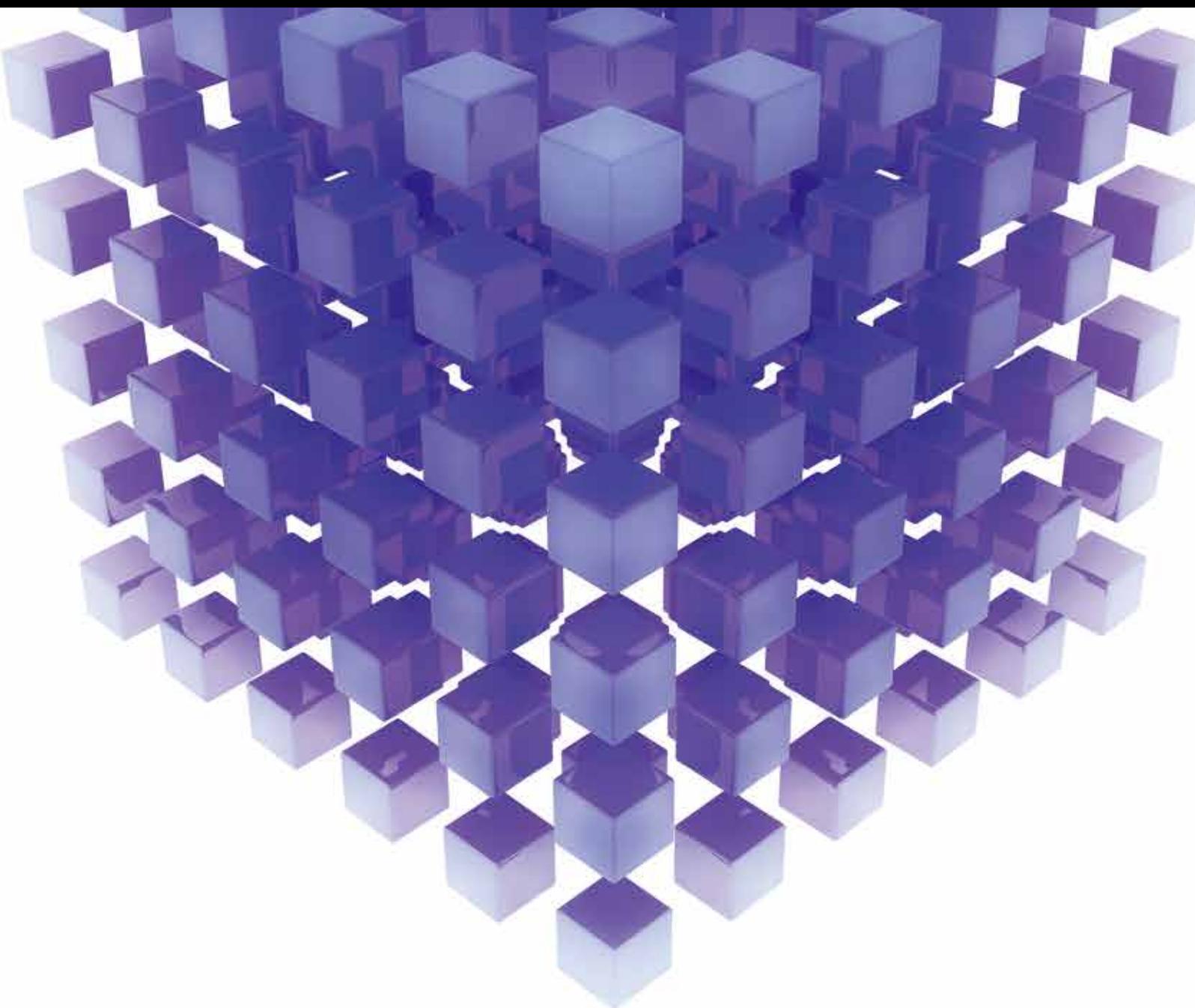


Modeling, Planning, and Control of Complex Logistic Processes

Guest Editors: Hamid Reza Karimi, Neil Duffie, Michael Freitag, Michael Lütjen, and Mohammed Chadli





Modeling, Planning, and Control of Complex Logistic Processes

Mathematical Problems in Engineering

Modeling, Planning, and Control of Complex Logistic Processes

Guest Editors: Hamid R. Karimi, Neil Duffie,
Michael Freitag, Michael Ltjen, and Mohammed Chadli



Copyright © 2015 Hindawi Publishing Corporation. All rights reserved.

This is a special issue published in “Mathematical Problems in Engineering.” All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Editorial Board

Mohamed Abd El Aziz, Egypt
Silvia Abrahão, Spain
Ramesh Agarwal, USA
Juan C. Agüero, Australia
Ricardo Aguilar-López, Mexico
Tarek Ahmed-Ali, France
Hamid Akbarzadeh, Canada
Muhammad N. Akram, Norway
Salvatore Alfonzetti, Italy
Tofiq Allahviranloo, Iran
Christian B. Allen, UK
Lionel Amodeo, France
Igor Andrianov, Germany
Sebastian Anita, Romania
Felice Arena, Italy
Sabri Arik, Turkey
Fumihiko Ashida, Japan
Seungik Baek, USA
Ezzat G. Bakhoun, USA
Laurent Bako, France
Stefan Balint, Romania
José M. Balthazar, Brazil
Alfonso Banos, Spain
Roberto Baratti, Italy
Abdel-Hakim Bendada, Canada
Rasajit K. Bera, India
Simone Bianco, Italy
Jonathan N. Blakely, USA
Daniela Boso, Italy
Taha Boukhobza, France
Francesco Braghin, Italy
Reyolando M. Brasil, Brazil
Michael J. Brennan, UK
Javier Bulduf, Spain
Salvatore Caddemi, Italy
Jose E. Capilla, Spain
Carlo Cattani, Italy
Marcelo M. Cavalcanti, Brazil
Mohammed Chadli, France
Shuenn-Yih Chang, Taiwan
Ching-Ter Chang, Taiwan
Michael J. Chappell, UK
Kacem Chehdi, France
Kui Fu Chen, China
Xinkai Chen, Japan

Jianbing Chen, China
Xuefeng Chen, China
Chunlin Chen, China
Zhang Chen, China
Singa W. Chiu, Taiwan
Jyh-Hong Chou, Taiwan
Slim Choura, Tunisia
Hung-Yuan Chung, Taiwan
Marcelo J. Colao, Brazil
Carlo Cosentino, Italy
Erik Cuevas, Mexico
Weizhong Dai, USA
Binxiang Dai, China
Purushothaman Damodaran, USA
Farhang Daneshmand, Canada
Swagatam Das, India
Fabio De Angelis, Italy
Filippo de Monte, Italy
Yannis Dimakopoulos, Greece
Baocang Ding, China
Zhengtao Ding, UK
Mohamed Djemai, France
Joao B. R. Do Val, Brazil
Alexandre B. Dolgui, France
Rodrigo W. dos Santos, Brazil
Alexander N. Dudin, Belarus
George S. Dulikravich, USA
Horst Ecker, Austria
Mehmet Onder Efe, Turkey
Karen Egiazarian, Finland
Elmetwally Elabbasy, Egypt
Alex E.-Zúñiga, Mexico
Fouad Erchiqui, Canada
Anders Eriksson, Sweden
Vedat S. Erturk, Turkey
Hua Fan, China
Ricardo Femat, Mexico
Jose R. Fernandez, Spain
Thierry Floquet, France
George Flowers, USA
Tomonari Furukawa, USA
Zoran Gajic, USA
Ugo Galvanetto, Italy
Xin-Lin Gao, USA
Zhong-Ke Gao, China

Laura Gardini, Italy
Alessandro Gasparetto, Italy
Oleg V. Gendelman, Israel
Paulo Batista Goncalves, Brazil
Rama S. R. Gorla, USA
Oded Gottlieb, Israel
Quang Phuc Ha, Australia
Masoud Hajarian, Iran
Zhen-Lai Han, China
Thomas Hanne, Switzerland
Xiao-Qiao He, China
Katica R. Hedrih, Serbia
M. Isabel Herreros, Spain
Wei-Chiang Hong, Taiwan
Jaromir Horacek, Czech Republic
Muneo Hori, Japan
Feng-Hsiang Hsiao, Taiwan
Fu-Shiung Hsieh, Taiwan
Changchun Hua, China
Zhenkun Huang, China
Chiung-Shiann Huang, Taiwan
Chuangxia Huang, China
Gordon Huang, Canada
Huabing Huang, China
Hai-Feng Huo, China
Asier Ibeas, Spain
Giacomo Innocenti, Italy
Nazrul Islam, USA
Reza Jazar, Australia
Khalide Jbilou, France
Linni Jian, China
Bin Jiang, China
Zhongping Jiang, USA
Jun Jiang, China
Jianjun Jiao, China
Ningde Jin, China
J. Joao Judice, Portugal
Tadeusz Kaczorek, Poland
T. Kalmar-Nagy, Hungary
T. Kapitaniak, Poland
Haranath Kar, India
C. Masood Khalique, South Africa
Do Wan Kim, Korea
Nam-Il Kim, Korea
Kyung Y. Kim, Republic of Korea

Manfred Krafczyk, Germany
V. Kravchenko, Mexico
Jurgen Kurths, Germany
Kyandoghere Kyamakya, Austria
Hak-Keung Lam, UK
Wen-Chiung Lee, Taiwan
Marek Lefk, Poland
Yaguo Lei, China
Valter J. S. Leite, Brazil
Stefano Lenci, Italy
Roman Lewandowski, Poland
Ming Li, China
Jian Li, China
Qing Q. Liang, Australia
Yan Liang, China
Teh-Lu Liao, Taiwan
Panos Liatsis, UK
Kim M. Liew, Hong Kong
Yi-Kuei Lin, Taiwan
Shueei M. Lin, Taiwan
Jui-Sheng Lin, Taiwan
Wanquan Liu, Australia
Yan-Jun Liu, China
Yuji Liu, China
Xian Liu, China
Peter Liu, Taiwan
Peide Liu, China
Paolo Lonetti, Italy
Vassilios C. Loukopoulos, Greece
Junguo Lu, China
Chien-Yu Lu, Taiwan
Jianquan Lu, China
Jinhu Lü, China
Tiedong Ma, China
Nazim I. Mahmudov, Turkey
Fazal M. Mahomed, South Africa
O. Daniel Makinde, South Africa
Didier Maquin, France
Rafael Martínez-Guerra, Mexico
Geffrard A. Maugin, France
Driss Mehdi, France
Khaled Saad Mekheimer, Egypt
Roderick Melnik, Canada
Pasquale Memmolo, Italy
Xiangyu Meng, Canada
Xinzhu Meng, China
Jose Merodio, Spain
Y. V. Mikhlin, Ukraine
Hiroyuki Mino, Japan
Pablo Mira, Spain
Nenad Mladenovic, Serbia
Ebrahim Momoniat, South Africa
Gisele Mophou, France
Rafael M. Morales, Spain
Giuseppe Muscolino, Italy
N. Bhujappa Naduvinamani, India
Trung Nguyen-Thoi, Vietnam
Hung Nguyen-Xuan, Vietnam
Ben T. Nohara, Japan
Sotiris K. Ntouyas, Greece
Maxim A. Olshanskii, Russia
Alejandro Ortega-Moñux, Spain
Erika Ottaviano, Italy
Alkiviadis Paipetis, Greece
Alessandro Palmeri, UK
Bijaya Ketan Panigrahi, India
Manuel Pastor, Spain
Pubudu N. Pathirana, Australia
Francesco Pellicano, Italy
Mingshu Peng, China
Zhike Peng, China
Haipeng Peng, China
Matjaz Perc, Slovenia
Maria do Rosário Pinho, Portugal
José R. C. Piqueira, Brazil
Antonina Pirrotta, Italy
Javier Plaza, Spain
Stanislav Potapenko, Canada
Sergio Preidikman, USA
Carsten Proppe, Germany
Hector Puebla, Mexico
Yuming Qin, China
Dane Quinn, USA
Jose Ragot, France
K. Ramamani Rajagopal, USA
Sellakkutti Rajendran, Singapore
Gianluca Ranzi, Australia
Bhairavavajjula N. Rao, India
Sivaguru Ravindran, USA
Alessandro Reali, Italy
Giuseppe Rega, Italy
Ricardo Riaza, Spain
Gerasimos Rigatos, Greece
José Rodellar, Spain
Rosana Rodriguez-Lopez, Spain
Ignacio Rojas, Spain
Carla Roque, Portugal
Debasish Roy, India
Rubén Ruiz García, Spain
Antonio Ruiz-Cortes, Spain
Ivan D. Rukhlenko, Australia
Abbas Saadatmandi, Iran
Kishin Sadarangani, Spain
Mehrddad Saif, Canada
Roque J. Saltarén, Spain
Miguel A. F. Sanjuan, Spain
Juan F. San-Juan, Spain
Ilmar Ferreira Santos, Denmark
Nickolas S. Sapidis, Greece
Evangelos J. Sapountzakis, Greece
Themistoklis P. Sapsis, USA
Andrey V. Savkin, Australia
Valery Sbitnev, Russia
Massimo Scalia, Italy
Mohammed Seaid, UK
Mohamed A. Seddeek, Egypt
Mathieu Sellier, New Zealand
Leonid Shaikhet, Ukraine
Cheng Shao, China
Sanjay K. Sharma, India
Bo Shen, Germany
Zhan Shu, UK
Dan Simon, USA
Luciano Simoni, Italy
Christos H. Skiadas, Greece
Delfim Soares Jr., Brazil
Davide Spinello, Canada
Victor Sreeram, Australia
Sri Sridharan, USA
Hari M. Srivastava, Canada
Ivanka Stamova, USA
Rolf Stenberg, Finland
Yuangong Sun, China
Xi-Ming Sun, China
Jitao Sun, China
Zhongkui Sun, China
Andrzej Swierniak, Poland
Wai Yuen Szeto, Hong Kong
Yang Tang, Germany
Alexander Timokha, Norway
Cristian Toma, Romania
Francesco Tornabene, Italy
Antonio Tornambe, Italy
Irina N. Trendafilova, UK

Jung-Fa Tsai, Taiwan
Chia-Cheng Tsai, Taiwan
George Tsiatas, Greece
Efstratios Tzirtzilakis, Greece
Francesco Ubertini, Italy
Kuppalapalle Vajravelu, USA
Robertt A. Valente, Portugal
Alain Vande Wouwer, Belgium
Pandian Vasant, Malaysia
Josep Vehi, Spain
Kalyana C. Veluvolu, Korea
Georgios Veronis, USA
Michael Vynnycky, Ireland
Cheng C. Wang, Taiwan
Yijing Wang, China
Qing-Wen Wang, China
Junwu Wang, China
Shuming Wang, Singapore
Yan-Wu Wang, China

Youqing Wang, China
Yongqi Wang, Germany
Moran Wang, China
Gerhard-Wilhelm Weber, Turkey
Hung-Yu Wei, Taiwan
J. A. S. Witteveen, Netherlands
Kwok-Wo Wong, Hong Kong
Zheng-Guang Wu, China
Yuqiang Wu, China
Dash Desheng Wu, Canada
Huang Xia, China
Gongnan Xie, China
Xuejun Xie, China
Guangming Xie, China
Lianglin Xiong, China
Hang Xu, China
Gen Qi Xu, China
Jun-Juh Yan, Taiwan
Xinggang Yan, UK

Chunyu Yang, China
Dan Ye, China
Peng-Yeng Yin, Taiwan
Mohammad I. Younis, USA
Bo Yu, China
Simin Yu, China
Jianming Zhan, China
Hongbin Zhang, China
Xu Zhang, China
Huaguang Zhang, China
Hong Zhang, China
Qingling Zhang, China
Hongyong Zhao, China
Lu Zhen, China
Jian Guo Zhou, UK
Quanxin Zhu, China
Zexuan Zhu, China
Mustapha Zidi, France

Contents

Modeling, Planning, and Control of Complex Logistic Processes, Hamid R. Karimi, Neil Duffie, Michael Freitag, Michael Lütjen, and Mohammed Chadli
Volume 2015, Article ID 184267, 2 pages

Optimal Scheduling of Logistical Support for Medical Resource with Demand Information Updating, Ming Liu and Yihong Xiao
Volume 2015, Article ID 765098, 12 pages

Simulation-Based Planning and Control of Transport Flows in Port Logistic Systems, Antonio Diogo Passos Lima, Frederico Werner de Mascarenhas, and Enzo Morosini Frazzon
Volume 2015, Article ID 862635, 12 pages

A Multiobjective Fuzzy Aggregate Production Planning Model Considering Real Capacity and Quality of Products, Najmeh Madadi and Kuan Yew Wong
Volume 2014, Article ID 313829, 15 pages

A Transient Queuing Model for Analyzing and Optimizing Gate Congestion of Railway Container Terminals, Ming Zeng, Wenming Cheng, and Peng Guo
Volume 2014, Article ID 914706, 13 pages

Service Capacity Reserve under Uncertainty by Hospital's ER Analogies: A Practical Model for Car Services, Miguel Ángel Pérez Salaverría and José Manuel Mira McWilliams
Volume 2014, Article ID 586236, 18 pages

A Framework to Determine the Effectiveness of Maintenance Strategies Lean Thinking Approach, Alireza Irajpour, Ali Fallahian-Najafabadi, Mohammad Ali Mahbod, and Mohammad Karimi
Volume 2014, Article ID 132140, 11 pages

Demand Management Based on Model Predictive Control Techniques, Yasser A. Davizón, Rogelio Soto, José de J. Rodríguez, Ernesto Rodríguez-Leal, César Martínez-Olvera, and Carlos Hinojosa
Volume 2014, Article ID 702642, 11 pages

Measuring Semantic and Structural Information for Data Oriented Workflow Retrieval with Cost Constraints, Yinglong Ma, Liying Zhang, and Moyi Shi
Volume 2014, Article ID 407809, 9 pages

Optimizing Gear Shifting Strategy for Off-Road Vehicle with Dynamic Programming, Xinxin Zhao, Wenming Zhang, Yali Feng, and Yaodong Yang
Volume 2014, Article ID 642949, 9 pages

A Generalized Minimum Cost Flow Model for Multiple Emergency Flow Routing, Jianxun Cui, Shi An, and Meng Zhao
Volume 2014, Article ID 832053, 12 pages

Effects of Integration on the Cost Reduction in Distribution Network Design for Perishable Products, Z. Firoozi, N. Ismail, S. Ariafar, S. H. Tang, M. K. M. A. Ariffin, and A. Memariani
Volume 2014, Article ID 739741, 10 pages

A Dynamic Optimization Strategy for the Operation of Large Scale Seawater Reverses Osmosis System,
Aipeng Jiang, Jian Wang, Wen Cheng, Changxin Xing, and Shu Jiangzhou
Volume 2014, Article ID 635434, 12 pages

A Study of Network Violator Interception Based on a Reliable Game Model, Shi An, Jianxun Cui,
Meng Zhao, and Jian Wang
Volume 2014, Article ID 124343, 8 pages

Unrecorded Accidents Detection on Highways Based on Temporal Data Mining, Shi An, Tao Zhang,
Xinming Zhang, and Jian Wang
Volume 2014, Article ID 852495, 7 pages

**Research on Coordinated Robotic Motion Control Based on Fuzzy Decoupling Method in Fluidic
Environments,** Wei Zhang, Haitian Chen, Tao Chen, Zheping Yan, and Hongliang Ren
Volume 2014, Article ID 820258, 10 pages

Multiobjective Order Assignment Optimization in a Global Multiple-Factory Environment,
Rong-Chang Chen and Pei-Hsuan Hung
Volume 2014, Article ID 673209, 14 pages

Application of Stochastic Regression for the Configuration of Microrotary Swaging Processes,
Daniel Rippel, Eric Moumi, Michael Lütjen, Bernd Scholz-Reiter, and Bernd Kuhfuß
Volume 2014, Article ID 360862, 12 pages

An Uncertain Programming for the Integrated Planning of Production and Transportation,
Deyi Mou and Xiaoding Chang
Volume 2014, Article ID 419358, 7 pages

Application of Fuzzy Optimization to Production-Distribution Planning in Supply Chain Management,
S. Ariaifar, S. Ahmed, I. A. Choudhury, and M. A. Bakar
Volume 2014, Article ID 218132, 8 pages

**An Integrated Model for Production and Distribution Planning of Perishable Products with Inventory
and Routing Considerations,** S. M. Seyedhosseini and S. M. Ghoreyshi
Volume 2014, Article ID 475606, 10 pages

Study on Application of T-S Fuzzy Observer in Speed Switching Control of AUVs Driven by States,
Xun Zhang, Xinglin Hu, Yang Zhao, and Zhaodong Tang
Volume 2014, Article ID 631396, 8 pages

**A Study of Bending Mode Algorithm of Adaptive Front-Lighting System Based on Driver Preview
Behavior,** Zhenhai Gao and Yang Li
Volume 2014, Article ID 939218, 11 pages

**Robust Production Planning in Fashion Apparel Industry under Demand Uncertainty via Conditional
Value at Risk,** Abderrahim Ait-Alla, Michael Teucke, Michael Lütjen, Samaneh Beheshti-Kashi,
and Hamid Reza Karimi
Volume 2014, Article ID 901861, 10 pages

Adaptability of the Logistics System in National Economic Mobilization Based on Blocking Flow Theory, Xiangyuan Jing, Qingmei Tan, Liusan Wu, and Xiaohui Li
Volume 2014, Article ID 843181, 6 pages

A Multiobjective Optimization Model of Production-Sourcing for Sustainable Supply Chain with Consideration of Social, Environmental, and Economic Factors, Zhixiang Chen and Svenja Andresen
Volume 2014, Article ID 616107, 11 pages

Global Conservative and Multipeakon Conservative Solutions for the Modified Camassa-Holm System with Coupling Effects, Huabin Wen, Yujuan Wang, Yongduan Song, and Hamid Reza Karimi
Volume 2014, Article ID 606249, 17 pages

Objective Attributes Weights Determining Based on Shannon Information Entropy in Hesitant Fuzzy Multiple Attribute Decision Making, Yingjun Zhang, Yizhi Wang, and Jingping Wang
Volume 2014, Article ID 463930, 7 pages

Fault Diagnosis for Compensating Capacitors of Jointless Track Circuit Based on Dynamic Time Warping, Wei Dong
Volume 2014, Article ID 324743, 13 pages

Editorial

Modeling, Planning, and Control of Complex Logistic Processes

**Hamid R. Karimi,¹ Neil Duffie,² Michael Freitag,³
Michael Lütjen,⁴ and Mohammed Chadli⁵**

¹*Department of Engineering, Faculty of Engineering and Science, University of Agder, 4898 Grimstad, Norway*

²*College of Engineering, University of Wisconsin System, Madison, WI, USA*

³*ArcelorMittal Bremen GmbH, Bremen, Germany*

⁴*BIBA-Bremer Institut für Produktion und Logistik GmbH, Planning and Control of Production Systems (PSPS),
University of Bremen, Hochschulring 20, 28359 Bremen, Germany*

⁵*University of Picardie Jules Verne, MIS-UPJV, 7 Moulin Neuf, 80000 Amiens, France*

Correspondence should be addressed to Hamid R. Karimi; hamidrk@uia.no

Received 26 August 2014; Accepted 26 August 2014

Copyright © 2015 Hamid R. Karimi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, complex logistic processes and supply chains have been considered to be among the most important activities in the modern economy and societies, and it is well known that they are subject to many disturbances such as rapid shifts in customer demand, changes in orders, transportation congestion, and communications and machine failures. Obviously, their complexity is becoming increasingly more difficult to manage due to increasing numbers of production facilities linked by expanding global networks of transportation services. There are many different technologies emerging for helping to improve complex supply chain logistics, for example, sensor networks, mass data management, and decentralized control. This special issue targets active research in logistic processes and supply chains to provide an up-to-date overview of the research directions and advanced methods in the field. Of particular interest are the development of mathematical methodologies for modeling, optimization, and control of complex logistics networks. Topics are as follows:

- (i) intelligence techniques, such as fuzzy logic and neural network approaches;
- (ii) intelligent production and logistics systems;
- (iii) effects of uncertainty in production networks;
- (iv) mathematical modeling, robustness, and stochastic optimization of complex logistic processes;
- (v) interrelation of demand forecasting and production planning;

(vi) decentralized control of logistic processes in production and distribution networks;

(vii) applications in complex logistics system.

The special issue attracted submissions from many institutes and countries. There are important results on the topic of production systems; for example, in the work entitled “Robust Production Planning in Fashion Apparel Industry under Demand Uncertainty via Conditional Value at Risk” by A. Ait-Alla et al., a mathematical model for robust production planning is introduced. The model helps fashion apparel suppliers in making decisions concerning allocation of production orders to different production plants characterized by different lead times and production costs and proper time scheduling and sequencing of these production orders. In the paper entitled “An Integrated Model for Production and Distribution Planning of Perishable Products with Inventory and Routing Considerations” by S. M. Seyedhosseini and S. M. Ghoreyshi, a new integrated production and distribution planning model for perishable products is formulated. The proposed model considers a supply chain network consisting of a production facility and multiple distribution centers. In the paper entitled “Application of Fuzzy Optimization to Production-Distribution Planning in Supply Chain Management” by S. Ariaifar et al., a production-distribution model is presented that not only allocates limited available resources and equipment to produce products over the required time

periods, but also determines the most economical distributors for dispatching the products to the distribution centers or retailers. The goal of the paper entitled “An Uncertain Programming for the Integrated Planning of Production and Transportation” by D. Mou and X. Chang is to tackle joint decisions in assigning production and organizing transportation for a single product in a production-transportation network system with multiple manufacturers and multiple demands. The work entitled “A Multiobjective Optimization Model of Production-Sourcing for Sustainable Supply Chain with Consideration of Social, Environmental, and Economic Factors” by Z. Chen and S. Andresen incorporates the three pillars of sustainability, economic, environmental, and social dimensions, into a supply chain. Moreover, a multiobjective programming model that jointly minimizes costs, emissions, and employee injuries in a supply chain is constructed. In the work entitled “Application of Stochastic Regression for the Configuration of Microrotary Swaging Processes” by D. Rippel et al. the cause-effect relationships between relevant process parameters are analyzed using stochastic regression models in order to determine cost-efficient process configurations for the manufacturing of bulk and tubular micro-components. In the paper “Multiobjective Order Assignment Optimization in a Global Multiple-Factory Environment” by R.-C. Chen and P.-H. Hung an effective method is presented for solving the order assignment problem of companies with multiple plants distributed worldwide. A multiobjective genetic algorithm (MOGA) is used to find solutions. To validate the effectiveness of the proposed approach, this study employs real data, provided by a famous garment company in Taiwan, as a base to perform experiments. In the paper “A Generalized Minimum Cost Flow Model for Multiple Emergency Flow Routing” by J. Cui et al., a novel generalized minimum cost flow model is presented for optimizing the distribution pattern of two types of flow in the same network by introducing the conflict cost.

In the context of logistics and transportation, in the paper entitled “A Transient Queuing Model for Analyzing and Optimizing Gate Congestion of Railway Container Terminals” by M. Zeng et al., the queue length and the average waiting time of the railway container terminal gate system, as well as the optimal number of service channels during the different time period, are investigated. In the paper entitled “Optimizing Gear Shifting Strategy for Off-Road Vehicle with Dynamic Programming” by W. Zhang et al., a dynamic programming (DP) algorithm is introduced to optimize the gear shifting schedule for an off-road vehicle by using an objective function that weighs fuel use and trip time. In the paper entitled “Adaptability of the Logistics System in National Economic Mobilization Based on Blocking Flow Theory” by X. Jing et al. a blocking-resistance optimum model and an optimum restructuring model based on blocking flow theories are constructed, both of which are illustrated by numerical cases and compared in characteristics and application. In the paper entitled “Unrecorded Accidents Detection on Highways Based on Temporal Data Mining” by S. An et al., a new method of detecting traffic accidents is proposed based on temporal data mining, which can identify unknown and unrecorded accidents by traffic police.

A time series model was constructed using ternary numbers to reflect the state of traffic flow based on a cellular telephone transmission model. In the work entitled “Service Capacity Reserve under Uncertainty by Hospital’s ER Analogies: A Practical Model for Car Services” by M. A. P. Salaverría and J. M. McWilliams, a capacity reserve model is introduced for dimensioning passenger car service installations according to the demographic distribution of the area to be serviced by using hospital’s emergency room analogies.

Of course, the above topics and papers are not a comprehensive list of those covered by this special issue. Nonetheless, they represent the rich and many-faceted knowledge that we have the pleasure of sharing with the readers.

Acknowledgments

We would like to express appreciation to the authors for their excellent contributions and also the reviewers for their hard work in reviewing the submissions. Meanwhile, Hamid Reza Karimi would like to acknowledge the support of DAAD (Deutscher Akademischer Austauschdienst) program in this research.

*Hamid R. Karimi
Neil Duffie
Michael Freitag
Michael Lütjen
Mohammed Chadli*

Research Article

Optimal Scheduling of Logistical Support for Medical Resource with Demand Information Updating

Ming Liu and Yihong Xiao

Department of Management Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

Correspondence should be addressed to Ming Liu; liumingsu@126.com

Received 1 April 2014; Revised 5 August 2014; Accepted 5 August 2014

Academic Editor: Michael Lütjen

Copyright © 2015 M. Liu and Y. Xiao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a discrete time-space network model for a dynamic resource allocation problem following an epidemic outbreak in a region. It couples a forecasting mechanism for dynamic demand of medical resource based on an epidemic diffusion model and a multistage programming model for optimal allocation and transport of such resource. At each stage, the linear programming solves for a cost minimizing resource allocation solution subject to a time-varying demand that is forecasted by a recursion model. The rationale that the medical resource allocated in early periods will take effect in subduing the spread of epidemic and thus impact the demand in later periods has been incorporated in such recursion model. A custom genetic algorithm is adopted to solve the proposed model, and a numerical example is presented for sensitivity analysis of the parameters. We compare the proposed medical resource allocation mode with two traditional operation modes in practice and find that our model is superior to any of them in less waste of resource and less logistic cost. The results may provide some practical guidelines for a decision-maker who is in charge of medical resource allocation in an epidemics control effort.

1. Introduction

Over the past few years, the world has grown increasingly concerned about the threat of different epidemics. Disastrous epidemic events such as SARS and H1N1 significantly impacted people's life. The outbreak of infections in Europe is another recent example. The infection, from a strain of *Escherichia coli*, can lead to kidney failure and death and is difficult to treat with antibiotics. It is now widely recognized that a large-scale epidemic diffusion can conceivably cause many deaths and more people of permanent sequela, which presents a severe challenge to local or regional healthcare systems.

After an epidemic outbreak, public officials are faced with many critical issues, the most important of which being how to ensure the availability and supply of medical resource so that the loss of life may be minimized and the rescue operation efficiency maximized. The medicine logistics in an epidemics controlling system is often complex and difficult. Hu et al. [1] compared public-health management mechanisms in both USA and China from the following three aspects, organizational structure, management system, and logistics network, and pointed out some deficient areas in the Chinese

public-health management mechanism. To date, medicine logistics operation in epidemic control activities in China has traditionally been done unsystematically and separately, based on the decision-makers' experience and disregarding the interrelationship between the time-varying demand and the logistics operation planning from a systematic perspective. Thus, this paramount life-saving and costly logistics problem opens up a wide range of applications of Operations Research/Management Science techniques and has motivated many recent research works.

In this paper, a time-space network model for the medical resource allocation problem in controlling epidemic diffusion is proposed. It couples a forecasting mechanism for the dynamic demand of the medical resource based on the epidemic diffusion pattern of susceptible-exposed-infected-recovered (SEIR) model [2] and a multistage programming model for optimal allocation and transport of such resource. The two dynamic processes are woven together and interactively proceed to model the epidemic diffusion and the medical resource allocation. Particularly, given the dynamic demand for the medical resource at each stage predicted by the forecasting mechanism, the linear programming problem

solves for the cost minimizing resource allocation pattern subject to related operating constraints. The optimal solution of the resource allocation will then determine their availability at each emergent district hospital, upon which the efficiency of rescuing effort is conditioned (assuming the other needed healthcare technologies and human resource are guaranteed). The efficiency of the rescuing effort will determine the recovering rate of the infected population, which, in turn, will generate the new forecast of the demand for medical resource by updating the SEIR diffusion model. The above described model is expected to be an effective decision-making tool that can help improve the efficiency of medicine logistics when an epidemic outbreaks. To the best of our knowledge, the dynamic and interactive optimization process has never been reported in the existing works.

The remainder of the paper is organized as follows: Section 2 is the literature review. Section 3 introduces the time-space network model, which combines a time-varying demand forecast model based on the epidemic diffusion rule, and a multistage programming model for cost minimizing allocation of the medical resource. The solution procedure for the optimization model is proposed in Section 4. A numerical example and a short sensitivity analysis are presented in Section 5. Finally, Section 6 discusses the limitations of the proposed approach and suggests future research directions.

2. Literature Review

Considering the relationship between the epidemic diffusion and the associated medical resource allocation, we review two streams of recent research efforts here: one is focused on the epidemic diffusion modeling and the other is related to the medical resource allocation modeling.

2.1. Epidemic Diffusion Modeling. Most analytical works on epidemic diffusion are concentrated on the compartmental epidemic models described by ordinary differential equations [3–5]. In these models, the total population is divided into several classes and each class of people is closed into a compartment. The sizes of the compartments are assumed to be large enough and the mixing of members to be homogeneous.

The second stream of research is on the development of epidemic diffusion models by applying complex network theory to the traditional compartment models [6–8]. Recently, Jung et al. [9] extended the previous studies on the prevention of the pandemic influenza to evaluate time-dependent optimal prevention policies, and they found that the quarantine policy was very important and more effective than the elimination policy, during the disease spread. Wang et al. [10] presented some suggestions for the epidemic prevention and infection control in the Wenchuan earthquake areas, Sichuan Province, China.

The third stream of research is on the development of epidemic diffusion models by applying simulation methods, including computer simulation and numerical computation [11–13]. For example, Samsuzzoha et al. [14] used a diffusive epidemic model to describe the transmission of influenza. The equations were solved numerically by using the splitting method under different initial distribution of population

density. Further, Samsuzzoha et al. [15] presented a vaccinated diffusive compartmental epidemic model to explore the impact of vaccination as well as diffusion on the transmission dynamics of influenza.

Recently and importantly, a robust data-driven fault detection approach is proposed with application to a wind turbine benchmark [16]. The main challenges of the wind turbine fault detection lie in its nonlinearity, unknown disturbances, and significant measurement noise. Sometimes the relative data may be missed [17, 18]. These works are constructive and helpful to understand and model the epidemic diffusion process in a very different way.

The above mentioned works represent some of the research on various differential equation models for epidemic diffusion and control. Although the emphasis of this paper is on the efficient allocation of medical resource, a basic component of our model, the forecasting mechanism for their dynamic demand, utilizes one of such epidemic diffusion models.

2.2. Medical Resource Allocation Modeling. To the best of our knowledge, a great deal of researches has been published with the topic on optimal allocation of medical resource [19–23]. To optimize the process of materials distribution in an epidemic diffusion system and to improve the distribution timeliness, Liu and Zhao [24] modeled the emergency materials distribution problem as a multiple traveling salesman problem with time window. Wang et al. [25] constructed a multiobjective stochastic programming model with time-varying demand for the emergency logistics network based on the epidemic diffusion rule. A genetic algorithm coupled with Monte Carlo simulation was adopted to solve the optimization model. Qiang and Nagurney [26] proposed a humanitarian logistic model for supply/distribution of critical needs in a disruption caused by a natural disaster. They considered a general network structure and disruptions that may have an impact on both network link capacities and product demand. The problem was studied in a bicriteria system optimization framework for network performance. Recently, Rachaniotis et al. [27] presented a deterministic resource scheduling model in epidemic control. In their work, a deterministic model, appropriate for large populations, where random interactions could be averaged out, was used for the epidemic's rate of spread. Besides, a case of the mass vaccination against H1N1 influenza in the Attica region, Greece, and a comparative study of the model's performance versus the applied random practice were presented.

To deal with the complexity and difficulty in solving the medical resource allocation problem, we observe a trend in solution methodologies, that is, decomposing the original problem, which can be a multicommodity, multimodal, or multiperiod model, into several mutually correlated subproblems, and then solve them systematically in same decision scheme. For instance, Barbarosolu et al. [28] proposed a bilevel hierarchical decomposition approach for helicopter mission planning during a disaster relief operation. The top-level model was formulated to deal with the tactical decisions, covering the issues of helicopter fleet management, crew assignment, and the number of tours undertaken by

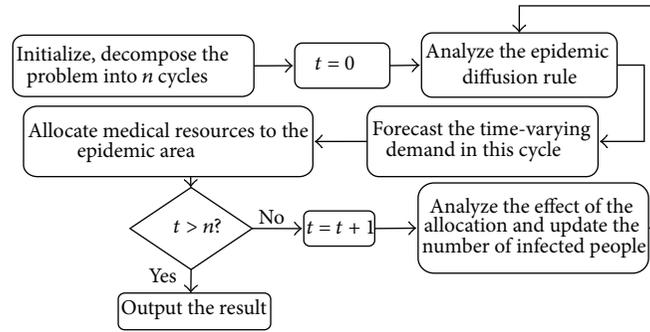


FIGURE 1: Operational procedure of the dynamic medicine logistics network.

each helicopter. The base-level model aimed to address the corresponding operational decisions, including routing, loading/unloading, and refueling scheduling. References [25, 29] were more recent works following this line.

Furthermore, we note that most of the previous works were carried out under the assumption that the relief demand is not time sensitive. While in reality, the demand for medical resource is dynamic, and the medical resource allocated in early cycles will affect the demand in later periods. In this paper, we will use a discrete time-space network to model the medical resource allocation problem when an epidemic outbreaks. In each decision cycle, the problem is constructed as a linear programming model to solve for the cost minimizing allocation solution subject to the time-varying demand that is predicted by the epidemic diffusion rule. As such, this paper attempts to bridge the two streams of literature, the epidemic diffusion and the medicine logistics, which were studied separately in existing literature.

3. The Mathematical Model

Epidemic diffusion process can be divided according to its development into three stages [30]. The first stage is the inception of the epidemic in very limited population, which if noticed in time and treated properly can be controlled effectively without causing a wide spread. In the second stage, the epidemic has broken out into a widespread diffusion. An important part of epidemic control and rescue campaign is to ensure the timely delivery of the needed medical resource according to the dynamic demand as determined by the progress of the epidemic spread. In the third stage, the epidemic diffusion has been controlled and the demand for medical resource has significantly declined. Liu et al. [31] proposed a model for studying medical resource distribution in the first stage. In this paper, we will concentrate on the logistics problem of medical resource allocation in the second stage. Particularly, we will study how the area distribution centers (ADC) should supply the district distribution centers (DDC) and how the DDCs should deliver the needed medical resource to the emergency designated hospitals (EDH) in the most efficient and cost-effective way. Here we assume there are several ADCs in the epidemic spread area, which can be divided into several municipal districts or towns. Each district will have one or more DDCs which supply the needed medical resource to the EDHs in that district.

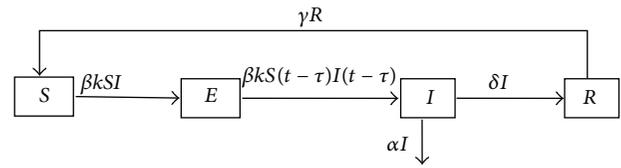


FIGURE 2: SEIRS model based on a small-world network.

Since demand from each EDH is determined by the number of patients hospitalized there and varies according to the progress of epidemic diffusion, the allocation of medical resource need to chase the demand over time. Figure 1 below gives a diagram of operations outlining the execution of the proposed model. The sequential operational routine continues until the epidemic diffusion gets under control. As Figure 1 shows, medical resource allocation process is decomposed into n decision-making cycles. Each decision-making cycle includes three phases: epidemic diffusion analysis, demand forecasting, and medical resource allocation. These three phases are executed iteratively. The effect of the medical resource allocation is analyzed, and the number of infected people is updated at each cycle during the entire distribution process.

In the sequel, we will introduce SIERS model, a well-recognized epidemic diffusion model, in Section 3.1, propose a forecasting model for the dynamic demand for the medical resource during the epidemic diffusion in Section 3.2, and a linear programming model for distribution of medical resource according to the forecasted dynamic demand.

3.1. SEIRS Epidemic Diffusion Model. The SEIR model has been widely adopted by researchers to study epidemic diffusion. It is based on small-world network theory and provides a good match to the actual social network [2]. Generally, the total population is divided into four classes, susceptible people (S), exposed people (E), infected people (I), and recovered people (R), and each class of people is closed into a compartment. Tham [32] showed that some of the recovered people who were discharged from hospitals might be reinfected. Figure 2 shows, without consideration of migration, the natural birth rate and death rate of the population, the epidemic process can be described by a SEIRS model based on a small-world network [25].

The dynamic system for the SEIRS diffusion model can be rewritten by the following ordinary differential equations:

$$\begin{aligned} \frac{dS}{dt} &= -\beta k S(t) I(t) + \gamma R(t), \\ \frac{dE}{dt} &= \beta k S(t) I(t) - \beta k S(t-\tau) I(t-\tau), \\ \frac{dI}{dt} &= \beta k S(t-\tau) I(t-\tau) - (\alpha + \delta) I(t), \\ \frac{dR}{dt} &= \delta I(t) - \gamma R(t). \end{aligned} \quad (1)$$

In the above system of equations, $S(t)$, $E(t)$, $I(t)$, and $R(t)$ represent, respectively, the number of susceptible people, the number of exposed people, the number of infected people, and the number of recovered people. k is the average degree of distribution for this small-world network, which can be interpreted as the average contact number of susceptible people of each infected person; β is the propagation coefficient of the epidemic; γ is the rate of the recovered people who are not immune and thus may be reinfected; δ is the recovery rate; α is the death rate; τ represents the incubation period of the disease. Consider $k, \beta, \gamma, \delta, \alpha, \tau > 0$.

ODE (1) states the following: (i) the growth rate of the susceptible population is determined by the returning population who are recovered but not immune and the losing population who actually get exposed to the disease and thus are counted towards the class of $E(t)$. The latter is in proportion to the propagation coefficient β , the average contact number of susceptible people of each infected person, k , and both of the current mass of the susceptible population and the current mass of the infected population. (ii) The growth rate of the exposed population is determined by the difference between the entering population, those of susceptible people who actually get exposed to the disease, and the exiting population, those of exposed population who get sick after the incubation period of the disease; (iii) the growth rate of the infected population is determined by the difference between the entering population, those of exposed population who get sick, and the exiting population who are either recovered or dead; and, finally (iv) the growth rate of the recovered population is determined by the difference between the joining population of the newly recovered and the losing population of the reinfected people.

Particularly, as we noted in (iii), the number of infected people, $I(t)$, is determined by the population of the recovered people and the onset exposed people at the end of the incubation period. Hence, improving the recovery rate, δ , and reducing the propagation coefficient, β , are the two effective measures to take in suppressing the growth of $I(t)$. In the context of epidemic controlling operation, that means sufficient medical resource should be allocated to the emergent designated hospitals (EDH).

3.2. The Forecasting Model for the Time-Varying Demand. Demand for medical resource has been studied in a variety of forms in the literature, such as a time-varying value [33] or obeying some stochastic distribution [29]. However,

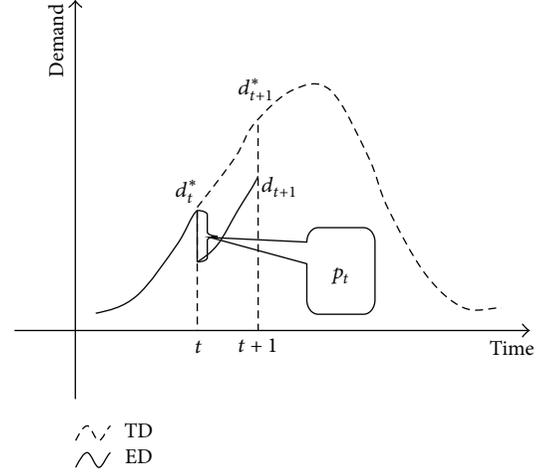


FIGURE 3: Diagrammatic sketch of the time-varying demand.

the impact of earlier resource allocation to the demand in later periods has basically been ignored in these approaches.

To address this deficiency, we propose the following linear relationship between the demand for medical resource and the number of infected people at time t based on the SEIRS epidemic diffusion model:

$$d_t = aI(t), \quad (2)$$

where d_t refers to the demand for medical resource at time t and a is the proportionality coefficient. In our interviews with the public-healthcare administrative personnel about controlling the epidemic spread, we found this linear forecasting function is the one they commonly adopted. Here we define it as the traditional demand (TD). However, a lag effect of earlier medicine allocation should be taken into account in the current demand forecast. As shown in Figure 3, the horizontal axis represents the decision-making cycle, and the vertical axis stands for demand in an epidemic area. The dotted line is a trajectory of (2), and the solid curve is the expected demand (ED). For instance, if the demand for medical resource at cycle t is d_t^* , and according to (2), the demand at cycle $t+1$ would have been d_{t+1}^* . However, a certain amount of medical resource, p_t , had been allocated to the disaster area during cycle t , and it would be taking effect in cycle $t+1$ in curing the infected patients in hospitals and thus subduing the diffusion. Hence, the expected demand for medical resource at cycle $t+1$ should be d_{t+1}^* , instead of d_{t+1}^* .

The following growth factor is introduced by the above observation to account for the lag effect:

$$\eta_t = \frac{d_{t+1}^* - d_t^*}{d_t^*}. \quad (3)$$

Herein, the growth factor η_t can be either positive (increasing demand) or negative (decreasing demand) and may vary in different cycles for the different demand d_t^* . As mentioned before, part of the recovered people who are discharged from the healthcare department may be reinfected. Thus, we define the effective cure rate as θ as the percent

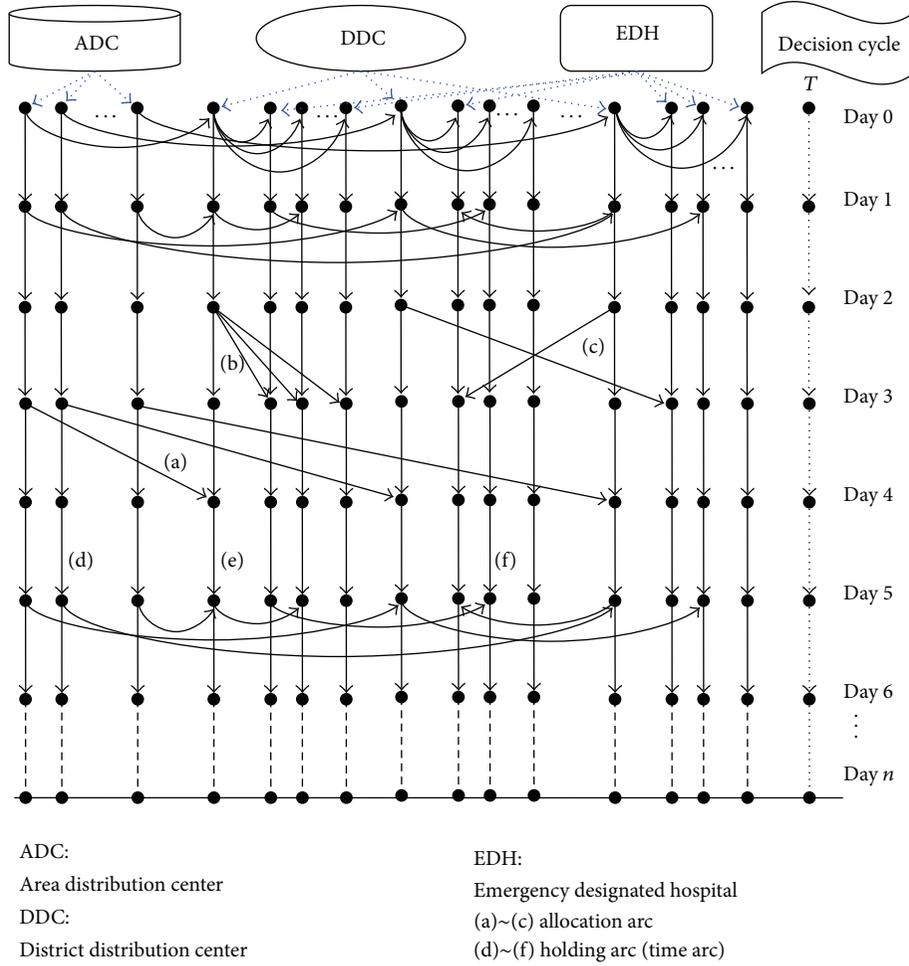


FIGURE 4: Time-space network of medical resource allocation.

of recovered people who are not reinfected. Considering that each infected person needs a period of time to receive treatment and get cured, herein we denote the treatment cycle as Γ and we assume it to be an integral multiple of the decision cycle. Then, the commuted effective cure rate in each decision cycle can be obtained as θ/Γ . Such an assumption would be feasible if the decision cycle is small enough, for example, one day. Hence, it helps us get the following recursion formulas:

$$\begin{aligned}
 \text{when } t = 1, \quad d_1 &= (1 + \eta_0) \left(1 - \frac{\theta}{\Gamma}\right) d_0; \\
 \text{when } t = 2, \quad d_2 &= (1 + \eta_1) \left(1 - \frac{\theta}{\Gamma}\right) d_1 \\
 &= (1 + \eta_0)(1 + \eta_1) \left(1 - \frac{\theta}{\Gamma}\right)^2 d_0; \quad (4) \\
 &\vdots \\
 \text{when } t = n, \quad d_n &= \prod_{i=0}^{n-1} (1 + \eta_i) \left(1 - \frac{\theta}{\Gamma}\right)^n d_0.
 \end{aligned}$$

Herein, $\prod_{i=0}^{n-1} (1 + \eta_i) = (1 + \eta_0)(1 + \eta_1) \cdots (1 + \eta_{n-1})$. $d_0 = aI(0)$ is the initial demand for medical resource in the epidemic area, and $I(0)$ is the initial number of infected people in the epidemic area. Recursion formulas (4) are our prescribed forecast model for the demand of medical resource. In what follows, we will propose a medicine logistics operation model to minimize the total allocation cost based on the forecasting model.

3.3. Time-Space Network of the Medicine Logistics. In this subsection, a multistage programming model for cost minimizing allocation of the medical resource is built upon a time-space network. Figure 4 is the schematic diagram of the network. The vertical axis represents the time duration. The horizontal axis represents the area distribution center (ADC), the district distribution center (DDC), and the emergency designated hospital (EDH), respectively. The allocation arcs are defined as follows: (a) represents that medical resource is transported from ADC to DDC; (b) stands for that medical resource is allocated from DDC to EDH in the same district; (c) refers to that medical resource is allocated from DDC to EDH in the other district; (d)~(f) are time duration arcs for different departments.

3.3.1. Assumptions. The following assumptions are needed to facilitate the model formulation in the following sections.

- (1) In the event of an epidemic outbreak, it is paramount for the government and the entire society to control the spread and rescue the infected. Thus it is reasonable to assume that the government can ensure the adequate supply of the needed medical resource either from domestic pharmaceutical companies or imported. Hence, there is enough medical resource in ADCs during the entire operation process.
- (2) Once an epidemic outbreak, the government will take strict control measures so that each epidemic area can be isolated from other areas to avoid the cross spread of the disease. In each epidemic area, the government will appoint a hospital to be the EDH, to be responsible for the rescue work in such an isolated area.
- (3) Medical resource in this paper is an assembled product, which may include water, vaccine, antibiotic, and so forth.

3.3.2. Notations. Notations used in the following programming model are specified as follows:

cc_{ij} : unit transportation cost of medical resource from ADC i to DDC j ,

cr_{ij} : unit transportation cost of medical resource from DDC i to EDH j ,

es_{it} : the available quantity of medical resource in ADC i in decision cycle t ,

zr_{it} : quantity of medical resource allocated to DDC i in decision cycle t ,

x_{ijt} : medical resource transported from ADC i to DDC j in decision cycle t ,

y_{ijt} : medical resource transported from DDC i to EDH j in decision cycle t ,

d_{it} : demand for medical resource in EDH i in decision cycle t ,

T : set of decision cycles,

C : set of ADCs,

R : set of DDCs,

H : set of EDHs.

3.3.3. Model Formulation. Let $F(x, y)$ be the objective function of the total cost of medical resource allocation. Based

on the above assumptions and descriptions, the proposed problem can be formulated as follows:

$$\text{Min} \quad F(x, y) = \sum_{t \in T} \sum_{i \in C} \sum_{j \in R} x_{ijt} cc_{ij} + \sum_{t \in T} \sum_{i \in R} \sum_{j \in H} y_{ijt} cr_{ij} \quad (5)$$

$$\text{s.t.} \quad \sum_{i \in C} x_{ijt} = zr_{jt}, \quad \forall j \in R, t \in T \quad (6)$$

$$\sum_{j \in R} x_{ijt} \leq es_{it}, \quad \forall i \in C, t \in T \quad (7)$$

$$\sum_{i \in R} y_{ijt} = d_{jt}, \quad \forall j \in H, t \in T \quad (8)$$

$$\sum_{j \in H} y_{ijt} \leq zr_{it}, \quad \forall i \in R, t \in T \quad (9)$$

$$d_{i0} = aI_i(0), \quad \forall i \in H \quad (10)$$

$$d_{it} = \prod_{t=0}^{t-1} (1 + \eta_{it}) \left(1 - \frac{\theta}{\Gamma}\right)^t d_{i0}, \quad (11)$$

$$\forall i \in H, t \in \{T, t \neq 0\}$$

$$x_{ijt} \geq 0, \quad \forall i \in C, j \in R, t \in T \quad (12)$$

$$y_{ijt} \geq 0, \quad \forall i \in R, j \in H, t \in T. \quad (13)$$

In this optimization model, x_{ijt} and y_{ijt} are the decision variables. The objective function (5) is to minimize the total cost of medical resource allocation, which is the transportation cost for delivering the medical resource from ADCs to DDCs and from DDCs to EDHs. Constraints (6)–(9) are the flow conservation equations. Particularly, constraint (6) suggests that each DDC can obtain medical resource from all ADCs. Constraint (7) ensures that the total shipments from any ADC cannot exceed the available amount of the resource in this ADC. Constraint (8) states that the period demand generated by the forecasting model in Section 3.2 at each EDH must be satisfied. That is, the shipments from all DDCs to each EDH must be equal to the demand at this EDH. Constraint (9) implies that the total shipments from any DDC cannot exceed the available stock in this DDC. Constraints (10)–(11) are forecasting model for the time-varying demand (Section 3.2). Herein, η_{it} is the growth factor (can be either positive or negative) of the demand for medical resource in EDH i in decision cycle t . Finally, (12) and (13) are the nonnegativity of the flows. Such model is a dynamic and multistage programming model.

