different countermeasures observed in the real world [18]. Compared to Fig. 5, our simulations show a similar behavior (i.e., more than one peak during the epidemic). It is easy to see that the countermeasures presented in [18] directly influence the course of the epidemic. The same property can be observed in our results (cf. Fig. 5). One can see that depending on the countermeasure level (indicated by the activated/used level), the number of infections increases or decreases. Note that although our figures show the number of infected individuals and not the death rate as shown in Fig. 5, a comparison is still possible, since this deviation can be normalized using a scaling factor.

Furthermore, we observe that small adjustments of the two parameters  $\alpha$  and  $\kappa$  entail a significant impact on the number of overall infections. Among others we already proved that if the power law exponent (and  $\kappa$  as well) is assumed to be some large constant, then even a very aggressive epidemic with  $\gamma = 100\%$  will affect no more than a polylogarithmic number of the population. Our findings now back up these observations.

In conclusion, we showed the impact of different countermeasures on the behavior of a population w.r.t. our model. Although some complexity of the real world is lacking, the similarities to real-world observations are still present. Starting with settings for the environment, and therefore implicitly the individuals' behavior, based on real-world observations, relatively low level countermeasures were sufficient to embank or at least significantly suppress an outbreak. Essentially the same properties were already observed in reality (cf. [18]). This underlines the importance of behavioral and environmental models based on power law distributions.

## 4 Conclusions

Agent based simulators offer various possibilities to perform very detailed experiments. However, the parameters used in these experiments highly influence the results one might obtain. As we have seen, even the number of agents has a significant impact on the quality of the results. This includes the reliability of different simulation runs with an identical parameter setting. By using the right parameter settings and a proper number of agents, it is possible to approximate the course of a disease as observed in the real world. Furthermore, our experiments indicate that the algorithmic framework presented in this paper is able to describe, to some extent, the impact of certain non-pharmaceutical countermeasures on the behavior of an epidemic.

## References

1. Adamic LA, Huberman BA (2000) Power-law distribution of the world wide web. *Science* 287(5461):2115
2. Ajelli M, Goncalves B, Balcan D, Colizza V, Hu H, Ramasco J, Merler S, Vespignani A (2010). Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. *BMC Infect Dis* 10:190
3. Amaral LA, Scala A, Barthelemy M, Stanley HE (2000) Classes of small-world networks. *PNAS* 97(21):11149–11152
4. Arbeitsgemeinschaft Influenza (2011). Bericht zur Epidemiologie der Influenza in Deutschland Saison 2010/2011
5. Balcan D, Hu H, Goncalves B, Bajardi P, Poletto C, Ramasco JJ, Paolotti D, Perra N, Tizzoni M, den Broeck WV, Colizza V, Vespignani A (2009) Seasonal transmission potential and activity peaks of the new influenza A(H1N1): a Monte Carlo likelihood analysis based on human mobility. *BMC Med* 7:45
6. Borgs C, Chayes J, Ganesh A, Saberi A (2010) How to distribute antidote to control epidemics. *Random Struct Algorithms* 37:204–222
7. Chowell G, Hyman JM, Eubank S, Castillo-Chavez C (2003) Scaling laws for the movement of people between locations in a large city. *Phys Rev E* 68(6):661021–661027
8. Elsässer R, Ogierman A (2012) The impact of the power law exponent on the behavior of a dynamic epidemic type process. In: SPAA'12, pp 131–139
9. Elsässer R, Ogierman A, Meier M (2013) Agent based simulations of epidemics on a large scale. In: SIMULTECH'13, pp 263–274
10. Eubank S, Guclu H, Kumar V, Marathe M, Srinivasan A, Toroczkai Z, Wang N (2004) Modelling disease outbreaks in realistic urban social networks. *Nature* 429(6988):180–184
11. Faloutsos M, Faloutsos P, Faloutsos C (1999) On power-law relationships of the internet topology. In: SIGCOMM'99, pp 251–262
12. Germann TC, Kadau K, Longini IM, Macken CA (2006) Mitigation strategies for pandemic influenza in the United States. *PNAS*, 103(15):5935–5940
13. Grassberger P (1983) On the critical behavior of the general epidemic process and dynamical percolation. *Math Biosci* 63(2):157–172
14. Hethcote HW (2000) The mathematics of infectious diseases. *SIAM Rev* 42(4):599–653
15. Jaffry SW, Treur J (2008) Agent-based and population-based simulation: a comparative case study for epidemics. In: Louca LS, Chrysanthou Y, Oplatkova Z, Al-Begain K (eds) ECMS'08, pp 123–130
16. Lee BY, Bedford VL, Roberts MS, Carley KM (2008) Virtual epidemic in a virtual city: simulating the spread of influenza in a us metropolitan area. *Transl Res* 151(6):275–287
17. Lee BY, Brown ST, Cooley PC, Zimmerman RK, Wheaton WD, Zimmer SM, Grefenstette JJ, Assi T-M, Furphy TJ, Wagener DK, Burke DS (2010) A computer simulation of employee vaccination to mitigate an influenza epidemic. *Am J Prev Med* 38(3):247–257
18. Markel H, Lipman HB, Navarro JA, Sloan A, Michalsen JR, Stern AM, Cetron MS (2007) Nonpharmaceutical interventions implemented by US cities during the 1918–1919 Influenza Pandemic. *JAMA* 298(6):644–654
19. Newman MEJ (2002) Spread of epidemic disease on networks. *Phys Rev E* 66(1):016128
20. Newman MEJ (2003) The structure and function of complex networks. *SIAM Rev* 45(2):167–256
21. Ripeanu M, Foster I, Iamnitchi A (2002) Mapping the Gnutella network: properties of large-scale peer-to-peer systems and implications for system. *IEEE Internet Comput J* 6(1):50–57
22. Robert Koch Institute (2012). *SurvStat@RKI*. A web-based solution to query surveillance data in Germany

# Hybrid Multilinear Modeling and Applications

Georg Pangalos, Annika Eichler and Gerwald Lichtenberg

**Abstract** Tensor systems are a framework for modeling of multilinear hybrid systems with discrete and continuous valued signals. Two examples from building services engineering and multi-agent systems show applications of this framework. A tensor model of a heating system is derived and approximated by tensor decomposition methods first. Second, a tensor model of a multi-agent system with a structure already given in a decomposed form is reduced further by the same decomposition methods.

**Keywords** Multilinear systems · Hybrid systems · Tensor decomposition · Heating systems · Multi-agent systems

## 1 Introduction

Multilinear models are a subclass of nonlinear models which is interesting for engineering applications because several systems behave multilinear, as shown by the examples. In most applications the processes are driven by some automation device with fixed sampling time. Thus, only discrete time models are investigated here.

---

G. Pangalos · G. Lichtenberg (✉)  
Faculty of Life Sciences, Hamburg University of Applied Sciences,  
Ulmenliet 20, Hamburg, Germany  
e-mail: gerwald.lichtenberg@haw-hamburg.de

G. Pangalos  
e-mail: georg.pangalos@haw-hamburg.de

A. Eichler  
Institute of Control Systems, Hamburg University of Technology,  
Eissendorfer Str. 40, Hamburg, Germany  
e-mail: annika.eichler@tuhh.de



The class of multilinear models obviously includes linear continuous valued models but moreover, all discrete valued models are inherently multilinear, see [6]. As multilinearity is mathematically closely related to tensor calculus, there are tensor representations of multilinear systems. Of special interest are discrete time hybrid systems consisting of a multilinear continuous valued subsystem and an arbitrary discrete valued subsystem connected by a quantizer and an injector. These can be represented by so-called *Hybrid Tensor Systems* defined in [7]. This modeling approach was first applied to the identification problem of discrete valued models from continuous data [8].

Over the last years simulation of building and heating systems has attracted a lot of interest, see e.g. [10] or [16]. In these approaches nonlinear models are derived and simulated, mostly for performance evaluation. Heating systems are application examples for hybrid tensor systems because of their physical structure, see [12]. A multilinear one zone model of a large scale building is used in this paper.

Another field of applications are distributed network systems like Multi-Agent Systems (MAS), which have received considerable attention over the past 10 years, because of their broad variety of applications, see [13, 14]. To perform tasks like Decision-Making or Policy Formulation, hybrid MASs are required [15]. These can be represented by the same class of hybrid tensor systems.

Numerical tensor calculus with special interest on decomposition methods is an active research field in applied mathematics, see e.g. [3, 4]. Tensor models can be approximated by standard tensor decomposition techniques—similar to model reduction techniques for linear systems. Most of the latter techniques like PCA use singular value decomposition (SVD) of matrices as workhorse for the reduction process. For tensor systems, so-called *higher-order SVD* decomposition algorithms are appropriate reduction methods, see [5].

The paper is organized as follows. After this introduction the class of hybrid tensor systems is described in Sect. 2. The model of a complex heating system is represented by a hybrid tensor model and decomposed in Sect. 3. In Sect. 4 multilinear modeling and transformation to a tensor system is shown for a hybrid MAS. Final conclusions are drawn in Sect. 5.

## 2 Hybrid Tensor Systems

Tensor representations of multilinear systems are introduced in [7] and reviewed in this section. All linear systems can be represented by this class, which on the other hand is a subset of the class of polynomial systems.

**Definition 1** A Tensor

$$\mathbf{X} \in \mathbb{D}^{I_1 \times I_2 \times \dots \times I_n}$$

of order  $n$  is an  $n$ -way array where elements  $x_{i_1 i_2 \dots i_n}$  indexed by  $i_j \in \{1, 2, \dots, I_j\}$  for  $j = 1, \dots, n$  can take values from some domain  $\mathbb{D}$ .

**Definition 2** A Canonical Polyadic (CP) tensor

$$\mathbf{K} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n] \cdot \boldsymbol{\lambda} \in \mathbb{D}^{r_1 \times r_2 \times \dots \times r_n}$$

is a tensor of dimension  $(r_1, \dots, r_n)$ , with elements given by

$$K_{jk\dots p} = \sum_{i=1}^r \lambda_i (\mathbf{X}_1)_{ji} (\mathbf{X}_2)_{ki} \dots (\mathbf{X}_n)_{pi},$$

i.e. sums of outer products of the column vectors of so-called *factor matrices*  $\mathbf{X}_i \in \mathbb{D}^{r_i \times r}$ , weighted by the elements of the so-called *weighting vector*  $\boldsymbol{\lambda}$ . A weighting vector of ones can be omitted.

**Definition 3** The tensors  $\mathbf{Y} \in \mathbb{D}^{I_1 \times \dots \times I_n}$  and  $\mathbf{X} \in \mathbb{D}^{I_1 \times \dots \times I_n \times I_{n+1} \times \dots \times I_{n+m}}$  have a contracted product

$$\langle \mathbf{X} | \mathbf{Y} \rangle (k_1, \dots, k_m) = \sum_{i_1}^{I_1} \dots \sum_{i_n}^{I_n} x_{i_1, \dots, i_n, k_1, \dots, k_m} y_{i_1, \dots, i_n}, \quad (1)$$

which is a tensor of dimension  $I_{n+1} \times \dots \times I_{n+m}$ .

With the monomial CP tensor (weighting vector  $\boldsymbol{\lambda} = 1$  is omitted)

$$\mathbf{M}(\mathbf{x}, \mathbf{u}) = \left[ \begin{pmatrix} 1 \\ \mathbf{u}_m \end{pmatrix}, \dots, \begin{pmatrix} 1 \\ \mathbf{u}_1 \end{pmatrix}, \begin{pmatrix} 1 \\ x_n \end{pmatrix}, \dots, \begin{pmatrix} 1 \\ x_1 \end{pmatrix} \right],$$

the state transition equation of a multilinear state space model

$$\Phi(\mathbf{x}) = \langle \mathbf{F} | \mathbf{M}(\mathbf{x}, \mathbf{u}) \rangle \quad (2)$$

can be written in terms of a contracted tensor product, where  $\Phi(x)$  denotes the next state vector. The transition tensor  $\mathbf{F} \in \mathbb{D}^{\times^{(n+m)}2 \times n}$  contains all parameters of the model while the structure is given by the monomial tensor  $\mathbf{M}(\mathbf{x}, \mathbf{u}) \in \mathbb{D}^{\times^{(n+m)}2}$ , where the following notation for tensor spaces is used

$$\mathbb{D}^{\times^{(n+m)}2} := \mathbb{D}^{\overbrace{2 \times \dots \times 2}^{n+m}}.$$

The transition tensor can be represented either element wise or in CP form

$$\mathbf{F} = [\mathbf{F}_{u_m}, \dots, \mathbf{F}_{u_1}, \mathbf{F}_{x_n}, \dots, \mathbf{F}_{x_1}, \mathbf{F}_\phi] \cdot \boldsymbol{\lambda}_f. \quad (3)$$

The latter usually has important advantages for engineering applications with sparse structure of the transition tensor where most of the  $n2^{n+m}$  elements are zero.

Moreover, contracted products can be efficiently computed if both tensors are given in CP form, like for the next state equation

$$\langle \mathbf{F} | \mathbf{M}(\mathbf{x}, \mathbf{u}) \rangle = \mathbf{F}_\phi^T \left( \lambda_f \otimes \left( \mathbf{F}_{u_m}^T \begin{pmatrix} 1 \\ u_m \end{pmatrix} \right) \otimes \dots \otimes \left( \mathbf{F}_{x_1}^T \begin{pmatrix} 1 \\ x_1 \end{pmatrix} \right) \right) \quad (4)$$

holds, where  $\otimes$  denotes the (element-wise) Hadamard product.

So long, no assumptions have been made for the domains of states, inputs and parameters. Tensors can be defined for the domain  $\mathbb{R}$  of real values but the approach works similarly for the domain  $\mathbb{B}$  of Boolean values. The state transition function of a Boolean tensor state space model

$$\Phi(\underline{\mathbf{x}}) = \langle \underline{\mathbf{F}} | \underline{\mathbf{L}}(\underline{\mathbf{x}}, \underline{\mathbf{u}}) \rangle \quad (5)$$

uses the binary state  $\underline{\mathbf{x}} \in \mathbb{B}^N$  and input  $\underline{\mathbf{u}} \in \mathbb{B}^M$  and defines a Boolean transition tensor  $\underline{\mathbf{F}} \in \mathbb{B}^{\times(N+M)2 \times N}$ . It can describe any binary dynamical system by a contracted product with the Boolean literal tensor

$$\underline{\mathbf{L}}(\underline{\mathbf{x}}, \underline{\mathbf{u}}) = \left[ \begin{pmatrix} \underline{u}_m \\ \underline{u}_m \end{pmatrix}, \dots, \begin{pmatrix} \underline{u}_1 \\ \underline{u}_1 \end{pmatrix}, \begin{pmatrix} \underline{x}_n \\ \underline{x}_n \end{pmatrix}, \dots, \begin{pmatrix} \underline{x}_1 \\ \underline{x}_1 \end{pmatrix} \right] \in \mathbb{B}^{\times(N+M)2}. \quad (6)$$

given that  $\bar{x}$  denotes negation of  $x$ .

The Literal Tensor can also be defined over the real domain  $\mathbb{R}^{\times(N+M)2}$  as

$$\underline{\mathbf{L}}(\underline{\mathbf{x}}, \underline{\mathbf{u}}) = \left[ \begin{pmatrix} 1 - \underline{u}_m \\ \underline{u}_m \end{pmatrix}, \dots, \begin{pmatrix} 1 - \underline{u}_1 \\ \underline{u}_1 \end{pmatrix}, \begin{pmatrix} 1 - \underline{x}_n \\ \underline{x}_n \end{pmatrix}, \dots, \begin{pmatrix} 1 - \underline{x}_1 \\ \underline{x}_1 \end{pmatrix} \right],$$

which uses the undertilde symbol to indicate variables with continuous domain  $\mathbb{R}$ . It is easy to see that the basic properties of (6) still hold by mapping the Boolean FALSE to real zero and Boolean TRUE to real one.

If this mapping is done inside a contracted product (1), this is denoted by  $\langle \cdot | \cdot \rangle^+$  and the state transition of an *algebraic Boolean tensor state space model*

$$\Phi(\underline{\mathbf{x}}) = \left\langle \underline{\mathbf{F}} \left| \underline{\mathbf{L}}(\underline{\mathbf{x}}, \underline{\mathbf{u}}) \right. \right\rangle^+, \quad (7)$$

can be defined. Its state and input are from continuous domains, the transition tensor is Boolean, and the resulting next state is a continuous vector. If all initial states and inputs are Boolean the state trajectories only take values (0) and (1) and thus, show a behavior of the binary system.

Hybrid systems connect continuous-valued subsystems with discrete-valued subsystems in a predefined way. Without loss of generality, it is assumed that all discrete-valued signals are encoded Boolean. The hybrid state and input vectors

$$\mathbf{x} = \begin{pmatrix} \underline{\mathbf{x}} \\ \underline{\mathbf{x}} \end{pmatrix} \in \mathbb{R}^n \times \mathbb{B}^N, \quad \mathbf{u} = \begin{pmatrix} \underline{\mathbf{u}} \\ \underline{\mathbf{u}} \end{pmatrix} \in \mathbb{R}^m \times \mathbb{B}^M \quad (8)$$

are given by appending the Boolean signal vectors to the continuous ones forming hybrid domains  $\mathbb{H}$ .

Note, that any discrete-valued subsystem has an inherent multilinear structure, because any Boolean dynamics can be represented by a Tensor model (5). Combining this Boolean discrete-valued subsystem with a multilinear continuous-valued subsystem needs interfaces between the subsystems as shown by the block diagram in Fig. 1.

The continuous- and discrete-valued subsystems are connected via a quantizer and an injector given in the next definitions, see [7].

**Definition 4** The injector  $\alpha : \mathbb{H}^{I_1 \times \dots \times I_N} \rightarrow \mathbb{R}^{I_1 \times \dots \times I_N}$  is a function which is given for all elements with index vector  $\mathbf{i} \in \mathbb{N}^N$  by

$$(\alpha(\mathbf{x}))_{\mathbf{i}} = \begin{cases} 1 \in \mathbb{R} & \text{if } x_{\mathbf{i}} = \text{TRUE} , \\ 0 \in \mathbb{R} & \text{if } x_{\mathbf{i}} = \text{FALSE} , \\ x_{\mathbf{i}} & \text{if } x_{\mathbf{i}} \in \mathbb{R} . \end{cases} \quad (9)$$

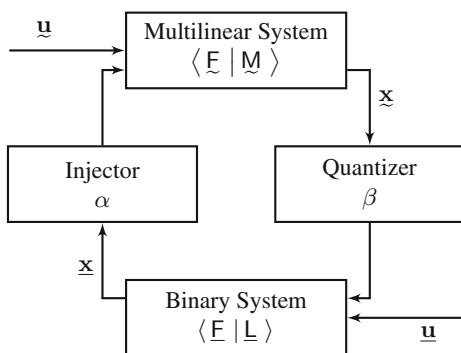
**Definition 5** The quantizer  $\beta : \mathbb{R}^{I_1 \times \dots \times I_N} \rightarrow \mathbb{B}^{I_1 \times \dots \times I_N}$  is a function which is given for all elements with index vector  $\mathbf{i} \in \mathbb{N}^N$  by

$$(\beta(\mathbf{x}))_{\mathbf{i}} = \sigma(x_{\mathbf{i}} - \frac{1}{2}) , \quad (10)$$

where the *Heaviside* function is given as

$$\sigma(x) = \begin{cases} 1 & \text{if } x \geq 0 , \\ 0 & \text{otherwise} . \end{cases} \quad (11)$$

**Fig. 1** Block diagram of a multilinear hybrid system



As both subsystems of the hybrid system are multilinear, a common representation can be found which leads to a hybrid tensor state space model

$$\Phi(\mathbf{x}) = \langle \mathbf{F} | \mathbf{M}(\mathbf{x}, \mathbf{u}) \rangle^{\boxplus}. \quad (12)$$

The *hybrid contracted tensor product* denoted by  $\langle \cdot | \cdot \rangle^{\boxplus}$  is a short notion for the application of injectors  $\alpha$  and standard quantizers  $\beta$  in a contracted tensor product such that all values are mapped to the correct domain of  $\Phi(\mathbf{x})$ . This will be illustrated by two application examples in the following sections.

### 3 Heating Systems Example

The heating system mentioned in the introduction will now be modeled as a hybrid tensor system. This is possible because of the inherit multilinear nature of heating systems. The heating system consists of a boiler with burner, a consumer, where a building temperature can be measured and a pump, which is controlled (see Fig. 2).

For each of the elements of the heating system, the time-discrete state space equations, which are based on thermal balances are stated. For the derivation of these equations see [11]. The factor matrices and the parameter vector of the model are given such that the system can be represented as a tensor system.

The continuous valued states of the system are the supply temperature  $T_s$ , the return temperature  $T_r$  and the buildings temperature  $T_b$ . Day oscillation and year oscillation of the ambient temperature are modeled by four continuous valued states  $T_{a1} \dots T_{a4}$ , two for each oscillation. Two Boolean states  $X_l$  and  $X_u$  determine whether or not the buildings temperature is in the interval where the thermostatic valves act linear, in the lower or in the upper saturation. Two further Boolean states  $H_l$  and  $H_u$  determine a supply temperature above an upper limit or below a lower limit and a Boolean state  $U$  that states if the boiler is running or not. The state vector is given by

$$\mathbf{x} = [T_s \ T_r \ T_b \ T_{a1} \ T_{a2} \ T_{a3} \ T_{a4} \ X_l \ X_u \ H_l \ H_u \ U]^T \in \mathbb{R}^7 \times \mathbb{B}^5. \quad (13)$$

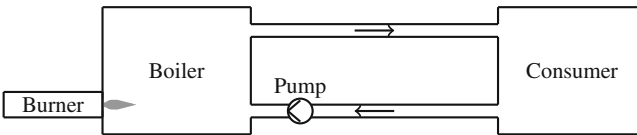
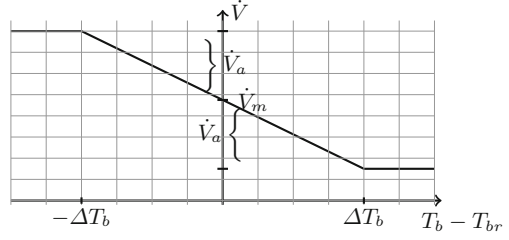


Fig. 2 Heating system

**Fig. 3** Flow rate over control error of the room temperature



### 3.1 Controlled Pump

The flow rate  $\dot{V}$  is a consequence of the room temperature  $T_b$  and the reference temperature for the room  $T_{br}$ . The relationship is shown in Fig. 3, where  $\dot{V}_m$  is the mean flow rate for  $T_b = T_{br}$  and  $\dot{V}_a$  is the amplitude from the mean flow rate at reference room temperature to its extrema. The extrema in the flow rate are reached at a room temperature that is at least  $\Delta T_b$  different from the reference room temperature.

To state the flow rate  $\dot{V}$  in terms of the discrete states and the room temperature, the discrete states

$$X_l = \begin{cases} 1 & \text{if } T_b > T_{br} - \Delta T_b \\ 0 & \text{else} \end{cases}, \quad X_u = \begin{cases} 1 & \text{if } T_b > T_{br} + \Delta T_b \\ 0 & \text{else} \end{cases} \quad (14)$$

are introduced. Thus the flow rate  $\dot{V}$  is given as

$$\dot{V} = \dot{V}_m + \dot{V}_a + \left( \frac{T_{br}\dot{V}_a}{\Delta T_b} - \dot{V}_a \right) X_l - \left( \frac{T_{br}\dot{V}_a}{\Delta T_b} + \dot{V}_a \right) X_u - \frac{\dot{V}_a}{\Delta T_b} T_b X_l + \frac{\dot{V}_a}{\Delta T_b} T_b X_u. \quad (15)$$

### 3.2 Boiler

The state equation of the supply temperature is

$$T_s(k+1) = T_s(k) - \frac{t_s}{V_b} T_s(k) \dot{V}(k) + \frac{t_s}{V_b} T_r(k) \dot{V}(k) + \frac{t_s}{V_b \rho c} P_{max} U(k). \quad (16)$$

Dropping the time index, denoting the next state by the operator  $\Phi$  and substituting (15) for the flow rate leads to

$$\begin{aligned}
\Phi(T_s) = & T_s \left( 1 - \frac{t_s(\dot{V}_a + \dot{V}_m)}{V_b} \right) + T_r \frac{t_s(\dot{V}_a + \dot{V}_m)}{V_b} + T_s X_l \frac{\dot{V}_a t_s (\Delta T_b - T_{br})}{\Delta T_b V_b} \dots \\
& - T_r X_l \frac{\dot{V}_a t_s (\Delta T_b - T_{br})}{\Delta T_b V_b} + T_b T_s X_l \frac{\dot{V}_a t_s}{\Delta T_b V_b} - T_b T_r X_l \frac{\dot{V}_a t_s}{\Delta T_b V_b} + T_s X_u \frac{\dot{V}_a t_s (\Delta T_b + T_{br})}{\Delta T_b V_b} \dots \\
& - T_r X_u \frac{\dot{V}_a t_s (\Delta T_b + T_{br})}{\Delta T_b V_b} - T_b T_s X_u \frac{\dot{V}_a t_s}{\Delta T_b V_b} + T_b T_r X_u \frac{\dot{V}_a t_s}{\Delta T_b V_b} + P_{max} U \frac{t_s}{V_b c \rho}.
\end{aligned} \tag{17}$$

### 3.3 Ambient Temperature

The ambient temperature is modeled as an oscillator with two frequencies. One for the temperature oscillation over the day and one for the oscillation of the year. The sinusoidal curves estimate the ambient temperature roughly.

The state equation in discrete time with sampling time  $t_s$  is given by

$$\begin{aligned}
\Phi \left( \begin{pmatrix} T_{a1} \\ T_{a2} \\ T_{a3} \\ T_{a4} \end{pmatrix} \right) &= \begin{pmatrix} 1 & t_s & 0 & 0 \\ -\omega_d^2 & 1 & 0 & 0 \\ 0 & 0 & 1 & t_s \\ 0 & 0 & -\omega_y^2 & 1 \end{pmatrix} \begin{pmatrix} T_{a1} \\ T_{a2} \\ T_{a3} \\ T_{a4} \end{pmatrix}, \quad T_a \\
&= (1 \quad 0 \quad 1 \quad 0) \begin{pmatrix} T_{a1} \\ T_{a2} \\ T_{a3} \\ T_{a4} \end{pmatrix} + T_{a,m}.
\end{aligned}$$

Note that a multilinear state space model allows a constant term, here  $T_{a,m}$ .

### 3.4 Building

The temperature of the building  $T_b$  is introduced as the average temperature of all rooms. The state equation of the building temperature can be given as

$$\Phi(T_b) = T_b \left( 1 - \frac{t_s(k_{r,b} + k_{b,a})}{k_b} \right) + T_r \frac{t_s k_{r,b}}{k_b} + T_{a1} \frac{t_s k_{b,a}}{k_b} + T_{a3} \frac{t_s k_{b,a}}{k_b} + \frac{t_s k_{b,a} T_{a,m}}{k_b},$$

where  $k_b$  is the capacity of the building,  $k_{r,b}$  the heat transfer coefficient from the radiators to the building and  $k_{b,a}$  the heat transfer coefficient from the building to the outside.

### 3.5 Consumer

The heat demand  $\dot{Q}_d$  can be calculated in terms of the return temperature and the building temperature as

$$\dot{Q}_d = (T_r - T_b)k_{r,b} . \quad (18)$$

Using (18) and (15) the state equation of the return temperature reads

$$\begin{aligned} \Phi(T_r) = & T_s \frac{t_s(\dot{V}_a + \dot{V}_m)}{V_c} + T_r \left( 1 - \frac{t_s(k_{r,b} + c\dot{V}_a\rho + c\dot{V}_m\rho)}{V_c c\rho} \right) \dots \\ & + T_b \frac{k_{r,b}t_s}{V_c c\rho} - T_s X_l \frac{\dot{V}_{at_s}(\Delta T_b - T_{br})}{\Delta T_b V_c} + T_r X_l \frac{\dot{V}_{at_s}(\Delta T_b - T_{br})}{\Delta T_b V_c} \dots \\ & - T_b T_s X_l \frac{\dot{V}_{at_s}}{\Delta T_b V_c} + T_b T_r X_l \frac{\dot{V}_{at_s}}{\Delta T_b V_c} - T_s X_u \frac{\dot{V}_{at_s}(\Delta T_b + T_{br})}{\Delta T_b V_c} \dots \\ & + T_r X_u \frac{\dot{V}_{at_s}(\Delta T_b + T_{br})}{\Delta T_b V_c} + T_b T_s X_u \frac{\dot{V}_{at_s}}{\Delta T_b V_c} - T_b T_r X_u \frac{\dot{V}_{at_s}}{\Delta T_b V_c} . \end{aligned} \quad (19)$$

### 3.6 Bang Bang Controller

The discrete states  $H_l$  and  $H_u$  are used to determine whether or not the supply temperature  $T_s$  is below a lower limit, here: 75 °C or above an upper limit, here: 95 °C. The signal  $U$  steering the boiler is then computed by

$$\Phi(U) = \begin{cases} U & \text{if } 75^\circ\text{C} < T_s < 95^\circ\text{C} \\ 1 & \text{if } T_s \leq 75^\circ\text{C} \\ 0 & \text{if } T_s \geq 95^\circ\text{C} \end{cases} . \quad (20)$$

The model of the building has been validated, see [11]. Also, there are given all parameters of the model. With these difference equations it is now straight forward to construct the system tensor. Here just the first entries of the system tensors factor matrices, that correspond to the next state of the supply temperature  $T_s$  are given. The factor matrices, in order of the states (13) are



$$\begin{aligned}
\mathbf{F}_{T_s} &= \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & \dots \\ 1 & 0 & 1 & 0 & 1 & \dots \end{pmatrix}, \mathbf{F}_{T_r} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & \dots \\ 0 & 1 & 0 & 1 & 0 & \dots \end{pmatrix}, \mathbf{F}_{T_b} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & \dots \\ 0 & 0 & 0 & 0 & 1 & \dots \end{pmatrix} \\
\mathbf{F}_{T_{a1}} &= \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots \end{pmatrix}, \mathbf{F}_{T_{a2}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots \end{pmatrix}, \mathbf{F}_{T_{a3}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots \end{pmatrix} \\
\mathbf{F}_{T_{a4}} &= \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots \end{pmatrix}, \mathbf{F}_{T_{X_1}} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & \dots \\ 0 & 0 & 1 & 1 & 1 & \dots \end{pmatrix}, \mathbf{F}_{T_{X_u}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots \end{pmatrix} \\
\mathbf{F}_{T_{H_1}} &= \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots \end{pmatrix}, \mathbf{F}_{T_{H_u}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots \end{pmatrix}, \mathbf{F}_u = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots \end{pmatrix}
\end{aligned}$$

Note, that the factor matrices are not complete but just given for the first five summands of the first next state. The corresponding part of the matrix  $\mathbf{F}_\phi$  and the parameter vector  $\lambda_f$  are given by

$$\mathbf{F}_\phi = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \lambda_f = \begin{pmatrix} 1 - \frac{t_s(\dot{V}_a + \dot{V}_m)}{V_b} \\ \frac{t_s(\dot{V}_a + \dot{V}_m)}{V_b} \\ \frac{\dot{V}_a t_s (\Delta T_b - T_{br})}{\Delta T_b V_b} \\ - \frac{\dot{V}_a t_s (\Delta T_b - T_{br})}{\Delta T_b V_b} \\ \frac{\dot{V}_a t_s}{\Delta T_b V_b} \\ \vdots \end{pmatrix}. \quad (21)$$

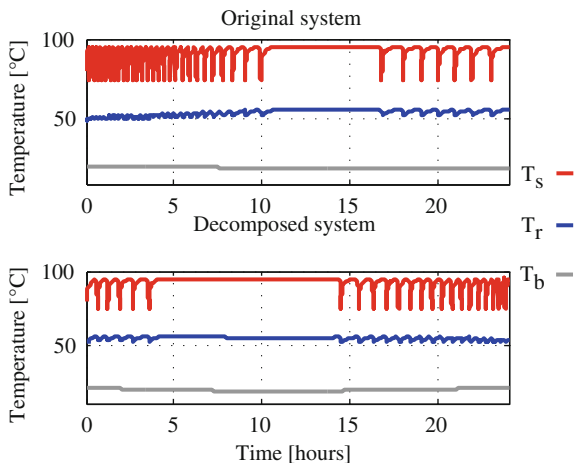
The matrix  $\mathbf{F}_\phi$  has dimension  $12 \times 46$  to assign the 46 summands of the system to the corresponding next state. Since all entries that were looked at correspond to the first next state, the first five entries of the first row of  $\mathbf{F}_\phi$  are 1.

In the following a reduction of the tensor rank is performed using tensor decomposition techniques. With the derived factor matrices and the parameter vector a state transition tensor

$$\mathbf{F} = [\mathbf{F}_{T_s}, \mathbf{F}_{T_r}, \mathbf{F}_{T_b}, \mathbf{F}_{T_{a1}}, \mathbf{F}_{T_{a2}}, \mathbf{F}_{T_{a3}}, \mathbf{F}_{T_{a4}}, \mathbf{F}_{X_1}, \mathbf{F}_{X_u}, \mathbf{F}_{H_1}, \mathbf{F}_{H_u}, \mathbf{F}_U, \mathbf{F}_\phi] \cdot \lambda_f$$

is constructed. The Tensor Toolbox [1] and the command `cp_als` is used to reduce the rank of the tensor  $\mathbf{F}$ , which was constructed to have a tensor rank of 46. Using the Toolbox, the rank of the system tensor is reduced to 23, which has a fit of over 99 %. The first five entries of the first two factor matrices are

$$\begin{aligned}
\mathbf{F}_1 &= \begin{pmatrix} -0.999 & 0.999 & 1.000 & 1.000 & 1.000 & \dots \\ 0.045 & -0.054 & 0.009 & 0.007 & 0.000 & \dots \end{pmatrix}, \\
\mathbf{F}_2 &= \begin{pmatrix} 0.988 & 0.984 & 0.999 & 1.000 & 1.000 & \dots \\ -0.152 & -0.179 & 0.040 & 0.031 & 0.000 & \dots \end{pmatrix}. \quad (22)
\end{aligned}$$

**Fig. 4** Original and decomposed system

The corresponding entries in the parameter vector are

$$\lambda = (10.443 \quad 6.967 \quad 4.322 \quad 2.598 \quad 2.587 \quad \dots)^T. \quad (23)$$

The results of the simulation with the original and the decomposed state transition tensor are given in Fig. 4. One can see, that the main system dynamics are captured with the decomposed system. The supply temperature drops to 75 °C if 95 °C is reached and rises again. Also the offset and the dynamics of the return temperature show similar behavior for both state transition functions.

Note that for numerical reasons, the states have been transformed to be in the interval  $[0, 0.5]$ , see [11] for details.

## 4 Multi-agent Systems Example

A MAS consists of a group of agents (e.g. vehicles with their computing entities) that exchange information and share knowledge to solve a common task [2]. The demand of autonomous MAS's with the ability to perform tasks as Decision-Making or Policy Formulation leads to hybrid MAS's. There are various different applications of hybrid MAS's and many different ways to describe them, e.g. [15]. How a hybrid MAS is modeled as tensor system is shown in the following.

A directed Graph  $\mathcal{G}(\mathcal{V}, \mathcal{E})$  consists of a set of  $N$  vertices, representing the single agents, and a set of edges  $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ , describing the communication topology of a MAS, see [9]. If agent  $i$  receives information from agent  $k$ , there is an edge  $(k, i) \in \mathcal{E}$ . The normalized Laplacian of a graph is [13]

$$\mathcal{L}_{ik} := \begin{cases} 1 & \text{if } i = k \text{ and } |\mathbf{N}_i| \neq 0 \\ -\frac{1}{|\mathbf{N}_i|} & \text{if } k \in \mathbf{N}_i \\ 0 & \text{otherwise.} \end{cases} \quad (24)$$

Here  $\mathbf{N}_i$  is the set of neighbors of agent  $i$ , that is defined as  $\mathbf{N}_i = \{k | (k, i) \in \mathcal{E}\}$ , and  $|\mathbf{N}_i|$  is its cardinality. Consider a MAS of  $N$  equal agents, where each agent  $i = 1 \dots N$  is described by the switched system

$$\begin{aligned} \Phi(x_i) &= ax_i + u_i \text{ if } \Phi^{-1}(x_i) \geq 0.5 \\ \Phi(x_i) &= bx_i + u_i \text{ if } \Phi^{-1}(x_i) < 0.5, \end{aligned} \quad (25)$$

with  $x_i, u_i \in \mathbb{R}$  being continuous signals and  $\Phi^{-1}(x_i) = x_i(k-1)$  denoting the previous state. By introducing the discrete signal  $z_i = \beta(\Phi^{-1}(x_i))$ , being the quantized state variable, system (25) can be written as the multilinear system

$$\Phi(x_i) = (a-b)z_i x_i + bx_i + u_i. \quad (26)$$

This can be described as a tensor system for one agent  $i$  by

$$\Phi(x_i) = \langle F^i | M(x, z, u) \rangle = \langle [\mathbf{F}_u^i, \mathbf{F}_z^i, \mathbf{F}_x^i] | M(x, z, u) \rangle \quad (27)$$

$$\text{with } \mathbf{F}_u^i = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \mathbf{F}_z^i = \begin{pmatrix} b & 1 \\ a-b & 0 \end{pmatrix}, \mathbf{F}_x^i = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The control input is determined by the communication topology as  $\mathbf{u} = -\mathcal{L}\mathbf{x}$  like simple consensus algorithms propose [9] with  $\mathbf{x} = (x_1 \dots x_N)^T$ ,  $\mathbf{u}$  and  $\mathbf{z}$  defined respectively. The discrete system is described by

$$\Phi(\mathbf{z}) = \beta(\mathbf{x}). \quad (28)$$

The continuous part can be described as a continuous tensor system of the form  $\Phi(\mathbf{x}) = \langle \mathbf{F}\mathbf{M}(\mathbf{x}, \mathbf{z}) \rangle^+$ . It is obvious that the continuous system only contains multilinear terms of the form  $x_i z_i$  introduced by the system equations of the single agents in (26), whereas the control input is linear in  $\mathbf{x}$ . Thus the standard monomial tensor of dimension  $\mathbb{R}^{\times(2N)2}$  is very sparse. A special formation monomial tensor is introduced here as

$$\mathbf{M}^F(\mathbf{x}, \mathbf{z}) = \left[ \begin{pmatrix} 1 & \dots & 1 \\ z_1 & & z_N \end{pmatrix}, \begin{pmatrix} 1 & \dots & 1 \\ x_1 & & x_N \end{pmatrix}, \mathbf{I}_N \right] \in \mathbb{H}^{2 \times 2 \times N}. \quad (29)$$

With (29) a state transition function of a hybrid tensor system for the overall system, with state vector  $\mathbf{h} = (\mathbf{x}, \mathbf{z})^T$  is defined as

$$\Phi(\mathbf{h}) = \langle \mathbf{F}^F | \mathbf{M}^F(\mathbf{x}, \mathbf{z}) \rangle^{\boxplus} \quad (30)$$

with  $\mathbf{F}^F = [\mathbf{F}_z^F, \mathbf{F}_x^F, \mathbf{F}_I^F, \mathbf{F}_\Phi^F] \in \mathbb{R}^{2 \times 2 \times N \times 2N}$  having the factor matrices consisting of  $2N$  factors

$$\begin{aligned} \mathbf{F}_\Phi^F &= \begin{pmatrix} \mathbf{I}_N & \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_N \end{pmatrix}, & \mathbf{F}_I^F &= (\mathbf{I}_N - \mathcal{L}\mathbf{I}_N), \\ \mathbf{F}_x^F &= \left( \mathbf{1}_{3N}^T \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right), & \mathbf{F}_z^F &= \left( \mathbf{1}_N^T \begin{pmatrix} b \\ a-b \end{pmatrix} \quad \mathbf{1}_{2N}^T \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right). \end{aligned} \quad (31)$$

First the continuous state vector  $\mathbf{x}$  is considered. Due to the identity matrices in  $\mathbf{F}_\Phi^F$  and  $\mathbf{F}_I^F$ , the first  $N$  columns of the factor matrices describes the internal dynamics of the agents. This can easily be seen by comparing the first  $N$  columns of  $\mathbf{F}_x^F$  and  $\mathbf{F}_z^F$  with  $\mathbf{F}_x^i$  and  $\mathbf{F}_z^i$  in (27). The Laplacian in  $\mathbf{F}_I^F$  describes the communication between the different agents. The fact, that the communication is linear in the states  $x^i$ , explains the second  $N$  vectors in  $\mathbf{F}_x^F$  and  $\mathbf{F}_z^F$  since

$$\left\langle \left[ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right], \left[ \begin{pmatrix} 1 \\ z_i \end{pmatrix}, \begin{pmatrix} 1 \\ x_i \end{pmatrix} \right] \right\rangle = x_i.$$

The discrete signal  $\mathbf{z}$  is simply calculated by  $\Phi(\mathbf{z}) = \beta(\mathbf{x})$ , which explains the choice of the last  $N$  columns in  $\mathbf{F}_x^F$  and  $\mathbf{F}_z^F$ .

Next, a hybrid tensor system with  $N = 100$  agents is modeled. The considered connected graph is directed and contains 473 random edges. The internal dynamics of the single agents is described by (26) with  $a = 0.5$  and  $b = 1$ . The number of factors of the exact system tensor is  $300 = 3N$ . CP decompositions of the system tensor from rank 1 to 299 have been calculated with `cp_als`. The Frobenius error between the exact and approximated hybrid tensor model is shown for tensor approximations of different rank in Fig. 5. A tensor of rank  $200 = 2N$  describes the system dynamics almost exactly. The Frobenius error for graphs of every tested different size shows qualitatively the same shape. Thus it can be supposed that the system is in fact not a tensor system of rank  $3N$  but of  $2N$ . In Fig. 6, continuous state trajectories of the rank 200 approximation are shown. The maximal mean

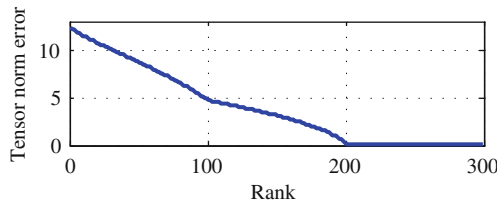
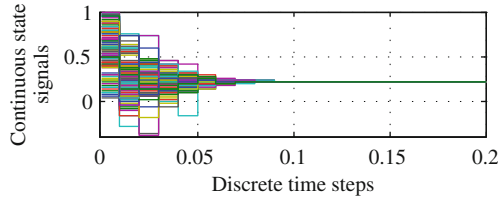
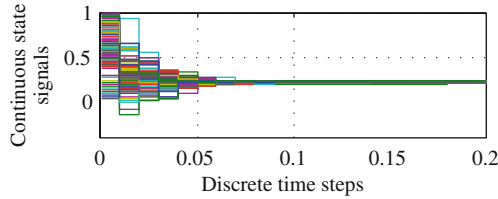


Fig. 5 Frobenius norm of error



**Fig. 6** Continuous valued state trajectories for rank 200 approximation



**Fig. 7** Continuous valued state trajectories for rank 150 approximation

squared error between the original and approximated trajectories for a rank 200 approximation is with  $<10^{-17}$  negligible and thus approximation and original system match exactly. For comparison the rank 150 approximation is shown on Fig. 7. Here the same consensus value of approximately 0.2 is reached with the same convergence shape. However, the states does not exactly converge to the real consensus value but reach an envelope around it, which gets small for higher order approximations. For approximations with smaller order instability of the convergence process may occur.

## 5 Conclusions

Hybrid tensor systems are an adequate framework for modeling discrete time multilinear systems with continuous and discrete valued states and inputs. A hybrid tensor model of a complex heating system is derived as a real-world example. All parameters are stored in a state transition tensor which in a second step is reduced to a low rank CP tensor using standard decomposition techniques. As simulations show, this decomposed tensor is still capable to capture the main dynamics of the heating system.

As a second application example from a quite different application domain, a hybrid tensor model of a Multi-Agent System (MAS) is derived. Structural constraints are imposed in an easy way, a reduced rank state transition tensor of the system is again computed by tensor decomposition algorithms and simulations are carried out based on the decomposed model. The reduced rank model converges to

the same final values as the original one. Moreover, the original  $N$ -agents system can be modeled exactly by a hybrid tensor system with a small rank of  $2N$ .

Further research will be done to investigate the stability behaviour of hybrid tensor systems. Another focus will be the derivation of tensor representations for nonlinear normal forms—as well for general nonlinear as for multilinear systems. Tensor decomposition techniques will play an important role in these fields and extensions of the continuous algorithms to hybrid spaces would be essential tools for analysis and design of multilinear hybrid systems.

**Acknowledgments** This work was partly supported by the project ModQS of the Federal Ministry of Economics and Technology, Germany.

## References

1. Bader B, Kolda T (2012) MATLAB tensor toolbox version 2.5. Available online
2. Balaji P, Srinivasan D (2010) An introduction to multi-agent systems, vol 310 of Studies in computational intelligence, Chapter 1, pp 1–27. Springer
3. Cichocki A, Zdunek R, Phan A, Amari S (2009) Nonnegative matrix and tensor factorizations. Wiley, Chichester
4. Hackbusch W (2012) Tensor spaces and numerical tensor calculus, vol 42 of Springer series in computational mathematics. Springer, Berlin Heidelberg
5. Kolda T, Bader B (2009) Tensor decompositions and applications. *SIAM Rev* 51(3):455–500
6. Lichtenberg G (2010) Tensor representation of boolean functions and Zhegalkin polynomials. In: International workshop on tensor decompositions
7. Lichtenberg G (2011) Hybrid tensor systems. Hamburg University of Technology, Habilitation
8. Lichtenberg G, Eichler A (2011) Multilinear algebraic boolean modelling with tensor decompositions techniques. In: 18th IFAC world congress, p TuC01.2
9. Mesbahi M, Egerstedt M (2010) Graph theoretic methods for multiagent networks. Princeton University Press, Princeton
10. Noudui TS, Zuo KPW, Wetter M (2012) Validation and application of the room model of the Modelica buildings library. In: 9th international Modelica conference, pp 727–736
11. Pangalos G, Eichler A, Lichtenberg G (2013) Tensor systems: multilinear modeling and applications. In: 3rd international conference on simulation and modeling methodologies, technologies and applications
12. Pangalos G, Lichtenberg G (2012) Approach to boolean controller design by Algebraic Relaxation for heating systems. In: 4th IFAC conference on analysis and design of hybrid systems
13. Ren W, Beard R (2007) Distributed consensus in multi-vehicle cooperative control—theory and applications. Communications and control engineering. Springer Publishing Company, Incorporated, New York
14. Ren W, Cao Y (2011) Distributed coordination of multi-agent networks. Springer, London Limited
15. Srinivasan D, Choy M (2010) Hybrid multi-agent systems, vol 310 of Studies in computational intelligence, Chapter 2, pp 29–42. Springer
16. Wetter M (2006) Multizone building model for thermal building simulation in Modelica. In: 5th international Modelica conference, pp 517–526

# Modeling Interdependent Socio-technical Networks: The Smart Grid—An Agent-Based Modeling Approach

Daniël Worm, David Langley and Julianna Becker

**Abstract** The aim of this paper is to improve scientific modeling of interdependent socio-technical networks. In these networks the interplay between technical or infrastructural elements on the one hand and social and behavioral aspects on the other hand, plays an important role. Examples include electricity networks, financial networks, residential choice networks. We propose an Agent-Based Modeling approach to simulate interdependent technical and social network behavior, the effects of potential policy measures and the societal impact when disturbances occur, where we focus on a specific use case: the smart grid, an intelligent system for matching supply and demand of electricity.

**Keywords** Agent based modeling · Interdependent socio-technical networks · Smart grid

## 1 Introduction

Rapidly increasing computer processing power means that computer modeling techniques now enable scientists to attempt to couple different models together and investigate the effects of their interdependency. The objective of this paper is to improve scientific modeling of interdependent socio-technical networks. This is important in the field of designing critical infrastructures. Failures in these systems are rare events which may have catastrophic consequences. Society requires resilient

---

D. Worm (✉) · J. Becker  
TNO, Delft, The Netherlands  
e-mail: daniel.worm@tno.nl

J. Becker  
e-mail: julianna.becker@tno.nl

D. Langley  
TNO, Groningen, The Netherlands  
e-mail: david.langley@tno.nl

infrastructure which can cope with a wide variety of threats. Examples of failures include natural disasters like Hurricane Sandy, and technical failures like cable burnout in the energy network in Germany which has a highly distributed renewable energy production. According to the German Federal Network Agency, at the end of March 2013 the electricity network threatened to collapse: “The security of the network can no longer be guaranteed. [...] We have had to intervene more than forty times to prevent surges in wind and solar power from compromising the entire electricity system. The stress generated by these situations is becoming increasingly difficult to handle.”

Since Holling’s [15] seminal work on the resilience of systems, engineering scientists have endeavored to design critical infrastructures capable of coping with disturbances [6, 22]. However, social components are often missing in models of critical infrastructure. This is a problem for two reasons. First, human behavior can influence the system, and thus the likelihood of failure. Second, effects of disturbances have human dimensions, whereby strategic decisions can best take account of the disruption that people experience (and the perceived effect thereof) rather than use solely technical parameters. This provides our motivation for this research into interdependent socio-technical modeling.

In this study we propose an Agent-Based Model (ABM) approach to simulate interdependent technical and social network behavior, the effects of potential policy measures and the societal impact when disturbances occur. The use of individual or agent based approaches are common in the study of complex adaptive systems, especially where the interactions between the agents are complex, nonlinear, discontinuous, or discrete, where the population is heterogeneous and where the topology of the interactions is heterogeneous and complex [7]. This applies increasingly to networks, whether physical or social. Using ABM, its structure and behavior have potential to resemble reality better than simple mathematical models, especially when the underlying real relationships are complex [29].

In order to obtain our objective, we focus on a specific use case: the smart grid, a future intelligent system that helps to match demand and supply of electricity in a sustainable and secure manner. In such a system, both social and technical aspects play an important role. The model we obtain for this use case, and that we will describe in this paper, helps to give insight in certain effects arising from the interplay between these aspects. Furthermore, from it we obtain generic insights into interdependent socio-technical network modeling, contributing to our main objective.

## 2 Related Work

The topic of simulation models for interdependent socio-technical networks is receiving attention in a wide range of scientific domains. This development is based on the enormous amount of data related to social, economic, technology and biological networks, which is increasingly available for research, as well as readily



accessible computing power for carrying out the necessary computations [16, 18, 19, 26]. We briefly address a number of the most relevant streams of literature.

In the field of resilience engineering a main focus is on the effects of natural disasters on a range of infrastructural networks. There is a general recognition that interdependencies between networks are both an important driver of cascading failures and a significant modeling challenge [27]. Recent work is including the ‘human factor’ as one of the interdependent networks, in recognition of the importance of modeling the socio-technical system as a whole, e.g. Johnsen and Veen’s assessment [17] of the key communication infrastructure used in emergency communication in railways in Norway, although this is not yet widespread practice.

A second relevant scientific domain is the sociology of the housing market, where methods are developed for analyzing housing price dynamics [11], urban sprawl and individual choices about where to live, and the implications of these choices for residential patterns [10]. Individual choices respond to the relative attractiveness of residential areas, but they also change that attractiveness [8]. ABM have been used to model these choices [5, 21] including the interdependencies of different market segments, such as racial residential segregation [34].

Finally, in direct relation to the case study we address in this paper is the smart electricity grid. Much literature on this topic which implements ABM is focused on multi-agent systems to control distributed smart grid technology, rather than simulate the socio-technical networks including household choices. Studies which do include human behavior include simulating load profiles for households equipped with smart appliances under conditions of real-time variable-price tariffs [13], and micro-level models of household capacity adaptation allowing for occupants to vary their achieved comfort by foregoing electricity when the price is too high [14]. Whether such behavior is realistic in the real world has yet to be demonstrated. ABMs of the smart grid demonstrate herding behavior where many agents independently converge their loads the time intervals they expect to have lower prices, thus leading to undesirable load peaks which can cause network failure [25]. To prevent such herding behavior developing, simulations have shown that introducing inertia can help, for example by imposing penalties for deviation from past behavior [32] or more complex algorithms for spreading load across a number of expected future low-price time intervals [28].

The model we present in this paper builds on the work from these scientific domains, adding particularly to the theoretical grounding of the social model from psychology as a way of improving the combined socio-technical approach.

### 3 Objectives

Our aim is to model, in a quantitative manner, interdependent socio-technical networks and the effects that failure cascades can have. Of key importance is the link between infrastructural, (technical) networks and human behavior. In this paper we focus on the smart grid case and in future research we consider other cases and

attempt to uncover generic elements one should take into account when modeling socio-technical interdependent networks and their societal impact.

These models need not be highly accurate at this stage, but they should be able to generate the types of network behaviors arising from the interdependency between the social and technical systems based on the characteristics of the different networks and on potential policy interventions.

Our research questions read:

- How can we model cascading effects between a technical and a social network model, whereby changes and disturbances in a technical network affect human behavior and vice versa?
- How can we model the societal impact of these mutually interacting networks, both in hard, financial terms and in a soft, reputational sense?

In the long term our objectives are twofold. First, to examine to what extent (re) routing /steering /consolidating human behavior is possible when a disturbance in a network occurs. Second, to examine how policy interventions can influence failure cascades between interdependent networks so as to minimize negative societal impact.

## 4 Socio-technical Smart Grid Model

In this use case we model a future residential-level electricity network, including smart grid elements. A smart grid is an electricity network that intelligently reacts to the behavior of different stakeholders, such as generators, consumers and those that do both, with the aim of efficiently supplying sustainable, economic and secure electricity and coping with disruptions [9]. An important element in achieving this is flexible pricing, triggering adaptive consumer behavior.

This use case is relevant in relation to our objectives described in Sect. 3, because of the strong interdependency between the technical and the human element in this socio-technical system. The behavior of consumers plays a key role in the performance of the future electricity network, since this behavior directly determines electricity demand and decentralized supply, which then affects the pressures placed on the physical electricity network.

We choose to model this human element at the individual level, rather than the aggregated level, so that we can include heterogeneous effects per household, such as the price each pays, the comfort (i.e. the fulfilled demand) and power failures each experiences, as well as the peer influence working via the social network. Therefore, the electricity network at the residential level (low voltage) is relevant to our purposes, although the results can be extrapolated to the neighborhood and regional level.

### 4.1 Description of Model Framework

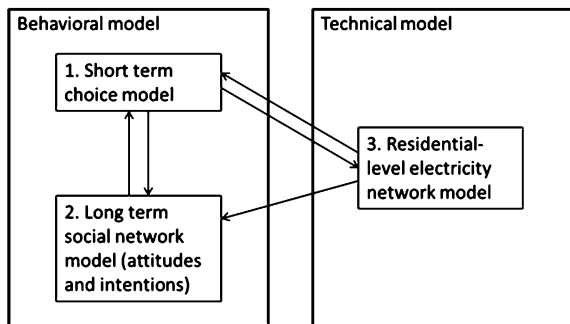
To model the interaction between a residential-level electricity network and human behavior, we made three separate models. These models interact with each other as shown by Fig. 1. The behavioral model is split up into two models:

1. *Short Term Choice Model* This model covers the short term choices consumers make based on their electricity needs and fluctuating electricity prices. We assume a power-management application adapts demand real-time and that the consumers can choose one of three profiles: maximum comfort (electricity is used irrespective of the price), medium comfort (a price cap is selected but only for a limited time) and minimum comfort (a price cap is selected and usage is halted above that price). Besides this, consumers can choose two other one-off measures: to install a solar panel and to insulate their home. Time steps in this model are intervals of 15 min and the model calculates how much electricity each household demands (or supplies) per time step.
2. *Long Term Behavioral Model* This model determines the attitudes and behavioral intentions of the consumers, which in turn influence their behavioral choices in the short term model. We model five attitudes which are influenced by both that household’s own experiences and by the attitudes of others in their social network.

The five attitudes are: Attitude about price paid for electricity, attitude about personal comfort (i.e. the willingness to forego electricity), attitude about personal energy efficiency, attitude about renewable energy production, and attitude about confidence in the electricity supply. These attitudes are continuous variables with value between 0 and 1. Time steps in this model are days, weeks, or months (set by an adjustable parameter) and the model calculates the five attitudes per household and what this means for their behavioral intention.

This model is based on the psychological Theory of Reasoned Action [1, 2, 4, 12] which states that behavioral intention is driven by attitude and social influence. Social influence is the person’s perception that most people who are important to him think he should or should not perform the behavior in question. Later theories of

**Fig. 1** Graphical overview of the relationships between the three models



social influence go beyond this normative pressure to include other forms of influence, such as imitation [20]. As for the link between behavioral intention and actual behavior, in a meta-analysis of 87 studies, [30] reports an average correlation between intention and behavior to be 0.53, which means that on average consumers' answers to questions about their intentions account for only 28 % of the variance in their actual behavior. For low-involvement products, such as electricity, this link may be even weaker [24]. Therefore, we introduce a probability for linking behavioral intention in the long term model to choice behavior in the short term model.

We do not include the extended Theory of Planned Behavior, which includes the role of perceived behavioral control, as the behaviors in our model are well within the behavioral control of the agents [1].

Finally we have a technical network model:

3. *Electrical Network Model* This model computes power flows in a residential-level. (Low voltage) network, based on demand and supply. It also determines if and where disruptions occur in the electrical network, for example if the supply in a given part of the electricity network exceeds the demand whereby a physical cable burns through. Time steps in this model are intervals of 15 min.

These models are connected as follows: Each 15 min interval, the short term choice model is executed, computing the demand (or supply) of each consumer, based on the price at that moment (which in turn is influenced by total demand and supply), their comfort profile, their devices needing electricity, their insulation level and the production of their solar panel, if applicable. The output is passed to the electrical network model, that determines how the demand is met and if any disruptions in the network occur. This is communicated back to the short term choice model, because disruptions affect the remaining demand of each consumer.

This process continues until one time step of the long term model has been reached. Then the long term behavioral model is executed, taking into account output from both the short term behavioral model as well as the electrical network model over the past time.

## 4.2 Model Specifics

### 4.2.1 Agents with Their Social Network

The agents in our model are 208 households in a fictional residential area, divided into 13 streets. Each household has a number of electrical devices that require different amounts of energy and have different time windows within a day in which the demand of the device should be fulfilled.

The agents are linked with each other via a social network ('friends'), which is randomly drawn via the following principles: The number of friends of each household is Poisson distributed with mean  $\lambda$ , and distributed in such a way that

two households in the same street ('neighbors') are  $n$  times more likely to be friends than two households in different streets. (We chose  $\lambda = 8$  and  $n = 4$ ) This social network will influence the agents' attitudes. We randomly divide the agents into three different types, which fixes initial attitudes of the agents: Comfort (willing to pay for high comfort), Budget (wants to pay as little as possible), Eco (aims towards sustainable energy).

#### 4.2.2 Technical Network

The electricity model consists of a network with 14 nodes, taken from an actual low-voltage network. One of the nodes is the main generator that connects the low-voltage network to the medium-voltage network and thus supplies all demanded electricity which the households do not produce themselves via solar panels. The model computes the power flows over each link needed to fulfill the demands, based on DC power flow methods [33]. Each link is endowed with a maximum capacity which, if crossed, will cause the link to break, leading to rerouted power flows which may cause new failures in turn and possibly derive households of electricity.

The behavioral and technical models connect via the electricity demands of the agents. The demand and supply within a street (16 households per street, 13 streets in total) are aggregated and communicated as input to one of the nodes in the electrical network. In turn, failures in the electrical network influence the behavioral model, by changing attitudes due to unfulfilled electricity demands.

#### 4.2.3 Implementation and Verification

The two behavioral models were implemented in Repast Symphony 2.0 Beta, a java-based toolkit for agent-based modeling and simulation [23]. For the technical electricity network model, an existing load balancing low-voltage model is used, which has been implemented in a MATLAB package called MATPOWER [35]. The Java package MatlabControl enables the connection between the different models.

In order to verify if the implemented models correspond with our conceptual design and work in the desired way, we have followed the verification process proposed by Rand and Rust [26], which includes documentation, programmatic testing, and test cases. This process was essential and helped in identifying incorrectly implemented components, which have a higher risk of occurring in these complex interdependent networks. One issue which we experienced in the verification process is interesting to note: that some of the proposed verification steps are more challenging to carry out in the case of modeling interdependent networks. For example, one of the test case approaches these authors recommend is the use of corner cases, whereby extreme values are used as inputs and the behavior of the model is examined for unexpected output. However, due to the interdependencies

incorporated into our model, interpreting the output of corner cases is non-trivial. A suggestion to deal with this is to take not only output variables, but also internal variables (that are not directly visible in the output of the model) to check in relation with the input parameters or other internal variables. The main reason for this is that in these complex socio-technical models it may not always be straightforward to see what the intended effect between one input and one output variable should be, given that relations may be very complex. Cutting this ‘problem’ into manageable pieces by considering a sequence of internal variables where intended effect of one on the other *is* known, may simplify the procedure.

### 4.3 Scenario Analysis

In order to address our research questions we ran a number of scenario’s whereby different conditions were assessed. We highlight a number of the most interesting results here.

#### 4.3.1 Crossover Effects

We ran the model in a ‘default’ setting (Fig. 2) and in a setting where network cables are more likely to fail (Fig. 3), in order to investigate crossover effects from the technical model into the behavioral model.

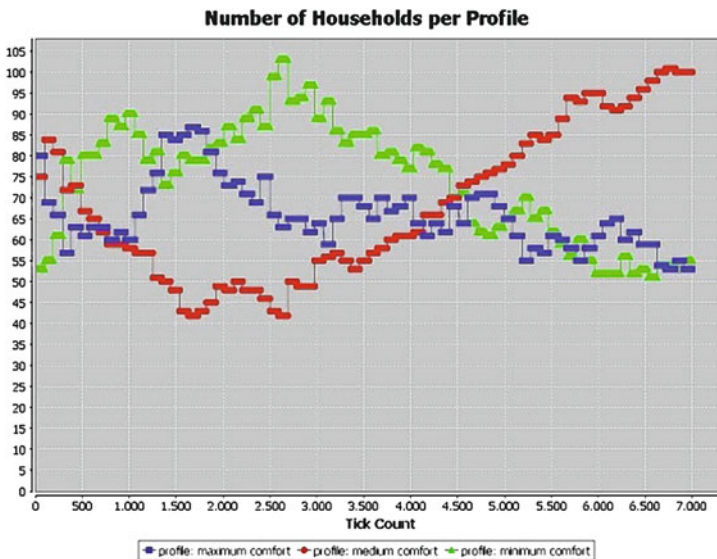


Fig. 2 Comfort profiles in the default scenario

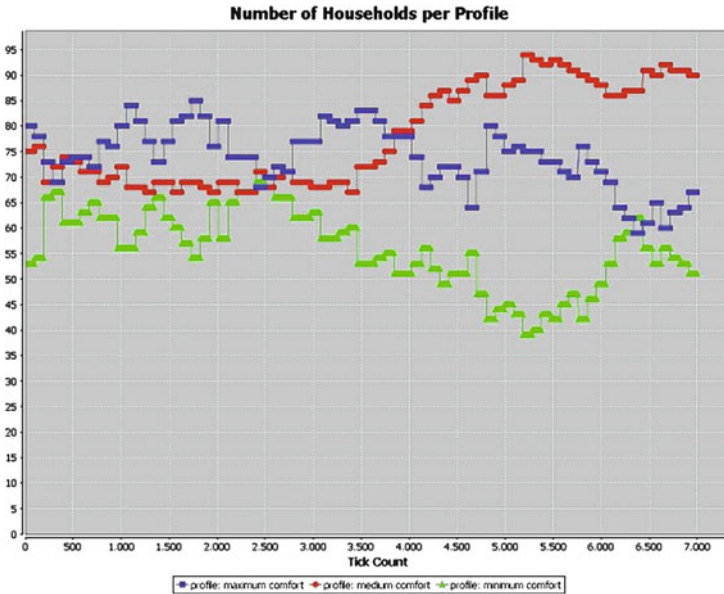


Fig. 3 Comfort profiles in the scenario with a more fragile electricity network

The figures show the dynamics in the behavioral model regarding agents choices for the different power-management comfort profiles. We see a clear distinction between these two cases: in the more fragile network setting, the minimum comfort profile is less popular than in the default setting, and there is an increase in the maximum comfort profiles. A reason for this is that disruptions lead to less fulfilled electricity demands than usual, causing more people to wish higher comfort. This in turn may cause even more disruptions in the network, due to increased demand, which shows a crossover effect from the behavioral model back into the social model.

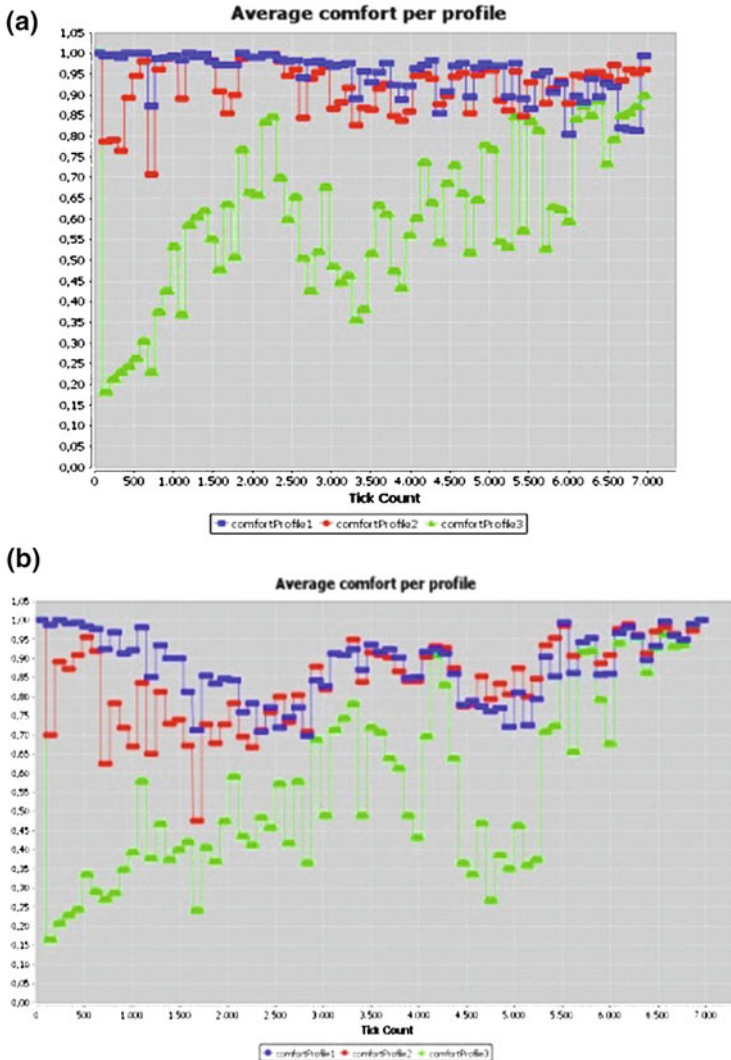
### 4.3.2 Policy Interventions

We can use the smart grid model to investigate the effects of policy interventions. For instance, by increasing solar panel subsidies, making it easier for agents to install their own energy producing devices. Using our model it is possible to investigate the effects of different levels of subsidy. To begin with we see an increase in agents selecting the intermediate comfort profile. The increase in solar panels means that the agents’ energy needs are more easily fulfilled (so their comfort is sufficient enough), and energy prices for them will drop also due to solar panels (so the price is cheap enough). However, after some time the level of solar power production reaches a level where it begins to have a negative effect on the stability of the network. On sunny days, a concentrated area with many solar panels



can add to the network load and thereby cause more power failures. When solar panel installation reaches very high levels, this disruption can lead to many consumers' energy demand not being fulfilled (see Fig. 4).

Therefore we see that well-intentioned policy can also have an unwanted effect on the electrical network: more disruptions occur in the network compared to the default scenario (where hardly any disruptions occur), throughout the timeline.



**Fig. 4** Effects of increasing subsidy for solar panels on the level of comfort that agents achieve. Subsidy is five (a) and ten (b) times higher than the default scenario. (note when subsidy levels increase to a high level, network overloads lead to many power outages, thereby reducing the amount of demand which is satisfied)



### 4.3.3 Societal Effects

Another crucial aspect in the smart grid case is the impact that disruptions have on society, in ‘hard’ financial terms as well as a ‘soft’ reputational sense, like trust. There are several ways to look at financial impact. In [3] several formulas are derived for financial impact for individual households and companies based on both frequency and duration of disruptions. Both of these may differ per agent in our model, so applying the formulas give insight in e.g. the variability of financial impact in a residential area, which turns out to be quite high in a scenario with many disruptions.

Trust (in the electricity system) is more difficult to measure in real life. Surveys can help to give an estimation for trust. In our model, we use the variable attitude about confidence in the electricity supply as a measure for trust. We use this to assess the relationship between fraction of disturbances and trust levels, in particular the impact trust has on the operation of the network. The nature of the model will reflect a level of distrust in the network when there are more failures due to the behavioral aspects built into the modeling. Therefore the impact looks at the relationship between the two in terms of what happens to the fraction of disruptions as distrust increases and how do agents adapt to this.

For this analysis a conditional probability was used focusing on the probability of failure given that there is high distrust in the network,  $P(F|D > 0.5)$ , compared to the overall probability of failure,  $P(F)$ . Looking at a run with many failures we found that the probability of failure is higher when there are high levels of distrust in the network, as would be expected. However, this also suggests something on the behavioral impact of these failures: When there is lower trust in the network agents are more likely to demand energy whenever they have access to it, as if there were a sense of urgency to use the energy before it goes out again as opposed to behaving in an energy efficient way to safeguard their energy levels (for instance by adapting a maximum comfort profile). The relationship between trust and behavior in this model implies a more irrational actor when trust is lost, increasing the probability of a network failure which would only perpetuate the cycle as represented in the Fig. 5.

However, we also encounter other scenarios where a loss of trust in the network occurs at a certain point in time, but where the system was able to overcome that to provide stable energy supply. These types of scenarios are interesting to model in terms of exploring alternative scenarios to restore or redirect trust.

Overall, the societal impact of the smart grid can be modeled to show how disruptions affect agents under various scenarios, and, in turn, to see how this influences the behavior of agents. This provides a foundation for further exploration into these interactions.

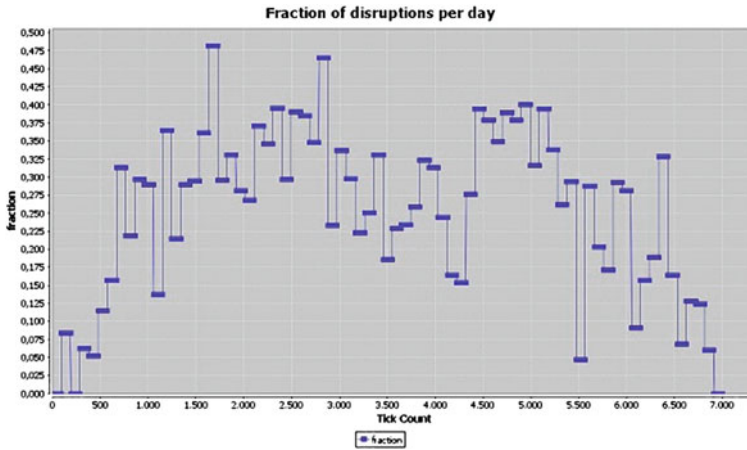


Fig. 5 Fraction of disruptions per day in extreme case with many disruptions

## 5 Conclusions and Future Research

Some observations can be made in the smart grid setting:

- A consumer's individual actions, e.g. to compensate for a fragile network, may cause a worsening effect on system level, in the end causing more damage for the individual.
- Policies with good intentions (e.g. subsidizing increased solar production) may lead to unwanted effects (disruptions in the network).
- Possibly, other pricing strategies might enable policy makers to better obtain the effect they want (e.g., a stable network).

Sensitivity analysis can be applied to gain more insight in the effects of (combinations of) parameters on the outcome.

The ABM approach seems suitable to investigate interdependency between social and technical networks. It allows to observe unforeseen (possibly unwanted) effects arising from this interdependency and certain policy interventions. It also allows to investigate impact on society. The java-based toolkit Repast Symphony is flexible for this purpose, and allows for connections with other programming languages, which is useful for embedding a specific technical model into a social/behavioral ABM.

Our case highlights the need for a multi-disciplinary approach to using ABM for socio-technical networks. Each research domain has its own ontology which typically does not readily combine with that from other [31]. For example, the concepts and entities generally included in system models of technical electricity grids are incompatible with social psychological models of human behavior. And yet we attempt to combine both ontologies in a single ABM.

An essential next step to take is validation of a socio-technical network. Because we need to make many simplifications (compared to reality) in both the social and the technical network model, the question is whether the combined model actually fits reality reasonably well. If unforeseen events arise from the socio-technical model, one would like to know if these events are plausible in reality or come from an oversimplification or wrong specification of the model. It should be stressed that our aim is not to create perfect accurate predictive models at this phase; instead we would like to use our models to generate the types of network behaviors arising from the interdependency between the social and technical systems based on the characteristics of the different networks and on potential policy interventions. In the smart grid case, the setting is futuristic, therefore we had to use fictional data and could not directly validate the complete system, though we need to take further steps in this direction.

Another relevant research direction is balancing the level of required detail or complexity in both the social and the technical network models, in order to make them fit together best.

For both themes our future research focuses on obtaining guidelines that are as generic as possible, i.e. that should be applicable also to other socio-technical networks. We aim to obtain these goals through the study of different use cases, like residential choice models and financial networks.

## References

1. Ajzen I (1991) The theory of planned behavior. *Organ Behav Hum Decis Process* 50 (2):179–211
2. Ajzen I, Fishbein M (1980) *Understanding attitudes and predicting social behavior*. Prentice-Hall, Englewood Cliffs
3. Baarsma B, Berkhout P, Hop P (2004) *Op prijs gesteld, maar ook op kwaliteit*. Technisch rapport, SEO, Amsterdam
4. Bagozzi RP (1992) The Self-Regulation of Attitudes. *Intent Behav Soc Psychol Q* 55 (2):178–204
5. Benenson I (2004) Agent-based modeling: from individual residential choice to urban residential dynamics, in *spatially integrated social science: examples in best practice*. Oxford Univ. Press, Oxford
6. Boin A, McConnell A (2007) Preparing for critical infrastructure breakdowns: the limits of crisis management and the need for resilience. *J Conting Crisis Manag* 15(1):50–59
7. Bonabeau E (2002) Agent-based modeling: methods and techniques for simulating human systems. *PNAS* 99:7280–7287
8. Bruch E, Mare D (2012) Residential mobility, and neighborhood change. *Sociol Methodol* 42:103–154
9. Clastres C (2011) Smart grids: another step towards competition, energy security and climate change objectives. *Energy Policy* 39(9):5399–5408
10. Devisch O, Timmermans H, Arentze T, Borgers A (2009) An agent-based model of residential choice dynamics in nonstationary housing markets. *Environ Plan A* 41(8):1997–2013
11. Erlingsson E, Raberto M, Stefánsson H, Sturluson J (2013) Integrating the housing market into an agent-based economic model. *Manag Mark Complex Lect Notes Econ Math Syst* 662:65–76

12. Fishbein M, Ajzen I (1975) *Belief, attitude, intention, and behavior: an introduction to theory and research*. Addison-Wesley, Reading
13. Gottwalt S, Ketter W, Block C, Collins J, Weinhardt C (2011) Demand side management—a simulation of household behavior under variable prices. *Energy Policy* 39:8163–8174
14. Guo Y, Li R, Poulton G, Zeman A (2008) A simulator for self-adaptive energy demand management. In: *IEEE conference on self-adaptive and self-organizing systems*
15. Holling C (1973) Resilience and stability of ecological systems. *Annu Rev Ecol Syst* 4:1–23
16. Jackson M (2008) *Social and economic networks (economics, physics, sociology)*. Princeton University Press, New Jersey
17. Johnsen S (2013) Veen, M: Risk assessment and resilience of critical communication infrastructure in railways. *Cogn Technol Work* 15(1):95–107
18. Khanin R, Wit E (2006) How scale-free are biological networks. *J Comput Biol* 13:810–818
19. Kleinberg J (2008) The convergence of social and technological networks. *Commun ACM* 51:66–72
20. Langley DJ, Bijmolt THA, Ortt JR, Pals N (2012) Determinants of social contagion during new product adoption. *J Prod Innov Manage* 29(4):623–638
21. Macy M, Wilier R (2002) From factors to actors: computational sociology and agent-based modeling. *Annu Rev Sociol* 28:143–166
22. McDaniel T, Chang S, Cole D, Mikawoz J, Longstaff H (2008) Fostering resilience to extreme events within infrastructure systems: characterizing decision contexts for mitigation and adaptation. *Glob Environ Change* 18(2):310–318
23. North MJ, Tatara E, Collier NT, Ozik J (2007) Visual agent-based model development with repast symphony. In: *Agent 2007 conference: complex interaction and social emergence*, Argonne National Laboratory, Argonne, pp 173–192
24. Quester P, Lim AL (2003) Product involvement/ brand loyalty: is there a link? *J Prod Brand Manag* 12(1):22–38
25. Ramchurn S, Vytelingum P, Rogers A, Jennings N (2011) Agent-based control for decentralised demand side management in the smart grid. In: *International conference on autonomous agents and multiagent systems*
26. Rand W, Rust RT (2011) Agent-based modeling in marketing: guidelines for rigor. *Int J Res Mark* 28:181–193
27. Reed D, Kapur K, Christie R (2009) Methodology for assessing the resilience of networked infrastructure. *IEEE Syst J* 3(2):174–180
28. Reddy P, Veloso M (2012) Factored models for multiscale decision-making in smart grid customers. In: *Proceedings of the twenty-sixth AAAI conference on artificial intelligence*, pp 363–369
29. Remondino M (2004) *Analysis of agent based paradigms for complex social systems simulation*, Ph. D. Thesis, Università di Torino (2004)
30. Sheppard B, Hartwick J, Warshaw P (1988) The theory of reasoned action—a meta-analysis of past research with recommendations for modifications and future-research. *J Consum Res* 15 (3):325–343
31. van Dam K, Nikolic I, Lukszo Z (2012) *Agent-based modelling of socio-technical systems*. Springer, Dordrecht
32. Voice T, Vytelingum P, Ramchurn S, Rogers A, Jennings N (2011) Decentralised control of micro-storage in the smart grid. In: *AAAI conference on artificial intelligence (AAAI-11)*
33. Wood A, Wollenberg B (1996) *Power generation, operation, and control*, 2nd edn. Wiley, New York
34. Zhang J (2004) A dynamic model of residential segregation. *J Math Sociol* 28:147–170
35. Zimmerman R, Murillo-Sánchez C, Thomas R (2011) MATPOWER: steady-state operations, planning and analysis tools for power systems research and education. *IEEE Trans Power Syst* 26(1):12–19

# A Heuristic Bidding Price Decision Algorithm Based on Cost Estimation Accuracy Under Limited Engineering Man-Hours in EPC Projects

Nobuaki Ishii, Yuichi Takano and Masaaki Muraki

**Abstract** In this paper, we develop a heuristic bidding price decision algorithm in consideration of cost estimation accuracy under limited engineering Man-Hours (MH) in Engineering, Procurement, Construction (EPC) projects. It allocates engineering MH for cost estimation, which determines the cost estimation accuracy, to each order under the limited volume of MH, and then determines the bidding price for maximizing the expected profit based on cost estimation accuracy under the deficit order probability constraint. Numerical examples show that the bidding price decision in consideration of cost estimation accuracy and deficit order probability is essential for the contractor in making a stable profit in EPC projects, and that the developed algorithm is effective for making such bidding price decision.

**Keywords** Competitive bidding · Cost estimation accuracy · EPC contract · Project management

## 1 Introduction

Among various types of project contracts, the importance of Engineering, Procurement, Construction (EPC) projects [14], where contractors design and build unique products or services based on the client requirements, is widely recognized in practice [15] in the field of construction, civil engineering, plant engineering, and so on. In EPC projects, the contractor has a single responsibility for project cost,

---

N. Ishii (✉)

Faculty of Information and Communications, Bunkyo University, Kanagawa, Japan  
e-mail: ishii@shonan.bunkyo.ac.jp

Y. Takano

School of Network and Information, Senshu University, Kanagawa, Japan  
e-mail: ytakano@isc.senshu-u.ac.jp

M. Muraki

Graduate School of Decision Science and Technology, Tokyo Institute of Technology, Tokyo, Japan  
e-mail: m.muraki8511@gmail.com

quality, and schedule under a fixed-price that is determined before the start of the project as a lump-sum contract. Thus, a reduced project cost and shorter schedule are expected [8].

Although several shortcomings, for instance, decisions on relatively detailed issues have to be made early on in the project delivery process, have been pointed out e.g., in Elfving et al. [3], competitive bidding is widely used for selecting a contractor who carries out the project. In the competitive bidding, the client usually evaluates contractors on the basis of the multi-attribute bid evaluation criteria, such as bidding price, past experience, past performance, company reputation, and the proposed method of delivery and technical solutions [21]. Then, the client basically selects the contractor who proposes the lowest price if there is not much difference in other criteria.

In EPC projects, accordingly, it is necessary for any contractor to determine the bidding price based on precise cost estimation. If the contractor's bidding price, which is obtained as a sum of the estimated cost and the target profit, is higher than that of the competitor due to cost estimation error, then the contractor could not accept the order and hence obtain no profit. In contrast, the contractor would increase the chance of accepting the order if the estimated cost is low due to cost estimation error. In this case, however, the profit could be below the contractor's expectation because of being over-budget, and he possibly suffers a loss on this order.

Namely, for stable profit from EPC projects, the contractor must determine the bidding price in consideration of cost estimation accuracy and deficit order probability. Cost estimation, however, is a highly intellectual task of predicting the costs of products or services to be provided in the future based on the analysis of the client's requirements and his tacit knowledge. Thus, experienced and skilled human resources, represented as engineering Man-Hours (hereafter referred to as MH), are required for accurate cost estimation. Those resources, however, are limited for any contractor. For these reasons, it is important to realize appropriate allocation of MH for cost estimation to each order to maximize the profits under the constraints on the volume of total MH. In addition, contractors should consider the possibility of realizing a loss due to cost estimation error and a competitive relationship with bidders. For example, the bidding price needs to be cut to some extent to accept the order successfully under a severe competitive environment; however, a low bidding price would reduce profit, or even worse, would create a large loss. Moreover, just a few deficit orders would result in the significant reduction of realized profits when the number of accepted orders is limited. (Note, in this paper, that we refer to the order creating an eventual loss as a deficit order).

In this paper, we develop a heuristic bidding price decision algorithm in consideration of cost estimation accuracy and deficit order probability under limited MH in EPC projects. The algorithm assumes that the costs are estimated at the same time for all orders. At the first step, the algorithm allocates MH for cost estimation to each order according to the ranking of orders under the constraints on the volume of total MH. The MH allocation determines the cost estimation accuracy of each

order. At the second step, it determines the bidding price for maximizing the expected profit based on cost estimation accuracy under the deficit order probability constraint.

We develop a mathematical model for simulating competitive bidding. Through the numerical results obtained by using this model, we show that the bidding price decision in consideration of cost estimation accuracy and deficit order probability is essential for the contractor in making a stable profit in EPC projects, and that our heuristic bidding price decision algorithm is effective for making such bidding price decisions.

## 2 Related Work

A variety of studies, such as bidding theory, bidding model, and auction design, have been conducted on competitive bidding [1]. In particular, a number of papers regarding the competitive bidding strategy date back to Friedman [4], who presented a method to determine an optimal bidding price based on the distribution of the ratio of the bidding price to cost estimate. However, little attention has been paid to profit volatility risk, which cannot be ignored in EPC projects. When, for instance, the number of accepted orders is limited, the realized total profit from the projects might be sharply lower than expected because the profit is significantly affected by a few deficit orders. Accordingly, the deficit order probability should be considered in the bidding price decision.

In addition to the profit volatility risk, we consider the allocation of MH for cost estimation to each order when making a decision on the bidding because certain MH is necessary to estimate cost accurately in EPC projects. Several papers have analysed the problem of allocating scarce resources in competitive bidding (see Rothkopf and Harstad [16] for detailed references). Among them, Kortanek et al. [10] considered sequential bidding models where the obtained contracts require the use of restricted resources, such as production capacity, at the time of actual production. Ishii et al. [7] developed an order acceptance strategy under limited MH. Takano et al. [18] considered the sequential bidding models where the obtained contracts require the use of restricted MH. However, the allocation of limited MH for cost estimation to each order, which affects the expected profits from orders significantly, has not been investigated in those studies.

Regarding cost estimation accuracy, various types of research have been performed. Oberlender and Trost [13] studied determinants of cost estimation accuracy and developed a system for predicting cost estimation accuracy. Bertisen and Davis [2] analysed costs of 63 projects and evaluated the accuracy of capital cost estimation statistically. In addition, several researchers have studied cost estimation methods and their accuracy. For example, Towler and Sinnott [19] studied relations among cost estimation methods, cost estimation data, and their accuracy in the field of plant engineering. More crucially, they suggested that the cost estimation accuracy is positively correlated with the volume of MH for cost estimation.

In EPC projects, the bidding price decision affects the expected profit and the deficit order probability. Since the bidding price is determined based on the project cost estimated before starting the project, cost estimation accuracy is clearly a major factor to lead an EPC project to a successful conclusion. Nevertheless, as stated above, few studies have ever attempted to analyse the bidding price decision problem in terms of cost estimation accuracy and deficit order probability under limited MH in EPC projects.

### **3 Features of the Bidding Price Decision Problem in EPC Project**

There are several ways to select a contractor from bidders in competitive bidding [3, 5, 17, 20]. In a generic competitive bidding process [6], the client prepares a Request For Proposal (RFP) and invites several potential contractors to the bidding. The contractor first carries out the preliminary evaluation followed by the bid or no-bid decision. In the preliminary evaluation, the contractor evaluates the RFP and estimates the preliminary cost based on limited information, such as the order information provided by the RFP and the past project data of the contractor. In the bid or no-bid decision, the contractor evaluates the order from the viewpoints of profitability, technical feasibility and so on, and makes a decision whether to bid or not. If the contractor decides to place the bid, he then starts the bidding price decision process, that is, he estimates the cost more accurately and determines the bidding price. At the end of the competitive bidding, the client assesses the proposals offered by contractors and selects one contractor as a successful bidder.

The bidding price decision, for which this paper develops an algorithm in Sect. 4, is made based on order information, such as estimated cost, target profit rate, and competitive environment, so that the contractor can accept profit-making orders successfully. Since the contractor must determine the bidding price using the limited information above, he should consider the following features of the competitive bidding.

The first feature is relevant to the accuracy of cost estimation. The bidding price is basically determined by adding the target profit to the estimated cost. However, the contractor cannot estimate the precise cost in the process of determining the bidding price because of limited information and restricted time. Thus the bidding price, which is affected by estimation errors, has a probability distribution. We define the cost estimation accuracy as the standard deviation of the estimated cost or the bidding price depending on the context. A lower deviation indicates a higher accuracy.

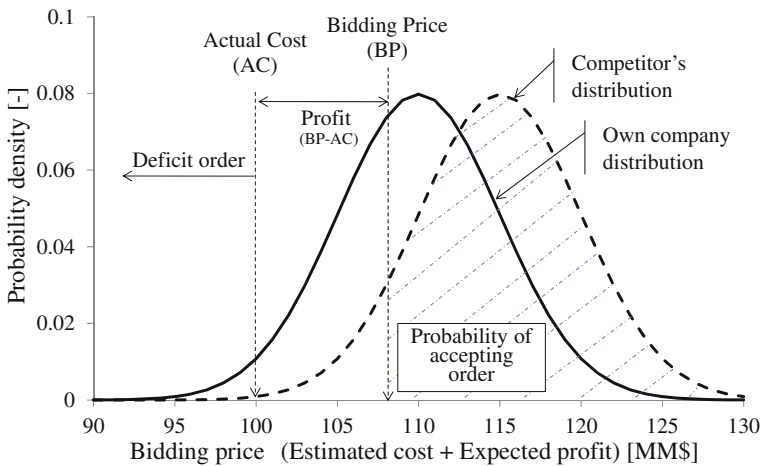
The bidding price with the lower cost estimation accuracy is likely to be accepted as the deficit order, from which the contractor suffers an eventual loss. The bidding price with the low accuracy also has a tendency to be very high compared to the other; however, the chance of the order being accepted becomes smaller as



the bidding price increases under a competitive environment where many competitors would offer low bidding prices. Based on these observations, it can be seen that consideration of cost estimation accuracy and deficit order probability is essential for the contractor in making a stable profit in EPC projects, and the bidding price decision process needs to include all these factors.

The second feature is the MH allocation for cost estimation. Cost estimation is a series of activities where experienced engineers analyse the client’s requirements, propose solutions, make a preliminary design, and estimate the required volume of works and quantities of materials needed to carry out the project based on the design. Thus, the substantial volume of MH for cost estimation is required to estimate the cost accurately [6]. However, the contractor often has more than one order at the same time, and the volume of MH of experienced engineers is limited. Namely, the contractor needs to allocate MH to each order appropriately for cost estimation. Since the bidding conditions are different in each order, the contractor needs to prioritize orders and allocate more MH to the potential orders to improve the total expected profit from orders.

The third feature is the effectiveness of adjusting the bidding price. Figure 1 shows bidding price distribution of a contractor and that of a competitor. As shown in Fig. 1, the contractor’s profit (= BP-AC) increases as the bidding price rises. On the other hand, the probability of accepting the order, shown as dashed lines, increases as the bidding price goes down. This is because the contractor can basically accept the order when the contractor’s bidding price (BP) is lower than that of the competitor’s. However, the contractor would accept the deficit order when the bidding price is very low. Namely, we can see that there is a bidding price that maximizes the contractor’s expected profit under a competitive environment.



**Fig. 1** An example of bidding price distributions at competitive bidding. (The contractor’s own company: average bidding price (AC + Profit (10 % of AC)): 110 [MM\$], cost estimation accuracy: 5 [%]; Competitor: Average bidding price: 115 [MM\$], cost estimation accuracy: 5 [%])

Based on the above observations, we introduce a parameter for adjusting the bidding price in view of the cost estimation accuracy of one’s own company and that of a competitor’s, as well as the deficit order probability.

### 4 A Bidding Price Decision Process Model

Figure 2 shows a bidding price decision process model, which represents fundamental factors and their interactive processes, to determine the bidding price in EPC projects based on the observations in the previous section. The model consists of three kinds of factors, i.e., decision processes, constraints, and given conditions.

The model enables us to evaluate the expected orders, the expected profits, and the deficit order probability, based on the bidding price, the cost estimation accuracy, and the information on competitive environment. The bidding price is determined based on the estimated cost, the target profit rate, and the risk parameter for adjusting the bidding price. The estimated cost and the cost estimation accuracy are both determined depending on the MH allocated to each order for cost estimation. The MH allocation is determined according to the ranking of orders provided by the pre-evaluation of orders processed under the total MH constraint as shown in Fig. 2.

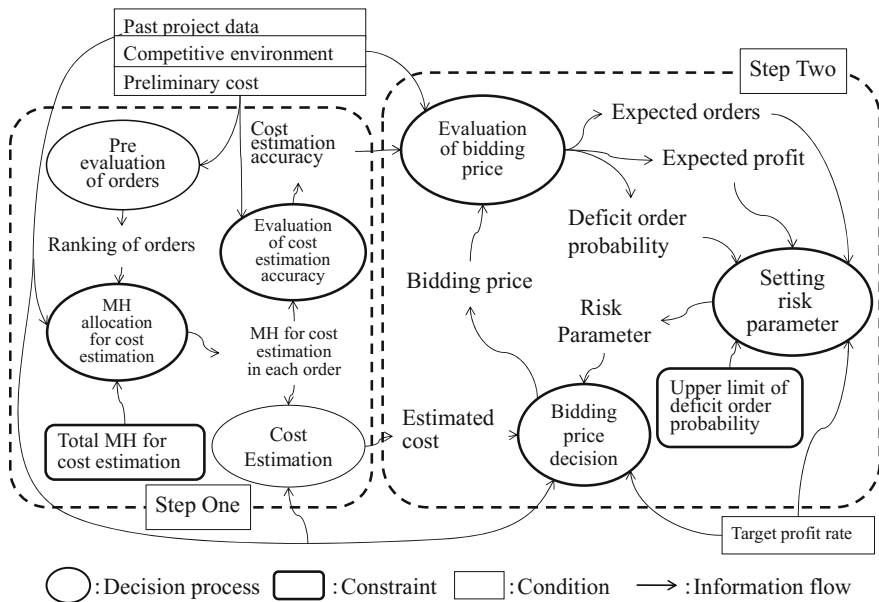


Fig. 2 A bidding price decision process model in EPC projects

### 4.1 A Mathematical Model on Bidding Price Decision

**Evaluation of Cost Estimation Accuracy.** Since cost estimation requires a detailed analysis conducted by experienced engineers, it can be seen that the MH for cost estimation significantly affects the cost estimation accuracy. In fact, Towler and Sinnott [19] suggest that the cost estimation accuracy is positively correlated with the volume of MH for cost estimation. It is also clear that the marginal rate of cost estimation accuracy approaches zero according to the increase of the volume of MH. Thus, in this paper, we define the cost estimation accuracy ( $\sigma$ ) as the function of the MH for cost estimation per order ( $PMH$ ) based on the logistic curve [6] as follows:

$$\sigma(PMH) = \sigma_{\min} \cdot \sigma_{\max} / \{ \sigma_{\max} + (\sigma_{\min} - \sigma_{\max}) \cdot e^{-C \cdot PMH} \} \quad (PMH > 0.0) \quad (1)$$

where  $\sigma_{\min}$  and  $\sigma_{\max}$  are the minimum and the maximum value of the standard deviation of the bidding price, and  $C$  is a parameter of the logistic curve. In practice, the contractor could determine these parameters from past project data.

**Evaluation of Bidding Price.** In the model, we consider  $n$  contractors ( $k = 1, 2, \dots, n$ ) and the bidding for  $L$  orders ( $i = 1, 2, \dots, L$ ). Particularly,  $k = 1$  represents one's own company, and  $k \geq 2$  are its competitors. In the model, based on standard order cost ( $STD$ ), each contractor ( $k$ ) sets a tentative bidding price ( $TBP$ ) of the order ( $i$ ) in consideration of the relative cost difference from  $STD$  ( $RC$ ) and target profit rate ( $e\_profit$ ) as follows:

$$TBP_k^i = STD_i \cdot (1 + RC_k^i) \cdot (1 + e\_profit_k^i) \cdot rp_k^i \quad (2)$$

where  $TBP$  can be adjusted by changing the value of risk parameter ( $rp$ ). If there is no difference in cost-competitiveness among contractors,  $RC$  is set to 0.

The expected volume of order ( $i$ ) in one's own company ( $k = 1$ ) is as follows:

$$\int_0^{+\infty} x_1^i \cdot p_1(x_1^i, TBP_k^i, \sigma_1^i) \cdot \prod_{k=2}^n \int_{x_1^i}^{+\infty} p_k(x_k^i, TBP_k^i, \sigma_k^i) dx_k^i \cdot dx_1^i \quad (3)$$

where  $p_k(x_k^i, TBP_k^i, \sigma_k^i)$  is the probability density function of the bidding price ( $x_k^i$ ) of the contractor ( $k$ ) for order ( $i$ ), and its average value and standard deviation are  $TBP_k^i$  and  $\sigma_k^i$ , respectively. As shown in Eq. (3), the expected order is the average value of one's own bidding price falling below those of all other contractors ( $k \geq 2$ ).

As shown in Eq. (4), the expected profit is the average excess of the bidding price over the standard order cost with the relative cost difference ( $STDR$ ) as defined in Eq. (5).

$$\int_0^{+\infty} (x_1^i - STDR_1^i) \cdot p_1(x_1^i, TBP_1^i, \sigma_1^i) \cdot \prod_{k=2}^n \int_{x_1^i}^{+\infty} p_k(x_k^i, TBP_k^i, \sigma_k^i) dx_k^i \cdot dx_1^i \quad (4)$$

$$STDR_k^i = STD_i \cdot (1 + RC_k^i) \quad (5)$$

In addition, as shown in Eq. (6), the deficit order probability is the probability of accepting the order whose bidding price is lower than  $STDR$ .

$$\int_0^{STDR_1^i} p_1(x_1^i, TBP_1^i, \sigma_1^i) \cdot \prod_{k=2}^n \int_{x_1^i}^{+\infty} p_k(x_k^i, TBP_k^i, \sigma_k^i) dx_k^i \cdot dx_1^i \quad (6)$$

We also assume that the data used in the above equations, such as the number of competitors ( $n-1$ ), standard order cost ( $STD$ ), relative cost difference over  $STD$  ( $RC$ ), probability density function of bidding price ( $p_k$ ), and so on, can be provided from RFP, past project data, several departments of the contractor, and published data. For example,  $STD$  can be specified in reference to the preliminary cost, which is estimated by scaling it from the cost data of past projects, which used similar technology [9]. Although a project is a temporary endeavor undertaken to create a unique product, similar parts can be found in functional units of past projects. Accordingly, even if the cost data of similar projects are not available, the preliminary cost estimate can be made by breaking down the project into functional units, and adding up the cost data of similar functional units in past projects. The cost data, the number of competitors, and so on, can also be estimated based on published data in many industries. For example, magazines related to the EPC business, such as Chemical Engineering, Hydrocarbon Processing, publish plant cost indexes, cost engineering data, EPC project news and surveys, periodically.

## 4.2 A Heuristic Bidding Price Decision Algorithm

In this section, we develop a heuristic algorithm for bidding price decision. As shown in Fig. 2, this algorithm determines the allocation of MH for cost estimation according to the ranking of orders at the first step, and searches the value of  $rp$  for maximizing the expected profit of each order under the deficit order probability constraint at the second step.

*Step One: Ranking of Orders and MH Allocation.* There are several procedures to rank orders. For example, pair-wise comparisons, scoring models, and analytical hierarchy process (AHP) are commonly used [11].

In this paper, we shall rank orders based only on the expected profit so as to assess the effectiveness of our algorithm from the perspective of profits. Specifically, we define the ranking score (*Score*) of the order ( $i$ ) as the expected profit based on the tentative bidding price ( $TBP_i$ ) estimated for the ranking at  $rp = 1$  as follows:

$$Score_i = (TBPFI_1^i - STDR_1^i) \cdot \prod_{k=2}^n \int_{TBPFI_k^i}^{+\infty} p_k(x_k^i, TBPFI_k^i, \sigma_k^i) dx_k^i \tag{7}$$

$$TBPFI_k^i = STD_i \cdot (1 + RC_k^i) \cdot (1 + e\_profit_k^i) \tag{8}$$

Note that we can modify the ranking score in consideration of multiple criteria besides the expected profit, such as technical feasibility, relationship with clients, and so on.

In the following MH allocation procedure, the order with the high *Score* is ranked high because such an order is expected to generate a large profit.

As described in the procedure below, we consider three grades of accuracy, A (high accuracy), B (average), and C (low accuracy), and we assign one of them to each order. The expected profit increases according to the increase of cost estimation accuracy, and hence, the following procedure results in the grade of high accuracy to high-ranking orders, and the grade of low accuracy to low-ranking orders in view of the allowable total MH.

---

*MH Allocation Procedure*

Step 0 [Parameter Setting]: Set the range of allowable total MH for cost estimation, and set the accuracy level from  $(\sigma_{min}, \sigma_{max})$  to each grade; A (high accuracy), B (average), and C (low accuracy)

Step 1 [Initial MH Allocation]: Set all the orders to grade B, and allocate the corresponding MH for cost estimation to each order based on Eq. (1)

Step 2 [Termination Condition]: Calculate the total MH required (*TMR*) by summing up all the MH allocated to each order. If *TMR* is within the range of allowable total MH, stop the procedure with the current MH allocation. If *TMR* is above the allowable range, go to Step 3. If *TMR* is below the allowable range, go to Step 4

Step 3 [Downgrading]: Choose the lowest-ranked one from grade B orders, and set it to grade C. If the grades of orders are all C, stop the procedure with the current MH allocation. Otherwise, go to Step 5

Step 4 [Upgrading]: Choose the highest-ranked one from grade B orders, and set it to grade A. If the grades of orders are all A, stop the procedure with the current MH allocation. Otherwise, go to Step 5

Step 5 [MH Reallocation]: According to the given grades, reallocate the MH for cost estimation to each order based on Eq. (1). Return to Step 2

---

*Step Two: Searching Risk Parameter Value for Profit Maximization.* Given the MH allocation determined by the procedure described above, we search the value of *rp* by solving the following optimization problem:

**Maximize**

$$\sum_{i=1}^L \int_0^{+\infty} (x_1^i - STDR_1^i) \cdot p_1(x_1^i, TBP_1^i, \sigma_1^i) \cdot \prod_{k=2}^n \int_{x_k^i}^{+\infty} p_k(x_k^i, TBP_k^i, \sigma_k^i) dx_k^i \cdot dx_1^i \tag{9}$$

subject to

$$TBP_k^i = STD_i \cdot (1 + RC_k^i) \cdot (1 + e\_profit_k^i) \cdot rp_k^i \quad (i = 1, 2, \dots, L; k = 1, 2, \dots, n) \quad (10)$$

$$\int_0^{STDR_1^i} p_1(x_1^i, TBP_1^i, \sigma_1^i) \cdot \prod_{k=2}^n \int_{x_1^i}^{+\infty} p_k(x_k^i, TBP_k^i, \sigma_k^i) dx_k^i \cdot dx_1^i \leq rprob_i \quad (i = 1, 2, \dots, L) \quad (11)$$

where  $rprob_i$  is the upper limit of the deficit order probability of the order ( $i$ ).

In the above optimization problem, the objective is to maximize the total expected profit from orders. Equation (10) defines  $TBP$ , and Eq. (11) is the upper limit constraint of the deficit order probability. Note that Eq. (10) can be eliminated from the problem by substituting Eq. (10) into Eqs. (9) and (11). Moreover, the problem can be separated into  $L$  problems ( $i = 1, 2, \dots, L$ ). As a result,  $rp$  of one's own company ( $k = 1$ ) is the single decision variable of each problem. In this paper, we use a simple iterative algorithm to search for a solution by gradually eliminating search space.

Given the MH allocation for cost estimation and the value of  $rp$ , the final bidding price is determined as follows:

$$NET_i \cdot (1 + e\_profit_1^i) \cdot rp_1^i \quad (12)$$

where  $NET$ , as shown in Fig. 2, is the estimated cost that is calculated by the allocated MH after the bid or no-bid decision.

## 5 Numerical Examples

In this section, we analyse and discuss the performance of the developed bidding price decision algorithm in EPC projects based on the numerical examples from the following perspectives: relations between cost estimation accuracy and expected profit, effectiveness of bidding price adjustment, and effect of the upper limit constraint of the deficit order probability.

### 5.1 Problem Setting

**Setting of Cases.** In this paper, we use the cases shown in Table 1 for numerical examples. Cases 0 and 1 are set to show the effectiveness of bidding price adjustment by the risk parameter. Cases 2 and 3 are set to show the effects of the

**Table 1** Cases for numerical examples

Case	$rp_1^i$	$\sigma_k^i (k \geq 2)$
Case 0	1.0	8 % of $STD_i$
Case 1	To be searched	8 % of $STD_i$
Case 2	To be searched	6 % of $STD_i$
Case 3	To be searched	10 % of $STD_i$

**Table 2** Conditions of orders

Order id ( $i$ )	1	2	3	4	5	6	7	8
$STD_i$	100			200			300	
$NBR (n)$	2	3	4	2	3	4	2	3
Order id ( $i$ )	9	10	11	12	13	14	15	16
$STD_i$		400			500		600	
$NBR (n)$	4	2	3	4	3	4	3	4

*NBR*: number of bidders

competitors’ cost estimation accuracy on the expected profit and deficit order probability of one’s own company. The competitors’ cost estimation accuracy  $\sigma_k^i (k \geq 2)$  in Table 1 are set based on the expected accuracy for bidding [12].

We set other parameter values through all the cases as follows:  $rp_k^i = 1.0 (k \geq 2)$ ,  $RC_k^i = 0.0 (k \geq 1)$ ,  $rprob_i = 1.0$ , and  $e\_profit_k^i = 0.1$ . We set  $rprob_i$  to 1.0 (100 %) to maximize the expected profit without the upper limit constraint of deficit order probability. The effect of the constraint is shown in Fig. 6.

Note that the value of  $\sigma_1^i$  is determined by Eq. (1) and the allocated MH. We suppose that the bidding price follows a normal distribution. Furthermore, we consider four conditions for the range of allowable total MH for cost estimation, i. e., (A) 70–80, (B) 80–90, (C) 90–100, and (D) 100–110 [M MH].

**Setting of Orders.** In this paper, we assume a mid-size EPC contractor in the chemical plant engineering business, and consider the conditions of 16 orders in each case as shown in Table 2.

Regarding the cost estimation accuracy of one’s own company (see Eq. (1)), we set  $C$  to  $0.25 * 100 / STD_i$ , and  $\sigma_{min}$  and  $\sigma_{max}$  to 0.5 % and 30 % of  $STD_i$ , respectively. In addition, we set the cost estimation accuracy level to 5 % of  $STD_i$  for grade A, 8 % of  $STD_i$  for grade B, and 15 % of  $STD_i$  for grade C when performing the MH allocation procedure.

## 5.2 Results of Numerical Calculations

**Cost Estimation Accuracy and Expected Profit.** As shown in Table 3, the significant difference in the total expected profits is caused by the total MH for cost

**Table 3** Expected orders (*EO*; Eq. (3)) and expected profits (*EP*; Eq. (4))

[MM\$]	The range of allowable total MH for cost estimation [M MH]			
	70–80	80–90	90–100	100–110
Case 0	Case 0.A	Case 0.B	Case 0.C	Case 0.D
<i>EO</i>	1858.2	1817.9	1823.3	1809.0
<i>EP</i>	28.6	46.3	51.7	61.5
Case 1	Case 1.A	Case 1.B	Case 1.C	Case 1.D
<i>EO</i>	1141.6	1238.1	1269.5	1357.2
<i>EP</i>	53.3	56.4	60.9	69.1
Case 2	Case 2.A	Case 2.B	Case 2.C	Case 2.D
<i>EO</i>	1275.2	1395.2	1437.6	1547.3
<i>EP</i>	48.0	51.3	56.3	65.5
Case 3	Case 3.A	Case 3.B	Case 3.C	Case 3.D
<i>EO</i>	1061.6	1143.7	1168.1	1236.6
<i>EP</i>	60.2	63.5	67.5	74.8

estimation for all the cases. For example, the expected profits in Case 0.A (70–80 [M MH]), Case 0.B (80–90 [M MH]), Case 0.C (90–100 [M MH]), and Case 0.D (100–110 [M MH]) are 28.6, 46.3, 51.7, and 61.5 [MM\$], respectively.

Since the cost estimation accuracy depends on the MH for cost estimation as shown in Eq. (1), the results indicate that the cost estimation accuracy affects the expected profit significantly. Namely, the contractor can expect a higher profit by increasing the cost estimation accuracy in EPC projects. However, there is usually a limit to the available MH for cost estimation. Thus we can conclude that an effective mechanism to allocate the MH for cost estimation to each order under the constraint of the volume of total MH is necessary in the bidding price decision process.

**Effectiveness of Bidding Price Adjustment by Risk Parameter.** Based on the results of Case 0 and Case 1, we analyse the effect of the bidding price adjustment on the expected profit. The bidding price is adjusted by  $rp$  to attain the maximum expected profits in Case 1, and the value of  $rp$  is fixed in Case 0.

As shown in Table 3, there is a significant difference in the expected profits between Case 0 and Case 1. For example, the total expected profits in Case 0.A and Case 1.A are 28.6 and 53.3 [MM\$], respectively. The bidding price adjustment also affects the expected orders and profit rate. In Case 0.A, for instance, the expected orders and profits are 1858.2 and 28.6; therefore the expected profit rate is 1.54 %. In contrast, in Case 1.A, the expected orders and profits are 1141.6 and 53.3; therefore the expected profit rate is 4.67 %, which is about three times as high as that in Case 0.A.

The deficit order probability is significantly decreased by the adjustment of the bidding price as shown in Table 4. For example, the range of deficit order probability in the orders is between 11.0 and 25.8 % in Case 0.A, and between 0.777 and 5.81 % in Case 1.A. In Case 0.A, the MH allocation procedure results in the low



**Table 4** Range of deficit order probability (Eq. (6)) [%]

	The range of allowable total MH for cost estimation [M MH]			
	70–80	80–90	90–100	100–110
Case 0	Case 0.A	Case 0.B	Case 0.C	Case 0.D
	11.0–25.8	11.0–12.1	3.20–12.1	2.98–12.1
Case 1	Case 1.A	Case 1.B	Case 1.C	Case 1.D
	0.777–5.81	4.33–5.81	1.77–5.81	1.77–5.81

cost estimation accuracy level (grade C) to the orders 2, 3, 6, and 9, and these orders result in negative earnings as shown in Table 5. However, in Case 1.A, the bidding price adjustment decreases the deficit order probabilities of these orders and improves the expected profits.

Table 6 shows the effects of the competitors’ cost estimation accuracy on the value of  $rp$ , the expected profit, and the deficit order probability of each order. Note that the competitors’ cost estimation accuracy of Case 2.B, Case 1.B, and Case 3.B is 6, 8, and 10 % of  $STD_i$ , respectively. As shown in Table 6, as the competitors’ cost estimation accuracy increases, the value of  $rp$  searched for by the algorithm decreases and the deficit order probability of each order increases. This is because the high accuracy of the competitors’ cost estimation reduces the chance of accepting the orders at high prices, and consequently, a small  $rp$  is chosen to accept such orders.

Figure 3 depicts the relation of the expected order and profit of the order id 10 with the value of  $rp$  in Case 1.B. In addition, Fig. 4 depicts the relation of the expected profits of the order id 10 with the value of  $rp$  in Case 1.B and Case 1.C, each of which corresponds to a different range of allowable total MH. Figure 3 shows that the expected order decreases as the value of  $rp$  increases. However, it is found from Figs. 3 and 4 that there is a value of  $rp$  that attains the maximum expected profit. Furthermore, Fig. 4 tells us the higher cost estimation accuracy, i.e., more MH for cost estimation, makes the maximum expected profit higher.

We can see that the higher cost estimation accuracy reduces the chance of accepting orders at very low price and thus increases the expected profit. However, the higher cost estimation accuracy also reduces chance of accepting profitable

**Table 5** Effectiveness of bidding price adjustment by risk parameter

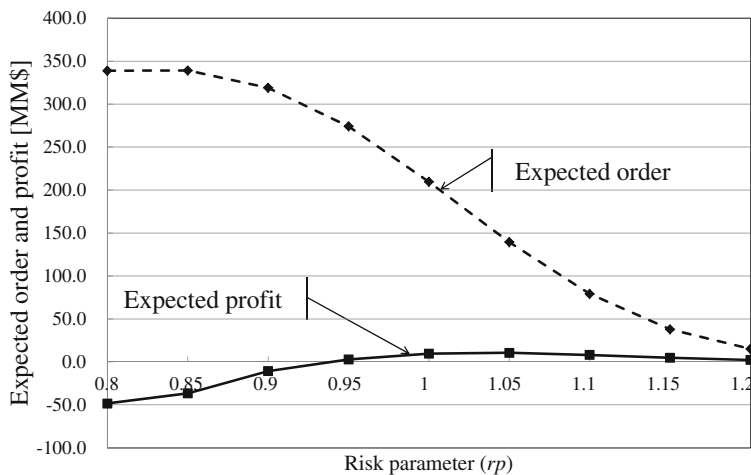
Order $id (i)$	Case 0.A			Case 1.A		
	$rp_1^i$	$EP$ [MM\$]	$DP$ [%]	$rp_1^i$	$EP$ [MM\$]	$DP$ [%]
2	1.0	-1.92	25.8	1.20	0.155	2.32
3	1.0	-2.25	25.2	1.26	0.0290	0.777
6	1.0	-4.50	25.2	1.26	0.0581	0.777
9	1.0	-6.75	25.2	1.26	0.0871	0.777

$EP$ : Expected profit,  $DP$ : Deficit order probability

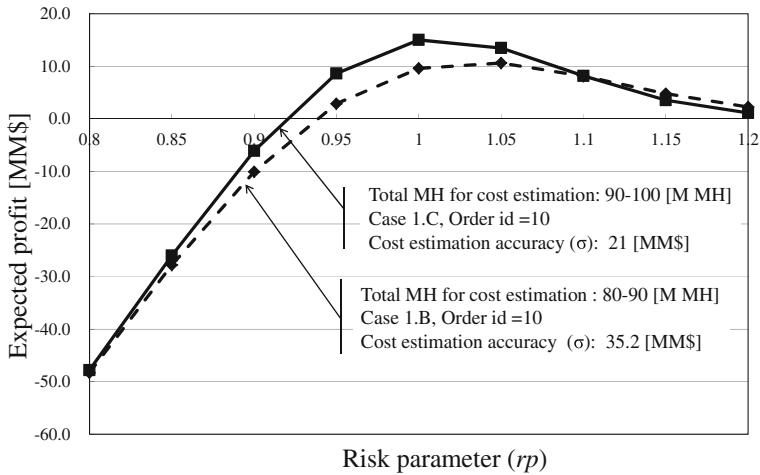
**Table 6** Bidding price adjustment with different competitors' accuracy (80–90 [M MH])

Order <i>id</i> ( <i>i</i> )	Case 2.B			Case 1.B			Case 3.B		
	$rp_1^i$	<i>EP</i> [MM\$]	<i>DP</i> [%]	$rp_1^i$	<i>EP</i> [MM\$]	<i>DP</i> [%]	$rp_1^i$	<i>EP</i> [MM\$]	<i>DP</i> [%]
1	1.026	2.27	7.12	1.035	2.70	5.44	1.045	3.14	4.09
2	1.026	0.916	6.99	1.030	0.983	5.81	1.035	1.09	4.73
3	1.035	0.444	5.44	1.040	0.418	4.32	1.042	0.436	3.63
4	1.026	4.54	7.12	1.035	5.40	5.44	1.045	6.29	4.09
5	1.026	1.83	6.99	1.030	1.97	5.81	1.035	2.18	4.73
6	1.035	0.888	5.44	1.040	0.836	4.32	1.042	0.872	3.63
7	1.026	6.81	7.12	1.035	8.11	5.44	1.045	9.43	4.09
8	1.026	2.75	6.99	1.030	2.95	5.81	1.035	3.27	4.73
9	1.035	1.33	5.44	1.040	1.25	4.32	1.042	1.31	3.63
10	1.026	9.08	7.06	1.035	10.8	5.54	1.044	12.6	4.13
11	1.026	3.67	6.99	1.030	3.93	5.81	1.035	4.36	4.73
12	1.035	1.78	5.44	1.040	1.67	4.32	1.042	1.74	3.63
13	1.026	4.58	6.99	1.030	4.91	5.81	1.035	5.45	4.73
14	1.035	2.22	5.44	1.040	2.09	4.32	1.042	2.18	3.63
15	1.026	5.50	6.99	1.030	5.90	5.81	1.035	6.54	4.73
16	1.035	2.66	5.44	1.040	2.51	4.32	1.042	2.62	3.63

*EP*: Expected profit, *DP*: Deficit order probability



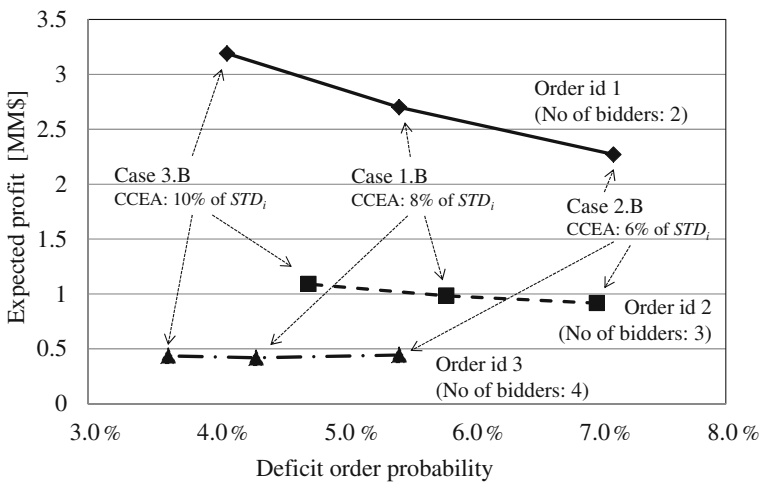
**Fig. 3** Relations among expected order, expected profit, and risk parameter. (Case 1.B; Order id = 10)



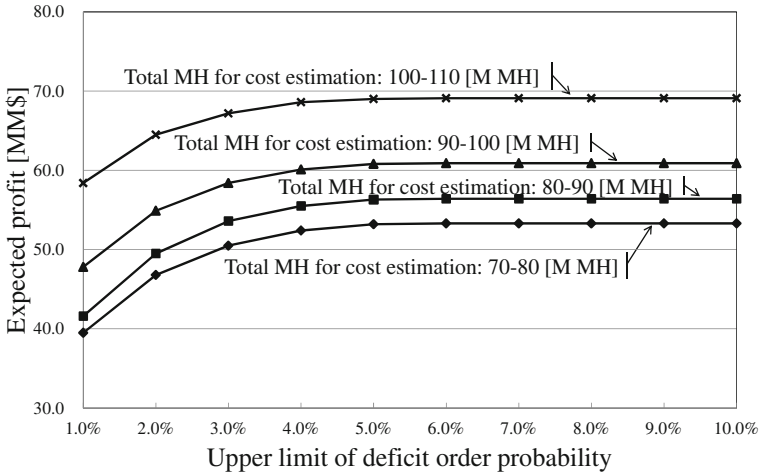
**Fig. 4** Relations among expected profit, total MH for cost estimation, and risk parameter. (Case 1. B, Order id = 10, Total MH for cost estimation: 80–90 [M MH]; and Case 1.C, Order id = 10, Total MH for cost estimation 90–100 [M MH])

orders when the value of  $rp$  is high. In Fig. 4, for instance, the expected profit in Case 1.C is lower than that in Case 1.B when  $rp$  is 1.15 or more.

**Effect of the Number of Bidders.** Figure 5 depicts the relation of the expected profit and the deficit order probability of the order id 1, 2 and 3 in Cases 2.B, 1.B, and 3.B. Note that the number of bidders is set to two, three, and four for the order id 1, 2 and 3, respectively. Also, in Cases 2.B, 1.B, and 3.B, the competitors’ cost



**Fig. 5** Relations among expected profits and deficit order probability. (Case 2B, 1B, and 3B; Order id 1, 2, and 3) (CCEA: Competitors’ Cost Estimation Accuracy)



**Fig. 6** Relations among expected profits, total MH for cost estimation, and upper limit of deficit order probability (Case 1)

estimation accuracy is set to 6, 8, and 10 % of  $STD_i$ . As shown in Fig. 5, the effect of the competitors’ cost estimation accuracy on the expected profit and the deficit order probability becomes smaller as the number of bidders increases.

For instance, in Order id 1, i.e., when the number of bidders is two, the difference of the expected profit between Case 3.B and Case 2.B is 0.87 [MM\$]. In contrast, in Order id 3, i.e., when the number of bidders is four, the difference of the expected profit between Case 3.B and Case 2.B is 0.008 [MM\$]. The difference in the deficit order probability between Case 3.B and Case 2.B is also reduced from 3.03 (in the case of Order id 1) to 1.81 [%] (in the case of Order id 3).

High degree of competition significantly reduces the chance of accepting orders at high prices as well as at low prices regardless of the competitors’ cost estimation accuracy. Consequently, it reduces the effect of the competitors’ cost estimation accuracy on the expected profit and the deficit order probability.

**Effect of Upper Limit Constraint of the Deficit Order Probability.** We examine how the upper limit constraint of the deficit order probability shown in Eq. (11) affects the expected profit. Figure 6 depicts the relation of the upper limit of the deficit order probability and the total expected profit in Case 1. As explained in Sects. 2 and 3, the risk of unexpected loss from the deficit orders should be avoided especially when only a small number of orders can be accepted. As shown in Fig. 6, the small upper limit of the deficit order probability decreases the total expected profit; however, it is found that the deficit order probability can be reduced from 5.0 to 1.0 % at the expense of the total expected profits of 10–15 [MM\$].

Bidding for a large-scale EPC project involves a substantial risk. Our framework developed for EPC projects will certainly be helpful for any contractor to avoid large deficit from accepted orders.

## 6 Conclusions

In this paper, we develop a heuristic bidding price decision algorithm based on the cost estimation accuracy under limited engineering MH in EPC projects. The algorithm allocates MH for cost estimation to each order under the limited volume of MH, and then determines the bidding price to maximize the expected profit under the deficit order probability constraint.

We develop a mathematical model for simulating competitive bidding. Through the numerical results obtained by using the model, we show that the bidding price decision in consideration of cost estimation accuracy and deficit order probability is essential for the contractor in making a stable profit in EPC projects, and that the heuristic bidding price decision algorithm developed in this paper is effective for making such bidding price decisions.

There are several issues that require further research. For example, the procedure for modifying the MH allocation and adjusting the bidding price dynamically in response to each order arrival is required for practical application. In addition, our heuristic algorithm does not consider the duration for estimating cost and for carrying out the project. The MH allocation procedure should consider the time cost-trade-off and its implication on the cost estimation accuracy and profit. It is also necessary to compare the performance of our procedure with other project scheduling methods dealing with the optimum allocation of resources for multiple projects.

## References

1. Ballesteros-Pérez P, González-Cruz MC, Cañavate-Grimal A (2013) On competitive bidding: scoring and position probability graphs. *Int J Proj Manage* 31:434–448
2. Bertisen J, Davis GA (2008) Bias and error in mine project capital cost estimation. *Eng Econ* 53:118–139
3. Elfving JA, Tommelein ID, Ballard G (2005) Consequences of competitive bidding in project-based production. *J Purchasing Supply Manage* 11:173–181
4. Friedman L (1956) A competitive-bidding strategy. *Oper Res* 4:104–112
5. Helmus FP (2008) *Process plant design: project management from inquiry to acceptance*. Wiley, Weinheim
6. Ishii N, Muraki M (2011) A strategy for accepting orders in ETO manufacturing with competitive bidding. In: *Proceedings of 1st international conference on simulation and modeling methodologies, technologies and applications*, pp 380–385
7. Ishii N, Takano Y, Muraki M (2014) An order acceptance strategy under limited engineering man-hours for cost estimation in engineering-procurement-construction projects. *Int J Proj Manage* 32:519–528
8. Jinru Z (2011) Study on cost management under EPC general contracting model. *Adv Mater Res* 181(182):49–53
9. Kerzner H (2009) *Project management: a systems approach to planning, scheduling, and controlling*. Wiley, New Jersey
10. Kortanek KO, Sodeni JV, Sodaro D (1973) Profit analyses and sequential bid pricing models. *Manage Sci* 20:396–417

11. Martino JP (1995) Research and development project selection. Wiley, New York
12. Oberlender GD (2000) Project management for engineering and construction. McGraw-Hill, New York
13. Oberlender GD, Trost SM (2001) Predicting accuracy of early cost estimates based on estimate quality. *J Constr Eng Manage* 127:173–182
14. Pritchard N, Scriven J (2011) EPC contracts and major projects, 2nd edn. Sweet & Maxwell, London
15. Ranjan M (2009) EPC: a better approach. *Chem Eng World* 44:46–49
16. Rothkopf MH, Harstad RM (1994) Modeling competitive bidding: a critical essay. *Manage Sci* 40:364–384
17. Steel G (2004) Tender management. In: Morris PWG, Pinto JK (eds) *The wiley guide to managing projects*. Wiley, New Jersey, pp 708–742
18. Takano Y, Ishii N, Muraki M (2014) A sequential competitive bidding strategy considering inaccurate cost estimates. *OMEGA* 42:132–140
19. Towler G, Sinnott R (2008) *Chemical engineering design—principles, practice and economics of plant and process design*. Elsevier, Burlington
20. Wang J, Xu Y, Li Z (2009) Research on project selection system of pre-evaluation of engineering design project bidding. *Int J Project Manage* 27:584–599
21. Watt DJ, Kayis B, Willey K (2009) Identifying key factors in the evaluation of tenders for projects and services. *Int J Project Manage* 27:250–260

# Intelligent Agents for Human Behavior Modeling as Support to Operations

Agostino G. Bruzzone, Marina Massei, Simonluca Poggi,  
Christian Bartolucci and Angelo Ferrando

**Abstract** Goal of the present paper is providing support to operations planning and management in complex scenarios. The authors are mainly focused on South Asia region, which is subject of experimental analysis by running an Intelligent Agents—driven HLA Federation. Simulation of investments and operations over an asymmetric mission environment with several parties, insurgents, terrorists and dynamic social framework is the aim. The scenario has various degrees of freedom and M&S enables evaluation of human behavior evolution and socio-psychological aspects. The presented models include Computer Generated Forces (CGF) driven by Intelligent Agents (IAs) that represents not only units on the battlefield, but also people and interest groups (i.e. Middle Class, Nomads, Clans). The study is focused on Civil Military Co-operations (CIMIC) and Psychological Operations (PSYOPs). The simulation is based on specific architecture that involves various federates playing different roles. Verification, Validation and Accreditation (VV&A) has been applied along the whole life cycle of the research, in order to determine the correctness and effectiveness of the results. The paper proposes experimental results obtained during the dynamic test of the federations.