4. Solution Methodology

To solve the above optimization model, (10)–(12) are adopted to calculate the time-varying demand firstly. After that, $\forall t \in T$, the research model can be converted as a two-stage linear programming model. The feature of such a two-stage programming problem is that both the input quantity and the output quantity of the medical resource in the

TABLE 1: Values of parameters in the SEIRS model.

	ADC 1				ADC 2			
	DDC 1		DDC 2		DDC 3		DDC 4	
	EDH 1	EDH 2	EDH 3	EDH 4	EDH 5	EDH 6	EDH 7	EDH 8
$S(0)$	5×10^3	4.5×10^3	5.5×10^3	5×10^3	6×10^3	4.8×10^3	5.2×10^3	4×10^3
$E(0)$	30	35	30	40	25	40	50	45
$I(0)$	5	6	7	8	4	7	9	10
$R(0)$	0							
β	5×10^{-5}							
$\langle k \rangle$	6							
δ	0.3							
d	1×10^{-3}							
γ	1×10^{-3}							
τ	5							

DDCs are unknown. There are many available techniques for solving such a problem, and a genetic algorithm is commonly used. Hence, a genetic algorithm coupled with MATLAB 7.0 mathematical programming solver is adopted to solve the model.

4.1. Chromosome Coding and Population Initializing. The first step of a genetic algorithm is to define the coding method of the chromosome. As is well known, the real number coding is superior to the binary coding in both aspects of quality and efficiency of the solution. Besides, the real number coding is closer to the actual problem findings and easier to interpret in the real world problem. Herein, the real number coding is adopted. For $\forall t \in T$, each chromosome contains R bit gene, where R is the number of DDC. The value of each bit refers to the available amount of medical resource in each DDC, which is also the quantity of medical resource replenished from all ADCs. Each individual in the initial population is generated by a random method, subject to the related resource constraints in the programming model.

4.2. Fitness Definition. The fitness of each individual is obtained by computing the objective function

$$F(x, y) = \sum_{t \in T} \sum_{i \in C} \sum_{j \in R} x_{ijt} cc_{ij} + \sum_{t \in T} \sum_{i \in R} \sum_{j \in H} y_{ijt} cr_{ij}. \quad (14)$$

Herein, the fitness function contains two parts. The first part is the total transportation cost between ADCs and DDCs. The second part is the total transportation cost between DDCs and EDHs. Obviously, the lower the total cost is, the better the fitness of the individual is.

4.3. Selection Operator. The best individual copy strategy is adopted in selection section. That means, each time when selection operator is iterated, the worst chromosome in the population will be replaced by the best one.

4.4. Crossover Operator. A crossover operator is one of the most important operators in a genetic algorithm. Different crossover operators are suitable for different kinds of chromosomes. According to the real number coding in this paper,

an arithmetic crossover is adopted. Let P_1 and P_2 represent the two parent chromosomes, and P_{c1} and P_{c2} stand for the two children chromosomes, respectively. The linear relationship between the parent and the children chromosomes can be formulated as follows:

$$\begin{aligned} P_{c1} &= \mu P_1 + (1 - \mu) P_2, \\ P_{c2} &= (1 - \mu) P_1 + \mu P_2. \end{aligned} \quad (15)$$

Herein, $\mu = U(0, 1)$ is a uniform random number between 0 and 1. Note that both of these two children chromosomes automatically satisfy the resource constraints in the multistage programming model. The range of the crossover probability p_c is 0.2~0.8.

4.5. Mutation Operator. A mutation operator is intended to simulate genetic mutation during biological evolution. Mutation is operated on some bits of individuals at a probability of p_m . This probability is generally very small and is set in the range $0.001 \leq p_m \leq 0.1$. When mutating, we exchange a pair of genes in the individual.

4.6. Termination Condition. As the optimal result is unpredictable, a max iteration is given for the termination.

5. Numerical Tests

5.1. A Numerical Example. We present a numerical example to illustrate the efficiency of the proposed model. Assume there is a smallpox outbreak in a city, which has two ADCs and four DDCs. Two hospitals are designated in each district, and each EDH can service a certain amount of patients. The values of the parameters in the epidemic diffusion model are given in Table 1.

Figure 5 depicts a numerical simulation of the epidemic model at EDH1 in this effected region. The four curves, respectively, represent the number of four groups of people (S, E, I, R) over time. As mentioned in Section 3, the process of epidemic diffusion is divided into three stages and our work in this paper is focused on the second stage. According to Figure 5, such a stage can be ranged from the 10th day

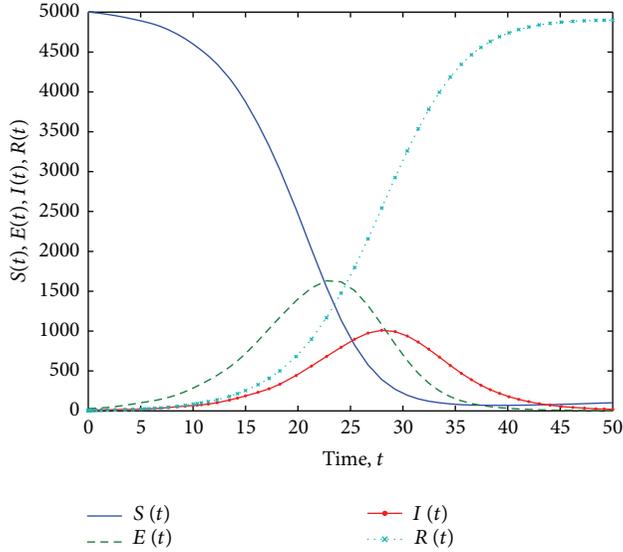


FIGURE 5: Solution of the SEIRS epidemic diffusion model (EDH1).

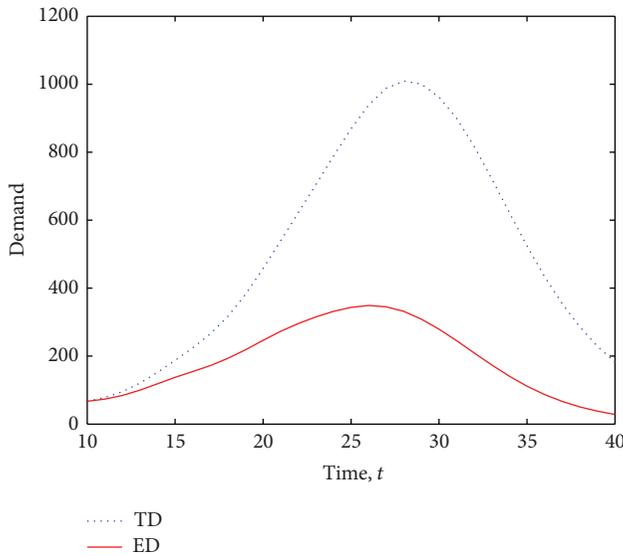


FIGURE 6: Demand in EDH1 by the two different methods.

(decision-making cycle $t = 0$) to the 40th day (decision-making cycle $t = 30$). Of course, when different emergency outbreak happens, the result can be adjusted correspondingly.

To facilitate the calculate process in the following sections, the decision-making cycle is assumed to be one day. Let $a = 1$; MATLAB 7.0 mathematical programming solver coupled with (1) and (2) is adopted to calculate the TD for medical resource in each EDH. Furthermore, given that $\theta = 90\%$ and $\Gamma = 15$ (days), the growth factor η_{it} in each decision-making cycle can be obtained. Then, the ED for medical resource in each EDH in each decision cycle can be forecasted according to (12)-(13). Taking EDH1 as an example, the demand for medical resource in each decision-making cycle by these two different methods is compared as shown in Figure 6.

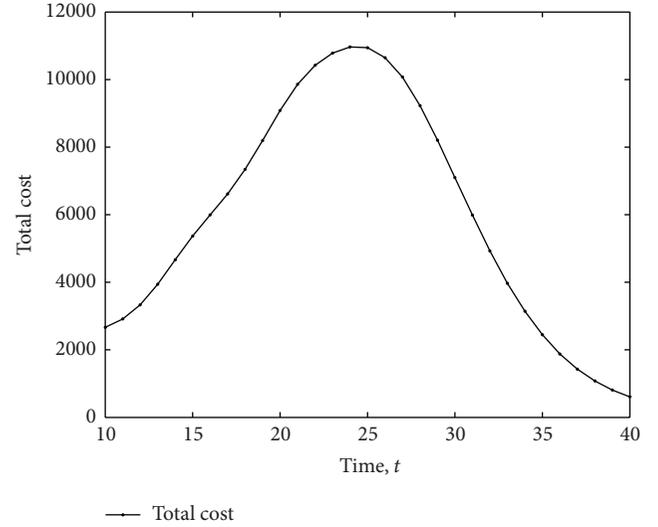


FIGURE 7: Total cost in each decision-making cycle (Unit: \$).

One can observe in Figure 6 that ED is way below TD, suggesting that the allocation of medical resource in the early periods will significantly reduce the demand in the following periods. The second observation is that both curves exhibit similar trends; namely, the demand will first increase along with the spreading the epidemic and then will decrease after the epidemic is brought under control.

We now proceed to illustrate the optimal allocation of the resource at each EDH. Table 2 shows the unit operation cost of medicine between two different departments.

Let $n = 200$, $p_c = 0.75$, and $p_m = 0.01$. The algorithm is set to terminate in 200 generations. Taking the allocation at cycle $t = 0$ as the initial example, we solve the above programming model (7)–(15) according to the solution procedure. The convergent allocation scheme is reported in Table 3 and the total operation cost is 2663.22.

To test the accuracy and stability of the algorithm, the computation process has been repeated for six times independently. As shown in Table 4, the convergent results in these six times are very close and the deviation is less than 0.065%. This proves the proposed algorithm is stable and accurate. We execute the solution procedure (Table 1) to find the dynamic allocation result of medical resource and show in Figure 7 the total cost at each decision-making cycle.

Comparing Figure 7 with Figure 6, one can find that the curve of the total operation cost matches well the demand curves in their variation pattern, suggesting that the cost of medical resource allocation mainly depends on the demand. The characteristics also reflect the hysteresis effect in an epidemic controlling system that medicine logistics lags behind the epidemic diffusion.

In next subsection, we will compare the proposed model with the two traditional allocation measurements that have been used in practice.

5.2. Model Comparison. Based on our interviews with the public healthcare administrative personnel in China, there are two traditional measurements in practice to predict

TABLE 2: Unit operation cost between two different departments (Unit: \$).

Cost	ADC1	ADC2	EDH1	EDH2	EDH3	EDH4	EDH5	EDH6	EDH7	EDH8
DDC1	3.5	2	1	2	4	2.5	5	5	2.5	1.5
DDC2	1.5	2	2	2.5	2	3	5	4	2.5	2
DDC3	3	1.5	2.5	3	5	2	1	3	1.5	4
DDC4	2.5	3	4	4	1.5	2.5	3	2	2	1.5

TABLE 3: Medical resource allocation result at decision-making cycle $t = 0$ (Unit: \$).

	DDC1	DDC2	DDC3	DDC4				
ADC1	0	186.6901	0	163.759				
ADC2	188.8323	0	191.9134	0				
	EDH1	EDH2	EDH3	EDH4	EDH5	EDH6	EDH7	EDH8
DDC1	67.1588	43.0695	0	31.1464	0	0	0	47.4576
DDC2	0	23.5330	82.9355	22.1931	0	0	31.4635	26.5650
DDC3	0	0	0	38.6671	74.9403	0	78.3061	0
DDC4	0	0	28.1169	14.6546	0	86.9697	18.6328	15.3849

TABLE 4: Total cost in cycle $t = 0$ (Unit: \$).

Run	1	2	3	4	5	6
Total cost	2664.97	2663.22	2663.22	2663.22	2664.97	2663.22

the demand for medical resource in case of an epidemic outbreak. Both of them utilize (2) as the basic forecasting method. In the first traditional measurement, referred as Traditional 1, the medical resource will only be allocated through administrative distribution. That is, an ADC will only service the DDCs in its own area, and a DDC will only service the EDHs in its own district. For instance, as Table 1 shows, ADC 1 will only service DDC1 and DDC2, and DDC1 will only replenish medical resource to EDH1 and EDH2. The second traditional measurement, referred here as Traditional 2, is based on the same forecasting method of (2) but allows cross area distribution. The total costs of these three different models are compared and illustrated in Figure 8.

Several interesting observations can be drawn from Figure 8. First, the total operation costs by model Traditional 2 are all time lower than those by Traditional 1, although the difference is not large, suggesting that cross area distribution has a definite advantage in saving allocation cost, which is of course not surprising from an optimization perspective. Secondly, the two cost curves by Traditional 1 and Traditional 2 behave consistently in their rising and falling trend and arrive at their maximum at the exact same time $t = 26$. This is because these two traditional models are based on the same demand forecasting mechanism for the medical resource, and the allocation cost is mainly determined by the allocation volume, that is, the demand. Thirdly, the cost curve generated by our time-space network model is the all time minimum and much lower than that by the two traditional measurements. The allocation cost generated by our model is obvious less than the traditional methods in most of the time. We attribute this significant cost reduction to our proactive forecasting that takes into account the positive impact of

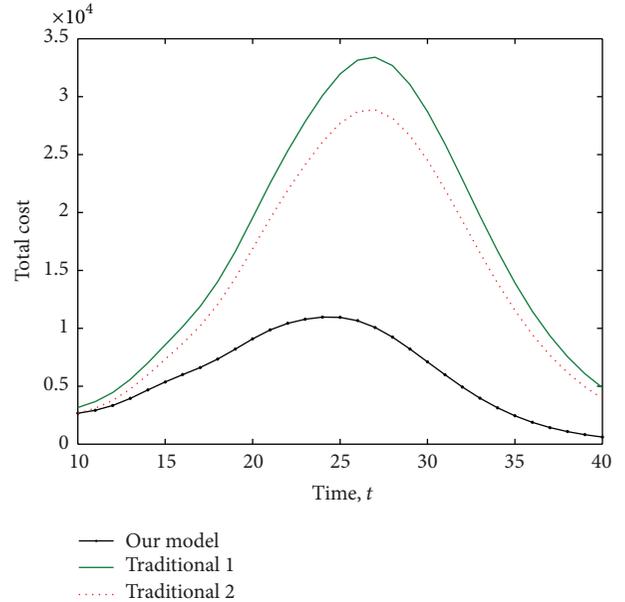


FIGURE 8: Total cost in three different patterns (Unit: \$).

the early allocation of medical resource to the demand in following periods. Finally and most importantly, one can notice that the cost curve by our model reaches its maximum at $t = 24.5$, comparing the two traditional measurements at $t = 26$. This suggests that by using our proposed model we can get control of the epidemic spread earlier, which stands for an invaluable social merit on top of the economic savings. To conclude our findings in this example, the cross area distribution that features our proposed model and Traditional 2 can reduce the logistic part of the allocation cost. The proactive forecasting model coupled with our time-space optimal allocation programming proposed in this paper can subdue the epidemic diffusion and thus significantly reduce the demand for the medical resource, resulting greater saving in the total operation cost.

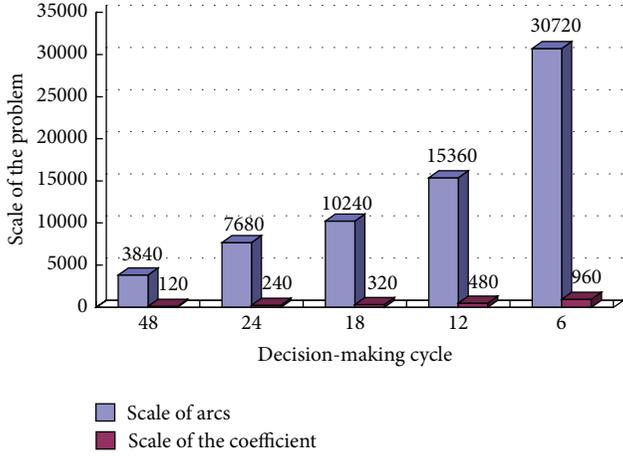


FIGURE 9: Relationship between the scale of problem and decision-making cycle.

5.3. *Sensitivity Analysis.* In this section, a sensitivity analysis of the three key parameters (η , θ and Γ) in the time-varying demand forecast model is conducted. According to the definition in Section 3.2, the parameter η is closely related to the decision-making cycle. In this paper, the decision-making cycle is set to be one day (24 hours); so we get a total of 240 η and 7680 arcs in the experiment. Figure 9 shows the relationship between the scale of the problem and the decision-making cycle (Unit: hour).

In practice, the decision-maker can choose the decision-making cycle according to the actual situation. Generally speaking, the shorter the cycle is, the better the forecast accuracy is, but the larger the scale of the problem and its complexity is. On the other hand, if the decision-making cycle is set too short to let the actual distribution operations uncomplete, then the accuracy of the model might be adversely affected. Therefore, the decision-making cycle should be selected appropriately in a practical problem.

As total cost of the proposed medicine logistics network mainly depends on the demand for medical resource, these two variables get a similar variation tendency. Thus, taking EDH1 as an example, we can hold all the parameters fixed, as in the numerical example given in Section 5.1, and let θ and Γ take on five different values, respectively. The demand for medical resource in each decision-making cycle is shown in Figures 10 and 11.

As Figure 10 shows, θ takes on five values ranging from 60% to 100%. The larger θ is, the lower the demand is. Accordingly, the lower the total cost would be. As Figure 11 shows, Γ takes on five values ranging from 10 to 20. The shorter Γ is, the lower the demand is, and thus the lower the total cost would be. The above analysis confirms that both of these two key parameters play important roles in medical resource allocation decisions. For a small change of θ and Γ , the final allocation decisions and the total operation cost in each cycle can change significantly. Unfortunately, the precise values for these two parameters in an epidemic control are difficult to get. As the accuracy of these two parameters is vital to the success of medicine logistics operation, it calls for more

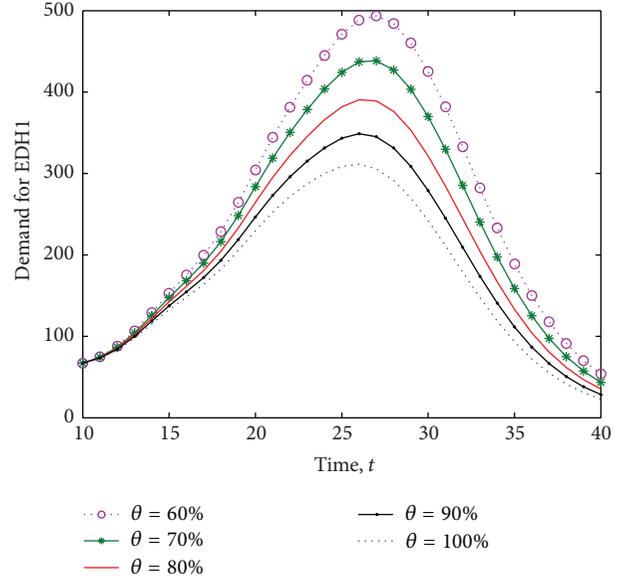


FIGURE 10: Demand in EDH1 with different value of θ .

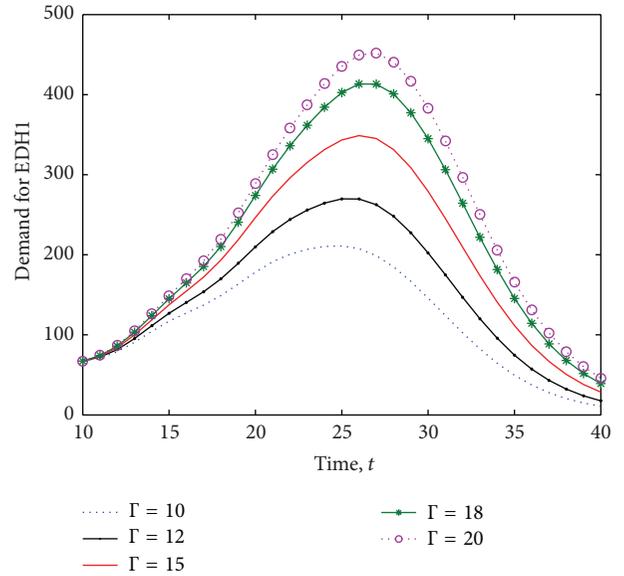


FIGURE 11: Demand in EDH1 with different value of Γ .

research work to scientifically estimate these two parameters for different epidemics.

6. Conclusions

In this paper, we develop a discrete time-space network model to study the medical resource allocation problem in an epidemic outbreak. In each decision-making cycle, the allocation of medical resource across the region from ADCs through DDCs to EDHs is determined by a linear programming model with the dynamic demand that is forecasted by an epidemic diffusion rule. The novelty of our model

against the existing works in literature is characterized by the following three aspects.

- (1) While most research on medical resource allocation studies a static problem which takes no consideration of the time evolution and dynamic nature of the demand, the model proposed in this paper addresses a time-series demand that is forecasted in match of the course of an epidemic diffusion.
- (2) The model couples a multistage linear programming for optimal allocation of medical resource with a proactive forecasting mechanism cultivated from the epidemic diffusion dynamics. The two dynamic processes are woven together and interactively proceed to model the epidemic diffusion and the medical resource allocation. The rationale that the medical resource allocated in early periods will take effect in subduing the spread of the epidemic spread and thus impact the demand in later periods has been for the first time incorporated into our model.
- (3) The computational results show that the proposed model remarkably outperforms the traditional measurements in both terms of cost reduction and epidemic control. Our model can significantly reduce the total operation cost of the medical resource allocation and may get the epidemic diffusion in control earlier than the traditional measurements.

Furthermore, the medicine logistics operation problem has been decomposed into several mutually correlated sub-problems and then been solved systematically in the same decision scheme. Thus, the result will be much more suitable for a real operation. As the limitation of the model, it is developed for the medical resource allocation in a geographic area where an epidemic disease has been spreading and it does not consider possible cross area diffusion between two or more geographic areas. We assume that once an epidemic outbreak exists, the government has effective means to separate the epidemic areas so that cross area spread can be basically prevented. However, this cannot always be guaranteed in reality.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work has been partially supported by the National Natural Science Foundation of China (no. 71301076), the MOE Project of Humanities and Social Sciences (no. 11YJCZH109), Natural Science Foundation of Jiangsu Province (BK20130771), and the Research Fund for the Doctoral Program of Higher Education of China (20133219120037). The authors also wish to thank Professor Ding Zhang from State University of New York, Oswego (USA), who read earlier versions of this paper. His helpful suggestions improved the presentation significantly.

References

- [1] J. Hu, A. Z. Zeng, and L. Zhao, "A comparative study of public-health emergency management," *Industrial Management and Data Systems*, vol. 109, no. 7, pp. 976–992, 2009.
- [2] J. Zhang and Z. Ma, "Global dynamics of an SEIR epidemic model with saturating contact rate," *Mathematical Biosciences*, vol. 185, no. 1, pp. 15–32, 2003.
- [3] B. K. Mishra and D. K. Saini, "SEIRS epidemic model with delay for transmission of malicious objects in computer network," *Applied Mathematics and Computation*, vol. 188, no. 2, pp. 1476–1482, 2007.
- [4] C. Sun and Y. Hsieh, "Global analysis of an SEIR model with varying population size and vaccination," *Applied Mathematical Modelling*, vol. 34, no. 10, pp. 2685–2697, 2010.
- [5] J. Zhang, J. Li, and Z. Ma, "Global dynamics of an SEIR epidemic model with immigration of different compartments," *Acta Mathematica Scientia B: English Edition*, vol. 26, no. 3, pp. 551–567, 2006.
- [6] J. Saramäki and K. Kaski, "Modelling development of epidemics with dynamic small-world networks," *Journal of Theoretical Biology*, vol. 234, no. 3, pp. 413–421, 2005.
- [7] X. Xu, H. O. Peng, X. M. Wang, and Y. H. Wang, "Epidemic spreading with time delay in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 367, pp. 525–530, 2006.
- [8] X. P. Han, "Disease spreading with epidemic alert on small-world networks," *Physics Letters A*, vol. 365, no. 1–2, pp. 1–5, 2007.
- [9] E. Jung, S. Iwami, Y. Takeuchi, and T. Jo, "Optimal control strategy for prevention of avian influenza pandemic," *Journal of Theoretical Biology*, vol. 260, no. 2, pp. 220–229, 2009.
- [10] J. X. Wang, W. S. Xu, R. Q. Zhang, and N. Wu, "Epidemic prevention and infection control of a field dressing station in Wenchuan earthquake areas," *International Journal of Infectious Diseases*, vol. 13, no. 1, p. 100, 2009.
- [11] M. E. Halloran, N. M. Ferguson, S. Eubank et al., "Modeling targeted layered containment of an influenza pandemic in the United States," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 12, pp. 4639–4644, 2008.
- [12] K. I. Kim, Z.-G. Lin, and L. Zhang, "Avian-human influenza epidemic model with diffusion," *Nonlinear Analysis: Real World Applications*, vol. 11, no. 1, pp. 313–322, 2010.
- [13] J. Liu and T. Zhang, "Epidemic spreading of an SEIRS model in scale-free networks," *Communications in Nonlinear Science and Numerical Simulation*, vol. 16, no. 8, pp. 3375–3384, 2011.
- [14] M. Samsuzzoha, M. Singh, and D. Lucy, "Numerical study of an influenza epidemic model with diffusion," *Applied Mathematics and Computation*, vol. 217, no. 7, pp. 3461–3479, 2010.
- [15] M. Samsuzzoha, M. Singh, and D. Lucy, "A numerical study on an influenza epidemic model with vaccination and diffusion," *Applied Mathematics and Computation*, vol. 219, no. 1, pp. 122–141, 2012.
- [16] S. Yin, G. Wang, and H. R. Karimi, "Data-driven design of robust fault detection system for wind turbines," *Mechatronics*, vol. 24, no. 4, pp. 298–306, 2014.
- [17] S. Yin, X. Li, H. Gao, and O. Kaynak, "Data-based techniques focused on modern industry: an overview," *IEEE Transactions on Industrial Electronics*, 2014.

- [18] S. Yin and G. Wang, "Robust PLS approach for KPI related prediction and diagnosis against outliers and missing data," *International Journal of Systems Science*, vol. 45, no. 7, pp. 1375–1382, 2014.
- [19] G. S. Zaric and M. L. Brandeau, "Resource allocation for epidemic control over short time horizons," *Mathematical Biosciences*, vol. 171, no. 1, pp. 33–58, 2001.
- [20] G. S. Zaric and M. L. Brandeau, "Dynamic resource allocation for epidemic control in multiple populations," *Journal of Mathematics Applied in Medicine and Biology*, vol. 19, no. 4, pp. 235–255, 2002.
- [21] M. L. Brandeau, G. S. Zaric, and A. Richter, "Resource allocation for control of infectious diseases in multiple independent populations: beyond cost-effectiveness analysis," *Journal of Health Economics*, vol. 22, no. 4, pp. 575–598, 2003.
- [22] G. S. Zaric, D. M. Bravata, J. Cleophas Holty, K. M. McDonald, D. K. Owens, and M. L. Brandeau, "Modeling the logistics of response to anthrax bioterrorism," *Medical Decision Making*, vol. 28, no. 3, pp. 332–350, 2008.
- [23] R. J. Duintjer Tebbens, M. A. Pallansch, J. P. Alexander, and K. M. Thompson, "Optimal vaccine stockpile design for an eradicated disease: application to polio," *Vaccine*, vol. 28, no. 26, pp. 4312–4327, 2010.
- [24] M. Liu and L. Zhao, "Optimization of the emergency materials distribution network with time windows in anti-bioterrorism system," *International Journal of Innovative Computing, Information and Control*, vol. 5, no. 11A, pp. 3615–3624, 2009.
- [25] H. Y. Wang, X. P. Wang, and A. Z. Zeng, "Optimal material distribution decisions based on epidemic diffusion rule and stochastic latent period for emergency rescue," *International Journal of Mathematics in Operational Research*, vol. 1, no. 1-2, pp. 76–96, 2009.
- [26] P. Qiang and A. Nagurney, "A bi-criteria indicator to assess supply chain network performance for critical needs under capacity and demand disruptions," *Transportation Research A*, vol. 46, no. 5, pp. 801–812, 2012.
- [27] N. P. Rachaniotis, T. K. Dasaklis, and C. P. Pappis, "A deterministic resource scheduling model in epidemic control: a case study," *European Journal of Operational Research*, vol. 216, no. 1, pp. 225–231, 2012.
- [28] G. Barbarosolu, L. Özdamar, and A. Çevik, "An interactive approach for hierarchical analysis of helicopter logistics in disaster relief operations," *European Journal of Operational Research*, vol. 140, no. 1, pp. 118–133, 2002.
- [29] S. Y. Yan and Y. L. Shih, "Optimal scheduling of emergency roadway repair and subsequent relief distribution," *Computers & Operations Research*, vol. 36, no. 6, pp. 2049–2065, 2009.
- [30] M. Liu and L. D. Zhao, "Analysis for epidemic diffusion and emergency demand in an anti-bioterrorism system," *International Journal of Mathematical Modelling and Numerical Optimisation*, vol. 2, no. 1, pp. 51–68, 2011.
- [31] M. Liu, L. Zhao, and H. Sebastian, "Mixed-collaborative distribution mode for emergency resources in an anti-bioterrorism system," *International Journal of Mathematics in Operational Research*, vol. 3, no. 2, pp. 148–169, 2011.
- [32] K. Y. Tham, "An emergency department response to severe acute respiratory syndrome: a prototype response to bioterrorism," *Annals of Emergency Medicine*, vol. 43, no. 1, pp. 6–14, 2004.
- [33] J. B. Sheu, "An emergency logistics distribution approach for quick response to urgent relief demand in disasters," *Transportation Research E: Logistics and Transportation Review*, vol. 43, no. 6, pp. 687–709, 2007.

Research Article

Simulation-Based Planning and Control of Transport Flows in Port Logistic Systems

**Antonio Diogo Passos Lima, Frederico Werner de Mascarenhas,
and Enzo Morosini Frazzon**

*Graduate Program in Production Engineering, Technology Center, Federal University of Santa Catarina (UFSC),
University Campus, P.O. Box 476, Trindade, 88040-900 Florianópolis, SC, Brazil*

Correspondence should be addressed to Enzo Morosini Frazzon; enzo.frazzon@ufsc.br

Received 4 April 2014; Accepted 26 June 2014

Academic Editor: Michael Freitag

Copyright © 2015 Antonio Diogo Passos Lima et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In highly dynamic and uncertain transport conditions, transport transit time has to be continuously monitored so that the service level is ensured at a proper cost. The aim of this research is to propose and to test a procedure which allows an agile planning and control of transport flows in port logistic systems. The procedure couples an agent-based simulation and a queueing theory model. In this paper, the transport scheduling performed by an agent at the intermodal terminal was taken into consideration. The decision-making agent takes into account data which is acquired in remote points of the system. The obtained results indicate the relevance of continuously considering, for the transport planning and control, the expected transit time and further waiting times along port logistic systems.

1. Introduction

Disturbances of internal or external origin can impact the reliability of freight transport. In this case, unreliable transport service generates inefficiencies associated with not only an increase in costs but also a reduction in the service level. Among the disorders which can impact the stability of transport flows two stand out: traffic congestion due to the increasing traffic of vehicles and delays in the decision making process. Notwithstanding these occurrences which indicate capacity constraints, it is likely that the overall logistic infrastructure is still not used optimally in full.

Infrastructure issues of road access can impair the operational efficiency of the entire port sector, thus becoming an obstacle for social and economic national development [1]. According to Hijjar and Alexim [2], when port access is inefficient, the entire cargo transport operation is damaged because bottlenecks in the entrance of the port terminal can result in delays and extra need for storage, therefore, increasing the total logistic cost. To sum up, issues in port access directly affect the efficiency of terminals, exporting companies, and transportation costs. Chin and Tongzong [3] affirm that logistic infrastructure makes up a vital link in

the overall chain of commerce, contributing to the international competition of a nation. Bittencourt [4] has stated that the logistic port model to send and receive cargo usually consists of port operators making time windows available, in which each shipper sends their vehicles into the port. However, once traffic conditions are uncertain and there is no access control to the port zone, shippers ultimately send in their vehicles at same time window, looking either for convenience or for guaranteeing goods delivery. Studies on new port logistic methods which enable rationalizing vehicle traffic and optimizing port cargo load meet the current need to improve road infrastructures, stimulating greater control and improvements in the national logistic scenario [5, 6].

Between two nodes in a logistics network, normally more than one possible route is available. In other words, more than one route choice linking the transport origin and destination may be chosen. In this case, the decision on which route should be used can reduce the impact caused by the large amount of traffic flow at a given time, mainly if the routing decision is taken quickly on the basis of the observation of changes in the environment. It is also pointed out that competitive markets require greater agility from transport operations so that they can respond quickly to fluctuations

[7–9]. Nevertheless, currently, most transport schedules take into account only local restrictions, ignoring dynamic environmental variables or external disturbances [10].

An important review of recent mathematical programming models that deal with this subject was made by Mula et al. [11]. Apart from that, noteworthy real-time approaches are being studied in order to address requirements of highly dynamic systems. Among the programming techniques being studied, robust programming can be highlighted. This type of programming is aimed at creating intelligent predictive schedules that minimize the effects of disturbances on the performance of the implemented schedule. Among various existing robust programming techniques, genetic algorithms, multiple predefined schedules, and fuzzy stand out [12, 13].

The application of an agent-based simulation in the preparation of the transport schedule can be an effective solution to mitigate the aforementioned impacts of raising transport volumes. Taking into account the complexity of the involved structures and the set of factors to be considered, the development of a procedure to decide the route to be used by a transport vehicle so that it can be applied on port logistic chains is quite challenging, both in scientific and in practical terms. In this context, this paper proposes a procedure, based on the application of stochastic queueing models, to reduce the impacts caused by increased transport time due to traffic congestion on the route. The analysis of the performance of the proposed procedure employs a test case, implemented through a simulation model representing a port logistic chain that includes transport flows, an intermodal logistics terminal, and a port terminal.

2. Literature Review

Several transport flow optimization problems can be solved using exact methods, such as Assad [14], for instance, route displacement minimization and minimization of number of used routes. However, despite the capability of exact methods to find a solution, these techniques are not the most used for solving NP-hard problems, since the processing time increases with the complexity of mentioned problems [15]. Apart from exact methods, other techniques available for solving transport flow optimization problems can be classified as [16] heuristic and metaheuristic methods. In a study by Bonasser and Gualda [17], several techniques are presented.

When there is a problem that takes into account only the length of the route to be used by a vehicle, this problem can be classified as a vehicle routing problem (VRP). In Bonasser and Gualda [17], the following variants of the vehicle routing problem are presented: heterogeneous fleet and sizing and allocation of fleet (fleet size and mix VRP). Furthermore, this study shows subvariants of the heterogeneous fleet and also techniques used to solve each variant of the VRP. Otherwise, Raff [15] classifies routing problem into the following types: single depot, multiple-vehicle, node routing problem; multiple depot, multiple-vehicle, node routing problem; and single depot, multiple-vehicle, node routing problem with stochastic demand. Regarding the complexity of a routing problem, this is enhanced through the inclusion of

the following constraints [15]: specification of a time period in which a vehicle must be in service before it returns to its point of origin; specification of tasks that can only be performed by certain types of resources; and specification of multiple garages where the vehicles may be stored. In addition to the concept of vehicle routing problems, it is necessary to differentiate it from the concept of scheduling problems. The last is mainly characterized by the routing mechanism of each vehicle that is established in the transport schedule and only takes into account time and space.

Concerning the existing queueing models, depending on the perspective utilized, those can receive different classifications. Thus, from the user point of view, the existing queues are classified as follows [18]: generative models—they provide the user with an optimal solution that satisfies the objective function; and evaluative models—although they do not provide an optimal solution that satisfies the user objective function, those models help in the evaluation of a set of decisions by providing performance metrics. Queueing networks models are classified as evaluative methods. Besides this form of classification, it is possible to classify queueing methods by the accuracy of obtained results [18]: (i) models that provide exact results—these models are difficult to obtain and only available for queueing network models of small size; (ii) models that provide approximate results. So, in this type of model, there is a tradeoff between complexity and accuracy. Still, if one wishes to efficiently analyse a network queue model with finite overall processing time, this approach is the most appropriate [18]. Other recent studies involving queueing models are described in Abdelkader and Al-Wohaibi [19] and Dragović et al. [20]. The first study proposed a new performance measure in a single server Markovian queueing system. The second one refers to a model that combines finite waiting areas, batch arrival queues, and identical and independent cargo-handling capacities.

The following dispatching methods were listed in a study by Wu and Chen [21]: dispatching rules; heuristics; data mining based approaches; agent technologies; and simulation. Moreover, the aforementioned work demonstrated that several studies have been conducted in an attempt to solve the problem of scheduling. Regarding the approaches used in these studies, Wu and Chen [21] highlight the following: (i) selection of the most appropriate dispatching rule among those that are available; (ii) inclusion of adjustable parameters that can optimize the dispatching rule; and (iii) best estimate of the schedule through the results of a limited number of simulations. As for the approach (ii), it is emphasized that this can only be used with factors that are defined previously and remain static [21]. Regarding dispatching rules, these can be considered a practical method for scheduling [22]. Different classifications of dispatching rules can be described [23]: static—the priority of waiting jobs does not change over time; dynamic—the priority of waiting jobs changes over time; local—decisions are taken only from the jobs that are waiting for service; and global—decisions are made with the use of additional information about jobs or other workstations machines.

In a study by Panwalker and Iskander [23], the ordering rules and their respective characteristics were also presented.

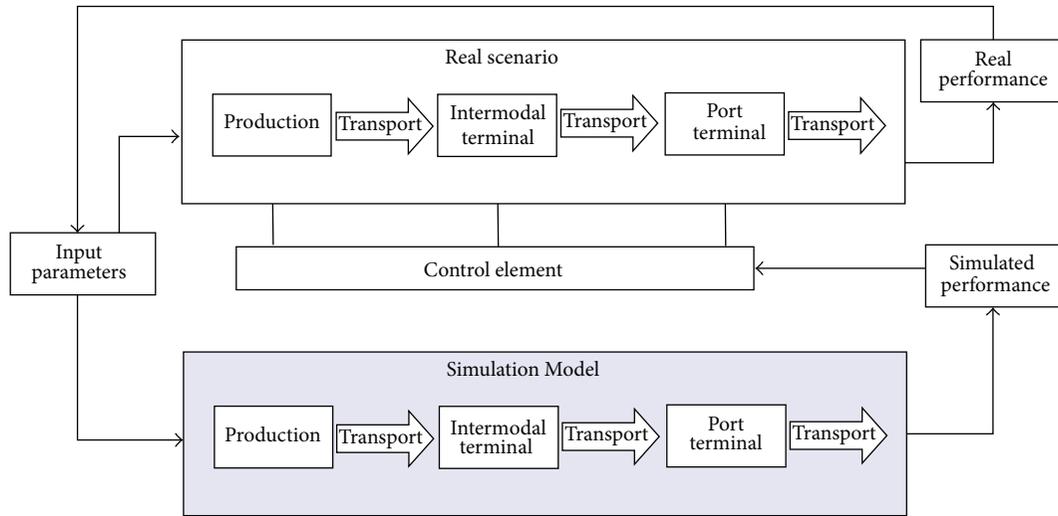


FIGURE 1: Approach to a simulation-based planning and control of port logistic system.

The rules listed in this study are classified as below: simple rules priority—based on information related to a single job; combination of simple rules of priority—two simple priority pieces of information are used when defining the priority; weighted priority rules—similar to the above classification, but with the use of weights; heuristic scheduling rules; and other rules, such as the use of specific rules for particular production sets. In the study of Azimi et al. [24], different kinds of dispatching rules were detailed. Although this study focused on a multiple-load AGV System, the concepts related to dispatching rules also apply to the transport scheduling problem. The study also differentiated categories of the route selection problem regarding the presence of stochastic events.

Due to the increasing complexity of transport flow problems as it integrates a larger number of elements in the model and the increase in the computing power, the use of simulation to solve this kind of problem has become an attractive option. In this case, simulations are usually employed when [25] a model presents variables that are not deterministic and when the problem requires both time and space integration. Models are the translation of operational requirements and constraints to the understanding of the process by the computer [26]. The representation quality of a model regarding the analysed logistics process influences directly the results provided by the model. Once the model of a logistics process is complete, the simulation is typically applied to one of the following purposes [25]: as a tool for identifying and evaluating the improvement of the operating performance or as a tool to gain a better understanding of the costs and performance potential of logistics operations. In this sense, customized simulations can be used to support reengineering decisions [27]. According to Swaminathan et al. [27], simulation is the only viable platform for detailed analysis of alternative solutions because the complex interactions between the entities of a supply chain do not allow the use of analytical solutions. Furthermore, nonprescriptive insights generated from qualitative analyzes, such as benchmarking, do not allow other conclusions that are not related to

the current trends. Swaminathan et al. [27] also point out that the major problems of simulation are associated with the time and effort required to develop specific models with sufficient fidelity to the actual supply chain of interest and the limited reuse of the simulation models.

3. Simulation-Based Planning and Control of Transport Flows

The approach is embedded in a procedure, which is based on the application of stochastic queueing models for improving dispatching rules and synchronizing transport flows in port logistic systems. The aim of this research is to propose and to test a procedure which allows for an agile planning and control of port logistics. The procedure couples an agent-based simulation and a queueing theory model. From the literature review, the proposed procedure considers the following:

- (i) maximization of the quantity of goods delivered as the objective function;
- (ii) a solution of the following vehicle routing problem: single depot, multiple-vehicle, and node routing problem; moreover, the problem addressed in this proposal presents a high degree of complexity, because it is constrained by a time window in which a vehicle must be in service before it returns to its point of origin [15];
- (iii) use of queueing theory models as evaluative models as well as to provide approximate results [18];
- (iv) use of dispatching rules which are classified as static at the bottleneck and global at the intermodal terminal [23];
- (v) use of simulation as a tool to identify and evaluate the performance improvement operations [25].

The conceptual view which motivated the proposed approach for the simulation-based planning and control of port logistic systems is presented in Figure 1. For a generic

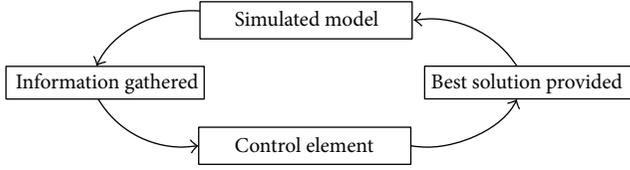


FIGURE 2: Approach to stages 2 and 3 of the simulation-based planning and control of transport flows.

export logistics chain, the concept embraces the following stages: (1) a representation of the real scenario, occurring in real time, subject to the input parameters and generating the output data (real performance) at each time interval; (2) a simulation model, subject to the input parameters and generating the output data (simulated performance), running in simulated time, to be executed numerous times, so that the results can be evaluated by an external control element; (3) a control element that evaluates the obtained simulated performance and, according to a predetermined control logic, allows for the effective implementation in the real scenario.

The research conducted in this paper comprises stages 2 (simulation model) and 3 (control element) for a simplified port logistic chain embracing an intermodal terminal, a transport operation, and the arrival at a port terminal, as represented in Figure 2.

As for the control element, this part has been accomplished through the use of predefined parameters. Thus, with the exception of the number of customers waiting for service on the bottleneck, the control assumed that all other parameter settings of the model remained constant. This case will be performed using a model that includes a scenario of an export logistics chain in which the cargo of an intermodal terminal should be dispatched at the lowest possible cost to a port terminal.

The simulation model represents the actual operation of a port logistic system comprising an intermodal terminal, a transport operation, and the arrival at a port terminal. Figure 3 illustrates the structure of the model.

The decision to dispatch the cargo is carried out early in the day and involves only the choice of the route to be held by transport vehicles. Besides that, the dispatching decisions vary in accordance to the following factors: quantity of goods in inventories of the intermodal terminal; traffic congestion of each road; and end time scheduled for the carriage of goods by the transport vehicle.

Furthermore, in an effort to address the above problem as a nondeterministic, this problem will not only have random components but also display queuing along the routes due to the low carrying capacity compared with the traffic demand. With this purpose, this study will evaluate the impact of the transport scheduling when the responsible agent takes decisions based on the information provided by a control element. The control can monitor the traffic on the routes that can be used by transport vehicles. The information that will be used as input to the control is the quantity of goods that are in the intermodal terminal; the estimated transport transit time on the routes due to the presence of traffic congestion; and the amount of time remaining for the cargo transportation by

TABLE 1: Notation of transport transit time components.

Constant parameters	
T_R	Travel time required to return the vehicle to the intermodal terminal from the port (hours)
T_L	Time to load the cargo at the vehicle (hours)
T_U	Time to unload the cargo at the vehicle (hours)
Variables	
T_{VT}	Travel time to the port terminal (hours)
T_P	Processing time at the bottleneck (hours)

the available transport vehicles. To calculate the transport transit time, the following equation will be used:

$$T_T = (T_P + T_{VT} + T_L + T_U + T_R). \quad (1)$$

Table 1 presents the description of the parameters of the above formula.

In an effort to calculate the estimated processing time (T_P) of the vehicles at the bottlenecks of the routes where traffic congestion occurs, the problem was modelled as the type $M/M/C$: (∞ , FIFO). In this modelling, C is the number of tracks of a given route. Thus, to support the transportation schedule of the day, the estimated transport transit time (T_T) on the routes was calculated from an estimate of the average waiting time of the vehicles in the system when a certain amount of vehicles in the queue was observed. For instance, this information could be obtained by monitoring cameras at the critical bottleneck of each route (Figure 3).

With the purpose of clarifying the understanding of the procedure used to calculate the estimated processing time (T_P) of the vehicles at the bottlenecks, the steps used by the control element are described below:

- (1) data-collection regarding the amount of vehicles at the bottleneck of each route;
- (2) the control element's identification of the class that includes, in its amplitude, the information collected in step 1, through the use of a table previously generated;
- (3) association of the situation observed in step 1 to a queue problem which presents a service rate (μ) previously raised and an arrival rate (λ) corresponding to the class identified in step 2;
- (4) utilization of the average waiting time (W) from the queue problem associated in step 3 as the estimative of the processing time (T_P) of the vehicles at the bottleneck.

Regarding the table used in step 2 of the aforementioned procedure, it was created using a computational algorithm. In this case, a range of data was defined for each bottleneck that would be used for the control decision. After that, this range of data was divided into 4 classes of equal amplitude. Using 4 classes of equal amplitude allows for the construction of the simulation model. It is important to remark that the use of

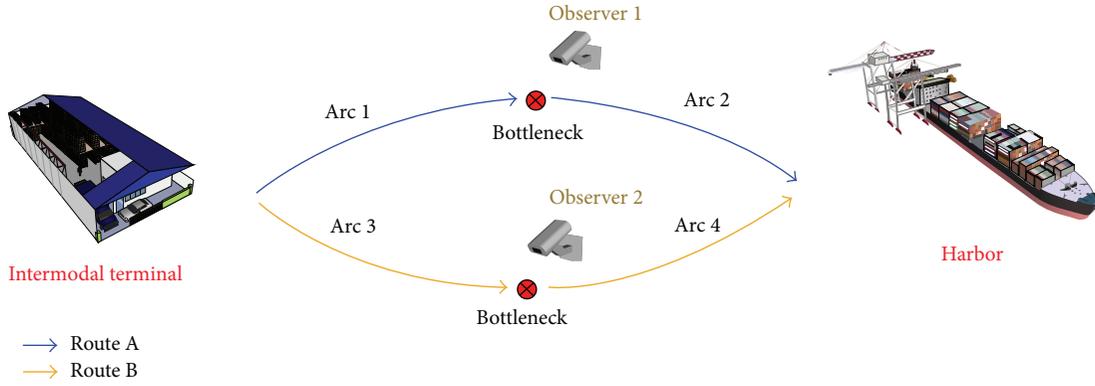


FIGURE 3: Model structure.

```

Input:  $\mu, \lambda_p, P_C, N_C, P_M, \text{step}$ 
 $\lambda_p = \mu - 1/W_{\min}$ 
aux = 0
While aux = 0
  Ro1 =  $\lambda_p/\mu$ 
   $P = 1 - \text{ro1}$ 
   $P_M = 1 - P$ 
  For  $i = 1, \dots, |N_C|$ 
     $P = P * \text{ro1}$ 
     $P_M = 1 - P$ 
  End For
  If  $((P_M > P_C) \text{ or } (\lambda_p \geq \mu))$ , aux = 1
  If  $(\lambda_p < \mu)$ , best =  $\lambda_p$ 
  Else
    best =  $\lambda_p$ 
     $\lambda_p = \lambda_p + \text{step}$ 
  End While
If  $(\text{best} < \mu)$ , Return best
Else, Return  $\lambda_p$  not found
    
```

ALGORITHM 1

a greater amount of data classes increases the sensitivity of the control element.

In the sequence, for each class, several iterations were performed, in order to identify the queue model that possesses the following characteristics: service rate (μ) corresponding to the bottleneck in analysis and probability of finding a number of customers at the top of the central element of the class greater than 50%. After that, the information concerning the arrival rate (λ) and average waiting time (W) of the identified queue model was recorded in the respective classes. This computational algorithm is presented in Algorithm 1.

Algorithm 1 starts with the following input data:

- (i) service rate of the modelled bottleneck (μ),
- (ii) lowest average waiting time in the system (W_{\min});
- (iii) thenumber of customers that, above this, has a probability of occurrence greater than P_C (N_C);
- (iv) minimum probability of finding more than N_C clients (P_C);
- (v) step size used by the algorithm to find the arrival rate (step).

Algorithm 1 calculates the initial value of the arrival rate of customers in the queue (λ_p) that will later be incremented, throughout the execution of the algorithm. In doing so, while the probability of occurrence of more than N_C customers in the queue is less than P_C , the value of the variable λ_p will be increased by the step value. However, if the value of λ_p becomes greater than or equal to μ , the value of the variable λ_p will no longer receive any increment. This way, in order to make λ_p receive the least amount of increment, a variable of control (aux) was included in the algorithm. Thus, while this control variable remains at zero, the algorithm through the while loop calculates the probability of finding more than one customer in the queue when it presents an arrival rate λ_p and a service rate μ . The result of this calculation is stored in the variable P_M . Next, the algorithm uses the For loop to calculate the probability of finding more than N_C customers in the queue. This calculation involves an update on the value of the variable P_M .

At the end of the For loop, the algorithm makes a comparison of the value of P_M found with the value of P_C , as well as the value of λ_p with the value of the service rate of the modeled bottleneck (μ). In this case, the control variable aux is updated with the value 1 if it satisfies one of the following conditions: the value of the probability of finding more than N_C customers in the queue (P_M) is greater than P_C and the λ_p is greater than or equal to μ , in which the algorithm cannot determine an arrival rate that meets the input requirements without a modification on the value of μ or step. Furthermore, if the value of P_M is greater than the value of P_C and λ_p is less than μ , the algorithm stores the value of the variable λ_p on the variable best. If the required conditions to an update of the variable aux to 1 were not met, the algorithm stores the value of the variable λ_p on the variable best and increments the λ_p value in a step value. At the end of the While loop, the algorithm returns the value of λ_p found in the last iteration of this loop if that variable was lower than μ . Otherwise, the algorithm reports that it failed to find a λ_p that meets the entry restrictions.

Finally, the goal is to evaluate transport scheduling performed by an intermodal terminal agent seeking the lowest transport cost. Therefore, agent's objective function is

TABLE 2: Notation of the objective function.

Index	
W	Days (1, . . . , 10)
Constant parameters	
c_I	Storage cost in the intermodal (mu/hour)
ct_D	Cost of the transportation (mu/hour)
ct_N	Additional cost of transportation added after 9 p.m. at the cost of the transportation (mu/hours)
c_F	Cost of not transported goods
Variables	
t_I	Total time in which the products were at intermodal (hours)
t_D	Total time in which the products were in transit (hours)
t_N	Total travel time of vehicles that were still in use after 9 p.m. (hours)
Q	Quantity of goods that were not transported

the minimization of all costs perceived by that agent. The respective cost formula is presented in

$$\min Z = \sum_{w=1}^{10} (c_I * t_{wI} + ct_N * t_{wN} + ct_D * t_{wD} + c_F * Q). \quad (2)$$

Table 2 presents parameters description.

The objective function will be used in the analysis of each simulation scenario, enabling the performance assessment of the proposed procedure. It is noteworthy that the decision of the control element will be based on estimates of the traffic arrival rate (λ) at the bottleneck. Moreover, as this decision is made based on information gathered by an observer who is far from the decision maker, differences are expected between the results obtained and those predicted processing time at the bottleneck.

The computational algorithm used by the control element is presented in Algorithm 2. Algorithm 2 receives as input the results obtained by the use of Formula 1 for each one of the routes to the port. Thus, Algorithm 2 starts with the following input data:

- (i) expected transport transit time for route A;
- (ii) expected transport transit time for route B.

Algorithm 2 checks which route has the lowest transport transit time (T_T). This verification by the control element is required every time a transport vehicle is available at the intermodal terminal and cargo needs to be transported to the port terminal. In the sequence, the algorithm checks if the transport of the goods can be performed before 9 p.m. (21:00 h). If possible, the transportation will be assigned to the vehicle. Otherwise, there will be no more transport assigned for that vehicle on that day. Besides that, once a cargo transport is assigned to a vehicle, it will be carried out.

```

Input:  $T_{Ta}, T_{Tb}$ 
If ( $T_{Ta} < T_{Tb}$ )
  If (Clock < (21 -  $T_{Ta}$ )), Return Route A
  Else, Return No route
Else
  If (Clock < (21 -  $T_{Tb}$ )), Return Route B
  Else, Return No route

```

ALGORITHM 2

The control element makes its decisions based on an expectation of transport transit time (T_T) due to the uncertainty regarding the processing time (P_T) at the bottleneck. The P_T used is based on an estimate of the average waiting time (W) if this problem was modelled as a queue of type M/M/C: (∞ , FIFO). Still, since the decisions of the control element are based on expectation that may change by the time the transport vehicle reaches the bottleneck, it is possible that some transportation vehicles continue in operation after 9 p.m. In this case, it is noteworthy that while the W used by the control element depends on the traffic observed in the bottleneck when the transport vehicle is still in the intermodal terminal, the ideal W depends on the arrival rate (λ) in the bottleneck when the transport vehicle reaches it.

4. Test Case of Port Logistic Systems

This research comprised the modelling of agents involved in the chain through the use of process modelling techniques. After that, the impact of decisions of an agent in the rest of the chain was analysed, using a simulation model. The simulation model was implemented using Simio LLC software [28]. Initially, the model was calibrated so that, depending on the traffic flow, a queue could be formed in any of the routes used to transport the cargo from the intermodal terminal to the port terminal. Moreover, an additional calibration was performed in the traffic flow to eliminate the possibility of a modelled route becoming predominantly better than others due only to their travel time (T_{VT}). In addition, the following operational constraints were established in the model: the transport vehicle should return to its place of departure before the end of the day; the carrying capacity of each vehicle is limited to only 1 unit; refusal of transport services occurs if the total journey is set to finish after 9 p.m.; and an additional cost is set when the transport vehicle cannot finish its journey before 9 p.m. due to the elapsed time in the queue.

Figure 4 illustrates the simulation model that was implemented in Simio LLC to evaluate the impact of decisions of an agent in the rest of the export logistics chain.

Concerning the modelling, not only of the delivery of goods to the port but also of the end of the traffic congestion after the bottleneck of each route, two sinks were included at the end of the routes: one traffic sink and one port sink. While the traffic sink absorbs all vehicles originating traffic on each route, the port sink absorbs all goods shipped at the intermodal terminal.

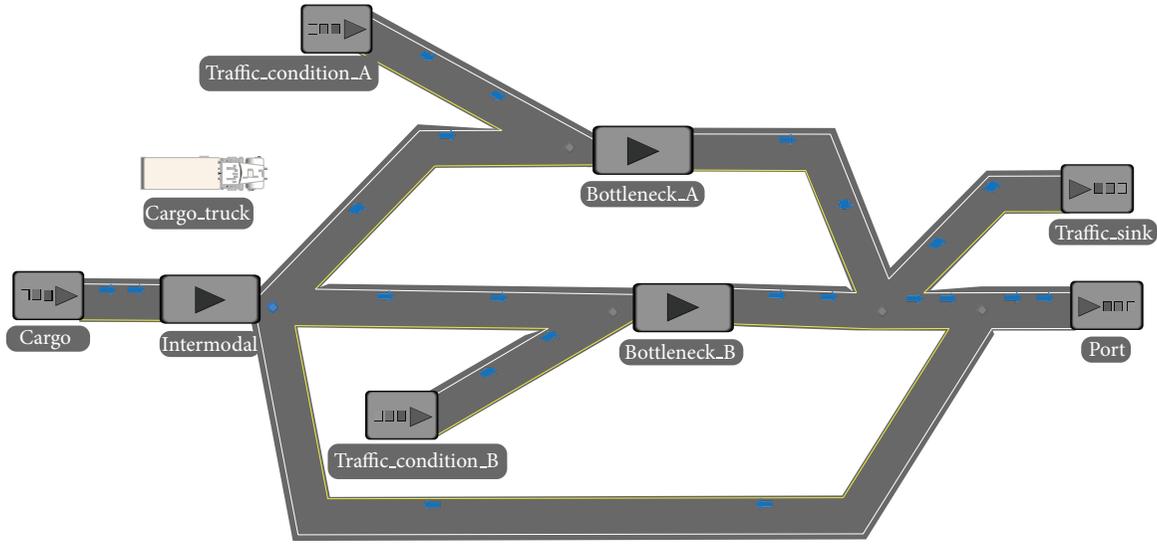


FIGURE 4: Simulation model implemented in Simio LLC.

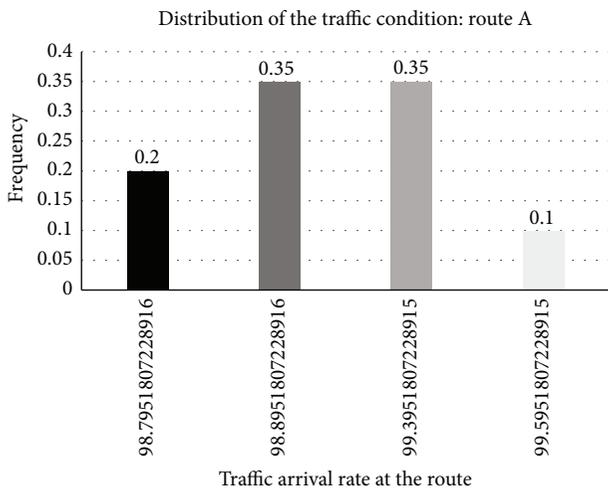


FIGURE 5: Frequency distribution of the traffic arrival rate (λ) at route A.

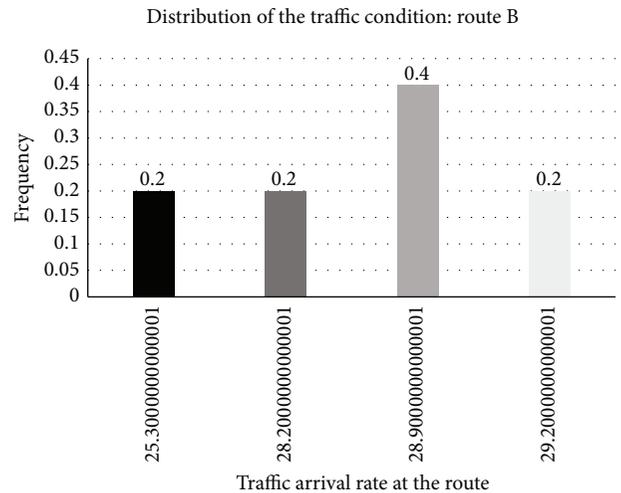


FIGURE 6: Frequency distribution of the traffic arrival rate (λ) at route B.

In order to enable the comparison of results, the amount of products available to be shipped at the intermodal terminal inventories will be set to the value of 9 units at the beginning of each day. Besides that, 3 vehicles were also used, each one with the capacity to transport one cargo at a time.

With regard to the traffic congestion representation in the model, this was accomplished with the inclusion of other vehicles in each of the routes used by the cargo transportation. In this case, the arrival rate of these vehicles at the bottlenecks of each route can assume values varying in accordance to a frequency distribution. Figures 5 and 6 present, respectively, the frequency distribution of the possible values of the traffic arrival rate (λ) that were used to induce traffic congestion at routes A and B.

The values regarding the traffic arrival rate (λ) at the bottleneck were used in order to allow the processing time

(P_T) of the vehicle on the bottleneck to match the average waiting time (W) calculated numerically.

The decisions to be taken by the control element will be performed with the use not only of information regarding the number of vehicles at the bottlenecks of each route but also of information stored in its database. Tables 3 and 4 show, respectively, the information stored in the control element database that is used to assess the situation observed at the bottlenecks of routes A and B.

Regarding the values presented at the column “average waiting time” of Tables 3 and 4, they were previously calculated. These values were obtained not only by modelling the bottleneck problem as a queue of type M/M/1: (∞ , FIFO) but also by considering knowledge of the arrival rate and service rate. It is pointed out that the control element decides which route should be used without being certain about the amount

TABLE 3: Control element database used to evaluate the traffic conditions of route A.

Lower limit of customers	Upper limit of customers	Route A		
		Number of customers whose probability of finding this amount in the queue is greater than 50%	Number of vehicles per hour (λ)	Average waiting time in the system in hours (W)
0	79	40	98,7951807228916	0,8300
80	159	120	98,8951807228916	1,9809
160	239	200	99,3951807228915	3,2806
240	319	280	99,5951807228915	4,8824

TABLE 4: Control element database used to evaluate the traffic conditions of route B.

Lower limit of customers	Upper limit of customers	Route B		
		Number of customers whose probability of finding this amount in the queue is greater than 50%	Number of vehicles per hour (λ)	Average waiting time in the system in hours (W)
0	6	3	25,3000000000001	0,2128
7	13	10	28,2000000000001	0,5556
14	20	17	28,9000000000001	0,9091
21	27	24	29,2000000000001	1,2500

of vehicles which will be waiting at the time the transport vehicle reaches the bottleneck.

This uncertainty arises from the following reasons: (i) there is a time difference between the time when the control element takes its decision and the time when the transport vehicle arrives at the bottleneck, since this decision occurs when the vehicle is still in the intermodal terminal and (ii) the traffic arrival rate (λ) of each route follows an exponential distribution.

Given the above reasons, the decision on the route that should be used is based on an estimation about the number of vehicles per hour (λ) on each route. Furthermore, in order to create a comparison among the proposed scenarios, the corresponding values of the variables that follow a frequency distribution were selected from random numbers. Table 5 illustrates the values that were randomly selected for these variables in each day contemplated in the simulation period. In this table, the random values obtained for the selection of the traffic arrival rate of routes A and B are shown in columns "A.N. flow of vehicles," where A.N. means aleatory number.

Regarding the return of the vehicle to the intermodal terminal, it was assumed that the vehicle would perform this route without any delay caused by traffic. Furthermore, as a consequence of modelling traffic arrival rate (λ) at the bottlenecks as an exponential distribution, the model needed to be run 100 times, for each day, so that the stochastic results converged. As for the other parameters required for the implementation of the model, their values did not vary in the days contemplated in the simulation period. The values of these parameters are shown in Table 6.

Concerning the cost factors of the proposed model, it is pointed out that the costs relating to ct_D and ct_N are

associated with variables a_1 , a_2 , a_3 , a_4 , and a_5 . In reference to the costs related to the unloading cargo at the port terminal, they are considered to be zero in this model.

5. Results and Discussion

In addition to the simulation model developed in Simio LLC for the impact analysis of the transport scheduling performed by an agent using the proposed procedure, comparisons of the results for the following scenarios were conducted.

- (i) 1st Scenario: transport schedule considers that all transport is carried out through the arch which presents a shorter route, but high traffic.
- (ii) 2nd Scenario: transport schedule considers that all transport is carried out through the arch which presents a longer route, but low traffic.
- (iii) 3rd Scenario: transport schedule is adjusted throughout the day according to an expectation of the time spent by the available vehicles to perform a given transportation.

5.1. 1st Scenario: Transport Schedule—All Vehicles Use Route A. In this scenario, the transport schedule is performed such that all transport to the port is made at route A (route that has the shortest path but high traffic). Moreover, the transport schedule did not take into account an evaluation of the expected transport transit time, in other words, did not evaluate the processing time at the bottleneck. Thus, the information taken into account in the transport schedule is referring to (i) number of vehicles available; (ii) amount of each type of product to be transported; (iii) the remainder of

TABLE 5: Selected values for each random variable comprised in the period of simulation.

Days	A.N. flow of vehicles: route A	Traffic arrival rate (λ) at route A (no. of vehicles/hour)	A.N. flow of vehicles: route B	Traffic arrival rate (λ) at route B (no. of vehicles/hour)
1	0,72	99,3951807228915	0,56	28,9000000000001
2	0,94	99,5951807228915	0,33	28,2000000000001
3	0,91	99,5951807228915	0,69	28,9000000000001
4	0,25	98,8951807228916	0,79	28,9000000000001
5	0,1	98,7951807228916	0,31	28,2000000000001
6	0,66	99,3951807228915	0,93	29,2000000000001
7	0,59	99,3951807228915	0,07	25,3000000000001
8	0,53	98,8951807228916	0,94	29,2000000000001
9	0,58	99,3951807228915	0,42	28,9000000000001
10	0,94	99,5951807228915	0,96	29,2000000000001

TABLE 6: Value of the constant parameters.

Parameter	Value	Description
a_1	0,9167	Travel time of arc 1 from route A (hours)
a_2	0,9167	Travel time of arc 2 from route A (hours)
a_3	1,25	Travel time of arc 3 from route B (hours)
a_4	1,25	Travel time of arc 4 from route B (hours)
a_5	1,833	Travel time of the vehicle return path (hours)
c_I	1,875	Storage cost in the intermodal (mu/hour)
ct_D	11,25	Cost of the transportation (mu/hour)
ct_N	16,25	Additional cost of transportation added after 9 p.m. at the cost of the transportation (mu/hour)
c_F	250,00	Cost of goods that were not transported
μ_a	100	Service rate (no. of vehicles/hour) of the bottleneck of the route A
μ_b	30	Service rate (no. of vehicles/hour) of the bottleneck of the route B
Λ_A	3	Loading/unloading time of transport vehicles (minutes)

the transport vehicle; and (iv) time needed for the transport vehicle which can carry out the collection and delivery of the product. Furthermore, in every day of the simulation period, the amount of goods available in the intermodal terminal, 9 in total, can be sent to the port by 3 vehicles at the time. Therefore, from this consideration and also from the considerations above, the transportation schedule for this scenario presents no difficulties. Using the above transportation schedule as input to the simulation model implemented in Simio LLC, the results were obtained in Table 7.

5.2. 2nd Scenario: Transport Schedule—All Vehicles Use Route B. Similar to the 1st simulation described, a new simulation was implemented, but the transport was performed only through route B (route that has the longest route, but little traffic) during the transport of goods from the intermodal

terminal to the port terminal. Using the above transport schedule as input to the simulation model implemented in Simio LLC, the results were obtained in Table 8.

5.3. 3rd Scenario: Transport Schedule—Adjusted throughout the Day according to an Expectation of the Time Spent by the Available Vehicles to Perform a Given Transportation. In this scenario, a change was made in the process that generates the transport schedule. At other scenarios, the transport schedule was used as input to the simulation model in Simio LLC. However, in this scenario, the transportation schedule was also generated at Simio LLC.

The transport schedule at this scenario was generated, during the simulation, while the need for transport and the vehicle availability were presented. Besides that, this transportation schedule did also take into account the traffic conditions that were observed at each bottleneck.

Furthermore, a restriction in the model was included so that only freights with the expected completion time before 9 p.m. were scheduled. Thus, during the simulation period of 10 days, in the simulation model implemented in Simio LLC, the results were obtained in Table 9.

5.4. Analysis. It was observed, in all scenarios, that transport was performed after 9 p.m. In spite of that, it was also possible to visualise different costs for each scenario. In the 1st Scenario, due to shorter travel time in the days when the traffic at route A was not intense, the transport transit time on this route became significantly lower when compared to the transport transit time on route B. Regarding the amount of goods that were not transported and the total travel time after 9 p.m., this scenario presented the worst results, due to those days of high traffic at route A. Concerning the sum of average costs after simulation replications of this scenario, it displays the highest values when compared to other scenarios.

In the 2nd Scenario, the dispatch of all goods was carried with the use of route B. In this case, with the exception of the total time in which the products were in transit, all other cost factors were lower when compared to the results of the 1st Scenario. Thus, despite the fact that the travel time of route B is longer than the respective time of route A, it is pointed

TABLE 7: Results when all transports are made with the use of route A.

	Average results from 1st Scenario after simulation replications					Sum of average Costs after simulation replications
	Total time in which the products were in transit (h)	Number of vehicles in use after 9 p.m.	Total travel time after 9 p.m.	Total time in which the products were found stored (h)	Amount of products that were not transported	
Average results after simulation replications	295,74	17	88,25	515,17	5,87	—
Average costs after simulation replications	3768,39	—	1434,11	965,94	1466,4	6200,73

TABLE 8: Results when all transports are conducted through route B.

	Average results from 2nd Scenario after simulation replications					Sum of Average Costs after simulation replications
	Total time in which the products were in transit (h)	Number of vehicles in use after 9 p.m.	Total travel time after 9 p.m.	Total time in which the products were found stored (h)	Amount of products that were not transported	
Average results after simulation replications	306,12	17	66,38	502,07	2,84	—
Average costs after simulation replications	3775,8	—	1078,7	941,38	710,2	5427,38

TABLE 9: Results when the decision making process uses the proposed control.

	Average results from 3rd Scenario after simulation replications					Sum of average Costs after simulation replications
	Total time in which the products were in transit (h)	Number of vehicles in use after 9 p.m.	Total travel time after 9 p.m.	Total time in which the products were found stored (h)	Amount of products that were not transported	
Average results after simulation replications	249,84	12	65,89	461,44	1,32	—
Average costs after simulation replications	3140,15	—	1070,64	865,19	331,05	4336,39

out that the difference between the results of the 1st and 2nd Scenarios is due to the low processing time (T_p) for transport vehicles at the bottleneck of route B. Compared to the sum of average costs after simulation replications of the 1st Scenario, the 2nd Scenario presented a lower value.

In the 3rd Scenario, the decision on the route that was used to carry out the shipping of the goods was made with the aid of the proposed control element. In this case, it was observed that all cost factors were lower compared to the other two scenarios. The sum of average costs after simulation replications of this scenario, in comparison to the cost of all other scenarios, presented the lowest value.

6. Conclusion

A greater consistency of goods carried was identified when the transport scheduling decision takes into account the expectation about the time calculated by applying queues stochastic models, which would be spent in queues. Furthermore, it was also possible to observe a reduction in storage

time and the time of goods in transit after the change in the decision process. The improvement performed at the decision process could be implemented in real-world port logistic systems through the application of a customised version of the proposed model.

Regarding reductions in storage time and the time of goods in transit after the change in the decision process, although this study did not demonstrate these benefits, they are related to the reduction in the level of safety stock held along the port logistic system. Thus, this work demonstrated the possibility of achieving a better performance of a port logistic system without necessarily making a change in its structure or infrastructure, only by improving the synchronization of transport flows through the application of a queueing model for supporting scheduling decision making.

Nevertheless, some limitations have to be pointed out. The model assumes the occurrence of only one queue along the transport route. When there is more than one bottleneck on the same route, the resulting flow of goods is determined

by the bottleneck with the lowest capacity. During a transport journey, it is possible that the vehicle passes through several traffic congestions. In this case, the elapsed time in the traffic congestion is associated with the service capacity limit of different segments along a route. The control element takes into account an expectation about the time that would be spent in queues. If there is more than one queue on a route, it would be necessary to monitor and inform the size of these queues to the control element.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This paper was supported by (i) a Grant from Simio LLC covering software licenses; (ii) two Grants from CNPq (the Brazilian National Council for Scientific and Technological Development): a Scientific Initiation Grant (*Bolsista de Iniciação Científica—PIBIC*) and a Senior Researcher Grant (*Bolsista de Produtividade em Pesquisa do CNPq—PQ Nível 2, Processo 303879/2012-2*); and (iii) one Master Student Grant from CAPES (*Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*). The authors would also like to thank the reviewers for the important comments and suggestions which contributed to the continuous improvement of this research paper.

References

- [1] A. M. Valente, E. Passaglia, J. A. Cruz et al., *Qualidade E Produtividade Nos Transportes*, Cengage Learning, São Paulo, Brazil, 2008.
- [2] M. F. Hijjar and F. M. B. Alexim, "Avaliação do acesso aos terminais portuários e ferroviários de contêineres no Brasil. Coppead/UFRJ, Centro de Estudos em Logística," http://www.coppead.ufrj.br/pt-br/upload/publicacoes/ArtLog_SET_2006.pdf/.
- [3] A. Chin and J. Tongzon, "Maintaining Singapore as a major shipping and air transport hub," in *Competitiveness of the Singapore Economy*, T. Toh, Ed., pp. 83–114, Singapore University Press, 1998.
- [4] R. Bittencourt, *Modelo de apoio ao controle de acesso de veículos pelo modal rodoviário junto aos portos utilizando tecnologia RFID. [Dissertação Mestrado em Engenharia Civil]*, Universidade Federal de Santa Catarina, Florianópolis, Brazil, 2012.
- [5] O. Sharif, N. Huynh, and J. M. Vidal, "Application of El Farol model for managing marine terminal gate congestion," *Research in Transportation Economics*, vol. 32, no. 1, pp. 81–89, 2011.
- [6] G. Chen, K. Govindan, and Z. Yang, "Managing truck arrivals with time windows to alleviate gate congestion at container terminals," *International Journal of Production Economics*, vol. 141, no. 1, pp. 179–188, 2013.
- [7] B. Yang and N. Burns, "Implications of postponement for the supply chain," *International Journal of Production Research*, vol. 41, no. 9, pp. 2075–2090, 2003.
- [8] B. Scholz-Reiter, E. M. Frazzon, and T. Makuschewitz, "Integrating manufacturing and logistic systems along global supply chains," *CIRP Journal of Manufacturing Science and Technology*, vol. 2, no. 3, pp. 216–223, 2010.
- [9] E. M. Frazzon, S. Loureiro, J. R. Orlando Fontes Lima, and B. Scholz-Reiter, "A knowledge management approach for the integration of manufacturing and logistics in global production networks," in *Proceedings of the 44th CIRP Conference on Manufacturing Systems*, Madison, Wis, USA, June 2011.
- [10] A. Scholl, *Robuste Planung und Optimierung: Grundlagen, Konzepte und Methoden; Experimentelle Untersuchungen*, Physica, Heidelberg, Germany, 2001.
- [11] J. Mula, D. Peidro, M. Díaz-Madroñero, and E. Vicens, "Mathematical programming models for supply chain production and transport planning," *European Journal of Operational Research*, vol. 204, no. 3, pp. 377–390, 2010.
- [12] D. Huisman, R. Freling, and A. P. M. Wagelmans, "A robust solution approach to the dynamic vehicle scheduling problem," *Transportation Science*, vol. 38, no. 4, pp. 447–458, 2004.
- [13] W. Herroelen and R. Leus, "Project scheduling under uncertainty: survey and research potentials," *European Journal of Operational Research*, vol. 165, no. 2, pp. 289–306, 2005.
- [14] A. A. Assad, "Models for rail transportation," *Transportation Research A: General*, vol. 14, no. 3, pp. 205–220, 1980.
- [15] S. Raff, "Routing and scheduling of vehicles and crews. The state of the art," *Computers and Operations Research*, vol. 10, no. 2, pp. 63–211, 1983.
- [16] C. B. Cunha, *Uma contribuição para o problema de roteirização de veículos com restrições operacionais*, EPUSP, Departamento de Engenharia de Transporte, São Paulo, Brazil, 1997.
- [17] U. O. Bonasser and N. D. F. Gualda, "Um problema de roteirização nas operações de transporte da força aérea brasileira," in *Proceedings of the 18th ANPET Congresso Nacional de Pesquisa e Ensino em Transporte*, Florianópolis, Brazil, November 2004.
- [18] H. T. Papadopoulos and C. Heavey, "Queueing theory in manufacturing systems analysis and design: a classification of models for production and transfer lines," *European Journal of Operational Research*, vol. 92, no. 1, pp. 1–27, 1996.
- [19] Y. H. Abdelkader and M. Al-Wohaibi, "Computing the performance measures in queueing models via the method of order statistics," *Journal of Applied Mathematics*, vol. 2011, Article ID 790253, 12 pages, 2011.
- [20] B. Dragović, N. Park, N. Zrnić, and R. Meštrović, "Mathematical models of multiserver queueing system for dynamic performance evaluation in port," *Mathematical Problems in Engineering*, vol. 2012, Article ID 710834, 19 pages, 2012.
- [21] H. Wu and T. Chen, "A fuzzy-neural ensemble and geometric rule fusion approach for scheduling a wafer fabrication factory," *Mathematical Problems in Engineering*, vol. 2013, Article ID 956978, 14 pages, 2013.
- [22] H. Tsai and T. Chen, "A fuzzy nonlinear programming approach for optimizing the performance of a four-objective fluctuation smoothing rule in a wafer fabrication factory," *Journal of Applied Mathematics*, vol. 2013, Article ID 720607, 15 pages, 2013.
- [23] S. S. Panwalkar and W. Iskander, "A survey of scheduling rules," *Operations Research*, vol. 25, no. 1, pp. 1–18, 1977.
- [24] P. Azimi, H. Haleh, and M. Alidoost, "The selection of the best control rule for a multiple-load AGV system using simulation and fuzzy MADM in a flexible manufacturing system," *Modelling and Simulation in Engineering*, vol. 2010, Article ID 821701, 11 pages, 2010.

- [25] D. J. Bowersox and D. J. Closs, "A simulation in logistics: a review of present practice and a look to the future," *Journal of Business Logistics*, vol. 10, no. 1, pp. 133–148, 1989.
- [26] A. Micu, A. E. Micu, V. Mazilescu, and C. Nistor, "Simulation in logistics network context: theoretical approach of the optimization process," *Universitatea Petrol-Gaze Ploiesti*, vol. 60, pp. 95–104, 2008.
- [27] J. M. Swaminathan, S. F. Smith, and N. M. Sadeh, "Modeling supply chain dynamics: a multiagent approach," *Decision Sciences*, vol. 29, no. 3, pp. 607–631, 1998.
- [28] SIMIO LLC, Sewickley, Pa, USA, 2014; <http://www.simio.com/>.

Research Article

A Multiobjective Fuzzy Aggregate Production Planning Model Considering Real Capacity and Quality of Products

Najmeh Madadi and Kuan Yew Wong

Department of Manufacturing and Industrial Engineering, Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, Skudai, Malaysia

Correspondence should be addressed to Kuan Yew Wong; wongky@fkm.utm.my

Received 3 April 2014; Revised 29 July 2014; Accepted 31 July 2014; Published 11 September 2014

Academic Editor: Michael Freitag

Copyright © 2014 N. Madadi and K. Y. Wong. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this study, an attempt has been made to develop a multiobjective fuzzy aggregate production planning (APP) model that best serves those companies whose aim is to have the best utilization of their resources in an uncertain environment while trying to keep an acceptable degree of quality and customer service level simultaneously. In addition, the study takes into account the performance and availability of production lines. To provide the optimal solution to the proposed model, first it was converted to an equivalent crisp multiobjective model and then goal programming was applied to the converted model. At the final step, the IBM ILOG CPLEX Optimization Studio software was used to obtain the final result based on the data collected from an automotive parts manufacturing company. The comparison of results obtained from solving the model with and without considering the performance and availability of production lines, revealed the significant importance of these two factors in developing a real and practical aggregate production plan.

1. Introduction

Since the introduction of aggregate production planning (APP) problem in 1950s, it has been studied vastly by many researchers. The interest in APP has a root in the ability that it provides to companies for effective control of production and inventory costs as the two substantial portions of the overall cost of manufacturers [1]. It also helps to identify decisions concerning layoff and hiring of workers, overtime production quantities, backorder and inventory levels, subcontracting, and all the required resources [2, 3].

The overall goal of APP is to set the overall production rates for each product category to meet the fluctuation of customers' demand in a cost-effective manner and for a certain time horizon [4]. While APP is considered as an upper level planning in the process of production management, other forms of disaggregation plans (e.g., master production schedule, capacity plan, and material requirements plan) are all related and dependent on APP in a hierarchical way [5]. The costs associated with APP mostly consist of costs related to payroll, inventory, backordering, hiring and layoff

of workers, overtime, and regular time production [6]. The time horizon for developing an APP is often from 3 to 18 months forward [7].

Taking into account the beneficial impact of applying APP in companies, different approaches and methodologies such as linear programming [8, 9], mixed integer linear programming [6, 10–13], goal programming [14–17], or, in the case of providing solutions to the developed mathematical models, some heuristic algorithms such as genetic algorithm [7, 18, 19] and tabu search [17, 20–22] have been applied. In most of the previously performed studies, APP has been introduced as the method by which decision makers trade off between incurring cost and increasing capacity, having inventory or backlog orders. Although, a successful combination of such trade-offs can bring beneficial results to the company, there may be some other influential factors that should be taken into account when developing an APP. For instance, the quality of products has not received adequate attention in previous studies in the area of APP. Ignoring the quality of products and just focusing on minimizing operational cost in the process of production planning lead to poor customers'

perception about the products and consequently sales loss in the long term.

The other issue in developing an aggregate production plan is the uncertain environment that the planner has to deal with. The uncertainty in production planning process is due to imprecise input data such as the size and amount of required or on-hand resources, customers' demand, or uncertainty about the aspiration levels of a decision maker's goals and objectives [23]. Therefore, suitable approaches should be followed to incorporate these uncertainties when developing an APP. Fuzzy programming is one of the main approaches used vastly in the literature to incorporate uncertainty into production plans [24]. This study also applies the fuzzy programming approach to consider the aforementioned uncertainties in developing the APP. In addition, because of the lack of attention given to the quality of products in previous related research, this study incorporates quality as an objective function in developing the APP.

The other contribution of this paper is the capability of the developed APP model in considering the real capacity of production lines through incorporating their availability and performance percentage into the capacity constraint of the mathematical model. Considering performance and availability levels of production lines, which takes into account their down time and speed loss, provides the planner with a realistic view of on-hand capacity and consequently results in a more practical APP.

This paper contains eight sections. The next section provides some information about fuzzy programming together with a brief review on some relevant previous studies. Section 3 provides information about the construction of the mathematical model. In Section 4, the approach applied for providing the solution to the model is discussed. Section 5 confirms the applicability of the proposed model using data collected from the company under study. In Section 6, a comparison will be made in order to signify the consequences of ignoring the performance and availability of production lines in developing a practical APP. Sensitivity analysis and conclusions will be presented in Sections 7 and 8, respectively.

2. Literature Review

Application of fuzzy set theory in the conventional linear programming model was first introduced by Zimmermann [25, 26] in 1978. The purpose was to deal with the uncertainties in the input data which are mostly due to the lack of decision makers' knowledge and unavailability of required data. Due to the nonrandomness nature of such uncertainties, they cannot be expressed by probability distributions and considering conventional stochastic programming models leads to inefficient and impractical results. On the other hand, assuming the uncertain parameters to be deterministic can also lead to unreal and impractical results [27].

In the fuzzy model proposed by Zimmermann [25] both the objective function and constraints were formulated in a fuzzy environment where imprecise parameters are processed as fuzzy numbers and imprecise constraints as fuzzy sets. Using the min-operator, he showed that there is an equivalent

linear programming model to the original constructed fuzzy multiobjective model.

Narasimhan [28] demonstrated how fuzzy subset concept can be incorporated into a goal programming model in a fuzzy environment. Hannan [29] illustrated the application of piecewise membership functions in quantifying fuzzy aspiration levels. He considered a fuzzy goal-programming model with preemptive priorities and Archimedean weights and solved the model by maximization of the membership function of the minimum goal.

Introduction of fuzzy set theory into linear programming models by Zimmermann [26] has opened new windows for researchers who are dealing with a fuzzy environment in different areas of operations management. Based on a survey conducted by Wong and Lai [30], between the years 1998 and 2009, the number of applications of fuzzy set theory in different areas of operations and production management was about 400. Among them, the area of long term capacity planning had the highest share of application, namely, 16.13%, followed by short term capacity planning (14.4%) and inventory control (11.17%). However, the least number of applications was found in the areas of job design and long term forecasting with almost no application. In the area of APP, fuzzy set theory has its own significant place [5, 28, 31–33]. Jamalnia and Soukhakian [23] developed a fuzzy multiobjective nonlinear APP model to address the fuzzy aspiration levels of objective functions. In their model, three quantitative objective functions, namely, minimization of production, inventory, and changes in workers costs, and one qualitative objective function of maximizing customer satisfaction were taken into account. The nonlinearity of the model was due to the effect of the learning curve in decreasing production time as workers achieve more experience. By applying the triangular fuzzy membership function, the model was converted to a crisp one and finally it was solved using a branch of genetic algorithm.

Wang and Fang [34] developed an aggregate production plan with some fuzzy parameters including product price, unit cost of subcontracting, workforce level, production capacity, and market demand along with fuzzy aspiration levels of objective functions. Providing a systematic framework, the proposed approach supports decision makers through an interactive way until the satisfactory results are obtained. At the final step, an aggregation operator was employed in order to obtain the compromised solution of the proposed system.

Wang and Liang [35] also proposed a fuzzy linear programming model and developed an APP. Minimization of total production costs, carrying and backordering costs, and rates of changes in labor levels are the three objective functions of that study. In order to convert the problem into an ordinary linear programming problem, the piecewise linear membership function of Hannan [29] (to represent the fuzzy goals of decision makers) along with the fuzzy decision making approach of Zadeh [36] were taken into account.

The review of several studies in the APP area described in this section obviously shows that almost all researchers have considered cost as the first objective to be minimized. Minimization of inventory, backorder level, and changes in workforce are the other main objectives that have been taken

into consideration as the second or third objective function. However, enhancing the quality of products as an objective function has been widely ignored in developing APP models. Particularly, there is no study that has taken into account the quality of products when developing an APP in the fuzzy environment. Another deficiency in this area was the overestimation of the real capacity of production lines due to the ignorance of two important factors, that are, performance and availability of production lines in the capacity constraints of the constructed APP models. In this study, an effort has been made to bridge these gaps through developing a multiobjective integer linear programming model in a fuzzy environment in which the quality of products as well as the real capacity of production lines have been taken into consideration. Since, in the real world, planners are faced with qualitative as well as quantitative goals, this study tries to present the application of both qualitative and quantitative objective functions in the construction of the proposed APP model. To test the applicability of the proposed model, it has been applied to a case study chosen from Iran's automotive industry.

3. Construction of the Mathematical Model

In this section a multiperiod, multiobjective integer fuzzy linear programming model is constructed based on the operational conditions of an automotive parts manufacturing company in Iran which is producing three types of products for local customers.

3.1. Operational Conditions and Assumptions of the Model. The operational conditions together with the assumptions of the model are as follows.

- (i) Forecasted customers' demand, production cost, inventory carrying cost, cost of training, cost of purchasing raw materials, reject rate of raw materials, and performance and availability percentages of all production lines are assumed to be imprecise and are modeled by fuzzy numbers.
- (ii) Production lines are balanced.
- (iii) Constant number of operators and workers has been dedicated to each production line throughout the specified time horizon.
- (iv) The cost of hiring is not included in the overall production cost, since all the operators and workers are hired at the beginning of the time horizon.
- (v) Firing the hired workers is not allowable. Workers are trained to acquire the required level of skills in each time period.
- (vi) Regular and overtime production and warehouse space cannot exceed their maximum levels.
- (vii) Backordered demand must be satisfied in the next time period.
- (viii) All customers' demand for all types of products should be fulfilled at the end of the time horizon.

- (ix) Outsourcing is not allowable for any type of products.
- (x) The number of workers with a certain skill level in a time period is not reduced in the next time period.
- (xi) Work in process (WIP) inventory cost is not considered.
- (xii) The specified time horizon contains six monthly periods.
- (xiii) Two separated warehouses are used for final products storage; one is assigned to the type one and type two products and the other warehouse is for storing products of type three.
- (xiv) The level of inventory is assumed to be zero at the beginning of the first period.
- (xv) Each type of products is assigned to just one production line.
- (xvi) All components can be purchased from all suppliers. In total, three suppliers are considered.
- (xvii) Reject rates and costs of raw materials, purchased from different suppliers, are different.
- (xviii) Rejected raw materials are sent back to the relevant suppliers and the company does not pay for them.
- (xix) Higher skilled workers are paid more.
- (xx) The salary of workers is not included in the overall production cost and is considered separately.

3.2. Objective Functions

3.2.1. Quantitative Objective Function. Quantitative objective functions are as follows:

- (i) minimization of total cost,
- (ii) maximization of product quality (through minimization of quality degradation).

3.2.2. Qualitative Objective Function. Customer service level should be "rather high."

The desired service level, that is, "rather high," in the above objective has been identified by the decision maker (production planner in the company under study).

3.3. Parameters Definition. Consider the following:

- T = planning time horizon including six monthly periods;
- t = time period;
- m = type of product ($m \in M$);
- k = supplier index ($k \in K$);
- i = raw material component ($i \in I$);
- n_{im} = usage coefficient of component i in product m ;
- h = regular ($h = 1$) or overtime ($h = 2$) production hours ($h \in H$);
- sk = skill level of a worker (ordinary ($sk = 1$), good ($sk = 2$), and excellent ($sk = 3$));

S_{sk} = salary of a worker with skill level of sk ;

\widetilde{C}_{ik} = cost of component i purchased from supplier k (cost of placing and receiving orders);

$\widetilde{\theta}_{ik}$ = reject rate of component i purchased from supplier k ;

\widetilde{D}_{mt} = forecasted demand of product type m in period t ;

τ_m = ideal production cycle time for product type m ;

\widetilde{C}_{mh}^P = cost of producing one unit of product type m in production time h ;

\widetilde{C}_{mt}^I = inventory carrying cost per unit of product m in period t ;

C_m^B = backorder cost per unit of product m ;

\widetilde{C}_{ht} = cost to train one worker in period t ;

\widetilde{Pr}_m = performance percentage of production line m ;

\widetilde{Av}_m = availability percentage of production line m ;

L_{mt}^W = number of workers to be assigned for producing product type m in period t ;

Max^{wm1} = maximum warehouse space for storage of product types 1 and 2;

Max^{wm2} = maximum warehouse space for storage of product type 3;

Max^{ht} = maximum allowable regular time ($h = 1$) or overtime ($h = 2$) production in period t ;

W_m = required warehouse space per unit of product m .

All input data related to the parameters will be made available upon request.

3.4. Decision Variables. Decision variables (outputs of the model) are as follows:

P_{mht} = unit of product type m to be produced in production time h (regular time or overtime) in period t ;

L_{mt}^h = number of workers to be trained in period t for product type m ;

B_{mt} = backorder level at the end of period t for product type m ;

I_{mt} = available inventory level of product type m at the end of period t ;

Q_{ikt} = quantity of component i to be purchased from supplier k in period t ;

L_{skmt} = number of workers with skill level of sk to produce product type m in period t .

3.5. Formulation of Objective Functions. The objective functions are formulated as follows.

3.5.1. Quantitative Objective Functions

(i) Minimization of cost is as follows:

$$\begin{aligned} \text{Min } Z_1 &= \sum_{t=1}^T \sum_{m=1}^M \sum_{h=1}^H \widetilde{C}_{mh}^P P_{mht} + \sum_{t=1}^T \sum_{m=1}^M \left(C_m^B B_{mt} + \widetilde{C}_{mt}^I I_{mt} + \widetilde{C}_{ht} L_{mt}^h \right) \\ &+ \sum_{t=1}^T \sum_{m=1}^M \sum_{sk=1}^3 S_{sk} L_{skmt} + \sum_{t=1}^T \sum_{k=1}^K \sum_{i=1}^I \widetilde{C}_{ik} Q_{ikt}. \end{aligned} \quad (1)$$

This objective function tries to minimize operational costs including production cost, backorder cost, inventory cost, training cost, and costs associated with salary and raw materials procurement.

(ii) Maximization of quality of products (minimization of quality degradation) is as follows:

$$\text{Min } Z_2 = \frac{\sum_{i=1}^I \sum_{k=1}^K \sum_{t=1}^T Q_{ikt} \widetilde{\theta}_{ik}}{\sum_{i=1}^I \sum_{m=1}^M \sum_{t=1}^T \widetilde{D}_{mt} n_{im}} + \frac{\sum_{m=1}^M \sum_{sk=1}^3 \sum_{t=1}^T L_{skmt}}{\sum_{m=1}^M \sum_{t=1}^T L_{mt}^W}. \quad (2)$$

The aim of this objective function is to maximize quality by minimizing quality degradation of products using two strategies. The first strategy, as has been formulated in the first element, is to decrease the number of rejected raw materials purchased from suppliers and the second strategy represented in the second element of the objective function is to decrease the number of hired workers with a lower skill level. These two elements have a significant effect on the quality of finished products produced during the specified time horizon. One can add other elements that are effective on the quality of finished products, depending on the situation of the company under study.

3.5.2. Qualitative Objective Functions

(i) ‘‘Customer service level should be rather high.’’

In order to measure customer service level, backorder level is used. The term ‘‘rather high’’ has been defined by the decision maker as the desired service level that the company aims to provide to its customers. Figure 1 shows the corresponding membership functions [27] constructed based on the definition of the decision maker and assigned to each linguistic term in Set A , where $A = \{\text{very low, low, rather low, medium, rather high, high, very high}\}$.

To formulate the required linguistic term, the part that has been assigned to ‘‘rather high (RH)’’ and highlighted in Figure 1 has been selected and formulated as shown in

$$\mu_{\text{BLP}_t} = \begin{cases} 1 & \text{BLP}_t \leq 25 \\ \frac{30 - \text{BLP}_t}{5} & 25 < \text{BLP}_t \leq 30 \\ 0 & \text{BLP}_t > 30, \end{cases} \quad (3)$$

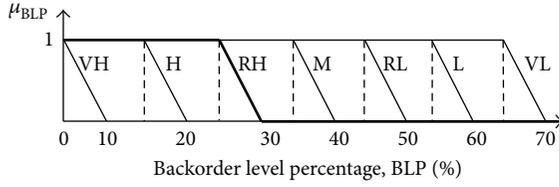


FIGURE 1: Membership functions for different linguistic terms.

where

$$\begin{aligned} \text{BLP}_t & \text{ (Backorder level percentage at the end of period } t) \\ & = \frac{\text{Backorder at the end of period } t}{\text{Demand in period } t} \times 100. \end{aligned} \quad (4)$$

Therefore the final formula for the third objective function can be constructed as presented in

$$\text{Max } Z_3 = \sum_{t=1}^T \mu_{\text{BLP}_t} = \sum_{t=1}^T \frac{30 - \text{BLP}_t}{5}. \quad (5)$$

3.6. Constraints. The following constraints have been formulated based on the assumptions and the company's operational conditions defined earlier:

$$\sum_{h=1}^2 P_{mht} = \bar{D}_{mt} - I_{m(t-1)} - B_{mt} + I_{mt} + B_{m(t-1)}; \quad \forall t > 1, \quad \forall m \quad (6)$$

$$\sum_{h=1}^2 P_{mht} = \bar{D}_{mt} - B_{mt} + I_{mt}; \quad \forall m, \quad t = 1 \quad (7)$$

$$\sum_{sk=1}^3 (L_{skmt} - L_{skm(t-1)}) = L_{mt}^h; \quad \forall t > 1, \quad \forall m \quad (8)$$

$$L_{skmt} \geq L_{skm(t-1)}; \quad \forall t > 1, \quad \forall m, \quad \forall sk \quad (9)$$

$$\sum_{sk=1}^3 L_{skmt} = L_{mt}^W; \quad \forall m, \quad \forall t \quad (10)$$

$$P_{mht} \tau_m \leq \text{Max}^{ht} \times \frac{\bar{Pr}_m \times \bar{Av}_m}{\bar{PA}_m}; \quad \forall m, \quad \forall t, \quad \forall h \quad (11)$$

$$B_{mt} \leq \bar{D}_{mt}; \quad \forall t, \quad \forall m \quad (12)$$

$$B_{mt} = 0; \quad \forall m, \quad \forall t = 6 \quad (13)$$

$$\sum_{m=1}^2 W_m I_{mt} \leq \text{Max}^{wm1}; \quad \forall t \quad (14)$$

$$W_m I_{mt} \leq \text{Max}^{wm2}; \quad \forall t, \quad \forall m = 3 \quad (15)$$

$$\sum_{k=1}^3 Q_{ikt} = \sum_{m=1}^3 \sum_{h=1}^2 (P_{mht} \times n_{im}); \quad \forall t, \quad \forall i. \quad (16)$$

Constraints (6) and (7) determine the production quantities in regular and overtime production hours. Constraint (8) defines the number of workers to be trained in each time period. Constraint (9) is generated based on the company's layoff strategy that forbids firing the hired workers and hiring higher skill workers. Due to the company's training strategy, workers are trained to achieve the required skills. Therefore, as formulated in constraint (9), the number of workers with higher skills in a time period should increase or remain constant compared to the former time period. Constraint (10) ensures that the total number of workers with different skill levels for producing a certain type of product is equal to the number of dedicated workers to the corresponding production line. Constraint (11) considers the limitation associated with the capacity of production lines taking into account performance and availability percentages of production lines. To be more illustrative, the formulas used for obtaining the values of performance and availability percentages of production lines are presented as follows:

$$\begin{aligned} \text{Availability percentage of production line } m (Av_m) \\ & = \left(\frac{\text{Operating time}}{\text{Planned production time}} \right) \times 100, \end{aligned} \quad (17)$$

in which operating time is obtained from subtracting shut down durations (such as equipment failure, material shortage, and changeover time) from the planned production time.

$$\begin{aligned} \text{Performance percentage of production line } m (Pr_m) \\ & = \left(\frac{\text{Net operating time}}{\text{Operating time}} \right) \times 100, \end{aligned} \quad (18)$$

where net operating time is the result of removing speed loss durations (such as machine wear, substandard material, and operator inefficiency) from the operating time.

Constraint (12) states that the backorder level must not exceed the customer demand in each time period. Constraint (13) imposes an obstacle on having any backorder level at the end of the specified time horizon. Constraints (14) and (15) emphasize the limits on the warehouses' capacity for storing finished products and finally constraint (16) identifies the quantity of raw materials to be purchased from the suppliers.

4. Providing a Solution to the Model

In general, in the process of solving a possibilistic fuzzy programming problem, the uncertain nature of parameters imposes two main issues: managing the relationship between the fuzzy sides of constraints and obtaining the optimal value for the objective function which involves some fuzzy parameters. Based on Jiménez et al. [37] the answers for these two questions are related to the process of ranking fuzzy numbers. Many approaches have been introduced in the literature addressing the problem of ranking fuzzy numbers [38]. This study applies the method introduced by Jiménez et al. [37] to rank the fuzzy constraints and objectives. The method uses two main concepts, that are, feasibility and optimality, for dealing with inequality relations in constraints

and ranking fuzzy objective functions, respectively. Unlike some ranking approaches that do not agree with each other, this approach verifies all properties used in other ranking approaches applied for ranking fuzzy numbers [37]. In addition, the method preserves the linearity of a linear programming model that makes it computationally efficient. It also does not increase the number of objective functions and inequality constraints [27]. Therefore, it is suitable for solving large scale fuzzy linear programming models. The method uses fuzzy relation for comparison of fuzzy numbers, while many other relevant methods use comparison relation that does not provide any information about likely violation of constraints (feasibility concept) and just simply state that a fuzzy number is bigger or smaller than others [37]. The feasibility and optimality concepts in this method allow the decision maker to interactively make a trade-off between the degree of violation of constraints (feasibility degree) and the degree of accomplishment of her/his targeted goal. This method is based on an expected interval and expected value of fuzzy numbers, which are considered as the two strong mathematical concepts [39] and were initially introduced by Yager [40] and Dubois and Prade [41] and continued by Heilpern [42] and Jiménez et al. [37]. Prior to explaining the methodology used for solving the constructed fuzzy mathematical model, some relevant terms are defined in the following section.

4.1. Definition of Terms

Fuzzy Number. A fuzzy number is a fuzzy set \tilde{a} on the real line R with the membership function shown in

$$u = \mu_{\tilde{a}}(x) = \begin{cases} 0 & \forall x \in (-\infty, a_1] \\ f_a(x) & \text{increasing } \forall x \in [a_1, a_2] \\ 1 & \forall x \in [a_2, a_3] \\ g_a(x) & \text{decreasing } \forall x \in [a_3, a_4] \\ 0 & \forall x \in [a_4, +\infty) \end{cases} \quad (19)$$

in which $\tilde{a} = (a_1, a_2, a_3, a_4)$.

An α -Cut of a Fuzzy Number \tilde{a} . It is a slice through the fuzzy number \tilde{a} which produces a nonfuzzy set and is defined as $a_\alpha = \{x \in R; \mu_{\tilde{a}}(x) \geq \alpha; 0 < \alpha \leq 1\}$. Based on this definition, it can be written as $a_\alpha = [f_a^{-1}(\alpha), g_a^{-1}(\alpha)]$. In such cases when f_a and g_a are linear functions, the membership function presented in (19) is the membership function of a trapezoidal fuzzy number, denoted by (a_1, a_2, a_3, a_4) . If $a_2 = a_3$, the trapezoidal fuzzy number is converted to the triangular fuzzy number denoted by (a_1, a_2, a_3) [37].

Expected Interval and Expected Value of a Fuzzy Number. Expected interval of a fuzzy number was first introduced by Heilpern [42]. Considering (19), an expected interval of a triangular fuzzy number can be represented as

$$EI(\tilde{a}) = [E_1^a, E_2^a] = \left[\int_0^1 f_a^{-1}(u) du, \int_0^1 g_a^{-1}(u) du \right]. \quad (20)$$

In addition, based on Heilpern's [42] definition, an expected value of a fuzzy number is half point of its expected interval. Then, we have

$$EV(\tilde{a}) = \frac{E_1^a + E_2^a}{2}. \quad (21)$$

Therefore, for a triangular fuzzy number $\tilde{a} = (a_1, a_2, a_3)$, the resulting interval and expected value would be

$$EI(\tilde{a}) = \left[\frac{a_1 + a_2}{2}, \frac{a_2 + a_3}{2} \right], \quad (22)$$

$$EV(\tilde{a}) = \frac{a_1 + 2a_2 + a_3}{4}.$$

It is notable that in some literature, the degree of a decision maker's optimism has been incorporated in calculating the expected interval of a fuzzy number as well. In addition, based on Dubois and Prade [41], for two fuzzy numbers \tilde{a} and \tilde{b} , the following equalities are used:

$$EI(\rho\tilde{a} + \sigma\tilde{b}) = \rho EI(\tilde{a}) + \sigma EI(\tilde{b}), \quad (23)$$

$$EV(\rho\tilde{a} + \sigma\tilde{b}) = \rho EV(\tilde{a}) + \sigma EV(\tilde{b}).$$

Definition 1 (see [43]). For any pair of fuzzy numbers \tilde{a} and \tilde{b} , the degree in which \tilde{a} is bigger than \tilde{b} can be defined as follows:

$$\mu_M(\tilde{a}, \tilde{b}) = \begin{cases} 0 & \text{if } E_2^a - E_1^b < 0 \\ \frac{E_2^a - E_1^b}{E_2^a - E_1^b - (E_1^a - E_2^b)} & \text{if } 0 \in [E_1^a - E_2^b, E_2^a - E_1^b] \\ 1 & \text{if } E_1^a - E_2^b > 0, \end{cases} \quad (24)$$

where $[E_1^a, E_2^a]$ and $[E_1^b, E_2^b]$ are the expected intervals of \tilde{a} and \tilde{b} .

When $\mu_M(\tilde{a}, \tilde{b}) = 0.5$, it is said that \tilde{a} and \tilde{b} are indifferent, and when $\mu_M(\tilde{a}, \tilde{b}) \geq \alpha$ it is said that \tilde{a} is bigger than, or equal to, \tilde{b} at least in degree α and we indicate it by $\tilde{a} \geq_\alpha \tilde{b}$.

Definition 2 (see [43]). Given a decision vector $x \in R^n$, it is feasible in degree α (or α -feasible) if

$$\min_{i=1, \dots, m} \{ \mu_M(\tilde{a}_i x, \tilde{b}_i) \} = \alpha, \quad (25)$$

where $\tilde{a}_i = (\tilde{a}_{i1}, \tilde{a}_{i2}, \dots, \tilde{a}_{im})$; it can be said that

$$\tilde{a}_i x \geq_\alpha \tilde{b}_i, \quad i = 1, \dots, m. \quad (26)$$

Referring to (24), it is equivalent to

$$\frac{E_2^{a_i x} - E_1^{b_i}}{E_2^{a_i x} - E_1^{a_i x} + E_2^{b_i} - E_1^{b_i}} \geq \alpha, \quad i = 1, \dots, m, \quad (27)$$

or

$$[(1 - \alpha) E_2^{a_i} + \alpha E_1^{a_i}] x \geq \alpha E_2^{b_i} + (1 - \alpha) E_1^{b_i}. \quad (28)$$

Definition 2 answers the feasibility issue. In other words, $1 - \alpha$ is the risk of unfeasibility of the solution. However, for the optimality issue, we should refer to Definition 3.

Definition 3 (see [43]). Consider the following classical fuzzy linear programming problem with all fuzzy parameters:

$$\begin{aligned} & \text{Min } \tilde{c}^t x, \\ & x \in \{x \in R^n \mid \tilde{a}_i x \geq \tilde{b}_i, i = 1, \dots, m, x \geq 0\}, \end{aligned} \quad (29)$$

in which $\tilde{c} = (\tilde{c}_1, \tilde{c}_2, \dots, \tilde{c}_n)$ and $\tilde{a}_i = (\tilde{a}_{i1}, \tilde{a}_{i2}, \dots, \tilde{a}_{in})$. All the parameters are described based on triangular fuzzy numbers and decision vector $x \in R^n$ is assumed to be crisp. A vector $x_\alpha^* \in R^n$ is an α -acceptable optimal solution for model (29), if it is an optimal solution to the following crisp linear programming model:

$$\begin{aligned} & \text{Min EV}(\tilde{c}^t) x \\ & x \in \{x \in R^n \mid \tilde{a}_i x \geq \tilde{b}_i, i = 1, \dots, m, x \geq 0\}, \end{aligned} \quad (30)$$

in which $\tilde{c} = (\tilde{c}_1, \tilde{c}_2, \dots, \tilde{c}_n)$ and $\tilde{a}_i = (\tilde{a}_{i1}, \tilde{a}_{i2}, \dots, \tilde{a}_{in})$.

On the other hand, the set of inequality fuzzy constraints $\tilde{a}_i x \geq \tilde{b}_i, i = 1, \dots, m$, can be converted to their equivalent crisp ones based on Expression (28). Also based on Jiménez et al.'s [37] approach, equality constraints of $\tilde{a}_i x = \tilde{b}_i$ are converted to the following crisp inequalities:

$$\begin{aligned} & \left[\left(1 - \frac{\alpha}{2}\right) E_2^{a_i} + \frac{\alpha}{2} E_1^{a_i} \right] x \geq \frac{\alpha}{2} E_2^{b_i} + \left(1 - \frac{\alpha}{2}\right) E_1^{b_i} \\ & \left[\left(1 - \frac{\alpha}{2}\right) E_1^{a_i} + \frac{\alpha}{2} E_2^{a_i} \right] x \leq \left(1 - \frac{\alpha}{2}\right) E_2^{b_i} + \frac{\alpha}{2} E_1^{b_i}. \end{aligned} \quad (31)$$

4.2. Application of Jiménez et al.'s [37] Approach to the Proposed Mathematical Fuzzy Model. As described before, to provide a solution to the model, two main issues of feasibility and optimality of solution must be taken into account. The solution feasibility means to what degree it violates none of the model constraints while the solution optimality implies to what extent the solution achieves the fuzzy goals [44]. In order to answer these main issues, the novel approach of Jiménez et al. [37], with all those details clarified in Section 4.1, are applied to the proposed fuzzy APP model. In brief, the solution to the model will be provided after passing through these phases:

- (i) modeling the imprecise data using triangular fuzzy numbers;
- (ii) converting the multiobjective fuzzy linear model to an equivalent crisp one;
- (iii) solving the resulting multiobjective crisp linear programming model using fuzzy goal programming approach.