**Keywords** Intelligent agents · Human behavior modeling · Simulation of operations · Country reconstruction · CIMIC · PSYOPS

---

A.G. Bruzzone (✉) · M. Massei  
Simulation Team, DIME University of Genoa, via Opera Pia 15, 16145 Genoa, Italy  
e-mail: agostino@itim.unige.it

M. Massei  
e-mail: massei@itim.unige.it

S. Poggi · A. Ferrando  
MAST Srl, Corso Podestà 8, 16128 Genoa, Italy  
e-mail: simonluca.poggi@mastsrl.eu

A. Ferrando  
e-mail: angelo.ferrando@mastsrl.eu

C. Bartolucci  
Simulation Team, Via Molinero 1, 17100 Savona, Italy  
e-mail: christian.bartolucci@simulationteam.com

## 1 Introduction

The human factors are a critical element when investments and operations are planned over a region; indeed the impact of population point of view and the interests of related social layers is often affecting effectiveness and efficiency of operations, it could introduce risks and/or provide opportunities; these elements are obviously pretty difficult to be investigated therefore they strongly affect the overall success; normally it is fundamental to identify all the stakeholders and to consider their interests and their attitude, this require to consider for instance economic, religious, ethnic groups as well towns, villages, local leaders as key factor to be consider in planning.

In several geo-political areas it is required to plan new investments and activities devoted to stabilize or normalize the situation respect previous critical conditions (i.e. civil war, insurgency, terrorism, etc.); in these case it is common to develop initiatives devoted to get support of the local population as well as to improve the quality of life from several point of view (i.e. economy, health, civil rights, security, education, etc.); in order to achieve these results Civil Military Co-operations (CIMIC), Information Operations (INFOPS) and Psychological Operations (PSYOPs) need to be prepared and carried out properly [4].

So it is evident the interest in being able to model these activities as well as the dynamic interaction with the population over a specific framework; indeed such interactions could be pretty complex involving many interest groups representing the different social layers of the population as well as their distribution over the terrain; obviously these elements need to be considered even in reference to the existing situation of the area from many point of view: infrastructures (i.e. roads, hospitals, schools), environmental conditions (i.e. weather), specific actions (i.e. strikes, demonstrations, intimidation activities).

The authors currently have developed models for these context by creating a new generation of intelligent agents [19] representing population and interest groups to drive complex simulation [20] related to these mission environments; therefore it is important to state that these models and the previous considerations could be effectively tailored and applied also in relation to civil scenarios where new investment (i.e. industries and infrastructures) have to be plan over a domestic region or a district as well as during promotional campaign in marketing initiatives.

## 2 Simulation Objectives

Therefore the international context in unstable areas introduce a good motivation to investigate these population models (i.e. country reconstruction operations); currently there it is expected that these models could be very useful for evaluating alternative planning considering risk, opportunities, times, resources costs over a complex and stochastic framework, vice versa the predictive capability of these



simulators is still pretty limited due to the high degree of uncertainty affecting human elements and the high influence of specific spot events [5]. Due to these considerations, the proposed agent-driven simulations are devoted to conduct experimental analysis and decision support by providing reliable estimations and useful risk analysis, but not the of the exact time and location of a new riot; indeed these events are generated usually by an ignition factor that is highly unpredictable (i.e. a single phrase or shot in a specific moment).

Considering the proposed context it is evident that nowadays military mission environments, especially within countries characterized by different cultures, society organization and changeable political situations, require a new approach to tactical and strategic operations which not only appreciates military engagements, but also relationship between civil population, military forces as well as community evolution and interest groups.

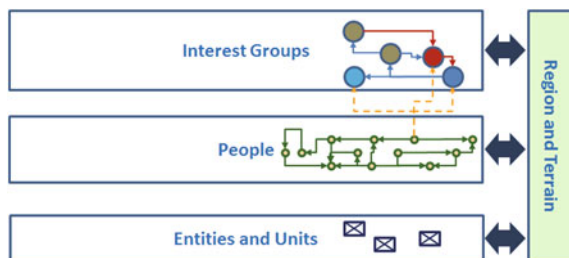
The problem of this analysis is that there are not universally accepted simulation models and that the human behavior modifiers (HBM) are very difficult to be represented; in addition it is even necessary to create models of specific operations that are currently not covered by the existing simulators in order to take of Civil-Military Cooperation, INFOPS, PSYOPS as well as of psychological consequences over population during mission execution; therefore some existing model/simulator is currently taking into account these not-kinetic operations, but usually it is just a qualitative on/off parameters or a manual script affecting the scenario evolution; this obviously don't allow to consider the complex dynamic of the interaction among different interest groups that is the basis for situation evolution.

### **3 Application Framework and Proposed Approach**

The simulator should consider for instance that digging a well within an area could generate positive effects on some part of population (i.e. people hired to carried out the work, owner of the land) as well as negative effects on other ones (i.e. opposite clan respect well owner, opposite political party respect that one involved); these actions generated direct impact on element of the population living in the area as well as on the their related interest group and in addition produce a cascade effect on all the social networks among people and interest groups. In addition if due to weather conditions and/or lack of resource the well constructions result to be affected by delays this could produce negative impact on the people that expect the completion to get benefits of this asset.

All these elements as well as the cascade of effects could result positive or negative with a strong influence due to the dynamically evolving relationships among people and interest groups and also due to the importance of the specific actions, the cultural background and the communications [15].

Indeed the diffusion in the region and among the people and interest groups of the effects of the actions is modeled based on communications over different supports (face to face, media, phones) and considering specific factors; therefore



**Fig. 1** Multilayer model including population, interest groups and units

these communications introduces attenuation factors and delays; due to the computational workload (i.e. in our case 300,000 people and 60 interest groups) the cascade effect could slow down simulation on single workstations, for this reason it is possible to run the simulation with correct diffusion models or by considering that the diffusion happen with fixed stochastic delays along each single operation phase (this reduces of drastically the events to be considered); considering multiple actions on going concurrently and the main interest to measure final effects this simplification resulted acceptable, therefore if computation power is available it is possible to run the simulator using more correct models.

In the proposed models it was required to model these elements and to create a simulation able to reproduce a scenario where CIMIC and PSYOPS were conducted; the interest group and populations were modeled based on a multilayer approach that consider both population elements deployed over the terrain and interest groups; in addition these entities interact with the units of the different parties (i.e. coalition forces and insurgents) (Fig. 1).

Due to the fact that the intelligent agents reproduce population behavior within operations, the model allows the users to correctly evaluate the reaction of the civilians not only to military actions, but also during peace keeping and reconstruction phases.

The research has been developed and tested through the involvement of Subject Matter Experts (SME) from different countries; in particular the proposed scenarios was developed as demonstration for a R&D (Research and Development) project named CIMIC and Planning Research in Complex Operational Realistic Network (CAPRICORN).

## 4 Intelligent Agents and Simulation Development

These models have been created in order to support operational planning decisions and to be integrated with other systems previously developed; it is able to work as single user or within a federation architecture (HLA standard) [12]; this was motivated not only by the needs of respecting existing standards, but even for the opportunity provided by developing an open architecture to be further integrated

with other simulations in order to cover complex problems. Due to these reason interoperability requirements are pretty important and represent a strategic advantage of the proposed approach [1, 2, 4–18]. In fact IAs (Intelligent Agents) [10]; have a big potential in addressing these kind of problems [17].

The most important feature of the models is represented by the intelligent agents that are able to simulate human behavior of people modeling their characteristics; this not only in term of their party side (neutral forces, friends, enemies and civilians), but even in reference to their liaisons to different interest groups and social networks. The capability to use this approach it is an important support for applications involving federations of simulation to address complex scenarios and multiple threats for training and it is pretty interesting to investigate their use for supporting operational planning.

In the past the authors set up libraries of innovative models able to simulate different society attributes represented by riots, agitators and terrorists (RATS) and IA-CGF modules [1, 6]; in some case it was possible to simulate the whole population of a large area reacting to a natural disasters [3] of a big town respect humanitarian activities [7] or in relation to health care issues [9]; therefore in this case the intelligent agents were extended to cover, not only entities and units operating on the field as well as single individuals/families within the population, but even social objects such as interest groups.

In order to succeed in this process it becomes necessary to properly design, tailor and experiment the scenario considering the very large quantity of elements, variables and parameters; due to these reasons the M&S (Modeling and Simulation) process is formulated over three phases: simulation development, specific mission environment tailoring and simulation experimentation over the specific mission environment. The authors decided to develop an innovative model of a whole country, taking into account the features that involve agents able to correctly interact in the agent based environment; obviously considering the very broad spectrum of applications and elements affecting these operations it is critical to restrict the range of validity and the components to model based on a detailed analysis to be carried out among trans-disciplinary teams involving scientists and operative people [9]. Indeed considering the possibility to use these agents in order to support decision makers on the field for planning operations in overseas scenarios it could be very important to develop a simulator able to be used by people with no strong scientific background and using limited computing capability. Therefore it is necessary to develop models and simulators able to run correctly based on an installation and configuration that should be operated and maintained on field with the kind of resources that are expected to be available on that context; users should be able to configure and create a mission environment; therefore considering the complexity of these applications it is expected that a team of experts and analysts will proceed to create a configuration of the simulator for a specific region and timeframe (i.e. Kapisa District in Afghanistan 2010) to be used as reference by final users eventually deployed on site.

More in detail, the first step approached by the scientists consist of Modeling phenomena, actions and elements that are specific for a socio-political-cultural

framework. During the conceptual model creation, simulation expert contributions are essential to building a proper and effective set of models and to properly approach the problem thanks to their knowledge of specific operations and scenarios; for instance for a region could emerge the necessity to include nomad behaviors among the possible alternative occupation of the population.

The second phase regards precisely the knowledge management: info sources are used in order to achieve the knowledge basis in order to tailor parameters and entities of a specific region or context; indeed to the necessity to determine a reference scenario a specific mission environment has to be defined; for instance it is necessary to collect information about the different political and economic groups as well as to tailor the importance of religion and clan factors in term of their influence in creating a family respect the specific cultural area. Obviously along the entire M&S process the VV&A has been performed to ascertain their formal correctness and their usability effectiveness according to the imagined use; therefore during the simulation experimentation is the critical moment to dynamically test the validity of the models as well as the functionality of the simulator. A set of mission environment hypothesis are defined by planners and/or analysts in order to choose alternative friend course of action (F-COA) and opposite course of action (O-COA); by this approach it is possible to plan investments and operations; a COA could involve CIMIC or PSYOPS targeting different interest groups over a specific zone, affecting people in the area as well as social layers; in addition the decision maker could define the operation time plan of the investment, the assigned resource in term of money as well as equipment and people; obviously during the simulation multiple operations could be planned and carried out concurrently or sequentially and the simulator allow to consider availability of resources, influence of opposite force actions as well as weather condition influence (i.e. weather working days for external constructions).

Each CIMIC or PSYOPS evolves based on different phases (i.e. for a CIMIC planning, engineering, acquisition of resources, erection, commissioning) each one affected by specific needs in term of money, resources, boundary conditions (i.e. weather).

The agents are currently driving the behavior of population and interest groups respect their perception of the general situation and their “feelings” respect on-going activities; the models use fuzzy rules to estimate the effect of the different operations respect their nature and their attitude respect the actors.

Relationships among entities are usually defined usually by two functions (impact and influence) that could be defined in term of ownership to membership functions respect the following classes (low, medium, high and negative, indifferent and positive) as proposed in the following graph (Fig. 2). By this approach it is possible to express quantitative estimations about the effectiveness of the actions conducted; for instance it is possible to apply de-fuzzification in order to transform the relationships among different critical interest groups or people over an area in order to estimate their trustiness respect specific players; for instance the Overall Trustiness of the Population respect the Coalition could be estimated or that one of a specific religious group, village and/or district. For each mission environment the

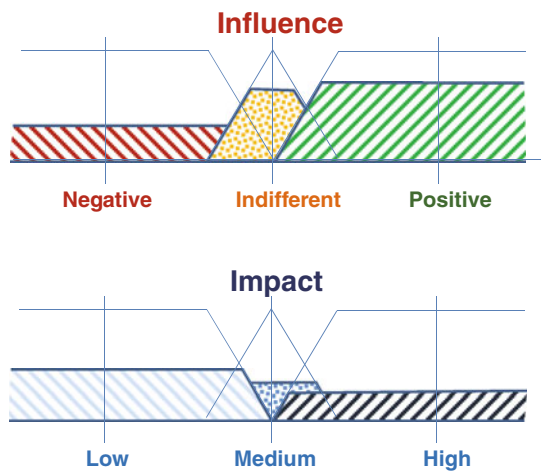
simulation parameters have to be set according to the initial conditions and hypothesis; then predefined settings represent the base for the execution of the simulator, which outputs has to be analyzed.

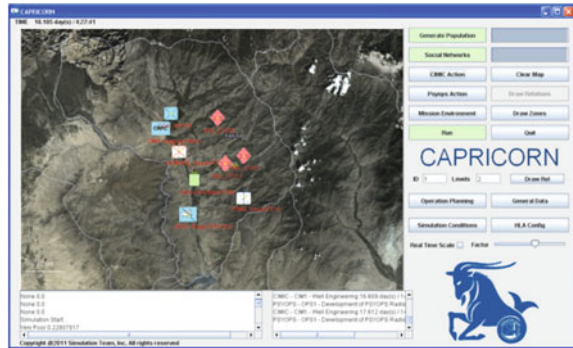
The simulation execution could run in different operative modes according to the context and user requirements: stand alone, federated with other simulators, multiple replicated runs etc.; at the end the outputs are collected for each single run and statistically analyzed; these results are evaluated during the simulation experimentation [16]; a classic and simple approach for analyzing the results it is based on what if analysis consists of the simulation of different hypothesis previously formulated. Another possibility that is allowed by simulator is to compare the Desired Final Effect (DFE) of a scenario with the Simulated Course of Action (COA).

The user can also first define Key Performance Indexes (KPIs) and then compare final results of different planning alternatives basing evaluation on them; an effective approach to perform a ranking of different alternatives consists of creating a target function which has to be able to appreciate and involve all these Key Performance Indexes. The latter also represent the reference in order to develop the cost-benefits analysis, which, together with risk analysis, gives to the user all the elements for choosing the best planning alternative. Considering the complexity of the mission environment usually Design of Experiments (DOE) is used in order to complete analysis and produce synthetic reports [11].

Through the feedback from military users with operational experiences and subject matter experts on the specific disciplines, it was possible to develop the models as well as to define the specific user needs; by this approach to develop and validate the conceptual models, to perform the definition of the specific mission environment created for CAPRICORN Demonstrator and to validate the functions; during the last phase related to use the Demonstrator for testing and analyzing simulation results it was possible to complete the dynamic VV&A of the proposed approach over a specific case study. In particular the involvement of the users for

**Fig. 2** Fuzzy membership for evaluating mutual relationships between two groups of interest



**Fig. 3** Simulation interface

VV&A was based on different phases; during the first one the focus was to review of key concepts and operational planning requirements; this allowed to generate a common synthesis about CIMIC and PSYOPS, decision making processes, scenario analysis methodologies, training and risk analysis requirements; most of the activities carried out in this phase was organized mostly by desk-top review and face validation performed through organization of meetings and workshops; therefore in this phase some preliminary simplified model was presented and even executed to share concepts and to validate and verify model assumptions and IA basic behaviors. In the following phases the work was based on running of the simulation in front of the users and on analyzing experimental results: during this phase operational planning was carried out through the cooperation of users, analysts, subject matter experts, development team and operational planners (Fig. 3).

## 5 Population Model

For a tactic scenario, such as the real recently warfare where Northern Atlantic countries are involved, is necessary to model civil status and characteristics like:

- Ethnic and Clan
- Religion;
- Cultural/educational level;
- Social and Economic Status
- Geographic location of object;
- Gender;
- Age;
- Health care status;
- Political party;
- Psychological status;

**Table 1** Classes included in the layers

Classes	Description
Terrain	Infrastructure Layers (i.e. Roads), Terrain Elements, Weather Conditions
People	The Population Entities on the Terrain (i.e. Mrs. Baran and/or Sakhi Family)
Groups	Interest Groups (i.e. Sunnis, Honey Producers, Hotaki Clan, Tajik Ethnic Group, Hamnazar Political Party)
Entities	Units on the Terrain (i.e. Coalition Platoon, Insurgent Group, Riot)

And especially are considered particular modifiers of person features, such as:

- Stress;
- Fear;
- Aggressiveness;
- Fatigue;
- Trustiness.

People relationships, friendships and social relationships are also considered in agents algorithms and distributed in a stochastic manner using Monte Carlo techniques based on consistency algorithms able to aggregate people respect the socio-cultural-economic background (Table 1).

People have their social networks interconnecting the population elements based on familiar and friendship relationships; while each people object is connected to multiple interest groups based on his nature with dynamic links representing his affiliation and the related strength; in addition groups and interconnected by mutual hostility and friendship over the social layer, obviously also these connections evolve dynamically during the simulation due the actions carried out; in addition the terrain and Entities affect the behavior of people and social layers, while the Intelligent Agents are in charge of directing the objects during the simulation for completing tasks and for reacting to stimuli and to their own situation awareness.

## 6 Demonstrator Results

The Capricorn Demonstrator consists of a Simulator including the Mission Environment Generator based on Monte Carlo technique applied on statistical database of the population; by this approach the whole people objects representing population and all related interest groups are created and interconnected by the reference relationships over the different layers.

It is proposed an example specific to a CIMIC/PSYOPS mission environment in the Kapisa Afghan region, considering the related COA and parameters concerning population, social networks and groups. The simulation paradigm is based on stochastic discrete-events simulation and it is federated within an HLA Federation (High Level Architecture) both referring to original and IEEE1516 standards; models were implemented in Java with different RTI (Run Time Infrastructures) were tested

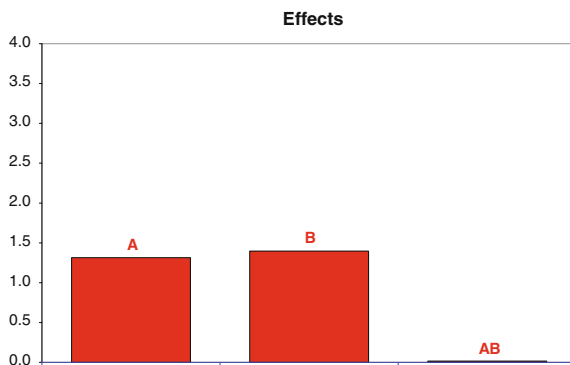
including Portico, Pitch, VT Mäk. During the test federation integrated CAPRICORN Simulator and IA-CGF E&U (Intelligent Agent Computer Generated Forces Entities and Units) developed by Simulation Team for modeling units on the battlefield; the simulation were carried out over Kapisa Region in South Asia considering presence of several companies of Coalition Forces, several units of Insurgents able to carried out O-COA (i.e. Intimidation), Demonstrations and Riots generated based on the population behaviors and simulated within IA-CGF E&U; the operations (i.e. CIMIC and PSYOPS) as well as the Interest Groups and Population were simulated by CAPRICORN Demonstrator over a timeframe corresponding to 1 year. ANOVA (Analysis of Variance) was applied in order to measure the confidence band on the controlled variables and the optimal duration time [14]. As anticipated the population and social networks, within the simulation, are generated by CAPRICORN Demonstrator basing on Monte-Carlo techniques; Groups and people so generated relate with entities and units as well as with PSYOPS/CIMIC operations; the actions and events are affected by stochastic factors considering time, cost and effectiveness elements as well as all human behavior factors.

The outputs of CAPRICORN are related to performances in term of times, costs and involved resources during the planned operations as well as a KPI concerning the evolution of the mutual behavior between the critical groups represented in the chosen scenario (i.e. trustiness of the target groups of the operation respect coalition forces, or overall trustiness of the population versus coalition forces).

A sensitivity analysis based on DOE was carried out respect different independent variables such as:

- A: Budget Allocated to the main Operation
- B: Staff and Resources Assigned to the main Operation

In the Fig. 4 the effect represents the influence expressed as ratio between contrast and the square pure error of trustiness scalar of the target function “overall trustiness”; the analysis propose the effect of single independent variables and of their combination respect this output.



**Fig. 4** Estimation of the effect of planner degrees of freedom (budget and resources) on the trustiness between the two critical Groups of Interest (Coalition and Target Group)



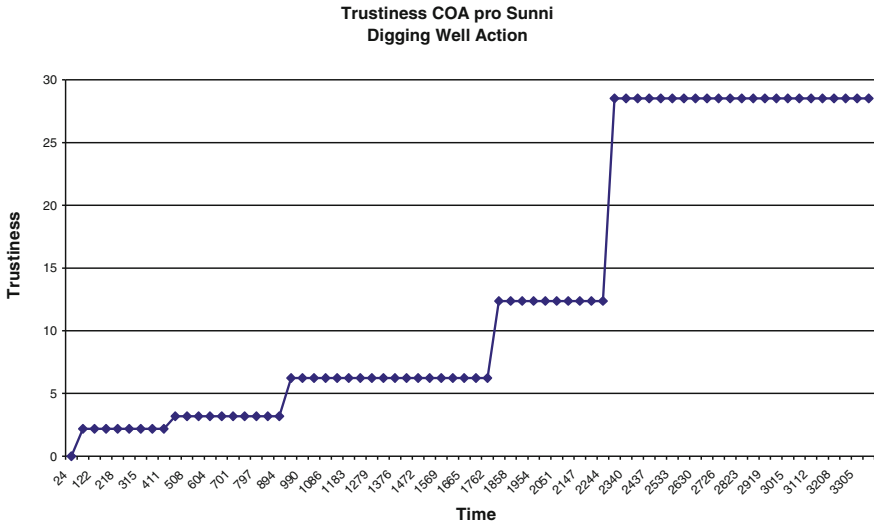


Fig. 5 Trustiness evolution with simplified model

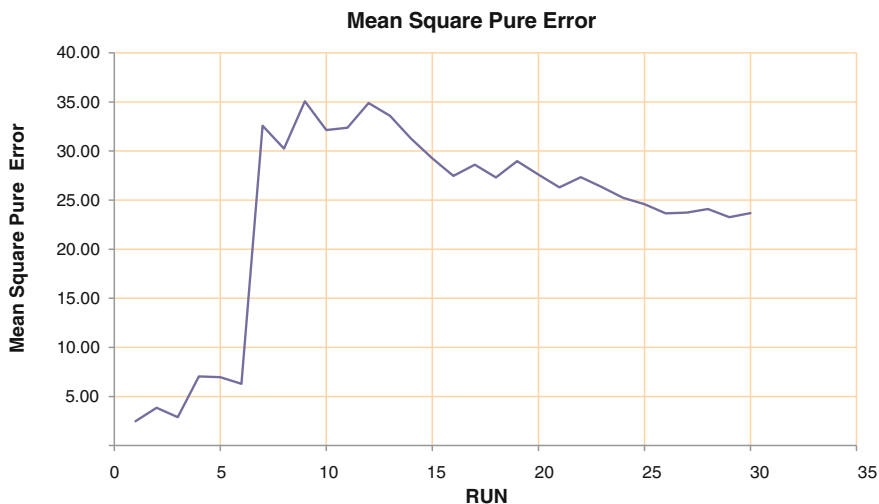
At the end of the simulation process the user knows the effective schedule of the different operation phases (i.e. planning, preparation, supplying, erection/execution, commissioning/follow-up) and the overall duration as well as costs, cash flow, impact on the population.

Figure 5 shows the evolution of trustiness during the simulation of the CIMIC action well digging COA; it is evident that the deliverables of the different phases introduce major changes; in fact the simulation in this case was executed with the simplified algorithm for diffusing of positive/negative reinforcement due to the action among the population the cascade.

The model allow simulating multiple CIMIC and PSYOPs actions within different zones and affecting different groups; these could be planned and studied over a single simulation run: in this case report concerns information about their changing along the time and their effect on the population. So for example the solution which determines the best impact on civilians could be identified and quantified as well as risk estimation could be used to support decision making.

The authors provide the experimental analysis carried out through the final reports of CAPRICORN simulation; in effect it is possible to evaluate and test different courses of actions by planning and simulating different CIMIC/PSYOPs in term of type, location, target group, promoting group, duration, budget, resources. Concerning with the experimental analysis, the authors applied Mean Square pure Error and Sensitivity Analysis for different CIMIC & PSYOP actions by considering a 2-level factorial design respect to the trustiness level as target function.

The figure below represents the mean square pure error diagram for a CIMIC action Digging Well from COA in term of variance of the trustiness respect the

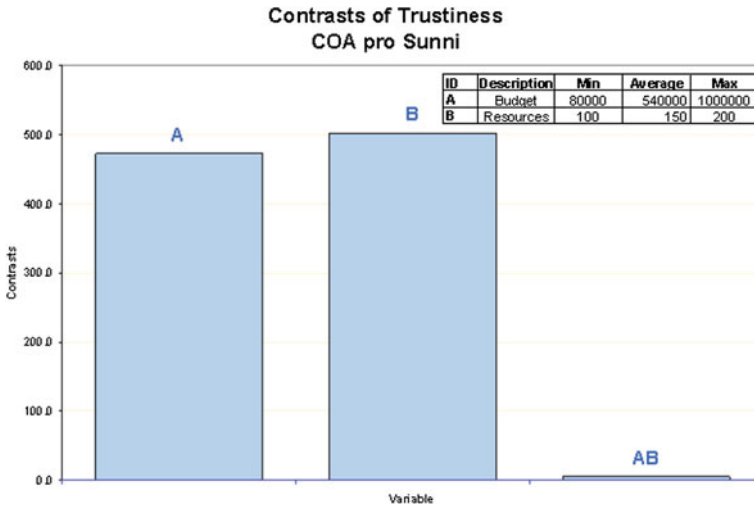


**Fig. 6** Mean square pure error diagram for a CIMIC action digging well from COA pro Sunni in Zone 0 (Run expresses the pure number of replications; MspE express the trustiness variance; Trustiness [-100 to 100])

replicated runs carried out by changing the random seeds of the statistical distributions; it is evident with 25 replications it is possible to obtain results stable with acceptable confidence in term of trustiness ( $\sim 15\%$ ). For instance, considering a CIMIC action Digging Well from COA pro Sunni in Zone 0 the MSpE (Mean Square pure Error) was the following:

The Experimental Error is stable between the 26th and the 30th run. So it is possible to state that 26 run are needed for a correct evaluation of the experimental error. By considering the sensitivity analysis related to the independent variables Resources and Budget, both are significant and have positive effects on the population trustiness level (Fig. 6).

In Fig. 7 the contrast represent the influence of factor expressed respect the target function “Trustiness among Coalition and Sunni Interest Group”. The model is available to be used for several different kind of investments and infrastructures such as school construction, police station installation and digging wells and it could be extended easily to irrigation infrastructures, buildings and apartment constructions, roads; in addition in term of INFOPS and PSYOPS it is already possible to simulate use of Radio and TV Media as well as leaflet campaign over a region.



**Fig. 7** Contrasts diagram related to trustiness target function respect of budget and resources (Contrast represent the influence of a factor expressed respect trustiness as scalar [-100 to 100]; Budget [Euro] and Resource [people])

## 7 Conclusions

This paper propose an approach to model operations devoted to create infrastructures and actions on a area to improve the social economic situation; the authors developed innovative models for the population and the interest groups devoted to reproduce their behavior and to estimate the impact of the new actions the context is referring to the case of CIMIC, INFOPS and PSYOPS conducted in South Asia and consider both economic and operational aspects as well as weather conditions and possible hostile actions by insurgents.

The research proposed by the authors represents a modeling approach for reproducing complex behavior among population and interest groups during specific operations. The experimental analysis provided interesting results and confirmed the potential of this approach; currently the authors are working for further extend the current models for different applications including industrial and civil cases over domestic scenarios.

**Acknowledgments** The authors are very glad to thank EDA—European Defence Agency, the Italian MoD (Ministry of Defence) and French MoD that sponsored CAPRICORN Project. For further information please visit [www.liophant.org/projects/capricorn](http://www.liophant.org/projects/capricorn).

## References

1. Bruzzone AG (2008) Intelligent agents for computer generated forces. Invited Speech at Gesi User Workshop, Wien, Italy, Oct 16–17
2. Bruzzone AG (2011) Human behavior modeling as an enabling technology for our world challenges. Invited Speech I3M2011, Rome, Italy, September
3. Bruzzone AG, Massei M, (2006) Modeling for estimating impact on road transportation of regional emergencies & disasters. In: Proceedings of HMS, Barcelona, Spain, October
4. Bruzzone AG, Massei M (2007) Polyfunctional intelligent operational virtual reality agent: PIOVRA final report, EDA Technical Report
5. Bruzzone AG, Massei M (2010) Intelligent agents for modeling country reconstruction operation. In: Proceedings of AfricaMS 2010, Gaborone, Botswana, Sept 6–8
6. Bruzzone AG, Scavotti A, Massei M, Tremori A (2008) Metamodeling for analyzing scenarios of urban crisis and area stabilization by applying intelligent agents. In: Proceedings of EMSS2008, Sept 17–19, Campora San Giovanni (CS), Italy
7. Bruzzone AG, Massei M, Tarone F, Madeo F (2011) Integrating intelligent agents & AHP in a complex system simulation. In: Proceedings of the international symposium on the AHP, Sorrento, Italy, June
8. Bruzzone AG, Tremori A, Massei M (2011) Adding smart to the mix. Model Simul Training: Int Defense Training J 3:25–27
9. Bruzzone AG, Novak V, Madeo F (2012) Agent based simulation model for obesity epidemic analysis. In: Proceedings of I3M2012, Wien, Austria, September
10. Calfee SH, Rowe NC (2004) Multi-agent simulation of human behavior in naval air defense. Naval Eng J 116(4):53–64
11. Hill D (1996) Object-oriented simulation. Addison Wesley, Reading
12. Kuhl F, Weatherly R, Dahmann J (1999) Creating computer simulation systems: an introduction to the high level architecture. Prentice Hall, Upper Saddle River
13. Montgomery DC (2000) Design and analysis of experiments. Wiley, New York
14. Mosca R, Giribone P, Bruzzone AG (1993) Optimum search techniques applied to studies relative to plant problems performed by means of simulation. In: Proceedings of SimTec93, November, San Francisco (CA)
15. Ören TI, Yilmaz L (2009) On the synergy of simulation and agents: an innovation paradigm perspective, special issue on agent-directed simulation. Int J Intell Control Syst (IJICS), vol 14, Nb. 1, March 4–19
16. Ören TI, Yilmaz L (2012) Agent-monitored anticipatory multisimulation: a systems engineering approach for threat-management training. In: Proceedings of the 24th European Modeling and Simulation Symposium, September 19–21, Austria, pp 277–282
17. Seck M, Frydman C, Giambiasi N, Oren TI, Yilmaz L (2005) Use of a dynamic personality filter in discrete event simulation of human behavior under stress and fatigue. Found Augmented Cogn 11:908–917
18. Spiegel MR, Schiller LJ (1999) Statistics. McGraw Hill, NYC
19. Yilmaz L, Ören T (2010) Intelligent agent technologies for advancing simulation-based systems engineering via agent-directed simulation. SCS M&S Magazine, July
20. Zacharewicz G, Frydman C, Giambiasi N (2008) G-DEVS/HLA environment for distributed simulations of workflows. Simulation 84(5):197–213

# Alternative to Multifractal Analysis of Scalable Random Variables Applied to Measured and Estimated Soil Properties at an Arizona Field Site

Alberto Guadagnini, Shlomo P. Neuman, Marcel G. Schaap and Monica Riva

**Abstract** Many earth, environmental, ecological, biological, physical, astrophysical and financial variables exhibit random space-time fluctuations; symmetric, non-Gaussian frequency distributions of increments characterized by heavy tails that often decay with separation distance or lag; nonlinear power-law scaling of sample structure functions (moments of absolute increments) in a midrange of lags, with breakdown in such scaling at small and large lags; extended power-law scaling at all lags; nonlinear scaling of power-law exponent with order of sample structure function; and pronounced statistical anisotropy. The literature has traditionally considered such variables to be multifractal. Previously we proposed a simpler and more comprehensive interpretation that views them as samples from stationary, anisotropic sub-Gaussian random fields or processes subordinated to truncated fractional Brownian motion or truncated fractional Gaussian noise. The variables thus represent mixtures of Gaussian components having random variances. We apply our novel approach to soil data collected at an Arizona field site and to corresponding hydraulic properties obtained by means of a neural network model

---

A. Guadagnini · M. Riva  
Dipartimento Di Ingegneria Civile E Ambientale, Politecnico Di Milano,  
Piazza L. Da Vinci 32, 20133 Milano, Italy  
e-mail: alberto.guadagnini@polimi.it

M. Riva  
e-mail: monica.riva@polimi.it

A. Guadagnini · S.P. Neuman (✉) · M. Riva  
Department of Hydrology and Water Resources, University of Arizona, Tucson  
AZ 85721, USA  
e-mail: neuman@hwr.arizona.edu

M.G. Schaap  
Department of Soil, Water and Environmental Science, University of Arizona, Tucson  
AZ 85721, USA  
e-mail: mschaap@ag.arizona.edu

and estimate their statistical scaling parameters by maximum likelihood. Our approach allows upscaling or downscaling statistical moments of such variables to fit diverse measurement or resolution and sampling domain scales.

**Keywords** Scaling · Multifractals · Sub-Gaussian · Soil properties · Neural network

## 1 Introduction

Many earth, environmental, ecological, biological, physical, astrophysical and financial variables exhibit intermittent space-time fluctuations; symmetric, non-Gaussian frequency distributions of increments characterized by heavy tails that often decay with separation distance or lag; nonlinear power-law scaling of sample structure functions (moments of absolute increments) in a midrange of lags, with breakdown in such scaling at small and large lags; extended power-law scaling at all lags; nonlinear scaling of power-law exponent with order of sample structure function; and pronounced statistical anisotropy. The literature has traditionally interpreted this to imply that the variables are multifractal. We propose a simpler and more comprehensive interpretation that views them as samples from stationary, anisotropic sub-Gaussian random fields or processes subordinated to truncated fractional Brownian motion (tfBm) or truncated fractional Gaussian noise (tfGn). The data thus represent mixtures of Gaussian components having random variances. This allows one to obtain maximum likelihood estimates of all statistical parameters characterizing these fields or processes and to upscale or downscale their statistical moments to fit diverse measurement or resolution and sampling domain scales. Our approach is described in Neuman et al. [4] and references therein.

In this paper we illustrate our novel method of analysis on soil data collected at an Arizona field site and on neural network estimates of soil hydraulic properties at the site. Hydraulic properties (such as volumetric water content, permeability and their functional relations to capillary pressure) required for subsurface flow and transport analyses can be measured in the field and/or the laboratory at a considerable investment of time and money. One alternative is to estimate these properties indirectly by means of pedotransfer functions (PTFs, for a review see [5]) on the basis of pedological indicators such as soil particle size distribution, bulk density and organic matter content that are much simpler and less costly to determine. One of the most powerful and increasingly popular tools of this kind is the nonlinear Rosetta neural network code of Schaap et al. [10], which comprises a set of five hierarchical PTFs tailored to varied circumstances ranging from data-poor to data-rich. Inputs may be limited to soil composition data such as per cent sand, silt and clay or include additional information about soil bulk density and one or two measured pairs of water content and capillary pressure data. Output consists of parameters defining the van Genuchten [13]–Mualem [3] constitutive relationships

between water content, hydraulic conductivity and capillary pressure. The code has been calibrated against pedological and hydraulic data obtained from laboratory analyses of 2,134 soil samples from across the United States. The calibration was combined with the non-parametric bootstrap method [1] to allow assessing Rosetta's predictive uncertainty. Assuming that the calibration data set of 2,134 samples represents correctly the underlying soil population, multiple random subsets (or replicas) of the original dataset were created through sampling with replacement: 100 replicates of saturated hydraulic conductivity and 50 replicates of van Genuchten–Mualem constitutive parameters [9]. Rosetta was calibrated separately against each replicate data set, each calibrated version was used to predict hydraulic parameters on the basis of the original 2,134 input data, and the results summarized in terms of sample mean and standard deviation of each predicted parameter [10]. The latter two statistics are taken to represent the mean and the uncertainty of the corresponding neural network outputs, which vary with each individual set of input data.

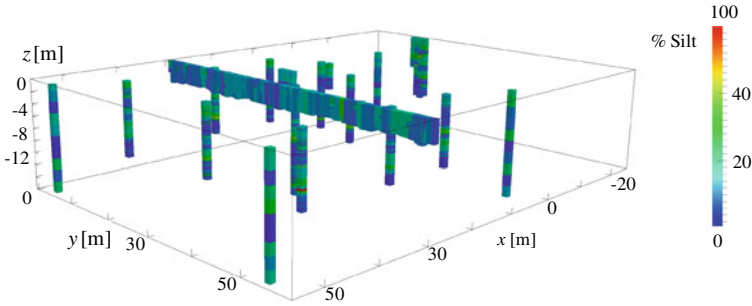
Due to their reliance on diverse data bases obtained using varied measurement techniques, it is not uncommon for different PTFs to produce mutually inconsistent outcomes [9]. Most PTFs have a modest accuracy when estimated hydraulic parameters are compared with experimental values [11]. In the case of Rosetta, correlation coefficients between experimental and estimated constitutive parameters of the van Genuchten water retention model range between 0.3 and 0.9 [10]. The root-mean square error between measured and estimated water contents range from 0.04 to 0.08 cm<sup>3</sup>/cm<sup>3</sup>, depending on model used. Correcting for capillary pressure-dependent bias reduces this error only slightly [11].

In this paper we use our novel approach of data interpretation to analyze the statistical scaling properties of Rosetta inputs and outputs using soil data from an experimental site near Maricopa, Arizona, USA [8].

## 2 Statistical Scaling of Neural Network Inputs

We start by analyzing the statistical scaling behavior of soil texture data measured to a depth of 15 m over an area of 3,600 m<sup>2</sup> at the Maricopa experimental site, operated by the University of Arizona (headquartered in Tucson). These data constitute inputs into the Rosetta neural network model. The sampling network, depicted in Fig. 1, comprises 1,029 measurement locations distributed along several vertical wells and a horizontal transect. A more complete description of the site and the network is given by Schaap [8].

Our texture data consist of relative fractions  $f_i$ ,  $0 \leq f_i \leq 1$ , of three texture categories  $i = sa, si$  and  $cl$  representing sand, silt and clay, respectively. In addition to the original measurements,  $f_i$ , we also consider two corresponding principal components, PC1 and PC2, as defined by Schaap [8]. Here we focus on statistical scaling of vertical increments in these variables.

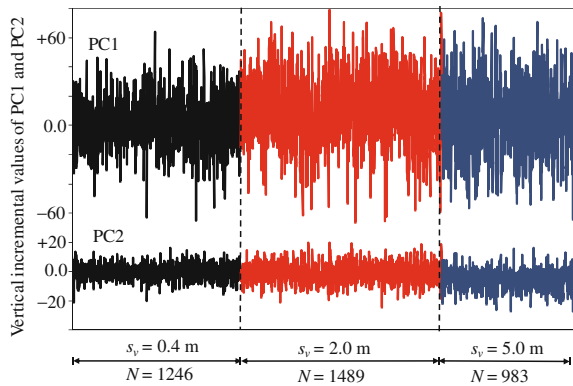


**Fig. 1** Spatial distribution of soil sampling network at Maricopa experimental site. Colour scale represents measured relative silt fraction,  $f_{si}$

Figure 2 concatenates sequences of vertical increments in PC1 and PC2 computed along the various sampling boreholes in Fig. 1, at vertical separation distances (lags)  $s_v = 0.4, 2.0$  and  $5.0$  m. Increments for a given lag in Fig. 2 are computed from top to bottom along each sampling borehole in Fig. 1 and arranged in sequence by placing values associated with all sampled boreholes next to each other. The increments are seen to vary randomly.

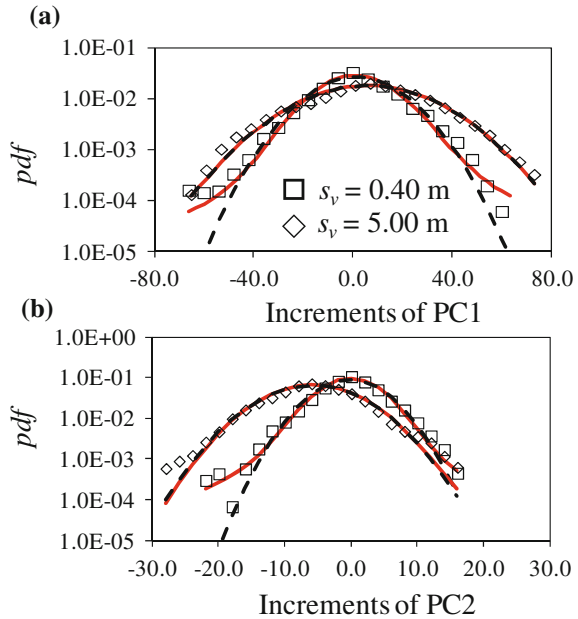
Frequency distributions of vertical increments, like those of the principal components PC1 and PC2 in Fig. 3, tend to be symmetric and exhibit heavy tails. As illustrated in Fig. 3, they can be fitted quite well by the maximum likelihood (ML) method to  $\alpha$ -stable probability density functions (pdfs) with stability indices  $\alpha \leq 2$ , where  $\alpha = 2$  corresponds to the normal (Gaussian) pdf. ML fits of normal pdfs to the empirical distributions are included in Fig. 3 for reference. Whereas the tails of  $\alpha$ -stable pdfs with  $\alpha < 2$  fall off as a power law, those of the normal pdf decay exponentially. ML estimates of  $\alpha$  associated with vertical increments of PC1 and PC2 increase from 1.85 at a lag of 0.4 m to 2 at lags exceeding 2 m. Kolmogorov–Smirnov and Shapiro–Wilk tests at significance level of 0.05 do not, in most cases, support a hypothesis that increments associated with estimates of  $\alpha > 1.9$  derive from a normal pdf.

**Fig. 2** Concatenated sequences of  $N$  vertical increments in PC1 and PC2 at vertical lags  $s_v = 0.4, 2.0$  and  $5.0$  m





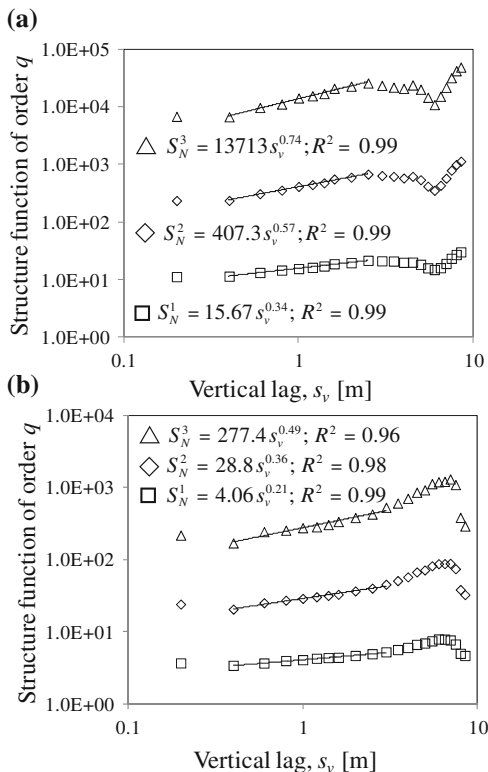
**Fig. 3** Frequency distributions of increments of (a) PC1 and (b) PC2 at two lags. Also shown are ML fits of  $\alpha$ -stable (red solid) and normal (dashed) pdfs



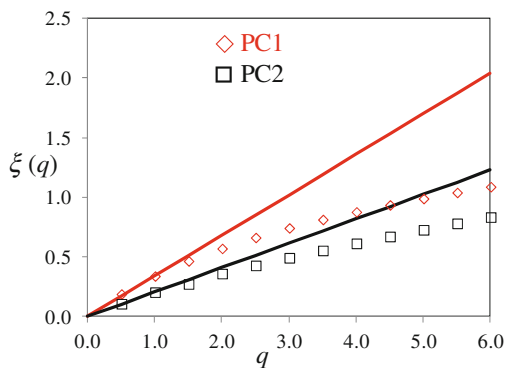
Next we compute structure functions  $S_N^q$  defined as  $q$ th order sample statistical moments of absolute vertical increments in a sample of size  $N$ . Figure 4 plots sample structure functions of orders 1, 2 and 3 associated with vertical increments of PC1 and PC2 as functions of vertical lag on logarithmic scale. In each case there is a mid-range of lags within which the data can be fitted by regression to straight lines at high levels of confidence as indicated by coefficients of determination,  $R^2$ , close to 1. This implies that, in a midrange of lags, each structure function scales as a power of lag; Fig. 4 lists corresponding power exponents, which we designate by  $\zeta(q)$ , ranging from 0.34 to 0.74 in the case of PC1 and from 0.21 to 0.49 in the case of PC2. We refer to this way of determining power scaling exponents for various orders  $q$  of a structure function as method of moments (M).

Figure 5 shows how the power-law scaling exponent,  $\zeta(q)$ , determined for PC1 and PC2 by the method of moments, varies with the order  $q$  of their structure functions up to  $q = 6$ . The exponent  $\zeta(q)$  is seen to scale in a nonlinear fashion with  $q$ , delineating a convex curve. Included in Fig. 5 are straight lines passing through  $\zeta(1)$  and the origin.

Power-law scaling of  $\alpha$ -stable increments such as those illustrated in Figs. 4 and 5, including breakdown in power-law scaling at small and large lags and nonlinear variation of the power-law scaling exponent  $\zeta(q)$  with  $q$ , have been shown by us elsewhere to be typical of samples from sub-Gaussian random fields or processes subordinated to truncated fractional Brownian motion (tfBm) and/or truncated fractional Gaussian noise (tfGn); for up-to-date descriptions consult Guadagnini et al. [2], Neuman et al. [4], Riva et al. [6, 7] and, Siena et al. [12]. Whereas

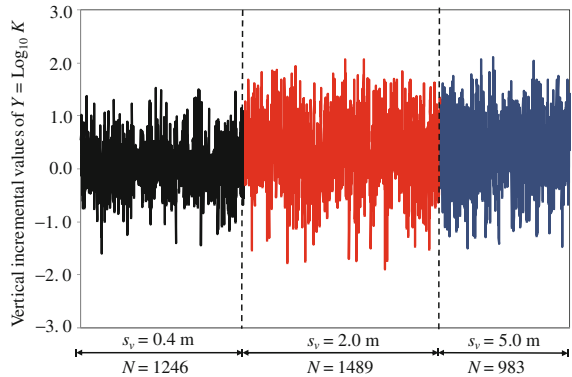


**Fig. 4** Structure functions of order  $q = 1, 2$  and  $3$  of vertical (a) PC1 and (b) PC2 increments versus lag. Regression lines ( $R^2$  values listed) indicate power-law scaling (equations listed) in midranges of lags



**Fig. 5** Variations of power-law scaling exponent  $\zeta(q)$  corresponding to PC1 and PC2 with order  $q$  of their respective structure functions obtained by the method of moments. Straight lines pass through  $\zeta(1)$  and the origin

**Fig. 6** Concatenated sequences of  $N$  vertical increments of log saturated hydraulic conductivity,  $Y = \log_{10}K$ , at lags  $s_v = 0.4, 2.0$  and  $5.0$  m



nonlinear variation of  $\zeta(q)$  with  $q$  had previously been attributed in the literature to multifractals and/or fractional Laplace motions, we note that fBm and/or fGn are monofractal self-affine.

Like fBm and fGn, their truncated tfBm and tfGn versions are characterized by a single power-law scaling exponent,  $H$ , known as the Hurst coefficient. One way to estimate  $H$  is from the slope of a straight line that passes through  $\zeta(1)$  and  $\zeta(0)$ . The two straight lines in Fig. 5 thus imply that PC1 is characterized approximately by a Hurst exponent  $H = 0.34$  and PC2 by  $H = 0.21$ . Both estimates are smaller than corresponding estimates of  $1/\alpha$ , implying that PC1 and PC2 are anti-persistent in the vertical direction, varying in a rough rather than in a smooth manner as indeed do the underlying textural indicators (not shown here), consistent with Fig. 2.

Similar statistical scaling behaviors are exhibited by other Rosetta input variables.

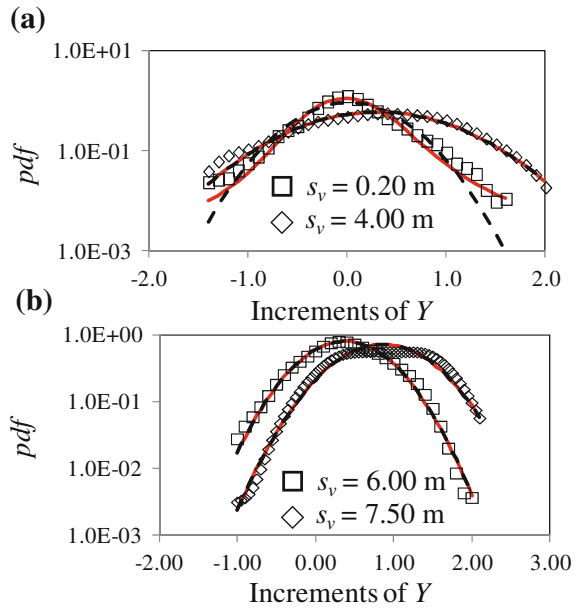
### 3 Statistical Scaling of Neural Network Outputs

Having characterized statistical scaling of Rosetta inputs, we now perform a corresponding analysis of selected outputs generated by the neural network model. Rosetta generates output hydraulic soil properties at all sampling locations at the Maricopa experimental site (Fig. 1). Here we focus on statistical scaling of vertical increments of log hydraulic conductivity,  $Y = \log_{10}K$ , at full soil saturation. Figure 6 concatenates sequences of such increments computed by Rosetta along the various sampling boreholes in Fig. 1, at vertical separation distances (lags)  $s_v = 0.4, 2.0$  and  $5.0$  m.

The increments are seen to vary randomly, as did the corresponding Rosetta inputs and their principal components in Fig. 2.

Frequency distributions of vertical  $Y = \log_{10}K$  increments in Fig. 7 tend to be symmetric and exhibit heavy tails, as did those of Rosetta input variables and PC1 and PC2 in Fig. 3. Like the latter, frequency distributions of Rosetta output estimates in Fig. 7 can be fitted reasonably well by ML to  $\alpha$ -stable pdfs with stability

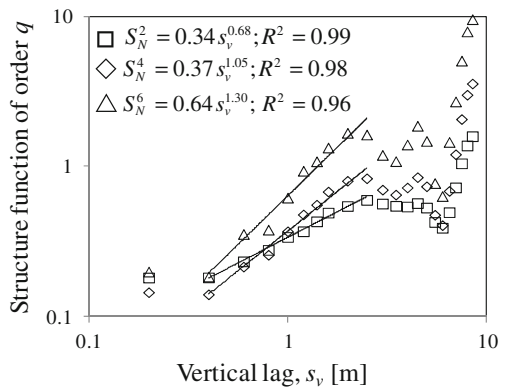
**Fig. 7** Frequency distributions of  $Y = \log_{10}K$  at four lags. Also shown are ML fits of  $\alpha$ -stable (red solid) and normal (dashed) pdfs

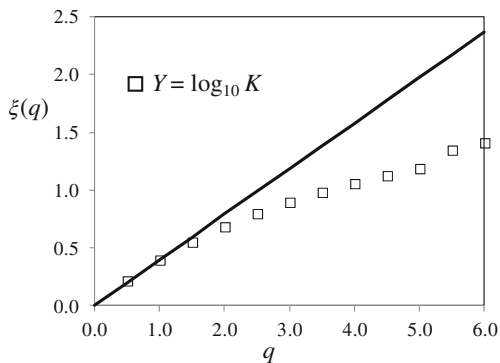


indices  $\alpha \leq 2$ . ML fits of normal pdfs to the empirical distributions are included in Fig. 7 for reference. ML estimates of  $\alpha$  associated with vertical  $Y = \log_{10}K$  increments increase from 1.68 at a lag of 0.2 m to 2.0 at lags exceeding 0.8 m. Kolmogorov–Smirnov and Shapiro–Wilk tests at significance level of 0.05 yield ambiguous results, neither overwhelmingly supporting nor clearly rejecting a hypothesis that increments associated with estimates of  $\alpha > 1.9$  derive from a normal pdf.

Figure 8 plots sample structure functions of integer orders  $q = 2, 4, 6$  associated with vertical  $Y = \log_{10}K$  increments as functions of vertical lag on logarithmic scale. As in the case of Rosetta inputs and PC1 and PC2 (Fig. 4), here again each sample

**Fig. 8** Structure functions of integer orders  $q = 2, 4, 6$  of vertical  $Y = \log_{10}K$  increments versus lag. Regression lines ( $R^2$  values listed) indicate power-law scaling (equations listed) in midranges of lags





**Fig. 9** Variations of power-law scaling exponent  $\zeta(q)$  corresponding to  $Y = \log_{10}K$  with order  $q$  of its structure function obtained by the method of moments. Straight lines pass through  $\zeta(1)$  and the origin

structure function exhibits a mid-range of lags within which it can be fitted by regression to a straight line at a high level of confidence as indicated by coefficients of determination,  $R^2$ , close to 1. In this midrange of lags, each structure function scales as a power of lag; Fig. 8 lists corresponding power exponents  $\zeta(q)$  ranging from 0.68 to 1.30.

Figure 9 shows how the power-law scaling exponent,  $\zeta(q)$ , determined for Rosetta output log saturated hydraulic conductivities by the method of moments, varies with the order  $q$  of its structure function up to  $q = 6$ . As in the case of Rosetta inputs and PC1 and PC2 in Fig. 5,  $\zeta(q)$  delineates a convex curve. Included in Fig. 9 are straight lines passing through  $\zeta(1)$  and the origin. The latter yields an estimated Hurst exponent  $H = 0.39$  which, like in the Rosetta input case, is smaller than corresponding estimates of  $1/\alpha$  and thus imply that  $Y = \log_{10}K$  is anti-persistent in the vertical direction, varying in a rough rather than in a smooth manner as do the Rosetta input variables.

Similar statistical scaling behaviors are exhibited by other Rosetta output variables.

## 4 Conclusions

We have analyzed, and presented selected examples of, the statistical scaling behaviors of soil pedological indicators at an experimental site in southern Arizona that have served as inputs into a neural network model of soil properties at the site. We have conducted a similar analysis on soil hydraulic property predictions by the same neural network model and illustrated them on log saturated hydraulic conductivity model outputs. We found that, like the neural network inputs (and many other earth, environmental as well as other variables), our neural network output predictions exhibited the following statistical scaling behaviours:

1. Symmetric frequency distributions of spatial increments (illustrated in vertical but observed also in horizontal directions) tending to possess heavy tails.
2. Good maximum likelihood fits of increment frequency distributions to  $\alpha$ -stable probability density functions with power-law tails.
3. Structure functions scaling as powers of separation distance, or lag, in intermediate ranges of lags.
4. Breakdown in such power-law scaling at small and large lags.
5. Nonlinear convex scaling of power-law exponents with order of the corresponding structure functions.
6. Highly irregular, anti-persistent spatial variability characterized by relatively small Hurst exponent estimates.

Such behavior has been shown by us elsewhere to be characteristic of samples from sub-Gaussian random fields or processes subordinated to truncated fractional Brownian motion (tfBm) and/or truncated fractional Gaussian noise (tfGn). Whereas nonlinear scaling of power-law exponents with structure function order had previously been attributed in the literature to multifractals and/or fractional Laplace motions, we note that fBm and/or fGn are monofractal self-affine.

Future work will focus on ways to condition sub-Gaussian random fields or processes on multiscale, space-time distributed earth and environmental measurements and on the statistical scaling of corresponding extreme values and/or events.

**Acknowledgments** Our work was supported in part through a contract between the University of Arizona and Vanderbilt University under the Consortium for Risk Evaluation with Stakeholder Participation (CRESP) III, funded by the U.S. Department of Energy.

## References

1. Efron B, Tibshirani R (1993) *An Introduction to the Bootstrap*. Chapman & Hall/CRC, Boca Raton
2. Guadagnini A, Riva M, Neuman SP (2012) Extended power-law scaling of heavy-tailed random air-permeability fields in fractured and sedimentary rocks. *Hydrol Earth Syst Sci* 16:3249–3260. doi:[10.5194/hess-16-3249-2012](https://doi.org/10.5194/hess-16-3249-2012)
3. Mualem Y (1976) A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resour Res* 12(3):513–522
4. Neuman SP, Guadagnini A, Riva M, Siena M (2013) Recent advances in statistical and scaling analysis of earth and environmental variables. *Recent advances in hydrogeology*. Springer, New York, pp 1–15
5. Pachepsky Y, Rawls WJ (eds) (2004) *Development of pedotransfer functions in soil hydrology*. Elsevier, Amsterdam, The Netherlands
6. Riva M, Neuman SP, Guadagnini A (2013a) Sub-Gaussian model of processes with heavy tailed distributions applied to permeabilities of fractured tuff. *Stoch Environ Res Risk Assess* 27:195–207. doi:[10.1007/s00477-012-0576-y](https://doi.org/10.1007/s00477-012-0576-y)
7. Riva M, Neuman SP, Guadagnini A, Siena M (2013b) Anisotropic scaling of Berea sandstone log air permeability statistics. *Vadose Zone J* 12(3). doi:[10.2136/vzj2012.0153](https://doi.org/10.2136/vzj2012.0153) (in press)

8. Schaap MG (2013) Description, analysis and interpretation of an infiltration experiment in a semi-arid deep vadose zone. *Recent advances in hydrogeology*. Springer, New York, pp 159–183
9. Schaap MG, Leij FJ (1998) Database related accuracy and uncertainty of pedotransfer functions. *Soil Sci* 163:765–779
10. Schaap MG, Leij FJ, van Genuchten MTh (2001) Rosetta: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *J Hydrol* 251:163–176
11. Schaap MG, Nemes A, Van Genuchten MTh (2004) Comparison of models for indirect estimation of water retention and available water in surface soils. *Vadose Zone J* 3:1455–1463
12. Siena M, Guadagnini A, Riva M, Neuman SP (2012) Extended power-law scaling of air permeabilities measured on a block of tuff. *Hydrol Earth Syst Sci* 16:29–42. doi:[10.5194/hess-16-29-2012](https://doi.org/10.5194/hess-16-29-2012)
13. van Genuchten MTh (1980) A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci Soc Am J* 44:892–898

# Adaptive Neuro Fuzzy Inference System Used to Build Models with Uncertain Data: Study Case for Rainfed Maize in the State of Puebla (Mexico)

Anáis Vermonden, Carlos Gay-García and Iván Paz-Ortiz

**Abstract** A model was built using Adaptive Neuro Fuzzy Inference System (ANFIS) to determine the relationship between the natural suitability index of rainfed maize and yield per hectare and percentage of production area lost for the state of Puebla. The data used to build the model presented inconsistencies. The data of the INEGI's land use map presented more municipalities without rainfed maize agriculture than the database of SAGARPA. Also the SAGARPA data, in terms of the percentage of production area lost, do not mark any distinctions of the loss. Even with data inconsistencies ANFIS produced a coherent output reviewed by experts and local studies. The model shows that higher the percentage of production area lost and high yields, the higher the suitability index is. According to local studies this is due to the high degradation of the soils and confirmed with the second model built adding soil degradation information.

**Keywords** Fuzzy logic · Agriculture · Adaptive neuro fuzzy inference system · ANFIS · Uncertain data · Mexico

## 1 Introduction

The municipality of Tehuacan in the state of Puebla, is considered to be one of the places where maize originated, since archaeologists have found some of the oldest maize fossil dating 5,500 y.b.p. [9]. This case of study for the state of Puebla was

---

A. Vermonden (✉) · C. Gay-García · I. Paz-Ortiz  
Programa de Investigación En Cambio Climático, Universidad Nacional Autónoma de México, Edificio de Programas Universitarios, Circuito Exterior CU,  
04510 Coyoacan, Mexico  
e-mail: anaisvt@gmail.com

C. Gay-García  
e-mail: cgay@servidor.unam.mx

I. Paz-Ortiz  
e-mail: ivnpaz@gmail.com



developed to understand the relationship between the natural suitability index for rain-fed maize, the yield per hectare and the percentage of the area lost. Today 62 % of the cultivated land in the state of Puebla is destined to grow maize, 70 % is rainfed, and up to 80 % are varieties of native maize [15].

The natural suitability index for rainfed maize was calculated using the mean temperature, mean precipitation, soil depth and slope. According to the experts, for an area to be suitable the four variables must be suitable, even if one of the variables is in another class then the index will belong to that class. The index of suitability for rainfed maize was calculated with a fuzzy model based on expert knowledge [14] and on the previous work of Monterroso [11]. The Index had a resolution of 1 km by 1 km. The natural suitability index helps determine which lands are best to practice rainfed agriculture. These can also be projected under different climate change scenarios to determine areas that will be more vulnerable as the suitability index changes. How is the natural suitability index related to other variables, and could these also be projected into a future made more uncertain due to climate change?

At state level the information was provided by two sources, the Secretaría de Agricultura, Ganadería, Desarrollo Social, Pesca y Alimentación (SAGARPA) and the Instituto Nacional de Estadística y Geografía (INEGI). The information from both government bodies came at different scales: SAGARPAs is at municipality level; INEGIs was obtained at 1 km × 1 km, making it coincide with the calculated natural suitability index for rainfed maize.

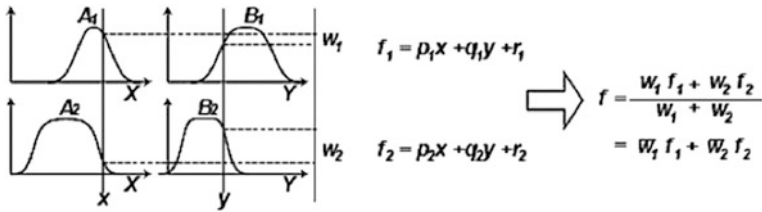
When examining the data from INEGI and SAGARPA, inconsistencies started to show. According to the data from SAGARPA [12] only three municipalities in the state of Puebla have no rainfed maize agriculture, Altepexi, Atzala and Zinacatepec, while the map of land use from INEGI [6] shows 16 other municipalities lacking rainfed agriculture, see Fig. 2. Due to the scale these small areas do not show on the map.

The data record available from SAGARPA at municipality level goes from 2003 to 2012. In previous years the municipalities were grouped in Rural Development Districts, but how they are grouped is not published.

In the study developed with only two variables, soon it became apparent a third variable was needed to explain the results, as higher suitability index had higher yields and higher percentage in area lost. Further research claimed soil degradation was a big factor in loss of rainfed maize production in the state. As it can be observed in Fig. 7, in the state of Puebla nearly 60 % of the land used for rainfed agriculture shows soil degradation.

## 2 Method

Adaptive Neuro Fuzzy Inference System (ANFIS), represents a Sugeno-type neuro-fuzzy system. By integrating both neural networks and fuzzy logic principles, it has the potential to capture the benefits of both in a single framework. It has the ability to construct sets of fuzzy if-then rules to approximate nonlinear functions. ANFIS



**Fig. 1** Diagram of a Sugeno model (evaluation of two fuzzy rules with two input variables, i.e. A and B)

can also build appropriate membership functions to generate the stipulated input-output pairs to be used in the model [8]. The Neuro-adaptive learning techniques provide a method for building a fuzzy model from the information contained in a data set. The fuzzy system enables flexibility in the variables and the representation of incomplete data, as membership to a fuzzy set is denoted by the degree of membership to the set. Since the ANFIS can deduce relations between the inputs/outputs, ANFIS forms an input output mapping based both on human knowledge (based on fuzzy if-then rules) and generated input/output data pairs by using a hybrid algorithm that is the combination of the gradient descent and least square estimates [7]. The main characteristic of the Sugeno inference system is that the consequent, or output of the fuzzy rules is a function, as shown in Eq. 1.

$$\begin{aligned}
 R1 : & \text{If } A \text{ is } A1 \text{ and } B \text{ is } B1 \text{ the } f1 = p1 * a + q1 * b + r1 \\
 R2 : & \text{If } A \text{ is } A2 \text{ and } B \text{ is } B2 \text{ the } f2 = p2 * a + q2 * b + r2
 \end{aligned}
 \tag{1}$$

Figure 1 graphically describes the inference process of a Sugeno model composed by the two rules described in Eq. 1.

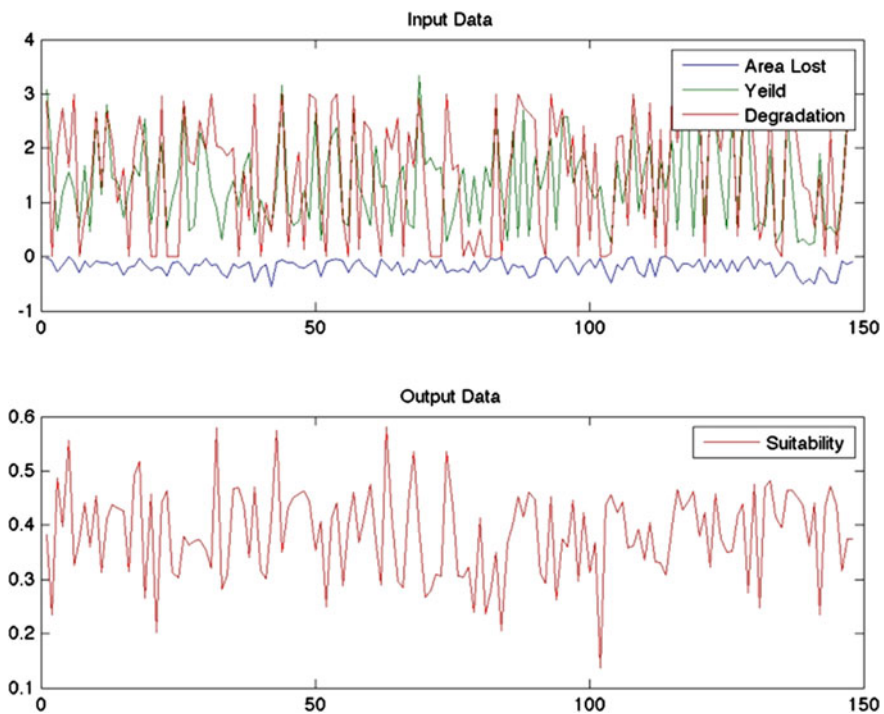
The first step combines a given input tuple (Fig. 1),  $x$  and  $y$ , through antecedent rules by determining the degree to which each input belongs to the corresponding fuzzy set. The min operator is used to obtain the weight of each rule, which is later used in the final output computation,  $f$ . Sugeno has two differentiated set of parameters, the first set corresponds to the input variable and the second to the output function of each rule, i.e.  $p_i, q_i$ , and  $r_i$ . ANFIS uses two optimization algorithms to automatically adjust the two sets of parameters. Back-propagation (gradient descent) to learn the parameters of the antecedents (membership functions) and least square estimation is used to determine the coefficients of the linear combinations in the rules' consequents.

### 3 Data

The examination of the data obtained from SAGARPA and the INEGI showed inconsistencies. According to the data from SAGARPA only three municipalities in the state of Puebla have no rainfed maize agriculture, Altepexi, Atzala and



**Fig. 2** Map of the state of Puebla showing the municipalities division and in green land use for rainfed agriculture



**Fig. 3** Input and output data for the first model. Input data on top, percentage of area lost (blue) and yield (green). Output data on the bottom natural suitability index (red)

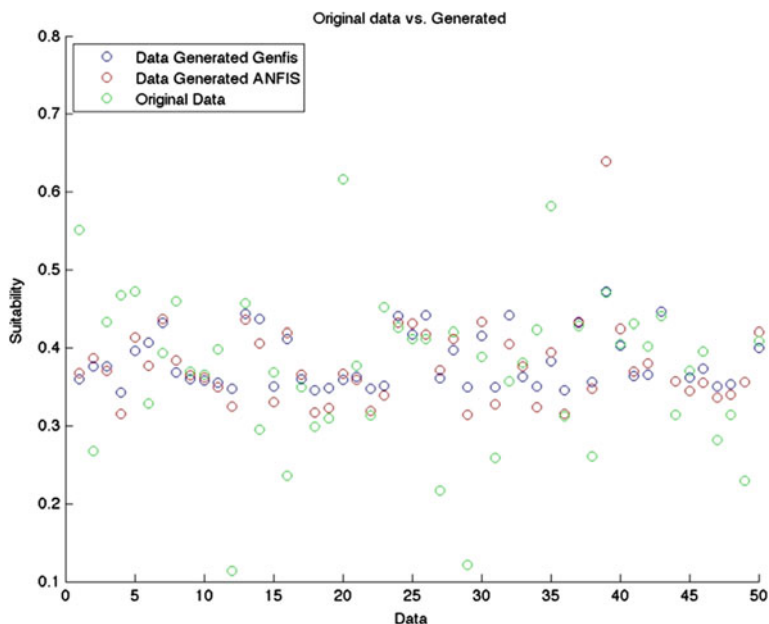
Zinacatepec, while the map of land use from INEGI [6] show another 16 municipalities without rainfed agriculture, see Fig. 2. Another assumption made for the land use map for rainfed agriculture is that in all the area presented maize is being cultivated. Maize is the most important cereal in the Mexican diet and 92 % of the farmers in Mexico that own between 0 and 5 ha produce 56.4 % of the countrys maize by the rainfed farming practice. In the state of Puebla 60 % of the land used for rainfed agriculture is used to cultivate maize; but how much of the land used for this crop varies between areas, in some of them, for example the municipality of Cohetzala, it is as high as 95 %. Therefore the assumption is: the areas presented as rainfed agriculture are rainfed maize agriculture.

Since the data from SAGARPA is at the municipality level for the period of 2003–2012, the map of land uses published by INEGI was used as a mask to extract the data of the suitability index, since it would correspond to the area marked where rainfed maize agriculture was being produced. An average was calculated to obtain the natural suitability index for rainfed maize to an equivalent scale of the following two variables. Figure 3 shows the values of percentage of production area lost and yield per hectare as well as the natural suitability index for rainfed maize. The land degradation map was obtained from INEGI, published in 2002. The mask from the land use map was once again used to calculate an average to obtain data at the same scale. In Fig. 6 a graphic representation of the values for the input data in addition the level of soil degradation is shown.

## 4 Results

To determine the relationship between the three variables, a subtractive clustering algorithm [1] was used to generate a fuzzy system. This algorithm allows to estimate the number of clusters and their centers, to later build the membership functions and the relations between the variables. First, the centers are established through subtractive clustering methods [2], once the centers are calculated their radius of influence are determined. For each data of the set a potential measure is calculated to check the center of the cluster using the density of surrounding data. This is done to identify natural groupings of data from a large set, allowing concise representation of embedded relationships. In this case four clusters were calculated, thus reducing the complexity of the sets and the analysis of the relationship between the variables. The clusters calculated were used in the training of the model.

The clusters were used to generate the if-then rules and membership functions. The information was added to the `genfis2` function [10] and 75 % of the data set was used to train and generate a fuzzy inference system (FIS) Sugeno type [13]. First, with the information obtained from the subclustering method, it determines the number rules, antecedent membership functions and uses the least square estimation to determine each rules consequent equations; then returning a FIS structure that contains a set of fuzzy rules to cover the featured space.



**Fig. 4** Graph showing the 25 % of the data used to verify the model, in *green circles* the original data, in *blue circles* the data generated by the first model with genfis2 and in *red circles* the data generated with ANFIS

The remaining 25 % of the data set is then used to verify the model. To verify the model the root mean square error of the system generated by the training data was calculated to be 0.0766. The root mean square error of the system used for both checking and testing the FIS parameters was 0.0953. Both very close to zero.

ANFIS is used to improve the capacity of the FIS to model data. Again 75 % of the data is used to train the neuro-adaptive network. In this case the hybrid optimization method was used, which combines gradient descent and the least squares method. The gradient descent is used on the premise parameters that define the membership functions; for the consequent parameters that define the coefficients of each output equations the least squares method is used. A hundred Epochs were used and the training error tolerance was set to zero. Stability of the training was achieved before Epoch 30. To verify the model the root mean square error of the system generated by the training data was calculated to be 0.0745. The root mean square error of the system used for both checking and testing the FIS parameters was 0.0937, improving the previous FIS generated by genfis2. See Fig. 4.

The fuzzy surface of the rules generated with the data (Fig. 5) show that the areas with the highest suitability index have the highest percentage of production area lost (Fig. 6).

A study of the municipality of Molcaxac [3], which has a high suitability index for the period of 2002–2003, informed that 35 % of the total production of the cereal was lost due to soil degradation. A third variable would further explain the

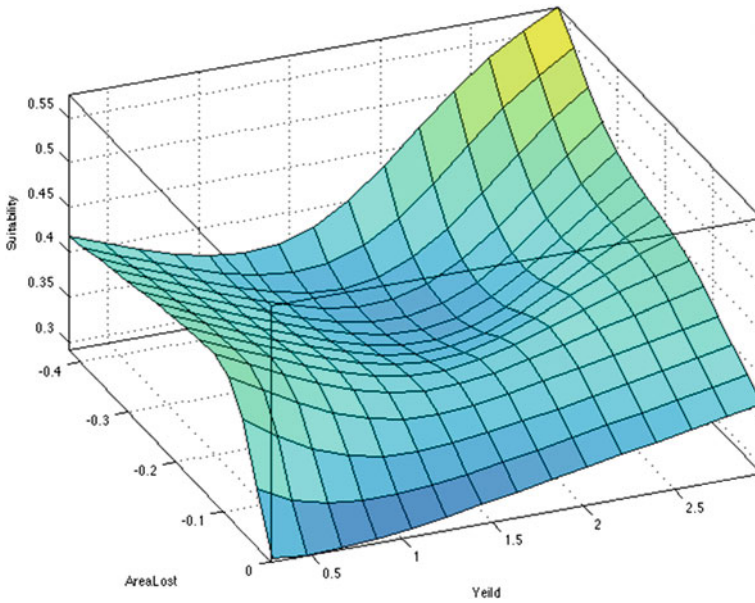


Fig. 5 Graph of the surface created by the rules of the FIS, using two variables: yields and percentage of area lost

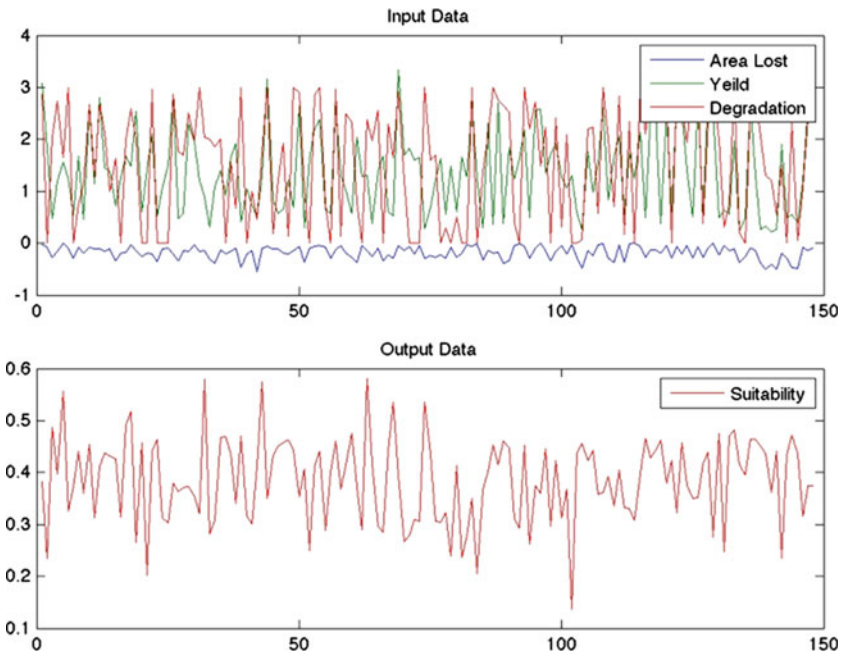


Fig. 6 Input and output data for the second model. Input data on top, percentage of area lost (blue), yield (green) and level of soil degradation (red). Output data on the bottom natural suitability index (red)

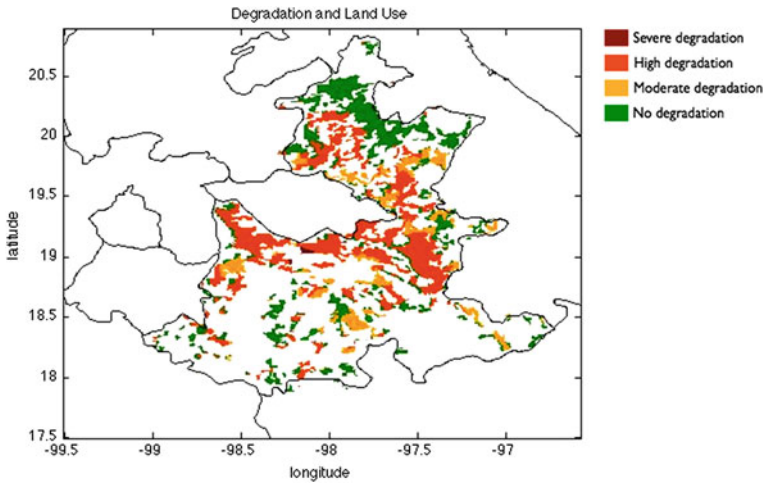


Fig. 7 Map of the state of Puebla showing level of degradation on land use for rainfed agriculture. With data from [5]

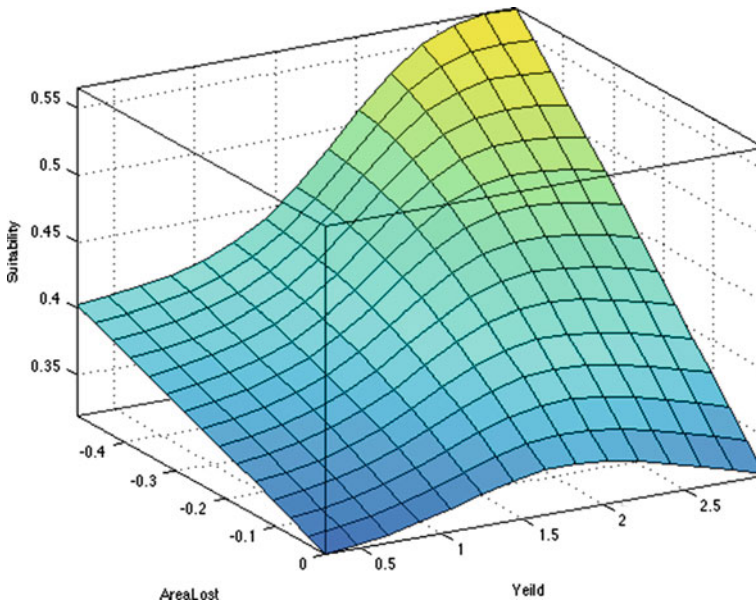
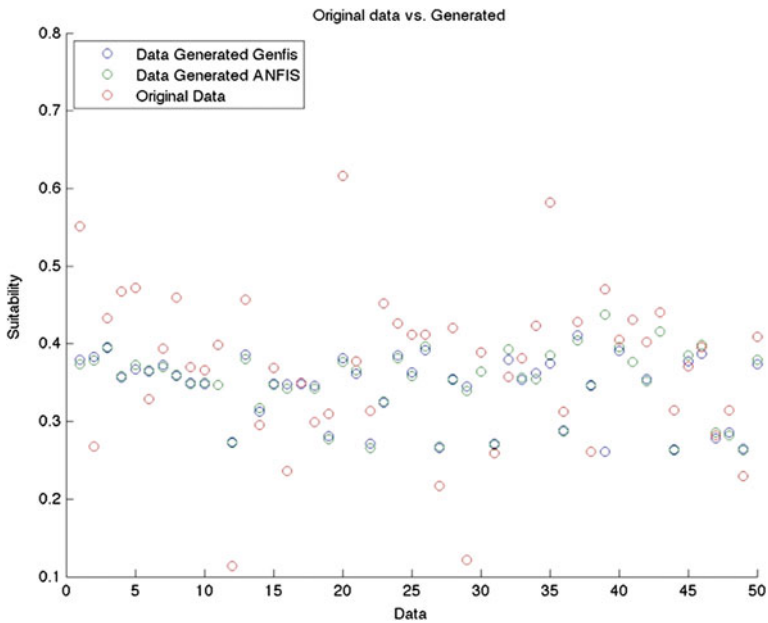


Fig. 8 Graph of the surface generated by the rules using three variables: yields, percentage of area lost and soil degradation





**Fig. 9** Graph showing the 25 % of the data used to verify the model with the three variables, in *green circles* the original data, in *blue circles* the data generated by the first model with genfis2 and in *red circles* the data generated with ANFIS

results obtained. Figure 7 shows how 60 % of the land used for rainfed agriculture presents some level of soil degradation. The two major types affecting these areas are chemical degradation and soil erosion. Both are directly linked to poor agricultural practices.

Repeating the procedures described above, but this time adding the soil degradation variable, confirmed that the areas of higher suitability, high yields and higher percentage of area lost were affected by soil degradation, as shown in Fig. 8. The signal of degradation was set to the max, showing the behavior obtained with the model with only two variables.

This model was also improved with the use of ANFIS since the root mean square error of the system generated by the training data was calculated to be 0.0757, and became 0.0752. The root mean square error of the system used for both checking and testing the FIS parameters was 0.0868, it improved slightly to 0.0814. The improvements can be seen on the graph in Fig. 9.

This model shows the importance of tackling soil degradation as a measure of adaptation to climate change, since today the most suitable areas are losing their productive capacity.



## 5 Conclusions

The state of Puebla is known for the origin of cultivated maize. The methodology used was the subtractive clustering analysis and ANFIS, to establish the relationships between the suitability index for rainfed maize and the other variables. This preliminary model shows that where suitability is higher the area lost is higher, soil degradation being one of the reasons. This has a very logic explanation as the most suitable areas have been used over the years to produce rainfed maize under poor agricultural practices. This can be verified, one of the major degradation types found over these lands is chemical degradation, which reduces or eliminates the biological productivity of the soil. The methodology used allows users to build models even when the data presents uncertainties. These models can always be improved with better information, expert knowledge and field validation. This first approach presents an important advance for case studies in regions were there is little information and with obvious restrictions. In this case the first restriction was the scale at which the two sources presented the data. It was important to use both to obtain an average of the suitability index at municipality level. If the land use mask had not been used the suitability index would have become insignificant. The other source of uncertainty was to assume that all the areas marked on the land use map for rainfed agriculture were maize. But the municipalities marked on the land use map had rainfed maize production according to the data obtained from SAGARPA, as many of the subsistence farmers in Mexico will grow maize in association with other crops, mainly beans and zucchinis.

SAGARPA data had other restrictions. The earliest data published at municipality level is from the year 2003. Older data are presented as Rural Development Districts, which groups together municipalities, but how they are grouped is unknown as this information is not published. The data used for this study is from 2003 to 2012. Also, the lost production area has no distinctions on why this happened; it could be climate, pests, even that farmers did not plant the area they had first declared or did not harvest all the area planted (which occurs when maize prices fall and no longer compensate the harvesting cost, this way they can claim insurance).

A case study at the municipality of Cohetzala, in the south west part of the state of Puebla [4], showed 98.5 % of the 1,308 ha were destined to grow maize and 95 % is under rainfed method. This municipality has a high suitability index and low degradation. This study found that most of the farmers in this municipality used traditional practices to grow maize: native seeds, association and crop rotation, water and soil conservation techniques, and use of manure as fertilizer (as they are more cost effective), instead of the recommendations made by the Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP) a subdivision of SAGARPA that promotes the use of fertilizers, pesticides, herbicides and hybrid seeds. The yields per hectare are low, but Huato et al. argue that what marked the biggest difference in yield was the use of irrigation (one of the reasons can be that the area has regosols soils), and the comparison between the farmers using traditional practices to the ones using INIFAP recommendations indicate that the

differences are insignificant. This kind of information could be added to the model and create distinctions between traditional practices and more modern practices to determine how it affects the yields, degradation and the percentage of area lost.

These models show that agriculture is a complex human system, but this first approach is showing there is an alarming trend in the state of Puebla due to soil degradation. Degradation must be a priority in the adaptation policies to climate change, since, in the present day scenario, it causes very high losses.

This kind of models facilitate decision making processes even in situations with uncertain data.

**Acknowledgments** The present work was developed with the support of the Programa de Investigación en Cambio Climático (PINCC) of the Universidad Nacional Autónoma de México (UNAM) and the Consejo Nacional de Ciencia y Tecnología (Conacyt). We would like to thank Dr. Cecilia Conde, Dr. Alejandro Monterroso and M. Guillermo Rosales for their valuable inputs and serving as the experts to validate the model.

## References

1. Chiu SL (1994) Fuzzy model identification based on cluster estimation. *J Intell Fuzzy Syst* 2 (3):267–278
2. Dubois D, Prade H (1980) *Fuzzy sets and systems: theory and applications*. Academic Press, New York
3. Gaspar Angeles E, Ortiz Torres E et al (2010) Caracterización y rendimiento de poblaciones de maíz nativas de Molcaxac, Puebla. *Revista fitotecnica mexicana* 33(4)
4. Huato MAD, Cruz Leon A et al (2012) Management of corn at Cohetzala, Puebla, Mexico: between the local and the global. *Estudios Sociales XX*(40)
5. INEGI (2002) Degradación de suelos. Degradación del medio ambiente, SEMARNAT and Colegio de Postgraduados
6. INEGI (2005) *Uso del Suelo y Vegetación*
7. Jang J-SR (1993) ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans Syst Man Cybern* 23(3):665–685
8. Jang J-SR, Sun C-T (1997) *Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence*. Prentice Hall, Upper Saddle River
9. Long A, Benz BF et al (1989) First direct AMS dates on early maize from Tehuacan, Mexico. *Radiocarbon* 31(3):1035–1040
10. MATLAB (2009) *Fuzzy logic toolbox. I. The MathWorks*
11. Monterroso AI, Conde C et al (2009) Assessing current and potential rainfed maize suitability under climate change scenarios in Mexico. *Atmósfera* 24(1):53–57
12. SAGARPA, SIAP (2003–2012). *Producción anual de maíz del estado de Puebla*
13. Sugeno M (1977) *Fuzzy measures and fuzzy integrals: a survey*. North-Holland, NY
14. Vermonden A (2012) *Modelo Difuso para la Evaluación de la Aptitud Actual y Potencial del Maíz de Temporal en México con Cambio Climático*. Mexico, Universidad Nacional Autónoma de México. Maestría, Posgrado de Ciencias de la Tierra
15. Viveros Flores CE (2010) *Estudio de la Dinámica de Aprovechamiento del Maíz en las Unidades de Producción Familiar en el Valle de Puebla, México*. Postgrado de estrategias para el desarrollo agrícola regional. Puebla, Colegio de Postgraduados. PhD, p 119

# Separation of Carbon Dioxide from Synthesis Gas Containing Steam by Pressure Swing Adsorption at Mid-high Temperature

Cheng-tung Chou, Yu-Hau Shih, Yu-Jie Huang and Hong-sung Yang

**Abstract** This study aimed to utilize a pressure swing adsorption (PSA) process to capture CO<sub>2</sub> from synthesis gas, which is the effluent stream of water-gas-shift reactor. The PSA process studied is a single-bed four-step process at mid-high temperature using K<sub>2</sub>CO<sub>3</sub>-promoted hydrotalcite as adsorbent. The breakthrough curve and desorption curve were verified against the simulation program which our group developed. It uses the method of lines combined with upwind differences, cubic spline approximation and LSODE of ODEPACK software to solve the equations. The optimal operating condition is obtained by varying the operating variables, such as feed pressure, bed length, etc. Furthermore, single-bed four-step process could achieve 98.49 % recovery of H<sub>2</sub> as the top product and 96.42 % purity and 96.57 % recovery of CO<sub>2</sub> as the bottom product.

**Keywords** Pressure swing adsorption · CO<sub>2</sub> capture · Synthesis gas

## 1 Introduction

Carbon dioxide is considered to be one of the major greenhouse gases that is directly influencing the global climate changes. The United Nations Intergovernmental Panel on Climate Change (IPCC) has studied these problems and a general conclusion has been achieved between researchers, industry leaders, and politicians that dramatic reduction in greenhouse gas emissions must be achieved in order to stop climatic changes [1, 4]. So using coal more efficiently and turning it into a clean energy source is an important issue for the whole world. An integrated

---

C.-t. Chou (✉) · Y.-H. Shih · Y.-J. Huang  
Department of Chemical and Materials Engineering, National Central University,  
Jhong-Li, Taiwan  
e-mail: t310030@ncu.edu.tw

H.-s. Yang  
Center for General Education, Hwa-Hsia Institute of Technology, New Taipei City, Taiwan

gasification combined cycle (IGCC) is a potential electric power technology that turns coal into synthesis gas, which can be burned to generate power.

The CO composition in syngas reacts with steam to generate CO<sub>2</sub> and H<sub>2</sub> via the water-gas-shift reaction,  $\text{CO} + \text{H}_2\text{O} \rightarrow \text{CO}_2 + \text{H}_2$ . In this study, pressure swing adsorption (PSA) is utilized to capture CO<sub>2</sub> from the effluent stream of water-gas-shift reaction at mid-high temperature, and the purified H<sub>2</sub> can be sent to gas turbine for generating electrical power or can be used for other energy source. This technology consists of gas adsorption at high pressure and desorption at low pressure to produce high-purity products. Conventionally, PSA is used to separate CO<sub>2</sub> and H<sub>2</sub> at ambient temperature. For traditional physical adsorbent, such as zeolite and activated carbon, the adsorbed amount of CO<sub>2</sub> is too low to separate CO<sub>2</sub>/H<sub>2</sub> at mid-high temperature. Because the outlet stream from water-gas-shift reactor is already at nearly 400 °C, in order to avoid separating CO<sub>2</sub> and H<sub>2</sub> at ambient temperature, and later raise the temperature of H<sub>2</sub> for power generation, which will cause energy waste, in this study PSA processes with K<sub>2</sub>CO<sub>3</sub>-promoted hydrotalcite adsorbent are studied to capture CO<sub>2</sub> from the outlet stream of water-gas-shift reactor at 400 °C. According to literature [6], K<sub>2</sub>CO<sub>3</sub>-promoted hydrotalcite is a chemisorbent that adsorbs CO<sub>2</sub> at mid-high temperature and it does not adsorb other gases, such as CO, H<sub>2</sub> and H<sub>2</sub>O. As required by the U.S. Department of Energy, it is important to be able to concentrate the captured CO<sub>2</sub> into >90 % concentration that is suitable for underground storage.

The feed gas entering the PSA process consists of CO, CO<sub>2</sub>, H<sub>2</sub> and H<sub>2</sub>O according to National Energy Technology Laboratory report [8].

Most PSA papers assume that steam is removed before entering CO<sub>2</sub>-capture PSA process. In this study we intend to consider the steam composition in feed gas into PSA process for real-case study.

## 2 Methodology

### 2.1 Mathematical Modelling

In the non-isothermal dynamic model, the following assumptions are made:

1. The linear driving force model is used because mass transfer resistance between the gas phase and solid phase exists.
2. Only CO<sub>2</sub> is adsorbed in K<sub>2</sub>CO<sub>3</sub>-promoted hydrotalcite.
3. The ideal gas law is applicable.
4. Adiabatic system.
5. Only axial concentration and temperature gradient are considered.
6. The pressure drop along the bed can be neglected due to large particle size.

These assumptions are used in the following equations:

Overall mass balance:

$$-\frac{\partial q}{\partial z} = \frac{\varepsilon A}{R} \frac{\partial(P/T)}{\partial t} + (1 - \varepsilon)A \sum_{i=1}^n \frac{\partial n_i}{\partial t} \quad (1)$$

Mass balance for component  $i$ :

$$\frac{\partial}{\partial z} \left( \frac{\varepsilon A D_{ax,i} P}{RT} \frac{\partial y_i}{\partial z} \right) - \frac{\partial(y_i q)}{\partial z} = \frac{\varepsilon A}{R} \frac{\partial}{\partial t} \left( \frac{y_i P}{T} \right) + (1 - \varepsilon)A \frac{\partial n_i}{\partial t} \quad (2)$$

Energy balance:

$$\begin{aligned} (A\bar{k}) \frac{\partial^2 T}{\partial z^2} - \frac{\partial}{\partial z} (\bar{C}_p q T) - \pi D h (T - T_\infty) \\ = \frac{\varepsilon A}{R} \frac{\partial}{\partial t} (\bar{C}_p P) + (1 - \varepsilon)A \sum_{i=1}^n \frac{\partial}{\partial t} [n_i (\bar{C}_{pi} T - H_i)] + (1 - \varepsilon) \rho_s \bar{C}_{ps} A \frac{\partial T}{\partial t} \end{aligned} \quad (3)$$

Linear driving force model:

$$\frac{\partial N_i}{\partial t} = K_{LDF} (N_i^* - N_i) \quad (4)$$

Axial dispersion coefficient [9]:

$$D_{ax,i} = \gamma D_{m,i} + \frac{0.5 \bar{u} d_p}{1 + \beta (D_{m,i} / \bar{u} d_p)} \quad (5)$$

$$\gamma \cong 0.75, \beta \cong 9.5, \bar{u} = \frac{u_0}{e}$$

for  $0.008 < \text{Re} < 400$  and  $0.25 < \text{Sc} < 2.2$

$D_{m,i}$  can be obtained by Chapman-Enskog equation [2]:

$$\frac{D_{m,i}}{D_{m,i}^0} = \left( \frac{P_0}{P} \right) \left( \frac{T}{T_0} \right)^{\frac{3}{2}} \quad (6)$$

where  $D_{m,i}^0$  is at  $P_0$  and  $T_0$

The adiabatic system means that there is no heat transfer between bed and surrounding so that  $h = 0$  in Eq. (3).

Boundary conditions can be assumed as follows:

At the inlet end:

$$c(t, 0) = c_{in}(t), T(t, 0) = T_{in}(t) \quad (7)$$

At the outlet end:

$$\frac{\partial c(t, L)}{\partial z} = 0, \frac{\partial T(t, L)}{\partial Z} = 0 \quad (8)$$

The flow rates at the two ends of the bed are estimated by using the valve equation recommended by Fluid Controls Institute Inc.:

$$q' = 16.05 C_v \sqrt{\frac{(P_1^2 - P_2^2)}{SG \times T}} \quad \text{for } P_2 > 0.53 P_1 \quad (9)$$

$$q' = 13.61 C_v P_1 \sqrt{\frac{1}{SG \times T}} \quad \text{for } P_2 \leq 0.53 P_1 \quad (10)$$

Twenty-one basic grid points are marked in the bed to set up the initial concentration, initial temperature, and initial pressure. The partial differential equations are converted to ordinary differential equations by the method of lines. The spatial derivatives of the concentration and the gas temperature are evaluated by upwind differences at every grid point. The cubic spline approximation is used to estimate the flow rates in the adsorptive bed. The concentration, temperature, and adsorption quantity in the bed are integrated with respect to time by LSODE of ODEPACK software with a time step of 0.1 s. The simulation is stopped by using Eq. 11 when the system reaches a cyclic steady state.

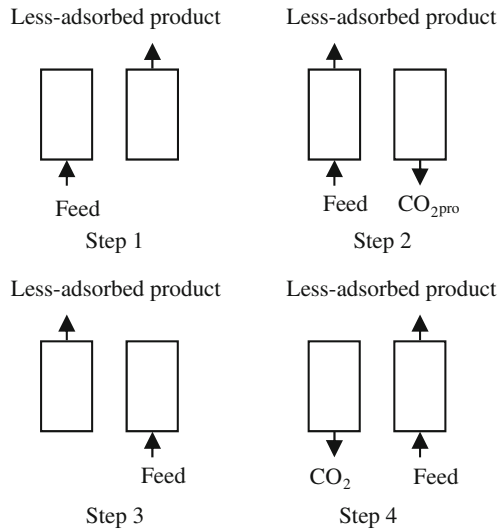
$$\sum \left( 1 - \frac{Y(\text{last cycle})}{Y(\text{this cycle})} \right)^2 < 1 \times 10^{-4} \quad (11)$$

where Y is the value of the mole fraction of each component and the amount of every flow stream.

## 2.2 PSA Process

The PSA process studied is a single-bed four-step process at mid-high temperature using  $K_2CO_3$ -promoted hydrotalcite. The feed gas is from the effluent stream in the water-gas-shift reactor which is cited in the report of National Energy Technology Laboratory [8]. The feed gas entering the PSA process consists of 27 %  $H_2O$ , 5 %  $CO$ , 28 %  $CO_2$  and 40 %  $H_2$ . The process is described as follows: feed pressurization (I), high pressure adsorption (II), cocurrent depressurization (III), vacuum desorption (IV). During step (I), the bed pressure increases from atmospheric pressure to high pressure, and less-adsorbed products are produced. Strongly adsorptive carbon dioxide is produced during step (IV) when the bed pressure decreases from high pressure to low pressure (0.1 atm). The procedure of the sing-bed four-step process is shown in Fig. 1. The step time and the physical parameters of adsorption bed are shown in Tables 1 and 2.

**Fig. 1** Procedure of single-bed four-step PSA process



**Table 1** Step time for single-bed four-step process

(I)	10 s
(II)	39 s
(III)	10 s
(IV)	39 s

**Table 2** Physical parameters of bed

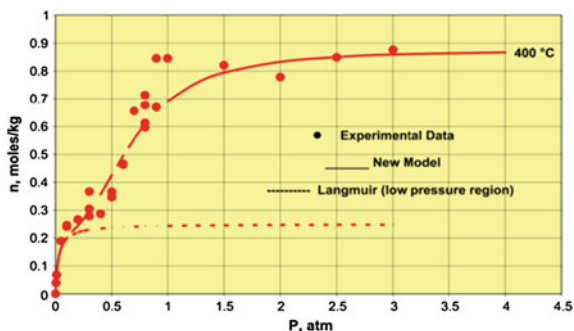
Bed length(cm)	100
Bed diameter(cm)	100
Adsorbent density(g/cm <sup>3</sup> )	1.563 <sup>a</sup>
Adsorbent heat capacity(J/g.K)	0.85 <sup>a</sup>
Bed void	0.37
Operating temperature(K)	673.14
Feed pressure(atm)	25
Vacuum pressure(atm)	0.1

<sup>a</sup> [3]

Isotherm of K<sub>2</sub>CO<sub>3</sub>-promoted hydrotalcite was measured at 400 °C in the pressure range of 0–3 atm by Lee et al. [6] and shown in Fig. 2. The figure also shows the best fit of the CO<sub>2</sub> chemisorption isotherms using the following Eq. (12). The parameters of model for sorption of CO<sub>2</sub> are given in Table 3.

$$n_i^*(P, T) = \frac{mK_C P [1 + (a + 1)K_R P^a]}{[1 + K_C P + K_C K_R P^{(a+1)}]} \quad (12)$$

**Fig. 2** CO<sub>2</sub> chemisorption isotherm on K<sub>2</sub>CO<sub>3</sub>-promoted hydrotalcite at 400 °C [6]



**Table 3** Parameters of model for sorption of CO<sub>2</sub> on K<sub>2</sub>CO<sub>3</sub>-promoted hydrotalcite

m(mole/kg)	0.25
a	2.5
q <sub>c</sub> (J/mole)	$2.098 \times 10^4$
$\Delta H_R$ (J/mole)	$4.210 \times 10^4$
k <sub>c</sub> (atm <sup>-1</sup> )	37.4
k <sub>R</sub> (atm <sup>-a</sup> )	2.5

## 3 Results and Discussion

### 3.1 Simulation Verification

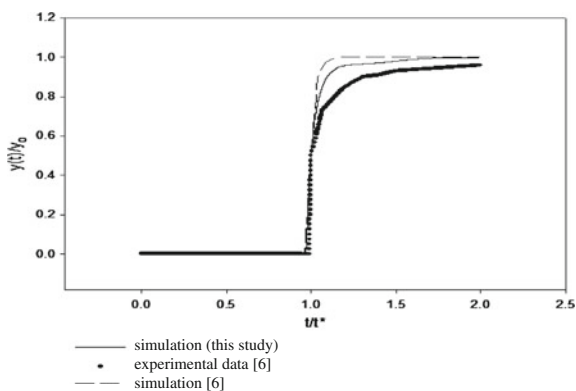
The breakthrough curve studied by [6] was used to verify the simulation program. A single-bed adsorption instrument was constructed to measure CO<sub>2</sub> chemisorption isotherm and CO<sub>2</sub> sorption breakthrough curves at temperature of 673.14 K on the K<sub>2</sub>CO<sub>3</sub>-promoted hydrotalcite. The single-bed was surrounded by heating elements, gas heating, cooling exchangers and flow measuring devices. A thick layer of insulation was surrounding around the column. Three thermocouples were used to measure the single-bed temperature. The exit end of column was equipped with CO<sub>2</sub> analyzers which was used to measure carbon dioxide concentration with time. The column was fully packed with the K<sub>2</sub>CO<sub>3</sub>-promoted hydrotalcite for measurement of CO<sub>2</sub> sorption dynamics. The sorbent was cleaned by flowing pure N<sub>2</sub>. At the same time the column was preheated to temperature at 673.14 K, pressure at 1 atm until no impurity flowed out in the effluent gas. Different compositions of CO<sub>2</sub> + N<sub>2</sub> mixtures were used as the feed gas. The operating conditions used are given in Table 4. The results are shown in Figs. 3, 4, 5. It shows that the simulation results are very close to experiment data. Therefore, the simulation program can be trusted.



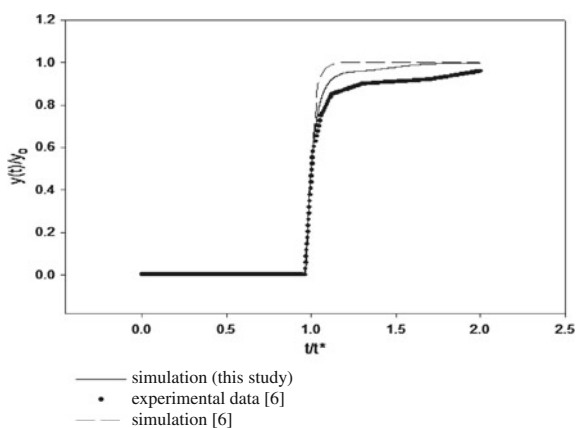
**Table 4** Operating parameters of breakthrough curve simulation

Operating pressure (atm)	1
Operating temperature (K)	673.14
Feed flow rate (L/min)	$5.0 \times 10^{-3}$
Bed length (cm)	101.6
Bed diameter (cm)	1.73
Bed volume (L)	0.238
Bulk density (g/cm <sup>3</sup> )	0.82
Adsorption Time Constant (min <sup>-1</sup> )	3.0
Feed composition	(40 % CO <sub>2</sub> , 60 % N <sub>2</sub> ) (50 % CO <sub>2</sub> , 50 % N <sub>2</sub> ) (60 % CO <sub>2</sub> , 40 % N <sub>2</sub> )

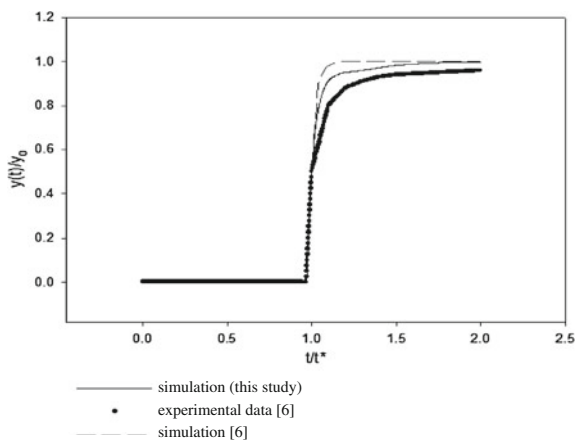
**Fig. 3** Simulation of breakthrough curve (inlet CO<sub>2</sub> mol fraction = 0.4)



**Fig. 4** Simulation of breakthrough curve (inlet CO<sub>2</sub> mol fraction = 0.5)



**Fig. 5** Simulation of breakthrough curve (inlet  $\text{CO}_2$  mol fraction = 0.6)

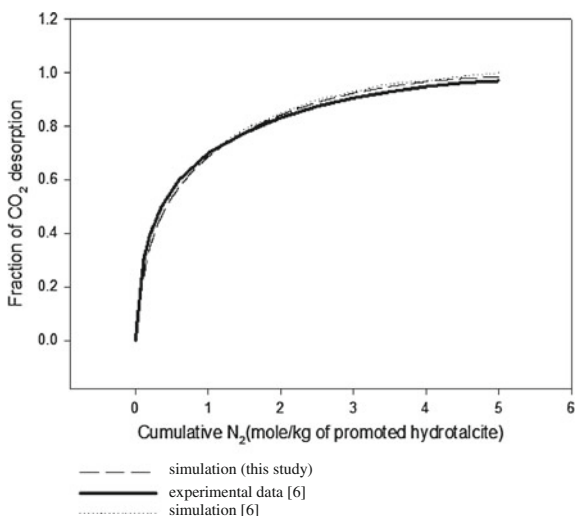


The desorption curve studied by [7] was also compared to our simulation. Figure 6 shows the column dynamic data (fraction of  $\text{CO}_2$  desorbed as a function of inlet  $\text{N}_2$  purge gas quantity) for desorption of 40 %  $\text{CO}_2 + \text{N}_2$  by  $\text{N}_2$  purge. The operating conditions used are given in Table 5. The agreement between our simulation and the experimental data is pretty good.

### 3.2 Single-Bed Four-Step PSA Process Simulation

In this study, the optimal operating conditions are discussed by varying the operating variables, such as feed pressure, bed length, vacuum pressure, feed flow rate, high pressure adsorption time and vacuum desorption time.

**Fig. 6** Simulation of desorption curve



**Table 5** Operating parameters of desorption curve simulation

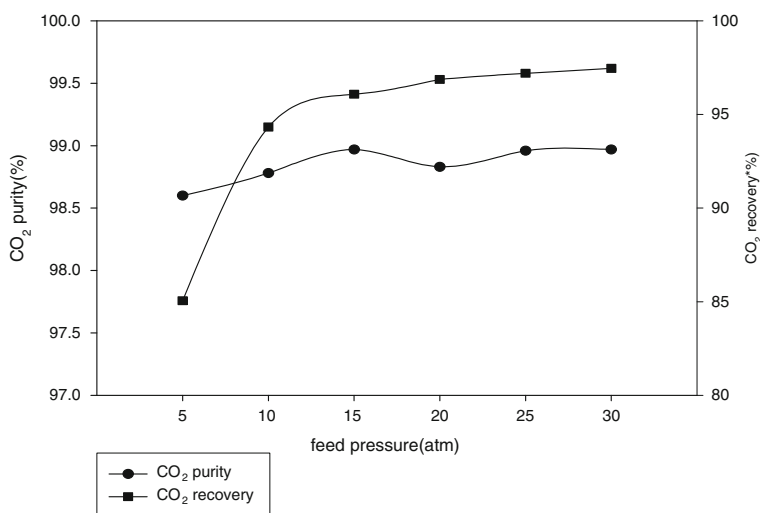
Operating pressure(atm)	1
Operating temperature(K)	793.14
Feed flow rate(L/min)	$6.667 \times 10^{-3}$
Bed length(cm)	100
Bed diameter(cm)	1.73
Bulk density(g/cm <sup>3</sup> )	0.82
Initial gas phase concentration	40 %CO <sub>2</sub> /N <sub>2</sub>

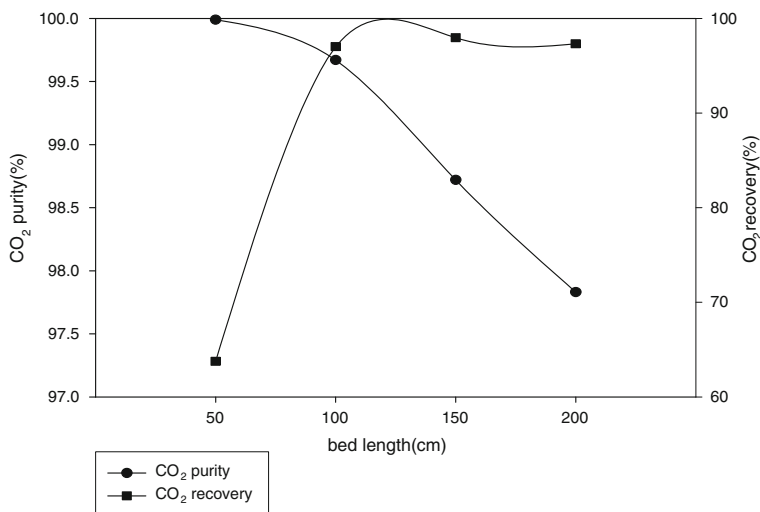
Definition of the recovery is:

$$\text{Recovery} = \frac{\text{product flow per cycle} \times \text{product concentration feed}}{\text{flow per cycle} \times \text{feed concentration}}$$

**Feed Pressure.** All the operating variables such as vacuum pressure, bed length, feed rate and step time are fixed, except feed pressure. Because the amount of gas adsorbed on K<sub>2</sub>CO<sub>3</sub>-promoted hydrotalcite increases as feed pressure increases, the flow of the strong adsorptive component to the bottom of the bed during desorption increases. Figure 7 shows that as feed pressure increases, the CO<sub>2</sub> purity and recovery in bottom product increases because CO<sub>2</sub> adsorption quantity becomes larger.

**Bed Length.** All the operating variables are fixed except bed length. With increasing bed length, the amount of adsorbent and the amount of adsorbed gas increase. Figure 8 shows that as bed length decreases, the CO<sub>2</sub> purity increases due

**Fig. 7** Effect of feed pressure on CO<sub>2</sub> in bottom product



**Fig. 8** Effect of bed length on CO<sub>2</sub> in bottom product

to that the amount of CO<sub>2</sub> flow to the top product increases. At the same feed flow rate CO<sub>2</sub> recovery decreases due to that the amount of CO<sub>2</sub> flow to the top product increases.

**Vacuum Pressure.** All the operating variables are fixed except vacuum pressure. As the vacuum pressure increases, the amount of desorption gas at desorption step decreases. Figure 9 shows that as the vacuum pressure increases, the CO<sub>2</sub> recovery decreases due to that the amount of adsorbed gas flow to the bottom product at desorption step decreases.

**Feed Flow Rate.** All the operating variables are fixed except feed flow rate. As the feed flow rate increases, the amount of CO<sub>2</sub> increases at high pressure adsorption step. Figure 10 shows that as the feed flow rate increases, the CO<sub>2</sub> recovery decreases due to that the amount of adsorption/desorption are approximately fixed. The CO<sub>2</sub> purity increases as the amount of adsorbed gas increases.

**High Pressure Adsorption Time and Vacuum Desorption Time.** All the operating variables are fixed except high pressure adsorption time/vacuum desorption time. The amount of CO<sub>2</sub> increases in the column as the pressure adsorption time increases. Therefore, Fig. 11 shows that CO<sub>2</sub> recovery decreases with the decreasing 2nd/4th step time.

The best operating conditions for the single-bed four-step PSA process at mid-high temperature is shown in Fig. 12. The results of best operating condition for the single-bed four-step process at mid-high temperature are 96.42 % purity and 96.57 % recovery of CO<sub>2</sub> as bottom product as shown in Fig. 12.

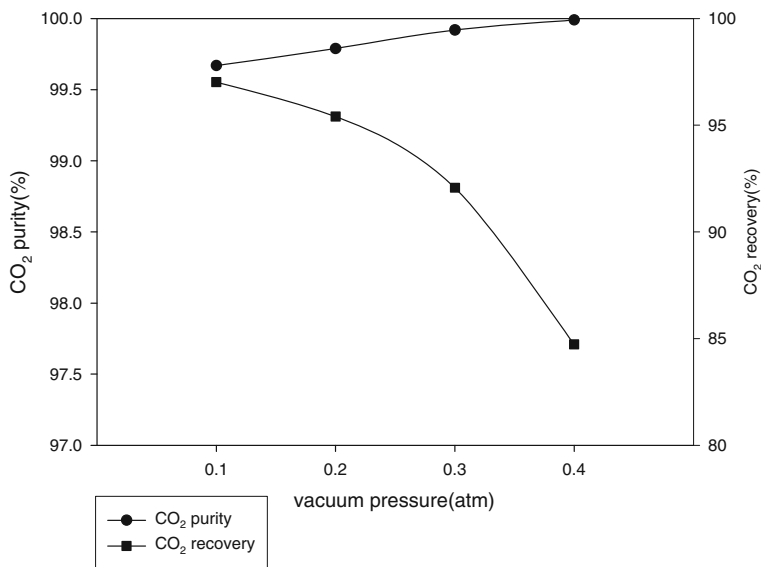


Fig. 9 Effect of bed length on CO<sub>2</sub> in bottom product

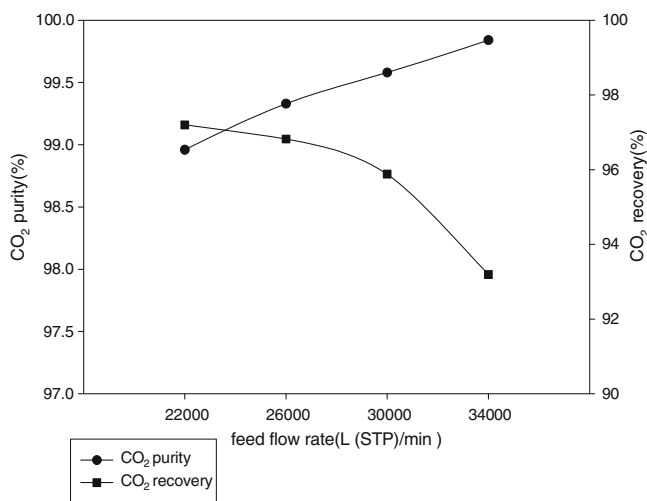


Fig. 10 Effect of feed flow rate on CO<sub>2</sub> in bottom product

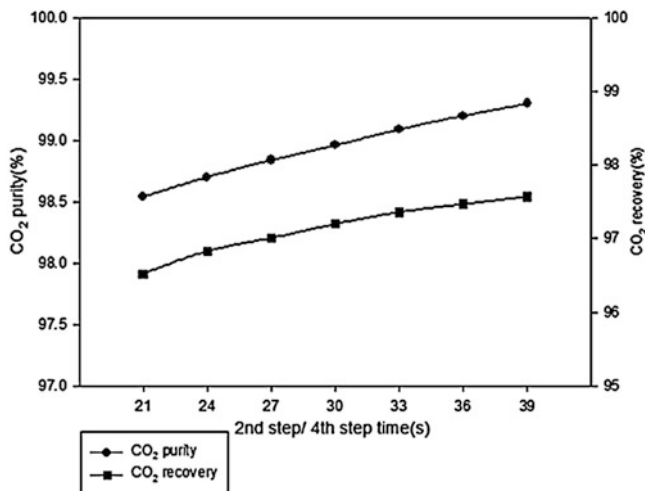
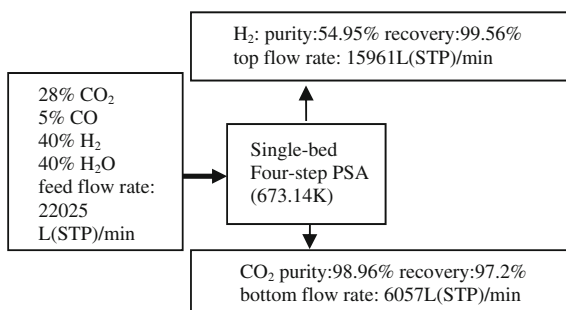


Fig. 11 Effect of 2nd/4th step time on CO<sub>2</sub> in bottom product

Fig. 12 Results of the single-bed four-step PSA process at mid-high temperature



### 4 Conclusions

A single-bed four-step pressure swing adsorption process is explored in this simulation study. The adsorbent K<sub>2</sub>CO<sub>3</sub>-promoted hydrotalcite was used [6]. The accuracy of the simulation program is verified by comparing our simulation results with the experimental data of breakthrough curve and desorption curve from Lee et al. [6, 7] and the single-bed five-step PSA results of Kim et al. [5]. The optimal operating condition is obtained by varying the operating variables, such as feed pressure, bed length, feed flow rate, etc. Furthermore, the optimal operating condition for inlet (27 % H<sub>2</sub>O, 5 % CO, 28 % CO<sub>2</sub> and 40 % H<sub>2</sub>) at mid-high temperature 673 K and bed diameter 100 cm is bed length 200 cm, feed pressure 25 atm, vacuum pressure 0.1 atm and step times at 10, 30, 10 and 30 s. The best results and the optimal operating condition for the single-bed four-step PSA process

**Table 6** Optimal operating condition for the single-bed four-step PSA process at mid-high temperature

Feed pressure (atm)	25
Bed length (cm)	100
Vacuum pressure (atm)	0.1
Feed flow rate (LSTP/min)	22000
High pressure adsorption time and vacuum desorption time (s)	30

at mid-high temperature are shown in Fig. 12 and Table 6. In the future, our research will proceed with dealing the top stream from CO<sub>2</sub>-PSA by a second stage H<sub>2</sub>-PSA to improve H<sub>2</sub> purity.

**Acknowledgments** The authors wish to thank the financial support from National Science Council, Taiwan under project number NSC 102-3113-P-008 -007.

## References

1. Abu-Zahra MRM, Feron PHM, Jansens PJ, Goetheer ELV (2009) New process concepts for CO<sub>2</sub> post-combustion capture process integrated with co-production of hydrogen. *Int J Hydrogen Energy* 34:3992–4004
2. Bird RB, Stewart WE, Lightfoot EN (1960) *Transport phenomena*. Wiley, New York
3. Ding Y, Alpay E (2000) Equilibria and kinetics of CO<sub>2</sub> adsorption on hydrotalcite adsorbent. *Chem Eng Sci* 55:3461–3474
4. IPCC (Intergovernmental Panel on Climate Change) (2005) *Carbon dioxide capture and storage*. Cambridge University Press, Cambridge
5. Kim WG, Yang J, Han S, Cho C, Lee CH, Lee H (1995) Experimental and theoretical study on H<sub>2</sub>/CO<sub>2</sub> separation by a five-step one-column PSA processes. *Korean J Chem Eng* 12:503–511
6. Lee KB, Caram HS, Verdooren A, Sircar S (2007) Chemisorption of carbon dioxide on potassium carbonate promoted hydrotalcite. *J Colloid Int Sci* 308:30–39
7. Lee KB, Beaver MG, Caram HS, Sircar S (2007) Reversible chemisorption of carbon dioxide: simultaneous production of fuel-cell grade H<sub>2</sub> and compressed CO<sub>2</sub> from synthesis gas. *Adsorption* 13:385–392
8. NETL (National Energy Technology Laboratory) (2009) Evaluation of alternate water gas shift configurations for IGCC systems, DOE/NETL-401/080509
9. Wen CY, Fan LT (1976) Models for flow systems and chemical reactors. *J AIChE* 22:412

# Fuzzy Climate Scenarios for Temperature Indicate that Things Could Be Worse Than Previously Thought

Carlos Gay García and Oscar Sánchez Meneses

**Abstract** Linear evolving emission scenarios are used instead of those of IPCC. They preserve, indeed they cover, the ranges of the corresponding IPCC values for the concentrations, forcings and global temperatures. Then, through fuzzy rules among concentrations, climate sensitivity and global temperature change, a fuzzy model has been conformed and used to explore uncertainties due to: not knowing what the emissions are going to be in the future, the one related to the climate sensitivity of the models (this has to do with different parameterizations of processes used in the models) and the uncertainties in the temperature maps produced by the models. Furthermore we show maps corresponding to 1, 2, etc., degrees centigrade of global and regional temperature increase and discuss the timing of exceeding them. Instead of talk about the uncertainty in temperature at a certain date we talk about the uncertainty in the date certain temperature is reached.

**Keywords** Temperature climate change scenarios · Uncertainty · Greenhouse gas emissions · Climate sensitivity

## 1 Introduction

In a recent publication, Gay et al. [2] simplified the emission scenarios developed by the Intergovernmental Panel on Climate Change (IPCC) using linear functions of time that after being fed to the Magicc model [6], produced the same wide range of concentrations of greenhouse gases (GHG) and aerosols, and the corresponding range of temperatures in 2100. These results show very clearly that higher

---

C. Gay García (✉) · O. Sánchez Meneses  
Centro de Ciencias de La Atmósfera, Universidad Nacional Autónoma de México,  
Ciudad Universitaria, México, DF, Mexico  
e-mail: cgay@unam.mx

O. Sánchez Meneses  
e-mail: casimiro@atmosfera.unam.mx



temperature increases correspond to higher emission of GHG and higher atmospheric concentrations. This fact can be transformed into linguistic rules that in turn are used to build a fuzzy model, which uses concentration values of GHG as input variables and gives, as output, the temperature increase projected for year 2100. Based on the same principles a second fuzzy model is presented that includes a second source of uncertainty: climate sensitivity.

It is our intention to extend these results and produce maps of temperature. It has been customary to ask what the temperature is going to be in 2030 or in 2050 and proceed to estimate the impacts that the changed temperature would have on social or economic sectors and activities that either may improve or most probably would be affected in a negative way. But in 2030 or in 2050 different models say different things so, what do we do? Use ensembles? Use the averages? Consider the standard deviation? Is the physics consistent? Here we propose to show temperature maps corresponding to global increases of 1, 2, 3, etc., degrees centigrade, give an idea of the uncertainty in timing, in contrast to the uncertainty in temperature for a certain date. This means that depending on the emissions, concentrations etc., the larger these variables, the sooner 1, 2, etc., degrees will be reached and considering other sources of uncertainty like the sensitivity, the pace of change may increase considerably. When we display the information in two dimensions produced by different models then the uncertainty due to different modeling strategies has to be considered.

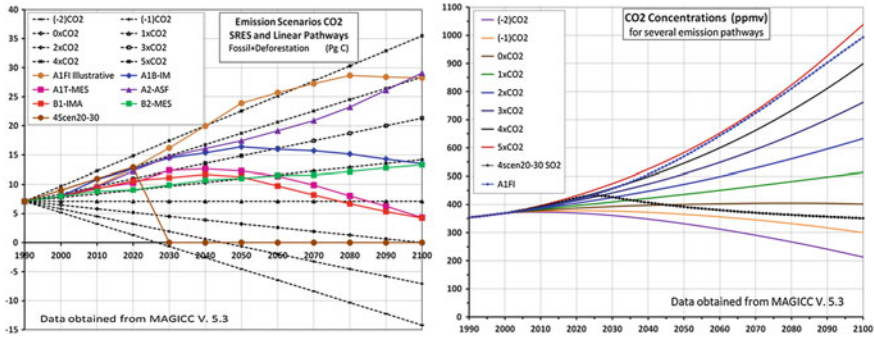
We think that it is easier to consider a degree by degree strategy than one based on dates. The question of what to do if the temperature increases one degree or what should we be doing right now because the temperature is reaching one degree by 2021 (in the worst of cases) and if we do nothing we will be two degrees warmer by 2039 with grave consequences for all.