4.2.1. Modeling the Imprecise Data Using Triangular Fuzzy Numbers. The first step in solving a fuzzy mathematical model is to represent the uncertain parameters by fuzzy

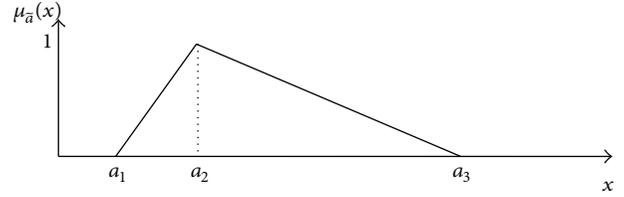


FIGURE 2: Triangular distribution of fuzzy number $\tilde{a} = (a_1, a_2, a_3)$.

numbers. The triangular possibilistic distribution is the most common tool to model the imprecise nature of the fuzzy parameters because of its computational efficiency and simplicity in acquisition of data [45]. In this phase, those imprecise data including forecasted demand, production cost, inventory carrying cost, cost of training workers, cost of purchasing raw materials, reject rate of raw materials, and performance and availability percentages of all production lines are modeled by fuzzy numbers. Figure 2 presents a triangular distribution corresponding to a fuzzy number $\tilde{a} = (a_1, a_2, a_3)$.

The lower bound value a_1 of fuzzy number \tilde{a} shows the most pessimistic value that has a small likelihood to belong to the set of available values (with a membership value of zero if normalized). The value a_2 of fuzzy number \tilde{a} shows the most possible value that certainly belongs to the set of available values (with a membership value of 1 after it is normalized). The upper bound value a_3 as the most optimistic value has a small likelihood to belong to the set of available values (with a membership value of zero if normalized) [46]. Therefore, fuzzy parameters of the proposed model are modeled as follows:

$$\begin{aligned} \tilde{D}_{mt} &= (D_{mt}^1, D_{mt}^2, D_{mt}^3); & \tilde{Pr}_m &= (Pr_m^1, Pr_m^2, Pr_m^3); \\ \tilde{Av}_m &= (Av_m^1, Av_m^2, Av_m^3); & \tilde{C}_{ik} &= (C_{ik}^1, C_{ik}^2, C_{ik}^3); \\ \tilde{\theta}_{ik} &= (\theta_{ik}^1, \theta_{ik}^2, \theta_{ik}^3); & \tilde{C}_{ht} &= (C_{ht}^1, C_{ht}^2, C_{ht}^3); \\ \tilde{C}_{mt}^I &= (C_{mt}^{I1}, C_{mt}^{I2}, C_{mt}^{I3}); & \tilde{C}_{mh}^p &= (C_{mh}^{p1}, C_{mh}^{p2}, C_{mh}^{p3}); \\ \tilde{PA}_m &= \tilde{Pr}_m \times \tilde{Av}_m = (PA_m^1, PA_m^2, PA_m^3). \end{aligned} \quad (32)$$

4.2.2. Converting the Multiobjective Fuzzy Linear Model to an Equivalent Crisp One. Since some of the parameters in the objective functions and constraints are fuzzy numbers, we are faced with both imprecise objectives and imprecise constraints (possibilistic programming). This phase involves the following:

- (i) treating imprecise objective functions (optimality issue);
- (ii) treating imprecise constraints (feasibility issue).

(a) Treating Imprecise Objective Functions. Since there are some triangular fuzzy parameters in the objective functions, we can express them based on a triangular possibilistic distribution. To obtain Z_i^1, Z_i^2, Z_i^3 , all fuzzy parameters in the

objective function Z_i are set at their pessimistic, most likely, and optimistic values, respectively. Therefore, the triangular fuzzy numbers for the first objective function (quantitative objective) can be stated as $Z_1 = (Z_1^1, Z_1^2, Z_1^3)$, in which

$$\begin{aligned} Z_1^1 = & \sum_{t=1}^T \sum_{m=1}^M \sum_{h=1}^H C_{mh}^{p1} P_{mht} \\ & + \sum_{t=1}^T \sum_{m=1}^M \left(C_m^B B_{mt} + C_{mt}^{I1} I_{mt} + C_{ht}^1 L_{mt}^h \right) \\ & + \sum_{t=1}^T \sum_{m=1}^M \sum_{sk=1}^3 S_{sk} L_{skmt} + \sum_{t=1}^T \sum_{k=1}^K \sum_{i=1}^I C_{ik}^1 Q_{ikt} \end{aligned} \quad (33)$$

$$\begin{aligned} Z_1^2 = & \sum_{t=1}^T \sum_{m=1}^M \sum_{h=1}^H C_{mh}^{p2} P_{mht} \\ & + \sum_{t=1}^T \sum_{m=1}^M \left(C_m^B B_{mt} + C_{mt}^{I2} I_{mt} + C_{ht}^2 L_{mt}^h \right) \\ & + \sum_{t=1}^T \sum_{m=1}^M \sum_{sk=1}^3 S_{sk} L_{skmt} + \sum_{t=1}^T \sum_{k=1}^K \sum_{i=1}^I C_{ik}^2 Q_{ikt} \end{aligned} \quad (34)$$

$$\begin{aligned} Z_1^3 = & \sum_{t=1}^T \sum_{m=1}^M \sum_{h=1}^H C_{mh}^{p3} P_{mht} \\ & + \sum_{t=1}^T \sum_{m=1}^M \left(C_m^B B_{mt} + C_{mt}^{I3} I_{mt} + C_{ht}^3 L_{mt}^h \right) \\ & + \sum_{t=1}^T \sum_{m=1}^M \sum_{sk=1}^3 S_{sk} L_{skmt} + \sum_{t=1}^T \sum_{k=1}^K \sum_{i=1}^I C_{ik}^3 Q_{ikt}. \end{aligned} \quad (35)$$

The same approach is applied to the second and third objective functions in order to transform them into their equivalent crisp ones. Based on the methodology explained earlier, this phase is continued by introducing expected values of these objective functions as follows:

$$\begin{aligned} EV_\gamma(\tilde{Z}_1) &= (1-\gamma) \frac{Z_1^1 + Z_1^2}{2} + (\gamma) \frac{Z_1^2 + Z_1^3}{2} \\ EV_\gamma(\tilde{Z}_2) &= (1-\gamma) \frac{Z_2^1 + Z_2^2}{2} + (\gamma) \frac{Z_2^2 + Z_2^3}{2} \\ EV_\gamma(\tilde{Z}_3) &= (1-\gamma) \frac{Z_3^1 + Z_3^2}{2} + (\gamma) \frac{Z_3^2 + Z_3^3}{2}. \end{aligned} \quad (36)$$

Parameter γ defines the degree of a decision maker's optimism and can be varied between zero and one [27]. This study takes a value of 0.3 for parameter γ .

(b) *Treating Imprecise Constraints.* This section addresses the issue of feasibility of solution. Here, based on the ranking approach of Jiménez et al. [37] as presented in (31), all

imprecise (fuzzy) constraints of the model are converted to their equivalent crisp ones as follows:

$$\begin{aligned} \sum_{h=1}^2 P_{mht} &\geq \left(\frac{\alpha}{2}\right) \frac{D_{mt}^2 + D_{mt}^3}{2} + \left(1 - \frac{\alpha}{2}\right) \frac{D_{mt}^1 + D_{mt}^2}{2} - I_{m(t-1)} \\ &\quad - B_{mt} + I_{mt} + B_{m(t-1)}; \quad \forall m, \forall t > 1 \\ \sum_{h=1}^2 P_{mht} &\leq \left(1 - \frac{\alpha}{2}\right) \frac{D_{mt}^2 + D_{mt}^3}{2} + \left(\frac{\alpha}{2}\right) \frac{D_{mt}^1 + D_{mt}^2}{2} - I_{m(t-1)} \\ &\quad - B_{mt} + I_{mt} + B_{m(t-1)}; \quad \forall m, \forall t > 1 \\ \sum_{h=1}^2 P_{mht} &\geq \left(\frac{\alpha}{2}\right) \frac{D_{mt}^2 + D_{mt}^3}{2} + \left(1 - \frac{\alpha}{2}\right) \frac{D_{mt}^1 + D_{mt}^2}{2} \\ &\quad - B_{mt} + I_{mt}; \quad \forall m, \quad t = 1 \\ \sum_{h=1}^2 P_{mht} &\leq \left(1 - \frac{\alpha}{2}\right) \frac{D_{mt}^2 + D_{mt}^3}{2} + \left(\frac{\alpha}{2}\right) \frac{D_{mt}^1 + D_{mt}^2}{2} \\ &\quad - B_{mt} + I_{mt}; \quad \forall m, \quad t = 1 \\ P_{mht} \tau_m &\leq \text{Max}^{ht} \times \left[(1-\alpha) \frac{PA_m^2 + PA_m^3}{2} \right. \\ &\quad \left. + \alpha \frac{PA_m^1 + PA_m^2}{2} \right]; \quad \forall m, \forall t, \forall h \\ B_{mt} &\leq (1-\alpha) \frac{D_{mt}^2 + D_{mt}^3}{2} + \alpha \frac{D_{mt}^1 + D_{mt}^2}{2}; \quad \forall m, \forall t. \end{aligned} \quad (37)$$

α is the feasibility degree [37] of the constraints that has been assigned by the decision maker based on the risk that he/she accepts about the violation of constraints imposed by the obtained solution [44]. In this study, a value of 0.8 has been considered for the parameter α .

4.2.3. *Solving the Resulting Multiobjective Crisp Linear Programming Model.* Passing through stages 1 and 2, as shown in the previous sections, a multiobjective crisp model is obtained as follows:

$$\begin{aligned} \text{Min} \left(EV_\gamma(\tilde{Z}_1) = (1-\gamma) \frac{Z_1^1 + Z_1^2}{2} + (\gamma) \frac{Z_1^2 + Z_1^3}{2} \right) \\ \text{Min} \left(EV_\gamma(\tilde{Z}_2) = (1-\gamma) \frac{Z_2^1 + Z_2^2}{2} + (\gamma) \frac{Z_2^2 + Z_2^3}{2} \right) \\ \text{Max} \left(EV_\gamma(\tilde{Z}_3) = (1-\gamma) \frac{Z_3^1 + Z_3^2}{2} + (\gamma) \frac{Z_3^2 + Z_3^3}{2} \right) \end{aligned} \quad (38)$$

Subject to: Equations (8) – (10);

Equations (13) – (16);

Equation (37).

TABLE 1: A typical payoff table for identification of positive and negative ideal solutions.

	v_1^*	v_2^*	v_3^*
Z_1	$Z_1(v_1^*)$	$Z_1(v_2^*)$	$Z_1(v_3^*)$
Z_2	$Z_2(v_1^*)$	$Z_2(v_2^*)$	$Z_2(v_3^*)$
Z_3	$Z_3(v_1^*)$	$Z_3(v_2^*)$	$Z_3(v_3^*)$

TABLE 2: Payoff table for obtaining positive and negative ideal solutions for the case study.

	v_1^*	v_2^*	v_3^*
$EV_{0.3}(\bar{Z}_1(x))$	6171886216	11746702890	6217647286
$EV_{0.3}(\bar{Z}_2(x))$	6.349	0.192	4.873
$EV_{0.3}(\bar{Z}_3(x))$	5.826	4.402	5.826

TABLE 3: Obtained positive and negative ideal solutions.

	PIS	NIS
$EV_{0.3}(\bar{Z}_1(x))$	6171886216	11746702890
$EV_{0.3}(\bar{Z}_2(x))$	0.192	6.349
$EV_{0.3}(\bar{Z}_3(x))$	5.826	4.402

To solve the resulting multiobjective model, the fuzzy goal programming approach has been applied. This approach involves three different steps including identifying goal values for the defined objective functions, constructing a membership function for each of the objective functions based on the defined goal values, and lastly transforming multiple objectives to a single one by applying an aggregation operator. Consider an objective function Z_i (minimization objective); the corresponding membership function is presented in Figure 3.

Z_i^{PIS} and Z_i^{NIS} are positive and negative ideal solutions of objective function Z_i , respectively. To determine the values of these two parameters, the approach used by Abd El-Wahed and Lee [47] has been followed in this study. Based on their approach, the maximum aspiration level Z_i^{PIS} is obtained by solving the model based on a single objective of Z_i and ignoring other objective functions. However, for obtaining the negative ideal solution of an objective function, one of the following equations should be applied:

$$Z_i^{\text{NIS}} = \max \{Z_i(v_j^*); i \neq j\}$$

in case of having a minimization objective

$$Z_i^{\text{NIS}} = \min \{Z_i(v_j^*); i \neq j\}$$

in case of having a maximization objective,

in which v_j^* is the positive ideal solution of objective function Z_j . A typical payoff table is shown in Table 1.

Once all the membership functions are constructed, the fuzzy goal programming model can be formulated. Here, as

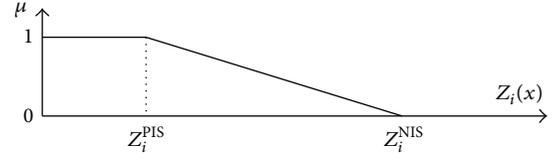


FIGURE 3: A typical membership function for a minimization objective.

the final step, the aim is to formulate an equivalent optimization model taking into account the goal values of all objective functions as well as the feasibility degree of constraints. To do so, the max-min operator of Bellman and Zadeh [48] was selected in order to convert the multiobjective linear model to an ordinary single objective linear programming model. By applying the max-min operator, the final model is derived as follows:

$$\begin{aligned} & \text{Max} && \varphi \\ & \text{Subject to:} && \varphi \leq \mu_{EV_\gamma(\bar{Z}_i(x))}; \quad i = 1, 2, 3 \\ & && \text{Equations (8) – (10);} \\ & && \text{Equations (13) – (16);} \\ & && \text{Equation (37);} \\ & && 0 \leq \varphi \leq 1; \quad x \geq 0. \end{aligned} \tag{40}$$

It is notable that the values of α and γ are assigned by the decision maker. In some cases when an unsatisfactory result is obtained, she/he can change them until a satisfactory result is achieved.

5. Applying the Proposed Model to the Company under Study

Using the IBM ILOG CPLEX Optimization Studio (version 12.4) software and applying the data gathered from the company under study as well as the approach described earlier, a payoff table for identifying positive and negative ideal solutions of each objective function was constructed as shown in Table 2.

Therefore, the positive and negative ideal solutions are obtained as shown in Table 3.

Applying the obtained PISs and NISs and referring to Figure 3, the membership functions can be constructed as depicted in Figures 4, 5, and 6 and formulated in

$$\mu_{EV_{0.3}(\tilde{Z}_1(x))} = \begin{cases} 1 & EV_{0.3}(\tilde{Z}_1(x)) \leq 6171886216 \\ \frac{11746702890 - EV_{0.3}(\tilde{Z}_1(x))}{11746702890 - 6171886216} & 6171886216 < EV_{0.3}(\tilde{Z}_1(x)) \leq 11746702890 \\ 0 & EV_{0.3}(\tilde{Z}_1(x)) > 11746702890 \end{cases}$$

$$\mu_{EV_{0.3}(\tilde{Z}_2(x))} = \begin{cases} 1 & EV_{0.3}(\tilde{Z}_2(x)) \leq 0.192 \\ \frac{6.349 - EV_{0.3}(\tilde{Z}_2(x))}{6.349 - 0.192} & 0.192 < EV_{0.3}(\tilde{Z}_2(x)) \leq 6.349 \\ 0 & EV_{0.3}(\tilde{Z}_2(x)) > 6.349 \end{cases}$$

$$\mu_{EV_{0.3}(\tilde{Z}_3(x))} = \begin{cases} 0 & EV_{0.3}(\tilde{Z}_3(x)) \leq 4.402 \\ \frac{EV_{0.3}(\tilde{Z}_3(x)) - 4.402}{5.826 - 4.402} & 4.402 < EV_{0.3}(\tilde{Z}_3(x)) \leq 5.826 \\ 1 & EV_{0.3}(\tilde{Z}_3(x)) > 5.826. \end{cases} \tag{41}$$

As the final step for solving the constructed fuzzy goal programming model, the max-min operator of Bellman and Zadeh [48] was applied as an aggregation operator to convert the multiobjective linear model to an equivalent single objective one. Therefore, the final model is shown as follows:

Max φ

Subject to: $\varphi \leq \mu_{EV_{0.3}(\tilde{Z}_1(x))}$

$\varphi \leq \mu_{EV_{0.3}(\tilde{Z}_2(x))}$

$\varphi \leq \mu_{EV_{0.3}(\tilde{Z}_3(x))}$

$$\mu_{EV_{0.3}(\tilde{Z}_1(x))} = \frac{11746702890 - EV_{0.3}(\tilde{Z}_1(x))}{11746702890 - 6171886216}$$

$$\mu_{EV_{0.3}(\tilde{Z}_2(x))} = \frac{6.349 - EV_{0.3}(\tilde{Z}_2(x))}{6.349 - 0.192}$$

$$\mu_{EV_{0.3}(\tilde{Z}_3(x))} = \frac{EV_{0.3}(\tilde{Z}_3(x)) - 4.402}{5.826 - 4.402}$$

$$EV_{0.3}(\tilde{Z}_1) = (1-0.3) \frac{Z_1^1 + Z_1^2}{2} + (0.3) \frac{Z_1^2 + Z_1^3}{2}$$

$$EV_{0.3}(\tilde{Z}_2) = (1-0.3) \frac{Z_2^1 + Z_2^2}{2} + (0.3) \frac{Z_2^2 + Z_2^3}{2}$$

$$EV_{0.3}(\tilde{Z}_3) = (1-0.3) \frac{Z_3^1 + Z_3^2}{2} + (0.3) \frac{Z_3^2 + Z_3^3}{2}$$

Equations (8) – (10);

Equations (13) – (16);

Equation (37);

$$0 \leq \varphi \leq 1; \quad x \geq 0,$$

(42)

where $\alpha = 0.8$ is substituted in (37).

6. Discussion of Results

Using the aforementioned software, the detailed solutions to the above model were obtained as presented in Tables 4, 5, and 6. As stated earlier, the values for parameters α and γ have been considered as 0.8 and 0.3, respectively.

Table 4 contains the optimal values obtained for each of the objective functions. As can be seen, the obtained optimal values of objective functions are relatively close to the aspiration levels that have been defined in (41) and presented in Figures 4 to 6. Table 5 presents the results obtained for some decision variables, including inventory and backorder levels, production quantities, and number of workers to be trained in different time periods. Investigating these results shows that incurring inventory and having overtime production are more preferred than having backorders in many time periods. The reason can be attributed to the higher penalty cost of backorders compared to the cost of carrying inventory or/and having overtime production in the company under study. In addition, considering the third objective function that aims to increase customer service level by reducing backorder levels has its own effect on obtaining such results.

Based on the results obtained for the number of workers to be trained (the last column of Table 5) and the results shown in Table 6 which suggest the number of workers with certain level of skill to be hired at the beginning of each time period, it is concluded that hiring workers with the required skill levels and avoiding training cost is more beneficial for the company under study.

Results related to purchasing raw materials Q_{ikt} which are not shown because of space limitation determine the

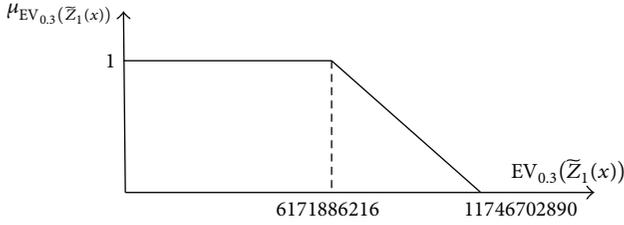


FIGURE 4: Membership function for the first objective.

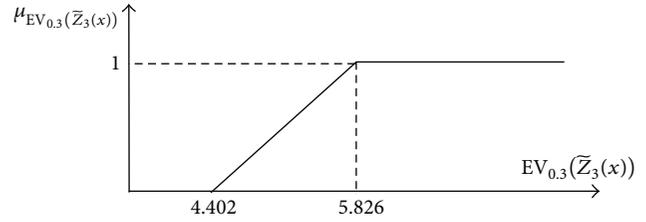


FIGURE 6: Membership function for the third objective.

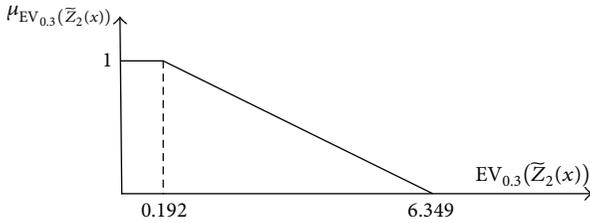


FIGURE 5: Membership function for the second objective.

TABLE 4: Results obtained for objective functions and φ .

1st objective	6277251051
2nd objective	0.31
3rd objective	5.83
φ	0.98

quantity of different types of raw materials to be purchased from different suppliers in various time periods, taking into account two important elements, that are, cost of purchasing and lower reject rate of purchased components as a measure of quality.

As one of the contributions of this research is to signify the importance of considering the performance and availability levels of production lines in developing an APP, a comparison was performed between the results obtained with and without considering these two parameters in the model.

Figures 7, 8, and 9 present a comparison between the results obtained for the three objective functions, that are, minimization of cost, maximization of quality (minimization of quality degradation), and maximization of customer service level, with and without considering the performance and availability of production lines. In addition, Figures 10 to 12 provide the same comparison between the results obtained for some decision variables including production quantity (in regular time and overtime), inventory level, and backorder level, respectively. The differences between the results obtained illustrate the role of the two factors (performance and availability levels) in developing a practical and real APP. Failure to take into account these two parameters can lead planners to inaccurate results which are completely different from the actual results obtained in their presence. The differences between the outputs in the two considered situations are due to the overestimation of production capacity stemming from ignoring the level of performance and

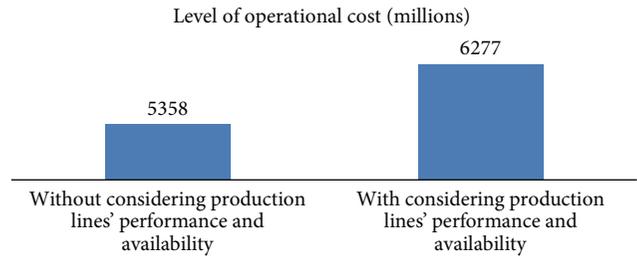


FIGURE 7: Comparison between the results obtained for the first objective function.

availability of production lines. In other words, unrealistic assumption about the capacity of production lines results in ineffective production planning. For example, the quantity of products that can be generated during regular time has been overestimated and the required quantity to be produced during overtime has been underestimated (Figure 10). Figures 11 and 12 also show the underestimation of inventory and backorder levels, respectively. The overestimation of production quantity in regular time is due to the overestimation of production capacity that allows generating more products in regular time in order to lower the production quantity in overtime and consequently the associated cost. The higher levels of total inventory and backorder, when considering the performance and availability parameters, can be associated with the lack of capacity for producing the required number of products in the corresponding time periods. Therefore, the company should compensate for the shortage by incurring higher inventory and carrying more backorder levels. Underestimating backorder and inventory levels can have at least two consequences. The first one results in underestimation of operational cost due to underestimation of backorder and inventory levels as the two main portions of the total operational cost (Figure 7) and the second consequence manifests itself in producing behind schedule and having inaccurate backorder levels which ultimately leads to customer dissatisfaction and sales loss. Therefore, in order to avoid developing an impractical APP with unrealistic estimation of resources on hand, it is necessary to take into account performance and availability factors when developing a practical APP.

TABLE 5: Results corresponding to inventory and backorder levels, production quantity, and number of workers to train.

Product	Time period	Inventory level	Backorder level	Regular time production quantity	Overtime production quantity	Number of workers to train
1	1	9031	0	8445	5278	0
1	2	4538	0	8445	5278	0
1	3	3081	0	8445	5278	0
1	4	0	6587	8445	5278	0
1	5	0	9424	8445	5278	0
1	6	0	0	7082	5278	0
2	1	3265	0	0	4579	0
2	2	7312	0	4800	4579	0
2	3	5018	0	0	0	0
2	4	0	0	0	0	0
2	5	2837	0	7326	0	0
2	6	0	0	0	3517	0
3	1	7981	0	7420	10929	0
3	2	19260	0	17487	10929	0
3	3	8679	0	17487	10929	0
3	4	20205	0	17487	10929	0
3	5	18433	0	17487	10929	0
3	6	0	0	17487	10929	0

TABLE 6: Number of workers with the required skill levels in different time periods.

Product	Skill level	Time period	Number of workers	Product	Skill level	Time period	Number of workers	Product	Skill level	Time period	Number of workers
1	1	1	0	2	1	1	0	3	1	1	0
1	1	2	0	2	1	2	0	3	1	2	0
1	1	3	0	2	1	3	0	3	1	3	0
1	1	4	0	2	1	4	0	3	1	4	0
1	1	5	0	2	1	5	0	3	1	5	0
1	1	6	0	2	1	6	0	3	1	6	0
1	2	1	10	2	2	1	10	3	2	1	6
1	2	2	10	2	2	2	10	3	2	2	6
1	2	3	10	2	2	3	10	3	2	3	6
1	2	4	10	2	2	4	10	3	2	4	6
1	2	5	10	2	2	5	10	3	2	5	6
1	2	6	10	2	2	6	10	3	2	6	6
1	3	1	0	2	3	1	0	3	3	1	0
1	3	2	0	2	3	2	0	3	3	2	0
1	3	3	0	2	3	3	0	3	3	3	0
1	3	4	0	2	3	4	0	3	3	4	0
1	3	5	0	2	3	5	0	3	3	5	0
1	3	6	0	2	3	6	0	3	3	6	0

7. Sensitivity Analysis

Figures 13 to 15 provide information about sensitivity analysis, which was performed on parameters α and γ . As clarified earlier, these two parameters were arbitrarily set by the

decision maker. In selecting these parameters, the decision maker ran into performing a trade-off between obtaining a better optimal value which is closer to the targeted aspiration level of the objective function (the set composed of those values whose membership functions are equal to one) and

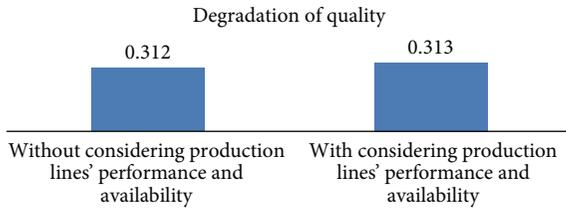


FIGURE 8: Comparison between the results obtained for the second objective function.

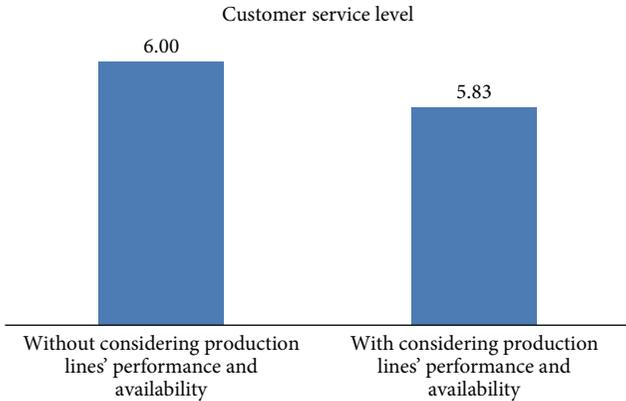


FIGURE 9: Comparison between the results obtained for the third objective function.

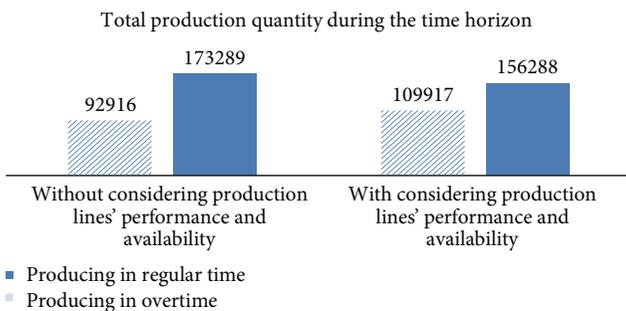


FIGURE 10: Comparison between total production quantities (assigned to all products) during the time horizon.

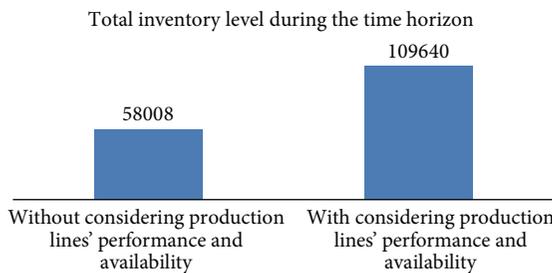


FIGURE 11: Comparison between total inventory levels (assigned to all products) during the time horizon.

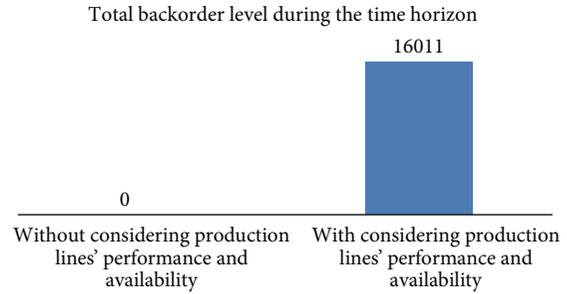


FIGURE 12: Comparison between total backorder levels (assigned to all products) during the time horizon.

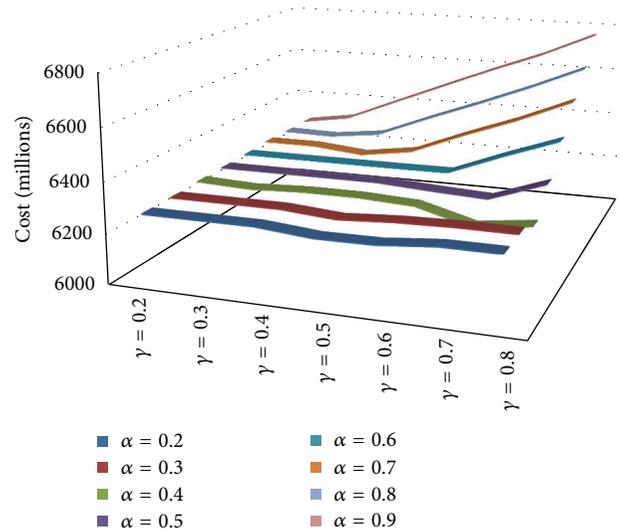


FIGURE 13: Sensitivity analysis on the first objective function.

having a higher degree of satisfaction of constraints, that is, selecting a higher feasibility degree. A general view of the results associated with the performed sensitivity analysis reflects the fact that a better optimal value of an objective function requires a lesser degree of feasibility (a higher risk of violating constraints) [37]. As can be generally observed from Figure 13, better optimal values for the first objective function, that are, values with higher membership degrees for the set $\{x \in R \mid 6171886216 \leq x \leq 11746702890\}$ (see Figure 4), are obtained at lesser degrees of feasibility (lesser values of parameter α). For example, for a certain value of γ in Figure 13, increasing the value of parameter α generally results in worsening the value of cost (higher cost) and obtaining optimal solutions with a lower membership degree for the set introduced above. Although the second and third objective functions (Figures 14 and 15, resp.) show relatively subtle sensitivity, the trends shown in these two figures do not reveal conflicts with the aforementioned fact.

8. Conclusions

In this paper, an attempt was made to propose an integer linear programming model for APP of an automotive parts

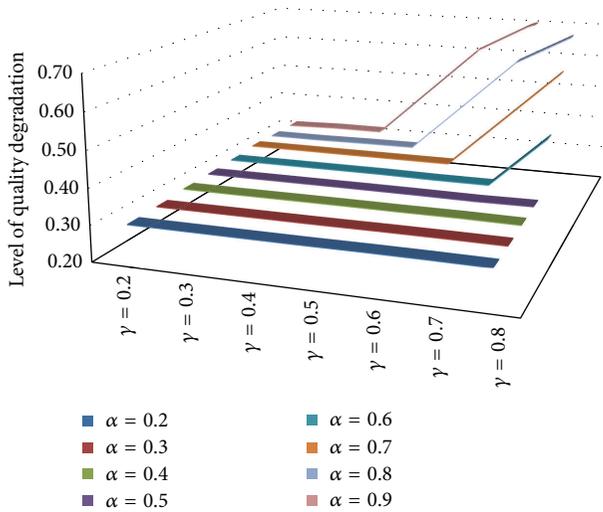


FIGURE 14: Sensitivity analysis on the second objective function.

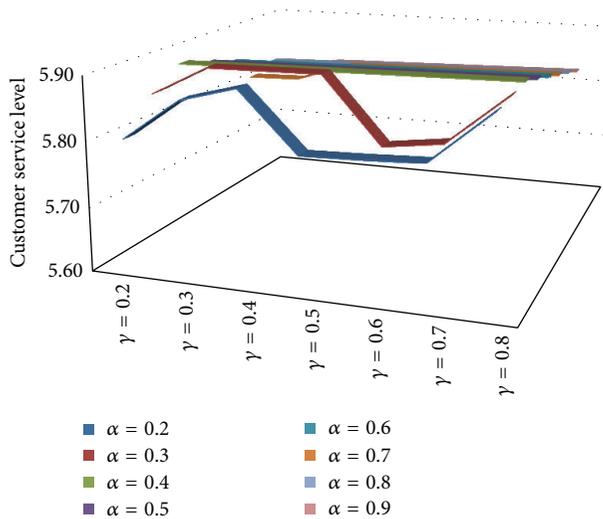


FIGURE 15: Sensitivity analysis on the third objective function.

manufacturing company in a fuzzy environment that takes into account quality of products as well as other common objectives used in the literature which are minimization of cost and maximization of customer service level. Since, in the real world, many planners have some qualitative objectives expressed in the form of linguistic terms, an effort was made to incorporate a qualitative objective function in the construction of the mathematical model. To make the plan more practical, it has taken into account the real capacity of production lines when formulating the model through the consideration of two important factors which are performance and availability of production lines. The model was solved using the IBM ILOG CPLEX software. To highlight the importance of considering the performance and availability of production lines, a comparison was made between the results obtained from solving the model with and without the consideration of these two factors. The

comparison obviously showed that ignoring these two factors can result in generating an inaccurate plan which is not practical in reality.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] A. Jain and U. S. Palekar, "Aggregate production planning for a continuous reconfigurable manufacturing process," *Computers and Operations Research*, vol. 32, no. 5, pp. 1213–1236, 2005.
- [2] R. C. Wang and T. F. Liang, "Application of fuzzy multi-objective linear programming to aggregate production planning," *Computers and Industrial Engineering*, vol. 46, no. 1, pp. 17–41, 2004.
- [3] R. Zhang, L. Zhang, Y. Xiao, and I. Kaku, "The activity-based aggregate production planning with capacity expansion in manufacturing systems," *Computers & Industrial Engineering*, vol. 62, no. 2, pp. 491–503, 2012.
- [4] S. Chopra and P. Meindl, *Supply Chain Management. Strategy, Planning, and Operation*, Gabler, 2007.
- [5] J. Tang, D. Wang, and R. Y. K. Fung, "Fuzzy formulation for multi-product aggregate production planning," *Production Planning and Control*, vol. 11, no. 7, pp. 670–676, 2000.
- [6] C. Gomes da Silva, J. Figueira, J. Lisboa, and S. Barman, "An interactive decision support system for an aggregate production planning model based on multiple criteria mixed integer linear programming," *Omega*, vol. 34, no. 2, pp. 167–177, 2006.
- [7] R. Ramezani, D. Rahmani, and F. Barzinpour, "An aggregate production planning model for two phase production systems: solving with genetic algorithm and tabu search," *Expert Systems with Applications*, vol. 39, no. 1, pp. 1256–1263, 2012.
- [8] A. P. Kanyalkar and G. K. Adil, "An integrated aggregate and detailed planning in a multi-site production environment using linear programming," *International Journal of Production Research*, vol. 43, no. 20, pp. 4431–4454, 2005.
- [9] V. Jayaraman, "Production planning for closed-loop supply chains with product recovery and reuse: an analytical approach," *International Journal of Production Research*, vol. 44, no. 5, pp. 981–998, 2006.
- [10] T. Sillekens, A. Koberstein, and L. Suhl, "Aggregate production planning in the automotive industry with special consideration of workforce flexibility," *International Journal of Production Research*, vol. 49, no. 17, pp. 5055–5078, 2011.
- [11] I. T. Christou, A. G. Lagodimos, and D. Lycopoulou, "Hierarchical production planning for multi-product lines in the beverage industry," *Production Planning and Control*, vol. 18, no. 5, pp. 367–376, 2007.
- [12] C. G. Vassiliadis, M. G. Vassiliadou, L. G. Papageorgiou, and E. N. Pistikopoulos, "Simultaneous maintenance considerations and production planning in multi-purpose plants," in *Proceedings of the Annual Reliability and Maintainability Symposium*, pp. 228–233, IEEE, 2000.
- [13] H. Suryadi and L. G. Papageorgiou, "Optimal maintenance planning and crew allocation for multipurpose batch plants," *International Journal of Production Research*, vol. 42, no. 2, pp. 355–377, 2004.
- [14] A. K. Rifai, "A note on the structure of the goal-programming model: assessment and evaluation," *International Journal of*

- Operations & Production Management*, vol. 16, no. 1, pp. 40–49, 1996.
- [15] S. C. H. Leung and S. S. W. Chan, “A goal programming model for aggregate production planning with resource utilization constraint,” *Computers and Industrial Engineering*, vol. 56, no. 3, pp. 1053–1064, 2009.
- [16] C. Romero, “A general structure of achievement function for a goal programming model,” *European Journal of Operational Research*, vol. 153, no. 3, pp. 675–686, 2004.
- [17] A. Baykasoglu, “MOAPPS 1.0: aggregate production planning using the multiple-objective tabu search,” *International Journal of Production Research*, vol. 39, no. 16, pp. 3685–3702, 2001.
- [18] Z. H. Che and C. J. Chiang, “A modified Pareto genetic algorithm for multi-objective build-to-order supply chain planning with product assembly,” *Advances in Engineering Software*, vol. 41, no. 7–8, pp. 1011–1022, 2010.
- [19] Z. Liu, D. K. H. Chua, and K. Yeoh, “Aggregate production planning for shipbuilding with variation-inventory trade-offs,” *International Journal of Production Research*, vol. 49, no. 20, pp. 6249–6272, 2011.
- [20] A. S. M. Masud and C. L. Hwang, “An aggregate production planning model and application of three multiple objective decision methods,” *International Journal of Production Research*, vol. 18, no. 6, pp. 741–752, 1980.
- [21] H. S. Yan, Q. Xia, M. Zhu, X. Liu, and Z. Guo, “Integrated production planning and scheduling on automobile assembly lines,” *IIE Transactions*, vol. 35, no. 8, pp. 711–725, 2003.
- [22] L. Pradenas, F. Peñailillo, and J. Ferland, “Aggregate production planning problem. A new algorithm,” *Electronic Notes in Discrete Mathematics*, vol. 18, pp. 193–199, 2004.
- [23] A. Jamalnia and M. A. Soukhakian, “A hybrid fuzzy goal programming approach with different goal priorities to aggregate production planning,” *Computers and Industrial Engineering*, vol. 56, no. 4, pp. 1474–1486, 2009.
- [24] S. M. J. Mirzapour Al-E-Hashem, H. Malekly, and M. B. Aryanezhad, “A multi-objective robust optimization model for multi-product multi-site aggregate production planning in a supply chain under uncertainty,” *International Journal of Production Economics*, vol. 134, no. 1, pp. 28–42, 2011.
- [25] H. J. Zimmermann, “Description and optimization of fuzzy systems,” *International Journal of General Systems*, vol. 2, no. 4, pp. 209–215, 1976.
- [26] H. J. Zimmermann, “Fuzzy programming and linear programming with several objective functions,” *Fuzzy Sets and Systems*, vol. 1, no. 1, pp. 45–55, 1978.
- [27] R. G. Yaghin, S. A. Torabi, and S. M. T. F. Ghomi, “Integrated markdown pricing and aggregate production planning in a two echelon supply chain: a hybrid fuzzy multiple objective approach,” *Applied Mathematical Modelling*, vol. 36, no. 12, pp. 6011–6030, 2012.
- [28] R. Narasimhan, “Goal programming in a fuzzy environment,” *Decision Sciences*, vol. 11, no. 2, pp. 325–336, 1980.
- [29] E. L. Hannan, “Linear programming with multiple fuzzy goals,” *Fuzzy Sets and Systems*, vol. 6, no. 3, pp. 235–248, 1981.
- [30] B. K. Wong and V. S. Lai, “A survey of the application of fuzzy set theory in production and operations management: 1998–2009,” *International Journal of Production Economics*, vol. 129, no. 1, pp. 157–168, 2011.
- [31] D. B. Rinks, “The performance of fuzzy algorithm models for aggregate planning under differing cost structures,” in *Fuzzy Information and Decision Processes*, M. M. Gupta and E. Sanchez, Eds., pp. 267–278, North-Holland, Amsterdam, The Netherlands, 1982.
- [32] Y. Y. Lee, *Fuzzy Sets Theory Approach to Aggregate Production Planning and Inventory Control*, UMI, 1990.
- [33] M. Gen, Y. Tsujimura, and K. Ida, “Method for solving multiobjective aggregate production planning problem with fuzzy parameters,” *Computers & Industrial Engineering*, vol. 23, no. 1–4, pp. 117–120, 1992.
- [34] R. Wang and H. Fang, “Aggregate production planning with multiple objectives in a fuzzy environment,” *European Journal of Operational Research*, vol. 133, no. 3, pp. 521–536, 2001.
- [35] R. Wang and T. Liang, “Application of fuzzy multi-objective linear programming to aggregate production planning,” *Computers and Industrial Engineering*, vol. 46, no. 1, pp. 17–41, 2004.
- [36] L. A. Zadeh, “Fuzzy sets,” *Information and Computation*, vol. 8, pp. 338–353, 1965.
- [37] M. Jiménez, M. Arenas, A. Bilbao, and M. V. Rodríguez, “Linear programming with fuzzy parameters: an interactive method resolution,” *European Journal of Operational Research*, vol. 177, no. 3, pp. 1599–1609, 2007.
- [38] Y. Lai and C. Hwang, “A new approach to some possibilistic linear programming problems,” *Fuzzy Sets and Systems*, vol. 49, no. 2, pp. 121–133, 1992.
- [39] M. S. Pishvaei and S. A. Torabi, “A possibilistic programming approach for closed-loop supply chain network design under uncertainty,” *Fuzzy Sets and Systems*, vol. 161, no. 20, pp. 2668–2683, 2010.
- [40] R. R. Yager, “A procedure for ordering fuzzy subsets of the unit interval,” *Information Sciences*, vol. 24, no. 2, pp. 143–161, 1981.
- [41] D. Dubois and H. Prade, “The mean value of a fuzzy number,” *Fuzzy Sets and Systems*, vol. 24, no. 3, pp. 279–300, 1987.
- [42] S. Heilpern, “The expected value of a fuzzy number,” *Fuzzy Sets and Systems*, vol. 47, no. 1, pp. 81–86, 1992.
- [43] M. Jiménez, “Ranking fuzzy numbers through the comparison of its expected intervals,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 4, no. 4, pp. 379–388, 1996.
- [44] M. M. Lotfi and S. A. Torabi, “A fuzzy goal programming approach for mid-term assortment planning in supermarkets,” *European Journal of Operational Research*, vol. 213, no. 2, pp. 430–441, 2011.
- [45] S. A. Torabi and E. Hassini, “An interactive possibilistic programming approach for multiple objective supply chain master planning,” *Fuzzy Sets and Systems*, vol. 159, no. 2, pp. 193–214, 2008.
- [46] T.-F. Liang, “Application of fuzzy sets to manufacturing/distribution planning decisions in supply chains,” *Information Sciences*, vol. 181, no. 4, pp. 842–854, 2011.
- [47] W. F. Abd El-Wahed and S. M. Lee, “Interactive fuzzy goal programming for multi-objective transportation problems,” *Omega*, vol. 34, no. 2, pp. 158–166, 2006.
- [48] R. E. Bellman and L. A. Zadeh, “Decision-making in a fuzzy environment,” *Management Science*, vol. 17, pp. 141–161, 1970.

Research Article

A Transient Queuing Model for Analyzing and Optimizing Gate Congestion of Railway Container Terminals

Ming Zeng, Wenming Cheng, and Peng Guo

School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China

Correspondence should be addressed to Ming Zeng; zengming1209@gmail.com

Received 3 April 2014; Revised 12 June 2014; Accepted 12 June 2014; Published 23 July 2014

Academic Editor: Michael Lütjen

Copyright © 2014 Ming Zeng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As the significant connection between the external and internal of the railway container terminal, the operation performance of the gate system plays an important role in the entire system. So the gate congestion will bring many losses to the railway container terminal, even the entire railway container freight system. In this paper, the queue length and the average waiting time of the railway container terminal gate system, as well as the optimal number of service channels during the different time period, are investigated. An $M/E_k/n$ transient queuing model is developed based on the distribution of the arrival time interval and the service time; besides the transient solutions are acquired by the equally likely combinations (ELC) heuristic method. Then the model is integrated into an optimization framework to obtain the optimal operation schemes. Finally, some computational experiments are conducted for model validation, sensitivity testing, and system optimization. Experimental results indicate that the model can provide the accurate reflection to the operation situation of the railway container terminal gate system, and the approach can yield the optimal number of service channels within the reasonable computation time.

1. Introduction

Nowadays, the container transportation has been fast developed throughout the world. It becomes the best mode of transportation for the commercial intercourse of various countries, and the containerization ratio of the general cargos in some countries has exceeded 80%. The statistics of the United Nations Conference on Trade and Development (UNCTAD) present that Germany is the first country of European Union containerized exports and imports, which has 3.04 million TEUS in export and 2.84 million TEUS in import in 2010. Meanwhile, China ranks as the first of all the exporters of containerized cargo in 2010, which has 31.3 million TEUS, and it also has 12.0 million TEUS import volume of containerized cargo in 2010 which is ranked the second. The first country is the United States which has 17.6 million TEUS [1]. The railway container transportation was firstly carried out in the United States in 1853, and China began this practice in the 1950s. In the recent years, the railway container transportation volume of China is on a rapid growth. Compared with 2000, the dispatch volume of

railway container cargos increased by 60.1% in 2005, and the average growth rate per annum of railway container transportation volume in China is 11.6% in 2000–2004. In 2006, the volume has a 16.7% increase to the year 2005, that is, 68.91 million tons. Furthermore, the container transportation volume is as much as 471 TEU which is equivalent to 90.65 million tons in 2012 [2].

According to the huge demand, China increases the investment to the relevant facilities and the infrastructure construction and has established a railway container transportation support system which takes 18 railway container terminals as the hub. Railway container terminals are always established in the main economic centers and the important ports, which are the collecting and distributing centre of the railway container transportation. Railway container terminal has many functions as marshalling, declassification, handling, warehousing, and some other logistics service. Nonetheless, with the constant development of the regional economy and the rapid growth of transportation volume, there still are many challenges to the railway container terminals. Generally, the maximum backlog time of the combined

type marshalling cargos in the container terminals may be up to 15–20 days. Hence, the improving of the operation efficiency in railway container terminal under the existing conditions is a serious problem at present. As one of the most important procedures in the railway container terminal operation system, the through capacity and service level of the gate system have significant impacts on the efficiency of the railway container terminal, even the whole railway container freight system. Currently, the operation management of the railway container terminal gate system relies on the empirical estimation and the real time information feedback in the most cases. However, these methods are not accurate and efficient in reality, so they are not able to solve this problem fundamentally.

Recently, the researchers pay more attention to the study of container terminals. Meanwhile, their research areas mainly focus on the operation management and the equipment scheduling. Mattfeld and Kopfer [3] developed an integral decision model for the operations of the terminal. Kim et al. [4] proposed a dynamic programming model to evaluate the delay time of outside trucks in port container terminals and discussed the different situations, respectively. Lee et al. [5] presented a mixed integer programming model together with a genetic algorithm to solve the quay crane scheduling problem and then extended the problem with the noninterference constraints [6]. Similar problem was also studied by using different constraints and algorithms [7–9]. In addition, the relevant layout planning [10, 11], the berth allocation [12], and the programming problems [13–18] are also researched comprehensively.

Similarly, the researches of the railway container terminals and the intermodal terminals are conducted in these fields, and the loading/discharging equipment scheduling problems have become the hot issues [19–21]. Guo et al. [22] presented a mixed integer programming model with a discrete artificial bee colony algorithm to solve the gantry crane scheduling problem in railway container terminals. Dorndorf and Schneider [23] studied the triple cross-over stacking cranes scheduling problem in order to increase the productivity and reduce the delays. Besides, the operation researches, for example, the loading planning [24, 25], the shunting of rail container wagons [26], the container stacking [27], and the layout planning like the crane areas determination [28, 29], are some other research interests in this field. The strategic planning and the scenario generating for the railway container terminals are usually implemented by the simulation methods [30–32]. Guo et al. [33] developed a discrete event simulation model for the container handling process of the railway container logistic center to make the evaluation and improvement. However, there exist some disadvantages in the simulation methods; for example, the quantitative solutions cannot be provided and they are unable to be nested with the optimization models. Therefore, Edmond and Maggs [34] have pointed out the importance of queuing theory in the related decision making of container terminals many years before. Recently, some literature began to use the queuing methods to analyze the problems. For instance, Canonaco et al. [35] proposed a queuing network model to solve the machine operations problem of the container terminal. But

the researches of queuing models for the railway container terminal gate system are still relatively small, because the railway container terminal gate system cannot be described completely and accurately by the traditional queuing models, which are also incapable of analyzing the instantaneous features of this system, so there are some limitations in this field.

The queuing models can be classified into stationary queuing models and transient queuing models. Because it is difficult to describe and calculate all the system states of the transient queuing models, the majority of the queuing models which are applied in various important research areas only take the steady solution into consideration [36, 37]. However, the theory and application research of transient queuing models has gone through some development in the past ten years. Ausin et al. [38] used the Bayesian analysis method to solve the optimized number of servers in the GI/M/c queuing system on the basis of the minimum total costs. Besides, they utilized the Bayesian inference to predict the transient features of the GI/G/1 system [39]. Czachórski et al. [40] studied the transient features of the G/G/1 queuing system from the diffusion approximation model. Parlar and Sharafali [41] created a time-related queuing model to optimize the airport security checkpoints and briefly discussed a general condition that the service time follows the Erlang distribution.

In this paper, a transient queuing model and an optimization model are established to analyze and optimize the gate congestion according to the arrival time interval of the external container trucks and the distribution regularity of the service time for the railway container terminal gate system. Then the equally likely combinations heuristic solution and the optimization solution methods are adopted to solve these models. After that, the system simulation and sensitivity analysis are conducted to verify the rationality and validity of our approach. Finally, the optimal numbers of the service channels in different time periods are determined by the optimization experiment. The remainder of this paper is structured as follows. Section 2 provides some background information to the case study. The queuing model and optimization model together with the solution methods are presented in Section 3. The model validation, sensitivity analysis, and the optimization experiments are given in Section 4. Finally, Section 5 concludes the paper with an outline of future study.

2. Chengdu Railway Container Terminal

Chengdu railway container terminal is located in the north-east of Chengdu, where it is near the Chengxiang railway station which connects with many key rail routes of China. Chengdu railway container terminal is the largest railway container terminal of Asia, which is also the container logistics hinge of southwest China. Therefore, it acts as an important part to the entire container transportation network. The terminal was built in 2010, which is 8.4 kilometers long and 850 meters wide. The annual cargo throughput is 1 million TEU in recent period, and the forward cargo handling capacity will be 2.5 million TEU.

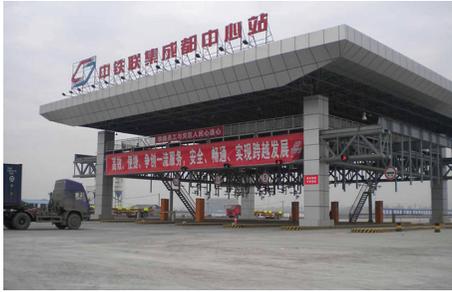


FIGURE 1: Chengdu railway container terminal gate system.

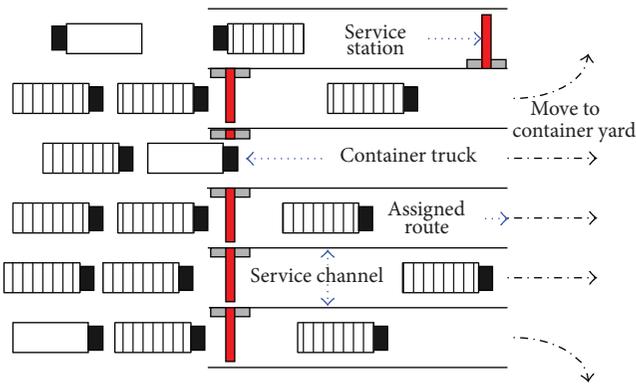


FIGURE 2: Schematic representation of the railway container terminal gate system.

Railway container terminal gate system is the access of the container trucks to go through. There are some service channels which have the technical facilities to conduct the necessary processes in the railway container terminal gate system. A picture of Chengdu railway container terminal gate system can be seen in Figure 1. The major works at the service channels are container number identification, information verification, inspection, and position assignment. The railway container terminal gate system is schematically depicted in Figure 2. However, because of the limitations of the facilities and the service conditions, the container trucks have to line up to complete the processes before they enter into or get out of the terminal. Therefore, the whole processes can be seen as a queuing process, in which the container trucks are the input flows and the channels are the service stations.

3. Railway Container Terminal Gate System Modelling

3.1. $M/E_k/n$ Queuing Model. All the data with respect to the time intervals of two container trucks that arrive in succession and the service time of the service channels were collected to ascertain the best fitting distributions. 560 arrival time intervals of container trucks and 350 service times of the service channels were field-measured in Chengdu railway container terminal gate system for the observation period of two weeks, and then probability information of the collected

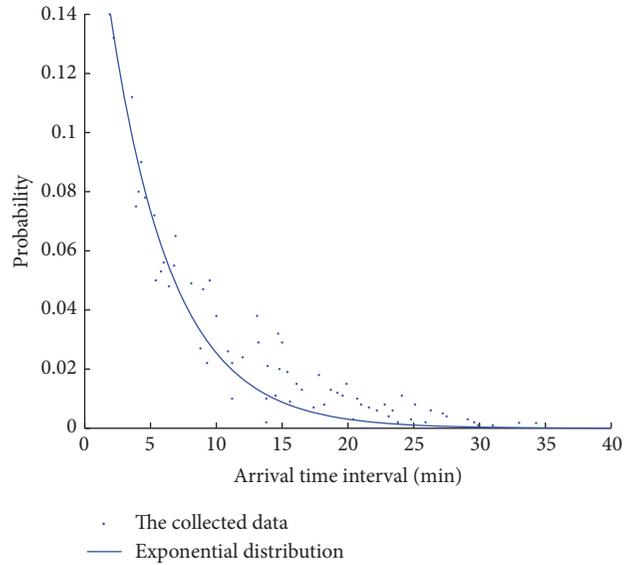


FIGURE 3: The distribution of arrival time interval.