## 2 Method

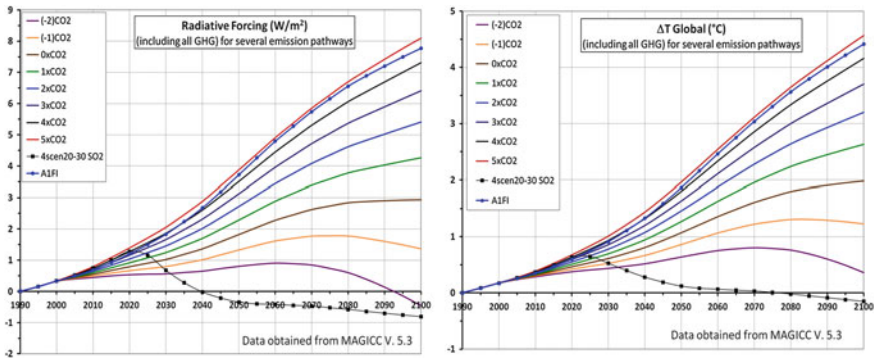
By using linear and non-intersecting emission trajectories, concentrations of GHG and global mean temperatures increases can be directly related as illustrated in Figs. 1 and 2 of Gay et al. [2].

With the linear emission pathways shown in Fig. 1 left panel, used as input for the Magicc model, Gay et al. [2] calculated the resulting concentrations (Fig. 1, right panel); radiative forcings (Fig. 2, left panel) and global mean temperature increments (Fig. 2, right panel) that we repeat here for clarity.

We would like to remark a statement made before (that can be directly observed in Fig. 2, right panel): if we want to keep temperatures at two degrees or less by the year 2100, we should have concentrations in 2100 consistent with the  $-2\text{CO}_2$ ,  $-1\text{CO}_2$  and  $0\text{CO}_2$  trajectories. The latter is a trajectory of decreasing emissions from the emissions in 1990 (5.9 Gt C) to 0 Gt C that gives us a temperature change of two degrees by year 2100. For a trajectory of constant emissions equal to the emissions in 1990 we obtain 2.63 °C.



**Fig. 1** Emissions scenarios CO<sub>2</sub>, illustrative SRES [5] and linear pathways (*left panel*) and corresponding concentrations (*right panel*). (-2) CO<sub>2</sub> means -2 times the emission (fossil + deforestation) of CO<sub>2</sub> of 1990 by 2100 and so for -1, 0, 1, to 5 CO<sub>2</sub>. All the linear pathways contain the emission of non CO<sub>2</sub> GHG as those of the A1F1. 4scen20-30 scenario follows the pathway of 4×CO<sub>2</sub> but at 2030 all gases drop to 0 emissions or minimum value in CH<sub>4</sub>, N<sub>2</sub>O and SO<sub>2</sub> cases [2]

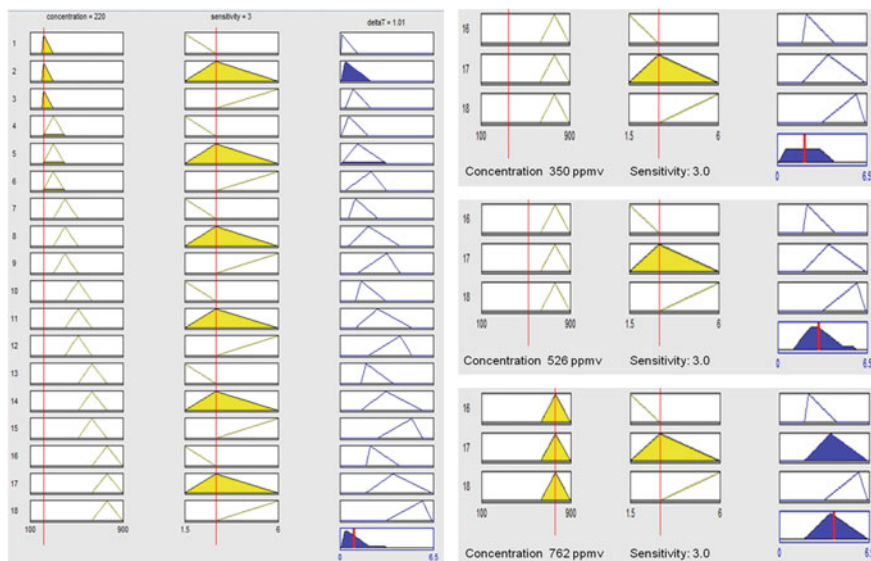


**Fig. 2** *Left panel*: Radiative forcings (all GHG included) for linear emission pathways and A1F1 SRES illustrative, the 4scen20-30 SO<sub>2</sub> only include SO<sub>2</sub>. *Right panel*: Global temperature increments for linear emission pathways, 4scen20-30 SO<sub>2</sub> and A1F1 as calculated using Magicc V. 5.3 with a fixed sensitivity of 3.0 °C/W/m<sup>2</sup> (Gay et al. [2])

From the linear representation, it is easily deduced (as mentioned earlier) that very high emissions correspond to very large concentrations, large radiative forcings and large increases of temperature.

These simple observations are basic for the formulation of the fuzzy model, based on linguistic rules of the **IF-THEN** form, capable of estimating increases of temperature. The fuzzy model was built using the results of the Magicc model [6] as crisp mathematical model, and Zadeh’s extension principle [7].

For illustrative purposes (the full rules are reported in Gay et al. [3]) we repeat here the first two rules of the 18 that were developed previously (see Fig. 3):



**Fig. 3** On *left panel*, the 18 rules of the fuzzy model for the estimation of global mean temperature increase  $\Delta T$ , for a concentration of  $\text{CO}_2$  of 220 ppmv and sensitivity of  $3.0 \text{ }^\circ\text{C/W/m}^2$ . The uncertainty interval is (0.08, 2.17) or (0.08, 3.27)  $^\circ\text{C}$  considering the elongated part. On *right panel*, the estimation of global mean temperature increase  $\Delta T$  and its uncertainty intervals (from the feet of the triangular fuzzy sets) for concentrations of  $\text{CO}_2$  of: 350 ppmv (*upper*) with  $2.01 \text{ }^\circ\text{C}$  (0.08–3.27); 526 ppmv (*middle*) with  $3 \text{ }^\circ\text{C}$  (1.07–5.02) or (1.07–5.75) considering the elongated part and 762 ppmv (*lower*) with  $3.98 \text{ }^\circ\text{C}$  (1.82–6.41). Data calculated with MATLAB only the last 3 rules are shown for simplicity

1. If (concentration is (-2)CO2) and (sensitivity is low) then (deltaT is T1) (1)
2. If (concentration is (-1)CO2) and (sensitivity is low) then (deltaT is T2) (1)

where the triangular fuzzy sets corresponding to T1 and T2 are (0.07, 0.07, 1.23) and (0.07, 0.61, 1.98) respectively.

The 18 rules were obtained from the combination of 6 concentrations, projected to 2100 and consistent with 6 linear emission trajectories, and 3 fuzzy values for the sensitivity of climate which are 1.5, 3.0 and  $6.0 \text{ }^\circ\text{C/W/m}^2$ , all the values were taken from the data previously generated by successive runs of Magicc software.

Once we have the global temperatures and an idea of the associated uncertainty due to different emission paths and sensitivities, we would like to convert this information to a two dimensional display of temperatures. The way to accomplish this is using the same idea for scaling employed in the Magicc/Scengen system [6]. This consists of scaling the value that results from running for a Global Circulation Model (GCM) option (one of 20 possible), for example with double  $\text{CO}_2$  at a certain grid point in the following way:

$$T_{\text{new}} = T_{\text{grid}}/T_{\text{map}} \times T_{\text{magicc}} \tag{1}$$

where  $T_{\text{new}}$  is the scaled temperature,  $T_{\text{grid}}$  is the value of the temperature given by the GCM at a certain position,  $T_{\text{map}}$  is the average temperature (global) of the map and  $T_{\text{magicc}}$  is the temperature given by the simple model.

However, emissions and sensitivity introduce uncertainties in the temperature that in turn must be reflected in the scaled temperature.

If we denote the uncertainty by a  $\Delta$  then we propose:

$$\Delta T_{\text{new}} = T_{\text{grid}}/T_{\text{map}} \times \Delta T_{\text{magicc}} \tag{2}$$

where  $\Delta T_{\text{magicc}}$ , is in fact a fuzzy number and consequently  $\Delta T_{\text{new}}$  also is.

We have to mention that another source of uncertainty is which GCM we use. We will try to illustrate this point too.

From the application of the Magicc/Scengen to the emission trajectories developed in the previous paper [2] we can extract the years in which the 1–4 °C thresholds are reached.

According to the IPCCs Fourth Assessment Report [4] the best estimate for the sensitivity is 3.0 however this parameter varies from 1.5 to 6, as mentioned before, so there is a source of uncertainty associated with this parameter. This is shown by the different values in the Tables 1 and 2. Dates for emission scenarios B1-IMA and A1FI-MI [5] are shown for reference.

For a threshold of 5 °C, there are only 5 emission trajectories achieving dates within 21st century, they are: 2CO<sub>2</sub> (2100), 3CO<sub>2</sub> (2088), 4CO<sub>2</sub> (2080), 5CO<sub>2</sub>

**Table 1** Dates to achieve the 1 and 2 °C thresholds following linear emission trajectories from –2CO<sub>2</sub> to 5CO<sub>2</sub> for climate sensitivities of 1.5, 3.0 and 6.0 °C/W/m<sup>2</sup>

Emission trajectory	1 °C threshold			2 °C threshold		
	Sensitivity					
	1.5	3.0	6.0	1.5	3.0	6.0
–2CO <sub>2</sub>			2049			
–1CO <sub>2</sub>		2057	2039			2073
0CO <sub>2</sub>	2079	2048	2033		2100 (1.98 °C)	2059
1CO <sub>2</sub>	2063	2042	2029		2072	2052
2CO <sub>2</sub>	2056	2038	2027		2064	2048
3CO <sub>2</sub>	2051	2035	2024	2093	2058	2045
4CO <sub>2</sub>	2047	2032	2023	2081	2054	2042
5CO <sub>2</sub>	2044	2030	2021	2053	2051	2039
B1-IMA	2090	2043	2027			2057
A1FI-MI	2046	2033	2024	2076	2053	2042

**Table 2** Dates to achieve the 3 and 4 °C thresholds following linear emission trajectories from  $-2\text{CO}_2$  to  $5\text{CO}_2$  for climate sensitivities of 1.5, 3.0 and 6.0 °C/W/m<sup>2</sup>

Emission trajectory	3 °C threshold			4 °C threshold		
	Sensitivity					
	1.5	3.0	6.0	1.5	3.0	6.0
$-2\text{CO}_2$						
$-1\text{CO}_2$						
$0\text{CO}_2$			2087			
$1\text{CO}_2$			2071			2095
$2\text{CO}_2$		2093	2064			2080
$3\text{CO}_2$		2081	2059			2073
$4\text{CO}_2$		2074	2055		2097	2068
$5\text{CO}_2$		2069	2052		2088	2064
B1-IMA			2095			
A1FI-MI		2070	2054		2090	2065

(2075) and the “pessimistic” SRES scenario A1FI (2077) all of them corresponding to the 6.0 climate sensitivity level.

Taking into account the opinion of the IPCC that the best estimate for the sensitivity is 3, it can be said that we would be exceeding the one degree threshold by 2030 (sensitivity of 3 and emission trajectory of  $5\text{CO}_2$ ). However due to the values that this parameter may assume (1.5–6) this threshold may be delayed to 2044 if the sensitivity is 1.5 or may be advanced to 2021 if the sensitivity is 6. These values for the threshold correspond to our worst emissions scenario  $5\text{CO}_2$ . If we continue mounted in the same scenario we could be reaching 6 °C by 2087 and almost 7 °C by 2100.

Again for the 3 °C threshold we could be surpassing it as early as 2052 and the “best estimate” would be 2069; if the sensitivity were 1.5 the 3 °C temperature would not be reached.

From these tables we can also learn that if the sensitivity is 6 there is no way of staying at two degrees unless the concentrations of  $\text{CO}_2$  had followed the  $-2\text{CO}_2$  trajectory: negative emissions that means very strong subtraction of  $\text{CO}_2$  from the atmosphere.

If we were lucky and the climate sensitivity had a value of three the concentration would have to be equivalent to the  $0\text{CO}_2$  path in 2100 this is about 300 ppmv.

There are obvious messages from the tables: the smaller the emissions the later the thresholds are exceeded, if we want small increases of temperature then we need to impose small emissions or more precisely small concentrations of  $\text{CO}_2$ .

Two sources of uncertainty are illustrated in the tables, the first coming from the emissions: large emissions large temperature changes and the second due to our imprecise knowledge of the climate sensitivity of the models. One uncertainty is for

the politicians because emissions depend on policy and the second for the scientists who may narrow the gap in the estimations of climate sensitivity.

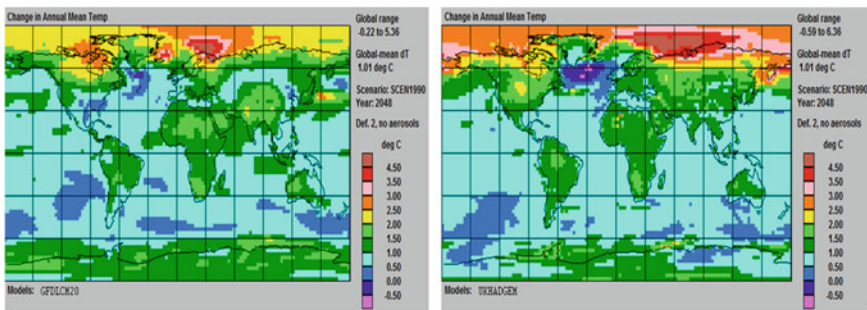
### 3 Results

The results of the fuzzy model that combines six levels of concentrations of CO<sub>2</sub>, from -2CO<sub>2</sub> to 3CO<sub>2</sub> in year 2100 (where 1CO<sub>2</sub> identifies the concentration associated to the emissions in 1990), and 3 levels of sensitivity: 1.5, 3 and 6 are presented here. The model, that incorporates the uncertainties mentioned above, consisting of 18 fuzzy rules [3], is run to obtain global temperatures increases in year 2100 and their corresponding uncertainty intervals. This information is then used to produce two-dimensional maps depicting physically consistent geographical distributions of temperatures which in turn are consistent with global temperatures obtained from our fuzzy model. That the temperatures are physically consistent can be justified by using the results of a physically consistent model, in the same way the Magicc/Scengen does: using the results of runs of different GCMs.

The fuzzy model with the best estimate for the sensitivity is used to get the uncertainty intervals for 1–4 °C.

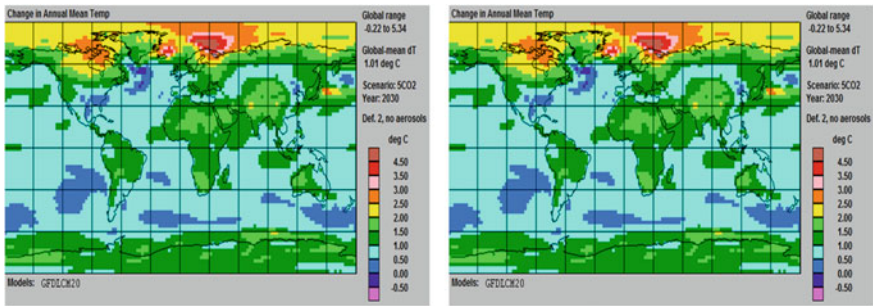
In the fuzzy model the value of the sensitivity is fixed at the best estimate of 3 and varying the concentration we try to get 1, 2, 3, etc. degrees. The temperature is a function of the concentration. In this way we obtain:

For an increase of one degree the concentration of CO<sub>2</sub> required is 220 ppmv and the uncertainty interval is from 0.08 to 2.17 degrees, based on the fuzzy sets feet presented in Gay et al. [3] and reproduced here as a graph (see Fig. 3, left panel). Therefore for a one degree global increase the uncertainty extends to more than two degrees, consequently for a 1 °C global increase, maps for one and two degrees (see ahead, Figs. 4, 5, 6 and 7) are to be considered.

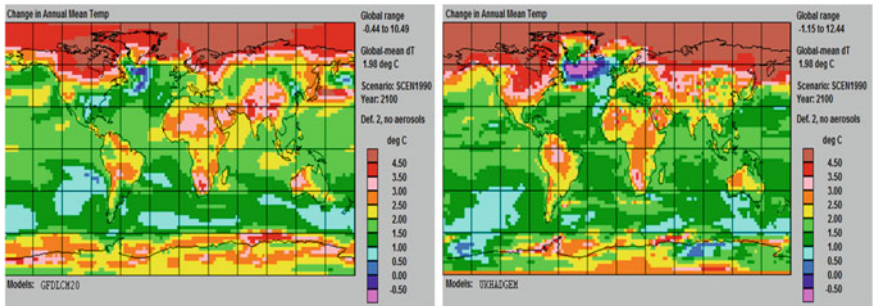


**Fig. 4** Spatial distribution of  $\Delta T = 1.01 \text{ }^\circ\text{C}$  according to GFDL 2.0 (*left panel*) and HADGEM1 (*right panel*) for OCO<sub>2</sub> emission trajectory (SCEN1990 in map). Maps were obtained using Magicc/Scengen V. 5.3

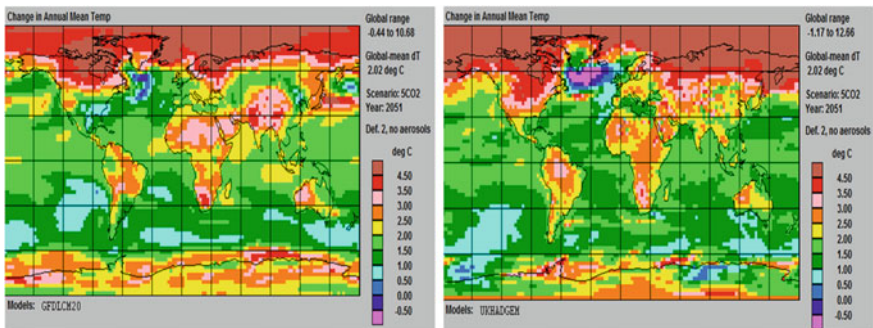




**Fig. 5** Spatial distribution of  $\Delta T = 1.01$  °C according to GFDL 2.0 (*left panel*) and HADGEM1 (*right panel*) for  $5CO_2$  emission trajectory. Maps were obtained using Magicc/Scengen V. 5.3



**Fig. 6** Spatial distribution of  $\Delta T = 1.98$  °C according to GFDL 2.0 (*left panel*) and HADGEM1 (*right panel*) for  $0CO_2$  emission trajectory (SCEN1990 in map). Maps were obtained using Magicc/Scengen V. 5.3



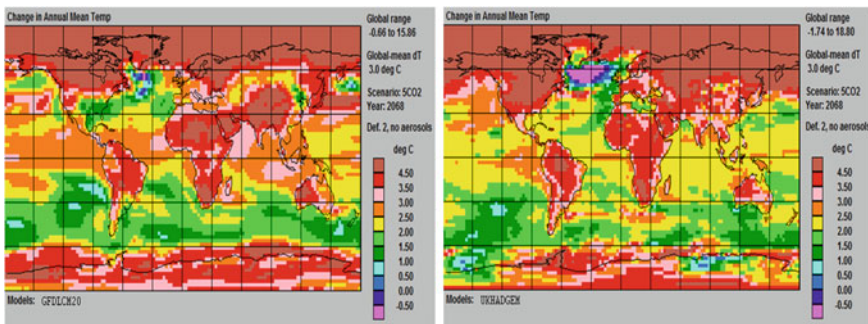
**Fig. 7** Spatial distribution of  $\Delta T = 2.02$  °C according to GFDL 2.0 (*left panel*) and HADGEM1 (*right panel*) for  $5CO_2$  emission trajectory. Maps were obtained using Magicc/Scengen V. 5.3

If  $\Delta T$  is 2 degrees the interval is from 0.08 to 3.27 °C; for 3 and 4 degrees the uncertainty intervals are from 1.07 to 5.02 °C and from 1.82 to 6.41 °C respectively (see Fig. 3, right panel). Therefore for a 3 °C global increase the uncertainty extends to 5 °C so, maps corresponding to 3–5 degrees should be considered.

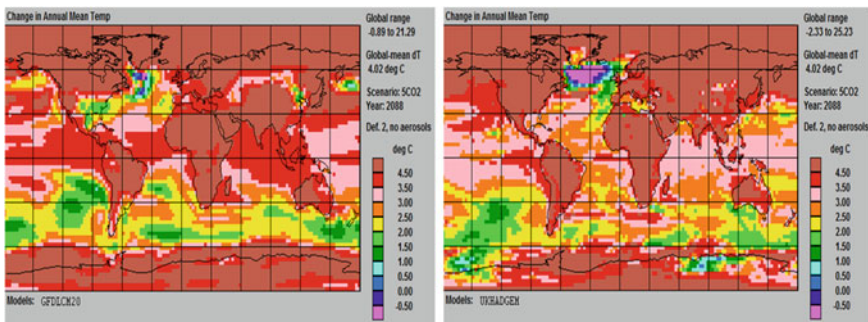
Now that we have the temperatures and the uncertainty intervals we use the Magicc/Scengen to obtain the maps for the temperatures referred above. This is done next.

As an example the results for the GCMs: (Geophysical Fluid Dynamics Laboratory Coupled Model, version 2.0) GFDL 2.0 and (Hadley Centre Global Environmental Model version 1) HADGEM1 for 1, 2, 3 and 4 °C are shown (Figs. 4, 5, 6, 7, 8 and 9).

The maps obtained with Magicc/Scengen for the HADGEM1 model for an increase of 1.01 °C (and for  $\Delta T \sim 2$  °C) with 5 and 0 CO<sub>2</sub>, are almost identical, as expected; the same for the GFDL2.0, i.e., they are independent from the emission trajectories.



**Fig. 8** Spatial distribution of  $\Delta T = 3.0$  °C according to GFDL 2.0 (left panel) and HADGEM1 (right panel) for 5CO<sub>2</sub> emission trajectory. Maps were obtained using Magicc/Scengen V. 5.3



**Fig. 9** Spatial distribution of  $\Delta T = 4.02$  °C according to GFDL 2.0 (left panel) and HADGEM1 (right panel) for 5CO<sub>2</sub> emission trajectory. Maps obtained using Magicc/Scengen V. 5.3



Once we have temperatures, uncertainty intervals and two dimensional maps we can go back to the original question, but put in different terms. When is the temperature going to be one degree warmer than today? The answer: as soon as 2021 but there is the possibility of a larger increase. A picture of the warming can be imagined between maps of left and right panels shown in Fig. 4 or 5. Now if the temperature is 2 degrees? The answer is that all the maps shown in the figures would become possible.

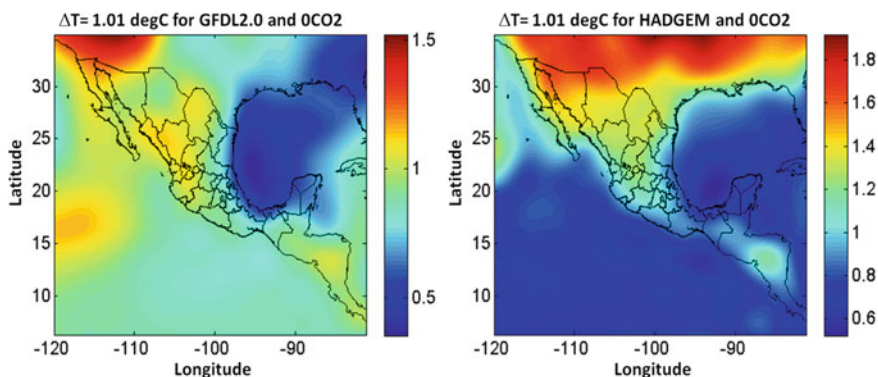
#### 4 A Regional Perspective of $\Delta T$ Thresholds

From a regional point of view, it can be made a similar analysis on the maps. Because of the size of the grid in data obtained from Magicc/Scengen, that is relatively big ( $2.5^\circ \times 2.5^\circ$ ), we proceed to reduce it by means of an interpolation algorithm, using splines. Afterwards, the size is reduced to a size of  $0.5' \times 0.5'$  (about  $10 \text{ km} \times 10 \text{ km}$ ). This methodology has been applied to propose regional climate scenarios for Mexico [1].

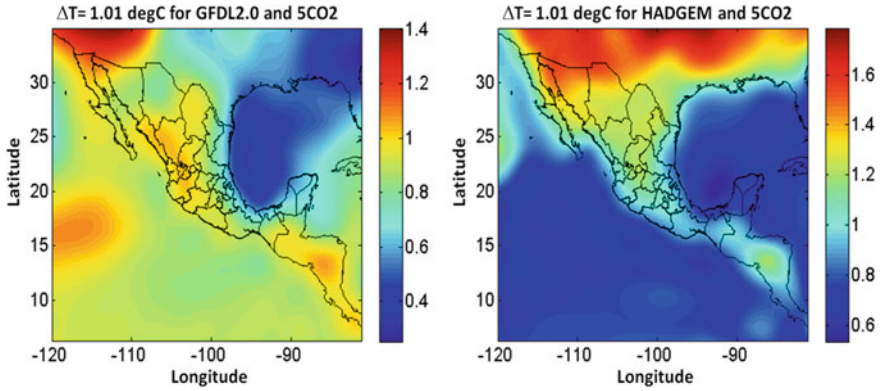
For purposes of comparison, maps for the same conditions as the ones shown in Figs. 4, 5, 6, 7, 8 and 9 are presented here in Figs. 10, 11, 12, 13, 14 and 15.

In these map is more evident that the maps are independent of the emission trajectory and, even, the structure of contour lines for each of the GCMs is the same, independently of the temperature threshold. Therefore, the uncertainty between GCMs has certain uniformity.

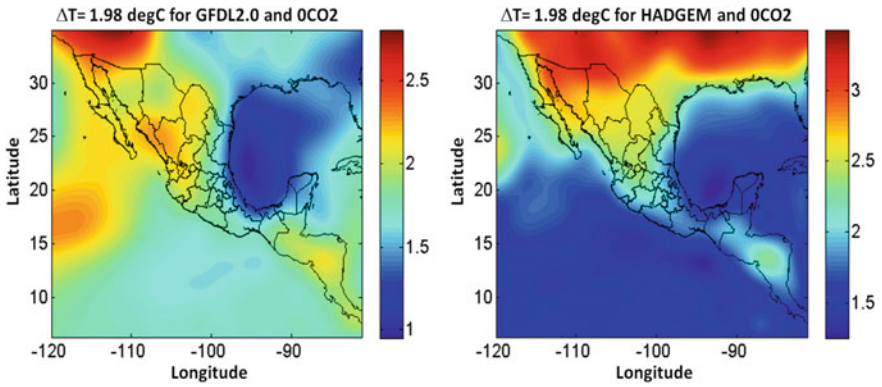
To regional scale decision makers and climate researchers it would be more adequate, for purposes of planning, to consider maps based on estimations of the thresholds of changes of climatic variables than effects projected to future time horizons as 2030, 2050 or 2100.



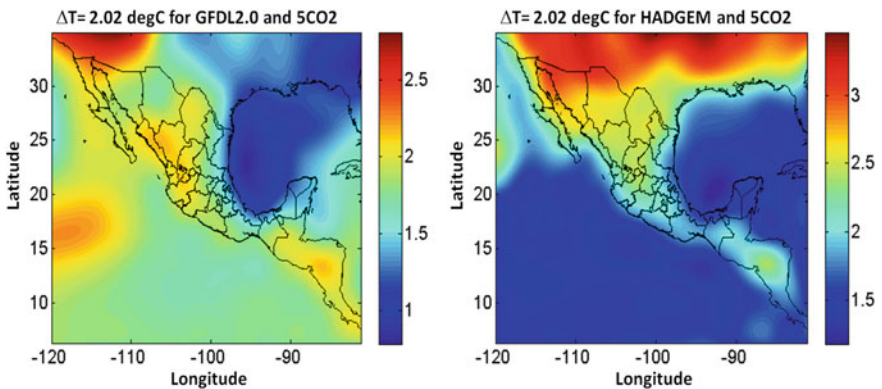
**Fig. 10** Regional scenario for Mexico for  $\Delta T_{\text{global}} = 1.01 \text{ }^\circ\text{C}$  according to GFDL 2.0 (*left panel*) and HADGEM1 (*right panel*) for  $0\text{CO}_2$  emission trajectory. Maps obtained using Magicc/Scengen V. 5.3 data and MATLAB routine



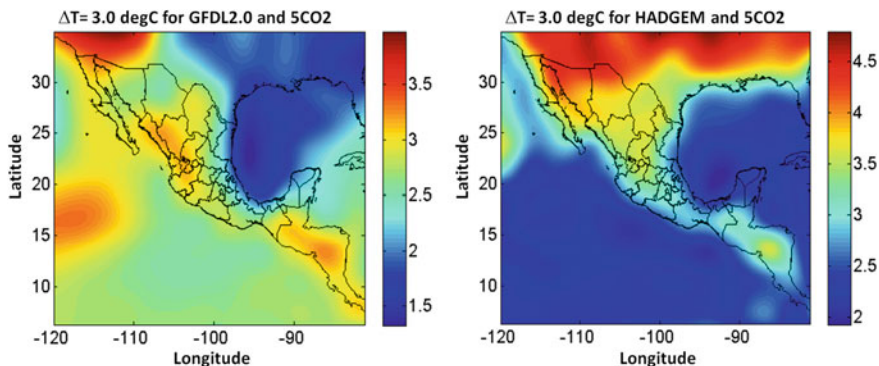
**Fig. 11** Regional scenario for Mexico for  $\Delta T_{\text{global}} = 1.01 \text{ }^\circ\text{C}$  according to GFDL 2.0 (*left panel*) and HADGEM1 (*right panel*) for 5CO<sub>2</sub> emission trajectory. Maps obtained using Magicc/Scengen V. 5.3 data and MATLAB routine



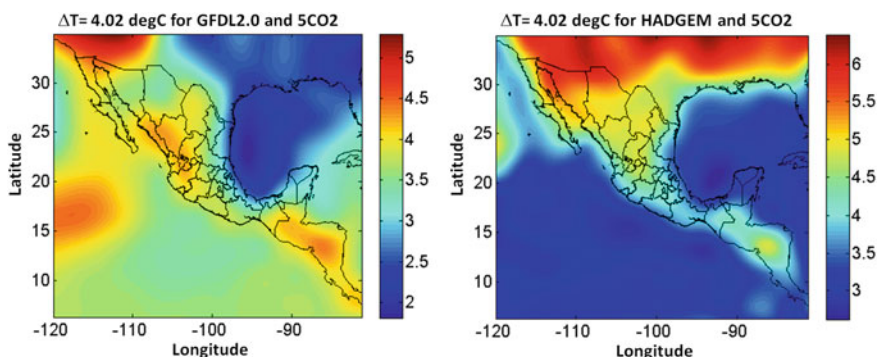
**Fig. 12** Regional scenario for Mexico for  $\Delta T_{\text{global}} = 1.98 \text{ }^\circ\text{C}$  according to GFDL 2.0 (*left panel*) and HADGEM1 (*right panel*) for 0CO<sub>2</sub> emission trajectory. Maps obtained using Magicc/Scengen V. 5.3 data and MATLAB routine



**Fig. 13** Regional scenario for Mexico for  $\Delta T_{\text{global}} = 2.02 \text{ }^\circ\text{C}$  according to GFDL 2.0 (*left panel*) and HADGEM1 (*right panel*) for 5CO<sub>2</sub> emission trajectory. Maps obtained using Magicc/Scengen V. 5.3 data and MATLAB routine



**Fig. 14** Regional scenario for Mexico for  $\Delta T_{\text{global}} = 3.0 \text{ }^{\circ}\text{C}$  according to GFDL 2.0 (*left panel*) and HADGEM1 (*right panel*) for  $5\text{CO}_2$  emission trajectory. Maps obtained using Magicc/Scengen V. 5.3 data and MATLAB routine



**Fig. 15** Regional scenario for Mexico for  $\Delta T_{\text{global}} = 4.02 \text{ }^{\circ}\text{C}$  according to GFDL 2.0 (*left panel*) and HADGEM1 (*right panel*) for  $5\text{CO}_2$  emission trajectory. Maps obtained using Magicc/Scengen V. 5.3. data and MATLAB routine

## 5 Discussion and Conclusions

Based on the fuzzy model presented by Gay et al. [3] and the simple climate model contained in Magicc/Scengen we show how the global mean temperature increase is distributed on the globe for the significant thresholds of 1–4  $^{\circ}\text{C}$ . The linear emission pathways include all the possibilities mentioned in successive reports of IPCC.

In this work we considered the possibility of analysing the impacts of temperature increase from the perspective of the year in which some temperature is reached. Two sources of uncertainty are taken into account, the emissions of GHG and the climate sensitivity.

The larger concentration and sensitivity the sooner the successive thresholds of temperature will be reached. If the sensitivity is 6 there is no way of staying at two

degrees unless the concentrations of CO<sub>2</sub> had followed the  $-2\text{CO}_2$  trajectory: negative emissions that means very strong subtraction of CO<sub>2</sub> from the atmosphere. We think that it is easier to consider a degree by degree strategy than one based on dates. For a one degree global increase the uncertainty extends to more than two degrees, then for a 1 °C global increase, maps for one and two degrees are to be considered. For 4 °C and sensitivity 3, uncertainty can extend to 6.41 °C.

We construct maps for 2 GCM's (as an example) with the necessary concentration to reach 1–4 °C limits to 2100. The maps show the spatial distribution of the temperature increase over the globe.

Emissions and sensitivity introduce uncertainties in the temperature that in turn must be reflected in the scaled temperature displayed in a map. Other source of uncertainty considered is the GCM. As expected, the map for any limit of temperature depends on the GCM but not on the emission trajectory. The maps constructed for different GCM's illustrate all possibilities for a region of the globe.

Additionally, we present maps on regional scale for Mexico, corresponding to the global maps mentioned above. It is evident that maps constructed from Magicc/Scengen for each GCM are independent of emission trajectory and the structure of its contour lines is independent of the temperature threshold.

Future work can be done to show how the GCM's introduce uncertainty in the estimates of precipitation change in global and regional scales, and to include a wider and complete selection of GCM's according to some regional criteria.

**Acknowledgments** This work was supported by the Programa de Investigación en Cambio Climático (PINCC, [www.pincc.unam.mx](http://www.pincc.unam.mx)) of the Universidad Nacional Autónoma de México.

## References

1. Conde C, Estrada F, Martínez B, Sánchez O, Gay C (2011) Regional climate change scenarios for México. *Atmosfera* 24(1):125–140. ISSN:0187-6236
2. Gay C, Sánchez O, Martínez-López B, Nébot Á, Estrada F (2012) Simple fuzzy logic models to estimate the global temperature change due to GHG emissions. In: 2nd international conference on simulation and modeling methodologies, technologies and applications (SIMULTECH). Special session on applications of modeling and simulation to climatic change and environmental sciences—MSCCEC 2012, July 28–31. Rome, Italy. Thomson Reuters conference proceedings citation index (ISI), INSPEC, DBLP and EI (Elsevier Index) <http://www.informatik.uni-trier.de/~ley/db/conf/simultech/simultech2012.html>
3. Gay C, Sánchez O, Martínez-López B, Nébot Á, Estrada F (2013) Fuzzy models: easier to understand and an easier way to handle uncertainties in climate change research. In: Pina N, Kacprzyk J, Filipe J (vol. eds.) *Simulation and modeling methodologies, technologies and applications. Advances in intelligent and soft computing*. Springer, Berlin (in review)
4. IPCC-WGI (2007) In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) *Climate change 2007: the physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge University Press, Cambridge, 996 pp
5. Nakicenovic N, Alcamo J, Davis G, de Vries B, Fenhann J, Gaffin S, Gregory K, Grübler A, Jung TY, Kram T, La Rovere EL, Michaelis L, Mori S, Morita T, Pepper W, Pitcher H, Price L,

- Riahi K, Roehrl A, Rogner H-H, Sankovski A, Schlesinger M, Shukla P, Smith S, Swart R, van Rooijen S, Victor N, Dadi Z (2000) Special report on emissions scenarios: a special report of working group III of the intergovernmental panel on climate change. Cambridge University Press, Cambridge 599 pp
6. Wigley TML (2008) MAGICC/SCENGEN V. 5.3: User manual (version 2). NCAR, Boulder. 80 pp. <http://www.cgd.ucar.edu/cas/wigley/magicc/>
  7. Zadeh LA (1965) Fuzzy sets. Inf Control 8(3):338–353

# Efficient Design of Inline E-Plane Waveguide Extracted Pole Filters Through Enhanced Equivalent Circuits and Space Mapping

Oleksandr Glubokov, Slawomir Koziel and Leifur Leifsson

**Abstract** A design procedure for inline waveguide extracted pole filters with all-metal E-plane inserts is presented. To achieve acceptable modeling accuracy for this class of filters, an enhanced schematic-circuit-based surrogate model is developed, accounting for parasitic effects between the neighboring elements of the structure. The proposed circuit representation is used as a coarse model in a surrogate-based optimization of the filter structure with the space mapping technique as the main engine in the optimization process. Feasibility of the modeling approach is demonstrated by two filter design examples. The examples show that a low computational cost (corresponding to a few evaluations of the high-fidelity EM simulations of the filter structure) is required to obtain an optimized design.

**Keywords** Inline filters · E-plane filters · Extracted pole filters · Waveguide · Modeling · Surrogate optimization

## 1 Introduction

Filters based on rectangular waveguide structures have been one of the most popular solutions for frequency selection in microwave and millimeter-wave systems in the past decades. However, massive housings, typically used for implementation of the waveguide structures, make them less attractive for specific applications such as satellite communications. It is, therefore, highly desirable to

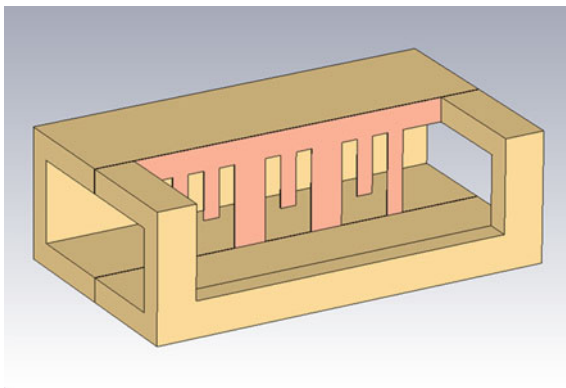
---

O. Glubokov · S. Koziel (✉) · L. Leifsson  
Engineering Optimization and Modeling Center, School of Science and Engineering,  
Reykjavik University, Menntavegur 1, Reykjavik 101, Iceland  
e-mail: koziel@ru.is

O. Glubokov  
e-mail: oleksandr@ru.is

L. Leifsson  
e-mail: leifurth@ru.is

**Fig. 1** Configuration of an inline E-plane waveguide extracted pole filter



miniaturize the components. This can be achieved by using an extracted pole filters concept. Inline E-plane extracted pole filters, which are cheap and easy-to-fabricate, allow for designing the devices with the required compactness, as well as preserving their filtering properties. Miniaturized inline E-plane extracted pole filters with high stopband performance using standard and enhanced extracted pole sections have been previously reported in [7, 8]. In [7], a generalized coupling coefficients extraction procedure is developed, allowing one to obtain a good initial design for this type of filters. However, further optimization of the filter structures is usually time consuming, as numerous analyses of densely meshed objects in full wave simulators are required. Surrogate-based optimization (SBO) offers a way to reduce the number of the required high-fidelity EM simulations [10].

The inline E-plane filters consist of several extracted pole sections containing touching and non-touching axial and transversal strips arranged within a rectangular waveguide, as shown in Fig. 1. Chang and Khan [3] have proposed equation-based surrogate models for these elements. Unfortunately, the model of the non-touching strip is not available in commonly used circuit simulators (e.g., Agilent ADS, [1]) as a component. At the same time, an explicit equation is only available for the transversally coupled strips [4], while couplings between the elements have a significant impact on the performance of the entire filter.

This chapter describes equation-based surrogate models for the touching and non-touching strips placed in the rectangular waveguide's E-plane. A surrogate circuit model of the filter is composed of the models where the coupling between adjacent strips are accounted for by tuning elements. The circuit model is subsequently used in an SBO scheme where space mapping [2, 9] is used as the optimization algorithm with a suitably corrected circuit model utilized as the surrogate. We demonstrate that the SBO approach allows us to achieve the optimal design at a low computation cost corresponding to a few high-fidelity EM simulations of the structures of interest using 2nd- and 4th-order EPS filter examples.



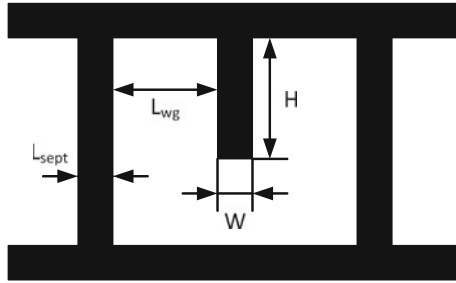


Fig. 2 E-plane waveguide insert with a resonating fin implementing an extracted pole section

## 2 Extracted Pole Section's Model

A single extracted pole section (EPS) can be implemented by an insert (here, all-metal) placed within the E-plane of a rectangular waveguide. The insert, shown in Fig. 2, consists of two strips connecting top and bottom ground planes (input/output septa) and a strip connected to the top ground plane with the other end left open (resonating fin); the septa and the fin are connected in series through uniform waveguide sections.

The EPS is capable of producing a single transmission zero at the fin's resonant frequency and up to two poles depending on the lengths of the waveguide sections  $L_{wg}$ . If  $L_{wg} < \lambda_g/2$ , the transmission zero appears above a single pole. In this case inductive and capacitive couplings between the fin and the septa are non-negligible and strongly affect the positions of the pole and zero, especially for small  $L_{wg}$ . The couplings may be neglected in the case of  $L_{wg} \approx \lambda_g/2$ , when two closely located poles can be obtained with the transmission zero located on either side of them, or even between the poles.

In order to create a surrogate model, the EPS is decomposed into septa, fin and waveguide sections. The septa and the fin are represented as circuit models; next, we compose the EPS of the developed models, and add tuning elements to the obtained circuit in order to take into account the couplings between the septa and the fin. The details are provided in the following subsections.

### 2.1 Septum Model

The circuit model of the septum is shown in Fig. 3. It contains four parameters: two inductances  $L_{s1}$  and  $L_{s2}$ , a capacitance  $C_s$  and a coupling coefficient  $k_s$ . It is implied for all the modeled structures in this paper that the waveguide is air-filled and its parameters, height and width, are kept constant for the design; therefore the septum's model parameters depend on the design parameter  $L_{sept}$  only.



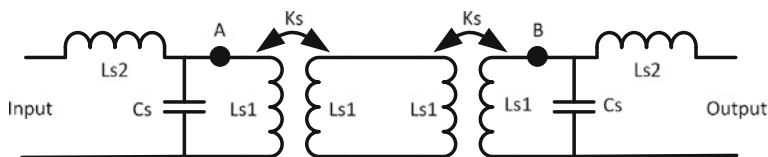


Fig. 3 Circuit model of the septum

Initial experiments indicate that the following set of equations is suitable to approximate the dependence of the model parameters on  $L_{sept}$  (denoted here as  $x$  in the interest of better readability):

$$L_{s1} = a_{L_{s1}}e^{-b_{L_{s1}}x} + c_{L_{s1}}e^{-d_{L_{s1}}x} \tag{1a}$$

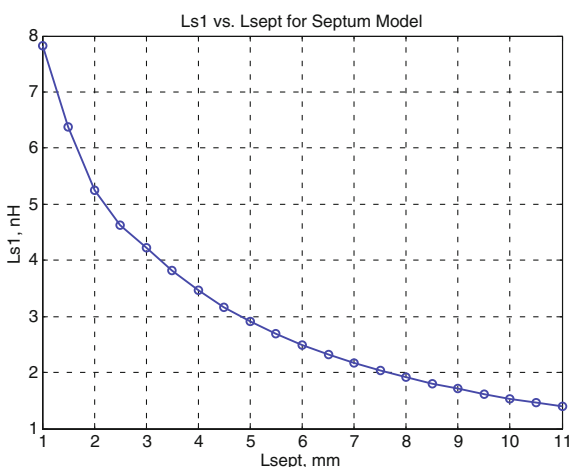
$$L_{s2} = a_{L_{s2}}x^2 + b_{L_{s2}}x + c_{L_{s2}} \tag{1b}$$

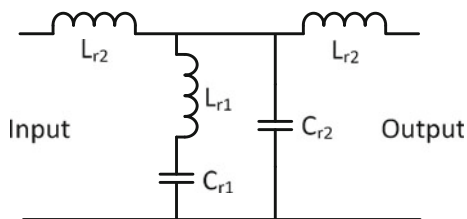
$$C_s = a_{C_s}e^{-b_{C_s}x} + c_{C_s}e^{-d_{C_s}x} \tag{1c}$$

$$K_s = a_{K_s}x^2 + b_{K_s}x + c_{K_s} \tag{1d}$$

In order to extract the model parameters, high-fidelity models of the septa are built and analyzed in CST Microwave Studio™ for  $L_{sept} = 1 \dots 11$  mm. The real and imaginary parts of the S-parameters of the proposed circuit model are subsequently fitted to the ones obtained from simulations using the least squares criterion within the frequency range 8...14 GHz. The extraction is illustrated in Fig. 4, where one of the model parameters  $L_{s1}$  is plotted against the design parameter  $L_{sept}$ .

Fig. 4 Extracted value of  $L_{s1}$  as a function of the design parameter  $L_{sept}$





**Fig. 5** Circuit model of the resonating fin

## 2.2 Resonating Fin Model

Figure 5 presents a circuit model of the resonating fin. Similarly to the previously considered model of the septum, this one also contains four model parameters: two inductances  $L_{r1}$  and  $L_{r2}$ , and two capacitances  $C_{r1}$  and  $C_{r2}$ . The resonating fin has two design parameters: height  $H$  and width  $W$ . The range of variation of the design parameters is determined based on the assumption that the transmission zeros are located within the frequency range of 8...14 GHz; this condition is sufficient for all the variety of filter design problems that can be solved by the structures under consideration. Consequently, the resonating fin model is created for  $H = 4.5...9.5$  mm and  $W = 1...6$  mm.

Following the investigation of the relationship between the design and model parameters, it has been determined that the best fit can be obtained by using the set of equations:

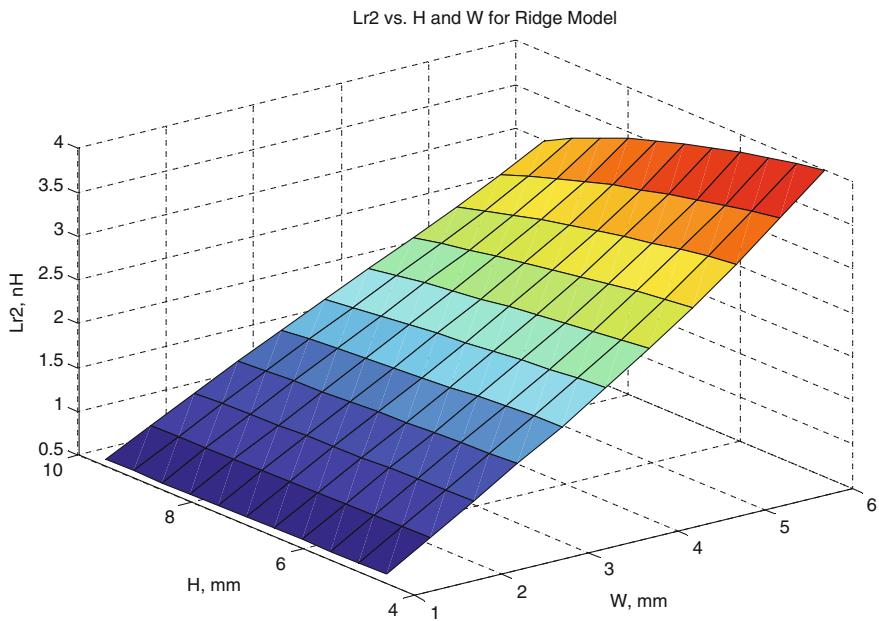
$$L_{r1} = (e^{a_{Lr1}H} + e^{b_{Lr1}W}) \left[ c_{Lr1} \left( \frac{H}{W} \right)^2 + d_{Lr1} \frac{H}{W} + f_{Lr1} \right] \quad (2a)$$

$$L_{r2} = a_{Lr2}HW + b_{Lr2}H + c_{Lr2}W + d_{Lr2} \quad (2b)$$

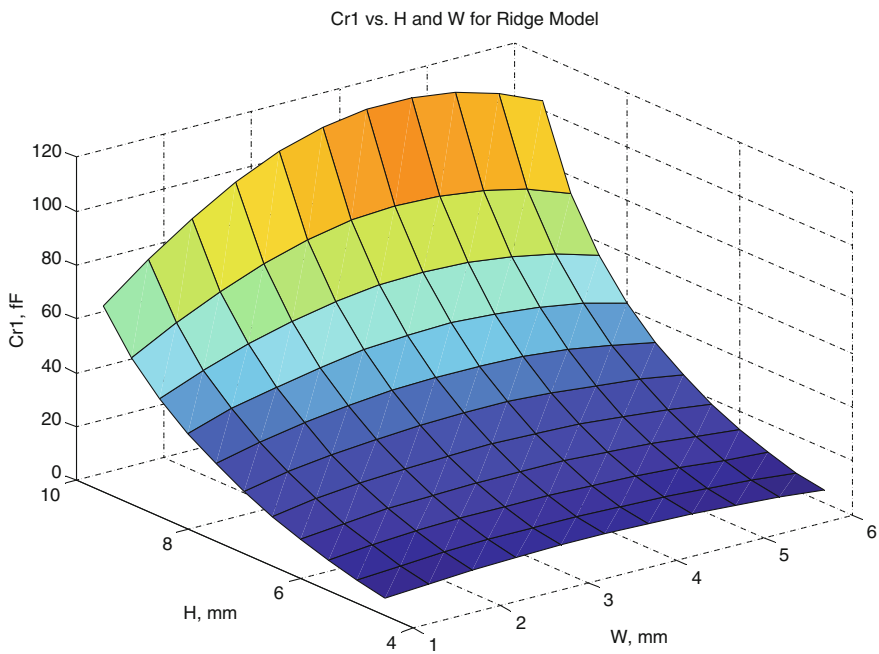
$$C_{r1} = (e^{a_{Cr1}H} + e^{b_{Cr1}W}) \left[ c_{Lr1} \frac{H}{W} + d_{Lr1} \right] \quad (2c)$$

$$C_{r2} = a_{Cr2}H^2 + b_{Cr2}HW + c_{Cr2}H + d_{Cr2}W + f_{Cr2} \quad (2d)$$

Extraction of the fin model parameters has also been carried out by fitting the real and imaginary parts of the S-parameters of the circuit model to the ones simulated in CST Microwave Studio™. Typical results of the fin modeling procedure are demonstrated in Figs. 6 and 7.



**Fig. 6** Extracted value of  $L_{r2}$  as a function of height  $H$  and width  $W$  of the resonating fin



**Fig. 7** Extracted value of  $C_{r1}$  with respect to the height  $H$  and width  $W$  of the resonating fin

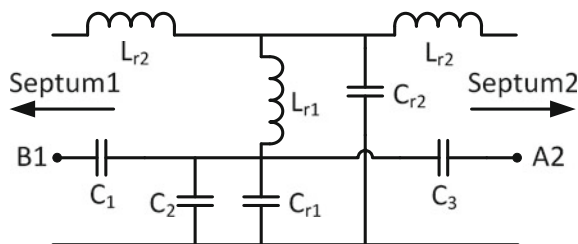


Fig. 8 Model of the capacitive coupling between septa and a fin

### 2.3 Couplings Model

In order to achieve better upper stopband performance of the filter one needs to place transmission zeros as close as possible to the filter's passband. In the inline E-plane waveguide filter this can be made by decreasing the lengths of the waveguide sections  $L_{wg}$ . On this condition, parasitic inductive and capacitive couplings between the adjacent elements significantly affect the frequency responses: poles and zeros drift to the new positions. Therefore, coupling effects between the septa and the resonating fins should be taken into account in the circuit model. In the model, we arrange the inductive coupling by the magnetic coupling coefficient  $K_{sr}$  introduced between the inductors  $L_{s1}$  and  $L_{r1}$ . At the same time, we take the capacitive coupling into consideration by adding three capacitors  $C_1$ ,  $C_2$  and  $C_3$ , as shown in Fig. 8.

$$K_{sr,i} \approx k_i H_i / L_{wg,i} \quad (3a)$$

$$C_{1,i} \approx c_{1i} L_{sept,i} / L_{wg,i} \quad (3b)$$

$$C_{2i} \approx c_{2i} H_i / L_{wg,i} \quad (3c)$$

$$C_{3i} \approx c_{3i} L_{sept,i+1} / L_{wg,i} \quad (3d)$$

## 3 Filter Optimization

The optimization procedure adopts the implicit space mapping technique [5]. Let us denote the response vector of the surrogate model by  $\mathbf{R}_s(\mathbf{x}, \mathbf{t})$  and the response vector of the fine model by  $\mathbf{R}_f(\mathbf{x})$ . The mode response vectors represent  $S$ -parameters of the filter structure as a function of frequency. Here,  $\mathbf{x}$  is the design parameters vector and  $\mathbf{t}$  is the tuning parameters vector. In our case, the design parameters are geometry dimensions of the filter structure, whereas tuning variables are certain parameters of the surrogate model that are adjusted in order to reduce misalignment between the surrogate and the fine model.

The overall goal is to find

$$\mathbf{x}_{opt} = \arg \min_{\mathbf{x}} U(\mathbf{R}_f(\mathbf{x})) \quad (4)$$

where  $U$  is an objective function that encodes the design specifications. Here, we use minimax objective function [2] that determines the maximum violation of the design requirements imposed on  $S$ -parameters over frequency bands of interest.

### 3.1 Algorithm

The optimization algorithm is illustrated by a diagram in Fig. 9, and can be described as follows:

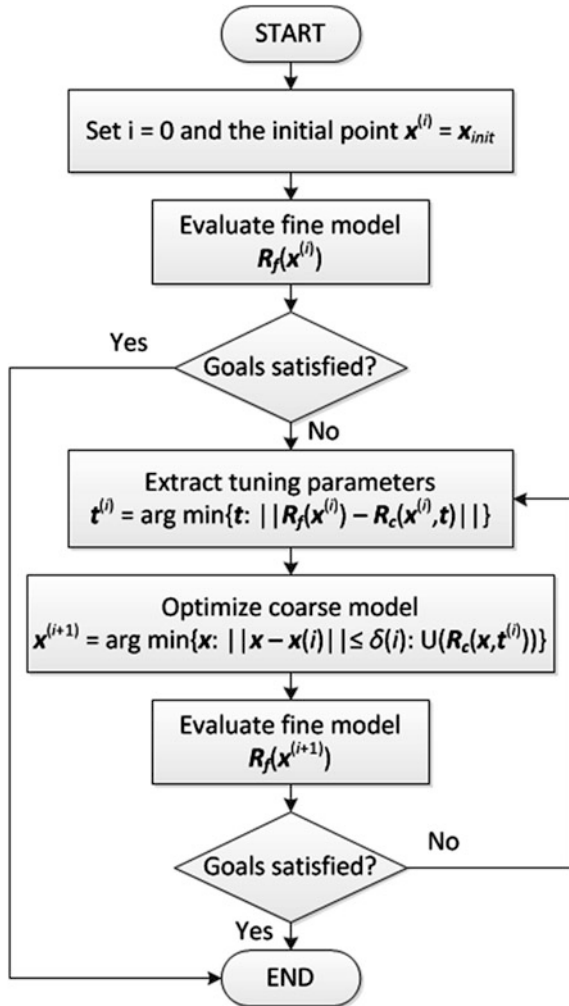
1. Set  $i = 0$  and the initial point  $\mathbf{x}^{(i)} = \mathbf{x}^{init}$ ;
2. Evaluate  $\mathbf{R}_f(\mathbf{x}^{(i)})$ ;
3. Calibrate the surrogate model by adjusting the tuning parameters  $\mathbf{t}^{(i)} = \arg \min \{t : \|\mathbf{R}_f(\mathbf{x}^{(i)}) - \mathbf{R}_s(\mathbf{x}^{(i)}, \mathbf{t})\|\}$ ;
4. Obtain the new design by optimizing the surrogate model  $\mathbf{x}^{(i+1)} = \arg \min \{\mathbf{x}, \|\mathbf{x} - \mathbf{x}^{(i)}\| \leq \delta^{(i)} : U(\mathbf{R}_s(\mathbf{x}, \mathbf{t}^{(i)}))\}$ ;
5. Evaluate the fine model  $\mathbf{R}_f(\mathbf{x}^{(i+1)})$ ;
6. If the termination conditions are not satisfied, set  $i = i + 1$  and go to step 3; else END.

In Step 3, the parameter extraction procedure is executed that identifies the values of the tuning parameters  $\mathbf{t}$  so that the surrogate model becomes as good representation of the fine model at the current iteration point  $\mathbf{x}^{(i)}$ . As indicated in Step 4, the surrogate model optimization is embedded in the trust region (TR) framework [6] to improve convergence properties of the algorithm;  $\delta^{(i)}$  is the TR radius updated in each iteration using conventional rules [6]. The procedure is terminated if the goals are satisfied or if the predefined number of iteration exceeded.

## 4 Illustration Examples

In this section, we illustrate the design methodology introduced in this Sect. 3 using two examples of E-plane filters of the second- and the fourth order.

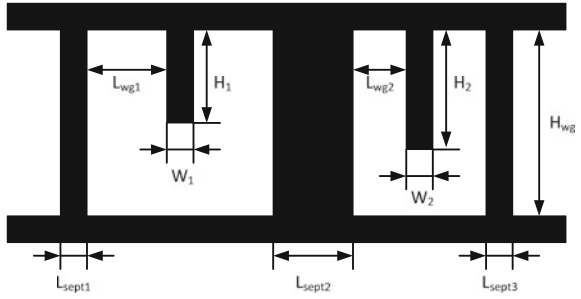
**Fig. 9** Optimization flow diagram



**4.1 Example 1: 2nd-Order Filter with Two EPS**

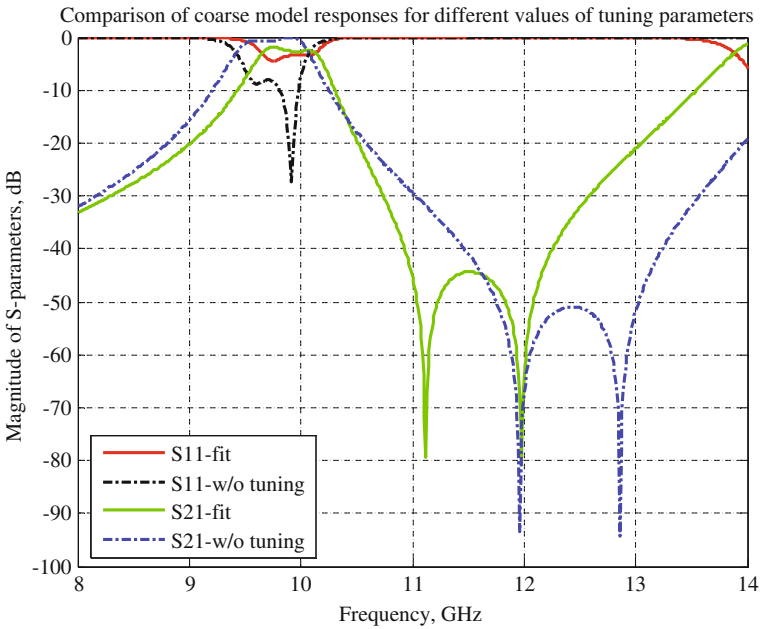
The models and the optimization algorithm were tested on a 2nd-order inline extracted pole bandpass filter with two symmetric EPS connected in series through a septum. Figure 10 presents the configuration of the E-plane insert for the filter. The design parameters are  $\mathbf{x} = [H_1 H_2 L_{sept1} L_{sept2} L_{sept3} L_{wg1} L_{wg2}]^T$ . Width of the fins is fixed to  $W = 2$  mm. The waveguide parameters are: width  $W_{wg} = 22.86$  mm, height  $H_{wg} = 10.16$  mm and  $\epsilon_r = 1.00059$ ; they remain constant throughout the entire optimization procedure. The corresponding specifications of the filter are given as follows:  $|S_{11}| \leq -20$  dB for  $9.92$  GHz  $\leq f \leq 10.08$  GHz, and  $|S_{21}| \leq -40$  dB for  $10.9$  GHz  $\leq f \leq 12.2$  GHz.

**Fig. 10** Configuration of the E-plane insert for the 2nd-order filter

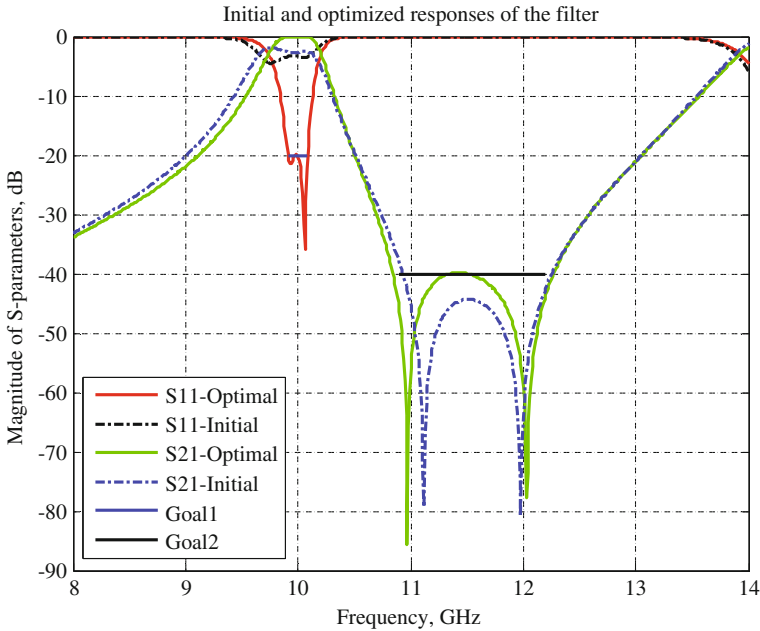


The initial design parameters  $\mathbf{x}_{init} = [6 \ 5.5 \ 1.5 \ 6.5 \ 2 \ 1.5 \ 3]^T$  mm are found by the generalized coupling coefficients extraction technique [7], which is helpful at attaining a good initial design.

The tuning parameters  $\mathbf{t} = [k_1 \ k_2 \ k_3 \ k_4 \ c_1 \ c_2 \ c_3 \ c_4 \ c_5 \ c_6]^T$  are obtained by fitting absolute values of the S-parameters by the least square criterion. Quality of extraction of the tuning parameters is crucial for the success of the procedure, as discrepancy between  $\mathbf{R}_s(\mathbf{x}_{init}, \mathbf{t}(0))$  and  $\mathbf{R}_s(\mathbf{x}_{init}, 0)$  is rather considerable, as illustrated in Fig. 11. On the other hand, the response  $\mathbf{R}_s(\mathbf{x}_{init}, \mathbf{t}(0))$  of the corrected surrogate is virtually indistinguishable from  $\mathbf{R}_f(\mathbf{x}_{init})$ , which indicates that the circuit model developed in Sect. 2 is a suitable tool to conduct space mapping optimization of the extracted pole filter.



**Fig. 11** Comparison of  $\mathbf{R}_s(\mathbf{x}_{init}, \mathbf{t}(0))$  and  $\mathbf{R}_s(\mathbf{x}_{init}, 0)$ .  $\mathbf{R}_s(\mathbf{x}_{init}, \mathbf{t}(0))$  is essentially identical to fine model response at  $\mathbf{x}_{init}$



**Fig. 12** Initial and optimized responses of the 2nd-order filter

Figure 12 shows the optimal response  $\mathbf{R}_f(\mathbf{x}_{opt})$  compared to the initial  $\mathbf{R}_f(\mathbf{x}_{init})$ , where  $\mathbf{x}_{opt} = [6.09 \ 5.45 \ 1.71 \ 6.42 \ 2.09 \ 1.42 \ 2.77]^T$  mm. The optimum has been obtained for 6 iterations and 9 evaluations of the fine model.

### 4.2 Example 2: 4th-Order Filter with Three EPS

A 4th-order inline extracted pole bandpass filters with three EPS, one of which produces two poles and a zero [8], is chosen as another test example. Configuration of the E-plane waveguide insert implementing the filter is shown in Fig. 13. The filter contains three symmetric EPS: the two standard ones, as considered in Sect. 4.1, are directly coupled to input and output through a septum, while the dual-mode EPS is located between them. All the EPS are connected in series by means of septa.

The design parameters are  $\mathbf{x} = [H_1 \ H_2 \ H_3 \ L_{sept1} \ L_{sept2} \ L_{sept3} \ L_{sept4} \ L_{wg1} \ L_{wg2} \ L_{wg3} \ W_2]^T$ . Widths of the first and third fins are fixed to  $W_1 = W_3 = 1.2$  mm. The waveguide parameters are: width  $W_{wg} = 22.86$  mm, height  $H_{wg} = 10.16$  mm and  $\epsilon_r = 1.00059$ , and they remain unchanged throughout the optimization.

The corresponding specifications of the filter are given as follows:

- $|S_{11}| \leq -20$  dB for  $10.925 \text{ GHz} \leq f \leq 11.075 \text{ GHz}$ ;
- $|S_{21}| \leq -60$  dB for  $10.58 \text{ GHz} \leq f \leq 10.62 \text{ GHz}$ ;



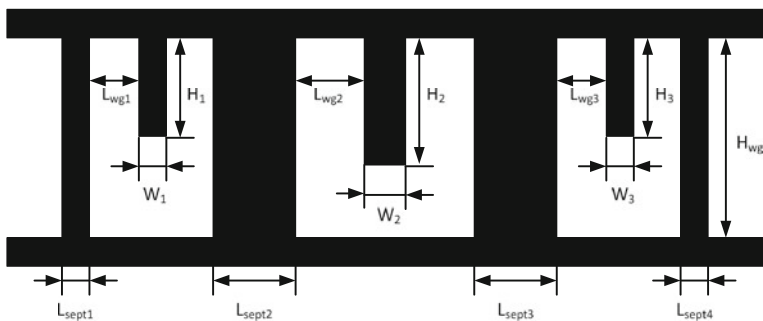


Fig. 13 Configuration of the E-plane insert for the 4th-order filter

- $|S_{21}| \leq -60$  dB for  $11.38$  GHz  $\leq f \leq 11.42$  GHz;
- $|S_{21}| \leq -70$  dB for  $11.68$  GHz  $\leq f \leq 11.72$  GHz.

The initial design parameters  $x_{init} = [6.1 \ 7.9 \ 5.85 \ 1.1 \ 10.5 \ 10.2 \ 1.4 \ 0.8 \ 12.5 \ 1.1 \ 3.5]^T$  mm are found by the generalized coupling coefficients extraction technique [7] and further tuning.

The tuning parameters  $t = [k_1 \ k_2 \ k_3 \ k_4 \ k_5 \ k_6 \ c_1 \ c_2 \ c_3 \ c_4 \ c_5 \ c_6 \ c_7 \ c_8 \ c_9 \ L_{add}]^T$  are obtained by fitting absolute values of the S-parameters by the least square criterion. Here, an additional waveguide section  $L_{add}$  has been added in series to  $L_{wg2}$  as a tuning parameter in order to improve the coarse model  $R_s(x_{init}, t(0))$  and enable fitting of the fine model response  $R_f(x_{init})$  with sufficient accuracy.

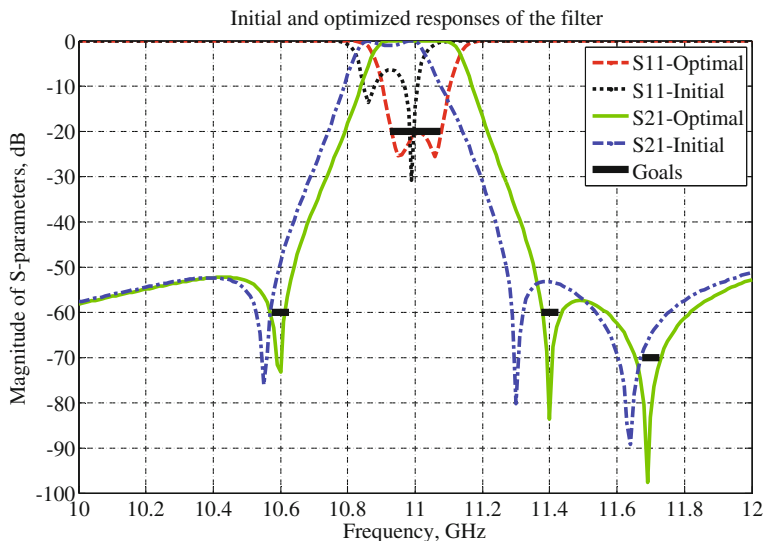


Fig. 14 Initial and optimized responses of the 4th-order filter

The optimized design has been obtained in 9 space mapping iterations with 14 evaluations of the fine model. The optimal response  $\mathbf{R}_f(\mathbf{x}_{opt})$  compared to the initial  $\mathbf{R}_f(\mathbf{x}_{init})$ , where  $\mathbf{x}_{opt} = [6.05 \ 7.89 \ 5.80 \ 1.19 \ 10.31 \ 10.39 \ 1.29 \ 0.74 \ 12.30 \ 1.26 \ 3.60]^T$  mm, is shown in Fig. 14.

## 5 Conclusions

A design procedure for inline E-plane waveguide extracted pole filters that includes surrogate modeling technique of this type of structures has been presented. The solution contains novel circuit models for septa, resonating fins and coupling effects between them, and enables use of surrogate-based optimization for this class of filters. Two design examples of a 2nd-order filter with 2 EPS and a 4th-order filter with 3 EPS connected in series have been demonstrated using implicit space mapping as an optimization engine. The optimum designs have been obtained at the cost of a few EM simulations of the filter structure.

**Acknowledgments** The authors thank Computer Simulation Technology AG for making CST Microwave Studio available. This work was supported in part by the Icelandic Centre for Research (RANNIS) Grants 120016021 and 13045051.

## References

1. Agilent ADS version (2011) Agilent Technologies, Santa Clara, CA, USA
2. Bandler JW, Cheng QS, Dakrouy SA, Mohamed AS, Bakr MH, Madsen K, Sondergaard J (2004) Space mapping: the state of the art. *IEEE Trans Microwave Theory Tech* 52 (1):337–361
3. Chang K, Khan PJ (1976a) Equivalent circuit of a narrow axial strip in waveguide. *IEEE Trans Microwave Theory Tech* 24(9):611–615
4. Chang K, Khan PJ (1976b) Coupling between narrow transverse inductive strips in waveguide. *IEEE Trans Microwave Theory Tech* 24(2):101–105
5. Cheng QS, Bandler JW, Koziel S (2008) An accurate microstrip hairpin filter design using implicit space mapping. *Microwave Mag* 9(1):79–88
6. Conn AR, Gould NIM, Toint PL (2000) Trust region methods. *MPS-SIAM Series on Optimization*
7. Glubokov O, Budimir D (2011a) Extraction of generalized coupling coefficients for inline extracted pole filters with nonresonating nodes. *IEEE Trans Microwave Theory Tech* 59 (12):3023–3029
8. Glubokov O, Budimir D (2011b) Novel inline waveguide e-plane filter using dual-mode extracted pole section. In: *European microwave conference (EuMC)*, pp 99–102
9. Koziel S, Bandler JW, Madsen K (2006) Space mapping framework for engineering optimization: theory and implementation. *IEEE Trans Microwave Theory Tech* 54:3721–3730
10. Koziel S, Cheng QS, Bandler JW (2008) Space mapping. *IEEE Microwave Mag* 9(6):105–122

# Decomposition and Space Mapping for Reduced-Cost Modeling of Waveguide Filters

Slawomir Koziel, Stanislav Ogurtsov and Leifur Leifsson

**Abstract** In this work, we present a technique for low-cost surrogate modeling of waveguide filters. The proposed methodology is based on the decomposition of the filter structure. Some of the decomposed parts are modeled using response surface approximations (RSAs). The RSA models are subsequently combined with analytical models of the waveguide sections to form an initial filter surrogate. As a result of electromagnetic couplings between the decomposed parts, which are not accounted for by the initial surrogate, its accuracy is limited. This misalignment is reduced by applying space mapping at the level of the complete filter structure. Decomposition approach allows us to greatly reduce the computational cost of creating the surrogate because the time required to simulate the structure in parts is much lower than the time for simulating the entire filter. Moreover, the number of parameters describing each part is lower than for the entire filter. The presented technique is demonstrated using two test cases. Application examples are also given.

**Keywords** Computer-aided design · Electromagnetic simulation · Surrogate modeling · Waveguide filters · Decomposition · Space mapping

## 1 Introduction

High-fidelity electromagnetic (EM) simulation offers an accurate but—at the same time—a computationally intensive way of evaluating microwave structures. It is often advantageous to replace EM models by computationally cheap surrogates

---

S. Koziel (✉) · S. Ogurtsov · L. Leifsson  
Engineering Optimization and Modeling Center, School of Science and Engineering,  
Reykjavik University, Menntavegur 1, Reykjavik 101, Iceland  
e-mail: koziel@ru.is

S. Ogurtsov  
e-mail: stanislav@ru.is

L. Leifsson  
e-mail: leifurth@ru.is

in situations that require numerous evaluations of a structure of interest, such as parametric design optimization or yield-driven design.

A number of distinct methodologies allowing for building low-cost surrogates have been proposed in the literature. These include function approximation models such as artificial neural networks (ANNs) [6, 13], fuzzy systems [10], multidimensional Cauchy approximation [14], radial basis function interpolation [12], kriging [12], or support vector regression (SVR) [15]. All of these methods, unfortunately, require large amounts of training data obtained through numerous EM simulations and suffer from the so-called *curse of dimensionality*, i.e., the number of training samples necessary to ensure given accuracy grows exponentially with the number of designable parameters. This is a problem when the surrogate is to be created for a one-time design of a given structure.

More recent approaches exploit physics-based models. Methods such as space mapping (SM) [1, 2] or shape-preserving response prediction (SPRP) [7], construct the surrogate by aligning the underlying low-fidelity model (e.g., a circuit equivalent or a coarse-mesh EM simulation) to the high-fidelity EM-simulation data at a limited number of training points. By utilizing the knowledge embedded in the low-fidelity model, physics-based surrogates can be more efficient. The low-fidelity model itself has to be sufficiently reliable and fast to ensure accuracy and computational efficiency.

In the particular case of waveguide filters, a possible modeling approach is by using decomposition [3], where the structure is split into smaller sections, each one modeled separately, and then combined into a surrogate using wave transfer matrices of each section computed from S-parameters evaluated with the dominant mode excitation. One of the problems here is that this straightforward decomposition neglects possible EM interactions between the sections through the higher order modes, which might result in reduced accuracy of the surrogate. Taking into account interactions through the higher order modes [11] slows down the evaluation of the filter response. Furthermore, accuracy of S-parameters below cut-off evaluated with a coarse discretization at different points of the modeling space needs a special investigation.

In this work, we develop a decomposition-based modeling procedure that uses space mapping as a way of correcting the initial surrogate model. The decomposed parts are modeled by kriging interpolation [8] (iris discontinuities) as well as analytical formulas (waveguide sections). The resulting surrogate model is very fast, reliable, and created at a low computational cost. We demonstrate the accuracy of the proposed technique, as well as emphasize the importance of space mapping correction, using two filter examples. Applications for filter optimization are also discussed.

## 2 Modeling Using Decomposition and Space Mapping

We describe the proposed modeling procedure in this section. The main components are: (i) structure decomposition, (ii) local response surface approximation models, and (iii) space mapping correction.

## 2.1 Formulation of the Surrogate Modeling Problem

Let  $\mathbf{R}_f : X \rightarrow R^m$ ,  $X \subseteq R^n$  denote the response vector of the microwave device under consideration (fine model). For example,  $\mathbf{R}_f(\mathbf{x})$  may represent  $|S_{21}|$  at  $m$  chosen frequencies,  $\omega_1$  to  $\omega_m$ . The task is to build a surrogate model  $\mathbf{R}_s$  of  $\mathbf{R}_f$  so that the matching between the two models is as good as possible in the region of interest  $X_R \subseteq X$ . Typically,  $X_R$  is an  $n$ -dimensional interval in  $R^n$  with center at the reference point  $\mathbf{x}^0 = [x_{0,1} \dots x_{0,n}]^T \in R^n$  and size  $\delta = [\delta_1 \dots \delta_n]$  [1].

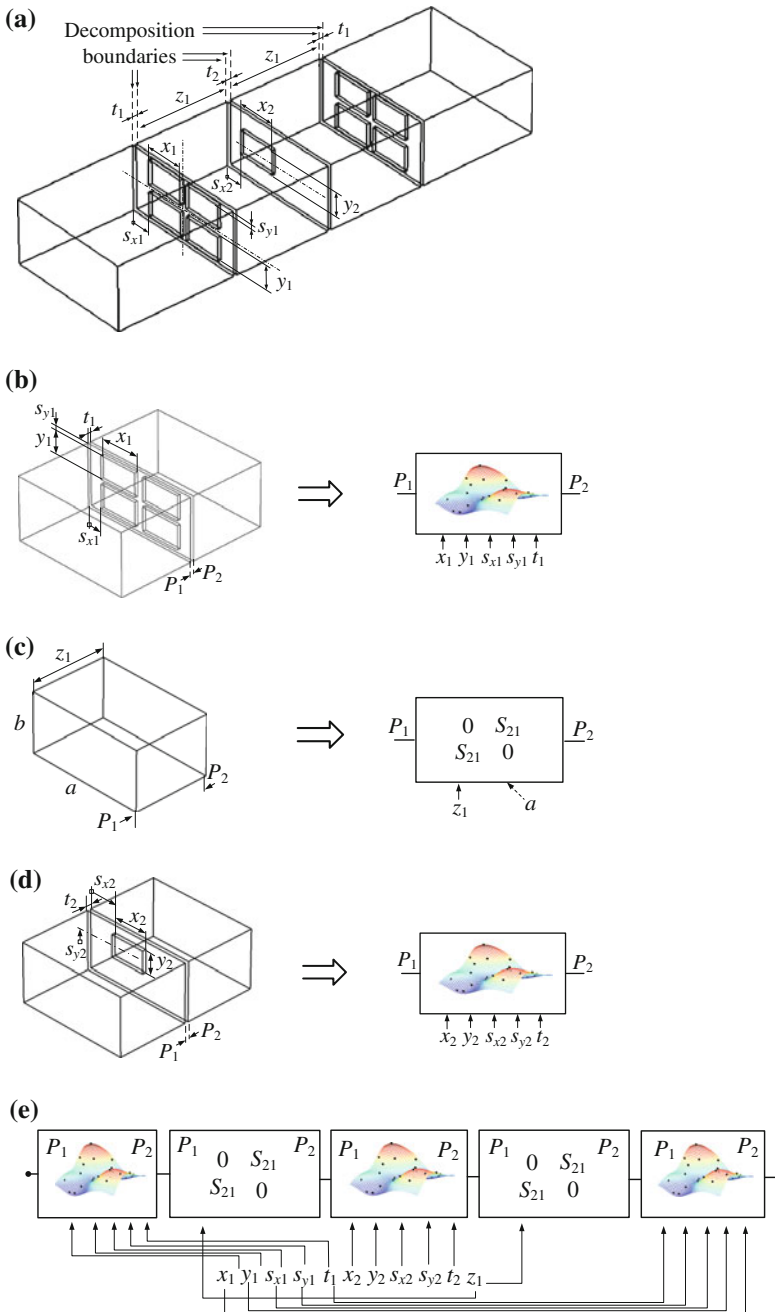
## 2.2 Filter Modeling Using Decomposition

Our modeling technique exploits decomposition and kriging surrogates. It is explained here with a waveguide band-pass filter example [5], shown in Fig. 1a, with 11 design parameters,  $\mathbf{x} = [z_1 \ u_{x1} \ u_{x2} \ u_{y1} \ u_{y2} \ s_{x1} \ s_{x2} \ s_{y1} \ s_{y2} \ z_{t1} \ z_{t2}]^T$ .

The first step of the procedure is to decompose the filter into smaller parts as shown in Fig. 1b through 1d. It is assumed that interactions via the dominant mode of the waveguide have a major contribution on the filter response. This assumption can be verified by comparing the reconstructed response with the response computed for the whole filter at test points of the modeling space. In the case of misalignment of the two responses, e.g., due to interactions via the higher order modes, a correction described in Sect. 2.3 is applied. For every part, a separate RSA model is created using kriging interpolation [8]. Because each subpart of the filter in Fig. 1 has up to five parameters, the number of training points necessary to set up the RSA surrogate is substantially smaller than for the entire filter. Also, the total simulation time of all the subparts, Fig. 1b and d is lower than the simulation time of the entire structure, Fig. 1a. Furthermore, the evaluation of the waveguide section of Fig. 1c described analytically is negligible even for the all simulation bandwidth. The initial RSA surrogate of the entire structure  $\mathbf{R}_{sd}$  is obtained by combining all models of the parts as shown in Fig. 1e by cascading the wave-amplitude transmission matrices.

## 2.3 Space Mapping Correction

The initial surrogate model  $\mathbf{R}_{sd}$  created as described in Sect. 2.2 has to be corrected because EM interactions between the decomposed parts are not taken into account. Here, we use space mapping (SM) [2] as a correction method. SM aims at reducing the misalignment between  $\mathbf{R}_{sd}$  and the high-fidelity model at selected designs. Let  $X_{SM} = \{\mathbf{x}_{sm}^1, \mathbf{x}_{sm}^2, \dots, \mathbf{x}_{sm}^K\}$  be the base set, such that the high-fidelity model response is known at all points  $\mathbf{x}_{sm}^j$ ,  $j = 1, 2, \dots, K$ . The final surrogate  $\mathbf{R}_s$  is defined as



**Fig. 1** Modeling with decomposition and kriging interpolation: **a** example filter and decomposition boundaries; **b** EM models of the first and fifth sections and their kriging models; **c** analytical model of the second and fourth sections (uniform waveguides) and its kriging model; **d** EM model of the third section and its kriging model; **e** complete initial surrogate  $R_{sd}$

$$\mathbf{R}_s(\mathbf{x}) = \mathbf{R}_{sd,f}(\mathbf{B} \cdot \mathbf{x} + \mathbf{c}) \quad (1)$$

where

$$[\mathbf{B}, \mathbf{c}] = \arg \min_{[\beta, \gamma]} \sum_{k=1}^K \|\mathbf{R}_f(\mathbf{x}_{sm}^k) - \mathbf{R}_{sd,f}(\beta \cdot \mathbf{x}_{sm}^k + \gamma)\| \quad (2)$$

and  $\mathbf{R}_{sd,f}$  is a frequency-scaled  $\mathbf{R}_s$ .

Frequency scaling is implemented in the following way. As  $\mathbf{R}_s$  is an evaluation of the filter structure at a discrete set of frequencies (e.g.,  $S$ -parameters vs. frequency), we use the notation  $\mathbf{R}_{sd}(\mathbf{x}) = [\mathbf{R}_{sd}(\mathbf{x}, \omega_1) \dots \mathbf{R}_{sd}(\mathbf{x}, \omega_m)]^T$ . A frequency scaling of the form  $F(\omega) = f_0 + f_1 \cdot \omega$  is utilized here, where the parameters  $f_0$  and  $f_1$  are obtained in the parameter extraction process similar to (2). The  $\mathbf{R}_{sd}$  data at frequencies  $f_0 + f_1 \cdot \omega_k$ ,  $k = 1, \dots, m$ , is obtained by interpolating the data pairs  $\{\omega_j, \mathbf{R}_{sd,j}\}$ .

### 3 Verification Examples

In this section, we apply the procedure described in Sect. 2 to create the surrogate models of two examples of waveguide filters, a coupled-iris filter and a fifth-order Chebyshev bandpass filter.

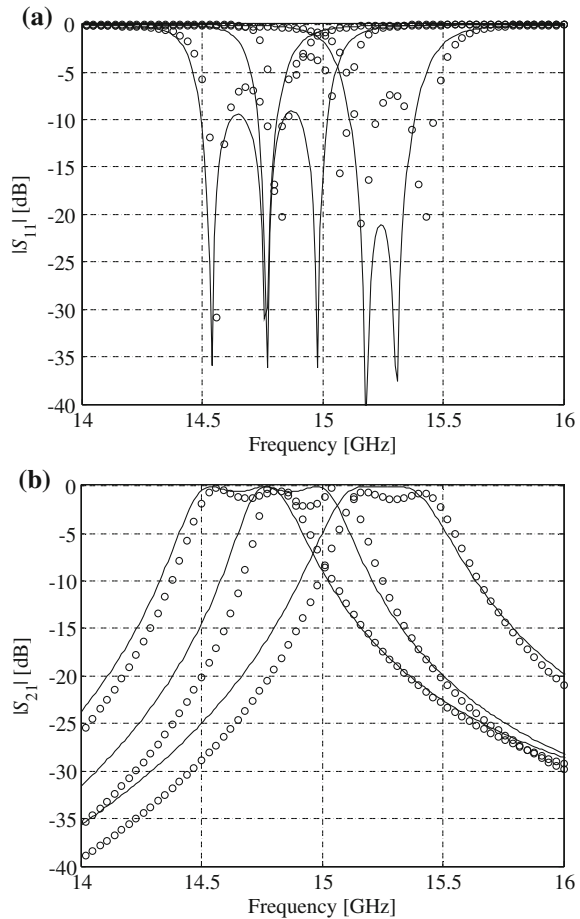
#### 3.1 Coupled Iris Waveguide Filter (Filter 1)

Consider the coupled iris filter [5] shown in Fig. 1a. The modeling/design variables are  $\mathbf{x} = [z_1 \ u_{x1} \ u_{x2} \ u_{y1} \ u_{y2} \ s_{x1} \ s_{x2} \ s_{y1} \ s_{y2} \ z_{t1} \ z_{t2}]^T$ , where  $z_1$  is for the length of the waveguide sections;  $u_{x1} \ u_{x2} \ u_{y1} \ u_{y2}$  are the aperture dimensions;  $s_{x1} \ s_{x2} \ s_{y1} \ s_{y2}$  are the offsets of the apertures, namely, from the walls for the first three as shown in Fig. 1 and the fourth one is the offset from the half height of the central iris;  $z_{t1} \ z_{t2}$  are the iris thicknesses. Apertures of the first and last irises, Fig. 1b are symmetrically located.

All EM models are evaluated using the CST MWS frequency domain solver with adaptive frequency sweep over the simulation bandwidth [4]. Response of the EM model of the first iris (4,600 tetrahedrons on a nominal design after adaptive meshing) is simulated in 18 s. The central section (6,250 tetrahedrons) is simulated in 20 s. The high-fidelity model of the whole filter (60,880 tetrahedrons) is simulated in 5 min.

The RSA surrogate of the filter is created with the procedure described in Sect. 2. Responses of the surrogate at selected test point are shown in Fig. 2. The impact of the higher order modes is visible in Fig. 2 as a discrepancy between the surrogate and the high-fidelity model. Notice that this discrepancy is removed from the surrogate using the SM correction as illustrated in Fig. 3. The total CPU cost of creating the surrogate (only  $\sim 51 \times \mathbf{R}_f$ ) is given with Table 1.

**Fig. 2** Filter 1: responses of the high-fidelity model (—) and the decomposition surrogate (before SM correction) (o) at the selected test points: **a**  $|S_{11}|$ , **b**  $|S_{21}|$



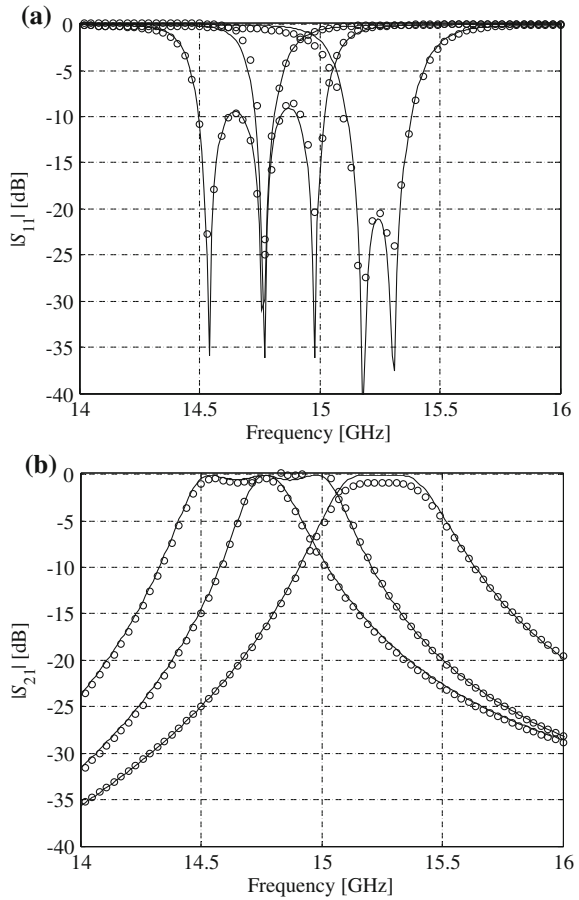
The use of the corrected surrogate for filter optimization is shown with application examples in Sect. 4.

### 3.2 *Fifth-Order Chebyshev Bandpass Waveguide Filter (Filter 2)*

Consider an X-band waveguide filter with nonsymmetrical irises [9] which shown in Fig. 4. The modeling/design variables are  $\mathbf{x} = [z_1 \ z_2 \ z_3 \ d_1 \ d_2 \ d_3 \ t_1 \ t_2 \ t_3]^T$ . Notice the symmetry of the filter so that only one EM model is needed to build the iris surrogates. Responses at selected test points are shown in Fig. 5 for the initial surrogate and in Fig. 6 for the corrected one. The cost of creating the surrogate,  $\sim 20 \times R_f$ , is given in Table 2. The EM models are simulated using the CST



**Fig. 3** Filter 1: responses of the high-fidelity model (—) and the decomposition surrogate (after SM correction) (o) at the selected test points: **a**  $|S_{11}|$ , **b**  $|S_{21}|$



**Table 1** Filter 1: modeling cost

Modeling stage	Number of model evaluations	CPU cost	
		Absolute (min)	Relative to $R_f^a$
RSA model of Section I	$256 \times R_{c1}$	76.8	15.4
RSA model of Section II	$256 \times R_{c1}$	85.3	17.1
Space mapping correction	$19 \times R_f$	95.0	19.0
Total cost	N/A	257.1	51.5

<sup>a</sup> Equivalent number of  $R_f$  evaluations

MWS frequency domain solver [4]. The EM model of the single iris (3,711 tetrahedrons on a nominal design after adaptive meshing) is simulated in 5 s. The high-fidelity model of the entire filter (136,984 tetrahedrons) is simulated in 8 min.

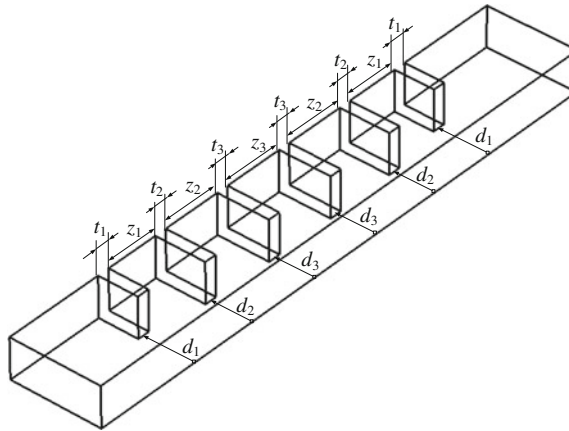


Fig. 4 Fifth order waveguide band-pass filter (Filter 2)

## 4 Application Examples

In this section, we present applications of the decomposition-based surrogate models for filter optimization.

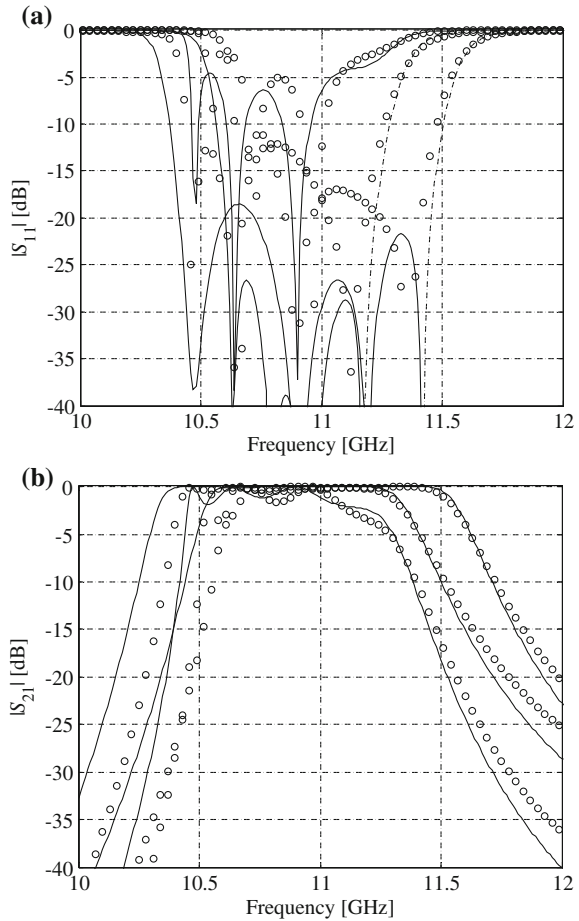
### 4.1 Optimization of Filter 1

The filter is optimized for the following set of design specifications:

- $|S_{11}| \leq -20$  dB for 14.85 GHz to 15.15 GHz,
- $|S_{11}| \geq -1$  dB for 13.0 GHz to 14.4 GHz,
- $|S_{11}| \geq -1$  dB for 15.6 GHz to 16.5 GHz, and
- $|S_{21}| \geq -1$  dB for 14.85 GHz to 15.15 GHz.

Initial design is the center of the region of validity  $\mathbf{x}^{(0)} = [12 \ 4.65 \ 5 \ 1.5 \ 2 \ 0.375 \ 0.6 \ 0.25 \ 0.25]^T$  mm. Final design obtained by optimizing the surrogate model obtained in Sect. 3.1 is  $\mathbf{x}^* = [11.2 \ 5.23 \ 6.106 \ 2.500 \ 1.014 \ 0.298 \ 0.151 \ 0.358 \ 0.234]^T$  mm. The responses of the high-fidelity model at the initial and at the final designs are shown in Fig. 7. The optimized response indicates that the surrogate model is sufficiently accurate to be used for optimizing the filter geometry parameters.

**Fig. 5** Filter 2: responses of the high-fidelity model (—) and the decomposition surrogate (before SM correction) (o) at the selected test points: **a**  $|S_{11}|$ , **b**  $|S_{21}|$



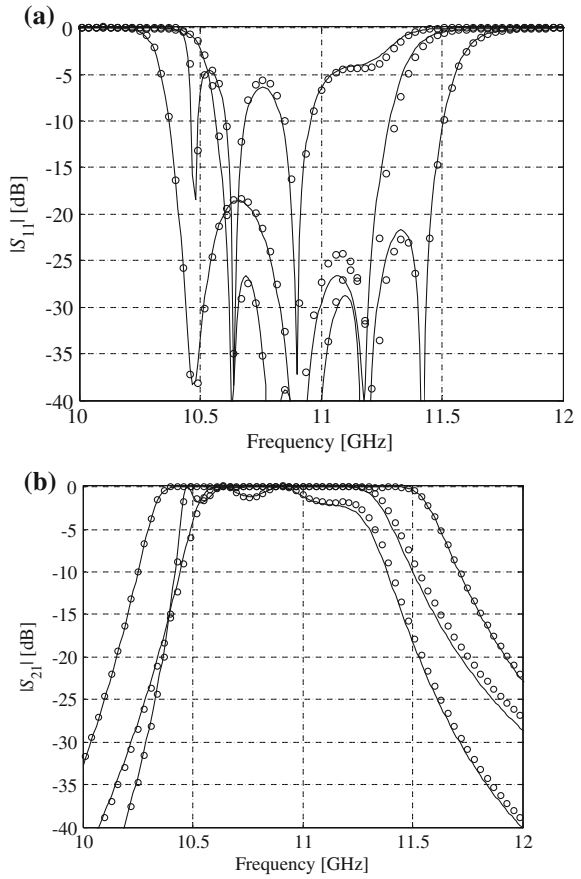
### 4.2 Optimization of Filter 2

The filter is optimized for the following set of design specifications:

- $|S_{11}| \leq -24$  dB for 10.6 GHz to 11.4 GHz,
- $|S_{11}| \geq -1$  dB for 9.5 GHz to 10.3 GHz,
- $|S_{11}| \geq -1$  dB and for 11.7 GHz to 12.5 GHz,
- $|S_{21}| \geq -1$  dB for 10.6 GHz to 11.4 GHz.

Initial design is the center of the region of validity  $\mathbf{x}^{(0)} = [14.25 \ 14.25 \ 14.25 \ 11.5 \ 11.5 \ 11.5 \ 3 \ 3 \ 3]^T$  mm. Final design obtained by optimizing the surrogate is  $\mathbf{x}^* = [12.11 \ 14.36 \ 14.87 \ 14.01 \ 11.55 \ 10.58 \ 1.686 \ 3.036 \ 2.317]^T$  mm. The responses of the high-fidelity model at the initial and at the final designs are shown in Fig. 8.

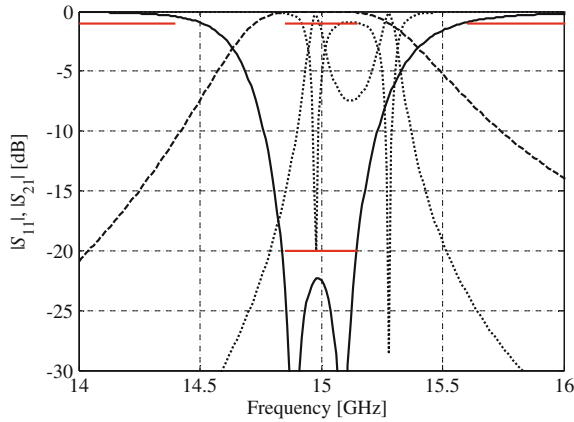
**Fig. 6** Filter 2: responses of the high-fidelity model (—) and the decomposition surrogate (after SM correction) (o) at the selected test points: **a**  $|S_{11}|$ , **b**  $|S_{21}|$



**Table 2** Filter 2: modeling cost

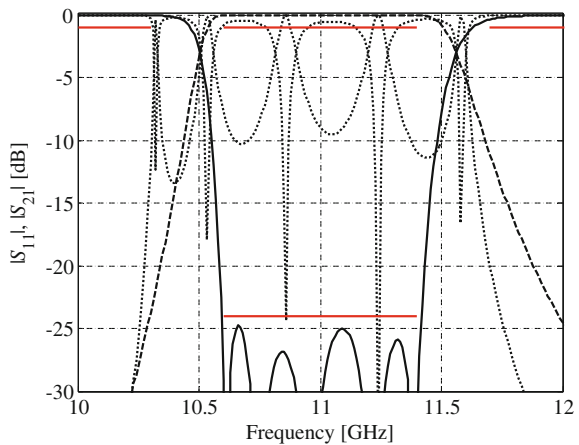
Modeling stage	Number of model evaluations	CPU cost	
		Absolute (min)	Relative to $R_f^a$
RSA model of Section I	$100 \times R_{c1}$	8.3	1.0
Space mapping correction	$19 \times R_f$	152	19.0
Total cost	N/A	160	20

<sup>a</sup> Equivalent number of  $R_f$  evaluations



**Fig. 7** Filter 1 (high-fidelity model responses):  $|S_{11}|$  and  $|S_{21}|$  at the initial design ( $\cdots$ ), and  $|S_{11}|$  ( $\text{---}$ ) and  $|S_{21}|$  ( $\text{- - -}$ ) at the design obtained by optimizing the surrogate model. Design specifications marked using thick horizontal lines

**Fig. 8** Filter 2 (high-fidelity model responses):  $|S_{11}|$  and  $|S_{21}|$  at the initial design ( $\cdots$ ), and  $|S_{11}|$  ( $\text{---}$ ) and  $|S_{21}|$  ( $\text{- - -}$ ) at the design obtained by optimizing the surrogate model. Design specifications marked using thick horizontal lines



## 5 Conclusions

Surrogate modeling of microwave filters using decomposition, response surface approximations and space mapping has been presented. The proposed approach allows for accurate modeling of filter  $S$ -parameters at a low computational cost. As demonstrated through examples, the space-mapped decomposition surrogates can be successfully used for filter design.

**Acknowledgments** The authors thank Computer Simulation Technology AG for making CST Microwave Studio available. This work was supported in part by the Icelandic Centre for Research (RANNIS) Grants 120016021 and 13045051.

## References

1. Bandler JW, Georgieva N, Ismail MA, Rayas-Sánchez JE, Zhang QJ (2001) A generalized space mapping tableau approach to device modelling. *IEEE Trans Microwave Theory Tech* 49:67–79
2. Bandler JW, Cheng QS, Koziel S (2006) Simplified space mapping approach to enhancement of microwave device models. *Int J RF Microwave Comput Aided Eng* 16:518–535
3. Cameron RJ, Kudsia CM, Mansour RR (2007) *Microwave filters for communication systems*. Wiley-Interscience, Hoboken
4. CST Microwave Studio (2012) CST AG, Bad Nauheimer Str. 19, D-64289 Darmstadt, Germany
5. Hauth W, Keller R, Papziner U, Ihmels R, Sieverding T, Arndt F (1993) Rigorous CAD of multipost coupled rectangular waveguide components. In: *Proceedings of 23rd European microwave conference*, Madrid, Spain, pp 611–614
6. Kabir H, Wang Y, Yu M, Zhang QJ (2008) Neural network inverse modeling and applications to microwave filter design. *IEEE Trans Microwave Theory Tech* 56:867–879
7. Koziel S, Leifsson L (2012) Generalized shape-preserving response prediction for accurate modeling of microwave structures. *IET Microwaves Ant Prop* 6:1332–1339
8. Lophaven SN, Nielsen HB, Søndergaard J (2002) DACE: a Matlab kriging toolbox. Technical University of Denmark, Copenhagen
9. Mediavilla A, Tazon A, Pereda JA, Lazaro M, Pantaleon C, Santamaria I (2001) High speed analysis and optimization of waveguide bandpass filter structures using simple neural architecture. *Microwave J* 44(6):86–99
10. Miraftab V, Mansour RR (2006) EM-based microwave circuit design using fuzzy logic techniques. *IEE Proc Microwaves Antennas Propag* 153:495–501
11. Patzelt H, Arndt F (1982) Double-plane steps in rectangular waveguides and their application for transformers, irises, and filters. *IEEE Trans Microwave Theory Tech* 30:771–776
12. Queipo NV, Haftka RT, Shyy W, Goel T, Vaidynathan R, Tucker PK (2005) Surrogate-based analysis and optimization. *Prog Aerosp Sci* 41:1–28
13. Rayas-Sánchez JE, Gutiérrez-Ayala V (2006) EM-based Monte Carlo analysis and yield prediction of microwave circuits using linear-input neural-output space mapping. *IEEE Trans Microwave Theory Tech* 54:4528–4537
14. Shaker GSA, Bakr MH, Sangary N, Safavi-Naeini S (2009) Accelerated antenna design methodology exploiting parameterized Cauchy models. *J Prog Electromagn Res (PIER B)* 18:279–309
15. Xia L, Meng J, Xu R, Yan B, Guo Y (2006) Modeling of 3-D vertical interconnect using support vector machine regression. *IEEE Microwave Wireless Comp Lett* 16:639–641

# Quasi-Monte Carlo and RBF Metamodeling for Quantile Estimation in River Bed Morphodynamics

Tanja Clees, Igor Nikitin, Lialia Nikitina and Sabine Pott

**Abstract** Four generic methods for quantile estimation have been compared: Monte Carlo (MC), Monte Carlo with Harrel-Davis weighting (WMC), quasi-Monte Carlo with Sobol sequence (QMC) and quasi-random splines (QRS). The methods are combined with RBF metamodel and applied to the analysis of morphodynamic—hydrodynamic simulations of the river bed evolution. The following results have been obtained. Harrel-Davis weighting gives a moderate 10–20 % improvement of precision at small number of samples  $N \sim 100$ . Quasi-Monte Carlo methods provide significant improvement of quantile precision, e.g. the number of function evaluations necessary to achieve  $\text{rms} \sim 10^{-4}$  precision is reduced from 1,000,000 for MC to 100,000 for QMC and to 6,000 for QRS. On the other hand, RBF metamodeling of bulky data allows to speed up the computation of one complete result in the considered problem from 45 min (on 32CPU) to 20 s (on 1CPU), providing rapid quantile estimation for the whole set of bulky data.

**Keywords** Stochastic analysis · Response surfaces · Fluid dynamics simulation

## 1 Introduction

The large impact of civil water engineering to nature and society imposes high requirements for the precision of numerical simulations used in planning and evaluation of river engineering concepts. In particular, coupled morphodynamic—

---

T. Clees · I. Nikitin (✉) · L. Nikitina · S. Pott  
Fraunhofer Institute for Algorithms and Scientific Computing, Sankt Augustin, Germany  
e-mail: Igor.Nikitin@scai.fraunhofer.de

T. Clees  
e-mail: Tanja.Clees@scai.fraunhofer.de

L. Nikitina  
e-mail: Lialia.Nikitina@scai.fraunhofer.de

S. Pott  
e-mail: Sabine.Pott@scai.fraunhofer.de

hydrodynamic simulation uses models of river bed evolution possessing uncertain parameters. The sources of uncertainty can be the natural variability, the deficient description of the physical processes in the model and the imprecision of the model parameters. The propagation of these uncertainties to the variance of the model result can be quantified with the aid of stochastic analysis. Precise evaluation of stochastic characteristics normally requires a huge amount of samples, which can be provided by surrogate-based modeling of simulation results. In this paper we present our advances in quantile estimation of morphodynamic simulations of river bed evolution. We use metamodeling of bulky simulation results with radial basis functions (RBF), quasi-Monte Carlo sampling (QMC) and efficient quantile estimator (QE). Four different quantile estimators have been tested. A realistic application case is used to demonstrate the efficiency of the approach.

Numerical simulations define a mapping  $y = f(x): \mathbb{R}^n \rightarrow \mathbb{R}^m$  from an  $n$ -dimensional space of simulation parameters to an  $m$ -dimensional space of simulation results. In morphodynamic simulation the dimensionality of simulation parameters  $x$  is moderate ( $n \sim 10\text{--}30$ ), while simulation results  $y$  are dynamical fields sampled on a large grid, typically containing  $\sim 10^5$  nodes and  $\sim 10$  time steps, resulting in values of  $m \sim 10^6$ . High computational complexity of morphodynamic models restricts the number of simulations available for analysis (typically  $N_{\text{exp}} < 10^3$ ), and this number shall be as small as possible. Metamodeling is an approximation technique allowing efficient representation of these large datasets for the purpose of data analysis, robust optimization and real time visualization. The metamodeling naturally involves in the analysis the uncertainties in optimization variables and other control parameters influencing the simulation.

Metamodeling with radial basis functions (RBF) is a representation of the form

$$f(x) = \sum_{i=1 \dots N_{\text{exp}}} c_i \Phi(|x - x_i|), \quad (1)$$

where  $x_i$  are the points with known function values  $y_i = f(x_i)$ . A suitable choice for the RBF is the multi-quadric function  $\Phi(r) = (b^2 + r^2)^{1/2}$ , which provides non-degeneracy of interpolation matrix  $\Phi_{ij} = \Phi(|x_i - x_j|)$  for all finite datasets of distinct points and all dimensions  $n$  [1]. The result can be written in a form of weighted sum  $f(x) = \sum_j w_j(x) y_j$ , with the weights

$$w_i(x) = \sum_j \Phi_{ij}^{-1} \Phi(|x - x_j|). \quad (2)$$

RBF interpolation can be extended by adding polynomial terms to (1), allowing reconstructing exactly polynomial (including linear) dependencies and generally improving precision of interpolation. Adaptive sampling and hierarchy of metamodels with appropriate transition rules are used for further precision improvement [16]. RBF metamodel is directly applicable for interpolation of high dimensional bulky data, e.g. complete simulation results can be interpolated at a rate linear in the size of data, and even faster in combination with PCA-based dimensional reduction



techniques [2]. The precision can be controlled via the cross-validation procedure: the data point is removed, data are interpolated to this point and compared with the actual value at this point, which for an RBF metamodel leads to a direct formula [9]:

$$\text{err}_i = f_{\text{interpol}}(x_i) - f_{\text{actual}}(x_i) = -c_i / (\Phi^{-1})_{ii}. \tag{3}$$

Metamodeling performed at controlled precision can replace simulation results in computationally intensive procedures, such as optimization and quantile determination. Various quantile estimators are known [5, 7, 13], all they can profit from the combination with metamodels. The other approach has been used in [15], where Monte Carlo method has been combined with Kriging metamodel for setting up additional simulations, used for corrections of the result. All the above methods show theoretical convergence rate  $O(N^{-1/2})$  with increasing number of metamodel evaluations  $N$ , making them hardly applicable for analysis of bulky data. The convergence can be improved by means of quasi-Monte Carlo integration schemes [11]. The behavior of such quantile estimators at high dimension has not been studied yet. In this paper we present our recent results in comparing different quantile estimators and developing their efficient combination with RBF metamodel.

## 2 Quantile Estimation

Quantile estimation is determination of quantiles  $Q_p$  for simulation results:  $P(y < Q_p) = p$ , where  $P$  is the probability measure and  $p$  is a user specified threshold value. For example, the median corresponds to 50 % of distribution, i.e.  $P(y < \text{med}) = 0.5$ ; 68 % of distribution is located in the interval  $[Q_{0.16}, Q_{0.84}]$ , etc. Several approaches for quantile estimation have been developed. In our previous paper [3] we have tested sensitivity based methods and Monte Carlo methods combined with RBF metamodeling. They are briefly described here in Sects. 2.1, 2.2. In Sects. 2.3–2.5 we present further improvements of the methods.

### 2.1 Sensitivity Based

1st order approximation, applicable for a linear mapping  $y = f(x)$  and a normal distribution of simulation parameters, with  $x_0 = \langle x \rangle$  a mean value of  $x$  and  $(\text{cov}_x)_{ij} = \langle dx_i dx_j \rangle$  a covariance matrix of  $x$  and  $dx = x - x_0$ . In this case,  $y$  is also normally distributed with mean value  $y_0 = \langle y \rangle = \text{med}(y) = f(x_0)$  and covariance matrix  $\text{cov}_y = J \text{cov}_x J^T$ , where  $J_{ij} = \partial f_i / \partial x_j$  is the Jacobian matrix of  $f(x)$ . The diagonal part of  $\text{cov}_y$  gives standard deviations  $\sigma_y^2$  directly defining  $Q_p(y)$ , e.g.  $Q_{p(68 \%)} = \langle y \rangle \pm \sigma_y$ ,  $Q_{p(99.7 \%)} = \langle y \rangle \pm 3\sigma_y$ . A finite difference scheme used to

compute the Jacobian matrix of  $f(x)$  requires  $N_{exp} = O(n)$  simulations, e.g.  $2n$  for central difference scheme plus one experiment at  $x_0$ ,  $N_{exp} = 2n + 1$ . The algorithm has a computational complexity  $O(nm)$  and can be implemented efficiently since data from  $N_{exp}$  open data streams can be read and processed simultaneously for writing quantiles to a single output data stream. In this way, memory requirements can be minimized, and parallelization can be done in a straightforward manner.

2nd order approximation applies correction associated with Hesse matrix  $H_{jk}^i = \partial^2 f_i / \partial x_j \partial x_k$ . In particular, the average of this correction  $\langle H_{jk} dx_j dx_k \rangle = \text{Tr}(H \text{cov}_x)$  can be directly used e.g. as robustness measure in optimization, distinguishing sensitive maximum (large  $H$ ) from a wide one (small  $H$ ) being viewed in comparison with the size of scatter (measured by  $\text{cov}_x$ ). Also higher order moments as  $\langle (y_{\max} - y)^2 \rangle$  can be directly computed, as well as cumulative distribution function  $\text{CDF}(f) = P(y < f)$  and all quantiles. The method requires the full Hesse matrix, which needs  $N_{exp} = O(n^2)$  simulations. Practically, the usability of 2nd order method is limited, because strongly non-linear functions can involve higher order terms and because distributions of simulation parameters can strongly deviate from normal ones.

## 2.2 Monte Carlo

Monte Carlo methods can be applied for generic non-linear mapping  $f(x)$  and arbitrary distribution  $\rho(x)$ , to estimate the probability  $P_N(y < Q_p) = \text{num.of } (y_n < Q_p) / N$  for a finite sample  $\{y_1, \dots, y_N\}$ . By the central limit theorem, the error of such estimation  $\text{err}_N = F_N - F$  at large  $N$  is distributed normally with zero mean and standard deviation  $\sigma \sim (F(1-F)/N)^{1/2}$ . The algorithm of quantile estimation performs sorting of  $m$  samples  $\{y_1, \dots, y_N\}$  in ascending order and selecting of  $k$ -th item in every sample with  $k = \lfloor (N-1)p + 1 \rfloor$  as a representative for  $Q_p$ . The algorithm possesses a computational complexity of  $O(mN \log N)$  and can be efficiently implemented using data stream operations similar to the sensitivity based methods.

The standard lower/upper quantiles are defined from a condition that a given portion  $\alpha$  of the distribution is located in the interval  $[Q_{\min}, Q_{\max}]$ :

$$P(y < Q_{\min}) = (1 - \alpha)/2, \quad P(y \geq Q_{\max}) = (1 - \alpha)/2, \quad P(Q_{\min} \leq y < Q_{\max}) = \alpha \tag{4}$$

For interpretation of the results it is convenient to define deviations subtracting from the upper/lower quantiles the median value:

$$dQ_{\min} = \text{med} - Q_{\min}, \quad dQ_{\max} = Q_{\max} - \text{med} \tag{5}$$

The choice of median as a central value has an advantage that in this definition both deviations are positive:  $dQ_{\min} > 0, dQ_{\max} > 0$ . In other definitions, e.g.

subtracting a mean value, for strongly distorted distributions one of two deviations can become negative.

RBF metamodel can represent the mapping  $f(x)$  in this method, reducing the number of required simulations. While the metamodel can be constructed with a moderate number of simulations, e.g.  $N_{exp} \sim 100$ , determination of  $Q_p$  can be done with  $N \gg N_{exp}$ . Algorithmically, it consists in precomputation of RBF weight matrix  $w_{ik} = w_i(x_{sk})$  and its multiplication with the data matrix  $y_{di}$ , comprising  $O(m N N_{exp})$  operations. Effort for the latter usually prevails over  $O(m N \log N)$  operations needed for sorting the interpolated samples. Here  $x_{sk}$ ,  $k = 1..N$  are sample points,  $x_j$ ,  $j = 1..N_{exp}$  are simulation points, and the data matrix  $y_{di}$  stores the whole simulation result of size  $m$  per every experiment  $i = 1..N_{exp}$ ,  $d = 1..m$ .

### 2.3 Weighted Monte Carlo

Harell and Davis [5] propose to use a weighted sum of order statistics to improve the precision of quantile estimation:  $Q_p = \sum_{i=1}^N \omega_i y_{(i)}$ . Here  $y_{(i)}$  denote  $i$ -th point in sample sorted in ascending order and the weights  $\omega_i$  are defined in terms of incomplete beta function  $I_x(a,b)$ :

$$\omega_i = I_{i/N}(p(N + 1), (1 - p)(N + 1)) - I_{(i-1)/N}(p(N + 1), (1 - p)(N + 1)). \quad (6)$$

### 2.4 Quasi-Monte Carlo

Low-discrepancy sequences, such as [14, 4] etc., can be used to improve the precision of quantile estimation. It is well known [10] that numerical integration of smooth functions by Monte Carlo method can profit essentially from the usage of low-discrepancy sequences, providing an approximation error of  $O(N^{-1} \log^n N)$ . Here  $N$  is the number of points in sample and  $n$  is the dimension of parameter space. The problem of quantile determination is related with integration of non-smooth functions. Indeed, quantile can be found by inverting cumulative distribution function, which can be expressed as

$$CDF(y) = \int d^n x \rho(x) \theta(y - f(x)). \quad (7)$$

Here  $\theta(t) = \{1, t \geq 0; 0, t < 0\}$  is Heaviside step function. For discontinuous functions quasi-Monte Carlo has a weaker performance [10]. In particular, the error of integration of theta-function can be estimated as  $O(N^{-1/2-1/(2n)})$ .

## 2.5 Quasi-Random Splines

Here we propose a version of QMC integration method, where  $n-1$  variables are generated as quasi-random points, while the dependence of the function  $f(x)$  on the remaining 1 variable is represented by a cubic spline. In this paper we will concentrate on practically important special case, when the dependence on the remaining variable is monotonous and can be unambiguously inverted. In this case an isosurface  $f(x_1, x_2 \dots x_n) = y$  can be represented as a graph of a function  $x_1 = f^{-1}(y, x_2 \dots x_n)$ . In practically relevant cases this function turns out to be sufficiently smooth and integration over  $x_2 \dots x_n$  variables by QMC method provides the desired convergence  $O(N^{-1} \log^{n-1} N)$ .

In details, the following construction is used. At first, a transformation  $x \rightarrow u$  to new variables possessing uniform probability distribution  $\rho(u) = 1$  is performed. In the case of independent random variables, i.e. when the original distribution is composed as a product  $\rho(x) = \rho_1(x_1) \dots \rho_n(x_n)$ , the cumulative distribution functions of the input variables give the desired transformation:  $u_i = \text{CDF}_i(x_i)$ . The integral is now transformed to

$$\text{CDF}(y) = \int d^n u \theta(y - f(u)). \quad (8)$$

In the case when the function  $f(u)$  is monotonous w.r.t. one variable, say  $u_1$ , a surface  $f(u) = y$  can be re-expressed as a graph of a function  $u_1 = f^{-1}(y, u_2 \dots u_n)$  and integration over  $u_1$  leads to

$$\text{CDF}(y) = \int du_2 \dots du_n f^{-1}(y, u_2 \dots u_n). \quad (9)$$

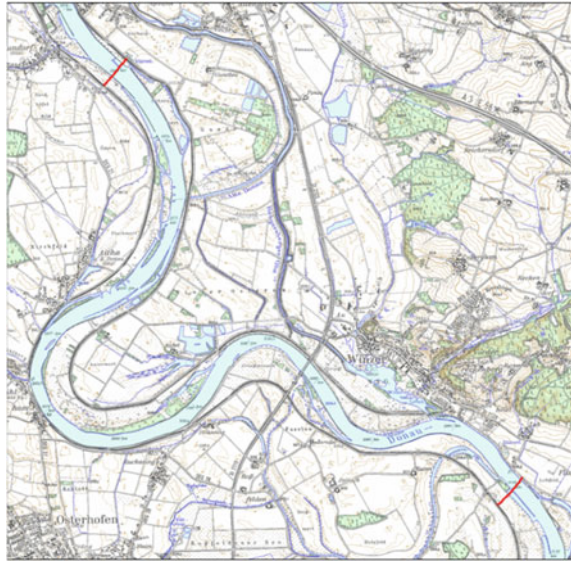
Here the integrand is the function  $f(u_1)$ , inverted at fixed  $u_2 \dots u_n$  and extended as follows:  $u_1 = f^{-1}(y)$  for  $f(0) \leq y \leq f(1)$ , 0 for  $y < f(0)$ , 1 for  $y > f(1)$ . For definiteness, we consider the case of monotonously increasing function  $f(u_1)$ , while for monotonously decreasing function similar formulae can be derived.

The function  $y = f(u_1)$  can be represented by a cubic spline to speed up the inversion, avoiding too many computationally intensive function evaluations. For this purpose  $u_1$  is sampled regularly in  $[0,1]$  with  $K$  points,  $f_k = f(u_{1k})$  are evaluated by call to the metamodel and a sequence  $(u_{1k}, f_k)$  is interpolated with a cubic spline. For the monotonous functions one can revert the pairs and use the sequence  $(f_k, u_{1k})$  directly for spline representation of inverse function  $u_1 = f^{-1}(y)$ .

Finally, one introduces a regular sampling  $y_p$ ,  $p = 1..K$  in  $[y_{\min}, y_{\max}]$ , uses the above constructed splines to find  $f^{-1}(y_p)$  and QMC averaging to obtain  $\text{CDF}(y_p)$ .

The QRS algorithm requires  $NK$  evaluations of the metamodel, where  $N$  is the number of quasi-random samples. This effort  $O(NKN_{\text{exp}})$  prevails over the own computational complexity of the algorithm  $O(NK)$ . Processing of bulky data with

**Fig. 1** River Danube model area in morphodynamic simulation application case

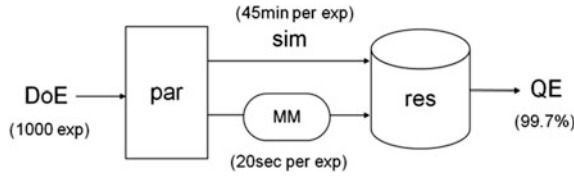


this algorithm can be done straightforwardly by the data stream operations described above.

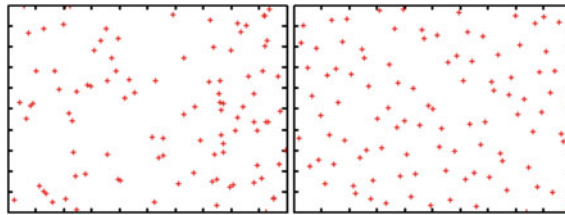
The precision of the algorithm is balanced between the quality of spline interpolation  $O(K^{-4})$  and the quality of QMC integration, where for sufficiently smooth integrand  $O(N^{-1} \log^{n-1} N)$  precision is expected.

### 3 Application Test Case

The methods above have been applied for the analysis of morphodynamic simulations of river bed evolution. A 10 km long stretch of the river Danube including a 270° bend (see Fig. 1) is modeled with the open source software *telemac* ([www.opentelemacs.org](http://www.opentelemacs.org)) including the morphodynamic module *Sisyph* [17] coupled with the hydrodynamic module *Telemac2D* [6]. Nearly 100,000 grid elements were used for the discretization with mean node distances of about 6 m in the river channel and up to 30 m at the flood planes. A synthetic hydrograph of 9 days was simulated including two high flood events. Using 32 processors one simulation run needed about 45 min. For the analysis 13 simulation parameters have been considered to be uncertain, from which five most influencing parameters have been detected. Probability distribution of distorted normal type has been used for modeling (Figs. 2 and 3):



**Fig. 2** Quantile estimation (QE) by means of Monte Carlo method. Either evaluation of simulation runs directly (sim) or of a constructed metamodel (MM) is used for quantile computation. For its construction, the metamodel requires a restricted set (~150) of simulations. Employing a metamodel allows significant speedup over a direct Monte Carlo method



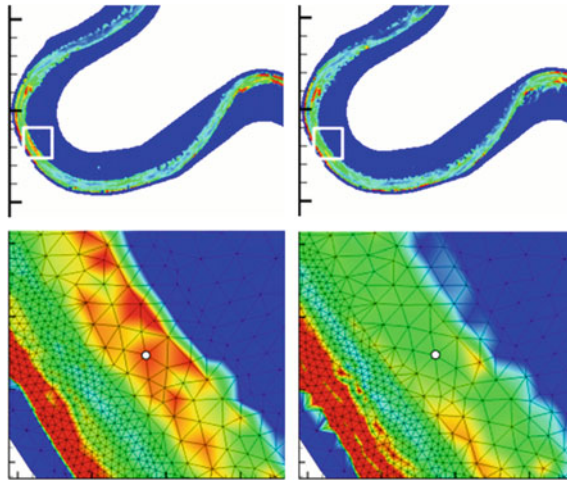
**Fig. 3** Examples of random sequences in dimension  $n = 2$ , standard pseudo-random on the left, Sobol quasi-random on the right. Quasi-random sequences provide more regular, gap-free sampling of design space

$$\rho(x) \sim \exp(-(x - x_0)^2 / 2\sigma_-), \quad x < x_0, \quad \rho(x) \sim \exp(-(x - x_0)^2 / 2\sigma_+), \quad x \geq x_0 \tag{10}$$

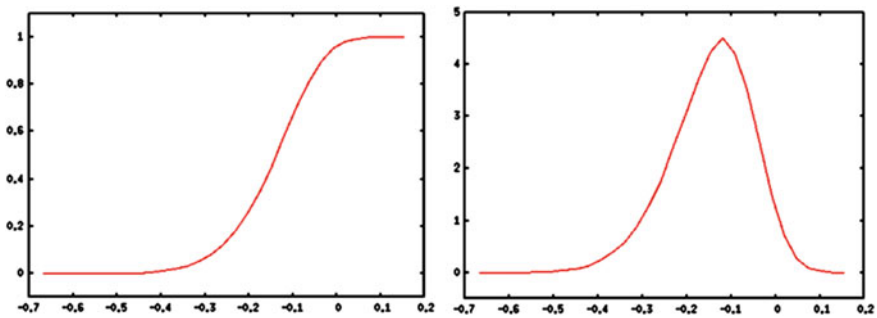
Detailed description of simulation parameters can be found in [3, 8, 12].