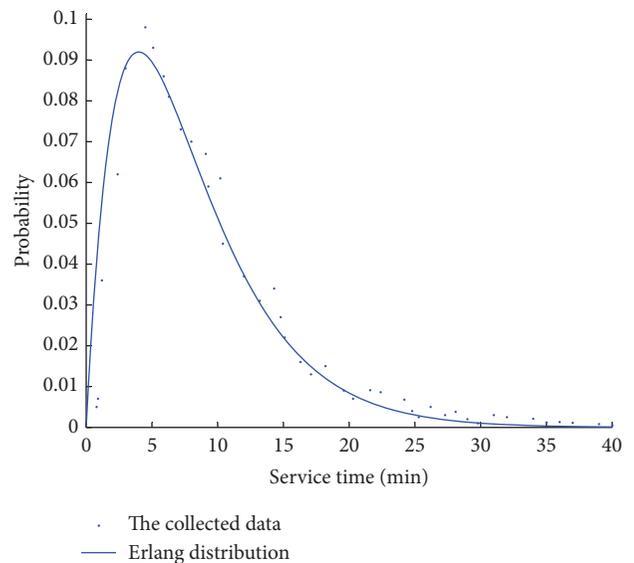


FIGURE 4: The distribution of service time.

data along with the relevant distribution curves was acquired as follows.

The best fitting distribution of the arrival time intervals as displayed in Figure 3 is the exponential distribution $f(x) = 0.212e^{-0.212x}$, $x \geq 0$, and the average arrival time interval is 4.714 minutes. The fit of arrival time interval distribution to the collected data is calculated by R -square which equals 0.9542 and RMSE (root mean square error) which is 0.012. The distribution of the service time as shown in Figure 4 is best fitted by the Erlang distribution $f(x) = 0.0625 \cdot x \cdot e^{-0.25x}$, $x \geq 0$, with k equal to 2, and the mean service rate is 0.25. The R -square value for fitting the service time distribution to the collected data is 0.9078, and the RMSE is 0.096. Thus, the appropriate queuing model for the railway

container terminal gate system is exponential arrival intervals with Erlang service times and multiple service stations, which is represented as $M/E_{k=2}/n$. The solution for this model can be divided into two types: the exact solution method and the approximate solution method. However, the former has a large number of possible states and the state transitions are complicated; the calculation is quite difficult. So here we utilize the approximate solution method to solve the $M/E_k/n$ queuing model. Equally likely combinations (ELC) heuristic solution technique has been put forward by Escobar et al. [42], the basic assumption of which is the combinations of the stages unfinished in the servers are equally likely. The detailed description of this heuristic solution method is as follows.

3.1.1. System States Description. For the sake of simplifying the Erlang distribution and then forming an efficient solution, the system states are represented by three elements and (s, v, m) ; s means the stages have to be finished by the container trucks in the system at this moment; v is the total number of container trucks in the system at present; m refers to the mode number. The reason for introducing the third element is to distinguish the different modes with the same s and v . For example, the state $(7, 4)$, as shown in Figure 5, either means four trucks in the system with two trucks having one stage unfinished, one having two stages unfinished, and the other one just arriving (i.e., Mode 1) or represents one truck having one stage left over and the other three all having two stages (i.e., Mode 2).

3.1.2. State Transition Probability. Before doing the description of the state transition probability, firstly let $P_{s,v}$ represent the probability of state (s, v) , because the relevant calculation will be conducted, so the mode number is discarded for conciseness. Compared with the original ELC solution which considers new trucks will arrive at each time step, in this paper, we take the average arrival time interval into account, in order to reduce the state space dimension and calculation difficulty.

Suppose there are v container trucks and s stages have to be finished at current time t for the next time step (i.e., $t + 1$); the possible states and state transition are as below.

Let t_a denote the average time interval between two trucks arriving at the gate system and let t_N represent the N th truck had arrived. For the next time step, before the $(N + 1)$ th truck arrived, there are two possible transitions that may happen when a truck finishes one stage.

(a) The stage is the last one to finish for a truck, and the truck leaves the gate system.

Under this situation, the state is transferred from (s, v) to $(s - 1, v - 1)$. As mentioned in Section 3.1.1, there are different modes for the same state and different combinations for the same mode. Therefore, the combinations for one given mode I can be calculated as follows:

$$C_i = \frac{\omega!}{\omega_1! \omega_2! \cdots \omega_p!}, \quad (1)$$

where ω is the $\min\{v, n\}$, which means the active service channels; p is the number of combinations of stages in

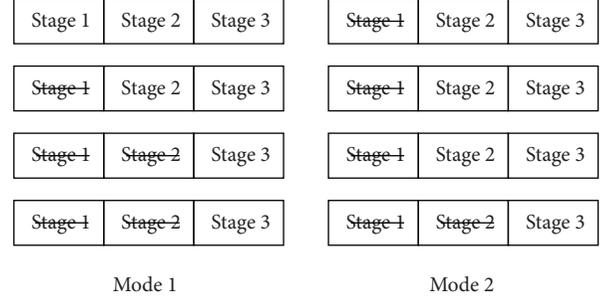


FIGURE 5: An example of different modes and stages for the same state.

the service channels; ω_j is the number of channels that have the same number of unfinished stages; $j = 1, 2, \dots, p$.

Accordingly, the total number of combinations for a specific state is

$$C_{\text{total}} = \sum_{i \in M} C_i, \quad (2)$$

where M is a set of all the mode numbers for this state.

Then the probability of this situation (i.e., the last stage is finished and the truck leaves the gate system) is calculated by

$$\sigma_{s,v} = \frac{1}{\omega C_{\text{total}}} \sum_{i \in M} E_{1,i} C_i, \quad (3)$$

where C_i/C_{total} means the probability of mode i ; $E_{1,i}$ is the number of service channels with one stage unfinished in mode i , so $E_{1,i}/\omega$ means the probability of service channels with only one stage left in mode i .

Suppose the service rate is $\lambda(t)$; the transition probability from state (s, v) to $(s - 1, v - 1)$ can thus be given by

$$P_{(s,v) \rightarrow (s-1,v-1)} = \sigma_{s,v} \times \lambda(t) \times k \times v. \quad (4)$$

(b) The truck is still in the gate system to finish the other stages.

Because the probability that the last stage is finished and the truck leaves the gate system is $\sigma_{(s,v)}$, the probability that one more stage is completed but there is no truck leaving the gate system can be simply represented by

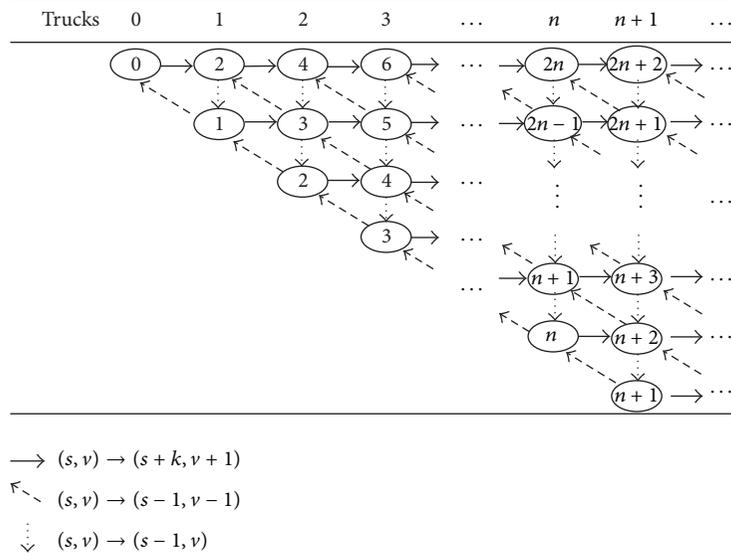
$$\delta_{s,v} = 1 - \sigma_{s,v}. \quad (5)$$

Consequently, the transition probability from state (s, v) to $(s - 1, v)$ is calculated as

$$P_{(s,v) \rightarrow (s-1,v)} = \delta_{s,v} \times \lambda(t) \times k \times v. \quad (6)$$

On the other hand, when the next time step (i.e., $t + 1$) exactly equals $t_N + t_a$, which means the $(N + 1)$ th truck has just arrived, the only possible state transition is from (s, v) to $(s + k, v + 1)$; k is the total number of stages for a container truck to finish in the gate system. Hence, the transition probability is $P_{(s,v) \rightarrow (s+k,v+1)} = 1$.

The entire state transition mentioned above for $M/E_{k=2}/n$ queuing model can be explained by the state transition diagram of Figure 6.


 FIGURE 6: State transition diagram of the $M/E_{k=2}/n$ queuing model.

For Figure 6, the numbers in the first row denote the total number of trucks in the system (i.e., v), and the numbers in the circles represent the unfinished stages in the system (i.e., s); a state (s, v) can thus be expressed. The solid arrows that point to the right indicate the transitions that one truck just arrives. The slanted arrows represent the transitions that one truck finishes the last stage and leaves the system. And the downward dotted arrows are used to denote the transitions that one more stage is finished but the truck is still in the system.

3.1.3. State to State Transition Probabilities Calculation. On the basis of what is discussed in Section 3.1.2, the state to state transition probabilities at each time step t_i then can be calculated and updated accordingly by the following equations. For the time steps, $t_i = t, \dots, t + t_a - 1$, that is, the time period when the N th truck had arrived and one time step before the $(N + 1)$ th truck arrives. Let $P_{s,v}(t_i)$ represent the probability of state (s, v) at current time t , and let $P'_{s,v}(t_i)$ represent the variation of $P_{s,v}(t_i)$ at each time step, so obviously the state probabilities of the next time step can be given by

$$P_{s,v}(t_i + 1) = P_{s,v}(t_i) + P'_{s,v}(t_i). \quad (7)$$

The calculation processes of the variation of probabilities are as below.

(a) When the number of trucks is less than the number of service channels in the system, $v < n$, we have the following.

For the state $(0, 0)$,

$$P'_{0,0}(t_i) = \lambda(t_i) \times k \times P_{1,1}(t_i). \quad (8)$$

For $v = 1, 2, \dots, n - 1$,

$$P'_{vk,v}(t_i) = -\lambda(t_i) \cdot k \cdot v \cdot P_{vk,v}(t_i) + \sigma_{vk+1,v+1} \cdot \lambda(t_i) \cdot k \cdot (v+1) \cdot P_{vk+1,v+1}(t_i). \quad (9)$$

For $v = 1, 2, \dots, n - 1, a = 1, 2, \dots, v \cdot (k - 1)$,

$$P'_{vk-a,v}(t_i) = -\lambda(t_i) \cdot k \cdot v \cdot P_{vk-a,v}(t_i) + \delta_{vk-a+1,v} \cdot \lambda(t_i) \cdot k \cdot v \cdot P_{vk-a+1,v}(t_i) + \sigma_{vk-a+1,v+1} \cdot \lambda(t_i) \cdot k \cdot (v+1) \cdot P_{vk-a+1,v+1}(t_i). \quad (10)$$

(b) When the number of trucks equals or is greater than the number of service channels in the system, $v \geq n$, we have the following.

For $v = n, n + 1, \dots, N$,

$$P'_{vk,v}(t_i) = -\lambda(t_i) \cdot k \cdot n \cdot P_{vk,v}(t_i) + \sigma_{(n-1)k+1,n} \cdot \lambda(t_i) \cdot k \cdot n \cdot P_{vk+1,v+1}(t_i). \quad (11)$$

For $v = n, n + 1, \dots, N, a = 1, 2, \dots, (n - 1) \cdot (k - 1)$,

$$P'_{vk-a,v}(t_i) = -\lambda(t_i) \cdot k \cdot n \cdot P_{vk-a,v}(t_i) + \delta_{nk-a+1,n} \cdot \lambda(t_i) \cdot k \cdot n \cdot P_{vk-a+1,v}(t_i) + \sigma_{(n-1)k-a+1,n} \cdot \lambda(t_i) \cdot k \cdot n \cdot P_{vk-a+1,v+1}(t_i). \quad (12)$$

For $v = n, n + 1, \dots, N, b = (n - 1) \cdot (k - 1) + 1, (n - 1) \cdot (k - 1) + 2, \dots, n(k - 1)$,

$$P'_{vk-b,v}(t_i) = -\lambda(t_i) \cdot k \cdot n \cdot P_{vk-b,v}(t_i) + \delta_{nk-b+1,n} \cdot \lambda(t_i) \cdot k \cdot n \cdot P_{vk-b+1,v}(t_i). \quad (13)$$

Then for the time point $t_i = t + t_a$, the $(N + 1)$ th truck arrives in the system. As mentioned in Section 3.1.2, the probability of state transition from (s, v) to $(s + k, v + 1)$ is equal to 1. Therefore, the state probabilities under this situation can be calculated as below.

For $v = 0, 1, \dots, N, N + 1, a = 0, 1, \dots, (n - 1) \cdot (k - 1), \dots, n \cdot (k - 1)$, if the states $(vk - a - k, v - 1)$ exist, then

$$P_{vk-a,v}(t_i) = P_{vk-a-k,v-1}(t_i). \quad (14)$$

The probabilities of the states which are not included in (8)–(14) are all equal to 0. The processes mentioned above are repeated until the time step reaches the end of analysis time period. Then all the transient state probabilities can be acquired using the decoding procedure in Algorithm 1.

3.1.4. The Calculation of Gate Congestion Indicators. After calculating the entire probabilities for each time step in the analysis time period, some significant performance indicators such as the probability of v container trucks in the system, the queue length, and the average waiting time for the v container trucks at any time point t all could be calculated accordingly. Moreover, these results can be obtained promptly without doing the unnecessary or repetitive calculation as the simulation models. The value of the performance indicators is provided by the following equations separately.

(a) The probability of v container trucks in the system at time t :

$$P_v(t) = \begin{cases} \sum_{s=v}^{kv} P_{s,v}(t), & 0 \leq v \leq n \\ \sum_{s=n+k(v-n)}^{kv} P_{s,v}(t), & v > n. \end{cases} \quad (15)$$

(b) The queue length at time t :

$$L_v(t) = l \times \sum_v (P_v(t) \times v), \quad 0 \leq v \leq N, \quad (16)$$

where l is the average length of the container trucks.

(c) The average waiting time for the v container trucks in the system at time t :

$$W_v(t) = \mu \sum_v (P_v(t) \times v). \quad (17)$$

3.2. Optimization Model. As mentioned before, this $M/E_k/n$ transient queuing model is easy to be integrated into an optimization framework in order to get some useful results to support the related decision making. So in this section, a simple and valid optimization model for the railway container terminal gate system is proposed.

The objective of the optimization is to minimize the total cost of the railway container terminal gate system in a certain analysis time period T . The total cost here consists of two parts: the operation expenses of the supply side and the waiting cost of the demand side. Therefore, in other words, the goal of this optimization problem is to find the optimal number of service channels which strike a balance between the supply and demand. Then the objective function and the constraints can be determined accordingly. As shown in function (18), the operating cost of the service channels is expressed by the multiplication of operation cost per hour of a service channel, the number of service channels, and the certain analysis time period T . The waiting cost of the container trucks in the queue is calculated by multiplying the hourly waiting value of each container truck by the number of container trucks waiting in the queue by the analysis time period T . There are two constraints represented in functions

(19) and (20) separately. All the above mentioned expressions are presented as below.

Let C_t denotes the total cost of the gate system during the time period t . Then

$$\min \quad C_t = (C_n \times n_t + C_w \times R_t) \times T \quad (18)$$

$$\text{s.t.} \quad \frac{\mu \times R_t}{n_t} \leq \varepsilon, \quad (19)$$

$$n_{\min} \leq n_t \leq n_{\max}, \quad (20)$$

where C_n is the operation cost per hour of a service channel; C_w is the waiting cost per hour of a container truck; n_t is the number of service channels opened in the time period t ; n_{\min} and n_{\max} are the minimum and maximum number of service channels available in the system; R_t is the number of container trucks waiting in the system in the time period t ; T is the time span of time period t ; μ is the average service time of the service channel; ε is the threshold of the average waiting time; $(\mu \times R_t)/n_t \leq \varepsilon$ ensures the average waiting time of each container truck does not exceed the threshold. $n_{\min} \leq n_t \leq n_{\max}$ indicates that the number of service channels opened in the time period t cannot go beyond the available range.

For the solution of this optimization problem, first integrate the transient queuing model into the optimization framework, and then conduct an optimization calculation in each time period t . That is to say, the optimal number of the service channels that satisfied the objective function and constraints is searched at each round of the transient calculation procedure in Section 3.1.3. This computation process will be carried out in MATLAB R2008a, and the detailed results and discussions are listed in Section 4.3.

4. Computational Experiments

4.1. Model Validation. For the purpose of verifying the transient queuing model and the corresponding solution method, a simulation method is adopted simultaneously to acquire the estimators of the railway container terminal gate system. eM.Plant is a professional object-oriented simulation software for management, industrial engineering, and system engineering. It is able to analyze the operation conditions of the railway container terminal gate system by running the visual simulation models. One simulation model of the railway container terminal gate system is shown in Figure 7.

In the simulation model, the Container trucks arrival is the producer of container trucks which is set at the exponential arrival time interval to represent the truck arrival situations. The Queuing place and waiting signs represent the container trucks in the queue. The Service station denotes the service channels in the railway container terminal gate system which is commanded to satisfy the Erlang service time. The Container trucks move is the saver of the produced container trucks and indicates the trucks move to the container yard. Event controller is used to determine the beginning and ending time of the simulation.

The experiments are conducted by using the $M/E_k/n$ transient queuing model and the eM.Plant simulation model, respectively. Take 30 trucks per hour as a possible volume

```

Procedure: the state to state transition probabilities transient calculation
Initialization:
    Input the initial value of  $k, n, t_{\max}, t_a, \lambda$ . Set  $t = 0, N = 0, s = 0, P_{0,0}(0) = 1, \text{isend} = 0$ 
    Initialize the  $P_{s,v}(t_i)$  and  $P'_{s,v}(t_i)$ 
while isend = 0 do
    for  $t_i = t : t + t_a - 1$  do
        if  $N < n$  then
            calculate  $P'_{0,0}(t_i)$  by (8) and  $P_{0,0}(t_i + 1)$  by (7)
            for  $v = 1 : N$  do
                calculate  $P'_{v,k,v}(t_i)$  by (9) and  $P_{v,k,v}(t_i + 1)$  by (7)
                for  $a = 1 : v(k - 1)$  do
                    calculate  $P'_{vk-a,v}(t_i)$  by (10) and  $P_{vk-a,v}(t_i + 1)$  by (7)
                end for
            end for
        else if  $N \geq n$  then
            calculate  $P'_{0,0}(t_i)$  and  $P_{0,0}(t_i + 1)$ 
            for  $v = 1 : n - 1$  do // calculate the part of probabilities that  $v < n$  //
                calculate  $P'_{v,k,v}(t_i), P'_{vk-a,v}(t_i)$  and  $P_{v,k,v}(t_i + 1), P_{vk-a,v}(t_i + 1)$ 
            end for
            for  $v = n : N$  do // calculate the part of probabilities that  $v \geq n$  //
                calculate  $P'_{v,k,v}(t_i)$  by (11) and  $P_{v,k,v}(t_i + 1)$  by (7)
                for  $a = 1 : (n - 1)(k - 1)$  do
                    calculate  $P'_{vk-a,v}(t_i)$  by (12) and  $P_{vk-a,v}(t_i + 1)$  by (7)
                end for
                for  $b = (n - 1)(k - 1) + 1 : n(k - 1)$  do
                    calculate  $P'_{vk-b,v}(t_i)$  by (13) and  $P_{vk-b,v}(t_i + 1)$  by (7)
                end for
            end for
        end if
    end for
     $t_i = t + t_a, v = N + 1$ 
    calculate the probabilities when a new container truck arrives  $P_{v,k-a,v}(t_i)$  by (14)
     $v = v - 1, t = t_i, N = N + 1$ 
    if  $t \geq t_{\max}$  then
        isend = 1
    end if
end while
    
```

ALGORITHM 1: The state transition probabilities transient calculation procedure.

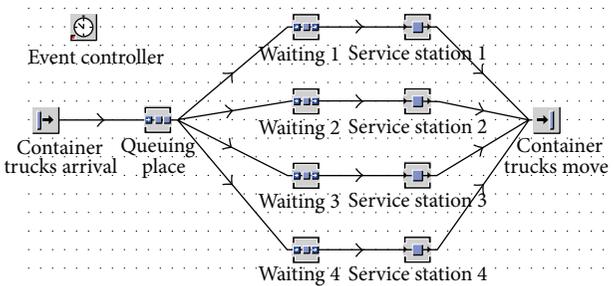


FIGURE 7: Simulation model in eM_Plant.

of the container trucks for the experiment. The service time follows the $E_{k=2}$ distribution and the mean value is set as 8 minutes. The analysis time period is instructed to be 60 minutes. Besides, to relieve the impact of randomness, the simulation model needs to be run several times to get

the mean value and standard variance for each case. Meanwhile, in consideration of the running time, the simulation model is run 20 times and all the results are shown in Table 1.

It can be seen, from the comparison experiment results, that the average number of trucks in the queue at the end of 60 minutes calculated from the $M/E_{k=2}/n$ transient queuing model is approximate to the simulation model. This means the transient queuing model is correct and the equally likely combinations (ELC) heuristic method is effective. Therefore, the model together with the solution method can be applied to analyze the queuing situation of the railway container terminal gate system.

4.2. Sensitivity Analysis. The sensitivity tests of the $M/E_k/n$ transient queuing model are carried out in three ways, which are to analyze the changes of the results, while the number of service channels, the mean arrival volume of the container trucks, and the average service time are different. From

TABLE 1: Results of the queuing method and the simulation method.

Arrival volume (trucks/h)	Number of service channels	Queuing model	Average number of trucks in the queue		Relative error
			Simulation model Mean value	Standard variance	
30	1	23.00	22.16	6.94	0.038
	2	16.28	14.73	5.75	0.105
	3	9.96	8.68	3.62	0.147
	4	4.73	4.41	2.28	0.073

these tests, the results of the model can be demonstrated and can provide some insights to the railway container gate system management. The calculated values of the mean arrival volume of the container trucks and the average service time are determined in accordance with the distributions in Section 3.1, that is, the mean arrival time interval equals 4.714 minutes (i.e., the mean arrival volume of the container trucks is 12 trucks per hour) and the average service rate is 0.25 (i.e., the average service time is 4 minutes). Besides, in order to confirm the results and present the trend more obviously, another set of severe values are also adopted as a comparison. The mean arrival volume of the container trucks and the average service time are set as 30 trucks per hour and 10 minutes, respectively.

4.2.1. The Influence of Increasing the Number of Service Channels. This sensitivity analysis test is on the purpose of observing the change of the average waiting time when increasing the number of service channels. As mentioned above, the calculated mean arrival volume of the container trucks is 12 trucks per hour and the average service time is 4 minutes which followed the mean values of the distributions in Section 3.1. Meanwhile, the comparison analysis which has the mean arrival volume of 30 trucks per hour and the average service time of 10 minutes is also conducted. According to the practical condition of the current railway container terminal gate system, the maximum number of service channels is 8. Therefore, let the number of service channels range from 1 to 8 to see the changes of the average waiting time for the container trucks at the end of 60 minutes. The results of the calculated value and the comparison analysis tests are shown in Figures 8 and 9, respectively.

It can be observed from the results in Figure 8 that the average waiting time has a dramatic decrease while the number of service channels increases from 1 to 2, and there is also an obvious decrease when another service channel is added; the average waiting time goes down to 2.3 minutes. However, there has hardly been any change when the number of service channels increases further. A similar situation can be seen in Figure 9. There is 148.92 minutes' reduction of the average waiting time when the number of service channels increases from 1 to 2, and the time decreases 48.76 minutes while one more service channel is increased. Then the declines slow down and the average waiting time has been reduced to 11.44 minutes when 5 service channels are opened.

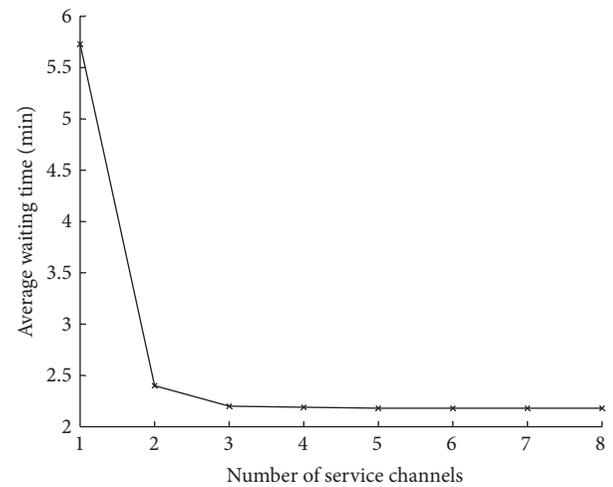


FIGURE 8: Results of the calculated value analysis test.

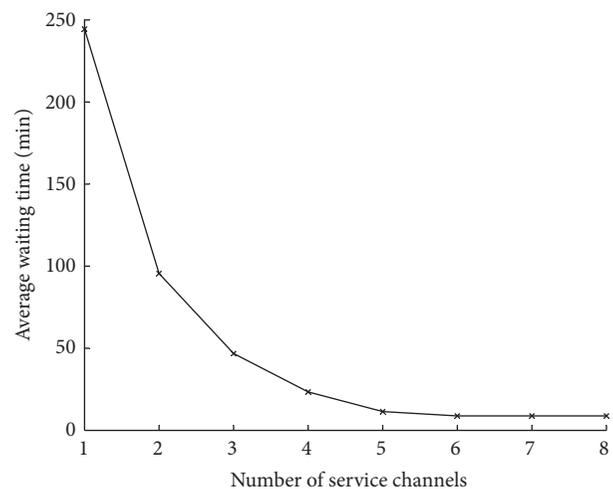


FIGURE 9: Results of the comparison analysis test.

After this, the changes are not very obvious. Consequently, the results can be used to decide the number of service channels under a certain average waiting time limitation. For example, the most reasonable number of service channels must be 5 if the average waiting time needs to be around 10 minutes in the second case.

TABLE 2: Results of the calculated value analysis test.

Mean arrival volume (trucks/h)	10	12	15	20	30	60
Average number of trucks	1.11	1.20	1.35	1.63	2.50	31.3
Average waiting time (minutes)	2.23	2.40	2.7	3.27	5.0	62.5

TABLE 3: Results of the comparison analysis test.

Mean arrival volume (trucks/h)	10	12	15	20	30	60
Average number of trucks	2.33	3.14	5.10	9.50	19.12	48.8
Average waiting time (minutes)	11.65	15.7	25.51	47.48	95.6	244

4.2.2. *The Impact of Raising the Mean Arrival Volume of the Container Trucks.* The objective of this sensitivity analysis test is to research the variation of the queue length and the average waiting time when raising the mean arrival volume of the container trucks. The average service time is also set by the calculated value (i.e., 4 minutes) and the comparison value (i.e., 10 minutes), but the mean arrival time interval in this test is reduced from 6 to 1 minute in decrement of 1 minute (i.e., the mean arrival volumes of the container trucks are 10, 12, 15, 20, 30, and 60 trucks per hour) by considering the mean arrival time interval of the exponential distribution (i.e., 4.714 minutes). Assume the number of service channels equals 2. Then obtain the number of the trucks in the queue and the average waiting time 60 minutes later. The results of the calculated value analysis tests and the comparison analysis tests can be found in Tables 2 and 3.

The results in Table 2 show a linear increasing trend along with the raise in the mean arrival volume of container trucks. The average waiting time is 2.23 minutes at the end of 60 minutes when the mean arrival volume is 10 trucks per hour. However, there exists a severe congestion (the average waiting time is as long as 62.5 minutes) when the mean arrival volume comes to 60 trucks per hour under this condition. The same trend but with a more significant increase appears in the comparison analysis test. It can be seen from Table 3 that the average waiting time quadruples (from 11.65 minutes to 47.48 minutes) while the mean arrival volume is increased from 10 trucks per hour to 20 trucks per hour. And more notably, the average waiting time reaches up to 244 minutes which cannot be accepted by the people in practice. Therefore, the relevant authorities can use the results to make some decisions to control the average waiting time of the container trucks in the railway container terminal gate system. For example, in the second case, if we want to maintain the average waiting time less than 20 minutes, then more service channels are required or the service rate must be increased when the mean arrival volume is equal to or greater than 15 trucks per hour.

4.2.3. *The Effect of Altering the Average Service Time.* The last sensitivity analysis test is aimed at investigating the influence of varying the average service time. To be more specific, this test is to seek the appropriate number of service channels when altering the average service time and also ensuring that the average waiting time remains at the same level. The mean arrival volume of the container trucks is assigned as well by

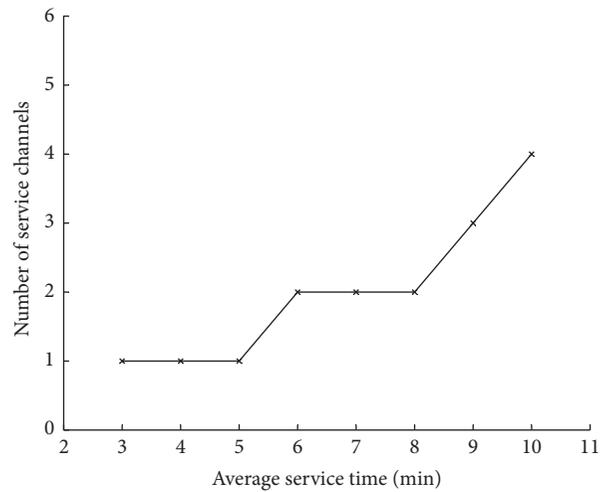


FIGURE 10: Results of the calculated value analysis test.

the calculated value of 12 trucks per hour and the comparison value of 30 trucks per hour, respectively. Considering the calculated value (i.e., 4 minutes) and the comparison value (i.e., 10 minutes) of the average service time, the range of the average service time in this test is set from 3 to 10 minutes in increment of 1 minute. Make the average waiting time of the container trucks around 20 minutes, and then determine the number of service channels needed at the end of 60 minutes. The computation results of the calculated value analysis test and the comparison analysis test are presented in Figures 10 and 11 as below.

A stepped growth trend can be seen from the results of the calculated value analysis test in Figure 10. There is only need for 1 service channel to be open when the average service time is less than 5 minutes under the first condition. Another service channel is needed while the average service time is more than 5 minutes but less than 8 minutes. However, the number of service channels shows a linear increasing when the average service time increases more; that is, with one-minute increase in average service time one more service channel is required. The growth trend of the comparison analysis test in Figure 11 is corresponding to the calculated value analysis test in Figure 10; there is also a rapid increase after the stepped growth with the average service time going up. On the whole, the necessary number of service channels is

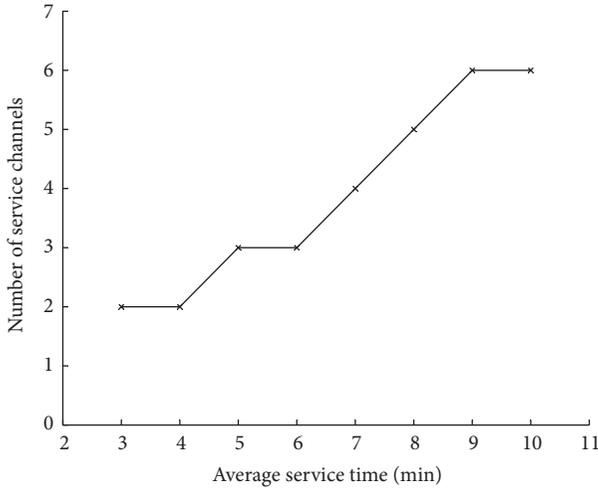


FIGURE 11: Results of the comparison analysis test.

threefold when the average service time varies from 3 minutes to 9 minutes in the second case. It can be concluded from all of these results that if the average waiting time of the container trucks should be controlled in a certain range without varying the mean arrival volume, more service channels are needed when the average service time increases. Meanwhile, the results are capable of helping the administrators to make some relevant policies for these cases. For example, the appropriate number of service channels is equal to 3 when the mean arrival volume of the container trucks is 30 trucks per hour and the acceptable average waiting time is around 20 minutes.

4.3. Optimization Results and Discussion. The model validation and the analysis tests have verified the established $M/E_k/n$ transient queuing model is reasonable and effective. Hence, a real data experiment is put forward accordingly for the solution of the optimization model. Moreover, through the discussion of the experiment results, some features in practical operations of the railway container terminal gate system are also revealed at the end.

The aim of the optimization experiment here is to find the optimal number of service channels for each time period and make some conclusions from the results. In this experiment, the hourly arrival volume of container trucks is the real data collected from 4:00 a.m. to 6:00 p.m. of Chengdu railway container terminal gate system for a certain day; the details can be seen in Figure 12. Thus, the analysis time period is 14 hours and the time period t is 1 hour (i.e., the system performance indicators such as the queue length, the average waiting time, and the optimal number of service channels are calculated every one hour until all the results of 14 hours have been output). The average service time of Chengdu railway container terminal gate system is the mean value of Erlang distribution in Section 3.1. For comparison, the operation cost per hour of a service channel, C_n , is set as two different values: one is \$40 per hour and another is \$80 per hour. Assume the waiting cost per hour of a container truck, C_w ,

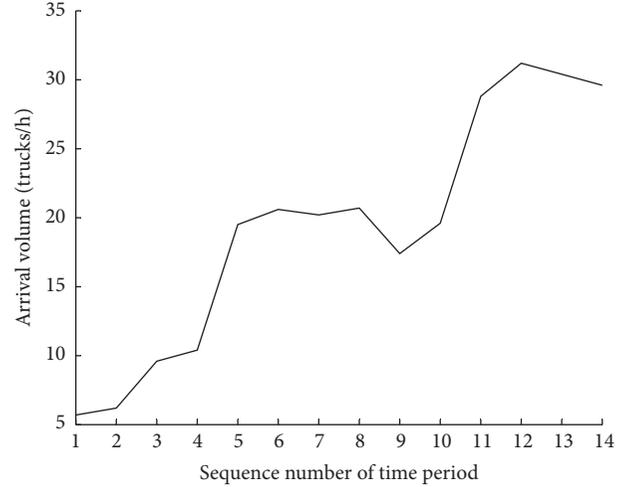


FIGURE 12: Arrival volume during the analysis time period.

is \$15 per hour. Meanwhile, based on the practical situation of the Chengdu railway container terminal gate system, the minimum and maximum number of service channels n_{\min} and n_{\max} are set as 1 and 8. In addition, the acceptable average waiting time for each container truck is about 10 minutes in practice. Therefore, the threshold ε is set as 10 minutes. All the optimization results for the duration of 14 hours are presented in Figure 13.

The sequence number of time period in Figure 13 is corresponding to the above mentioned time period t , the initial number indicates the time period from 4:00 a.m. to 5:00 a.m., and the rest are expressed in the same way. Graph (a) is the minimum total cost of Chengdu railway container terminal gate system in different time period, which is calculated from the established optimization model. The solid line and the marked dash line are generated from the two cases that the operation cost per hour of a service channel is \$80 and \$40, respectively. Graph (b) is the optimal numbers of service channels during all the time periods under these two cases. On the whole, the optimization results have a similar trend with the arrival volume of the container trucks for the analysis time period, which is the same as the intuition. When the arrival volume of the container trucks is low, the number of service channels and the total cost are small. With the growth of the arrival volume, the number of service channels and the total cost are increased. It also can be concluded from this optimization experiment that when the operation cost of the service channel is relatively high compared to the waiting cost of the container truck, the average waiting time of the container trucks should be sacrificed so as to obtain the minimum total cost of the system. As seen in Graph (b), the optimal number of service channels when the hourly cost of operating a service channel is \$80 is less than the number of service channels in the other case (i.e., the operation cost per hour of a service channel is \$40). By comparison between the two Graphs, although opening more service channels causes more expenses, the total cost is still less when the level of operation cost of a service channel for an hour is comparatively low.

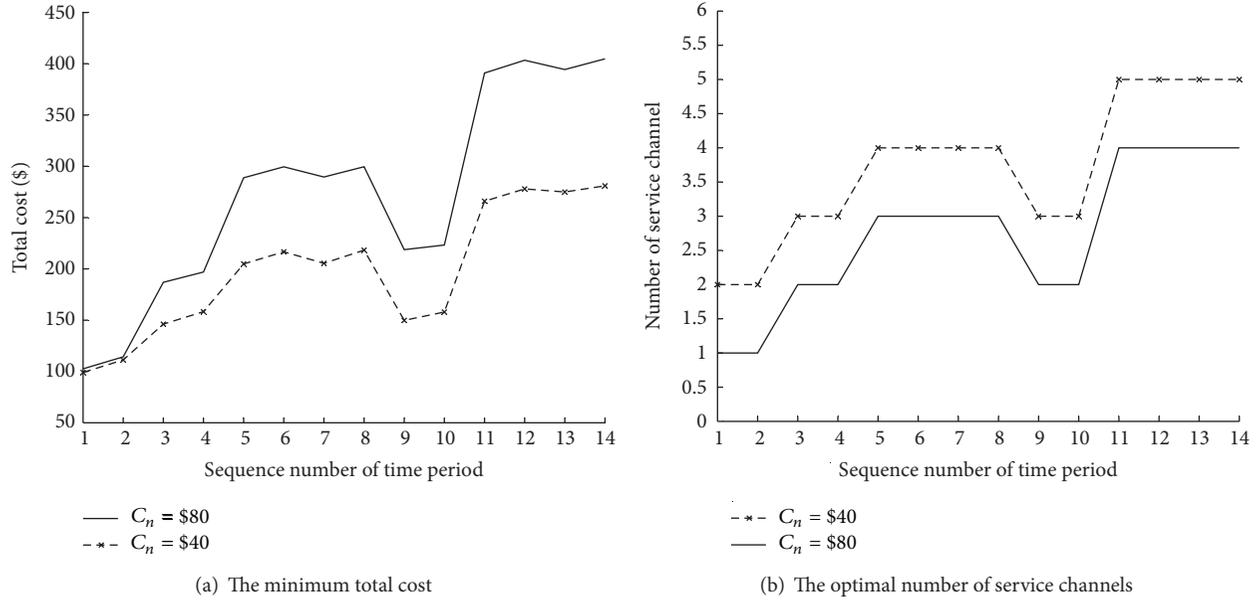


FIGURE 13: Optimization results for the analysis time period.

TABLE 4: Optimization results of the railway container terminal gate system in different time sections.

Time section	Number of service channels	Average number of trucks in the queue	Average total cost (\$)
4:00–6:00	1	1.56	103.41
6:00–8:00	2	1.80	186.93
8:00–12:00	3	3.29	289.38
12:00–14:00	2	2.20	192.99
14:00–18:00	4	4.89	393.36

Beyond the analysis conclusions above, the other useful information also can be obtained from the further study of this optimization experiment. Take the case that the operation cost of a service channel is \$80 per hour as an example, although the optimal number of service channels is altered in time with the variation of the service demand at different time period, a general pattern still can be summarized for a typical day of Chengdu railway container terminal gate system. Then accordingly, the analysis time period is divided into several sections on the basis of the same optimal number of service channels. For example, the optimal numbers of service channels are all equal to 1 for the time periods 1 and 2 (i.e., 4:00–5:00 and 5:00–6:00); therefore they are taken as one separate time section. And the rest are treated in the same way. All the number of service channels, the average number of trucks in the queue, and the average total cost of different time sections are provided in Table 4. As shown in the table, the number of service channels is the least in the early morning of the railway container terminal gate system and is gradually increasing until noon; then after a little decrease, the number reaches the peak in the late afternoon. These results make sense to the relevant optimization of the railway container terminal gate system and give the decision makers some reference to grasp the pattern of the variation

and help in forming the most economic and reasonable operation scheme.

5. Conclusion

This paper contributes to analyzing the railway container terminal gate system by the $M/E_k/n$ transient queuing model with the equally likely combinations (ELC) heuristic solution. The comparison experiment was then conducted with the eM.Plant simulation method. The experiment results indicate that the $M/E_k/n$ model is capable of providing a precise reflection to the queuing process of the railway container terminal gate system, and the relevant performance indicators of gate congestion such as the number of container trucks in the system, the queue length, and the average waiting time can be obtained by the ELC transient solution dynamically and effectively. Moreover, a variety of sensitivity tests were carried out for analysis. It turned out that the model is efficient under different conditions with low computational consumption, so it is able to be applied to assist decision making. In addition, because of the flexibility of the analytic method, this model can be easily integrated in an optimization framework. Therefore, the optimization of the railway container terminal gate system was implemented ultimately. All these results

can provide some support to the relevant authorities, so the model and method of this paper have the value of practical application.

Furthermore, there are some related works that can be done for the following research. For example, the applicability of the model and methods for the other railway container terminals needs further observations. In addition, take the railway container terminal system as a whole; combine the gate system optimization to the other parts of the entire system.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research is partially supported by the National Natural Science Foundation of China (no. 51175442 and no. 51205328), the Fundamental Research Funds for the Central Universities (2010ZT03 and 2682014BR019), and the Special Fund of the Chinese Scholarship Council of the Ministry of Education of China ([2012]3013).

References

- [1] "The global trade statistics about the industry in world shipping council," <http://www.worldshipping.org/about-the-industry/global-trade/trade-statistics#1>.
- [2] Railway Statistical Bulletin, Ministry of Railway, China, 2014, (Chinese) http://www.xinhuanet.com/zhengfu/zf_temp4.htm.
- [3] D. C. Mattfeld and H. Kopfer, "Terminal operations management in vehicle transshipment," *Transportation Research A*, vol. 37, no. 5, pp. 435–452, 2003.
- [4] K. H. Kim, K. M. Lee, and H. Hwang, "Sequencing delivery and receiving operations for yard cranes in port container terminals," *International Journal of Production Economics*, vol. 84, no. 3, pp. 283–292, 2003.
- [5] D. Lee, H. Q. Wang, and L. Miao, "Quay crane scheduling with handling priority in port container terminals," *Engineering Optimization*, vol. 40, no. 2, pp. 179–189, 2008.
- [6] D. Lee, H. Q. Wang, and L. Miao, "Quay crane scheduling with non-interference constraints in port container terminals," *Transportation Research Part E: Logistics and Transportation Review*, vol. 44, no. 1, pp. 124–135, 2008.
- [7] Y. Zhu and A. Lim, "Crane scheduling with non-crossing constraint," *Journal of the Operational Research Society*, vol. 57, no. 12, pp. 1464–1471, 2006.
- [8] E. K. Bish, "A multiple-crane-constrained scheduling problem in a container terminal," *European Journal of Operational Research*, vol. 144, no. 1, pp. 83–107, 2003.
- [9] K. H. Kim and Y. Park, "A crane scheduling method for port container terminals," *European Journal of Operational Research*, vol. 156, no. 3, pp. 752–768, 2004.
- [10] M. E. H. Petering, "Effect of block width and storage yard layout on marine container terminal performance," *Transportation Research E: Logistics and Transportation Review*, vol. 45, no. 4, pp. 591–610, 2009.
- [11] J. Wiese, L. Suhl, and N. Kliewer, "Mathematical models and solution methods for optimal container terminal yard layouts," *OR Spectrum*, vol. 32, no. 3, pp. 427–452, 2010.
- [12] A. Imai, E. Nishimura, and S. Papadimitriou, "The dynamic berth allocation problem for a container port," *Transportation Research Part B: Methodological*, vol. 35, no. 4, pp. 401–417, 2001.
- [13] M. Kia, E. Shayan, and F. Ghotb, "Investigation of port capacity under a new approach by computer simulation," *Computers and Industrial Engineering*, vol. 42, no. 2–4, pp. 533–540, 2002.
- [14] W. Huang, T. Kuo, and S. Wu, "A comparison of analytical methods and simulation for container terminal planning," *Journal of the Chinese Institute of Industrial Engineers*, vol. 24, no. 3, pp. 200–209, 2007.
- [15] B. Ha, E. Park, and C. Lee, "A simulation model with a low level of detail for container terminals and its applications," in *Proceedings of the Winter Simulation Conference (WSC '07)*, pp. 2003–2011, Piscataway, NJ, USA, December 2007.
- [16] B. K. Lee, S. O. Park, and J. H. Seo, "A simulation study for designing a rail terminal in a container port," in *Proceedings of the Winter Simulation Conference (WSC '06)*, pp. 1388–1397, Monterey, Calif, USA, December 2006.
- [17] S. P. Sgouridis, D. Makris, and D. C. Angelides, "Simulation analysis for midterm yard planning in container terminal," *Journal of Waterway, Port, Coastal and Ocean Engineering*, vol. 129, no. 4, pp. 178–187, 2003.
- [18] P. Legato and R. M. Mazza, "Berth planning and resources optimisation at a container terminal via discrete event simulation," *European Journal of Operational Research*, vol. 133, no. 3, pp. 537–547, 2001.
- [19] N. Boysen, F. Jaehn, and E. Pesch, "Scheduling freight trains in rail-rail transshipment yards," *Transportation Science*, vol. 45, no. 2, pp. 199–211, 2011.
- [20] N. Boysen, F. Jaehn, and E. Pesch, "New bounds and algorithms for the transshipment yard scheduling problem," *Journal of Scheduling*, vol. 15, no. 4, pp. 499–511, 2012.
- [21] G. Froyland, T. Koch, N. Megow, E. Duane, and H. Wren, "Optimizing the landside operation of a container terminal," *OR Spectrum. Quantitative Approaches in Management*, vol. 30, no. 1, pp. 53–75, 2008.
- [22] P. Guo, W. Cheng, Z. Zhang, M. Zhang, and J. Liang, "Gantry crane scheduling with interference constraints in railway container terminals," *International Journal of Computational Intelligence Systems*, vol. 6, no. 2, pp. 244–260, 2013.
- [23] U. Dorndorf and F. Schneider, "Scheduling automated triple cross-over stacking cranes in a container yard," *OR Spectrum*, vol. 32, no. 3, pp. 617–632, 2010.
- [24] F. Bruns and S. Knust, "Optimized load planning of trains in intermodal transportation," *OR Spectrum: Quantitative Approaches in Management*, vol. 34, no. 3, pp. 511–533, 2012.
- [25] P. Corry and E. Kozan, "An assignment model for dynamic load planning of intermodal trains," *Computers and Operations Research*, vol. 33, no. 1, pp. 1–17, 2006.
- [26] I. A. Hansen, "Automated shunting of rail container wagons in ports and terminal areas," *Transportation Planning and Technology*, vol. 27, no. 5, pp. 385–401, 2004.
- [27] R. Dekker, P. Voogd, and E. van Asperen, "Advanced methods for container stacking," *OR Spectrum*, vol. 28, no. 4, pp. 563–586, 2006.
- [28] N. Boysen, M. Fliedner, and M. Kellner, "Determining fixed crane areas in rail-rail transshipment yards," *Transportation Research E: Logistics and Transportation Review*, vol. 46, no. 6, pp. 1005–1016, 2010.

- [29] N. Boysen and M. Fliedner, "Determining crane areas in intermodal transshipment yards: The yard partition problem," *European Journal of Operational Research*, vol. 204, no. 2, pp. 336–342, 2010.
- [30] M. Gronalt, T. Benna, and M. Posset, "Strategic planning of hinterland container terminals: a simulation based procedure," in *Proceedings of Informatik Trifft Logistik. Band 1. Beiträge der 37. Jahrestagung der Gesellschaft für Informatik e.V. (GI) (INFORMATIK '07)*, pp. 425–428, Bremen, Germany, September 2007.
- [31] S. Hartmann, "A general framework for scheduling equipment and manpower at container terminals," *OR Spectrum*, vol. 26, no. 1, pp. 51–74, 2004.
- [32] A. E. Rizzoli, N. Fornara, and L. M. Gambardella, "A simulation tool for combined rail/road transport in intermodal terminals," *Mathematics and Computers in Simulation*, vol. 59, no. 1–3, pp. 57–71, 2002.
- [33] P. Guo, W. Cheng, and M. Zhang, "Simulation on the operation influence by external factors for railway container logistic center," in *Proceedings of the International Conference of Logistics Engineering and Management: Logistics for Sustained Economic Development—Infrastructure, Information, Integration (ICLEM '10)*, pp. 2915–2922, Chengdu, China, October 2010.
- [34] E. D. Edmond and R. P. Maggs, "How useful are queue models in port investment decision for container berths," *The Journal of the Operational Research Society*, vol. 29, no. 8, pp. 741–750, 1978.
- [35] P. Canonaco, P. Legato, R. M. Mazza et al., "A queuing network model for the management of berth crane operations," *Computers and Operations Research*, vol. 35, no. 8, pp. 2432–2446, 2008.
- [36] Z. G. Zhang, H. P. Luh, and C. H. Wang, "Modeling security-check queues," *Management Science*, vol. 57, no. 11, pp. 1979–1995, 2011.
- [37] S. Kim, "The toll plaza optimization problem: design, operations, and strategies," *Transportation Research E: Logistics and Transportation Review*, vol. 45, no. 1, pp. 125–137, 2009.
- [38] M. C. Ausin, R. E. Lillo, and M. P. Wiper, "Bayesian control of the number of servers in a $GI/M/c$ queueing system," *Journal of Statistical Planning and Inference*, vol. 137, no. 10, pp. 3043–3057, 2007.
- [39] M. C. Ausin, M. P. Wiper, and R. E. Lillo, "Bayesian prediction of the transient behaviour and busy period in short-and long-tailed $GI/G/1$ queueing systems," *Computational Statistics & Data Analysis*, vol. 52, no. 3, pp. 1615–1635, 2008.
- [40] T. Czachórski, T. Nycz, and F. Pekergin, "Transient states of priority queues—a diffusion approximation study," in *Proceedings of the 5th Advanced International Conference on Telecommunications (AICT '09)*, pp. 44–51, Venice, Italy, May 2009.
- [41] M. Parlar and M. Sharafali, "Dynamic allocation of airline check-in counters: a queueing optimization approach," *Management Science*, vol. 54, no. 8, pp. 1410–1424, 2008.
- [42] M. Escobar, A. R. Odoni, and E. Roth, "Approximate solution for multi-server queueing systems with Erlangian service times," *Computers and Operations Research*, vol. 29, no. 10, pp. 1353–1374, 2002.

Research Article

Service Capacity Reserve under Uncertainty by Hospital's ER Analogies: A Practical Model for Car Services

Miguel Ángel Pérez Salaverría¹ and José Manuel Mira McWilliams²

¹ Jaguar Land Rover SLU, Torre Picasso, Plaza Pablo Ruiz Picasso 1, Planta 42, Complejo Azca, 28020 Madrid, Spain

² Universidad Politécnica de Madrid, Avenida Ramiro de Maeztu 7, 28040 Madrid, Spain

Correspondence should be addressed to José Manuel Mira McWilliams; josemanuel.mira@upm.es

Received 23 January 2014; Revised 19 May 2014; Accepted 21 May 2014; Published 10 July 2014

Academic Editor: Michael Lütjen

Copyright © 2014 M. Á. Pérez Salaverría and J. M. Mira McWilliams. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We define a capacity reserve model to dimension passenger car service installations according to the demographic distribution of the area to be serviced by using hospital's emergency room analogies. Usually, service facilities are designed applying empirical methods, but customers arrive under uncertain conditions not included in the original estimations, and there is a gap between customer's real demand and the service's capacity. Our research establishes a valid methodology and covers the absence of recent researches and the lack of statistical techniques implementation, integrating demand uncertainty in a unique model built in stages by implementing ARIMA forecasting, queuing theory, and Monte Carlo simulation to optimize the service capacity and occupancy, minimizing the implicit cost of the capacity that must be reserved to service unexpected customers. Our model has proved to be a useful tool for optimal decision making under uncertainty integrating the prediction of the cost implicit in the reserve capacity to serve unexpected demand and defining a set of new process indicators, such as capacity, occupancy, and cost of capacity reserve never studied before. The new indicators are intended to optimize the service operation. This set of new indicators could be implemented in the information systems used in the passenger car services.

1. Introduction and Literature Review

Today, the passenger car industry is one of the world's most important industries encompassing investment groups and manufacturers. All passenger car brands operate in a global competitive marketplace with commercial brands that must offer a wide range of products, including repair and maintenance services.

Historically, passenger car services were intended to fix product issues and carry out the scheduled maintenance routines. However, at present time, after-sale services have evolved, becoming an indispensable part of the business to ensure current customer retention and new customer conquest. The after-sales service market has ballooned to four to five times the size of the original equipment business [1].

Under the above scope, any after-sales service opportunity is taken into account, not only to fix or maintain the car but also to respond to customer demands and increase company's revenue. Escalating customer expectations for

rapid, flawless service support has increased the opportunity for firms to profit from appropriately priced differentiated service products targeted to meet the needs of particular market segments [1].

Thus, customer demands are not exclusively related to product issues; therefore, services are conveniently designed to suit customer needs and exceed initial expectations to make sure clients remain loyal to the brand and keep purchasing new products. On the other hand, services are usually planned in advance and customers are required to arrange an appointment prior to visiting the workshop, but whenever there is a breakdown, servicing unexpected visits introduces a random component and its resolution will always depend on the workshop availability.

As a consequence of the additional challenges in the after sales we have described in the previous paragraphs that passenger car companies embrace commercial relationships with a focus on maximizing revenue. That revenue will be obtained only if the direct result of the customer lifetime value

is positive. With this in mind, passenger car brands adopt the motto “a happy customer is a returning customer,” but there is an important difference between the meanings of satisfaction and retention [2]. Nevertheless, even when everything has been carefully planned, an unexpected customer might appear, and independently of how the service manages the customer, the emergency will affect the service revenue.

A very recent research [3] has demonstrated that the integrated nature of the after-sales quality in the passenger car service is strongly associated with the retention rate of the customers. In that study, the authors confirmed that when customers perceive the poor service quality, immediately they switch to another service centre. In this study it is also proved that, in this highly competitive environment, it is the service quality only by which brands can retain their customers. This confirms there is a real gap [4] between customer’s real demand and actual service capacity.

The above gap is also studied by other authors. Literally citing an article published by Cohen et al. [1] in 2006, “Customers don’t expect products to be perfect, but they do expect manufacturers to fix things quickly when they break down. Not surprisingly, customers are usually unhappy with the quality of after-sales support.” According to the same publication, “That’s mainly because after-sales support is notoriously difficult to manage, and only companies that provide services efficiently can make money from them.”

Essentially, passenger car breakdowns are unexpected and do not adhere to planned schedules, like maintenance. Only those passenger car brands that manage after-sales service skilfully make money from it [1, 5].

1.1. Service Quality and Service Capacity. Service quality as a generic concept is well defined by various researchers in several ways [6–9]; technical quality, functional quality, and reputation are identified as the most frequent components of service quality. Usually, passenger car brands measure service quality by comparing initial customer expectations before the service with the perception after it has been delivered.

While service quality is a popular term in the passenger car industry, service capacity is not. Only limited research has been published on service capacity in the passenger car industry with respect to the extra capacity required to serve unexpected demand. Recent service capacity studies focus mainly on the specific situation in emerging markets, such as China and India, but no new researches have been published in regard to mature markets, such as Europe or USA.

Although there are no new specific publications in the passenger car field from an operations research approach, there are other studies from a marketing perspective [3, 10, 11]. This means that the issue of service capacity in the passenger car service industry has hardly been dealt with [7].

1.1.1. A Very Different Approach: Hospital’s Service Capacity. As opposed to the situation in the passenger car industry, hospitals often reserve capacity for patients arriving to their hospital’s emergency room (ER) in response to demand uncertainty. Reserving part of the hospital’s capacity ensures enough flexibility for urgent admissions. Particularly, this is

the usual scenario for premium passenger car manufacturers and traders, but it is not limited to them. Passenger car brands set up processes to ensure that customers are taken care of with the maximum convenience, which allows us to compare workshops with hospital’s emergency room (ER). Under this scope, workshop bays are intended as ER beds, technicians are like the medical staff and the service reception and foreman must act like the hospital’s emergency room (ER) capacity planner. Thus, there is a tradeoff to be optimized between service efficiency and the capability to admit unexpected customers in the process by reserving some of the service capacity.

The seminal references in the healthcare sector for the present document are based on the work of Kamenetzky et al. [12], who studied, in 1982, how to estimate necessities and demands for prehospital care. Subsequently in 1993, Badri and Hollingsworth [13] published a simulation model for scheduling in the hospital’s emergency room (ER). Later, in 1996, Gerchak et al. [14] studied a reservation planning under uncertain demand for emergency surgery. Additionally, in 1998, Bazargan et al. [15] set an initial approach to hospital’s emergency room (ER) and hospital services utilization using a theoretical model from historical data (kind of patient, demography, etc.).

Other authors approached hospital’s capacity problem from an operational research point of view. In 2004, Brailsford et al. [16] developed a model for emergency and on-demand health care for large complex systems. Also in 2004, Beraldi et al. [17] created a stochastic programming routine with probabilistic constraints aimed to solve a location and dimensioning problem. Then in 2006, Green et al. [18] developed a model to manage patient service in a diagnostic medical facility.

Unlike the passenger car industry, hospitals usually manage their hospital’s emergency room (ER) capacity by making the distinction between elective and emergency (nonelective) admissions and highlight the importance of an accurate forecast on both [19], estimating how fixed capacity on the nonelective admission expectations of unexpected demand turns into effective demand.

In other words, if a hospital capacity requires a number of available beds to assign to incoming patients, health managers have capacity for those patients who might enter the hospital as elective demand; that is, after a specialist diagnoses and retains some of the whole hospital capacity for those patients entering from the emergency room, because as depending on the severity of their disease they might not be rejected.

In contrast to what is described as a standard process in the passenger car arena, hospitals usually reserve some capacity in response to demand uncertainty to support the specification for optimal capability, incorporating cost derived from capacity reservation. By reserving some “empty” beds in the hospital, capacity planners ensure the required capacity to serve “emergency” admissions of patients.

Unfortunately, the literature review, in regard to health management, confirms that seldom estimations of hospital cost structures have taken production into account by incorporating the impact of nonelective demand on hospital

cost structures. The same papers establish that hospitals are in control of the output decisions, in response to such unexpected demand [20]. In these studies the emphasis has been on estimating (and minimizing) the cost of maintaining reserve capacity rather than using nonelective demand as part of a decision support system. Our research will lean on the work referred to in this paragraph to apply the proposed methodology to the unexpected demand in the passenger car service industry.

1.2. Usual Tools and Models in the Passenger Car Industry and the Service Sector. An extended tool along the passenger car industry is an IT system called “dealer management system,” known as dealer management system (DMS). Generically, dealer management system (DMS) includes inventory tool kits to manage parts availability information while arranging appointments. Often, the same solution is applied to book available dates in the service diary, but inventory techniques are intended for basic control and future decisions are not supported properly.

Inventory models, particularly “newsvendor” solutions, built in the current dealer management systems (DMS) have been studied intensively. Unfortunately, inventory models have an important limitation as they are intended to obtain a point forecast [21], a mathematical expression to help in determining the economic order quantity [22], or the ordering frequency, to keep goods or services flowing to the customer without interruption [23] or delay [24]. Therefore, inventory models only deal with part of the problem of capacity reserve.

The second limitation of inventory models is, even when they can incorporate uncertain demand, the main application that is directed to quantify decisions or to estimate profit (or loss) of unsold units, and so forth, but, whatever the model is, there is a common goal: to maximize the expected profit [25]. An important constraint of inventory models is, according to the reviewed literature [26], the resource requirements that are not fully known when a decision about the service resource distribution is taken due to the nature of customer behaviour. Thus, a strategy that balances service quality and cost yields must be found [27].

A third limitation of inventory models is the dating process that frequently does not work properly given that customers are not always able to arrive within the appointed time window, delaying the reception operation and creating a bullwhip effect in upstream dates [28]. Therefore, inventory models cannot respond to the car manufacturer problem since they do not cover potential emergencies or capacity reserve. Additionally, in the passenger car industry, profit is not always related to stock trade but to a customer long-term relationship.

Yield management [29] is another methodology intended to manage the capacity of service systems. An important limitation is yield management that focuses on service pricing instead of service constraints and system capacity. This approach seems to be interesting for other service sectors, such as hotels and airlines, where the service duration is well known (i.e., one night, 3 hours flight) and service prices vary

with the demand. Therefore, yield management models are not specifically developed to respond to the questions we aim to answer with our suggested service model for the passenger car industry.

In addition to the above methods, a common optimization methodology used in the service operation consists in running a simple forecast to estimate future demand values, without estimating uncertainty by means of a probability model [5]. Forecasts are then used to feed a mathematical expression that can be derived to minimize or maximize a variable. Nonetheless, there is no uncertainty quantification incorporated in the above optimization method as input demand is taken as an aggregated value, without differentiating between elective and nonelective demands.

In other cases [4], previous service’s research mainly focuses on understanding and measuring customer expectations and perceptions about the quality of service being provided. This would result in ascertaining the gap between customers’ expectation and perception. The obvious next stage is to identify the reasons for the gap between customer’s expectation and service capacity and finally provide suggestions for bridging this gap and a follow-up of the effectiveness of the actions taken.

1.3. Research Purpose. In the context of demand uncertainty, resolution of optimal capacity is very strongly dependent on an appropriate specification of the service outputs. One limitation of previous studies is that they have used aggregate measures of service to define outputs [30]; a second limitation is the reliance on annual or quarterly fluctuations in demand to system responses to nonelective demand [31], but failing to take account nonelective demand leads to a misspecification of system cost output relation [32].

With the input desegregation in mind [33], just after the beginning of a time period, when the aggregated demand for this period is known, a decision can be made, but this works against optimal capacity. Thus, when capacity reserve is expensive or the rejection rate is high, any further increase of its value will cause a decrease in optimal capacity [34].

An important property of the time series is that constraints on elective and nonelective demand are separated from other constraints [35]. On the other hand, in hospital emergency room (ER) applications it is being assumed that all hospitals have similar patient stream structure and that patients arrive at the hospital according to a Poisson flow [36], but without taking into account how the stochastic nature of demand is related to the type of case being serviced, while in this paper we will incorporate this relationship. Thus, our research will implement queuing theory to study arrival patterns at the service reception, waiting lines and servers, waiting times, and tasks completed [37].

To summarize, this paper is about the stochastic simulation of the process of service capacity reserve in the passenger car industry. A stochastic model has been implemented in a Monte Carlo simulation code written in Matlab and has proved to be a very useful tool for optimal decision making under uncertainty, involving an optimization process to define and maximize new key process indicators (KPIs).

The major contributions of this paper are the definition of new key process indicators (KPIs) and the development of valid integrated capacity model to respond to the needs of the passenger car service industry.

2. Our Integrated Approach: Capacity Reserve Model

There is a very scarce literature on applying simulation techniques to capacity reserve in the car industry, and inventory models do not include customer expectations but are intended to define some constraints related to supply chain specifics. Our simulation process has been developed to fill this gap, finding the conditions for maximum average service occupancy and average minimum capacity reserve cost; the probability distributions are obtained from the Monte Carlo simulations.

According to specific studies [38], in practice, service operation algorithms are ultimately carried out by computer simulations. Therefore, the Markov chain usually simulated is only an approximation to the true chain. Such limitations affect the simulation process, reducing the final results and raising questions about the validity of the previous algorithms used to build the model.

In this paper, we propose a new methodology which substantiates the integration of existing strategies used in medical installations to develop a valid model to be used in the passenger car industry. The model will be used to predict unexpected service demand and set a decision support system to estimate (in accordance with the tradeoff of above) the optimal operational costs (service efficiency mentioned above) and optimum service capacity reserve (servicing unexpected customers) by coupling discrete events with simulation techniques.

Additional references [39–41] will be cited below when describing the methodology.

2.1. Basics of the Methodology. We propose a major innovation which implements changes in demand estimation for the initial inputs, providing definition of new outputs and the development of a stochastic simulator for the whole process.

Our methodology incorporates the risk in quantitative analysis and decision making; thus, we are able to provide service managers with a range of possible outcomes and the probability for each of them. Thus, we can select different simulated variables and compare with the logical solution of going for the most conservative decision; this is, keeping the service workshop layout as is, but considering the impact of increasing service staff.

The methodology follows the 5-stage process flow as displayed in Figure 1.

- (i) *1st Stage: Service Demand Estimation.* We split the total demand (TV) in two major types, elective and nonelective, each with its own probabilistic distribution; therefore, we can integrate the stochastic models and sources of uncertainty of demand (elective and nonelective) and propagate this uncertainty to the output, thus adding value to predictions and allowing

for statistical interpretation. We do this by applying ARIMA models to step ① in Figure 1, for both types, and specific models are developed for each. Input data is gathered from the dealer management system (DMS).

- (ii) *2nd Stage: Service Times Definition.* Dealer's historic service data are gathered from the dealer management system (DMS) to estimate the probability distribution of each service time variable (reception, parts, and workshop). A stochastic queuing system is thus defined and fed with the adequate probabilistic models at step ② in Figure 1.
- (iii) *3rd Stage: Service KPI Definition.* We define a new set of output key process indicators (KPIs) at step ③ in Figure 1, which are functions of the random variables defined in stages 1 and 2.
- (iv) *4th Stage: Monte Carlo Simulation.* We run simulations to generate samples of the random variables of stages 1 and 2 and then propagate this uncertainty to obtain samples of the joint distribution of the different key process indicators (KPIs). If we do this for a number of scenarios (changing the number of technicians and of work bays), we will obtain different samples of the KPI joint distribution, one sample for each scenario. From each Monte Carlo sample we produce a report which displays a probabilistic analysis.
- (v) *5th Stage: Results Analysis and Optimization:* Here we analyse the simulation results to identify the service operation conditions for maximum average service occupancy and average minimum capacity reserve cost; the different scenarios are given in terms of the number of work bays and technicians. During the optimization stage, our methodology incorporates other variables, that will be defined later, which are used to identify the optimal scenario to assess the current system's effectiveness and improve ability to anticipate the impact of various changes in the service settings, similarly to previous researches [13]. This will be discussed in Section 3.

2.2. Stage 1: Service Demand Variable Definition and Modelling. A common way to manage service times is to request the customer to arrange a valid date for the next visit in advance. We define elective demand (ED) as the total number of prearranged visits to the service. This variable increases service reception managing capacity and saves time and money by arranging parts and skilled staff in advance.

As discussed in the introduction, the dating process does not work properly if customers do not arrive in time or change their minds, delaying the reception operation and creating a bullwhip effect in upstream dates [28]. Some customers dislike the dating process due to the inflexibility and lack of same or next day availability. Elective demand (ED) is thus a stochastic variable subject to high variability, partially dependable on customer requests which are only known with certainty after the arrival. We define nonelective

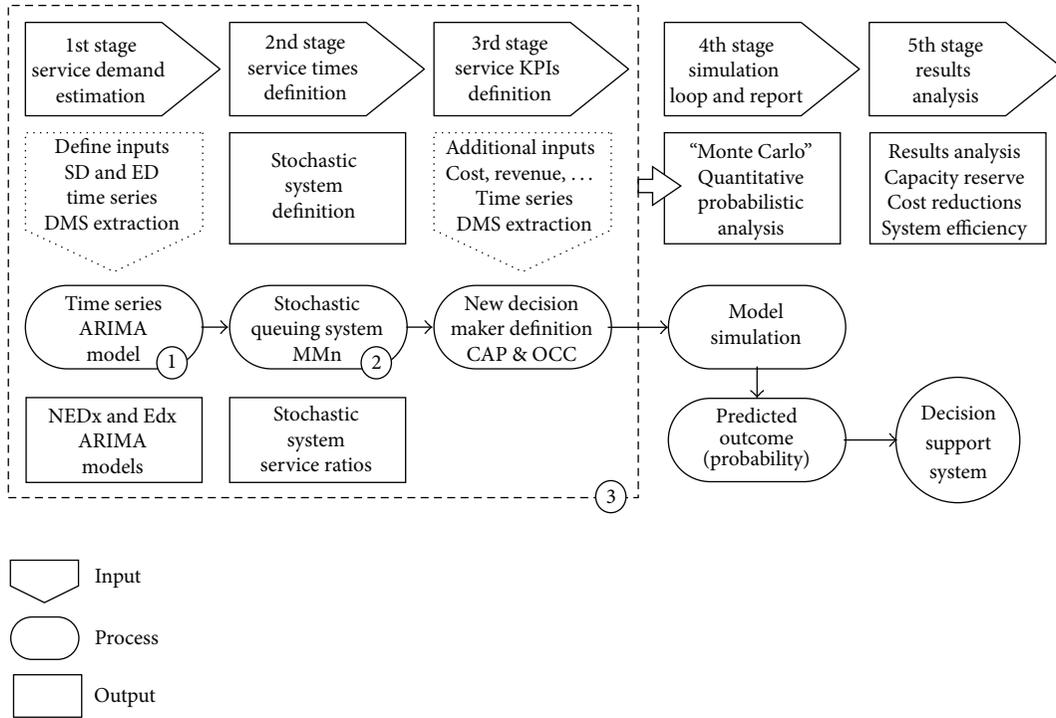


FIGURE 1: Methodology.

demand (NED) as the total number of unexpected services accepted in the system without a previous appointment.

As stated before, [26], resource requirements are not fully known at the time when a decision about the service resource distribution is taken. Therefore, in order to obtain valid forecasts, elective demand (ED) and nonelective demand (NED) should be the result of different random processes and can be expressed in a mathematical form as a probabilistic concept used to describe a sequence of random variables (stochastic process) that evolves in terms of another variable (index), usually time. Each of the random variables of the full service process has its own probability distribution, and variables may be correlated or not.

Hence, we suggest using ARIMA models for elective demand (ED) and nonelective demand (NED) forecasting. It should be noted there is a clear difference between both variables, and that is the reason to use separated variables (ED & NED) and feed the queuing system with each variable unique probabilistic distribution. This is required to demonstrate our methodology obtains valid results, integrating uncertainty and adding statistical value to the simple forecasting processes.

As discussed previously, nonelective demand (NED) has never been studied or estimated in the passenger car industry yet; therefore, there are no time series data to work with, neither the cost implicit in reserving capacity to service unexpected demand has been part of any research. This cost is a new concept which we will define below, in stage 3, as

the capacity reserve cost to serve all nonelective demands (CAPRNED).

In the current economic situation, customers are demanding prompt and flexible service; thus, nonelective demand (NED) is becoming a huge issue for all passenger car brands. Elective demand (ED) and nonelective demand (NED) balance has a cost related to the capacity reserve to suit nonelective demand (NED) service needs.

If capacity reserve for nonelective demand (NED) is high, services could lose elective demand's (ED) service income and profit would be lower than expected. Also, if service is full with elective demand (ED) only, there is no capacity to suit nonelective demand (NED) arrivals and customers will turn away.

Else, operating at full capacity sets the optimal reserve capacity levels compatible with economically efficient utilization but imposes a cost, however, in the form of production inflexibility, leading to patients being queued or turned away. At the same time, failing to take account nonelective demand leads to a misspecification of system cost output relation [32].

As explained in the literature review, emphasis has been on estimating (and minimizing) the cost of maintaining reserve capacity rather than using nonelective demand as part of a decision support system. Our research will lean on the work referred to in this paragraph and extend it to apply the proposed methodology to unexpected demand at the car industry, according to the reviewed literature [12–18, 20].

Even when dealer management systems (DMS) are not focused on time series analysis, we can gather sufficient data

to create time series to work with, which will include the following information:

- (i) period,
- (ii) total demand (TV),
- (iii) elective demand (ED),
- (iv) monthly work time,
- (v) monthly invested time,
- (vi) monthly invoiced time,
- (vii) total part sales,
- (viii) cost of part sales,
- (ix) vehicle sales,
- (x) cumulative vehicle car park.

Now, we can estimate the following parameters with the previous data:

- (i) average invested hours per vehicle (invested time/car park),
- (ii) average invoiced hours per vehicle (invoiced time/car park),
- (iii) average sold parts per vehicle (total part sales time/car park),
- (iv) average technician employment (invested time/work time),
- (v) average technician productivity (invoiced time/work time),
- (vi) average technician efficiency (invoiced time/invested time).

A simpler modelling could be developed by using an aggregated ARIMA model for total demand (TV) time series data only, but this would not allow us to relate the model to unexpected visits. Thus, in order to obtain a valid time series for nonelective demand we must gather the relevant data from the dealer management systems (DMS).

Dealer management systems (DMS) usually register all service visits but also manage the appointment process effectiveness by registering elective demand separately. We define total demand (TV) as the total number of visits in a given time period:

$$TV = \sum (ED + NED). \quad (1)$$

Total demand (TV) is thus the sum of elective and nonelective demands. Since this paper is intended to set a valid methodology to estimate the unexpected demand reserve costs, we need to differentiate between elective demand (ED) and nonelective demand (NED) rather than using a single nondisaggregated variable, as total demand (TV) is.

Once a time series is available for total demand (TV) and elective demand (ED), we will be able to obtain the nonelective demand (NED) time series to predict future values with specific ARIMA models for each. This analysis will produce a forecast with uncertainty bands and confidence intervals that can be used to confirm if both time series forecasts are confident simultaneously.

The sample data time series from a real service is shown in Table 1; also data graph is displayed in Figure 2. Data will be used to estimate valid service demand ARIMA models, for elective demand (ED) and nonelective demand (NED), and to feed the stochastic queuing system.

An important property of the observed samples of time series is that constraints on elective demand (ED) and nonelective demand (NED) are independent of other constraints [35]. One axiom of our research is that elective demand (ED) and nonelective demand (NED) are independent. To check this, we must confirm from the data that elective demand (ED) and nonelective demand (NED) are not correlated.

Thus, individual ARIMA models and forecasts for elective demand (ED) and nonelective demand (NED) will have the form of a parametric expression that relates the future value to previous ones, plus the noise. Given an ARIMA model of consumer demand and the lead times at each stage, it has been proven that the orders and inventories at each stage are also ARIMA [28], and closed-form expressions for these models are given.

2.3. Stage 2: Service Times. We now deal with the second stage: we estimate the probability distributions required to feed a stochastic queuing system and emulate the whole service operation.

We build the queuing model to analyse the behaviour of the system along time and the reaction to different stimuli and waiting times for a queue in which customers require simultaneous service from a variable number of servers [41]. In previous studies, the service systems considered are centralized and controllable and do not generate labour at a constant rate [40].

Tasks are admitted upon generation and processed by the system and completed labour is ejected from the system that has the capability of dealing with as many jobs per unit time on average as possible. Under this generic framework the system capability is measured as the maximum rate of work arrivals for which the system has a steady state [39].

We differ from the previous statement since we are considering service operation as a complete unit; that is, we include additional departments and not just service's workshop. This is, we are considering Parts and Reception times, including reception delays due to customer unavailability to arrive, "elective" customers changing to "nonelective," and other delays related to parts ordering and delivery.

We propose to measure the full service system capability (as shown in Figure 3) by running a queuing model built in the simulation loop at a constant arrival rate of work (arrival rate λ : shown in Figure 3 and defined in Table 2). Therefore, our methodology will cover a unique service cycle (as seen by customers) with the following phases, shown in Figures 2 and 3:

- (1) arrival,
- (2) reception,
- (3) parts,
- (4) service workshop.

TABLE I: Sample data.

Period	TV	ED	NED
1	107	91	16
2	94	91	3
3	131	119	12
4	116	113	3
5	172	172	0
6	184	182	2
7	154	148	6
8	75	75	0
9	107	97	10
10	125	107	18
11	129	108	21
12	103	91	12
13	149	135	14
14	130	130	0
15	165	147	18
16	249	248	1
17	137	137	0
18	179	179	0
19	145	145	0
20	117	117	0
21	195	179	16
22	149	131	18
23	205	169	36
24	135	122	13
25	150	121	29
26	149	115	34
27	144	140	4
28	132	127	5
29	252	243	9
30	165	159	6
31	169	144	25
32	105	94	11
33	172	142	30
34	131	109	22
35	180	144	36
36	128	115	13
37	180	141	39
38	148	127	21
39	155	127	28
40	138	124	14
41	258	222	36
42	168	152	16
43	151	135	16
44	114	99	15
45	128	97	31
46	151	140	11
47	139	111	28
48	97	85	12
49	160	136	24
50	173	149	24

TABLE I: Continued.

Period	TV	ED	NED
51	146	119	27
52	331	307	24
53	140	129	11
54	137	124	13
55	174	163	11
56	88	86	2
57	131	117	14
58	183	162	21
59	135	120	15
60	125	119	6
61	98	93	5
62	145	124	21
63	138	116	22
64	200	185	15
65	226	212	14
66	166	156	10
67	151	139	12
68	124	99	25
69	156	137	19
70	183	165	18
71	141	118	23
72	118	109	9
73	119	111	8
74	122	109	13
75	113	95	18
76	221	197	24
77	193	171	22
78	140	133	7
79	160	144	16
80	117	92	25
81	141	130	9
82	148	139	9
83	135	123	12
84	111	99	12
85	108	105	3
86	110	109	1
87	112	106	6
88	132	131	1
89	128	126	2
90	129	127	2
91	117	109	8
92	95	91	4
93	117	103	14
94	119	116	3
95	134	126	8
96	136	134	2
97	112	111	1
98	70	61	9
99	89	82	7
100	84	81	3

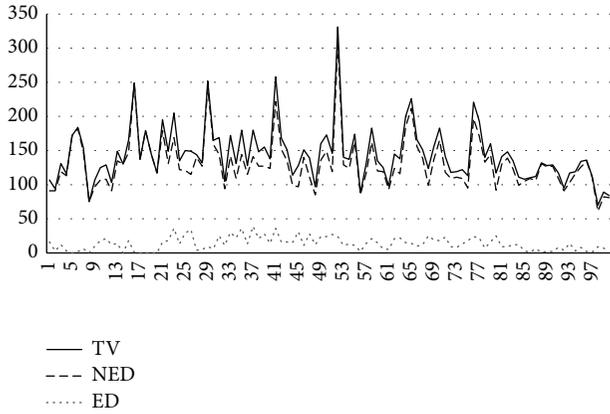


FIGURE 2: Sample data plot.

The arrival rate (λ) is given by the elective demand (ED) and nonelective demand (NED) forecasts.

Now the purpose of this section is to do the following.

- (1) We use the dealer management system (DMS) data to estimate the probability distributions of the following random variables:
 - (a) reception time,
 - (b) parts time,
 - (c) workshop time.
- (2) Subsequently, we use these distributions to feed our queuing system and estimate the distribution of its random variables, in terms of the reception time, parts time, and workshop time described referred in the previous point:
 - (a) customer time at the reception of the service,
 - (b) parts lead time,
 - (c) parts availability,
 - (d) working time per vehicle,
 - (e) additional specific queuing model parameters (defined below in Section 2.3.1 Queuing Methodology for Service Times).

We define the total preparation time (TSR) as the necessary time in minutes to deal with the customer, find the required parts, and take the vehicle to the technician. This variable is not measured in the industry and requires physical checks on the field and data sampling to understand its structure. The total preparation time (TSR) is thus the result of adding the time the customer is at the dealership and the time to get the parts physically:

$$\text{TSR} = \text{TCust} + \text{TParts}, \quad (2)$$

where

- (i) TCust (customer waiting time) is the time to manage customer request at the reception desk and raise a job card;

- (ii) TParts (parts lead time) is the time to get the parts supplied before being fitted to the car in the service.

Now, knowing the car fleet for a particular region we estimate from the dealer management system (DMS) data the distribution of the service time per vehicle (TSW), which will vary with the model, region, service skills and competency, workshop layout, and other parameters.

2.3.1. Queuing Methodology for Service Times. As we discussed in the introduction, queuing theory allows for the study of waiting lines and servers, including arriving patterns at the queue, waiting times, and tasks completed [37].

We build a queuing system into the Matlab code, with the following 6 characteristics.

- (1) Arrival pattern of customers: as mentioned before, it is a constant rate process, where the rate is a function of the total demand (TV) and therefore will depend on the elective demand (ED) and nonelective demand (NED) probability distributions.
- (2) Service pattern of customers: it depends on the number of customers queuing for service and will be a function of the distribution of the customer waiting time (TCust).
- (3) Queue discipline: in our research we set priorities in terms of part availability. If a part is backordered, the customer will be requested to wait. Therefore, this will be a function of the distribution of the parts lead time (TParts).
- (4) Queuing capacity: it is limited by the number of appointments plus the emergency visits. It depends on the customer waiting time (TCust) and the parts lead time (TParts). Thus, it is a function of the distribution of the total preparation time (TSR).
- (5) Number of servers: rather than considering a two-stage server system (reception and workshop), we set our system as a single-level server, where customers leave their vehicles at the reception but they do not physically wait until it is taken to the workshop. Therefore, this characteristic will be a function of the service time per vehicle (TSW). Maximum number of servers is given by the facility layout and could vary with time depending on the technician's availability, including holiday periods, sick leaves, and training courses. We will simulate this variability as part of our methodology.
- (6) Number of work phases along the complete service process: similar to hospital's emergency room (ER) we assume a single stage service for the whole service process, but we simulate the time variability due to the work complexity and different kinds of services. Thus, it is a function of the service time per vehicle (TSW).

In Figure 4, customers (C_1, C_2, \dots, C_c) arrive at the reception area; they could be part of elective demand (ED) or nonelective demand (NED) with their own probability

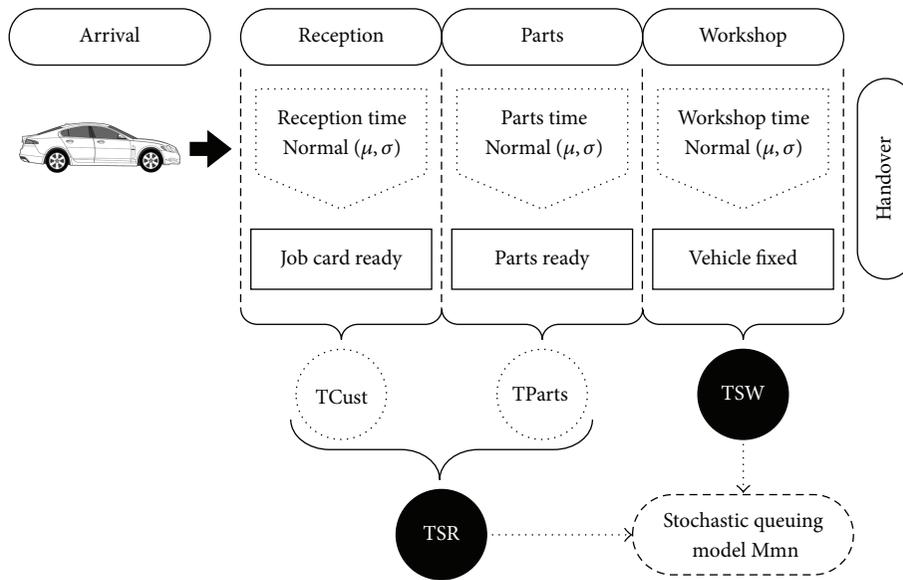


FIGURE 3: Service process times: total preparation time (TSR) and service time per vehicle (TSW).

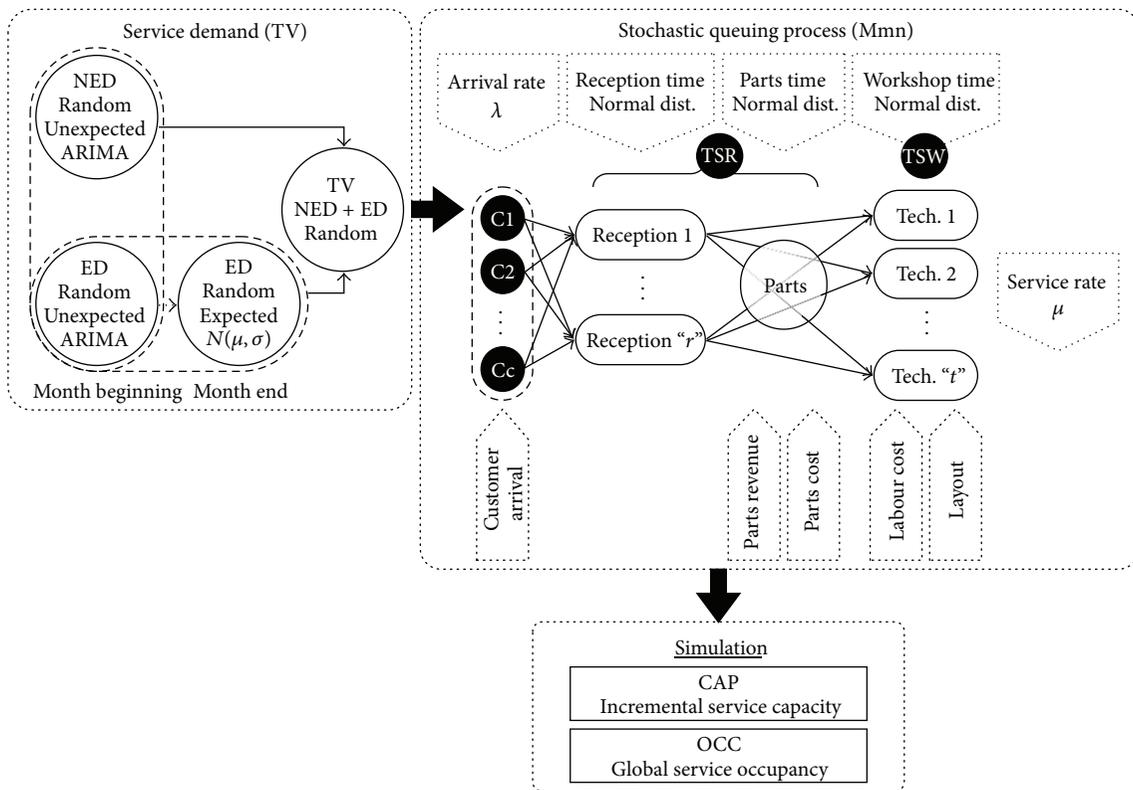


FIGURE 4: Theoretical process for capacity reserve and service occupancy estimation.

distribution, but once at the service reception, they travel through the network and are served at the reception nodes (receptions $1, 2, \dots, r$) through parts and to the workshop nodes (Tech. $1, 2, \dots, t$).

The first variable to be defined is the customer's arrival rate (λ), where C_c is the number of customers waiting at the

queue, where, as an open network, customers can join and leave the system as shown in Figure 4.

Arrival rate (λ) and service rate (μ) are function random variables: the total preparation time (TSR) and the service time per vehicle (TSW), respectively; therefore, they are random variables.

TABLE 2: Queuing system steady-state measures of effectiveness.

Steady-state measures of effectiveness	Variable	Name
$N_{veh} = \frac{TV}{WDM}$	Number of vehicles arriving to the service	N_{veh}
Simulation variable	Average number of vehicles in the Q-NTec	L
Simulation variable	Average number of vehicles in the Queue	L_q
$FTI = 1 - \frac{W_{tec}}{N_{Tec}} = P_0 + \left(\frac{s-1}{s}\right) \times P_1 + \dots + \left(\frac{1}{s}\right) \times P_{s-1}$	Fraction of time that a technician is idle	FTI
$VhW = N_{Veh} - L$	Average number of vehicles that are being worked	VhW
$W_{tec} = L - L_q = N_{Tec} - \sum_{n=0}^{s-1} (s-n) * P_n$	Average number of technicians that are working	W_{tec}
$EFC_{Tec} = \frac{VhW}{N_{Veh}}$	Operating efficiency per vehicle	EFC_{veh}
$EFC_{veh} = \frac{W_{tec}}{N_{Veh}}$	Operating efficiency per technician	EFC_{Tec}
$\mu = \frac{1}{TSW \times 60}$	Service rate [vehicles/hour]	μ
$\lambda = \frac{1}{TSR \times 60}$	Arrival rate [vehicles/hour]	λ
$P_n = \frac{N_{Veh}!}{h! \times (N_{Veh} - h)!} \times \left(\frac{\lambda}{\mu}\right)^h \times P_0$	Probability n vehicles are in the queuing system (Q-NTec)	P_n
$P_0 = \frac{1}{(P_0)^{-1}}$	Probability of no calling units in the queuing system.	P_0
Initial P_0 inverse = 0, then: $P_0 = (P_0)^{-1} + den_1$	P_0 inverse	P_0^{-1}
$f(den_1) = \frac{N_{Veh}!}{h! \times (N_{Veh} - h)!} \times \left(\frac{\lambda}{\mu}\right)^h$	Operator	$f(den_1)$

In queuing theory, the state of the system is given by a vector with different variables. The complete list of variables to be used in the stochastic queuing system (defined by the above characteristics) is detailed in Table 2.

The queuing system allows us to estimate the average cycle service time per vehicle (AT_{PV}), this time is not just calculated adding the service time per vehicle (TSW) and the total preparation time (TSR). It is the result of a complex forecast process to estimate the whole service time, including timing delays due to operational inefficiencies and other system limitations; that is, vehicle movements included in the Service Time per Vehicle (TSW) or some Elective Demand (ED) missing the time window appointment.

The time spent in each of the processes above is a random variable, with its own probability distributions and parameters.

2.4. Stage 3: Service Key Process Indicators (KPIs). This is an essential contribution of this research and consists of the definition of a set of new KPIs for the passenger car industry, in terms of the inputs defined above and of additional random variables and parameters to be defined as follows. A comprehensive list of key process indicators (KPIs) is detailed

in Table 3, showing input variables to feed the model with and output variables to be obtained from the Monte Carlo simulation.

Traditionally, service management organizes and measures the technician's time to administer the service department's labour availability and performance to maximize operational net profit. Technician performance and time control are basically monitored upon the 3 measurement ratios below.

- (i) Productivity: time the technician is physically present at work divided by the actual working hours.
- (ii) Efficiency: time spent working on a vehicle divided by the flat rate time received.
- (iii) Availability: flat rate time received divided by the time the technician is physically present.

Without doubt, these three indicators are useful to monitor technical service performance, but they do not incorporate delays on reception or parts backorder. Therefore, we need to define new measures of effectiveness which take into account inputs from service, parts, and reception and include customer "expectations" in our model. We now focus on the

TABLE 3: List of acronyms: parameters and variables-simulation inputs and outputs.

(a)			
Acronym	Definition	Source	Parameter
Lab	Retail labour rate	DMS data	Constant
Ntec	Number of available technicians		Constant
WBN	Total number of staffed work bays		Constant
Ψ	Number of work bays per technician (WBN/Ntec)		Constant (from 1 to 2)
(b) Simulation Random Inputs			
Acronym	Definition	Source	Prob. distribution
ED	Elective demand		ARIMA
HPRES	Working time per month (hours)		$N(399,65; 85,92)$
InvoT	Invoiced time		$N(373,76; 76,9)$
NED	Nonelective demand (NED = TV – ED)		ARIMA
Nveh	Number of vehicles arriving to the service per day	DMS data	$N(6,13; 1,85)$
PartsCost	Cost of parts sale		$N(21297,95; 11963,79)$
PS	Parts sale		$N(27278,41; 13517,82)$
TV	Total demand		ARIMA
WDM	Working days per month		$N(21; 1)$
WTD	Daily working time (hours per day)		$N(6,51; 1,13)$
(c) Simulation random outputs			
Acronym	Definition	Source	
ATPV	Average cycle service time per vehicle (hours)		
CAP	Service incremental capacity		
CAPRNED	Capacity reserve cost to serve all nonelective demands		
CAPRNED ₁	Capacity reserve cost to serve 1 unexpected vehicle		
CEW	Empty work bay cost estimation		
DMS	Dealer management system		
EFC	Service system efficiency		
EFC _{Tec}	Operating efficiency per technician		
EFC _{Veh}	Operating efficiency per vehicle		
GSR	Monthly gross service revenue estimation		
InvoTn	Invoiced time estimation		
OCC	Service occupancy	Queuing + simulation	
PartsCostn	Cost of parts sale estimation		
PFunit	Profit per vehicle in service		
PSn	Parts sale estimation		
Pstec	Parts sale per technician		
Tcust	Customer waiting time		
TEXP	Monthly workshop total cost estimation		
Tparts	Parts lead time		
TPF	Service total gross profit estimation		
TSR	Total preparation time. $TSR = TCust + TParts$		
TSW	Service time per vehicle		
VhW	Average number of vehicles that are being worked		
Wtec	Average number of technicians that are working		
(d) Other queuing system outputs			
Acronym	Definition	Source	
f(den.1)	Mathematical operator		
FTI	Fraction of time that a technician is idle	Queuing + simulation	
L	Average number of vehicles in the Q-NTec		

(d) Continued.

Acronym	Definition	Source
Lq	Average number of vehicles in the queue	
P_n	Probability n vehicles are in the queuing system (Q-Ntec)	
Po	Probability of no calling units in the queuing system	
Po^{-1}	Po inverse	
λ	Arrival rate [vehicles/hour]	
μ	Service rate [vehicles/hour]	

seven new key process indicators (KPIs) that we are going to define:

- (i) service incremental capacity (CAP),
- (ii) service occupancy (OCC),
- (iii) nonelective demand (NED),
- (iv) cost of capacity reserve (RCAP),
- (v) cost of empty work bay (CEW),
- (vi) capacity reserve cost to serve all nonelective demand (CAPRNED),
- (vii) profit per vehicle in service (PF_{UNIT}).

2.4.1. Service Incremental Capacity. We define the service incremental capacity (CAP) as the system potentiality to admit additional workload without interrupting on-going works. In other words, the service incremental capacity (CAP) is the capability to accommodate unexpected customers during the normal working time.

Service incremental capacity (CAP) is calculated as a percentage rate of the whole service process capacity that will decrease as long as the number of vehicles through the system increases:

$$CAP = \frac{WBN \times WTD}{ATPV} \times \Psi \times EFC. \quad (3)$$

Therefore, we express the service incremental capacity (CAP) as a function of the following.

- (i) WBN is the total number of staffed work bays. It is constant and depends on the workshop layout.
- (ii) WTD is the daily working time.
- (iii) ATPV is the average cycle service time per vehicle.
- (iv) Ψ is the number of work bays/technician, which usually can vary from 1 to 2. It is constant and will depend on the facility layout.
- (v) EFC is the average workshop efficiency rate.

2.4.2. Service Occupancy. We define the service occupancy (OCC) as the measurement of vehicles in-progress through the service system. It is related to the number of vehicles entering the system divided by the work bays and the system capability expressed as service incremental capacity (CAP).

Service occupancy (OCC) is calculated as a percentage rate of the whole service process workload that will grow as long as the number of vehicles through the system increases.

$$OCC = \frac{TV}{CAP \times WDM}. \quad (4)$$

OCC is the service occupancy expressed above as a function of the following.

- (i) Total demand (TV) is the total visit number in a given time period.
- (ii) WDM is the total working days in a given month.

Now, if we write nonelective demand (NED) as a function of (1) and (4), then

$$NED = (OCC \times CAP \times WDM) - ED. \quad (5)$$

2.4.3. Cost of Capacity Reserve. We define cost of capacity reserve (RCAP) as the opportunity cost in € to reserve service capacity in the form of empty work bay. It is expressed as the following equation:

$$RCAP = \frac{TPF}{TV} + \frac{TEXP}{TV}, \quad (6)$$

where

- (i) TPF is the monthly workshop gross profit,
- (ii) TEXP is the monthly workshop total cost.

The first component of the above expression is the profit per unit (PF_{UNIT}):

$$PF_{UNIT} = \frac{TPF}{TV}. \quad (7)$$

By replacing (7) with (6), we define the capacity reserve cost for a single nonelective demand ($CAPRNED_1$):

$$CAPRNED = PF_{UNIT} + \frac{TEXP}{NED}. \quad (8)$$

Also, by replacing (5) with (8),

$$CAPRNED_1 = PF_{UNIT} + \frac{TEXP}{(OCC \times CAP \times WDM) - ED}. \quad (9)$$

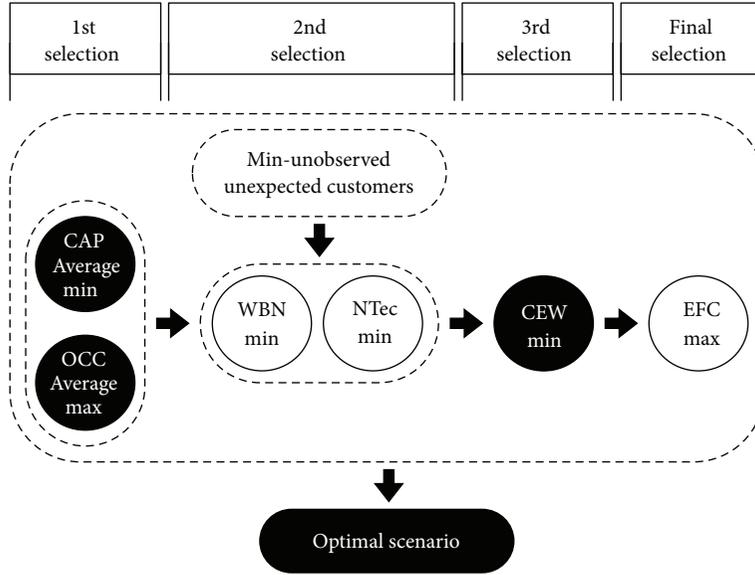


FIGURE 5: Optimal scenario selection criteria in 5 levels.

Therefore, the capacity reserve cost to serve all nonelective demand (CAPRNED) will be

$$\text{CAPRNED} = \text{NED} \times \text{CAPRNED}_1. \quad (10)$$

And then replacing (8) with (10),

$$\text{CAPRNED} = (\text{NED} \times \text{PF}_{\text{UNIT}}) + \text{TEXP}. \quad (11)$$

Given that service incremental capacity (CAP), service occupancy (OCC), nonelective demand (NED), and capacity reserve cost to serve all nonelective demand (CAPRNED) are function of the random variables detailed in Table 3, they are also random variables whose joint distribution will be estimated through Monte Carlo simulation.

2.5. Stage 4: Monte Carlo Simulation. The purpose of this section is to obtain Monte Carlo samples of the distributions of the key process indicators (KPIs) defined above.

As covered in the literature review, the queuing system, or Markov chain, actually simulated is only an approximation of the true chain. Such results with finite precision and range are introduced and pose further questions about the validity of these algorithms [38]. Thus, we apply the Monte Carlo methodology to generate samples of the input variables defined in Section 2 and propagate their uncertainty obtaining samples of the distributions of the key process indicators (KPIs).

This stage has been implemented in a Matlab code which runs a complete set of calculations to simulate capacity reserve with a queuing model to obtain samples of the same size for each KPI.

We repeat the simulations for a number of scenarios (changing number of technicians and of work-bays). The Matlab code will run two loops taking different values of both variables (number of technicians and of number of work-bays). The value of each variable will be increased 1 by 1 each

loop to obtain different samples of the joint distribution of the key process indicators (KPIs). The sequence is as follows:

number of technicians → number of work bays → queuing system loop.

For each KPI, the simulation will then store the results in a tridimensional matrix: the first index varies with the sample (i.e., from 1 to 1000), the second with the number of technicians, and the third with the number of work bays.

3. Statistical Analysis of Key Process Indicators and Optimization

The purpose of this section is twofold:

- (a) statistical analysis of the Monte Carlo samples of the joint distributions of our new key process indicators (KPIs);
- (b) set the optimization criteria to define the optimum scenario in terms of recommended number of technicians and number of work bays.

After running the simulation we can estimate the joint and marginal distributions of the key process indicators (KPIs) and identify if there are additional relationships among them. A further statistical analysis is discussed in Section 4.

After the previous stages have been fully completed, our Matlab code identifies the service optimal scenario as a tradeoff between the dealer's total demand (TV) and the capacity reserve cost to serve all nonelective demands (CAPRNED).

TABLE 4: Simulation-decision support system final report.

(a)						
Description	Constant					
Working days per month	WDM		21			
Nonelective demand	NED ₁		12			
Work bays per technician	WBtec ₂		1,25			
(b)						
Description	Variable	Average	Standard deviation			
Customer waiting time (min)	TCust	37,61	12,966			
Parts lead time (min)	TParts	3586,87	2080,052			
Total preparation time (min)	TSR	3624,48	2080,052			
Service time per vehicle (min)	TSW	6533,74	3730,76			
% Time a technician is idle	FTI	16,89%	0,24			
Working time per day (hours)	WTD	7,5	0,58			
Parts sold per technician (€)	PStec	4755,03	1458,17			
% Time sold at retail price	RLab	72,40%	0,072			
Effective labour price (€)	ELab	66,13	0,486			
(c)						
Description	Simulation results (Ntec)			Simulation results (Ntec + 1)		
	Variable	Average	Standard deviation	Variable	Average	Standard deviation
Efficiency per vehicle	EFCvh ₁	34,60%	0,25	EFCvh ₂	36,31%	0,24
Average service efficiency	EFC ₁	86,50%	0,23	EFC ₂	83,11%	0,24
Workshop available time (min)	HPRES ₁	1102,43	85,3	HPRES ₂	1259,92	97,48
Average cycle service time per vehicle (hours)	ATPV ₁	3,46	0,95	ATPV ₂	3,79	1,13
Service incremental capacity	CAP ₁	26,80%	—	CAP ₂	20,50%	—
Service occupancy	OCC ₁	49,00%	—	OCC ₂	64,00%	—
Profit per vehicle (€)	PFunit ₁	120,84	36,98	PFunit ₂	137,83	42,27
Estimation of parts sale (€)	PS ₁	33285,2	10207,19	PS ₂	38040,2	11665,36
Work bays per technician	WBtec ₁	1,43	0	WBtec ₂	6,65	1,917
Cost of empty bay (€)	CEW ₁	34,27	6,301	CEW ₂	39,07	7,191
Capacity reserve cost for all nonelective demands (€)	CAPRNED ₁	1861,32	449,97	CAPRNED ₂	2122,85	514,798
Capacity reserve cost to serve 1 unexpected customer (€)	CAPRNED ₁₁	155,11	37,498	CAPRNED ₂₁	176,9	42,9
Number of cars in the workshop	VhW ₁	4,5	3,206	VhW ₂	4,72	3,15

The optimal scenario will be selected following the hierarchical approach displayed in Figure 5 and detailed below.

(1) The code will select those scenarios with a maximum service occupancy (OCC) and minimum capacity reserve cost to serve all nonelective demands (CAPRNED).

(2) Then it will order the selected scenarios starting from

(a) the lowest number of work bays (layout constraints and cost),

(b) the lowest number of technicians (operational constraints and cost),

(c) the lowest number of unobserved unexpected customers.

(3) A 3rd level will filter those scenarios with the lowest cost for an empty server (work bay).

(4) The final level filters and selects the scenario with the highest whole service (reception, parts, and workshop) operational efficiency out of the previous selection. With the information from the “optimal

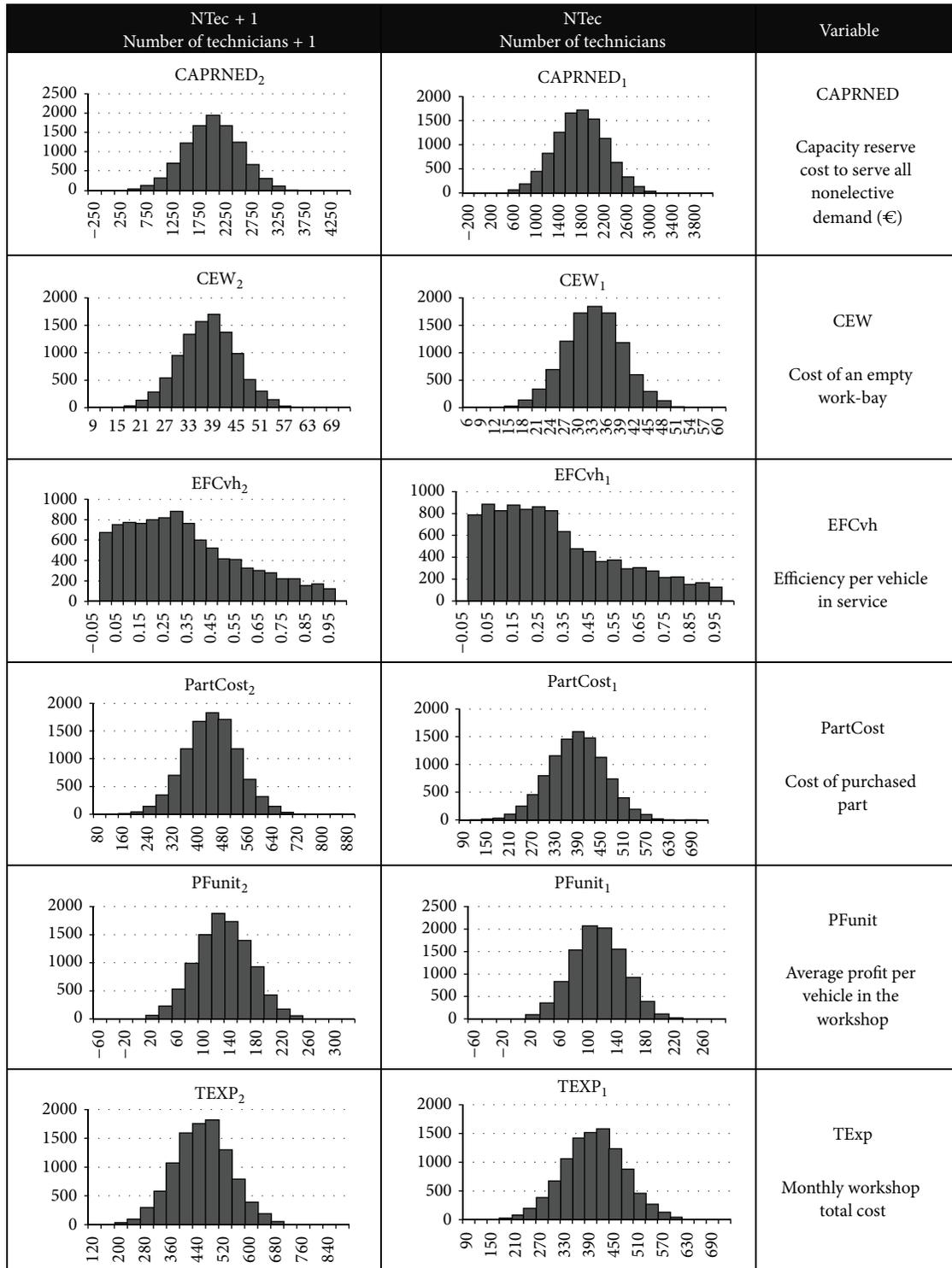


FIGURE 6: Simulation-output variables histograms.

TABLE 5: 95% confidence intervals.

(a)					
Description	Input variable	Mean	Standard error	95% Confidence intervals	
				Lower limit	Upper limit
Elective demand	ED	147,859	4,32304	139,305	156,414
Nonelective demand	NED	13,3672	0,824012	11,7366	14,9978
Total demand	TV	161,328	4,51834	152,286	170,168

(b)				
Description	Input variable	Sigma	Lower limit	Upper limit
Elective demand	ED	48,9096	43,5628	55,7652
Nonelective demand	NED	9,32263	8,30348	10,6294
Total demand	TV	51,1192	45,5308	58,2845

(c)					
Output variable	Output variable	Mean	Standard error	Lower limit	Upper limit
Capacity reserve cost for all nonelective demands (€)	CAPRNED ₁	1861,14	4,5	1852,31	1869,96
	CAPRNED ₂	2122,64	5,15	2112,54	2132,74
Cost of an empty work bay (€)	CEW ₁	34,26	0,06	34,14	34,39
	CEW ₂	39,07	0,07	38,93	39,21
Efficiency per vehicle in service (%)	EFCvh ₁	0,35	0,001	0,34	0,35
	EFCvh ₂	0,36	0,001	0,36	0,37
Cost of purchased parts (€)	PartCost ₁	402,27	0,76	400,79	403,75
	PartCost ₂	458,65	0,86	456,96	460,34
Average profit per vehicle in the workshop (€)	PFunit ₁	120,83	0,37	120,11	121,56
	PFunit ₂	137,82	0,42	136,99	138,65
Monthly workshop total cost (€)	TExp ₁	411,18	0,76	409,69	412,66
	TExp ₂	468,83	0,86	467,14	470,52

(d)				
Output variable	Output variable	Sigma	Lower limit	Upper limit
Capacity reserve cost for all nonelective demands (€)	CAPRNED ₁	450,33	444,18	456,66
	CAPRNED ₂	515,21	508,17	522,45
Cost of an empty work bay (€)	CEW ₁	6,31	6,22	6,4
	CEW ₂	7,2	7,1	7,3
Efficiency per vehicle in service(%)	EFCvh ₁	0,25	0,24	0,25
	EFCvh ₂	0,24	0,24	0,25
Cost of purchased parts (€)	PartCost ₁	75,54	74,51	76,6
	PartCost ₂	86,22	85,04	87,43
Average profit per vehicle in the workshop(€)	PFunit ₁	37	36,49	37,52
	PFunit ₂	42,29	41,71	42,88
Monthly workshop total cost (€)	TExp ₁	75,71	74,68	76,78
	TExp ₂	86,42	85,24	87,64

scenario,” we will produce a final report (Table 4) which displays the expected key process indicators (KPIs) for the recommended number of work bays (WBN) and the recommended number of technicians (NTec). It also displays a second simulation with an additional technician (NTec + 1) to support the decision making in the following scenarios:

- (i) Short term: identifying how indirect operational revenue or cost can be improved by increasing the operational staff.
- (ii) Medium term: assessing the effectiveness of the current service system and identifying the impact of applying changes to the original service settings.