For the application of quantile estimators, RBF metamodel has been constructed on 151 experiments for variation of five most influencing parameters ( $n = 5$ ). The typical result of determined quantiles is shown on Fig. 4. The asymmetry of the distribution is clearly visible at the closeups. In particular, for a probe point indicated by a white circle at the closeups:  $dQ_{min} = 0.48$  m,  $dQ_{max} = 0.25$  m. Figure 5 shows the cumulative distribution function (CDF) and probability distribution function (PDF) for the probe point. These plots also show significant asymmetry of the distribution. The asymmetry of the output distribution is related with the asymmetry of the input distribution, as well as with non-linear effects, which deform symmetric input distributions to asymmetric ones. The central limit theorem acts in opposite direction, trying to combine a set of input distributions, independently on their shape, to a normal one.

On Fig. 6 different methods for quantile determination are compared: MC—Monte Carlo based on a standard pseudo-random sequence, WMC—Monte Carlo with Harrel-Davis weights, QMC—quasi-Monte Carlo based on Sobol sequence, WQMC—quasi-Monte Carlo with Harrel-Davis weights. For this comparison the quantiles are estimated for the evolution criterion at the last timestep in the probe point. The estimation on Figs. 6, 7 is done for the median ( $p = 0.5$ ), while the other



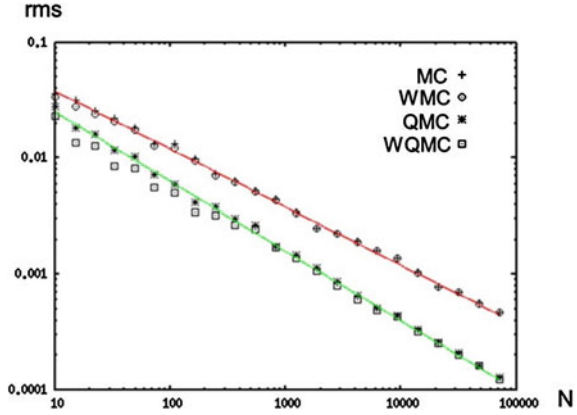
**Fig. 4** Quantiles of the resulting distribution for the river bed evolution, ranging from 0 m (blue) to 0.5 m (red), at the last time step. *Left* dQmin(99.7 %), *right* dQmax(99.7 %), *bottom* corresponding closeups, *white circle* indicates location of probe point



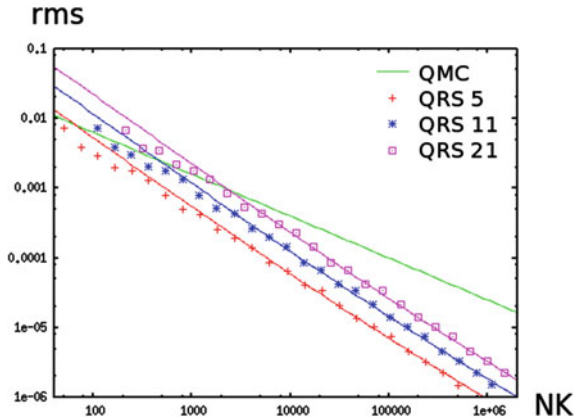
**Fig. 5** Asymmetry of the resulting distribution for the river bed evolution in the probe point. *Left* CDF, *right* PDF

quantiles show similar behavior. The root mean square error (rms) of the determined quantile is computed on 100 trials. The dependence on the sample length  $N$  is tracked in the range  $N = 10 \dots 100,000$ . Harrel-Davis weighting gives an improvement at small  $N$ , particularly near  $N \sim 100$  the improvement is 10 % for MC and 20 % for QMC. For comparison, [5] report improvements of 5–10 % for MC, obtained in the tests on model functions at  $N = 60$  [11] show the improvements of 10–40 % for MC achieved at  $N = 100$  for other model functions. Further, with increasing  $N$ , the effect of Harrel-Davis weighting becomes smaller. The observed asymptotic performance of the methods  $\text{rms}(\text{MC}) \sim N^{-0.5}$ ,  $\text{rms}(\text{QMC}) \sim N^{-0.6}$  coincides with their estimations from Sect. 2:  $\text{rms}(\text{MC}) \sim N^{-1/2}$ ,  $\text{rms}(\text{QMC}) \sim N^{-1/2-1/(2n)}$  at  $n = 5$ .

**Fig. 6** Convergence of different methods for quantile estimation



**Fig. 7** Convergence of different methods for quantile estimation (cont'd)

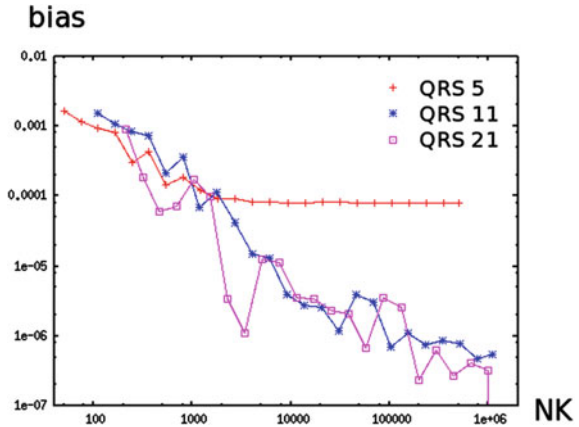


For the application of QRS method we have tested that criterion “evolution” is a monotonously decreasing function of parameter  $x_2$  “coefficient for slope effect”. The test has been done by numerical evaluation of derivative  $\partial f/\partial x_2$  in  $10^7$  randomly selected points.

Figure 7 compares the convergence of QMC and QRS methods. As expected, QRS method provides the asymptotic performance  $rms(QRS) \sim N^{-1} \log^{n-1} N$ , in particular,  $rms(QRS)$  can be well fitted by the dependence  $c_1 N^{-1} \log^4 N + c_2 N^{-1}$ . The tests of QRS method with three different  $K = 5, 11, 21$  have been performed. Smaller  $K$  require less metamodel evaluations, larger  $K$  correspond to better precision of spline interpolation. Figure 8 shows the bias between the median computed by the current and the best method (QRS  $K = 21$   $N = 10^5$ ). Saturation of the dependence corresponds to the quality balance between spline interpolation and QMC integration.



**Fig. 8** Bias of median for QRS method



## 4 Conclusions

Four generic methods for quantile estimation have been presented: Monte Carlo (MC), Monte Carlo with Harrel-Davis weighting (WMC), quasi-Monte Carlo with Sobol sequence (QMC) and quasi-random splines (QRS). The methods are combined with RBF metamodel and applied to the analysis of morphodynamic—hydrodynamic simulations of the river bed evolution. The comparison of the methods shows the following. Harrel-Davis weighting gives a moderate 10–20 % improvement of precision at small number of samples  $N \sim 100$ . Quasi-Monte Carlo methods provide significant improvement of quantile precision, e.g. the number of function evaluations necessary to achieve  $rms \sim 10^{-4}$  precision is reduced from 1,000,000 for MC to 100,000 for QMC and to 6,000 for QRS. On the other hand, RBF metamodeling of bulky data allows to speed up the computation of one complete result in the considered problem from 45 min (on 32CPU) to 20 s (on 1CPU), providing rapid quantile estimation for the whole set of bulky data.

Our further plans include the extension of QRS method for non-monotonous functions. In particular, a reconstruction of probability distribution in terms of moments possesses smooth integrands more suitable for QMC method.

**Acknowledgments** We are grateful to Rebekka Kopmann from Federal Waterways Engineering and Research Institute, Karlsruhe, Germany for providing us with simulation data. We also thank Slawomir Koziel and Leifur Leifsson from Reykjavik University for fruitful discussions at the conference SIMULTECH 2013, Reykjavik, Iceland.

## References

1. Buhmann MD (2003) Radial basis functions: theory and implementations. Cambridge University Press, Cambridge
2. Clees T, Nikitin I, Nikitina L, Thole C-A (2012) Analysis of bulky crash simulation results: deterministic and stochastic aspects. In: Pina N et al (eds) Simulation and modeling methodologies, technologies and applications, AISC 197, Springer 2012, pp 225–237
3. Clees T, Nikitin I, Nikitina L, Kopmann R (2012) Reliability analysis of river bed simulation models. In: Herskovits J (ed) CDROM Proceedings of EngOpt 2012, 3rd international conference on engineering optimization, Rio de Janeiro, Brazil, #267.
4. Halton JH (1964) Algorithm 247: radical-inverse quasi-random point sequence. Commun ACM 7(12):701–702
5. Harell FE, Davis CE (1982) A new distribution-free quantile estimator. Biometrika 69:635–640
6. Hervouet J-M (2007) Hydrodynamics of free surface flows: modelling with the finite element method. Wiley, Chichester
7. Jones MC (1990) The performance of kernel density functions in kernel distribution function estimation. Stat Probabil Lett 9(2):129–132
8. Kopmann R, Schmidt A (2010) Comparison of different reliability analysis methods for a 2D morphodynamic numerical model of River Danube. In: Dittrich A, Koll K, Aberle J, Geisenhainer P (eds) Proceedings of River Flow 2010—international conference on fluvial hydraulics, Braunschweig, 8–10. Sept. 2010, pp 1615–1620.
9. Nikitin I, Nikitina L, Clees T (2012) Nonlinear metamodeling of bulky data and applications in automotive design. In: Günther M et al (eds) Progress in industrial mathematics at ECMI 2010, Mathematics in Industry (17), Springer, 2012, pp 295–301
10. Press WH, Teukolsky S, Vetterling WT, Flannery BP (1992) Numerical Recipes in C, Cambridge University Press 1992, Chap. 7.7
11. Rhein B, Clees T, Ruschitzka M (2012) Robustness measures and numerical approximation of the cumulative density function of response surfaces. Commun Stat B-Simul
12. Riehme J, Kopmann R, Naumann U (2010) Uncertainty quantification based on forward sensitivity analysis in Sisyphé. In: Pereira JCF, Sequeira A (eds) Proceedings of V European conference on computational fluid dynamics ECCOMAS CFD 2010, Lisbon, Portugal, 14–17 June 2010
13. Sfakianakis ME, Verginis DG (2008) A new family of nonparametric quantile estimators. Commun Stat B-Simul 37:337–345
14. Sobol IM (1967) On the distribution of points in a cube and the approximate evaluation of integrals. USSR Comp Math Math Phy 7(4):86–112
15. Sudret B (2012) Meta-models for structural reliability and uncertainty quantification. In: Phoon K et al (eds) Proceedings of 5th Asian-Pacific symposium on structural reliability APSSRA 2012, Singapore, pp 53–76
16. van Bühren G, Clees T, Hornung N, Nikitina L (2012) Aspects of adaptive hierarchical RBF metamodels for optimization. J Comput Methods Sci Eng JCMSE 12(2012), Nr.1–2:5–23
17. Villaret C (2010) Sisyphé 6.0 user manual H-P73-2010-01219-FR, department national hydraulics and environment laboratory, Electricité de France

# Multi-objective Optimization by Using Modified PSO Algorithm for Axial Flow Pump Impeller

H.S. Park and Fu-qing Miao

**Abstract** Axial flow pumps are one type of blade pump with great flux, lower head, highly fluids flow. This type of pump can be used in agriculture, irrigation and massive water project widely. Impellers are the main and highly sensitive part of the pumps which performs the function by transferring energy to the fluid there by increasing pressure and velocity. In axial flow pump design process, in order to get high performance pump, designers usually try to increase the efficiency ( $\eta$ ) and decrease the required net positive suction head (NPSHr) simultaneously. In this paper, multi-objective optimization of axial flow pump based on modified Particle Swarm Optimization (MPSO) is performed. At first, the NPSHr and  $\eta$  in a set of axial flow pump are numerically investigated using commercial software ANSYS with the design variables concerning hub angle  $\beta_h$ , chord angle  $\beta_c$ , cascade solidity of chord  $\sigma_c$ , maximum thickness of blade H. And then, using the Group Method of Data Handling (GMDH) type neural networks in commercial software DTREG, the corresponding polynomial representation for NPSHr and  $\eta$  with respect to design variables are obtained. Finally, multi objective optimization based on modified Particle Swarm Optimization (MPSO) approach is used for Pareto based optimization. The result shows that an optimal solution of the axial flow pump impeller was obtained: NPSHr was decreased by 11.68 % and efficiency was increased by 4.24 % simultaneously. It means by using this method, better performance pump with higher efficiency and lower NPSHr can be got and this optimization is feasible.

**Keywords** Axial flow pump impeller · Multi-objective optimization · GMDH · Modified PSO algorithm (MPSO)

---

H.S. Park (✉) · F.-q. Miao  
Department of Mechanical and Automotive Engineering, University of Ulsan,  
Ulsan, Republic of Korea  
e-mail: phosk@ulsan.ac.kr

F.-q. Miao  
e-mail: miaofuqing120@gmail.com

## 1 Introduction

Axial flow pumps are huge consumers of energy in various industries. This kind of blade pumps has characters of great flux, lower head, and high fluids flow. Because of its broad use in agriculture, irrigation and massive water project, so many researchers pay attention to the area of axial flow pump design [1]. So it is essential to improve the efficiency of such equipment through the design optimization. Optimization of axial flow pump is a multi-objective optimization problem rather than a single objective optimization problem that has been considered so far in the literature. Luo et al. [14] investigated an multi-object optimum design for hydraulic turbine guide vane based on NSGA-II algorithm. They tried to minimize the loss of total pressure and maximize the minimal pressure in guide vane. Finally, for the optimized guide vanes, the loss was reduced, and the cavitations performance was improved. Zhang et al. [20] presented a multi-objective shape optimization of helico-axial multiphase pump impeller based on NSGA-II and ANN. They tried to maximum the pressure rise and pump efficiency. After the optimization using NSGA-II multi-objective genetic algorithm, the five stages of optimized compression cells were manufactured and applied in experiment test. The result shows that the pump pressure rise and the pump efficiency have increased which indicated that the method is feasible.

NPSHr and efficiency in axial flow pumps are important objective functions to be optimized simultaneously in a real world complex multi-objective optimization problem. These objective functions are either obtained from experiments or computed using very timely and high cost CFD approaches, which cannot be used in an iterative optimization task unless a simple but effective meta-model is constructed over the response surface from numerical or experimental data [18]. Therefore, modelling and optimization of the parameters is applied in this paper by using GMDH-type neural networks and modified multi-objective Particle Swarm Optimization in order to minimize the NPSHr and maximize the efficiency of the pumps simultaneously.

In this paper, pump NPSHr and efficiency are numerically investigated using commercial software ANSYS. Then based on the numerical results of geometrical parameters and pump two objective functions, GMDH neural networks are applied in commercial software DTREG in order to obtain polynomial models of NPSHr and efficiency. The obtained simple polynomial models are then used in a Pareto based modified multi-objective PSO optimization approach to find the best possible combinations of NPSHr and efficiency, known as Pareto front.

## 2 Definition of Variables and CFD Simulation of Axial Flow Pump

### 2.1 Definition of Objective Functions

As present in Sect. 1, both the NPSHr and efficiency in axial flow pump are important objective functions to be optimized simultaneously. The efficiency of a axial flow pump is defined by

$$\eta = \frac{P_e}{P} \quad (1)$$

where  $P_e$  is the useful power transferred from pump to the liquid.  $P_e$  can be given by

$$P_e = \rho g Q H \quad (2)$$

where  $P$  is shaft power.

NPSHr means the required net positive suction head which defines the cavitation characteristic of axial flow pump. It is the energy in the liquid required to overcome the friction losses the suction nozzle to the eye of the impeller without causing vaporization [15]. NPSHr varies with design, size, and the operating conditions [2]. It will lead to reduction or stop of the fluid flow and damage the pump with the increasing of the NPSHr. In the handbook of pump design, the NPSHr can be calculated with the following formula:

$$\text{NPSHr} = \frac{P_{\text{in}} - P_{\text{min}}}{\rho g} + \frac{v_0^2}{2g} \quad (3)$$

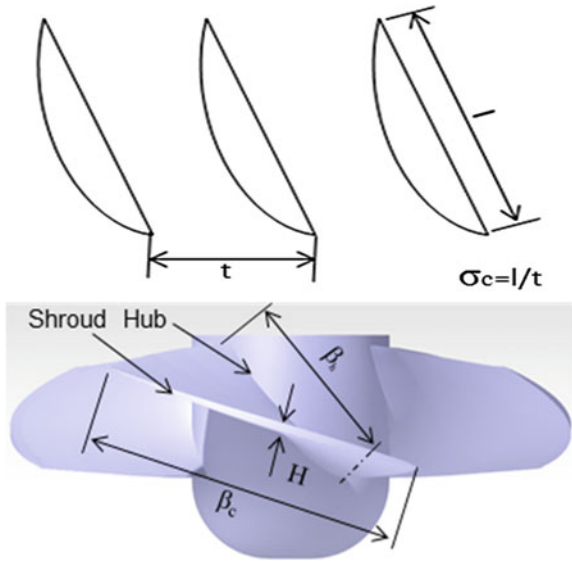
where  $P_{\text{in}}$  is the pressure at inlet,  $P_{\text{min}}$  is the minimum pressure at the whole blade surface which can be obtained from the post processing of ANSYS Fluent software simulation [6].  $\rho$  is the density of fluid and  $v_0$  is the inlet velocity.

### 2.2 Definition of Design Variables

The design variables in this paper are hub angle  $\beta_h$ , chord angle  $\beta_c$ , cascade solidity of chord  $\sigma_c$  ( $\sigma_c = l/t$ ), maximum thickness of blade  $H$ . There are two sections are defined in the blades, one on hub and another on shroud as shown in Fig. 1.

So there are four design variables namely:  $\beta_h$ ,  $\beta_c$ ,  $\sigma_c$  and  $H$ . The various designs can be generated and evaluated in ANSYS Fluent by changing the geometrical independent parameters as shown in Table 1. Consequently, some meta-models can be optimally constructed using the GMDH type neural networks in commercial software DTREG, which will be further used in multi-objective Pareto based design

**Fig. 1** Design variables of impeller blades



**Table 1** Design variables and their range

Design variables	Range from	Range to	Selected values
$\beta_h$	36°	54°	36°, 45°, 54°
$\beta_c$	21°	25°	21°, 23°, 25°
$\sigma_c$	0.75	0.85	0.75, 0.80, 0.85
H	7 mm	11 mm	7 mm, 9 mm, 11 mm

of axial flow pump using modified PSO method. In this way, 81 various CFD analyses have been performed.

### 2.3 Flow Analysis

Because of the incompressible fluid flow, the equations of continuity and balance of momentum are given as

$$\frac{\partial V_i}{\partial x_i} = 0 \tag{4}$$

$$\frac{DV_i}{Dt} = -\frac{1}{\rho} \frac{\partial p}{\partial x_i} + \nu \frac{\partial^2 V_i}{\partial x_j \partial x_j} - \frac{\partial}{\partial x_j} \overline{u_i u_j} \tag{5}$$

**Table 2** Operating conditions in simulation

Parameter	Value
Number of blades	4
Fluid temperature	20°C
Liquid density	998 kg/m <sup>3</sup>
Rotation speed	530 rmp
Mass flow rate	2 m <sup>3</sup> /s
Pump head	2.3 m

The physical model that used in the solver is the Reynlds-Averaged Navier-Stokes equations and the k-ε turbulence models is used [17]. The k-ε equations are given as

$$\frac{Dk}{Dt} = \frac{\partial}{\partial x_j} \left[ \left( C_k \frac{k^2}{\varepsilon} + \nu \right) \frac{\partial k}{\partial x_i} \right] - \overline{u_i u_j} \frac{\partial V_i}{\partial x_j} \quad (6)$$

$$\frac{D\varepsilon}{Dt} = \frac{\partial}{\partial x_j} \left[ \left( C_k \frac{k^2}{\varepsilon} + \nu \right) \frac{\partial \varepsilon}{\partial x_j} \right] - C_{\varepsilon 1} \frac{\varepsilon}{k} \overline{u_i u_j} \frac{\partial V_i}{\partial x_j} - C_{\varepsilon 2} \frac{\varepsilon^2}{k} \quad (7)$$

For the grid generation, tetragonal, hexagonal mesh type was used in software ANSYS Gambit 2.4.6. Pump and around of impeller areas used tetragonal mesh, the other areas were filled with hexagonal mesh [9].

Boundary conditions are as follows: non-slip conditions was applied all around the walls, mass flow rate at the pumps inlet, static boundary condition is used at the outlet. The simulation is continued until the solution converged with a total residual of less than 0.0001 [8].

Operating conditions are shown in the Table 2.

The results of numerical simulation using ANSYS Fluent are shown in Table 3, moreover a path line and total pressure contour of the simulation are shown in Figs. 2 and 3. Total 81 CFD simulation results can be used to build the response surface of both the efficiency and the NPSHr using GMDH type neural networks. Such meta models can be used for the Pareto based multi-objective optimization of the axial flow pump using modified PSO method.

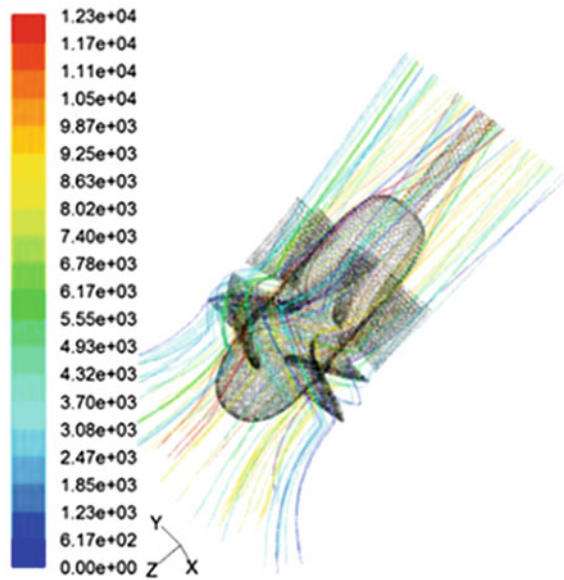
### 3 Meta-models Building Using GMDH-Type Neural Network

Group Method of Data Handling (GMDH) polynomial neural network are self-organizing approach by which gradually complicated models are generated based on the evaluation of their performances on a set of multi input-single output data

**Table 3** Numerical results of CFD simulation

No.	Design Variables				Objective	
	$\beta_h$ (°)	$\beta_c$ (°)	$\sigma_c$ (°)	H (mm)	NPSHr (m)	$\eta$ (%)
1	36	21	0.75	7	6.5	61.2
2	54	23	0.75	9	7	80.6
3	45	25	0.85	11	7.2	79.4
4	54	21	0.8	11	8.2	78.1
5	36	23	0.75	9	6.2	83.4
6	36	25	0.85	7	7.1	86.4
7	45	21	0.8	9	6.7	80.6
8	45	23	0.85	11	8.4	79.1
9	54	25	0.75	7	8.5	83.4
–	–	–	–	–	–	–
80	54	25	0.85	11	6.5	64.3
81	45	23	0.75	9	7.7	74.1

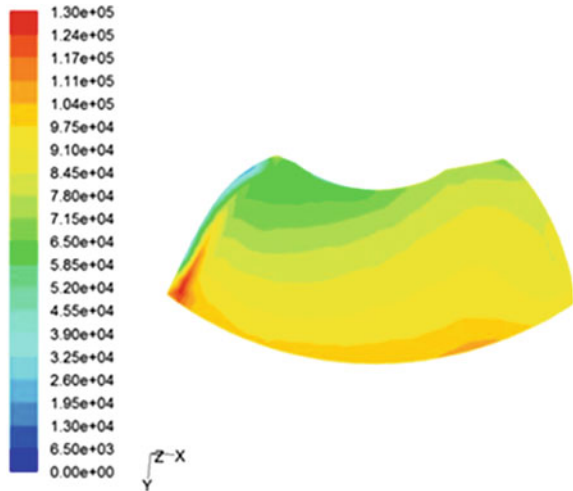
**Fig. 2** Path line of CFD simulation



pairs [12]. GMDH networks were originated in 1968 by Prof Alexey G. Ivakhnenko who was working at that time on a better prediction of fish population in rivers at the Institute of Cybernetics in Kyiv (Ukraine). This algorithm can be used to model complex system without having specific knowledge of the system. The main idea of GMDH is to build an analytical function in a feed forward network based on a



**Fig. 3** Total pressure contour of CFD simulation



quadratic node transfer function [5] whose coefficients are obtained using regression technique.

### 3.1 Structure of a GMDH Network

The meaning of self-organizing means the connections between neurons in the network are not fixed but rather are selected during training to optimize the network. The number of layer in the network also is selected automatically to produce maximum accuracy without over fitting.

As shown in Fig. 4, the first layer (at the left) presents one input for each predictor variable. Every neuron in the second layer draws its inputs from two of the input variables. The neurons in the third layer draw their inputs from two of the neurons in the previous layer and this progress through each layer. The final layer (at the right) draws its two inputs from the previous layer and produces a single value which is the output of the network [16].

The formal definition of identification problem is to find a function  $f$  that can be approximately used instead of the actual one  $f_a$ . In order to predict output  $y$  for a given input vector  $X = (x_1, x_2, x_3, \dots, x_n)$  as close as possible to its actual output  $y_a$ . For the given  $M$  observation of multi input-single output data pairs, there have

$$y_{ai} = f_a(x_{i1}, x_{i2}, x_{i3}, \dots, x_{in}) \quad (i = 1, 2, \dots, M) \tag{8}$$

For any given input vector  $X = (x_1, x_2, x_3, \dots, x_n)$ , there have

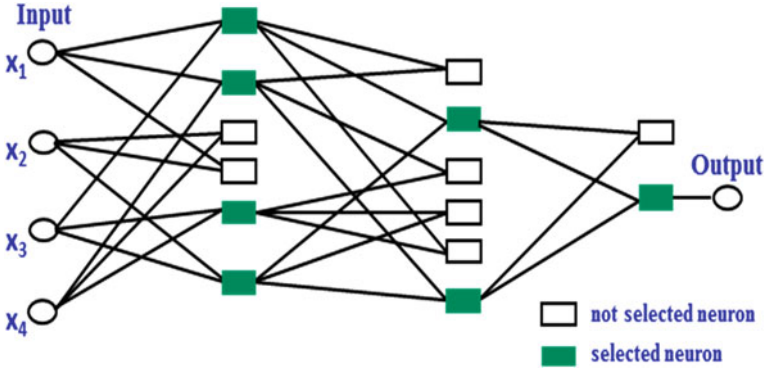


Fig. 4 The structure of a basic GMDH network

$$y_i = f(x_{i1}, x_{i2}, x_{i3}, \dots, x_{in}) \quad (i = 1, 2, \dots, M) \tag{9}$$

In order to determine the GMDH neural network, the square of difference between the actual output and the predicted one is minimized, there have

$$\sum_{i=1}^M [f(x_{i1}, x_{i2}, x_{i3}, \dots, x_{in}) - y_{ai}]^2 \rightarrow \min \tag{10}$$

The most popular base function used in GMDH is the Volterra functional series in the form of

$$y_a = a_0 + \sum_{i=1}^n a_i x_i + \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j + \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n a_{ijk} x_i x_j x_k + \dots \tag{11}$$

where  $y_a$  is the Kolmogorov-Gabor polynomial [5]. It is use complete quadratic polynomials of two variables as transfer functions in the neurons. These polynomials can be represented by the form as show below:

$$y = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2 \tag{12}$$

### 3.2 Meta-models Building in DTREG

The input and output data used in such modelling evolve two different data tables obtained from CFD simulation. Both of the tables consists four variables as inputs that is  $\beta_h, \beta_c, \sigma_c$  and  $H$  (as shown in Fig. 1) and two outputs that is efficiency  $\eta$  and NPSHr. There have 81 patterns which can be used to train and test GMDH neural network. The corresponding polynomial representation for NPSHr is as follows:

$$N(3) = -0.6432 - 0.015\beta_h + 0.17\beta_c + 0.0325\beta_h^2 - 0.0274\beta_c^2 + 2.3e - 6\beta_h\beta_c$$

$$N(1) = 3.524 - 0.134\beta_h + 0.0245\sigma_c + 0.0418\beta_h^2 + 2.056e - 12\sigma_c^2 + 2.01e - 5\beta_h\sigma_c$$

$$N(7) = 5.643 - 0.2134H - 0.02192 + 0.0021H^2 + 0.0004\beta_h^2 + 2.45e - 5H\beta_h$$

$$N(4) = -1.89 + 0.213\beta_c + 0.015\sigma_c - 0.004\beta_c^2 + 1.19e - 11\sigma_c^2 + 1.24 - 5\beta_c\sigma_c$$

$$N(9) = 6.941 - 3.7845 N(3) + 0.5418N(1) + 0.4723 N(3)^2 + 0.0245 N(1)^2 + 0.0475 N(3) N(1)$$

$$N(6) = 6.147 - 2.1873 N(7) - 0.7122 N(4) + 0.218 N(7)^2 + 0.0947 N(4)^2 + 0.4234 N(7) N(4)$$

$$NPSHr = -0.412 - 0.062N(9) + 1.148N(6) + 0.0812 N(9)^2 + 0.0412 N(6)^2 - 0.1271 N(9) N(6)$$

The corresponding polynomial representation for efficiency is as follows:

$$N(4) = -0.5412 - 0.411\beta_h + 2.143\beta_c + 0.016\beta_h^2 - 0.0124\beta_c^2 + 0.00012\beta_h\beta_c$$

$$N(6) = 16.895 + 1.2183H + 0.495\sigma_c - 0.008H^2 - 0.0041\sigma_c^2 + 0.0012H\sigma_c$$

$$N(1) = -15.03 + 1.96\beta_c + 0.6181\sigma_c - 0.023\beta_c^2 - 0.0041\sigma_c^2 + 0.0012\beta_c\sigma_c$$

$$N(7) = 34.107 + 1.254H - 0.49\beta_h - 0.0082H^2 + 0.0092\beta_h^2 + 0.00059H\beta_h$$

$$N(9) = 54.324 - 0.5912N(4) - 1.0126N(6) + 0.00521N(4)^2 + 0.0063N(6)^2 + 0.0191N(4)N(6)$$

$$N(3) = 61.801 - 0.5012 N(1) - 1.0125N(7) + 0.0043 N(1)^2 + 0.00751N(7)^2 + 0.014N(1)N(7)$$

$$\eta = 0.7241 - 3.0174N(9) + 5.0124N(3) - 2.3104 N(9)^2 - 2.5983N(3)^2 + 5.362N(9)N(3)$$

## 4 Apply Multi-objective Optimization by Using Modified PSO Method

Particle Swarm Optimization (PSO) is a computational method that optimize a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality. PSO is a population-based research algorithm. PSO is originally developed by Kennedy and Eberhart [10]. It was first intended for simulation social behaviour, as a stylized representation of the social behaviour of bird flock or fish school. This algorithm originally adopted for balancing weights in neural networks [4], PSO already became a popular global optimizer. There are one study reported in literature that extend PSO to multi-objective problem [3]. A dynamic neighbourhood particle swarm optimization (DNPSO) for multi objective problems was presented [7]. In their study, for each generation, particles of swarm find their new neighbours. The best local particle in the new neighbourhood is choice as gbest for each particle. A modified DNPSO is introduced by find the nearest n particles as the neighbour of the current particle based on the distances between the current particle from others [19].

PSO algorithm is similar to other algorithms based on the principles are accomplished according to the following equations:

$$v_{ij}^{t+1} = wv_{ij}^t + c_1r_1(pb_{ij}^t - x_{ij}^t) + c_2r_2(gb_{ij}^t - x_{ij}^t) \quad (13)$$

$$x_{ij}^{t+1} = x_{ij}^t + v_{ij}^t \quad (14)$$

$$i = 1, 2, \dots, N$$

$$j = 1, 2, \dots, n$$

where  $x$  is the particle current position,  $v$  is the particle current velocity,  $t$  is point of iterations (generations),  $w$  is inertia weight,  $c_1$  and  $c_2$  is acceleration constants,  $r_1$  and  $r_2$  is random values range  $[0,1]$ ,  $pb_{ij}$  is the personal best position of a given particle and  $gb_{ij}$  is the position of the best particle of the entire swarm.

The algorithm developed by Kennedy and Eberhart inspired by the insect swarm (or fish benches or bird flocks) and their coordinated movements. This algorithm pays attention to the information sharing of  $pb_{ij}$  and  $gb_{ij}$  but just considered the experience of  $pb_{ij}$  and  $gb_{ij}$  and ignored the communication of other particles. So an improved particle swarm optimization method (IPSO) was developed as the equations shown below:

$$v_{ij}^{t+1} = wv_{ij}^t + c_1r_1(pb_{ij}^t - x_{ij}^t) + c_2r_2(gb_{ij}^t - x_{ij}^t) + c_3r_3CR \quad (15)$$

$$CR = \begin{cases} pb_{kj}^t - x_{ij}^t & \text{if } ran < cp \\ 0 & \text{other} \end{cases} \quad (16)$$

where  $k$  means the  $k$ th particle and  $k \neq i$ ,  $cp$  is communication probability,  $ran$  is the random values range  $[0,1]$ . The communications between particles were considered and could supply much information in order to search optimal solutions in IPSO. Function test indicated that IPSO increase the ability in the search of optimal solutions [13]. But at the later evolution process, with the disappearance of the swarm diversity, the optimization is easier to be trapped into local optimum. For the disadvantages of IPSO, a modified PSO (MPSO) was developed inspired by bacterial foraging algorithm.

Passino originally proposed the Bacterial Foraging Algorithm [11] in 2002. It is inspired by the abstract and simulate of the food engulf of bacterium in human intestinal canal. There have three steps: chemotactic, reproduce and elimination-dispersal to guide the bacterium to the nutrient-rich area.

Elimination-dispersal happened when bacterium got stimulate from outside and then move to the opposite direction. Chemotactic is the action of bacterium gather to the nutrient-rich area. For example, an *E. coli* bacterium can move in two different ways. It can run (swim for a period of time) or it can tumble, and it alternates between two modes of operation its entire lifetime (i.e., it is rare that the flagella will stop rotating).

As introduced, if the flagella rotate clockwise, each flagellum pulls on the cell, and then the net effect is that each flagellum operates relatively independently of the others. Sometimes the bacterium does not have a set direction of movement and there is little displacement.

The bacterium has behaviour of climbing nutrient gradients. The motion patterns that the bacterium will generate in the presence of chemical attractants and repellants are called chemotactic. For *E. coli*, encounters with serine or aspartate result in attractant responses, whereas repellent responses result from the metal ions Ni and Co, changes in pH, amino acids like leucine, and organic acids like acetate [11]. So for the behaviour of draw on advantages and avoid disadvantages, bacterium can search better food source, increase the chance of surviving and enhance the adaptive capacity to varies environment. If there have harmful stimulus in the process of MPSO, it is easy to get rid of local optimal by applying chemotactic operation. The function as shown below:

$$v_{ij}^{t+1} = wv_{ij}^t - c_1r_1(pbest_{ij}^t - x_{ij}^t) - c_2r_2(gbest_{ij}^t - x_{ij}^t) - c_3r_3CR \quad (17)$$

The flow chart of MPSO as shown in Fig. 5:

The polynomial neural network models obtained in Sect. 3 are now employed in a multi-objective optimization procedure using modified PSO method in order to investigate the optimal performance of the axial flow pump. Two conflicting objectives efficiency  $\eta$  and NPSHr that to be simultaneously optimized with the design variables  $\beta_h$ ,  $\beta_c$ ,  $\sigma_c$  and  $H$ .

Design optimization problem of the objective function and constraints as a function of the following equation:

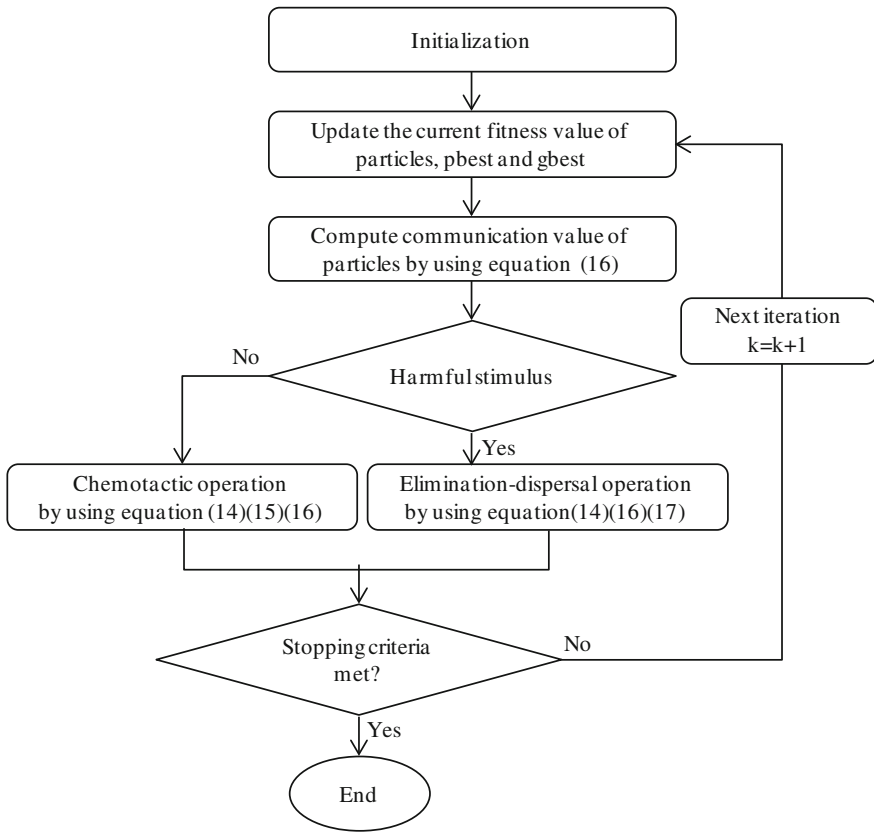
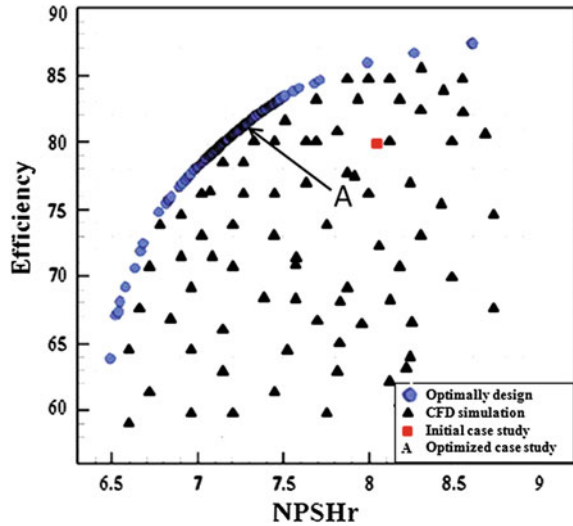


Fig. 5 Flow chart of MPSO

$$\left\{ \begin{array}{l} \text{Maximize} \quad \eta = f_1(\beta_h, \beta_c, \sigma_c, H) \\ \text{Minimize} \quad \text{NPSHr} = f_2(\beta_h, \beta_c, \sigma_c, H) \\ \text{Subject to :} \quad 36^\circ \leq \beta_h \leq 54^\circ \\ \quad \quad \quad 21^\circ \leq \beta_c \leq 25^\circ \\ \quad \quad \quad 0.75 \leq \sigma_c \leq 0.85 \\ \quad \quad \quad 7 \leq H \leq 11 \end{array} \right. \quad (18)$$

The obtained non-dominated optimum solutions based on Pareto front as shown in Fig. 6. These points demonstrate the trade-offs in objective function NPSHr and efficiency. We can find that all the optimum design points in the Pareto front are non-dominated and could be chosen as optimum pump. But choose a better value for any objective function would cause a worse value for another objective. The solutions shown in Fig. 6 are the best possible design points. If we chose any other decision variables, the corresponding values of objectives will be worse.

**Fig. 6** Pareto front of NPSHr and efficiency



**Table 4** Comparison of the results for traditional method and MPSO method

Method	Design variables				Objectives	
	$\beta_h$ (°)	$\beta_c$ (°)	$\sigma_c$	H (mm)	NPSHr (m)	$\eta$ (%)
Traditional	48	21	0.75	10.3	8.13	80.2
MPSO	53.7	22.5	0.81	8.2	7.28	83.6

We use the mapping method to find a trade-off optimum design points compromising both of the objective functions.

$$Mapped \text{ value} = \frac{f - f_{min}}{f_{max} - f_{min}} \tag{19}$$

In the mapping method, the value of objective functions of all non-dominated points are mapped into interval 0 and 1. Using sum of these values for each non dominated point, the trade-off point simply is the one having the minimum sum of those values. Consequently, the optimum design point A is the trade-off points which have been obtained from the mapping method. There have a comparison (as shown in Table 4) of the results for traditional method and MPSO method.

In Table 4, comparison of the obtained best compromise solution and the traditional solution as shown. It is clear that in this comparison, NPSHr was decreased by 11.68 % and efficiency was increased by 4.24 % simultaneously.

## 5 Conclusions

In this paper, multi objective optimization of axial flow pump based on modified Particle Swarm Optimization (PSO) approach are used for Pareto based optimization. Two different polynomial relations for NPSHr and efficiency have been found by GMDH type neural networks using experimentally validated CFD simulations. The obtained polynomial functions were used in an modified PSO optimization process and obtained Pareto front of NPSHr and efficiency. After the mapping method was applied, an optimal solution of the axial flow pump impeller was obtained: NPSHr was decreased by 11.68 % and efficiency was increased by 4.24 % simultaneously. It means this method is feasible and can be applied in impeller design.

## References

1. Alision JL, Jiang C et al (1998) Modern tools for water jet pump design and recent advances in the field. International conference on water jet propulsion, RINA, Amsterdam, 1998
2. Bachus L, Custodio A (2003) Know and Understand Centrifugal Pumps. Elsevier press, Amsterdam, p 14
3. Coello CA, Lchuga MS (2002) MOPSO: a proposal for multiple objective particle swarm optimization. In: Proceedings of the IEEE congress on evolutionary computation, Honolulu, Hawaii, USA, pp 1051–1056
4. Eberhart RC, Dobbins R, Simpson PK, (1996) Computational Intelligence PC Tools. Morgan Kaufmann Publishers
5. Farlow SJ (1984) Self-organizing method in modeling: GMDH type algorithm. Marcel Dekker Inc, New York
6. Guan X (1982) Research on development of high performance axial flow pump impellers. *Pump Technol* 1982(1):25–32
7. Hu X, Eberhart RC, (2002) Multi-objective optimization using dynamic neighborhood particle swarm optimization. In: Proceedings of the IEEE congress on evolutionary computation, Honolulu, Hawaii, USA, pp 1677–1681
8. Huang M, Gao H et al (2008) Numerical simulation and experimental validation of the flow in axial flow pump. *Trans Chinese Soc Agric Mach* 39(8):66–69
9. Jung IS, Jung WH et al (2012) Shape optimization of impeller blades for a bidirectional axial flow pump using polynomial surrogate model. *World Academy of Science, Engineering and Technology*, pp 775–781
10. Kennedy J, Eberhart RC (1995) Particle swarms optimization. In: Proceedings IEEE international conference neural networks, vol IV, Perth, Australia, pp 1942–1948
11. Kevin MP (2002) Biomimicry of bacterial foraging for distributed optimization and control. *IEEE Control Syst Mag* 22(3):52–67
12. Kristinson K, Dumont G (1992) System identification and control using genetic algorithms, *IEEE Trans Syst Manuf Cybern* 22(5):1033–1046
13. Lu X, Ding L (2012) An improved particle swarm optimization based on bacterial foraging to airfoil design. *Acta Aerodynamica Sinica* 30(4):533–538
14. Luo X, Guo P et al (2010) Multi-objective optimum design for hydraulic turbine guide vane based on NSGA-II algorithm. *J Drainage Irrig Mach Eng* 28(5):369–373
15. Nourbakhsh A (2006) Pump and pumping. University of Tehran Press, Tehran, p 337
16. Phillip HS (2003) DTREG Predictive Modeling Software. DTREG Inc



17. Riasi A, Nourbakhsh A, Raisee M (2009) Unsteady turbulent pipe flow due to water hammer using  $k-\epsilon$  turbulence model. *J Hydraul Res* 47(4):429–437
18. Safikhani H, Khalkhali A et al (2011) Pareto based multi-objective optimization of centrifugal pumps using CFD, neural networks and genetic algorithms. *Eng Appl Comput Fluid Mech* 5 (1):37–48
19. Toufik A, Mekki A (2012) Modified PSO algorithm for multi-objective optimization of the cutting parameters. *computer aided engineering, Prod Eng Res Devel* (2012) 6:569–576
20. Zhang J, Zhu H et al (2011) Multi-objective shape optimization of helicon-axial multiphase pump impeller based on NSGA-II and ANN. *Energy Convers Manage*, 538–546

# A Public Health Model for Simulating Policy Interventions to Reduce Nonmedical Opioid Use

Alexandra Nielsen, Wayne Wakeland and Amanuel Zimam

**Abstract** Reports on the development of a system dynamics simulation model of initiation and nonmedical use of pharmaceutical opioids in the US. The study relies on historical trend data as well as expert panel recommendations that inform model parameters and structure. The model was used to assess simulated public health interventions for reducing initiation and nonmedical use of opioids. Results indicate that interventions which reduce the likelihood of informal sharing of leftover medicine could significantly reduce initiation and nonmedical use. Less effective are supply restrictions, such as drug take-back days, and interventions aimed at reducing the likelihood that nonusers would decide to initiate nonmedical use based on their interactions with nonmedical users. We conclude that system dynamics is an effective approach for evaluating potential interventions to this complex system where the use of pharmaceutical opioids to treat pain can lead to unintended distal outcomes in the public sphere.

**Keywords** Substance abuse · System dynamics · Dynamic modelling · Public health · Epidemic modelling

## 1 Introduction

A dramatic rise in the nonmedical use of pharmaceutical opioids in the late 1990s and early 2000s created a substantial public health challenge for the United States [6]. Despite implementation of public health policies and regulations [11], the high

---

A. Nielsen (✉) · W. Wakeland · A. Zimam  
Systems Science Graduate Program, Portland State University, Portland, OR, USA  
e-mail: alexan3@pdx.edu

W. Wakeland  
e-mail: wakeland@pdx.edu

A. Zimam  
e-mail: azimam@pdx.edu

level and increasing prevalence of negative outcomes such as fatal and non-fatal overdoses remains largely unabated [5]. Resistance to policy interventions has been observed and likely stems from the complexity of the pharmaceutical opioid system, including multiple interactions between prescribers, pharmacists, persons obtaining opioids for medical or nonmedical use, opioid traffickers, and public health advocates. The resulting chains of cause and effect often result in feedback loops that diminish or even reverse well-intentioned interventions.

This paper presents a system dynamics model of the complex system surrounding the initiation and nonmedical use of pharmaceutical opioids in the United States. In addition to accounting for historical trends in the initiation and prevalence of nonmedical use, the model aims to clarify the processes by which nonmedical users acquire pharmaceutical opioids via friends and relatives (Substance Abuse and Mental Health Services Administration, SAMHSA) [27], and to facilitate the comparison of policy alternatives to help ameliorate this complex public health issue.

## ***1.1 Background and Problem***

The number of overdose deaths involving opioids tripled between 1999 and 2006 in the US, rising to 14,800 in 2008 [32]. As evidenced by the high fraction of opioid overdose decedents without prescriptions [16], nonmedical use of pharmaceutical opioids plays a significant role in the prevalence of overdose deaths. Estimates from the National Survey on Drug Use and Health (NSDUH) suggest that the rate of initiation of nonmedical use of pain relievers increased almost three-fold from 1995 to 2003 (SAMHSA) [26] and has continued at high rates. In 2010, an estimated 2.4 million individuals initiated nonmedical use of pain relievers [27] and 5.1 million individuals used opioids nonmedically within the month prior to the survey [27].

Diversion of opioids from prescription holders is a major source of supply for nonmedical use. Around 70 % of respondents to the 2010 NSDUH indicated that they received opioids from friends or relatives. And among those who received the drugs for free, 80 % identified their source as originally acquiring the drugs from a single doctor. Leftover opioid prescriptions are likely involved in much of this diversion [6]. A study of post-surgical patients discharged from a urology group practice found that 42 % of opioids prescribed were unconsumed, and that 67 % of patients had surplus opioids. Further, 91 % of patients with leftover medicine kept it in their homes rather than disposing of it [2]. Survey of dentists in West Virginia found 88 % prescribed opioids and 36 % of the dentists expected patients to have leftovers [29]. Similarly, in a nationally representative study, dentists were the major prescribers of opioids for patients aged 10–19 [30]. However, primary care physicians prescribed most of the opioids during a given year. Recent analysis of a 10 year trend indicated increases in opioid overdose deaths, treatment admissions, opioid sales [22] and prescriptions [12]. With age, opioid abuse and dependence begin to emerge [21] contributing to the escalation of the epidemic. A recent

National Drug Take-Back event in Madison, Wisconsin recovered approximately 100,000 opioid dosage units in one day [13]. These studies suggest that there is a large reservoir of unused opioids stored in homes, and the high fraction of individuals receiving drugs for free from friends and family is likely to be strongly correlated with the size of this reservoir.

Several studies have suggested that heroin users are more likely to have first used opioids nonmedically before transitioning to heroin use [18, 23] leading to potentially more severe consequences as a result of increased use of illicit drugs. Unless appropriate policies are implemented to reduce availability of opioids and willingness of prescription holders from freely providing opioids, the current upward trend of overdose incidences and deaths are forecasted to continue. System dynamics modelling offers an opportunity to study complex systems such as the opioid epidemic and test policy intervention scenarios to curb the initiation and nonmedical use of prescription opioids.

## 2 Method

### 2.1 A System Dynamics Simulation Model

The system dynamics modelling approach uses a set of differential equations to simulate system behaviour over time. This approach provides a framework in which to capture the underlying processes involved in a system and the feedback loops that generate its behaviour. When applied to public health problems, system dynamics modelling allows for the simulation of intervention alternatives in order to provide policymakers with a tool to assess interventions for magnitude of impact and potential for unintended consequences—information that is not available from research focused on individual aspects of a system [25].

In the current research, a system dynamics model complements and leverages results from existing research, primarily historical trends available from NSDUH [27], and holds promise for the simulation of intervention alternatives. Figure 1 provides a high level picture of the current model, which features one of the main pathways by which people may initiate nonmedical use of pharmaceutical opioids and transition from casual usage based on free access to paying for drugs through illicit channels. The ease of obtaining drugs for free depends in the model on the amount of leftover and undisposed pharmaceutical opioids that are stored in homes (“medicine cabinets”).

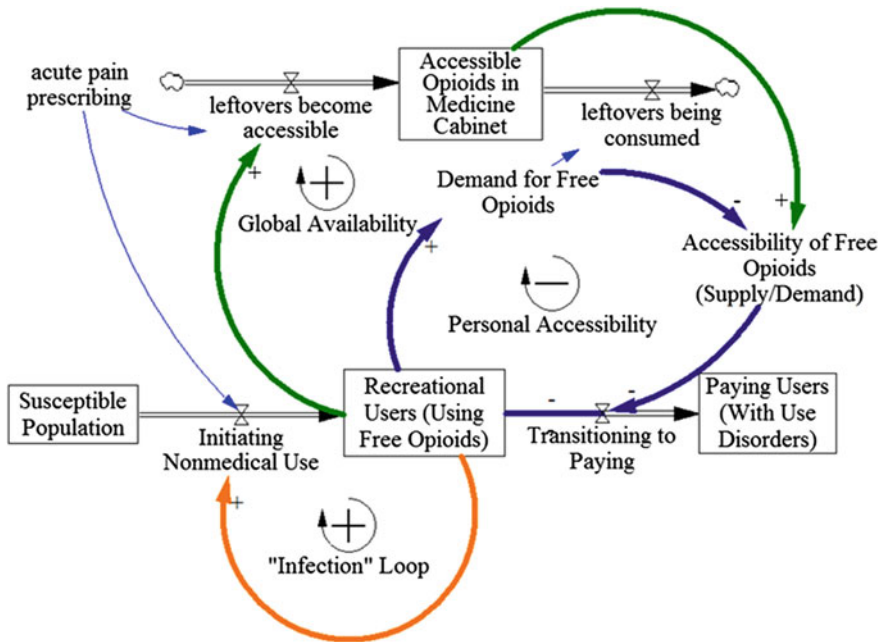


Fig. 1 High level diagram of the overall model structure

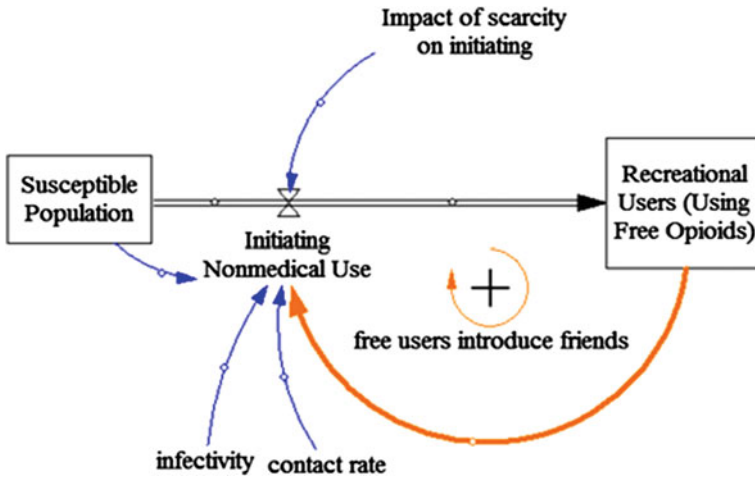
## 2.2 The Dynamics of Opioid Nonmedical Use Initiation

The rate of prescribed opioids for acute pain treatment is shown in the upper left corner of Fig. 1, which serves as a key exogenous input to the model. The model assumes that leftover prescriptions from acute pain conditions are more likely to constitute free sharing than prescriptions for chronic pain diagnoses.

The lower part of Fig. 1 depicts the progression of people from initiating non-medical use to paying for drugs, which implies the development of a use disorder (such as opioid abuse or addiction) and other increasingly risky behaviours.

Figure 2 focuses on the recruitment mechanism whereby recreational users, who acquire opioids for free from friends or family, influence their peers to initiate nonmedical opioid use. This recruitment is modelled as an infectious disease process using the SIR (susceptible, infected, recovered) epidemic modelling framework.

In SIR disease models, an infected party will make contact with susceptible individuals based on a contact rate. The infectivity of the disease determines whether contact results in infection of the susceptible. When the number of infected individuals becomes large, a susceptible is likely to have multiple contacts with infected people and infection becomes more likely. Thus the infected population becomes larger causing infection to spread more quickly, resulting in a disease epidemic.

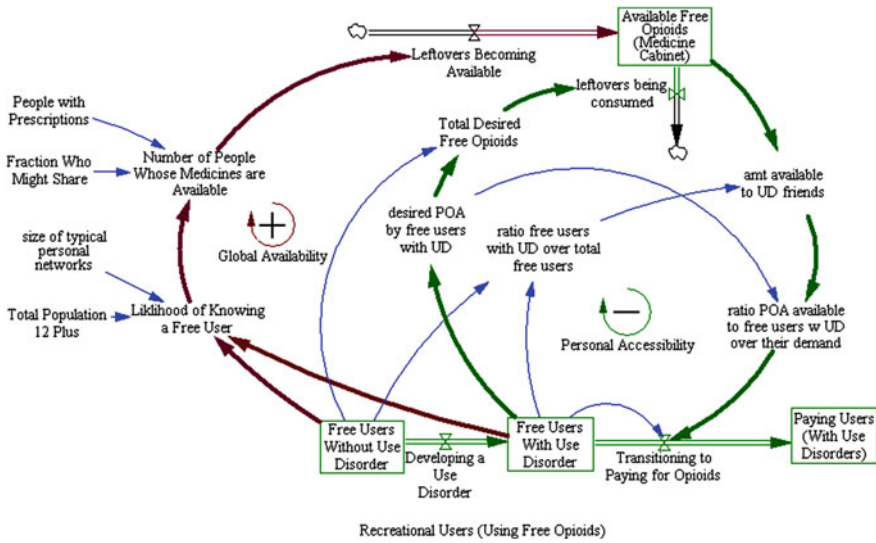


**Fig. 2** Additional details regarding the initiation feedback loop

While nonmedical pharmaceutical opioid use is not an infection per se, the SIR epidemic model is a compelling framework to explain initiation. Susceptibles in this case are people who have never used opioids nonmedically, and infected individuals are those who use opioids recreationally. When individuals in these two populations make contact, the idea of using opioids recreationally can spread to the susceptible who then initiate opioid use based on the “infectivity” of the idea. The infection of a susceptible by an infected individual could be active, as when a peer is pressured or persuaded to use drugs by other peers, or passive in which a susceptible observes drug use behaviours in peers, parents, or through the media and copies those behaviors [8, 1]. Increase in the number of nonmedical users increases the rate of initiation resulting in a positive feedback loop or vicious spiral. In order to initiate opioid use, a susceptible must have both the desire to use opioids and access to them. In this model, the initiation rate is mediated by the availability of free opioids, which is related to the ratio of supply and demand.

In the classic SIR disease model, people who recover from infection do not spread the disease, nor are they susceptible to reinfection. In this model nonmedical users are organized into three groups, recreational users with and without a use disorder (heretofore shown as the aggregated recreational user group), and people with use disorder who use more than they can obtain for free and have to pay for some of their drugs. Individuals in the third group are assumed not to participate in the recruitment process [33]. These users may no longer be peers of susceptibles as they become increasingly socially isolated, and instead of sending positive messages about drug use behaviour that susceptibles want to mimic, they may send negative messages.

Figure 3 describes the relationship between the free supply of opioids in medicine cabinets, and the progression of users from casual (free) use, to development

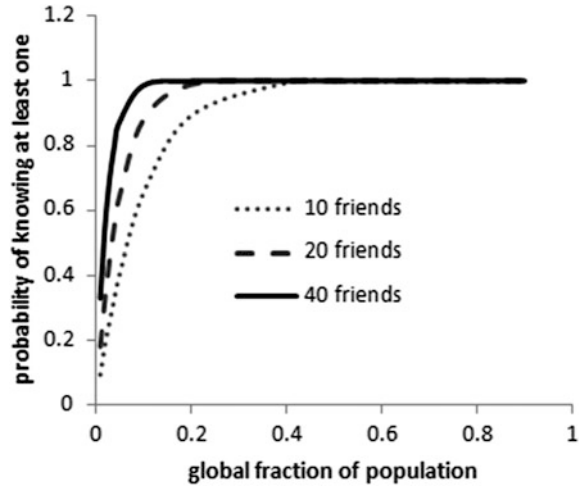


**Fig. 3** Impact of supply on user progression to paying for drugs

of a use disorder, to paid use. The outermost arrows represent the global dynamics of opioid availability: Much leftover medicine is not accessible because prescription holders may not have any desire to use it nonmedically or know anyone who does. However, as the population of recreational (or “free”) users increases (the stocks at the bottom of Fig. 3), the likelihood that prescription holders with leftover opioids know at least one person who would want to use them nonmedically increases. This likelihood depends on a binomial probability calculation based on the individual’s personal network size and the fraction of the total population that is seeking opioids (see Fig. 4). When the network size is fairly large, the likelihood that a prescription holder knows someone who is seeking free opioids is high even when a relatively small fraction of the total population has prescriptions and will share them, as in the solid plot in Fig. 4. However, even when the network size is small, the likelihood of knowing at least one person who will share is still greater than this global fraction. The probability curve is always bowed outward. Therefore, it doesn’t matter how many drug seekers the prescription holder knows; if she knows at least one, then her leftovers might be potential supply.

Some fraction of the prescription holders who know at least one person who is seeking free opioids for nonmedical use will actually choose to share them. Their leftover medicine then flows into the “Available Free Opioids” supply, here referred to as the “Medicine Cabinet.” Thus, increases in initiation lead to *increased* accessibility of leftover prescription opioids, which tends to increase the population of casual users because fewer of them transition to paid use due to loss of access to free opioids, constituting a second positive feedback loop.

**Fig. 4** Binomial probability calculation of likelihood parameters with 10, 20 and 40 friends. The likelihood that a prescription holder knows someone seeking free opioids depends on the fraction of the total population who are seeking free supply



On the other hand, the inner arrows shown in Fig. 3 also show the personal or local dynamics of opioid availability, which operate differently than the global dynamics. Frequent use of pharmaceutical opioids can lead to the development of opioid use disorders [10] and, with them, consumption levels that cannot be sustained by free leftover medicines prescribed to those in one's personal network. When the demand for opioids exceeds what these individuals can access through their personal contacts, they may begin purchasing opioids through the black market. This advancement to paid use is assumed to be associated with the development of an opioid use disorder and with a higher risk of adverse outcomes. Therefore, although an increase in the number of casual users "loosens up" opioid supply by increasing access to leftover prescription holders at the global scale, it also results in the exhaustion of sources of supply at the local scale. Because of these local dynamics, an increase in the population of casual users leads to decreases in accessibility, as represented by a balancing (negative) feedback loop.

### 2.3 Estimation of Key Parameter Values

A time series of the number of people receiving prescription opioid analgesics for acute pain was obtained from SDI Health data for the years 2002, 2006 and 2010. Script data from Governale [15] was used to estimate number of people receiving prescription opioid analgesics for acute pain in 1998. Using the 1998–2010 series, we extrapolated to obtain the 1995 and 2014 data points. Estimates for typical leftover medications were obtained from dentists' expectations of unused percentage of prescribed medications [29]. Regarding the number of doses taken per nonmedical usage event, Katz et al. [19] projected the number to range from one to



as many as five. In the present model, the average is 1.5 for recreational users and 2 for those with use disorder.

Perception of risk is believed to affect the probability of experimenting with prescription opioid analgesics (POA) [28]. The fraction of youth perceiving little risk in using POA was found to be 28 % in a recent Monitoring the Future Survey [17]. Since a person might be susceptible even if they perceive moderate risk, the model uses 0.4 as the annual fraction of population growth added to the susceptible stock.