The report identifies also capacity and occupancy levels for the optimal scenario and how they could be affected when the existing staffs are increased by 1 head, provided that there is at least 1 additional work bay to be used for servicing nonelective demand (NED).

Also, we noticed when service staff is increased and unexpected demand is part of the capacity reserve, the empty work bay cost estimation is also reduced. As stated before, service incremental capacity (CAP) is inversely proportional to the service productive headcount, as it drops as soon as staff increases, while service occupancy (OCC) is directly proportional to the service staff.

This makes sense and confirms the expected outcome; the potential capacity we could have in the system should be lower when an additional vehicle is processed in the service system, showing an increment in the system occupancy rate. The simulation results suggest that services generate costs when reserving service capacity to serve nonelective demand.

4. Simulation Results Analysis and Uncertainty

This section shows the simulation results report in Table 4, including the relevant key process indicators (KPIs) information, and a sample of variables histograms in Figure 6. As said before, we use probabilistic distributions to demonstrate that our methodology obtains valid results integrating uncertainty and adding value to the simple forecasting processes which are common in the passenger car service industry.

In addition to the standard deviations shown below, interquartile ranges or 95% intervals could be easily computed from the samples as complementary measures of uncertainty. This is done in Table 5, where we are showing 95% confidence intervals for the means and standard deviations of each of the selected variables.

Other key process indicators (KPIs) are used to understand how economic variables could change depending on the solution applied, comparing the current layout and situation with the possibility of increasing service staff in one head.

Thus, the cost of empty work bay (CEW_1 and CEW_2 in Table 5) is increased as long as the profit per unit (PF_{unit_1} and PF_{unit_2} in Table 4) raises, so the cost implicit in capacity reserve to suit customer needs affects all the economic factors as we wanted to demonstrate.

Now we will study six variables out of the total number displayed in Table 4 (see Figure 6 and Table 5).

5. Conclusions

This paper studies a new approach, where, by analysing nonelective demand (NED), the apparent inefficiency resulting from services operating within production limits is understood. This analysis could also help brand managers when setting efficiency objectives with adequate adjustment for unexpected demand and its impact on cost structures.

We confirm here how separating nonelective demand (NED) from elective demand (ED) when estimating service costs is of paramount importance, as well as for labour fees setting and service level, which will depend also on how accurate service demand and general costs predictions are. Furthermore, the leftmost column of the simulation report, as displayed in Table 4, identifies some apparent inefficiencies resulting from services operating within production limits.

With this information, the report compares several service process indicators to demonstrate how results can be affected by hiring additional technical staff.

The significance of the contribution of our research is the definition of new key process indicators (KPIs) to be used as a management tool for services. The capacity reserve strategy has been proved to be plausible and consistent, according to the reviewed literature of the hospital's emergency room (ER) field, with our conceptual arguments relating to production responses to demand uncertainty. Therefore, the information used allows for a more detailed specification of service output that can be applied to the passenger car industry to forecast service requisites and plan brand strategies which are aligned with the customer's real demand.

Future research could afford an exhaustive analysis to the data gathered after the Monte Carlo simulation. This could be done with the support of any of the existing statistical software packages to fully understand the existing relations among the multiple key process indicators (KPIs) and between inputs and outputs, like partial correlations and stochastic dependence between the new key process indicators (KPIs) defined in this paper.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] M. A. Cohen, N. Agrawal, and V. Agrawal, "Winning in the after market," *Harvard Business Review*, vol. 84, no. 5, pp. 129–138, 2006.
- [2] R. N. Bolton, "A dynamic model of the duration of the customer's relationship with a continuous service provider: the role of satisfaction," *Marketing Science*, vol. 17, no. 1, pp. 45–65, 1998.
- [3] V. Srivastava and A. Tyagi, "The study of impact of after sales services of passenger cars on customer retention," *International Journal of Current Research and Review*, vol. 5, no. 1, pp. 127–132, 2013.
- [4] R. Katarne, S. Sharma, and J. Negi, "Measurement of service quality of an Automobile Service Centre," pp. 286–291, Proceedings of the 2010 International Conference on Industrial Engineering and Operations Management, Dhaka, Bangladesh, January 2010.
- [5] M. A. Cohen, N. Agrawal, and V. Agrawal, "Achieving breakthrough service delivery through dynamic asset deployment strategies," *Interfaces*, vol. 36, no. 3, pp. 259–271, 2006.
- [6] S. S. AmbekarIndian, "Service quality gap analysis of automobile service centers," *Indian Journal of Research in Management, Business and Social Sciences*, vol. 1, no. 1, pp. 38–41, 2013.
- [7] A. Berndt, "Investigating service quality dimensions in South African motor vehicle servicing," *African Journal of Marketing Management*, vol. 1, no. 1, pp. 1–9, 2009.
- [8] M. Hanif, S. Hafeez, and A. Riaz, "Factors affecting customer satisfaction," *International Research Journal of Finance and Economics*, vol. 63, no. 60, pp. 44–52, 2010.
- [9] A. Herrmann, F. Huber, and A. Gustafsson, "From value orientated quality improvement to customer satisfaction: a

- case study for passenger cars," in *Customer Retention in the Automotive Industry*, pp. 93–115, Gabler, 1997.
- [10] A. A. Jahanshahi, M. A. Hajizadeh Gashti, S. A. Mirdamadi, K. Nawaser, and S. M. Sadeq Khaksar, "Study the effects of customer service and product quality on customer satisfaction and loyalty," *International Journal of Humanities and Social Science*, vol. 1, no. 7, pp. 253–260, 2011.
- [11] M. D. Johnson, A. Herrmann, F. Huber, and A. Gustafsson, "An introduction to quality, satisfaction, and retention—implications for the automotive industry," in *Customer Retention in the Automotive Industry*, pp. 1–17, Gabler, 1997.
- [12] R. D. Kamenetzky, L. J. Shuman, and H. Wolfe, "Estimating need and demand for prehospital care," *Operations Research*, vol. 30, no. 6, pp. 1148–1167, 1982.
- [13] M. A. Badri and J. Hollingsworth, "A simulation model for scheduling in the emergency room," *International Journal of Operations and Production Management*, vol. 13, no. 3, article 13, 1993.
- [14] Y. Gerchak, D. Gupta, and M. Henig, "Reservation planning for elective surgery under uncertain demand for emergency surgery," *Management Science*, vol. 42, no. 3, pp. 321–334, 1996.
- [15] M. Bazargan, S. Bazargan, and R. S. Baker, "Emergency department utilization, hospital admissions, and physician visits among elderly African American persons," *Gerontologist*, vol. 38, no. 1, pp. 25–36, 1998.
- [16] S. C. Brailsford, V. A. Lattimer, P. Tarnaras, and J. C. Turnbull, "Emergency and on-demand health care: modelling a large complex system," *Journal of the Operational Research Society*, vol. 55, no. 1, pp. 34–42, 2004.
- [17] P. Beraldi, M. E. Bruni, and D. Conforti, "Designing robust emergency medical service via stochastic programming," *European Journal of Operational Research*, vol. 158, no. 1, pp. 183–193, 2004.
- [18] L. V. Green, S. Savin, and B. Wang, "Managing patient service in a diagnostic medical facility," *Operations Research*, vol. 54, no. 1, pp. 11–25, 2006.
- [19] S. A. Jones, M. P. Joy, and J. Pearson, "Forecasting demand of emergency care," *Health Care Management Science*, vol. 5, no. 4, pp. 297–305, 2002.
- [20] B. Friedman and M. Pauly, "Cost functions for a service firm with variable quality and stochastic demand: the case of hospitals," *The Review of Economics and Statistics*, vol. 63, no. 4, pp. 620–624, 1981.
- [21] W. H. Wiersema, "Forecasting for inventory," *Electrical Apparatus*, vol. 12, article 40, no. 49, 1996.
- [22] S. C. Graves, "A single-item inventory model for a non-stationary demand process," Working Papers WP, 3944-97, Massachusetts Institute of Technology (MIT). Sloan School of Management, 1997.
- [23] P. Alfredsson and J. Verrijdt, "Modeling emergency supply flexibility in a two-echelon inventory system," *Management Science*, vol. 45, no. 10, pp. 1416–1431, 1999.
- [24] E. S. Gardner Jr., "Evaluating forecast performance in an inventory control system," *Management Science*, vol. 36, no. 4, pp. 490–499, 1990.
- [25] A. Federgruen and Y. Zheng, "Efficient algorithm for computing an optimal (r, Q) policy in continuous review stochastic inventory systems," *Operations Research*, vol. 40, no. 4, pp. 808–813, 1992.
- [26] C. Dance and A. A. Gaivoronski, "Stochastic optimization for real time service capacity allocation under random service demand," *Annals of Operations Research*, vol. 193, no. 1, pp. 221–253, 2012.
- [27] J. Hwang, L. Gao, and W. Jang, "Joint demand and capacity management in a restaurant system," *European Journal of Operational Research*, vol. 207, no. 1, pp. 465–472, 2010.
- [28] K. Gilbert, "An ARIMA supply chain model," *Management Science*, vol. 51, no. 2, pp. 305–310, 2005.
- [29] J. D. Dana, "Using yield management to shift demand when the peak time is unknown," *RAND Journal of Economics*, vol. 30, no. 3, pp. 456–474, 1999.
- [30] H. J. Finarelli Jr. and T. Johnson, "Effective demand forecasting in 9 steps," *Healthcare Financial Management*, vol. 58, no. 11, pp. 52–58, 2004.
- [31] M. Utley, S. Gallivan, T. Treasure, and O. Valencia, "Analytical methods for calculating the capacity required to operate an effective booked admissions policy for elective inpatient services," *Health Care Management Science*, vol. 6, no. 2, pp. 97–104, 2003.
- [32] M. Gaynor and W. B. Vogt, "Competition among hospitals," *RAND Journal of Economics*, vol. 34, no. 4, pp. 764–785, 2003.
- [33] K. A. Ariyawansa, C. Cacho, and A. J. Felt, "A family of stochastic programming test problems based on a model for tactical manpower planning," *Journal of Mathematical Modelling and Algorithms*, vol. 4, no. 4, pp. 369–390, 2005.
- [34] M. Armony, E. Plambeck, and S. Seshadri, "Sensitivity of optimal capacity to customer impatience in an unobservable M/M/S queue (why you shouldn't shout at the dmv)," *Manufacturing and Service Operations Management*, vol. 11, no. 1, pp. 19–32, 2009.
- [35] J. Chomicki, D. Goldin, G. Kuper, and D. Toman, "Variable independence in constraint databases," *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 6, pp. 1422–1436, 2003.
- [36] N. Litvak, M. van Rijsbergen, R. J. Boucherie, and M. van Houdenhoven, "Managing the overflow of intensive care patients," *European Journal of Operational Research*, vol. 185, no. 3, pp. 998–1010, 2008.
- [37] H. L. Lee and M. A. Cohen, "Multi-agent customer allocation in a stochastic service system," *Management Science*, vol. 31, no. 6, pp. 752–763, 1985.
- [38] G. O. Roberts and J. S. Rosenthal, "Markov-chain Monte Carlo: some practical implications of theoretical results," *Canadian Journal of Statistics*, vol. 26, no. 1, pp. 5–29, 1998.
- [39] M. P. Bailey, "Estimating service capacity using nonhomogeneous work arrivals," in *Proceedings of the 1993 Winter Simulation Conference*, pp. 408–413, ACM Press, 1993.
- [40] D. Hughes and A. McGuire, "Stochastic demand, production responses and hospital costs," *Journal of Health Economics*, vol. 22, no. 6, pp. 999–1010, 2003.
- [41] A. F. Seila, "On waiting times for a queue in which customers require simultaneous service from a random number of servers," *Operations Research*, vol. 32, no. 5, pp. 1181–1184, 1984.

Research Article

A Framework to Determine the Effectiveness of Maintenance Strategies Lean Thinking Approach

Alireza Irajpour,¹ Ali Fallahian-Najafabadi,¹
Mohammad Ali Mahbod,² and Mohammad Karimi³

¹ Faculty of Management and Accounting, Islamic Azad University, Qazvin Branch, Qazvin, Iran

² Department of Management, University of Isfahan, Iran

³ Islamic Azad University, South Tehran Branch, Tehran, Iran

Correspondence should be addressed to Ali Fallahian-Najafabadi; alifallahian@yahoo.com

Received 26 March 2014; Accepted 6 June 2014; Published 10 July 2014

Academic Editor: M. Chadli

Copyright © 2014 Alireza Irajpour et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The purpose of this paper is to provide a framework that can identify and evaluate the effectiveness of a given maintenance strategy and to rank components of maintenance system. The framework is developed using DEMATEL method on maintenance strategy as a guideline. To gain a richer understanding of the framework, a questionnaire is constructed and answered by experts. Then the DEMATEL method is applied to analyze the importance of criteria and the casual relations among the criteria are constructed. The scope of the paper is limited to performance measurement of maintenance strategies. It is found that the framework is applicable and useful for the strategic management of the maintenance function. It is observed that the influencing and preferred infrastructures for designing Learning and Training are three components, that is, optimal maintenance, CMMS, and RCM which are interdependent on each other and are the fundamental components to realize the designed goals of maintenance process. This paper provides an overview of research and developments in the measurement of maintenance performance. Many tools and techniques have been developed in other fields. However, the applicability of those tools to maintenance function has never been tried. In that respect this topic is novel. It helps in managing maintenance more effectively.

1. Introduction

Some challenges from modern competitors have provoked many industrial companies to implement new manufacturing approaches [1, 2]. Particularly salient among these is the concept of lean production [3, 4]. Lean production is an approach that includes a set of management practices, including just in time, quality systems, work teams, cellular manufacturing, and supplier management, in an integrated system. The main core of lean production is that these practices can work synergistically to create a high quality system for reaching customer demand with no waste.

Some articles on the topic of lean production system emphasize the relationship between implementation of lean and performance. While most of these studies have focused on a single aspect of lean and its performance implications (e.g., [6–8]), a few studies have explored the implementation and performance relationship with two aspects of lean (e.g.,

[8, 9]). Even fewer studies have investigated the simultaneous synergistic effects of multiple aspects of lean implementation and performance implication. A noteworthy exception is Cua et al.'s [10] investigation of implementation of practices related to just in time (JIT), total quality management (TQM), and total preventive maintenance (TPM) programs and their impact on operational performance. However, conceptual research continues to stress the importance of empirically examining the effect of multiple dimensions of lean production programs simultaneously [11].

Companies implement lean strategies to achieve better quality, designing the processes which meet customer requirements and expectations, waste elimination (waste is any activity that does not add value to the product or service) and lead time reduction (it helps a Lean enterprise deliver the products to the customer in a shorter time and reduce total costs, both direct and indirect) [12].

Since waste elimination is one of the Lean objectives, it is crucial for companies to identify wastes relevant to defects, waiting time, overproduction (producing more, earlier, or sooner than next workstation demand results in larger inventory and costs), transportation (transportation within Work-In-Process (WIP) resulting from weak plant layout and shortage in understanding of production or process flow), inventory (excess raw materials, finished products, and WIP), unused creativity (failure in exploiting the knowledge and unique abilities of the employees), movement (extra transportation due to wrong location of equipment and tools), and overprocessing (parts of processes that create no added-value to the product or service) [13, 14].

Lean principles have been originated from Toyota's production system known as just in time (JIT) production [15, 16]. The term lean has become widespread after the publication of a book titled *The Machine That Changed the World*. Then, the term lean production was widely used. Mason-Jones et al. [17] have matched various strategies of supply chain with product type. They have introduced a "leagile" approach which determines the decoupling point between lean and agile paradigms in a supply chain. Sullivan et al. [18] have presented the performance of equipment replacement decision problems within the context of lean manufacturing.

They utilized VSM as a road map for providing necessary information for the analysis of equipment replacement decision problem in lean manufacturing implementation.

Muda and Hendry [19] have proposed a world class manufacturing concept incorporated with lean principles for the make-to-order sector. Pavnaskar et al. [20] have presented a classification scheme for lean manufacturing tools. They have suggested that their classifications scheme enables companies to become lean and serve as a foundation for research into lean concepts. Many researchers have contributed to the definition of lean manufacturing. Shah and Ward [11] have provided a comprehensive definition of lean production which is an integrated sociotechnical system whose objective is to eliminate waste by reducing and minimizing the supplier, customer, and internal variability.

The tools and techniques of lean manufacturing include TQM, TPM, Kanban, Kaizen, SMED, Poka-Yoke, and visual control. Houshmand and Jamshidnezhad [21] have presented an extended model of design process of lean production system by means of process variables. They have used axiomatic design theory for developing hierarchical structure to model a design process of lean production system composed of functional requirements, design parameters, and process variables. Braglia et al. [22] have presented a new approach for a complex production system based on seven iterative steps associated with typical industrial engineering tools including VSM. Shah and Ward [23] have defined the measures of lean production. They have mapped the various conceptual lean strategies [24].

Shin et al. [25] have provided the basic data-driven methods including off-line design and on-line computation algorithms; original idea, basic assumption/condition, and

computation complexity were presented. Provided methods were implemented on an industrial benchmark.

1.1. Evolution of Equipment Management. To begin with, there is a requirement to improve an understanding of the basic perception of the maintenance role. Here, it is pertinent to note that the maintenance function has undergone serious change in the last three decades. The traditional perception of maintenance's role is to fix broken items. Taking such a narrow view, maintenance activities have been confined to the reactive tasks of repair actions or item replacement. Thus, this approach is identified as reactive maintenance, breakdown maintenance, or corrective maintenance. A more recent view of maintenance is defined by Gits [26] as "All activities aimed at keeping an item in, or restoring it to, the physical state considered necessary for the fulfilment of its production function." Clearly, the scope of this opinion also contains the proactive tasks such as the following:

- routine servicing and periodic inspection,
- preventive replacement,
- condition monitoring.

In order to maintain equipment, maintenance must carry out some further activities. These activities contain the planning of work, purchasing and control of materials, personnel management, and quality control [27]. This variety of responsibilities and activities convert maintenance from a simple function to a complex function to manage.

Maintenance should ensure equipment availability in order to produce products at the compulsory quantity and quality levels [28]. The scope of maintenance management includes every phase in the life cycle of technical systems (plant, machinery, equipment, and facilities), specification, acquisition, planning, operation, performance evaluation, improvement, and disposal [29].

1.1.1. Breakdown Maintenance (BM). This type of maintenance states the maintenance strategy, after the equipment failure equipment is repaired [30]. This maintenance strategy was mainly implemented in the manufacturing organizations before 1950. In this stage, machines are serviced only when repair is required. This idea has some weaknesses such as the following:

- unplanned stoppages,
- excessive damage,
- spare parts problems,
- high repair costs,
- excessive waiting and maintenance time,
- high trouble shooting problems [31].

1.1.2. Preventive Maintenance (PM). This concept is a type of physical checkup of the equipment to prevent equipment breakdown. Preventive maintenance includes activities which are started after a period of time or amount of machine use [32]. This type of maintenance depends on the estimated

probability that the equipment will break down in the specified interval. The preventive works are as follows:

- equipment lubrication,
- cleaning,
- parts replacement,
- tightening,
- adjustment.

1.1.3. Predictive Maintenance (PdM). Predictive maintenance is often mentioned as condition based maintenance (CBM). In this strategy, maintenance is started in response to a specific equipment condition or performance deterioration [33]. The analytic techniques are organized to measure the physical condition of the equipment such as temperature, noise, vibration, lubrication, and corrosion [34]. When one or more of these indicators reach a set deterioration level, maintenance initiatives are assumed to restore the equipment to desired condition. This means that equipment is taken out of service only when direct evidence exists that deterioration has happened. Predictive maintenance is based on the same principle as preventive maintenance. The advantages of predictive maintenance are based on the need to perform maintenance only when the repair is really necessary, not after a specified period of time [32].

1.1.4. Corrective Maintenance (CM). The main core of this concept is to prevent equipment failures. This type of maintenance system has been applied to the improvement of equipment; hence the equipment failure can be removed (improving the reliability) and the equipment can be simply maintained (improving equipment maintainability) [35]. The main difference between corrective and preventive maintenance is based on the time of maintenance action. In the corrective action system a problem must exist before corrective actions are taken [36]. The corrective maintenance is following some purposes such as

- improving equipment reliability,
- maintainability,
- safety,
- reducing design weaknesses (material, shapes),
- reducing deterioration and failures,
- aiming at maintenance-free equipment.

1.1.5. Maintenance Prevention (MP). This type of maintenance system is based on the design phase of equipment. Equipment is designed such that they are maintenance free and an ideal condition of “what the equipment and the line must be” is attained [35]. In the development of new equipment, MP activities must begin at the design stage of equipment [37]. Maintenance prevention often applies some earlier equipment failures and feedback from production areas to ensure equipment design for production systems.

1.1.6. Reliability Centered Maintenance (RCM). RCM can be defined as an organized, rational process for improving the maintenance requirements of a physical resource in its operating context to understand its “inherent reliability,” where “inherent reliability” is the level of reliability which can be attained with an effective maintenance program. RCM is a process implemented to determine the maintenance requirements of any machines or equipment in its operating context by recognizing their functions, the causes of failures, and the effects of the failures.

RCM has seven basic steps:

- (1) identify the equipment/system to be analyzed;
- (2) determine its functions;
- (3) determine what constitutes a failure of those functions;
- (4) identify the failure modes that cause those functional failures;
- (5) identify the impacts or effects of those failures' occurrence;
- (6) use RCM logic to select appropriate maintenance tactics;
- (7) document your final maintenance program and refine it as you gather operating experience [38].

The various tools employed for affecting maintenance improvement on these 7 steps include

- (1) failure mode and effect analysis (FMEA),
- (2) failure mode effect and criticality analysis (FMECA),
- (3) physical hazard analysis (PHA),
- (4) fault tree analysis (FTA),
- (5) optimizing maintenance function (OMF),
- (6) hazard and operability (HAZOP) analysis.

1.1.7. Productive Maintenance (PrM). The main aim of productive maintenance is to increase the productivity of a manufacturing unit by decreasing the total cost of the equipment over the whole life from design to equipment degradation. The significant features of this maintenance viewpoint are equipment maintainability and reliability focus, as well as cost reduction of maintenance actions. The maintenance strategy including all previous viewpoints to increase equipment productivity by applying preventive maintenance, corrective maintenance, and maintenance prevention is named productive maintenance [39, 40] (see Figure 3).

1.1.8. Computerized Maintenance Management Systems (CMMSs). Computerized maintenance management systems (CMMSs) are vigorous for the management of all activities related to the availability, productivity, and maintainability of complex systems. Modern computational facilities have offered a dramatic scope for improved effectiveness and efficiency in, for example, maintenance. Computerized maintenance management systems (CMMSs) have existed, in

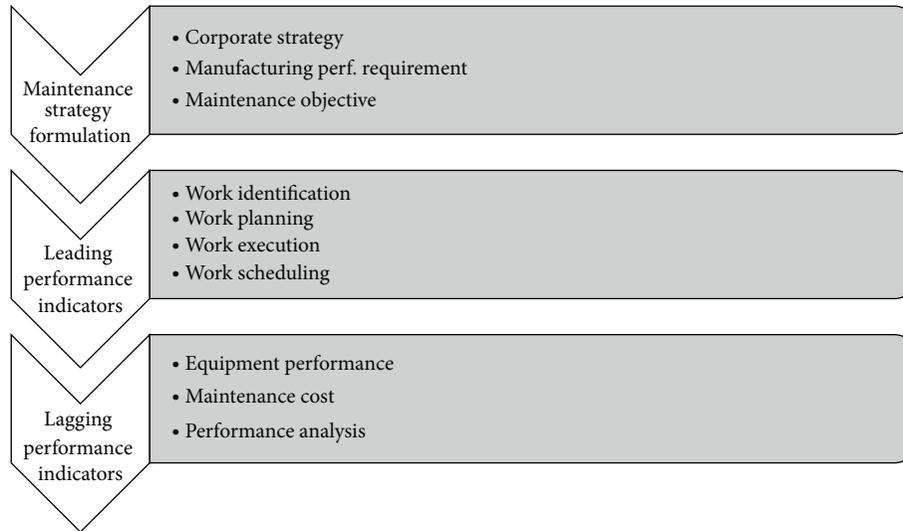


FIGURE 1: The performance measurement framework for the maintenance function.

one form or another, for several decades. CMMS can be used to mechanize the PM function and to help in the control of maintenance inventories and the buying of materials. CMMS can reinforce reporting and analysis capabilities [41, 42]. Accessibility and accuracy of information can provide more reliable decisions in CMMS because of closer working relationships between maintenance and production [43, 44].

1.1.9. Total Productive Maintenance (TPM). This methodology is linked to the maintenance systems designed and perfected by “Toyota family” companies including Denso and Aisin Seiki [45, 46].

TPM is an innovative approach to maintenance that optimizes equipment effectiveness, removes breakdowns, and promotes autonomous maintenance by operators through day-to-day activities involving total workforce [47]. A strategic approach to improve the performance of maintenance activities is to effectively implement strategic TPM initiatives in the manufacturing organizations. TPM brings maintenance into attention as a necessary part of the business. The TPM initiative is aimed at improving competitiveness of organizations TPM try to find to involve all levels and functions in an organization to optimize the overall effectiveness of production equipment. This method also tunes up existing processes and equipment by reducing mistakes and accidents. TPM is a world class manufacturing (WCM) initiative that pursues to optimize the effectiveness of manufacturing equipment [48].

1.2. Lean TPM and Maintenance Performance Framework. The integration of Lean Thinking [3] and total productive manufacturing (Lean TPM) applies the proven business models of “world class” manufacturing enterprise.

The maintenance performance conceptual framework proposed by Muchiri et al. [49] recognizes main processes

that lead the maintenance function to delivery of performance required by manufacturing objectives. The conceptual framework supports alignment of maintenance objectives with the manufacturing and corporate objectives. The conceptual framework has three main sections that include maintenance alignment with manufacturing, maintenance effort/process analysis, and maintenance of results performance analysis.

The first area of the conceptual framework pursues aligning the maintenance objectives with the manufacturing strategy. By studying the requirements of the stakeholders, the performance requirements of the manufacturing system can be well-defined. Cognitive mapping is a crucial tool for studying the cause and effect relationship between strategic essentials [50].

1.3. The Maintenance Performance Indicators. The maintenance performance framework summarizes the main essentials that are central in the management of the maintenance function. The essentials make sure that the right work is recognized and effectively implemented for definite results that are in line with the manufacturing performance requirements. Each step is important for effective management of the maintenance function. Both the maintenance process (leading) indicators and maintenance results (lagging) indicators are vital for measuring the performance of the maintenance function. For each essential, the main encounter is to recognize the performance indicators that will express whether the essentials are managed well. Efficient indicators should cover control and monitoring performance and support maintenance actions towards achievement of objectives. Muchiri et al. [49] have provided some indicators that appear often in literature. The classification of Muchiri et al. [49] is applied in this paper for facilitation of performance measurement of maintenance system (see Figure 1).

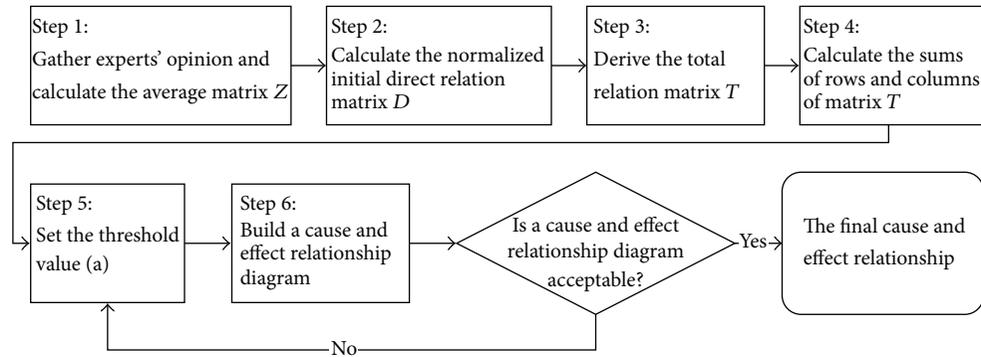


FIGURE 2: The process of the DEMATEL method (adopted from [5]).

1.3.1. Maintenance Process (Leading) Indicators. The maintenance leading indicators monitor some maintenance processes that are as follows:

- work identification,
- work planning,
- work scheduling,
- work execution.

Key performance indicators for each process are proposed by Muchiri et al. [49] to measure if requirements of each process are satisfied.

1.3.2. Maintenance Results (Lagging) Indicators. The results of the maintenance process can be divided into efficiency of technical systems and cost systems. The lagging indicators are used to measure maintenance results in terms of equipment performance and maintenance cost [49].

1.4. DEMATEL Method. DEMATEL method was developed between 1972 and 1979 by the Science and Human Affairs Program of the Battelle Memorial Institute of Geneva, with the purpose of studying the complex and intertwined problematic group.

It has been extensively recognized as one of the methods to explain the cause and effect relationship among the criteria [51–55]. This method is applied to analyze the relationship between cause and effects and among evaluation criteria [56] or to originate interrelationship among factors [55].

Based on Shieh et al. [57], the procedure of DEMATEL method is presented in Figure 2.

2. Research Methodology

This research includes both classes of fundamental and applied researches because it seeks to explain relations and indices and present model. In this research, survey method is used in terms of time and is descriptive (nonexperimental) research in terms of data collection and classification. Case and field research method is used for presentation of model and technique and final conclusion because the present research includes a series of the methods which

aims to describe conditions or relations between the studied phenomena on which basis the technique and model are presented by recognizing the previous position and presenting the present position completely.

The present research is conducted with field and library method such as library references including books, publications, theses, and sources available in the university, scientific centers have been used for theoretical information and review of the literature, and direct observation has been used for collecting data.

2.1. Research Domain (Theme, Space, and Time). Business management and plant management problems particularly problems of maintenance management are some of the multidimensional problems and have abundant complexities which may not be evident at the first look. But it becomes evident after deliberation that these problems have deep and extensive dimensions and their multilateral study may be time consuming. Therefore, it is necessary that each researcher first specify limits of his research to prevent confusion and waste of time and sources.

Thematic domain: in this research, researcher tries to study empowering and effective components of maintenance process optimization in manufacturing industries and presents a suitable and practical model for effective application of each component to achieve goals of organization.

Spatial domain: place of this research includes 9 plants, that is, Pol Film and Atlas Film (production of polymer packaging), Plot and Azin Chap (printing and production of heliogravure print cylinder), Zanjan and Bead Wire Industries in the field of wire and welding electrode production, Fardan Aryan Industries (production of food packs and preform), MEdisk (production of optical compact disc, CD and DVD) and Arya Kian Industries (automobile steering assembly line).

Time domain: time of the present research is year of 2012 in six-month period and information, statistics, and documents relate to this time period.

Requirement of each applied research is study and recognition of factors affecting working field of research. Information collection instrument is questionnaire. Different maintenance strategies and indices and key elements for

implementation of lean maintenance were identified and these concepts were used in two questionnaires.

In questionnaire 1, the experts were asked to specify effect of each lean maintenance element as pairwise comparison and announce in front of each row if these components have been applied in their plants and in what level these components are applied for its implementation.

In questionnaire 2, experts were asked to determine effect and significance of each component affecting making maintenance process lean based on four groups of leading indicators and two groups of lagging indicators. This effect is specified according to Saati's scale. It is necessary to note that the following scale is positive for the positive indicators. To determine effect and significance of the component relating to negative indicators, reverse Saati's scale is used.

The indices were determined considering application of interpretive structural modeling approach to ensure theoretical dominance, practical experience, and access due to time-consuming and different types of the questionnaire compared with the common questionnaires. To ensure comprehensiveness of attitudes, the following indices were obtained:

- relationship between working experience of experts and maintenance issue,
- the presence of experts as maintenance managers and senior experts,
- experts with the related academic education.

Study of the papers which have used interpretive structural method for analyzing results has suggested the number of experts to be between 4 and 64. 64 selected experts include maintenance managers of 9 plants selected for field studies and 5 experienced maintenance experts are working in these plants.

2.2. Data Analysis Method. Extraction of useful results from a research requires application of suitable scientific, accurate, and confirmed methods. In this regard, the following stages are used for analyzing data in this research.

Step 1. It is extraction of approximate agreement matrix for intensity of direct relations between lean maintenance process components adapted from data of questionnaire 6 which was filled by 64 selected experts.

Step 2. This is to structure effect of each lean production component on each other and study feedback and its relations and determine effect of DEMATEL method on lean maintenance process. Application of this method can give the research a suitable structure considering relations between lean maintenance components and make the optimal model policy possible for explaining strategy with lean approach in maintenance process. By extracting two influencing and influenced indices between components of lean maintenance process, effect of components on each other can be evaluated to analyze them properly for ranking these components considering significance of each of them:

- making direct relations matrix normal,
- creating general relations matrix,

creating cause and effect matrix,

creating Dependence matrix,

specifying influence order of elements on each other,

extracting influencing and influenced indices of lean production components.

Step 3. Approximate agreement matrix is extracted using data of questionnaire 2 and lean maintenance components are regarded as alternative and maintenance leading and lagging indices are regarded as criterion to which output of Step 2, that is, influencing and influenced indices of components, is added.

Step 4. Determining weight of indices using Shannon entropy method: to determine weight of the index, Shannon entropy method is used instead of experts' view due to uncertainty. In this analysis, it specifies weight and significance of each group of leading and lagging indices and influencing and influenced indices of the process lean production components.

Step 5. Ranking and determining weights of lean maintenance components using TOPSIS technique: TOPSIS technique has been used for ranking and weighting because this technique is a compensatory summative method which compares alternatives through weight of each criterion and normalized numbers on each criterion and calculation of the criterion. It is assumed that TOPSIS has an equal measure for increasing or decreasing values.

Normalization usually requires parameters and criteria which almost have heterogeneous dimensions in multicriteria problems. In compensatory methods, such as TOPSIS, exchange between criteria is compulsory so that weak results of a criterion are neutralized and compensated through good results. Compensatory methods preset more realistic form than noncompensatory methods which ignore the solutions obtained through cuts applied on them:

creating normalized decision matrix,

calculating weighted normal matrix,

determining positive and negative ideal point,

calculating Euclidean distance of positive and negative ideal solutions,

ranking: these numbers are ranked decreasingly to select the preferred option.

Based on output values of TOPSIS, weights of each component are specified using weighted mean method.

Step 6. This step encompasses studying the trend of indices considering implementation levels of lean maintenance components to determine relationship between components of lean maintenances process and trend of maintenances indices in 9 plants selected for field studies, to evaluate effectiveness or loss of efficiency of maintenance process in each one of the plants in six-month period for application or nonapplication of the lean maintenance components.

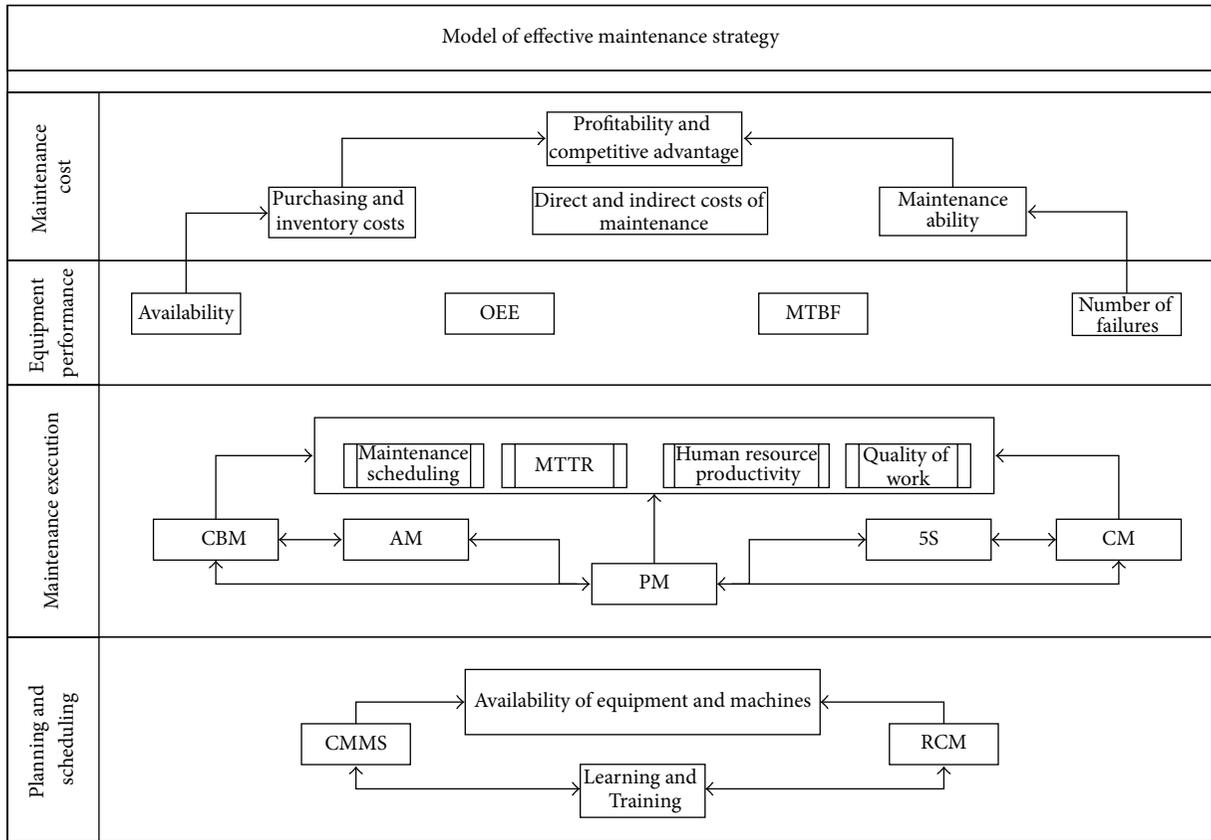


FIGURE 3: Model of effective maintenance strategy.

2.3. *Model Integration.* There is an integration of outputs of DEMATEL and TOPSIS methods that shows elements dependency and ranks that have been specified. Final results of the conducted field studies and analysis of trend between lean maintenances components and trend of maintenances indices have been achieved. An optimized model is presented for explaining maintenance strategy with lean thinking approach and depicted in Figure 3.

3. Research Findings and Analysis

In this way, 9 components have been identified as the factors of which proper implementation can make maintenance process lean and optimal. These components are as follows:

- inventory management with lean approach,
- 5S,
- CMMS,
- training and learning,
- CBM,
- RCM,
- AM,
- PM,
- CM.

Extraction of approximate agreement matrix for intensity of direct relationship between components of lean maintenance process.

9 × 9 the following pairwise comparison matrix of which components are taken from the studies conducted in review of the literature is a matrix 9 which has been taken from data of questionnaire 6 by 64 selected experts based on group decision making entitled approximate agreement matrix for intensity of direct relationship between components of lean maintenance process. The results have been shown is Table 1.

3.1. *Determining Effect of Lean Maintenance Components Using DEMATEL.* The following results are obtained by implementing DEMATEL technique as shown in Table 2.

In this step, we specify influence order of elements on each other. Elements of column d_i indicate hierarchy of the influencing elements and order of elements of column r_i indicates hierarchy of the influenced elements. $(d - j)$ indicates position of an element and this position will be certainly influencing in case of positivity $(d - j)$ and will be certainly influenced in case of negativity. $(d + j)$ indicates sum of intensity of an element in terms of influencing element and influenced element. Position of an element will be certainly influencing in case of positivity $(d - j)$ and will be certainly influenced in case of negativity. The above analysis shows that CMMS, RCM, education, and culture building are the most effective on their effective execution in

TABLE 1: Pairwise matrix of Lean elements of maintenance.

Lean elements of maintenance	Inventory management with lean approach	5S	CMMS	Training and learning	CBM	RCM	AM	PM	CM
Inventory management with lean approach	0	3	1	2	2	2	2	4	4
5S	4	0	3	3	3	2	4	3	3
CMMS	4	3	0	2	4	4	2	4	4
Training and learning	2	4	2	0	3	3	4	3	4
CBM	3	2	3	2	0	4	3	3	4
RCM	3	2	2	4	4	0	3	4	4
AM	1	2	2	2	3	3	0	4	4
PM	4	2	3	2	2	2	3	0	4
CM	4	2	2	1	1	2	2	2	0

TABLE 2: Determining effect of Lean maintenance components using DEMATEL.

	Row sum (di)	Influence matrix T : 9 criteria		
		Column sum (ri)	di + ri	di - ri
Inventory management with lean approach	4.33	5.54	9.87	-1.20
5S	5.42	4.38	9.80	1.04
CMMS	5.82	4.03	9.85	1.79
Training and learning	5.42	3.96	9.38	1.47
CBM	5.22	4.68	9.90	0.54
RCM	5.60	4.75	10.35	0.85
AM	4.61	4.98	9.58	-0.37
PM	4.75	5.80	10.54	-1.05
CM	3.57	6.63	10.20	-3.06

TABLE 3: Ranking based on influenced indices of lean production components.

Ranking based on influenced indices of lean production components	
Component	Column sum (ri)
CM	6.63
PM	5.80
Inventory management with lean approach	5.54
AM	4.98
RCM	4.75
CBM	4.68
5S	4.38
CMMS	4.03
Training and learning	3.96

TABLE 4: Ranking based on influencing indices of lean production components.

Ranking based on influencing indices of lean production components	
Component	Row sum (di)
CMMS	5.82
RCM	5.60
Training and learning	5.42
5S	5.42
CBM	5.22
PM	4.75
AM	4.61
Inventory management with lean approach	4.33
CM	3.57

the first level compared with other components. Such analysis specifies that CM and PM, spare parts store and purchase will be mostly influenced by application and execution of other components. It is necessary to note that these three components are certainly influenced by other components. Outputs of di will be used as influencing positive index and ri will be used as influenced negative index in the next step

for ranking lean maintenance components. The results have been shown in Tables 3 and 4.

3.2. Extraction of Approximate Agreement Matrix for Evaluating Lean Maintenance Components as Maintenance Indices. Approximate agreement matrix is extracted using data of questionnaire 2 and lean maintenance components are regarded as alternative and maintenance leading and lagging

TABLE 5: Determining weight of indices using Shannon entropy method.

Type	Group	Weight
Leading indicator	Work identification	12.32%
	Work planning	12.40%
	Work scheduling	13.24%
	Work execution	20.69%
Lagging indicator	Equipment performance	13.22%
	Maintenance cost	28.12%

TABLE 6: Ranking by TOPSIS.

Ranking by TOPSIS		
Lean production components	cli +	Cumulative weight
PM	0.6422	13.8%
CMMS	0.6286	13.5%
RCM	0.5918	12.7%
Inventory management with lean approach	0.5325	11.5%
CM	0.4955	10.7%
Training and learning	0.4758	10.3%
CBM	0.4479	9.6%
5S	0.4172	9.0%
AM	0.4104	8.8%

indices are regarded as criterion to which output of Step 2, that is, influenced and influencing indices of components, is added. The indices which are included in dark cells are regarded as positive indices and the indices which are in bright cells are regarded as negative indices.

3.3. Determining Weight of Indices Using Shannon Entropy Method. Considering the extracted weights, it is observed that weight of each main class of the maintenance leading and lagging indices is shown in Table 5.

Considering the analysis in Table 6, we see that 4 components, that is, PM, CMMS, RCM, and spare parts purchase and store, with lean thinking approach are the most important components for moving toward optimum in maintenance. These four components include 52% of the lean weight of the maintenance process.

4. Model Presentation

Outputs of integration methods have been applied in selecting optimal maintenance strategy, optimal model is presented for explaining maintenance with lean thinking approach.

In this model which was presented as bottom-up method, it is observed that the influencing and preferred infrastructures for designing Learning and Training are three components, that is, optimal maintenance, CMMS, and RCM which are interdependent on each other and are the fundamental components to realize the designed goals of

maintenance process. In the next step, other components of lean maintenance are given to realize other goals.

5. Conclusion

The presented model is an applied model which can be used in different plants and different production lines for optimizing maintenance process. In the performed analyses, it was observed that each one of the lean maintenance components should be valued differently; significance and weight of each of them should be included properly in budgeting for execution. To execute these components for realizing maintenance goals, one should start with the mentioned infrastructures and then apply other components.

In the further researches based on the methodology introduced in the research, one can replace ANP (analytical network process) with TOPSIS or utilize combinatory statistical analysis of Fisher's test and logistic regression test instead of multicriteria decision methods and compare their results with the present research.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] R. W. Hall, *Attaining Manufacturing Excellence: Just-in-Time, Total Quality, Total People Involvement*, Dow Jones-Irwin, Homewood, Ill, USA, 1987.
- [2] J. R. Meredith and R. McTavish, "Organized manufacturing for superior market performance," *Long Range Planning*, vol. 25, no. 6, pp. 63–71, 1992.
- [3] J. P. Womack and D. T. Jones, *Lean Thinking: Banish Waste and Create Wealth in Your Corporation*, Simon & Schuster, New York, NY, USA, 1996.
- [4] J. P. Womack, D. T. Jones, and D. Roos, *The Machine That Changed the World*, Harper Perennial, New York, NY, USA, 1990.
- [5] D. Sumrit and P. Anuntavoranich, "Using DEMATEL method to analyze the causal relations on technological innovation capability evaluation factors in Thai technology-based firms," *International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies*, vol. 4, no. 2, 2012.
- [6] J. R. Hackman and R. Wageman, "Total quality management: empirical, conceptual, and practical issues," *Administrative Science Quarterly*, vol. 40, no. 2, pp. 309–342, 1995.
- [7] D. Samson and M. Terziowski, "Relationship between total quality management practices and operational performance," *Journal of Operations Management*, vol. 17, no. 4, pp. 393–409, 1999.
- [8] K. E. McKone, R. G. Schroeder, and K. O. Cua, "Impact of total productive maintenance practices on manufacturing performance," *Journal of Operations Management*, vol. 19, no. 1, pp. 39–58, 2001.
- [9] B. B. Flynn, S. Sakakibara, and R. G. Schroeder, "Relationship between JIT and TQM: practices and performance," *Academy of Management Journal*, vol. 38, no. 5, pp. 1325–1360, 1995.

- [10] K. O. Cua, K. E. McKone, and R. G. Schroeder, "Relationships between implementation of TQM, JIT, and TPM and manufacturing performance," *Journal of Operations Management*, vol. 19, no. 6, pp. 675–694, 2001.
- [11] R. Shah and P. T. Ward, "Lean manufacturing: context, practice bundles, and performance," *Journal of Operations Management*, vol. 21, no. 2, pp. 129–149, 2003.
- [12] S. Bhasin and P. Burcher, "Lean viewed as a philosophy," *Journal of Manufacturing Technology Management*, vol. 17, no. 1, pp. 56–72, 2006.
- [13] J. Drew, B. McCallum, and S. Roggenhofer, *Journey to Lean: Making Operational Change Stick*, Palgrave Macmillan, New York, NY, USA, 2004.
- [14] M. Arashpour, M. R. Enaghani, and R. Andersson, "The rationale of lean and TPM," in *Proceedings of the International Conference of Engineering and Operations Management (IEOM '10)*, Dhaka, Bangladesh, 2010.
- [15] Z. Tang, R. Chen, and J. I. Xuehong, "An innovation process model for identifying manufacturing paradigms," *International Journal of Production Research*, vol. 43, no. 13, pp. 2725–2742, 2005.
- [16] F. K. Pil and T. Fujimoto, "Lean and reflective production: the dynamic nature of production models," *International Journal of Production Research*, vol. 45, no. 16, pp. 3741–3761, 2007.
- [17] R. Mason-Jones, B. Naylor, and D. R. Towill, "Lean, agile or league? Matching your supply chain to the marketplace," *International Journal of Production Research*, vol. 38, no. 17, pp. 4061–4070, 2000.
- [18] W. G. Sullivan, T. N. McDonald, and E. M. Van Aken, "Equipment replacement decisions and lean manufacturing," *Robotics and Computer-Integrated Manufacturing*, vol. 18, no. 3-4, pp. 255–265, 2002.
- [19] S. Muda and L. Hendry, "Proposing a world-class manufacturing concept for the make-to-order sector," *International Journal of Production Research*, vol. 40, no. 2, pp. 353–373, 2002.
- [20] S. J. Pavnaskar, J. K. Gershenson, and A. B. Jambekar, "Classification scheme for lean manufacturing tools," *International Journal of Production Research*, vol. 41, no. 13, pp. 3075–3090, 2003.
- [21] M. Houshmand and B. Jamshidnezhad, "An extended model of design process of lean production systems by means of process variables," *Robotics and Computer: Integrated Manufacturing*, vol. 22, no. 1, pp. 1–16, 2006.
- [22] M. Braglia, G. Carmignani, and F. Zammori, "A new value stream mapping approach for complex production systems," *International Journal of Production Research*, vol. 44, no. 18-19, pp. 3929–3952, 2006.
- [23] R. Shah and P. Ward, "Defining and developing measures of lean production," *Journal of Operations Management*, vol. 25, no. 4, pp. 785–805, 2007.
- [24] S. Vinodh and S. K. Chintha, "Leanness assessment using multi-grade fuzzy approach," *International Journal of Production Research*, vol. 49, no. 2, pp. 431–445, 2011.
- [25] Y. Shin, S. X. Ding, A. Haghani, H. Hao, and P. Zhang, "A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process," *Journal of Process Control*, vol. 22, no. 9, pp. 1567–1581, 2012.
- [26] C. W. Gits, "Design of maintenance concepts," *International Journal of Production Economics*, vol. 24, no. 3, pp. 217–226, 1992.
- [27] V. Priel, *Systematic Maintenance Organization*, McDonald and Evan, London, UK, 1974.
- [28] L. M. Pintelon and L. F. Gelders, "Maintenance management decision making," *European Journal of Operational Research*, vol. 58, no. 3, pp. 301–317, 1992.
- [29] M. Murray, K. Fletcher, J. Kennedy, P. Kohler, J. Chambers, and T. Ledwidge, "Capability assurance: a generic model of maintenance," in *Proceedings of the 2nd International Conference of Maintenance Societies*, pp. 1–5, Melbourne, Australia, 1996.
- [30] T. Wireman, *World Class Maintenance Management*, Industrial Press, New York, NY, USA, 1990.
- [31] A. D. Telang, "Preventive maintenance," in *Proceedings of the National Conference on Maintenance and Condition Monitoring*, K. Vijayakumar, Ed., pp. 160–73, Government Engineering College, Institution of Engineers, Cochin Local Centre, Thissur, India, February 1998.
- [32] F. Herbaty, *Handbook of Maintenance Management: Cost Effective Practices*, Noyes Publications, Park Ridge, NJ, USA, 2nd edition, 1990.
- [33] D. Vanzile and I. Otis, "Measuring and controlling machine performance," in *Handbook of Industrial Engineering*, G. Salvendy, Ed., John Wiley & Sons, New York, NY, USA, 1992.
- [34] R. Brook, "Total predictive maintenance cuts plant costs," *Plant Engineering*, vol. 52, no. 4, pp. 93–95, 1998.
- [35] H. R. Steinbacher and N. L. Steinbacher, *TPM for America*, Productivity Press, Portland, Ore, USA, 1993.
- [36] L. R. Higgins, D. P. Brautigam, and R. K. Mobley, *Maintenance Engineering Handbook*, McGraw-Hill, New York, NY, USA, 5th edition, 1995.
- [37] K. Shirose, *TPM for Operators*, Productivity Press, Portland, Ore, USA, 1992.
- [38] J. D. Campbell, *The Reliability Handbook*, Clifford-Elliott, Ontario, Canada, 1999.
- [39] Y. Wakaru, *TPM for Every Operator*, Productivity Press, Portland, Ore, USA, 1988.
- [40] B. Bhadury, *Total Productive Maintenance*, Allied Publishers, New Delhi, India, 1988.
- [41] R. Hannan and D. Keyport, "Automating a maintenance work control system," *Plant Engineering*, vol. 45, no. 6, pp. 108–110, 1991.
- [42] T. Singer, "Are you using all the features of your CMMS? Following this even-step plan can help uncover new benefits," *Plant Engineering*, vol. 53, no. 1, p. 32, 1999.
- [43] R. Dunn and D. Johnson, "Getting started in computerized maintenance management," *Plant Engineering*, vol. 45, no. 7, pp. 55–58, 1991.
- [44] A. Labib, "Computerised maintenance management systems," in *Complex System Maintenance Handbook*, pp. 417–435, Springer, London, UK, 2008.
- [45] S. Nakajima, *Introduction to Total Productive Maintenance (TPM)*, Productivity Press, Portland, Ore, USA, 1988.
- [46] D. McCarthy and N. Rich, *Lean TPM: A Blueprint for Change*, Butterworth-Heinemann, 2004.
- [47] B. Bhadury, "Management of productivity through TPM," *Productivity*, vol. 41, no. 2, pp. 240–251, 2000.
- [48] K. Shirose, *TPM Team Guide*, Productivity Press, Portland, Ore, USA, 1995.
- [49] P. Muchiri, L. Pintelon, L. Gelders, and H. Martin, "Development of maintenance function performance measurement framework and indicators," *International Journal of Production Economics*, vol. 131, no. 1, pp. 295–302, 2011.

- [50] F. Ackermann, C. Eden, and I. Brown, *The Practice of Making Strategy: A Step by Step Guide*, SagePublication, London, UK, 2005.
- [51] Y. J. Chiu, H. C. Chen, G. H. Tzeng, and J. Z. Shyu, "Marketing strategy based on customer behaviour for the LCD-TV," *International Journal of Management and Decision Making*, vol. 7, no. 2-3, pp. 143-165, 2006.
- [52] J. J. H. Liou, G. Tzeng, and H. Chang, "Airline safety measurement using a hybrid model," *Journal of Air Transport Management*, vol. 13, no. 4, pp. 243-249, 2007.
- [53] G. H. Tzeng, C. H. Chiang, and C. W. Li, "Evaluating intertwined effects in e-learning programs: a novel hybrid MCDM model based on factor analysis and DEMATEL," *Expert Systems with Applications*, vol. 32, no. 4, pp. 1028-1044, 2007.
- [54] W. W. Wu and Y. T. Lee, "Developing global managers' competencies using the fuzzy DEMATEL method," *Expert Systems with Applications*, vol. 32, no. 2, pp. 499-507, 2007.
- [55] C. Lin and G. Tzeng, "A value-created system of science (technology) park by using DEMATEL," *Expert Systems with Applications*, vol. 36, no. 6, pp. 9683-9697, 2009.
- [56] Y. P. Yang, H. M. Shieh, J. D. Leu, and G. H. Tzeng, "A novel hybrid MCDM model combined with DEMATEL and ANP with applications," *International Journal Operational Research*, vol. 5, no. 3, pp. 160-168.
- [57] J. I. Shieh, H. H. Wu, and K. K. Huang, "A DEMATEL method in identifying key success factors of hospital service quality," *Knowledge-Based Systems*, vol. 23, no. 3, pp. 277-282, 2010.