Where empirical support was not available to support parameter values, estimates under the guidance of expert panel members were adopted. The model was calibrated to fit three time series from the NSDUH for the years 1995–2005: total past year nonmedical opioid users, and total past year initiates of opioid use, total past year opioid users who meet the criteria for opioid abuse or addiction. A complete model and exact parameter values are available from the authors upon request.

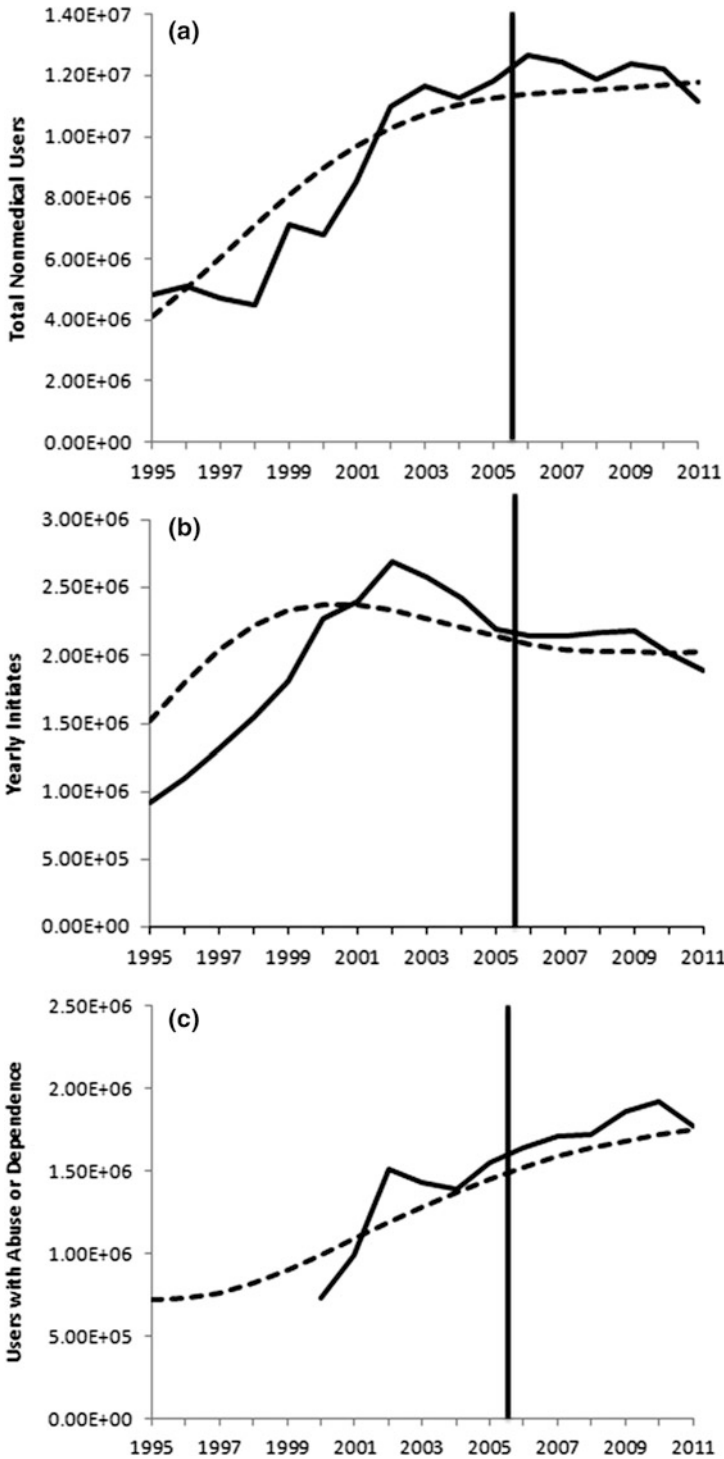
## 2.4 Model Testing

To build confidence in the model, outputs were tested for fit against 2006–2011 data. Results of calibration and tests of fit are shown in Fig. 5. Degree of fit to 2006–2011 data was calculated using mean absolute percent error (MAPE) (see [24] for a discussion of validation tests for SD models). MAPE for new initiates, total nonmedical users, and users with UD were 4.8, 6.2, and 6.8 %, respectively. These results, along with having passed face validity tests with expert panel members, led the team to conclude that the model was sufficiently plausible to begin exploratory policy analysis, on a relative comparison basis.

## 3 Model Results

Logic representing three initiation reduction scenarios was developed to illustrate the potential for evaluating policies for reducing the initiation and persistence of the nonmedical use of pharmaceutical opioids. The baseline run begins in 1995 and continues until 2011, and the policy scenarios are simulated to have been implemented in 2005 and analysed in terms of their relative impacts on the simulated

**Fig. 5** Model outputs (*dashed*) versus data (*solid*). Data prior to 2006 was used for calibration. 2006 onward used for tests of fit and confidence building. **a** Total past year nonmedical users (2006–2011 MAPE-6.21 %). **b** Total past year initiates of nonmedical use (2006–2011 MAPE = 4.82 %). **c** Total past year users who meet the criteria for abuse or addiction. Data prior to 2000 could not be obtained (2006–2011 MAPE 6.77 %) 



number of nonmedical opioid users for the six-year period from 2005 to 2011. The scenarios presented here were implemented as a step change in a specific model parameter.

### ***3.1 Prescription Drug Take-Back Initiative***

The first scenario, a prescription drug take-back initiative, simulates an expansion of the DEA's National Prescription Drug Take-Back Day program [9] to collect unneeded medications by asking individuals to bring leftover prescriptions to a disposal location. Disposal records from one Take-Back Day in Madison Wisconsin suggest that as many as 100,000 opioid dosage units can be collected in a major city on one day [13]. In the current policy test, the national take-back program is simulated as the removal of fifty million dosage units from the "medicine cabinet" supply of available opioids each year, starting in 2005. This amount is highly speculative, but could be possible if, for example, 25 states were to implement Take-Back Days in two major cities, approximately monthly, with the degree of success that was witnessed in the recent Madison Take-Back Day.

### ***3.2 Reducing Initiation Through Drug-Resistance Strategies***

The second simulated scenario features a reduction in the "infectivity" of opioids as a desirable substance for nonmedical use. Some interventions, such as "Keepin' it R.E.A.L." [14], may deter or delay initiation of nonmedical opioid use, even if opioids are freely available and recommended by peers, through teaching culturally specific drug-resistance strategies. In the current model, infectivity was reduced by 25 % in 2005, so that uninitiated individuals were 25 % less likely to initiate nonmedical use even if exposed to the idea.

### ***3.3 Reducing Willingness to Share Opioids***

The third simulated scenario features a reduction in the fraction of individuals who are willing to share their opioids with others for nonmedical use. Individuals with leftover prescriptions might also be educated about the risks involved in sharing medications, or might be encouraged to adopt safety features, such as locked medicine cabinets. This intervention is currently modelled as a 25 % reduction in the fraction of individuals who are willing to share their leftover opioid prescriptions with others, starting in 2005.

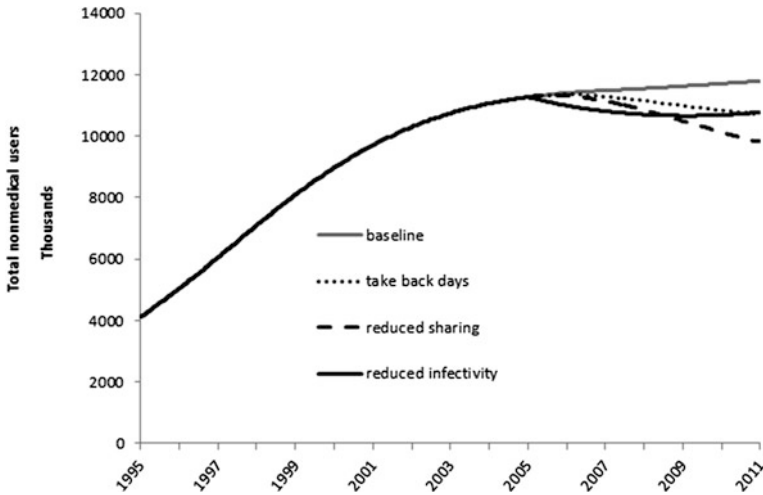


Fig. 6 Impact of interventions on the number of nonmedical users

### 3.4 Policy Scenario Comparison

Figure 6 shows a comparison of the three interventions in terms of their impact on total number of nonmedical users over time. This analysis suggests that behavioral interventions may be able to significantly reduce the number of people who use opioids nonmedically, with a change in the willingness of prescription holders to “share” their unused medicine having the largest effect of the three conceptual interventions tested. A reduction in the “infectivity” of the idea of using opioids nonmedically vis-à-vis susceptible individuals has greater short term impact, but effect reaches a plateau after 2 or 3 years. Restricting the supply via take-back days has only modest short-term impact, but the effects would be cumulative and therefore significant over the longer term.

Supply reduction achieved by reducing prescription holders’ willingness to share acts on two feedback processes, which acting together tend to amplify the beneficial effects in terms of reducing abuse and nonmedical use. The primary effect is to reduce the flow of medicines into the available free supply. However, the impact on the total number of nonmedical users is moderated because one effect of supply constriction would be to shift nonmedical users with use disorder from the “Free users with Use Disorder” stock to the “Paying users” stock. Nevertheless, over time, the slower replenishment rate of the free opioids stock would significantly reduce nonmedical opioid usage.

Demand reduction achieved by reducing the infectivity of the opioid use idea would also have a significant impact on the number of nonmedical users, and this effect is seen quickly because a 25 % reduction in infectivity results in an immediate 25 % reduction in the initiation. This reduction is offset somewhat over time,

however, because fewer nonmedical users means less demand, and therefore more supply, which reduces the negative impact of scarcity on initiation.

The prescription take-back scenario works similarly to the reduced willingness to share scenario, but size of the impact is less in the case tested. This is due to the particular experimental numbers used in the analysis, which have only very tentative empirical support.

## 4 Discussion

The model presented in this paper is useful primarily because it extends our understanding of the dynamics of pharmaceutical opioid abuse problem in the United States and comparatively demonstrates policy leverage points for intervention. The model proposes that nonmedical opioid use spreads in fashion similar to the spread of a disease. Some communities in the United States are deeply impacted by opioid abuse and others are not [3, 4]. Using the disease metaphor we might suggest that in some areas the opioid use *idea* had infected too few people for the idea to spread, while in others the infected population is large so the “disease” of opioid use has become endemic. The disease metaphor can be broadened to encompass possible additional intervention strategies. Reducing infectivity (of the idea of using opioids nonmedically) was shown to have a quick and significant impact. Other strategies including ‘immunization’ intervention potentially provide an approach that could inform the dynamics of nonmedical use of opioids, initiation, intervention alternatives and the resulting behaviour of the system. For instance, what might an “immunization” intervention strategy look like? How would it impact initiation? Could a policy be formulated that acts like a quarantine? Because the infectious disease metaphor has been formalized into a model and calibrated against historical data, these types of ideas may merit exploration.

The other two hypothetical interventions also showed promise. Further investigation seems warranted regarding ways to reduce the free supply, whether it is drug take back day programs, campaigns to reduce prescription holders’ willingness to share, or other ideas not yet considered. While the vast majority of nonmedical users use very little and do not develop use disorders, a small fraction do, and smaller fraction still buy opioids to support high levels of use (SAMHSA) [27]. Even though this fraction is small, it is included in this model because the high price of pharmaceutical opioids for those who cannot obtain them for free may be an important factor in the recent rise in heroin use [27]. The street price of pharmaceutical opioids is high compared to heroin, and qualitative studies suggest that many opioid users switch to heroin due to its lower cost [20, 34]. Modelling a progression of opioid use that includes a transition to paying may provide a jumping off point for an investigation of the recent rise in heroin use.

## ***4.1 Limitations***

This manuscript describes on-going research, and stronger empirical support is being sought for the many parameters in the model. Parameter validity tends to be the primary limitation in this type of study [31].

The scenarios presented in this preliminary analysis are too simple for a rigorous comparison of effectiveness. These scenarios compare the system level impact of hypothetical interventions with specific and stable proximal effects (such as a 25 % reduction in the infectivity of the idea of nonmedical drug use), on the number of people who become nonmedical users and on the number of people who escalate their usage and manifest use disorders. Framing scenarios in terms of their proximal effects leaves several important questions unanswered: How can these reductions be achieved? Are reductions of the desired magnitude achievable, given constraints such as limited budgets? How can we compare interventions if some are easy but low impact, and others are difficult but high impact? In order to compare the effectiveness of interventions themselves, model structures would need to be developed that transform exogenous inputs, such as dollars spent on drug resistance programs, into local outcomes that impact model parameters or structure, such as a 2 year delay in initiation. A more rigorous treatment of intervention strategies is necessary for this preliminary model to become a useful policy evaluation instrument.

Additionally, the population represented in the current model is derived from the NSDUH, which is known to be limited in its representation of hidden drug using communities such as the incarcerated, members of the armed forces, and the homeless [7]. The current model presents only one of several possible routes of initiation and does not include initiation of nonmedical use through medical exposure, or as a substitute for or complement to other illicit drug use. Furthermore, the potential impact of the availability of chronic pain medicine is not considered, and may be an important factor.

## ***4.2 Future Research***

Future work will include additional efforts to locate empirical support for model parameters and model structure to develop the model beyond the proof of concept stage. Expansion of the model logic for policy interventions is also planned. A variety of model testing techniques, including sensitivity analysis and more rigorous comparisons to reference behaviour, will help to strengthen the model's validity and credibility. In addition, model development is underway for several other aspects of the pharmaceutical opioid system, including the dynamics of black market opioid purchasing and the negative outcomes associated with nonmedical use, including transition to heroin use and fatal overdoses. Integration of the current model with these other sectors will enable future simulations to yield greater

insights regarding the likely magnitude of impact, and rigorous testing will increase confidence in the model's results.

## 5 Conclusion

Initiation and nonmedical use of pharmaceutical opioids has seen a dramatic rise from 1995 to 2005, and stabilization at a high level toward the end of the last decade [27]. The current model replicates historical trends in initiation and non-medical use, and in doing so provides increased understanding of underlying processes and feedback loops that may give rise to observed historical trends in the pharmaceutical opioid system. Based on initial simulation runs, the model also demonstrates the potential for the system dynamics approach to be useful in evaluating policy alternatives in terms of their likely impact on negative consequences. While further testing and elaboration of intervention logic are needed, preliminary results suggest that the public health interventions described here could potentially have sufficient leverage to appreciably decrease the number of individuals who use opioids nonmedically.

**Acknowledgments** Funding for this research was provided by NIDA grant number 5R21DA031361-02. We appreciate Dennis McCarty, Ph.D., Aaron Gilson, Ph.D., Lynn Webster, M.D., Todd Bodner, Ph.D., and Neal Wallace, Ph.D. who served as advisors and provided valuable expert judgement and insights.

## References

1. Andrews JA, Hops H, Duncan SC (1997) Adolescent modeling of parent substance use: the moderating effect of the relationship with the parent. *J Fam Psychol* 11:259
2. Bates C, Laciak R, Southwick A, Bishoff J (2011) Overprescription of postoperative narcotics: a look at postoperative pain medication delivery, consumption and disposal in urological practice. *J Urol* 185:551–555
3. Brownstein JS, Green TC, Cassidy TA, Butler SF (2010) Geographic information systems and pharmacoepidemiology: using spatial cluster detection to monitor local patterns of prescription opioid abuse. *Pharmacoepidemiol Drug Saf* 19:627–637
4. Butler SF, Venuti SW, Benoit C, Beaulaurier RL, Houle B, Katz N (2007) Internet surveillance: content analysis and monitoring of product-specific internet prescription opioid abuse-related postings. *Clin J Pain* 23:619–628
5. Centers for Disease Control and Prevention (2012) Prescription drug abuse and overdose: public health perspective. CDC's Primary Care and Public Health Initiative
6. Compton WM, Volkow ND (2006) Major increases in opioid analgesic abuse in the United States: concerns and strategies. *Drug Alcohol Depend* 81:103–108
7. Crum RM (2005) Epidemiology of opioid use, abuse, and dependence. *Treat Opioid Depend* 43
8. Dasgupta N, Mandl KD, Brownstein JS (2009) Breaking the news or fueling the epidemic? Temporal association between news media report volume and opioid-related mortality. *PLoS ONE* 4:e7758

9. Drug Enforcement Administration (2012) DEA's fifth national prescription drug take-back day results in another big haul, [www.justice.gov/dea/docs/results\\_final.pdf](http://www.justice.gov/dea/docs/results_final.pdf)
10. Fishbain DA, Cole B, Lewis J, Rosomoff HL, Rosomoff RS (2008) What percentage of chronic nonmalignant pain patients exposed to chronic opioid analgesic therapy develop abuse/addiction and/or aberrant drug-related behaviors? A structured evidence-based review. *Pain Med* 9:444–459
11. Food and Drug Administration (2013) Extended-release (er) and long-acting (la) opioid analgesics risk evaluation and mitigation strategy (REMS)
12. Fortuna RJ, Robbins BW, Caiola E, Joynt M, Halterman JS (2010) Prescribing of controlled medications to adolescents and young adults in the United States. *Pediatrics* 126:1108–1116
13. Gilson A (2012) Personal communication: findings of Madison WI drug take-back day
14. Gosin M, Marsiglia FF, Hecht ML (2003) Keepin'it REAL: a drug resistance curriculum tailored to the strengths and needs of pre-adolescents of the southwest. *J Drug Educ* 33:119–142
15. Governale L (2008) Outpatient drug utilization trends for oxycodone products
16. Hall AJ, Logan JE, Toblin RL, Kaplan JA, Kraner JC, Bixler D, Crosby AE, Paulozzi LJ (2008) Patterns of abuse among unintentional pharmaceutical overdose fatalities. *JAMA, J Am Med Assoc* 300:2613–2620
17. Johnson LD, O'Malley PM, Bachman JG, Schulenberg JE (2013) Monitoring the Future: National Results on Adolescent Drug Use: Overview of Key Findings, 2012. Ann Arbor Inst Soc Res Univ, Mich
18. Jones CM (2013) Heroin use and heroin use risk behaviors among nonmedical users of prescription opioid pain relievers—United States, 2002–2004 and 2008–2010. *Drug Alcohol Depend* 132:95–100
19. Katz NP, Birnbaum HG, Castor A (2010) Volume of prescription opioids used nonmedically in the United States. *J Pain Palliat Care Pharmacother* 24:141–144
20. Levy MS (2007) An exploratory study of OxyContin use among individuals with substance use disorders. *J Psychoact Drugs* 39:271–276
21. Martins SS, Keyes KM, Storr CL, Zhu H, Gruzca RA (2010) Birth-cohort trends in lifetime and past-year prescription opioid-use disorder resulting from nonmedical use: results from two national surveys. *J Stud Alcohol Drugs* 71:480–487
22. Paulozzi L (2011) Vital Signs: overdoses of prescription opioid pain relievers—United States, 1999–2008
23. Rosenblum A, Parrino M, Schnoll SH, Fong C, Maxwell C, Cleland CM, Magura S, Haddox JD (2007) Prescription opioid abuse among enrollees into methadone maintenance treatment. *Drug Alcohol Depend* 90:64–71
24. Sterman J (2000) Business dynamics: systems thinking and modeling for a complex world. Irwin/McGraw-Hill, Boston
25. Sterman JD (2006) Learning from evidence in a complex world. *J Inf* 96: 505–514
26. Substance Abuse and Mental Health Services Administration (2006) Results from the 2005 national survey on drug use and health: national findings. Department of Health and Human Services, Rockville
27. Substance Abuse and Mental Health Services Administration (2012) Results from the 2011 national survey on drug use and health: summary of national findings, [http://www.samhsa.gov/data/2k11/WEB\\_SR\\_088/WEB\\_SR\\_088.pdf](http://www.samhsa.gov/data/2k11/WEB_SR_088/WEB_SR_088.pdf) Accessed 2012
28. Sung HE, Richter L, Vaughan R, Johnson PB, Thom B (2005) Nonmedical use of prescription opioids among teenagers in the United States: trends and correlates. *J Adolesc Health* 37:44–51
29. Tufts health care institute program on opioid risk management (2010) The role of dentists in preventing opioid abuse tufts health care institute program on opioid risk management, [http://www.thci.org/opioid/opioid\\_mar10.asp](http://www.thci.org/opioid/opioid_mar10.asp) Accessed 2010
30. Volkow ND, McLellan TA, Cotto JH, Karithanom M, Weiss SRB (2011) Characteristics of opioid prescriptions in 2009. *JAMA, J Am Med Assoc* 305:1299–1301



31. Wakeland W, Fitzgerald J, Haddox JD (2010) Key data gaps for understanding trends in prescription opioid analgesic abuse and diversion among chronic pain patients and nonmedical users. Presented at the College on Problems of Drug Dependence, 72nd Annual Scientific Meeting, Scottsdale
32. Warner M, Chen LH, Makuc DM, Anderson RN, Minino AM (2011) Drug poisoning deaths in the United States, 1980–2008. Centers for Disease Control and Prevention (2011)
33. Winkler D, Caulkins JP, Behrens DA, Tragler G (2004) Estimating the relative efficiency of various forms of prevention at different stages of a drug epidemic. *Socioecon Plann Sci* 38:43–56
34. Young AM, Havens JR (2012) Transition from first illicit drug use to first injection drug use among rural Appalachian drug users: a cross-sectional comparison and retrospective survival analysis. *Addiction* 107:587–596

# Supervisory Fuzzy Cognitive Map Structure for Triage Assessment and Decision Support in the Emergency Department

Voula C. Georgopoulos and Chrysostomos D. Stylios

**Abstract** Soft Computing techniques, such as Fuzzy Cognitive Maps (FCMs), can handle uncertainties in modeling complex situations using abstract inference mechanisms; they have been successfully used to select among different suggestions, to lead to a decision and to develop Medical Decision Support Systems for many medical-discipline applications. FCM models have great ability to handle complexity, uncertainty and abstract inference as is the case in the health care sector. Here is examined the case of the triage procedure in the Emergency Department (ED), where a decision supporting mechanism is quite invaluable. A Hierarchical structure is proposed within an integrated computerized health system where the Supervisor is modeled as an abstract FCM to support the triaging procedure and assessment of the health condition of people with communication difficulties such as the elderly arriving at the ED. There is also the lower level of the hierarchical structure where a FCM-ESI DSS has been developed and used to assign the Triage ESI level of every patient. Here a new methodology for designing and developing the FCM-ESI DSS is presented so to ensure the active involvement of human experts during the FCM-ESI construction procedure.

**Keywords** Medical decision support system · Triage assessment modeling · Soft computing · Fuzzy cognitive maps

---

V.C. Georgopoulos (✉)

School of Health and Welfare Professions, Technological Educational Institute of Western Greece, Patras, Greece

e-mail: voula@teipat.gr

C.D. Stylios

Department of Computer Engineering, Technological Educational Institute of Epirus, Artas, Greece

e-mail: stylios@teiep.gr

© Springer International Publishing Switzerland 2015

M.S. Obaidat et al. (eds.), *Simulation and Modeling Methodologies, Technologies and Applications*, Advances in Intelligent Systems and Computing 319,

DOI 10.1007/978-3-319-11457-6\_18

## 1 Introduction

Hospital Emergency Departments (EDs) require prompt decisions that are significantly difficult in their making as they have to deal with situations characterized by inherent complexity, intrinsic uncertainty and dynamic nature. Frequently, the Hospital Emergency departments are overcrowded by elderly patients requiring critically urgent to non-urgent medical services while they may not be able to objectively describe and communicate their own health situation. Older patients (>65 years of age) account for 12–24 % of all ED attendees worldwide [25] and according to recent census data across Europe [9] and the US [29], adults over the age of 65-years-old comprise 17.7 and 13 % of the population making them one of the fastest growing segments of the population estimated to reach 25 % over the next 25 years. Given the constantly increasing age of a population this leads to increased numbers of visits of elderly patients to the ED and as a result to an increased burden on the EDs [32]. Older patients frequently have atypical clinical presentation and/or higher severity of illness, multiple comorbidities, increased frailty, a high prevalence of chronic-degenerative diseases which may include cognitive disorders, are susceptible to frequent exacerbations, all of which contribute to a higher risk of adverse outcomes [1, 26]. Taking into consideration the aging population, with its complexities, increasingly seeking services at the ED in combination with the limited resources and increased costs, this leads to the need for development of advanced decision making tools that will effectively and efficiently provide patient care in a timely fashion.

Although Emergency Departments (EDs) vary worldwide both in terms of range of services offered as well as patients arriving for care in EDs, a recent definition provides an accurate description of what goes on in EDs: “Emergency departments provide unscheduled care for a wide variety of persons for reasons that range from life-threatening conditions to problems that could be treated in a primary care setting” [22].

In all Emergency Departments, specific protocols are followed to evaluate the health condition and sort all the patients entering the ED. This procedure is called triaging and it involves an initial sorting of patients based on their health condition by rapidly identifying patients requiring immediate care due to urgent, life-threatening conditions, as well as assessing the severity of the problem so as to ensure that care is appropriate and timely [8]. According to general accepted protocols, patients are categorized to various levels of urgency [10, 28], based on their general condition as it is concluded by their appearance, their complaints about pain level along with a brief health examination which may include physiological factors when necessary. When available the patient record is also consulted. Following this procedure, the triage approach achieves minimizing of the waiting time for treatment of the most urgent patients, while those not in need of urgent treatment are placed in a waiting area.

A widely used tool for ED triage is the Emergency Severity Index (ESI). The ESI triage procedure yields rapid, reproducible, and clinically relevant stratification

of patients into five groups, from level 1 (most urgent) to level 5 (least urgent). The ESI provides a method for categorizing ED patients by both acuity and resource needs [17].

Triage decisions are often made with limited objective data, ambiguous information [16] while other times that there are numerous patient factors (some subjective), physiological measurements, and medical history variables that the ED triage nurse must consolidate during the decision making process [23].

In spite of its widespread use, it is significant to mention that, for older patients, it may present difficulty in accurate categorization by medical personnel not experienced in geriatrics. Also, communication as well as cognitive deficits may present a problem limiting the patient's ability to participate or cooperate during the initial triage process to identify chief complaints, symptoms and history [20]. This may result in potential under-triage or over-triage. In under-triage a patient is assigned a triage code level that is lower than their actual level of urgency based on objective clinical and physiological measures. Under-triage is defined as the underestimation of the severity of an illness or injury, resulting in a patient receiving lower levels of treatment (and/or with lower priority) than required. This decision has the potential to result in a prolonged waiting time to medical intervention for the patient and risks an adverse outcome [7, 30]. Of course, perfect triage is not possible in all cases and triage algorithms are usually designed with an under-triage rate of 5 % as an acceptable error rate trade-off in order to minimizing over-triage [5]. In a study of a sample of 50 randomly selected cases of ED admissions patients 65 years or older, discrepancies were found between the medical staff and expert nurses in 20 cases: where staff nurses had under-triaged 13 patients and over-triaged 7 patients [21]. According to another study [19] of 519 patients over 65, it was found that under-triage occurred in 117 cases, i.e. 22.5 % of the cases. In yet a third study of 4,534 geriatric trauma patients undergoing triage in an ED within a 10 year period, it was found that 15.1 % were under-triaged [24]. Therefore, there is a general consensus that under-triaging is not a result of ESI being a poor tool (quite the opposite actually), but rather the possible overlooking of high-risk situations and not appropriately considering vital signs.

Another difficulty arises from the fact that in various studies conducted of agreement between nurses in rating of triage levels to be only fair to moderate [18] even when conducted within the same hospital with the same group of patient case scenarios [6].

It is generally accepted that all patients arriving at the ED are not of equal severity and complexity requiring those that do not have a severe/and or life threatening condition to wait to receive medical care. These patients are triaged at levels 3–5 and do not normally receive immediate care even though those triaged as level 3 are treated with higher priority over those with levels 4 and 5, etc. When there are several patients waiting with the same ESI level, there are no clear differentiators to establish a prioritization [2]. Normally, after triage patients are prioritized within their level on a first-come-first served basis. For the elderly population where the complexity of problems is increased, a long wait may cause deterioration of their condition. This combined with under-triaging can lead to

adverse effects for these particular patients. Therefore, the problem at hand is two-fold, on one hand to be able to provide decision support in order to minimize, as much as possible, under-triaging and on the other hand it is important that patients are also prioritized after the triage classification within their classification category and not be tended to only on a first-come first-served basis.

Emergency departments are not only extremely complex because of the patient assessment and treatment protocols that are in place, but also due to the high level of automation and instrumentation, huge amounts of information, and interdisciplinary coordination that is necessary. Thus, the complex triage decision can be modeled using soft computing modeling techniques such as Fuzzy Cognitive Maps discussed in the next section.

In this work, a two-level Decision Support System is proposed to perform two complementary decisions: automatically assist in the triage classification as well as to suggest and update the priority for patients within their initial classification range.

## 2 Fuzzy Cognitive Maps

The soft computing technique of Fuzzy Cognitive Maps has been derived through the synergistic combination and integration of aspects of fuzzy logic, neural networks, semantic networks, expert systems and knowledge based systems. FCMs have been supplemented with other soft and hard computing methodologies in order to create advanced Decision Support Systems.

An FCM is illustrated as a causal graph representation consisting of interrelated weighted concepts. FCMs are fuzzy signed directed graphs permitting feedback, where the weighted edge  $w_{ij}$  from concept  $C_i$  to affected concept  $C_j$  describes the degree of causality by which the first concept influences the latter. It is mentioned that FCMs permit feedback and so they are characterized as fuzzy feedback models of causality, where the weighted interconnections between concepts of the FCMs stand for the influence between concepts and so they create an interconnected network of interrelated entities, similar to the abstract mental model than humans creates in their minds to model a complex situation and to infer decisions and suggestions. Feedback interconnections are permitted along with if-then inferencing; this is the main strength of the FCM to model any complex nonlinear dynamic system. Thus, FCMs have the ability to include hidden nonlinear dynamics.

In the Fuzzy Cognitive Model the key elements are concepts that stand for the main characteristics of an abstract mental model for any complex system. Each concept of the FCM model represents a granular entity representing a state, variable, input, output, event, action, goal, and/or trend of the real system that is modeled as an FCM. The value of every concept  $C_i$  is  $A_i$  and it results from the transformation of the fuzzy real value of the system's variable, for which this concept stands for, in the interval  $[0,1]$ . So the initial concept values are produced, which then are updated as they are computed through the interaction of the

interconnected concepts with the corresponding weight. Generally, between two concepts there are three possible types of causal relationships that express the type of influence of one concept to the other. The weight of the arc between concept  $C_i$  and concept  $C_j$  could be positive ( $W_{ij} > 0$ ) which means that an increase in the value of concept  $C_i$  leads to the increase of the value of concept  $C_j$ , and a decrease in the value of concept  $C_i$  leads to the decrease of the value of concept  $C_j$ . When there is negative causality ( $W_{ij} < 0$ ), this means that an increase in the value of concept  $C_i$  leads to the decrease of the value of concept  $C_j$  and vice versa. Finally, there can be no causality ( $W_{ij} = 0$ ).

The value  $A_i$  of every concept  $C_i$  expresses a fuzzy value of its corresponding physical value. FCMs are used to model the behavior of systems; during the simulation step, the value  $A_i$  of a concept  $C_i$  is calculated by computing the influence of the interconnected concepts  $C_j$ 's on the specific concept  $C_i$  following the calculation rule:

$$A_i^{(k+1)} = f(A_i^{(k)} + \sum_{\substack{j \neq i \\ j=1}}^N A_j^{(k)} \cdot w_{ji}) \quad (1)$$

where  $A_i^{(k+1)}$  is the value of concept  $C_i$  at simulation step  $k + 1$ ,  $A_j^{(k)}$  is the value of concept  $C_j$  at simulation step  $k$ ,  $w_{ji}$  is the weight of the interconnection from concept  $C_j$  to concept  $C_i$  and  $f$  is the sigmoid threshold function:

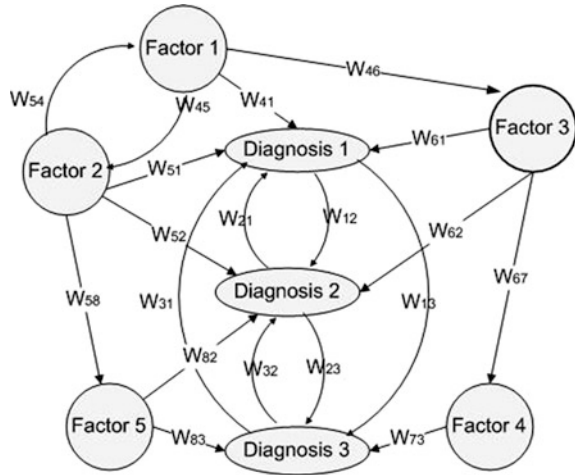
$$f = \frac{1}{1 + e^{-\lambda x}} \quad (2)$$

where  $\lambda > 0$  is a parameter that determines its steepness. In this approach, the value  $\lambda = 1$  has been used. This function is selected since the values  $A_i$  of the concepts must lie in the interval  $[0,1]$ .

Fuzzy Cognitive Maps have been used to develop Medical Decision Support Systems (MDSS). A specific type for Medical Diagnosis is the Competitive Fuzzy Cognitive Map (CFCM) [11, 12, 15] which consists of two main types of concepts: diagnosis-concepts and factor-concepts. Figure 1 illustrates an example CFCM model that is used to perform medical diagnosis. Here, the concepts of the FCM and the causal relations among them that influence concepts and determine the value of diagnosis concepts indicating the final diagnosis are illustrated.

In the CFCM model each diagnosis concept represents a single diagnosis, which means that these concepts must be mutually exclusive because the main intention is to always infer only one diagnosis. This is the case of most medical applications, where, according to symptoms, medical professionals conclude to only one diagnosis and then decide accordingly concerning the treatment. Actually, this comes from the medical axiom: "every patient has only one disease" but may represent many symptoms related to different diseases but all are results of the primitive disease. The general diagnosis procedure is a complex process that has to take

**Fig. 1** A CFCM model for medical diagnosis



under investigation a variety of interrelated factors, symptoms and functions. In accomplishing any diagnosis process, some of these factors are complementary, others are similar and even others are conflicting.

In the Competitive Fuzzy Cognitive Map model, the factor-concepts can be considered as inputs into the MDSS from patient data, observed symptoms, patient records, experimental and laboratory tests etc., which can be dynamically updated based on the system interaction, whereas the decision-concepts are considered as outputs where their estimated values outline the possible diagnosis for the patient.

### 3 ESI Triage System

When a patient first arrives in the Emergency Department, the first stop is triage where a trained and experienced registered nurse typically prioritizes each patient's condition into one of five general categories. This is done according to the Emergency Severity Index (ESI) which was designed for use in ED triage by the US Department of Health and Human Services. The ESI is a five-level categorization algorithm that prioritizes patients into five groups from 1 (most urgent) to 5 (least urgent) on the basis of severity and the number of resources that the patient may need to receive proper care [31].

In particular, the ESI uses the following scale based on decision points to determine its categories [3, 17]:

- ESI category 1-Emergent: patient intubated, without pulse or respiration, or unresponsive i.e. the patient requires immediate life-saving intervention so as to prevent loss of life, limb, or eyesight,
- ESI category 2-Urgent: patient is in a high-risk situation, or confused, lethargic or disoriented, or in severe pain, or danger zone of vital signs.

- ESI category 3-Acute: patient is in need of many resources to be taken care of. These may include, for example, Laboratory Tests, ECG, X-rays, CT-MRI-ultrasound-angiography, IV fluids, specialty consultation, complex procedures etc.
- ESI category 4-Routine: patient is in need of one resource.
- ESI category 5-Non urgent: patient is in need of no resources.

In comparison with other previously used triage severity scales along the dimensions of ease of implementation, ease of use, and in predicting resource demand, users reported that the ESI is much better or at least better [27]. Triage is a dynamic decision-making process and so continuous reassessment of relative factors and evolving information is necessary to ensure that whoever requires the most immediate receives it. It requires a continuously monitoring and keeping track of patients waiting to be seen and a prompt assessment of each new patient who arrives at the triage area. Therefore, due to the dynamic, complex and uncertain nature of the overall triage process in addition to the difficulty in differentiating severity levels, decision support methods are needed to help the triage nurse to be efficient in prioritizing patients with the same acuity classification [2, 23]. Therefore a Hierarchical Decision Support System for ESI Triage based on Fuzzy Cognitive Maps is developed in the next section: the first layer is a FCM-ESI Triage Model and the second layer is a supervisor assisting in patient prioritizing for those patient categorized as ESI 3–5.

#### **4 First Layer Fuzzy Cognitive Map Model for ESI Triage System**

In an ED triage system each patient is assigned one of the 5 ESI levels and therefore, the Fuzzy Cognitive Map ESI has to include 5 Decision Concepts (DC), each one corresponding to an ESI level: DC1-ESI Level 1, DC2-ESI Level 2, DC3-ESI Level 3, DC4-ESI Level 4, DC5-ESI Level 5.

The FCM development procedure is based on human experts who must define the main factor concepts that influence the triaging procedure and thus, they have to be represented in the FCM-ESI DSS. The well-known procedure of FCM construction is followed, where there is a group of experts whose knowledge and experience is exploited to design the FCM-ESI DSS and so they are asked to select the main factors based on which they usually conclude to an ESI triage level. Every expert replies with a set of factors that based on his experience are essential to conclude to a triaging decision, this is a blind procedure and no expert knows what the others have suggested. Then, based on the frequency with which each factor was chosen by the group of experts as a whole, the importance weight ( $iw$ ) between a factor concept and a decision concept is determined. It is obvious that the greater the the number of experts the more objective an FCM-ESI DSS will be constructed.



There are 23 factor concepts (FC1-FC23) that have been identified for the FCM-ESI DSS. The importance weights  $i_w$  for the first 22 have been detailed reported in [13, 14] and are mentioned in Table 1. Factor FC23 is a pre-existing communication/cognitive deficit which affects the patient's ability to explain and/or identify the chief complaint. This is an important factor to be assessed because studies have shown that unrecognized cognitive deficits are present in 30–40 % of older emergency department patients [4]. This is due the fact that these patients usually do well in basic communication functions making it not obviously detectable. Therefore, detecting cognitive dysfunction is important because it may affect the ESI level at which they are characterized as well as their thereafter waiting priority.

Moreover, the experts are asked to evaluate the triage stage of specific cases and based on their assessment we infer additional information, which leads to a complementary second weight, the “influence to specific decision” specific weight-  $sw$ ,

**Table 1** Factors of the FCM-ESI

Factor concept	Factor	Importance weight ( $i_w$ )
FC1	Life threatening	0.45
FC2	Limb threatening	0.40
FC3	Patient chief complaint	0.67
FC4	Vital signs	0.4
FC5	Medical history	0.35
FC6	Other factor	0.32
FC7	Expected # of resources	0.31
FC8	Patient age	0.16
FC9	Required timely intervention	0.15
FC10	Weakness	0.20
FC11	Additional symptoms other than chief complaint	0.14
FC12	Severe pain or distress	0.12
FC13	Patient referred to ED from outside	0.08
FC14	Behavioral or psychiatric issue	0.07
FC15	No additional symptoms to chief complaint	0.05
FC16	Absence of medical history	0.05
FC17	Patient medications	0.05
FC18	Hospital or ED discharge <3 days	0.04
FC19	Patient immune-compromised	0.04
FC20	Alcohol or illicit drug use	0.03
FC21	No recent change mental state	0.75
FC22	Patient can walk or sit	0.12
FC23	Pre-existing communication/cognitive deficits	0.10

which represents how much the specific factor leads towards a specific decision/diagnosis. The procedure to calculate the  $sw$  is the following: every expert who considers one factor as important and takes it into consideration, is asked to present the degree with which this specific factor leads the expert to select one decision. Every expert describes the degree of influence of one factor towards one decision using a linguistic variable, such as “very strong influence  $vs_i$ ”, “strong influence,  $s_i$ ”, “medium influence,  $m_i$ ”, “weak influence  $w_i$ ”, “very weak influence  $vw_i$ ”, as it is depicted in Fig. 2.

Thus, every expert describes the specific weight  $sw$  of each interconnection with a fuzzy linguistic variable from the above mentioned set, which stands for the relationship between the two concepts and determines the grade of causality between the two concepts. Then, all the proposed linguistic weights for one interconnection suggested by experts, are aggregated using the SUM method and an overall linguistic weight is produced. The overall linguistic weight with the defuzzification method of Center Of Gravity (COG), is transformed to a numerical weight belonging to the interval  $[-1, 1]$ .

Then, the overall weight describing the influence from one factor concept towards a decision concept is calculated using the form:

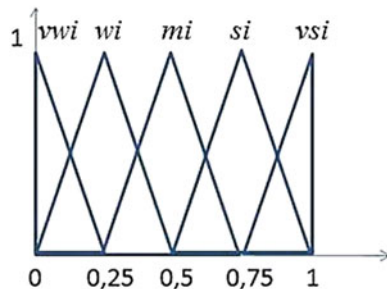
$$w_{ji} = \text{sgn}(sw)(l_1 * iw + l_2 * |sw|) \tag{3}$$

where the two parameters  $l_1, l_2$  are introduced to represent the participation of the *importance weight*  $iw$  and the *specific weight*  $sw$ , on the overall weight describing the influence of every factor concept towards the decision/diagnosis concept. It is mentioned that the value of  $w_{ji}$  has to be normalized in the interval  $[-1, 1]$ , where the weight takes values.

The current FCM-ESI DSS model has extended the previous developed one [14] to include interactions between the various factor concepts since the decision is complex and there are always cause-effect relationships between factors contributing to the triage decision. These are as follows:

- Vital signs (FC4) and Patient chief complaint (FC3); Vital signs (FC4) and Patient immuno-compromised (FC19)—for example patients that are experiencing fever and are on chemotherapy.

**Fig. 2** The positive fuzzy linguistic weights



- Over the counter medications (F15) and chief complaint (FC3)—patients using over the counter pain medication may have a decrease in their pain level and as a result the severity indicated concerning their chief complaint may be decreased
- Over the counter medications (F15) and vital signs (FC4)—over the counter medications may change vital signs, e.g. reduced fever, increased blood pressure etc.
- Pre-existing communication/cognitive deficits (FC23) and Chief Complaint (FC3)—there is increased inaccuracy of the description of the chief complaint with the existence of communication/cognitive deficits.
- Pre-existing communication/cognitive deficits (FC23) and Medical History (FC5)—the reliability of medical history can be compromised by communication/cognitive deficits of the patient.
- Pre-existing communication/cognitive deficits (FC23) and Severe Pain or Distress (FC12)—there may be on one hand inability to judge severity of pain and

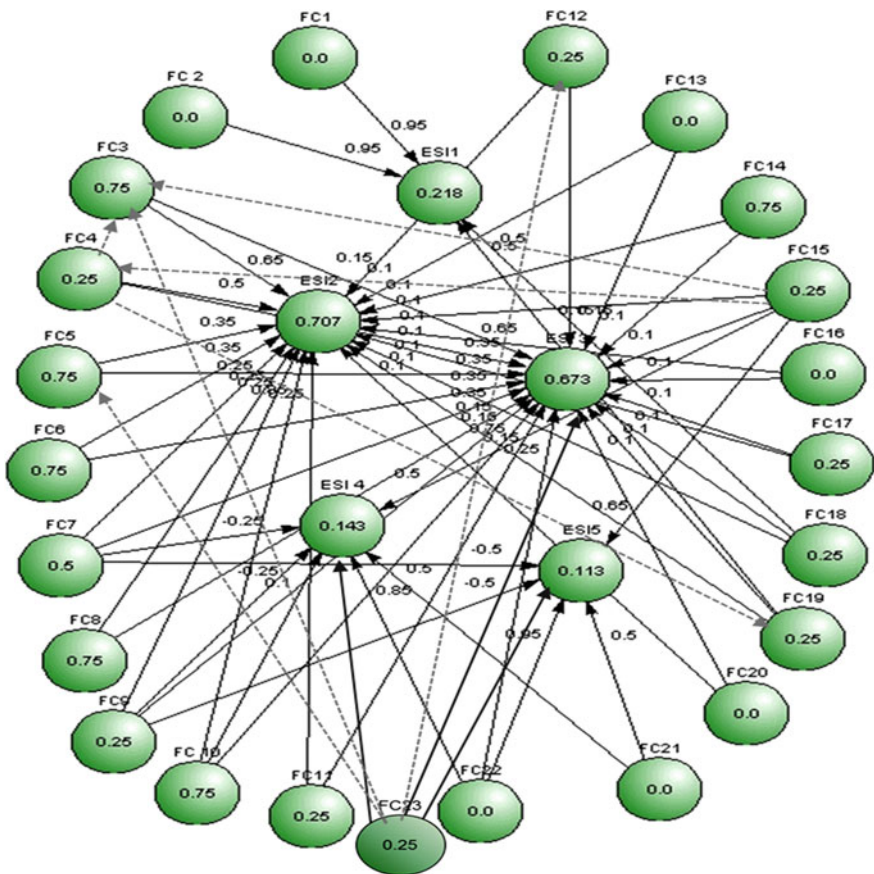


Fig. 3 23 factor FCM-ESI including factor interactions

on the other hand increased distress may be related to unfamiliar environmental change and not to actual medical condition.

Experts were asked to identify the possible interactions among Factor Concepts. At first, every expert is asked to determine the pair of concepts that are coupled. Thus, a set of possible paired factors concepts is created and then all the experts are asked to suggest the degree (using a linguistic weight) of coupling/influencing among the previously identified pairs of factor concepts. A set of five fuzzy linguistic values are used: “very strong”, “strong”, “medium”, “weak”, and “very weak”.

The complete FCM-ESI is illustrated in Fig. 3 where the 5 central nodes are the decision nodes (ESI levels). The decision node with the maximum value is the level at which a patient is triaged.

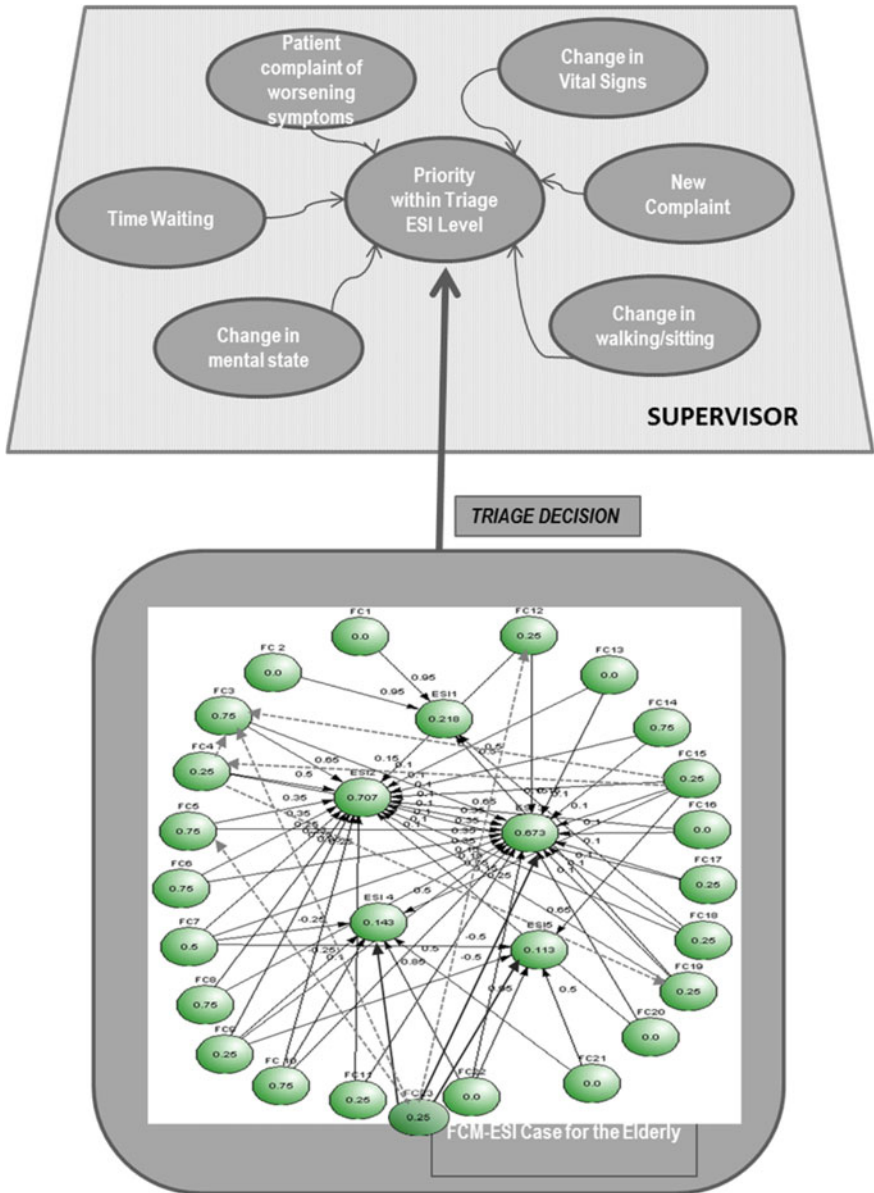
The FCM-ESI for every patient takes the concept factor values from measurements, laboratory test and examination and these values are transformed in the interval [0,1] where concepts take values and then the values of Decision Concepts are calculated, so that to infer the ESI level.

## 5 Supervisor System for Priority Between Equally Triage Patients for ESI Levels 3–5

An important issue after initial triage for patients with ESI levels 3–5 in an overcrowded ED is the priority with which patients receive care. Patients within a given triage level may end up being seen by a physician on a first come—first served basis. However, unfortunately, severity may change over time. For example, in some cases injuries and illnesses that need medical and nursing intervention are time sensitive. The longer the wait, the more damage may occur because of changes for example such in oxygen, blood, electrolytes (potassium, sodium, etc.), sugar, etc. Failure to prioritize triaged patients appropriately may result in very sick patients at risk for deterioration while waiting. Since this is also difficult and critical decision for the personnel in the ED, a supervisory level has been added on the FCM-ESI DSS model, where the outcome reflects changes in priority for patients within the same ESI level.

In order to develop the supervisor priority FCM-ESI Decision Support Systems, the most essential factor concepts are selected that may influence the patient status. The supervisor priority FCM-ESI consists of the concepts (Fig. 4):

- Change in vital signs (FC4).
- Patient report of worsening symptoms.
- Change in mental state (FC21).
- Change in patient can walk or sit (FC22).
- Triage ESI level.
- Time in waiting area.
- New symptoms.



**Fig. 4** Supervisory fuzzy cognitive map structure for triage assessment and decision support in the emergency department

Therefore, for each patient where a particular ESI Decision Concept had the maximum value (i.e. the ESI Level with which the patient was characterized) this decision node interacts dynamically with other FC nodes as new information is provided over time and is checked every half hour.

This leads to prioritizing of the patients that have equal or almost equal ESI status in order to avoid adverse events after triage due to long wait in overcrowding. Thus, the supervisor priority FCM-ESI is called to prioritize among patients with the same ESI level.

## 6 Case Example

A 72-year-old woman presented to a busy emergency department (ED). During triage she told the triage nurse that she experienced face and tongue swelling in the last 2 days. The vital signs at triage, including respiratory rate and oxygen saturation, were normal (BP 125/70 mm Hg, HR 72 beats/min, RR 12 breaths/min, and Body Temperature 36.5 °C). No previous history indicated this condition, the patient had not taken medications, was not in any pain, and there were no problems with the patient's mental state. Also the patient was able to sit and walk. Using the information collected at triage, both a triage nurse and the FCM-ESI resulted in ESI-Level 3.

The patient sat in the waiting room for more than 2 h after which she was placed in a room in the ED. After an additional hour a physician evaluated her. In the meantime, her tongue and throat had swollen substantially, and she was having difficulty breathing. She was diagnosed with angioedema and required emergency intubation, a potentially dangerous and high-risk procedure accompanied by aggressive treatment with intravenous epinephrine.

On the other hand, if the supervisor priority FCM-ESI DSS had been used, due to the change in state after 1.5 h when the patient was experiencing substantial swelling and difficulty breathing, the triage the patient priority would be increased to Very-Very High and thus, the patient would receive immediate care without the need for intubation.

## 7 Summary

This paper is an extension of previous work [13, 14] and it presents an integrated methodology for developing a hierarchical Decision Support System for ESI Triage. Usually at the Emergency Department (ED) of hospitals medical staff has to cope with many patients, asking for urgent treatment and so they have to assess their health condition under significant time constrains. The case of elderly patients has great importance as they usually are admitted quite frequently at the ED suffering from chronic problems, their health condition is characterized by complementarity and/or controversy and usually with a lack of interaction and low level communication ability. Thus, triaging of elderly people is characterized by high complexity and it makes the assessment and decision about the health condition a difficult task.

For such cases, Soft Computing methodologies are rather suitable and so Fuzzy Cognitive Maps (FCMs) are proposed here to model and develop a Supervisory Decision Support Systems for the ESI Triage during patient admission at the Emergency Department (ED) of hospitals. Generally, FCMs have been successfully applied to develop Medical Decision Support Systems for many discipline fields.

Here, there is further expansion of a recently proposed methodology to develop FCMs exploiting and combining knowledge and experience of human experts along with information and bibliographic data. Moreover, a hierarchical two level structure is introduced consisting of a FCM at each level. In the lower level a FCM-ESI system categorizes patients according to the 5 levels of the ESI. Then in the supervisor layer there is a prioritization of patients within ESI levels 3–5 which is continuously updated, as new information is received, in order to assist in preventing adverse outcomes while waiting.

**Acknowledgments** This work was supported by the joint research project “Intelligent System for Automatic CardioTocoGraphic Data Analysis and Evaluation using State of the Art Computational Intelligence Techniques” by the programme “Greece-Czech Joint Research and Technology projects 2011–2013”.

## References

1. Aminzadeh F, Dalziel WB (2002) Older adults in the emergency department: a systematic review of patterns of use, adverse outcomes, and effectiveness of interventions. *Ann Emerg Med* 39:238–247
2. Ashour OM, Kremer GEO (2013) A simulation analysis of the impact of FAHP–MAUT triage algorithm on the emergency department performance measures. *Expert Syst Appl* 40:177–187
3. Barbee GA, Berry-Cabán CS, Daymude ML, Oliver J, Gay S (2012) The effect of provider level triage in a military treatment facility emergency department. *J Emerg Primary Health Care* 8(4):2
4. Birrer R, Singh U, Kumar DN (1999) Disability and dementia in the emergency department. *Emerg Med Clin North Am* 17(2):505–517
5. Champion HR (1986) Triage. In Gales RH, Heilig RW Jr (ed) *Trauma care systems*. Aspen Publishers, Rockville, pp 79–109
6. Dallaire C, Poitras J, Aubin K, Lavoie A, Moore L (2012) Emergency department triage: do experienced nurses agree on triage scores? *J Emerg Med* 42(6):736–740
7. Department of Health and Ageing (2007) *Emergency triage education kit*. Australian Government, Canberra
8. Emergency Nurses Association (ENA) (2001) *Making the right decision: a triage curriculum*, 2nd edn. ENA, Des Plaines
9. Eurostat (2013) People by age group. [http://epp.eurostat.ec.europa.eu/portal/page/portal/population/data/main\\_tables](http://epp.eurostat.ec.europa.eu/portal/page/portal/population/data/main_tables). Last updated 26 Sept 2013
10. Fernandes C, Tanabe P, Gilboy N, Johnson LA, McNair RS, Rosenau AM et al (2005) Five-level triage: a report from the ACEP/ENA five-level triage task force. *J Emerg Nurs* 31(1):39–50
11. Georgopoulos VC, Stylios CD (2005) Augmented fuzzy cognitive maps supplemented with case based reasoning for advanced medical decision support. In: Nikraves M, Zadeh LA, Kacprzyk J (Eds) *Soft computing for information processing*. Springer, New York, pp 389–388

12. Georgopoulos VC, Stylios CD (2008) Complementary case-based reasoning, competitive Fuzzy cognitive maps for advanced medical decisions. *Soft Comput* 12:191–199
13. Georgopoulos VC, Stylios CD (2012) Introducing fuzzy cognitive maps for developing decision support system for triage at emergency room admissions for the elderly. In: *Proceedings of the 8th IFAC symposium on biological and medical systems*, 29–31 August, Budapest, Hungary
14. Georgopoulos VC, Stylios CD (2013) Fuzzy cognitive map decision support system for successful triage to reduce unnecessary emergency room admissions for the elderly. In: *Fuzziness and medicine: philosophical reflections and application systems in health care*. Springer, Berlin Heidelberg, pp 415–436
15. Georgopoulos VC, Malandraki GA, Stylios CD (2003) A fuzzy cognitive map approach to differential diagnosis of specific language impairment. *J Artif Intel Med* 29:261–278
16. Gerdtz MF, Bucknall TK (1999) Why we do the things we do: applying clinical decision-making frameworks to triage practice. *Accid Emerg Nurs* 7(1):50–57
17. Gilboy N, Tanabe T, Travers D, Rosenau AM (2011) Emergency severity index (ESI): a triage tool for emergency department care, Version 4. *Implementation Handbook 2012 Edition*. AHRQ Publication No. 12-0014
18. Göransson KE, Ehrenberg A, Marklund B, Ehnfors M (2006) Emergency department triage: is there a link between nurses' personal characteristics and accuracy in triage decisions? *Accid Emerg Nurs* 14(2):83–88
19. Grossmann FF, Zumbrunn T, Frauchiger A, Delpont K, Bingisser R, Nickel CH (2012) At risk of undertriage? Testing the performance and accuracy of the emergency severity index in older emergency department patients. *Ann Emerg Med* 60(3):317–325
20. Hampers LC, McNulty JE (2002) Professional interpreters and bilingual physicians in a pediatric emergency department: effect on resource utilization. *Arch Pediatr Adolesc Med* 156(11):1108–1113
21. McCall B., Travers D, Platts-Mills T, Biese K (2009) Mistriage of elderly in the emergency department. *Connecting the dots: geriatric nursing, education, and clinical simulation*. Chapel Hill, NC
22. Nawar EW, Niska RW, Xu J (2007) National hospital ambulatory medical care survey: 2005 emergency department summary. *Adv Data* 386:1–32
23. Patel VL, Gutnik LA, Karlin DR, Pusic M (2008) Calibrating urgency: triage decision-making in a pediatric emergency department. *Adv Health Sci Educ* 13(4):503–520
24. Rogers A, Rogers F, Bradburn E, Krasne M, Lee J, Wu D, Edavettal M, Horst M (2012) Old and undertriaged: a lethal combination. *Am Surg* 78(6):711–715
25. Salvi F, Morichi V, Grilli A, Giorgi R, De Tommaso G, Dessi-Fulgheri P (2007) The elderly in the emergency department: a critical review of problems and solutions. *Intern Emerg Med* 2:292–301
26. Samaras N, Chevalley T, Samaras D, Gold G (2010) Older patients in the emergency department: a review. *Ann Emerg Med* 56(3):261–269
27. Singer RF, Infante AA, Oppenheimer CC, West CA, Siegel B (2012) The use of and satisfaction with the emergency severity index. *J Emerg Nurs* 38(2):120–126
28. Travers DA, Waller AE, Bowling JM, Flowers D, Tintinalli J (2002) Five-level triage system more effective than three-level in tertiary emergency department. *J Emerg Nurs* 28(5):395–400
29. US Census Bureau (2011) 2010 Census Brief, Washington. US Department of Commerce, Economics, and Statistics Administrations, US Census Bureau, DC
30. Whiley SP, Alves H, Grace S (2013) Full-body X-ray imaging to facilitate triage: a potential aid in high-volume emergency departments. *Emerg Med Int* 2013:437078
31. Wuerz R (2001) Emergency severity index triage category is associated with six-month survival. ESI triage study group. *Acad Emerg Med* 8:61–64
32. Yim VW, Graham CA, Rainer TH (2009) A comparison of emergency department utilization by elderly and younger adult patients presenting to three hospitals in Hong Kong. *Int J Emerg Med* 2:19–24



# Author Index

## B

Bartolucci, Christian, 119  
Becker, Julianna, 87  
Bruzzone, Agostino G., 119

## C

Chéramy, Maxime, 37  
Chou, Cheng-tung, 157  
Clees, Tanja, 211

## D

Déplanche, Anne-Marie, 37

## E

Eichler, Annika, 71  
Elsässer, Robert, 55

## F

Ferrando, Angelo, 119

## G

Gay-García, Carlos, 145, 171  
Georgopoulos, Voula C., 255  
Glubokov, Oleksandr, 185  
Guadagnini, Alberto, 133

## H

Hladik, Pierre-Emmanuel, 37  
Huang, Yu-Jie, 157

## I

Ishii, Nobuaki, 101

## K

Kashevnik, Alexey, 3  
Koziel, Slawomir, 185, 199

## L

Langley, David, 87  
Leifsson, Leifur, 185, 199  
Levashova, Tatiana, 3  
Lichtenberg, Gerwald, 71

## M

Massei, Marina, 119  
Meier, Michael, 55  
Miao, Fu-qing, 223  
Muraki, Masaaki, 101

## N

Neuman, Shlomo P., 133  
Nielsen, Alexandra, 239  
Nikitina, Lialia, 211  
Nikitin, Igor, 211

## O

Ogierman, Adrian, 55  
Ogurtsov, Stanislav, 199

**P**

Pangalos, Georg, [71](#)  
Park, H.S., [223](#)  
Paz-Ortiz, Iván, [145](#)  
Poggi, Simonluca, [119](#)  
Pott, Sabine, [211](#)

**R**

Riva, Monica, [133](#)  
Rossmann, Juergen, [23](#)

**S**

Sánchez Meneses, Oscar, [171](#)  
Schaap, Marcel G., [133](#)  
Schluse, Michael, [23](#)  
Shih, Yu-Hau, [157](#)  
Shilov, Nikolay, [3](#)  
Smirnov, Alexander, [3](#)  
Stylios, Chrysostomos D., [255](#)

**T**

Takano, Yuichi, [101](#)

**V**

Vermonden, Anäis, [145](#)

**W**

Wakeland, Wayne, [239](#)  
Waspe, Ralf, [23](#)  
Worm, Daniël, [87](#)

**Y**

Yang, Hong-sung, [157](#)

**Z**

Zimam, Amanuel, [239](#)