Research Article

Demand Management Based on Model Predictive Control Techniques

Yasser A. Davizón, Rogelio Soto, José de J. Rodríguez, Ernesto Rodríguez-Leal, César Martínez-Olvera, and Carlos Hinojosa

School of Engineering, Tecnológico de Monterrey, 64849 Monterrey, NL, Mexico

Correspondence should be addressed to Yasser A. Davizón; a00534333@itesm.mx

Received 31 March 2014; Revised 30 May 2014; Accepted 30 May 2014; Published 26 June 2014

Academic Editor: Hamid Reza Karimi

Copyright © 2014 Yasser A. Davizón et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Demand management (DM) is the process that helps companies to sell the right product to the right customer, at the right time, and for the right price. Therefore the challenge for any company is to determine how much to sell, at what price, and to which market segment while maximizing its profits. DM also helps managers efficiently allocate undifferentiated units of capacity to the available demand with the goal of maximizing revenue. This paper introduces control system approach to demand management with dynamic pricing (DP) using the model predictive control (MPC) technique. In addition, we present a proper dynamical system analogy based on active suspension and a stability analysis is provided via the Lyapunov direct method.

1. Introduction

A supply chain is a complex system that involves suppliers, producers, distribution centers, retailers, and finally the customers. The main purpose of the supply chain management (SCM) is to optimize the entire chain. In other words, SCM manages efficiently the flow of products and information between elements of the chain in order to attain goals that cannot be reached in an isolated manner. The application of DM to SCM allows us to efficiently balance the customer's requirements and the capabilities of the supply chain (e.g., see [1]).

Nowadays, in such dynamic business world, the rapid learning about the behaviour of demand and its yield is required. The adequate management of these two variables is an important source of competitive advantage for any company in the supply chain. Moreover, traditional yield management is switching into the more complex activity of DM. Therefore our research focuses on the integration of yield or revenue management into the framework of DM. It is well known that hoteliers, airlines, and car rentals companies have implemented yield management activities successfully. According to [2], to include yield management activities into DM enhances profitability and long run sustainability.

Furthermore, DM allows managers to sell the right product or service to the right customer, at the right time, and for the right price with the main objective of maximizing the profits margins. This action can be made through the assignation of units of capacity to the available demand. For example, when dealing with perishable products, DM assumes a fixed capacity. However, in the case of wafer fabrication facilities, if product mix is highly predictable, or if all products use each piece of production equipment equally, all overall production forecast is needed to determine equipment requirements. However, demand has multiple dimensions such as the different products sold, the types of customers served (preferences and purchase behavior), and time. In this context, decisions made about the price or quantity of a product may affect the demand for related products and/or may also affect the future demand for the same product. This scenario represents an opportunity to develop a model that determines how much to sell at what price and to which market segment. Thus we propose the application of DM to SCM as a helpful tool which integrates capacity, inventory level, and demand in DP framework. Dynamic pricing (DP) is the core element of DM and refers to fluid pricing between the buyer and the seller, rather than the more traditional fixed pricing [3]. Based on this, the company's main goal is to maximize its

total expected revenues by making the correct decision. From here our premise is to show that DP can be modeled applying control systems and specifically addressed by the techniques of optimal control in real time, as the well-known model predictive control (MPC). This research work introduces a second order dynamical system which integrates capacity, inventory level, and bidding price in the framework of DM via DP. We utilize a dynamical 1/4 active suspension system to model and prove the applicability of DP to SCM. Moreover, by using the Lyapunov direct method we concluded that a sufficient and necessary condition for stability is the presence of inventory level along the supply chain.

The remainder of this paper's structure is as follows. Section 2 describes DM in semiconductor manufacturing; Section 3 presents sensitivity and stability analysis for DM; Section 4 considers model predictive control formulation for DM with respective simulations, while conclusions are provided in Section 5.

2. Demand Management Sector Applications

2.1. Demand Management Sector Applications: A Review of the Current Literature. The techniques of DM are relatively new and the first research that dealt directly with these issues appeared less than 20 years ago. The major sector of application of DM has been in the airline industry and different approaches have been presented in the literature for airlines and hotels [4–12] (see Table 1), manufacturing, services, and transportation [13–22] (see Table 2). So much research has been done in the DM field. Please refer to Tables 1 and 2 for an extended list of authors and a brief description of their work. There are many different areas in which DM has been applied successfully, as shown in Figure 1. To this day, most of the research has been performed in the airline sector (which is justified because DM was developed to solve problems in this area) and in the area of hotel management, where, by applying DM techniques, it has been possible to mitigate some of the booking limit problems based on reservations. We present a novel approach for dynamic pricing-inventory level based on DM, assuming that we have a pair of products subject to the same demand profile.

2.2. From Supply to Demand Management: The Damping Effect. Synchronization between supply and demand is the optimal scenario in any complex supply network. In dynamical systems, dissipation is considered the loss of energy over time; this is because of damping. It is common to take into account the damping effect in different situations like economics, finance, and electromechanical systems and much more in biological systems. In DM, the notion of damping is related to inventory level, basically to achieve stability conditions over the supply chain.

2.2.1. Finite Dimensional Formulation for Demand Management. The ordinary differential equation (ODE) for the dynamic system (DS) with damping is

$$M \frac{d^2 x}{dt^2} + \beta \frac{dx}{dt} + Kx = f(t). \quad (1)$$

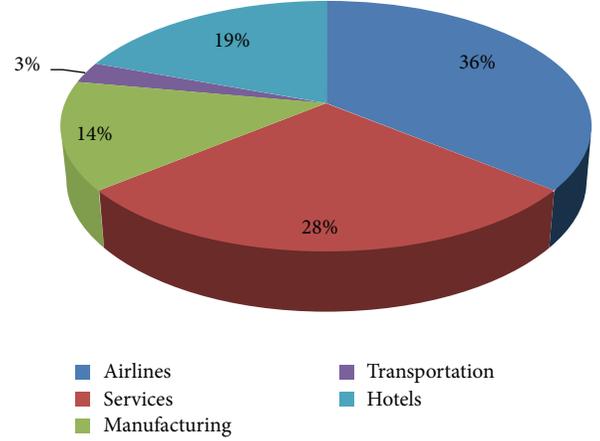


FIGURE 1: Demand management sector applications.

The present model in (1) describes the dynamics of mass-spring-damper second order dynamical system. Also, published by us [23] this model proposes a dynamic pricing for DM applying fast model predictive control approach. The present work is an extension of our previously published work.

For purposes of demand management we present the following model in the damping part of (1):

$$C \frac{d^2 p}{dt^2} + \left[\int_0^\infty \beta(t) dt \right] \frac{dp}{dt} + Kp = d(t). \quad (2)$$

Equation (2) presents p —price level, C —capacity, $d(t)$ —demand, K —bidding price restitution coefficient, and $\beta(t)$ —damping coefficient-inventory level.

The damping factor in (2) has the characteristic to be a function over time and then Lipschitz property can be applied and requires convergence in a scalar value (see Appendix A).

The demand management problem for optimal control has been addressed previously by different research groups [24]; dynamic pricing for optimum inventory level has been previously integrated as a first order ODE as shown in [25]. Moreover, [26] has extended this work by integrating dynamic pricing and demand level into a second order ODE. The interaction of dynamic pricing and demand level [27] presents a dynamic pricing model on web service, in which case the demand function is nonlinear. Optimization techniques have been shown [28] and have been used to quantify demand response in energy industry applying bid-pricing approaches; also, integer linear programming has been applied to model inventory level, production level, and capacity for dynamic pricing in [29].

2.2.2. Infinite Dimensional Formulation for Demand Management. An extended version of ODE in (2) is expressed as a generalized wave equation first proposed in [30] for flexible systems such as

$$C(p) \frac{\partial^2 \varphi}{\partial t^2} + \beta \frac{\partial \varphi}{\partial t} + K\varphi = D(p, t). \quad (3)$$

TABLE 1: Demand management for airlines and hotels: literature review.

Author	Sector application	Methodology	Contribution
Abdel Aziz et al. [4]	Hotels	Research article (case studies)	Application to address the problem of room pricing in hotels. Optimization techniques are present. Multiclass scheme similar to the one implemented in airlines is studied
de Boer [5]	Airlines	Research article (theoretical)	Determination of a revenue management policy in response to the realization of the demand. This is done by matching supply and demand for seats
Hung and Chen [6]	Airlines	Research article (case studies)	This study develops and tests two heuristics approaches: the dynamic seat rationing and the expected revenue gap to help the airline make a fulfillment or rejection decision when a customer arrives
Gosavi et al. [7]	Airlines	Research article (simulation based optimization)	Application of simulation based optimization for seat allocation in airlines. The model considers real life assumption such as cancelations and overbooking
Graf and Kimms [8]	Airlines	Research article (case studies)	The optimal transfer prices are determined by a negotiation process. Option-based capacity control process results combined with transfer price optimization are compared with a first come first served approach
Kimms and Müller-Bungart [9]	Airlines	Research article (simulation based case study)	Application of heuristics and optimal networks in revenue management. Assuming stochastic demand
Netessine and Shumsky [10]	Airlines	Survey	Analysis of fundamental concepts and trade-offs of yield management to describe the parallels between yield management and inventory management
Rannou and Melli [11]	Hotels	Research article (case study)	Analysis and application of revenue management in hospitality industry. The idea is to achieve a performance measure for the return of investment
Walczak and Brumelle [12]	Airlines	Research article (theoretical)	Problem formulation of semi-Markov model for dynamic pricing and revenue management in airline industry, assuming continuous demand

Equation (3) is called master damping equation for demand management, which can be written as partial differential equation (PDE) as follows:

$$v_p \frac{\partial^2 \varphi(p, t)}{\partial p \partial t} + (\beta + a_p) \frac{\partial \varphi(p, t)}{\partial p} + \alpha \beta \frac{\partial \varphi(p, t)}{\partial t} + K \varphi(p, t) = D(p, t), \quad (4)$$

where first derivative of price in time is v_p and second derivative in time is a_p . See Appendix B for a detailed proof of (3) and (4).

3. Robust and Stability Analysis for Demand Management

DM can be stated as follows: how to select the product's mix amount and price in order to satisfy a fluctuating demand with the main goal of maximizing the profit. For this reason, in this document we propose the use of MPC by comparing DM and DP using an active suspension with damping.

3.1. Physical Modeling in DM-DP Problem. The use of a physical model represents an approach to obtain a suitable model to be controlled. According to [31], DM is the creation across the supply chain and its markets of a coordinated flow of demand. Also, the role of DM is to decrease demand. The application of the active suspension analogy to track both problems is suitable, because the price level (which represents outputs) and demand (which represents input) converge to a stability condition that requires inventory level in the dissipation of the dynamical system. DM assumes fixed capacity (which in terms of active suspension is the mass), inventory level (damping of the system), and bidding price (restitution coefficient). In this case the idea is to present a model in which, by using DM, we can assess how demand does affect prices. We have chosen an active suspension system to approach the system model; see Figure 2.

The dynamic equations for the system are as follows.

For the capacity fixed C_f :

$$C_f \ddot{p}_f = k_f (w - p_f) - k_u (p_f - p_u) - \beta (\dot{p}_f - \dot{p}_u) - \alpha \cdot d. \quad (5)$$

TABLE 2: Demand management for manufacturing and services: literature review.

Author	Sector application	Methodology	Contribution
Bertsimas and Shioda [13]	Services	Research article (case studies)	Application of revenue management in restaurants. Using stochastic optimization via dynamic programming. Computational experiments are presented for reservation and without reservation case studies
Boyd and Bilegan [14]	Services	Research article (case study in e-commerce)	Analysis of revenue management applied to e-commerce in the airline industry. Optimization techniques are shown in inventory control mechanism
Defregger and Kuhn [15]	Manufacturing	Research article (case studies)	Application of revenue management to a make-to-order manufacturing company with limited inventory capacity. Classifying the underlying Markov decision process, by introducing a heuristic procedure
Kimes [16]	Services	Survey	Revenue management analysis via the application of optimization, forecasting, and overbooking
Kumar and Frederik [17]	Services	Survey	Application of revenue management in construction industry. The benefits of revenue management can be realized at manufacturing companies. Uncertain demand and pricing of available capacity are analyzed via this approach
Lee et al. [18]	Transportation	Research article (case studies)	Analysis and application of heuristic to solve a single revenue management problem with postponement, arising from the sea cargo industry. Optimization based case studies are addressed
Nair and Bapna [19]	Services	Research article	This paper studies optimal policies for allocating modems capacity among segments of customers using a continuous time Markov decision
Spengler et al. [20]	Manufacturing	Research article (case studies)	Revenue management approach for companies in the iron and steel industry is developed. The aim is to improve short-term order selection
Shin and Park [21]	Manufacturing	Research article (theoretical)	Yield improvement is analyzed in semiconductor manufacturing via machine learning
Steinhardt and Gönsch [22]	Services	Research article (case studies)	This paper addresses the problem of integrating revenue management capacity control with upgrade decision making. A new structural property for an integrated dynamic programming formulation is present

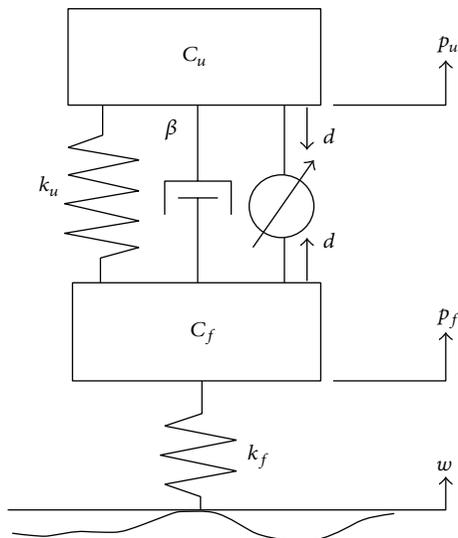


FIGURE 2: Active suspension analogy for DM-DP model.

For the capacity unfixed subject to active suspension performance C_u :

$$C_u \ddot{p}_u = k_u (p_f - p_u) + \beta (\dot{p}_f - \dot{p}_u) + (1 - \alpha) d. \quad (6)$$

Variables definition is as follows:

C_f : capacity fixed,

C_u : capacity fixed subject to active suspension,

p_f : price level for the C_f capacity fixed,

p_u : price level for the C_u subject to active suspension,

w : price level disturbance in the market,

d : demand,

k_f : bidding price restitution coefficient based on price for C_f ,

k_u : bidding price restitution coefficient based on price for C_u ,

β : damping-inventory level in the system,

α : embedding parameter $0 \leq \alpha \leq 1$.

As it is stated in DM, C_f and C_u are assumed fixed; the characteristic of this system is such that the damping coefficient is β . The nature of the input signal is the demand and the output is the price level for the capacity fixed and for the price level subject to active suspension. There exists a relation between bidding prices and restitution coefficients k_f and k_u ; this means the amount of product per unit of price. The model presented in the active suspension for demand and price modeling, from Figure 2, is based on the assumption that the analysis is for two products in the system. We plan to extend our model to a half-car active suspension with balanced capacity and unbalanced capacity for four products. Finally, in future work, we intend to extend it to a full-car model for a multiproduct system and analyze if the system is balanced (with the same amount of capacity in each active suspension).

3.2. Robust Policies in DM: A Sensitivity Analysis. The following approach can be used for the design of robust controllers, as suggested in [32], which analyzed the robustness for mass-spring system without damping. Consider a representation of an uncertain linear dynamical system in state-space form:

$$\dot{z} = A(q)z + B(q)u, \quad (7)$$

where $z \in \mathfrak{R}^n$, $u \in \mathfrak{R}^m$, and q is a vector of uncertain parameters. The sensitivity of the states with respect to parameter q is

$$\frac{dz}{dq_i} = \frac{\partial A}{\partial q_i} z + A \frac{\partial z}{\partial q_i} + \frac{\partial B}{\partial q_i} u. \quad (8)$$

Consider the quarter active suspension system; from Figure 2, the equations of motion, assuming that the uncertain parameter is the damping coefficient (β), and the partial derivative of (5) and (6) with respect to β are

$$\frac{\partial \ddot{p}_f}{\partial \beta} = \frac{1}{C_f} \left[(\dot{p}_u - \dot{p}_f) + \beta \left(\frac{\partial \dot{p}_u}{\partial \beta} - \frac{\partial \dot{p}_f}{\partial \beta} \right) \right], \quad (9)$$

$$\frac{\partial \ddot{p}_u}{\partial \beta} = -\frac{1}{C_u} \left[(\dot{p}_u - \dot{p}_f) - \beta \left(\frac{\partial \dot{p}_f}{\partial \beta} - \frac{\partial \dot{p}_u}{\partial \beta} \right) \right]. \quad (10)$$

Both equations are not independent of one another; the sensitivity states are expressed as

$$\frac{\partial \ddot{p}_f}{\partial \beta} = -\frac{C_u}{C_f} \left(\frac{\partial \ddot{p}_u}{\partial \beta} \right). \quad (11)$$

The following boundary conditions increase the robustness:

$$\frac{\partial p_f}{\partial \beta} = \frac{\partial p_u}{\partial \beta} = \frac{\partial \dot{p}_f}{\partial \beta} = \frac{\partial \dot{p}_u}{\partial \beta} = 0. \quad (12)$$

Integrating (11) in time and subject to boundary conditions (12), finally the relationship is such that

$$\frac{\partial \dot{p}_f}{\partial \beta} = -\frac{C_u}{C_f} \left(\frac{\partial \dot{p}_u}{\partial \beta} \right). \quad (13)$$

Substituting (13) into (10) permits achieving the general state sensitivity equation of the form:

$$\frac{\partial \ddot{p}_u}{\partial \beta} = -\frac{1}{C_u} (\dot{p}_u - \dot{p}_f) + \left(\frac{\beta}{C_f} + \frac{\beta}{C_u} \right) \left(\frac{\partial \dot{p}_u}{\partial \beta} \right). \quad (14)$$

The sensitivity analysis presented in Section 3.2 considers variations in the inventory level, which gives a capacity equivalent, $C_{eq} = C_u C_f / (C_u + C_f)$, to be used in the dynamical system for modeling consideration and it refers to robust policies. We conclude that a robust inventory policy is represented by the following equation: $\beta / C_{eq} = \beta((1/C_f) + (1/C_u))$.

3.3. Stability Analysis for DM. For the behavior of the damping effect in DM, this section proposes the stability criteria via Lyapunov direct method.

Definition 1. A linear system is stabilizable if all unstable modes are controllable.

DM-ODE equation (2) is necessary to analyze the stability performance. Assume the DM-ODE equation has the structure $\dot{x} = Ax + Bu$.

Theorem 2. For the dynamical system (2), and considering that (A, B) are controllable, a necessary and sufficient condition for stability in DM is $\beta > 0$.

Proof. The following condition in the demand of the system is assumed:

$$d(t) = 0. \quad (15)$$

Equation (2) can be formulated as

$$C_{eq} \ddot{p} + \beta \dot{p} + kp = 0. \quad (16)$$

From (16) we use energy as the Lyapunov function:

$$V(p, \dot{p}) = \frac{1}{2} C_{eq} \dot{p}^2 + \frac{1}{2} kp^2, \quad (17)$$

$$\frac{dV}{dt} = \frac{\partial V}{\partial p} \frac{dp}{dt} + \frac{\partial V}{\partial \dot{p}} \frac{d\dot{p}}{dt}, \quad (18)$$

$$\frac{dV}{dt} = kp \cdot \dot{p} + C_{eq} \dot{p} \cdot \ddot{p}, \quad (19)$$

$$\frac{dV}{dt} = kp \cdot \dot{p} + C_{eq} \cdot \dot{p} \left(\frac{-\beta}{C_{eq}} \dot{p} - \frac{k}{C_{eq}} p \right). \quad (20)$$

Eliminating terms from (20), the following condition is achieved: $-\beta \cdot \dot{p}^2 \leq 0$. Finally, the presence of inventory level in the system $\beta > 0$ is a necessary and sufficient condition for stability purposes. \square

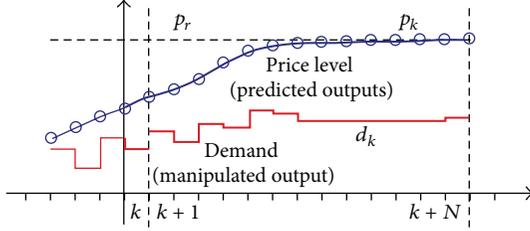


FIGURE 3: Schematic receding horizon control philosophy.

4. Model Predictive Control Analysis for Demand Management

4.1. Model Predictive Control: Towards Optimal Policies in DM. Model predictive control (MPC) is a real time optimal control strategy that has been applied in process control, aerospace, automotive, management science, and robotic applications. Besides these applications, MPC has other advantages such as good tracking performance, physical constraints handling, and extension to nonlinear systems [33]. MPC has been used to address production-inventory systems. For example, in adaptive MPC, the adapted model along with a smoothed estimation of the future customer demand is used to predict inventory levels over the optimization horizon [34]. MPC policy shows improved performance, greater flexibility, and higher functionality relative to an advanced order-up-to policy based on control engineering principles found in the literature [35]. The philosophy of MPC, also known as receding horizon control (RHC), has the advantage of handling control and state constraints [36]. The RHC strategy takes into account an objective function and many constraints; see Figure 3. In RHC to define an optimal strategy, the first step requires calculating a control law and repeating this process indefinitely, each sampling time.

With MPC, an optimization problem is solved at each time step to determine a plan of action over a fixed time horizon [37]. MPC is a nonlinear control policy that handles input and output constraints, as well as various control objectives. Using MPC, a system can be controlled near its physical limits, often outperforming linear control. It is well known that the computation of predictive control laws is a crucial task in every MPC application to systems with fast dynamics due to the fact that an optimization problem has to be solved online [38]. Other approaches, applying MPC techniques, have been developed such as simplified MPC algorithm for Markov jump systems [39] and constrained robust MPC effective for uncertain Markov jump systems, as previously shown [40].

For the active suspension analogy to DM-DP, we will use a linear MPC for purposes of analysis and control. Based on this, we are interested in optimal control problems of the form

$$\min_{d_k, \dots, d_{k+N-1}} J = \sum_{k=0}^{N-1} \|p_k - p_r\|^2 + \|d_k - d_r\|^2, \quad (21a)$$

TABLE 3: MPC design parameters for optimal and robust policies.

Policy case	T_s	N_p	N_u
Optimal policies	0.05	10	2
Robust policies	0.03	10	2

$$\begin{aligned} \text{s.t.} \quad & x_{k+1} = A \cdot x_k + B \cdot d_k, \\ & p_k = C \cdot x_k + D \cdot d_k, \\ & d_{\min} \leq d_k \leq d_{\max}, \\ & p_{\min} \leq p_k \leq p_{\max}, \\ & x_k = x(k), \quad k = 0, \dots, N-1. \end{aligned} \quad (21b)$$

Here $x : \mathfrak{R} \rightarrow \mathfrak{R}^n$ denotes the state, $d : \mathfrak{R} \rightarrow \mathfrak{R}^n$ the control input, and $p : \mathfrak{R} \rightarrow \mathfrak{R}^n$ the output of the system.

The conventional linear MPC law is based on the following algorithm.

Algorithm (linear MPC)

- (1) Achieve the new state p_k .
- (2) Solve the optimization problem (21a) and (21b).
- (3) Calculate the law $d(k) = d(k+0 | k)$.
- (4) $k \leftarrow k+1$. Go to (1).

Computational Complexity Analysis. The complexity of the solver for the optimization problem (21a) and (21b) depends on the choice of the performance index, the model, and constraints. For our algorithm the optimization problem is a quadratic program (QP).

Based on the nature of the proposed algorithm, which is a QP related to an active set method, each iteration has cost $O(n^2)$ floating point operations (flops), where n is the number of decision variables and n is proportional to horizon N_p . It is important to notice that active set methods are exponential in the worst case but show good practical performance. Also, active set methods work best for small and medium size problems.

4.2. Simulations Results. After the presented analysis two scenario simulations are shown for optimal policies in Figures 4 and 5. The robust policies for DM are presented in Figures 6 and 7. The main goal is to contribute in the performance of MPC for optimal policies and robust policies in DM and to compare it with linear quadratic regulator (LQR). The application of tracking MPC in DM approaches achieves a suitable response for optimal policies. In Figure 6 price level is shown taking into account capacity fixed and unfixed responses. It is important to notice that the demand response, in Figure 7, is constrained.

MPC design parameters for optimal and robust policies are present in Table 3. It is important to note that T_s is the sampling time, N_p is the prediction horizon, and N_u is the control horizon, where $N_u < N_p$.

In Figure 4, the MPC simulation reflects under optimal policies a performance with small overshoot in prices and

TABLE 4: Design parameter for DM-active suspension dynamical system.

Policy case	C_u	C_f	k_u	k_f	β
Optimal policies	300	50	2000	2000	500
Robust policies	500	300	10000	10000	1000

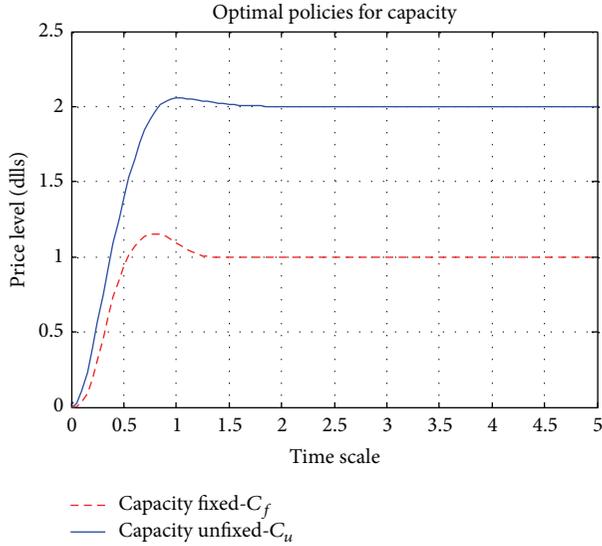


FIGURE 4: Simulation response for optimal policy price level-tracking MPC.

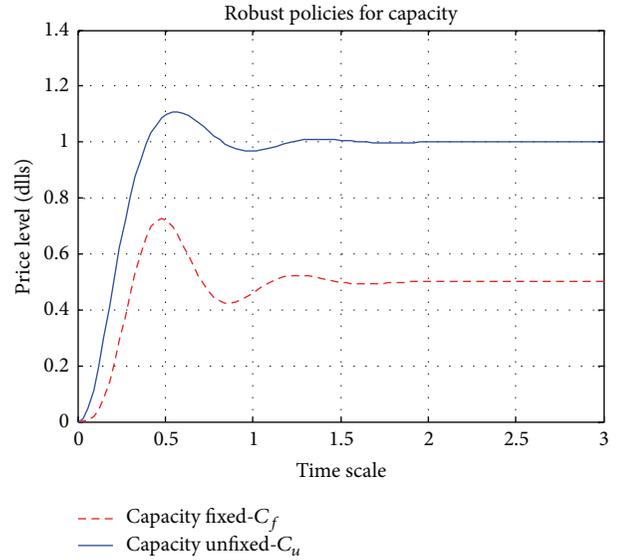


FIGURE 6: Simulation response for robust policy price level-tracking MPC.

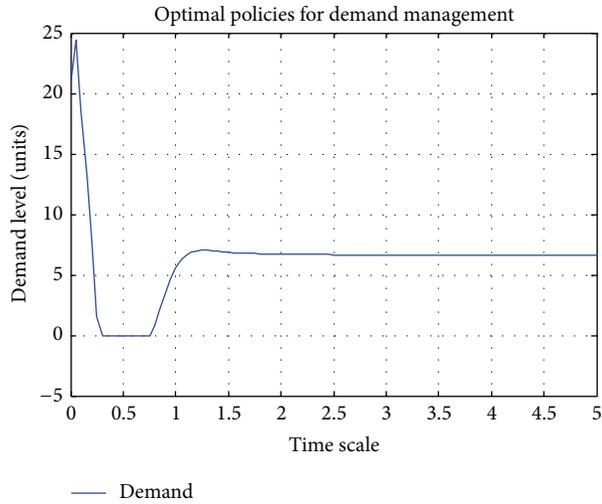


FIGURE 5: Optimal policies for demand management.

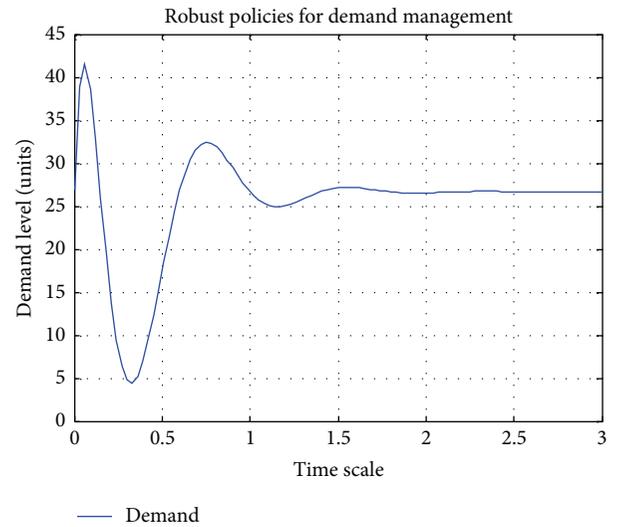


FIGURE 7: Robust policies for demand management.

considerable fluctuations in demand. It is important to remember that we have the price level as output for the dynamical system and demand as input.

The MPC simulation shows in Figure 6 the performance in the application of the robust policies obtained from the analysis showed in Section 3.2. It is important to note that the price level in both responses is reduced by half compared with scenario 1 from Figure 10. This reduction is based on the notion that there is a relation between inventory level and

capacity equivalent for the robust policy. Under variations in this relation, the price level is reduced and the demand presents small variations.

Design parameters for DM-active suspension are shown in Table 4, for optimal and robust policies. It is important to note that these parameters are the same over both policies for purposes of control via MPC and LQR.

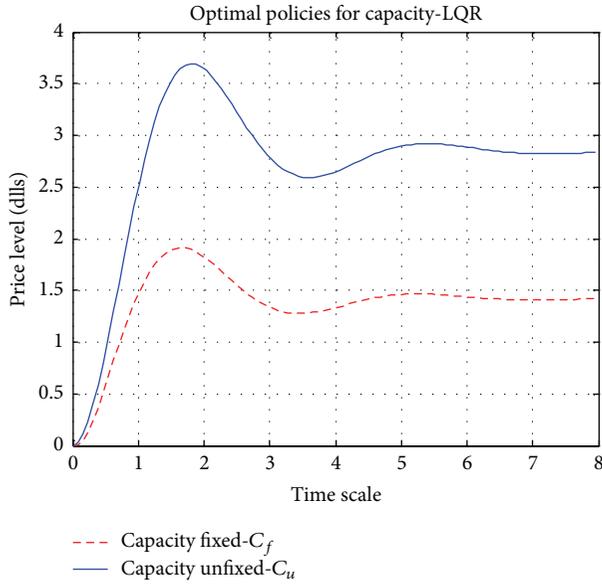


FIGURE 8: Optimal policies for capacity via LQR.

In summary, from demand curve in economics, which establish that if price is lower consumers are ready to buy more, based on this context, with an optimal policy, the prices are higher and demand is low. However for robust policies the prices are lower than in optimal policies and the demand is higher. In this case the relation between the price and the demand is satisfied from economic theory.

Linear Quadratic Regulator. Consider that MPC is an online solution of a LQR optimization problem. The approach here is to introduce the LQR methodology and compare it with the MPC results. In the state-feedback version of the LQR problem we assume that the whole state x can be measured and therefore it is available to control [41]. The state-feedback control law $u(t) = -Kx(t)$ minimizes the cost function:

$$J_{\text{LQR}} = \int_0^{\infty} (x^T Q x + u^T R u) dt, \quad (22)$$

where $K = R^{-1} B^T P$, and P is solved by an algebraic Riccati equation. The results achieved by the application of LQR in the optimal and robust policies are presented in Figures 8 and 9.

Once LQR is applied for optimal and robust policies, simulations present more oscillations when compared to MPC, using the same parameters from Table 4. In Figure 8, for optimal policies the price level achieves more overshoot and a larger setting time. Also in Figure 9 the demand level presents higher variation than in MPC optimal policies. The demand level is reduced at the end of time horizon, as it has been proposed in DM theory.

Robust policies achieve an oscillatory performance in the price level as it is noted in Figure 10; however, the price level is reduced considerably when compared to optimal policies via LQR. In Figure 11, demand level achieves a low value in units at the end of time horizon, which is expected, but the oscillation is persistent.

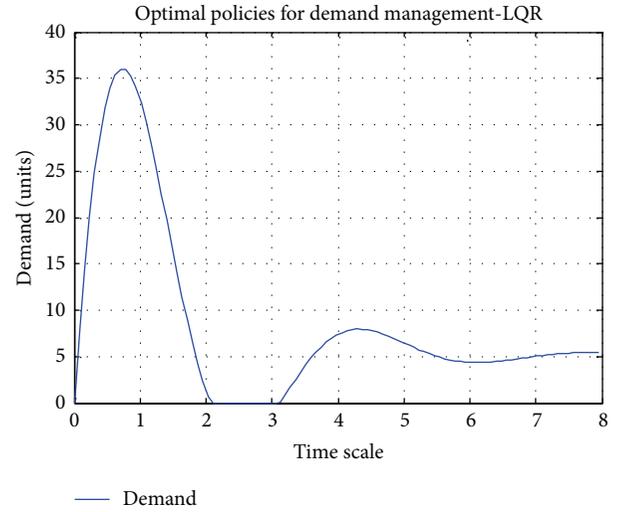


FIGURE 9: Optimal policies for demand management via LQR.

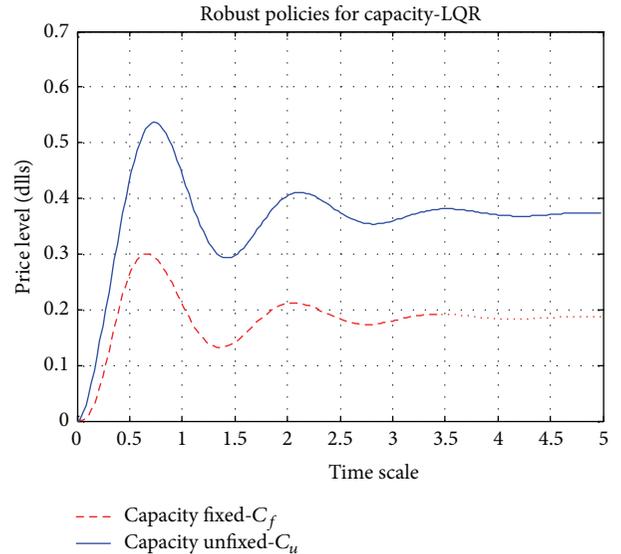


FIGURE 10: Robust policies for capacity via LQR.

In summary, MPC simulations from Figures 4 to 7 present better performance than LQR which are shown in Figures 8 to 11, for optimal and robust policies.

Based on this context and these results, we aim to explore other control oriented approaches for DM-active suspension analogy via dynamic pricing such as robust control techniques, as developed in [42, 43] with the goal to compare its performance with MPC techniques. Also we plan to explore adaptive control techniques [44], output feedback control [45], and sample data control present [46] and to develop a trade-off with optimal control techniques.

5. Conclusions

This research work presents a novel DM-DP approach based on MPC for second order systems with damping, which

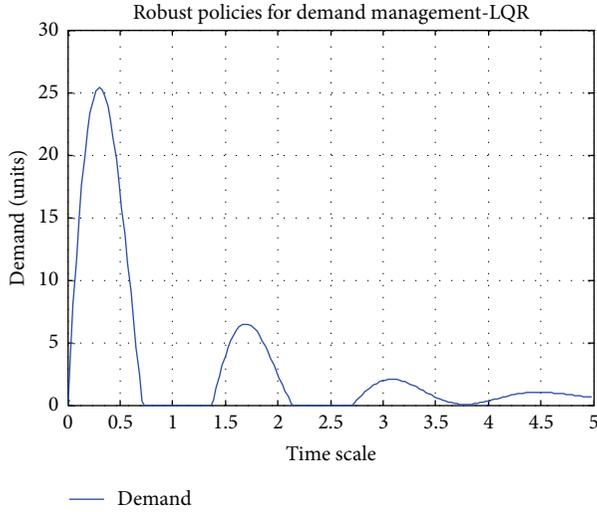


FIGURE 11: Robust policies for demand management via LQR.

in terms of DM tell us that dissipation process is present with inventory storage level in the system. Our results of the stability analysis from Lyapunov direct method show that an inventory level is needed in the system. The one-quarter active suspension model presents the analogy for a DM-DP process system behavior. Besides the sensitivity analysis shows us that an equivalent capacity level will provide reduced price level. Simulations results show that under robust policies the price level is reduced by half; this is because of the ratio between inventory level and capacity in the active suspension dynamical system. For future work, the intention is to develop a full active suspension model for multiple product system as well as to achieve inventory level dynamics via first order ordinary differential equations. Another future research area will focus on the exploration of a full active suspension model for one or multiple products in a multistage supply chain.

Appendices

A. Inventory Level Derivation

The derivation of the value of β requires the following analysis:

$$\text{Let } \beta(t) = \int_0^{\infty} e^{-t/\beta} dt, \quad \text{Finally } \beta(t) = \beta. \quad (\text{A.1})$$

B. Infinite Dimensional Analysis for Demand Management

We propose the following generalized wave equation which describes the behavior of flexible systems, presented first in [30]:

$$m(x) \frac{\partial^2 u(x,t)}{\partial t^2} + \beta \frac{\partial u(x,t)}{\partial t} + Ku(x,t) = F(x,t). \quad (\text{B.1})$$

The purpose is to obtain a PDE description for demand management via dynamic pricing. Based on this

$$C(p) \frac{\partial^2 \varphi(p,t)}{\partial t^2} + \beta \left[\frac{\partial \varphi(p,t)}{\partial p} + \alpha \frac{\partial \varphi(p,t)}{\partial t} \right] + k\varphi(p,t) = D(p,t), \quad (\text{B.2})$$

where

$$\frac{\partial \varphi}{\partial t} = \frac{\partial \varphi}{\partial p} \frac{\partial p}{\partial t} = v_p \frac{\partial \varphi}{\partial p}, \quad (\text{B.3})$$

substituting (B.3) in (B.2) which gives

$$C(p) \left[\frac{\partial}{\partial t} \left(v_p \frac{\partial \varphi(p,t)}{\partial p} \right) \right] + \beta \frac{\partial \varphi(p,t)}{\partial p} + \alpha \beta \frac{\partial \varphi(p,t)}{\partial t} + K\varphi(p,t) = D(p,t). \quad (\text{B.4})$$

Developing partial derivatives:

$$C(p) \left(v_p \frac{\partial^2 \varphi(p,t)}{\partial p \partial t} + a_p \frac{\partial \varphi(p,t)}{\partial p} \right) + \beta \frac{\partial \varphi(p,t)}{\partial p} + \alpha \beta \frac{\partial \varphi(p,t)}{\partial t} + K\varphi(p,t) = D(p,t). \quad (\text{B.5})$$

Define $C(p) \equiv 1$ and $0 \leq \alpha \leq 1$.

The PDE description for demand management is

$$v_p \frac{\partial^2 \varphi(p,t)}{\partial p \partial t} + (\beta + a_p) \frac{\partial \varphi(p,t)}{\partial p} + \alpha \beta \frac{\partial \varphi(p,t)}{\partial t} + K\varphi(p,t) = D(p,t), \quad (\text{B.6})$$

where $\beta + a_p > 0$.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to thank Professor Leopoldo E. Cárdenas-Barrón for his exhaustive reviews and constructive comments, which have helped to improve the contents of this paper. This work was supported in part by the Research Chair e-Robots and the Laboratorio Nacional del Noreste y Centro de Mexico supported by Conacyt and Tecnológico de Monterrey.

References

- [1] K. L. Croxton, D. M. Lambert, S. J. García-Dastugue, and D. S. Rogers, "The demand management process," *International Journal of Logistics Management*, vol. 13, no. 2, pp. 51–66, 2002.

- [2] F.-S. Chou and M. Parlar, "Optimal control of a revenue management system with dynamic pricing facing linear demand," *Optimal Control Applications & Methods*, vol. 27, no. 6, pp. 323–343, 2006.
- [3] M. Truffelli, "Dynamic pricing: new game, new rules, new mindset," *Journal of Revenue and Pricing Management*, vol. 5, no. 1, pp. 81–82, 2006.
- [4] H. Abdel Aziz, M. Saleh, M. H. Ramsy, and H. ElShishiny, "Dynamic room pricing model for hotel revenue management systems," *Egyptian Informatics Journal*, vol. 12, pp. 177–183, 2011.
- [5] S. de Boer, "The impact of dynamic capacity management on airline seat inventory control," *Journal of Revenue and Pricing Management*, vol. 2, no. 4, pp. 315–330, 2004.
- [6] Y.-F. Hung and C.-H. Chen, "An effective dynamic decision policy for the revenue management of an airline flight," *International Journal of Production Economics*, vol. 144, no. 2, pp. 440–450, 2013.
- [7] A. Gosavi, E. Ozkaya, and A. F. Kahraman, "Simulation optimization for revenue management of airlines with cancellations and overbooking," *OR Spectrum*, vol. 29, no. 1, pp. 21–38, 2007.
- [8] M. Graf and A. Kimms, "Transfer price optimization for option-based airline alliance revenue management," *International Journal of Production Economics*, vol. 145, no. 1, pp. 281–293, 2013.
- [9] A. Kimms and M. Müller-Bungart, "Simulation of stochastic demand data streams for network revenue management problems," *OR Spectrum*, vol. 29, no. 1, pp. 5–20, 2007.
- [10] S. Netessine and R. Shumsky, "Introduction to the theory and practice of yield management," *INFORMS Transactions on Education*, no. 3 1, pp. 34–44, 2002.
- [11] B. Rannou and D. Melli, "Measuring the impact of revenue management," *Journal of Revenue and Pricing Management*, vol. 2, no. 3, pp. 261–270, 2003.
- [12] D. Walczak and S. Brumelle, "Semi-Markov information model for revenue management and dynamic pricing," *OR Spectrum*, vol. 29, no. 1, pp. 61–83, 2007.
- [13] D. Bertsimas and R. Shioda, "Restaurant revenue management," *Operations Research*, vol. 51, no. 3, pp. 472–486, 2003.
- [14] E. A. Boyd and I. C. Bilegan, "Revenue management and e-commerce," *Management Science*, vol. 49, no. 10, pp. 1363–1386, 2003.
- [15] F. Defregger and H. Kuhn, "Revenue management for a make-to-order company with limited inventory capacity," *OR Spectrum*, vol. 29, no. 1, pp. 137–156, 2007.
- [16] S. E. Kimes, "Revenue management: a retrospective," *Cornell Hotel and Restaurant Administration Quarterly*, vol. 44, no. 5-6, pp. 131–138, 2003.
- [17] S. Kumar and J. L. Frederik, "Revenue management for a home construction products manufacturer," *Journal of Revenue and Pricing Management*, vol. 5, no. 4, pp. 256–270, 2007.
- [18] L. H. Lee, E. P. Chew, and M. S. Sim, "A heuristic to solve a sea cargo revenue management problem," *OR Spectrum*, vol. 29, no. 1, pp. 123–136, 2007.
- [19] S. K. Nair and R. Bapna, "An application of yield management for internet service providers," *Naval Research Logistics*, vol. 48, no. 5, pp. 348–362, 2001.
- [20] T. Spengler, S. Rehkopf, and T. Volling, "Revenue management in make-to-order manufacturing—an application to the iron and steel industry," *OR Spectrum*, vol. 29, no. 1, pp. 157–171, 2007.
- [21] C. K. Shin and S. C. Park, "A machine learning approach to yield management in semiconductor manufacturing," *International Journal of Production Research*, vol. 38, no. 17, pp. 4261–4271, 2000.
- [22] C. Steinhardt and J. Gönsch, "Integrated revenue management approaches for capacity control with planned upgrades," *European Journal of Operational Research*, vol. 223, no. 2, pp. 380–391, 2012.
- [23] Y. A. Davizón, J. J. Lozoya, and R. Soto, "Demand management in semiconductor manufacturing: a dynamic pricing approach based on fast model predictive control," in *Proceedings of the 7th IEEE Electronics, Robotics and Automotive Mechanics Conference (CERMA '10)*, pp. 534–539, October 2010.
- [24] P. L. Sacco, "Optimal control of input-output systems via demand management," *Economics Letters*, vol. 39, no. 3, pp. 269–273, 1992.
- [25] D. N. Burghes and A. D. Wood, *Mathematical Models in the Social, Management and Life Sciences*, Ellis Horwood, 1980, Mathematics and its Applications.
- [26] E. A. Ibijola and A. A. Obayomi, "Non-standard method for the solution of ordinary differential equations arising from pricing policy for optimal inventory level," *Canadian Journal on Computing in Mathematics, Natural Sciences, Engineering and Medicine*, vol. 3, no. 7, 2012.
- [27] W. Pan, L. Yu, S. Wang, G. Hua, G. Xie, and J. Zhang, "Dynamic pricing strategy of provider with different QoS levels in web service," *Journal of Networks*, vol. 4, no. 4, pp. 228–235, 2009.
- [28] A. G. Vlachos and P. N. Biskas, "Demand response in a real-time balancing market clearing with pay-as-bid pricing," *IEEE Transactions on Smart Grid*, vol. 4, no. 4, pp. 1966–1975, 2013.
- [29] V. Jayaraman and T. Baker, "The internet as an enabler for dynamic pricing of goods," *IEEE Transactions on Engineering Management*, vol. 50, no. 4, pp. 470–477, 2003.
- [30] M. J. Balas, "Feedback control of flexible systems," *IEEE Transactions on Automatic Control*, vol. AC-23, no. 4, pp. 673–679, 1978.
- [31] J. T. Mentzer, M. A. Moon, D. Estampe, and G. W. Margolis, *Demand Management: Handbook of Global Supply Chain Management*, SAGE, 2006.
- [32] M. Vossler and T. Singh, "Minimax state/residual-energy constrained shapers for flexible structures: linear programming approach," *Journal of Dynamic Systems, Measurement and Control*, vol. 132, no. 3, pp. 1–8, 2010.
- [33] X.-M. Tang, H.-C. Qu, H.-F. Xie, and P. Wang, "Model predictive control of linear systems over networks with state and input quantizations," *Mathematical Problems in Engineering*, vol. 2013, Article ID 492804, 8 pages, 2013.
- [34] E. Aggelogiannaki, P. Doganis, and H. Sarimveis, "An adaptive model predictive control configuration for production-inventory systems," *International Journal of Production Economics*, vol. 114, no. 1, pp. 165–178, 2008.
- [35] J. D. Schwartz and D. E. Rivera, "A process control approach to tactical inventory management in production-inventory systems," *International Journal of Production Economics*, vol. 125, no. 1, pp. 111–124, 2010.
- [36] Y.-F. Wang, F.-C. Sun, Y.-A. Zhang, H.-P. Liu, and H. Min, "Getting a suitable terminal cost and maximizing the terminal region for MPC," *Mathematical Problems in Engineering*, vol. 2010, Article ID 853679, 16 pages, 2010.
- [37] J. Mattingley, Y. Wang, and S. Boyd, "Receding horizon control: automatic generation of high-speed solvers," *IEEE Control Systems Magazine*, vol. 31, no. 3, pp. 52–65, 2011.

- [38] Y. Wang and S. Boyd, "Fast model predictive control using online optimization," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 2, pp. 267–278, 2010.
- [39] Y. Yin, Y. Liu, and H. R. Karimi, "A simplified predictive control of constrained Markov jump system with mixed uncertainties," *Abstract and Applied Analysis*, vol. 2014, Article ID 475808, 7 pages, 2014.
- [40] T. Yang and H. R. Karimi, "LMI-based model predictive control for a class of constrained uncertain fuzzy Markov jump systems," *Mathematical Problems in Engineering*, vol. 2013, Article ID 963089, 13 pages, 2013.
- [41] B. Yang, Y. He, J. Han, and G. Liu, "Rotor-flying manipulator: modeling, analysis, and control," *Mathematical Problems in Engineering*, vol. 2014, Article ID 492965, 13 pages, 2014.
- [42] H. R. Karimi and H. Gao, "New delay-dependent exponential H_∞ synchronization for uncertain neural networks with mixed time delays," *IEEE Transactions on Systems, Man, and Cybernetics B: Cybernetics*, vol. 40, no. 1, pp. 173–185, 2010.
- [43] H. R. Karimi, "Robust delay-dependent H_∞ control of uncertain time-delay systems with mixed neutral, discrete, and distributed time-delays and Markovian switching parameters," *IEEE Transactions on Circuits and Systems I*, vol. 58, no. 8, pp. 1910–1923, 2011.
- [44] H. Li, J. Yu, C. Hilton, and H. Liu, "Adaptive sliding-mode control for nonlinear active suspension vehicle systems using T-S fuzzy approach," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 8, pp. 3328–3338, 2013.
- [45] H. Li, X. Jing, and H. R. Karimi, "Output-feedback-based H_∞ control for vehicle suspension systems with control delay," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 1, pp. 436–446, 2014.
- [46] H. Li, X. Jing, H.-K. Lam, and P. Shi, "Fuzzy sampled-data control for uncertain vehicle suspension systems," *IEEE Transactions on Cybernetics*, vol. 44, no. 7, pp. 1111–1126, 2014.

Research Article

Measuring Semantic and Structural Information for Data Oriented Workflow Retrieval with Cost Constraints

Yinglong Ma, Liying Zhang, and Moyi Shi

School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China

Correspondence should be addressed to Yinglong Ma; gtmikema@gmail.com

Received 20 February 2014; Revised 13 April 2014; Accepted 28 April 2014; Published 23 June 2014

Academic Editor: M. Chadli

Copyright © 2014 Yinglong Ma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The reuse of data oriented workflows (DOWs) can reduce the cost of workflow system development and control the risk of project failure and therefore is crucial for accelerating the automation of business processes. Reusing workflows can be achieved by measuring the similarity among candidate workflows and selecting the workflow satisfying requirements of users from them. However, due to DOWs being often developed based on an open, distributed, and heterogeneous environment, different users often can impose diverse cost constraints on data oriented workflows. This makes the reuse of DOWs challenging. There is no clear solution for retrieving DOWs with cost constraints. In this paper, we present a novel graph based model of DOWs with cost constraints, called constrained data oriented workflow (CDW), which can express cost constraints that users are often concerned about. An approach is proposed for retrieving CDWs, which seamlessly combines semantic and structural information of CDWs. A distance measure based on matrix theory is adopted to seamlessly combine semantic and structural similarities of CDWs for selecting and reusing them. Finally, the related experiments are made to show the effectiveness and efficiency of our approach.

1. Introduction

Data oriented workflows (DOWs) nowadays have been adopted in diverse areas such as high quality computation [1–4] and supply chain process [5, 6]. The reuse of DOWs can reduce the cost of workflow system development and control the risk of project failure and is crucial for accelerating the automation of business processes. For example, managers want to find and customize a logistic workflow such that customers can quickly perform the tasks of inquiry/quotation, order conformation, payment, and delivery within the least cost. Because each step in a logistic workflow system has to spend some cost to perform, managers are concerned about how to make the least cost of the whole logistic workflow by selecting suitable tasks from those developed logistic workflows and assembling a new one instead of developing a new workflow. The reuse of workflows with cost constraints is a challenging problem.

Reusing workflows can be achieved by measuring the similarity among candidate workflows and selecting the workflow satisfying requirements of users from them. Most existing approaches of retrieving DOWs are made based

on their structures [7–9] because of most real-life workflows such as [10]. However, the structure based approaches often concentrate on data flows and invocation relations between services, but they neglect the semantic information that services and data rely on. As a result, the accuracy of retrieving DOWs is low, which is hard to satisfy the users' requirements indeed. Semantic based approaches make full use of semantic information of services and data in workflows and retrieve workflows by comparing them from the perspective of semantics [8], where ontologies are used to represent semantic information of services and data [11, 12]. These approaches have illustrated their potential in reusing DOWs.

However, due to DOWs being often developed based on an open, distributed, and heterogeneous environment, different users often can impose diverse cost constraints on data oriented workflows. This makes the reuse of DOWs challenging. There is no clear solution for retrieving DOWs with diverse cost constraints. On one hand, there is no formal representation model that not only represents both semantic and structure information within DOWs, but also encodes diverse cost constraints that are imposed on DOWs by

different users. The vendors of DOWs cannot represent and publish more semantic enriched information for their DOWs that will be reused by users. This will straightforwardly bring about the poor accuracy in retrieving DOWs. On the other hand, we lack a feasible method to seamlessly combine the semantic and structure based approaches for retrieving and reusing DOWs. This makes the retrieval results full of uncertainty. These problems will impede the more accurate DOW retrieval and therefore have a significance to the reuse of DOWs.

In this paper, we present a novel graph based model of DOWs with diverse cost constraints, called constrained data oriented workflow (CDW), which can express cost constraints that users are often concerned about. An approach is proposed for retrieving CDWs, which seamlessly combines semantic and structural information of workflows by computing the similarities between CDWs. A distance measure based on matrix theory is adopted to seamlessly combine semantic and structural similarities of CDWs for selecting and reusing them. Finally, the related experiments are made to show the effectiveness and efficiency of our approach.

This paper is organized as follows. Section 2 is the related work of the development for data oriented workflow retrieval. An overview of our work will be discussed in Section 3. Section 4 is some basic notations about constrained data oriented workflows. Semantic similarity computation of task nodes will be discussed in Section 5. In Section 6, we discuss CDW structure based similarity and comparison. Section 7 is the experiments and evaluation with other methods. Section 8 is the conclusion and future work.

2. Related Work

The main actuating force of workflow retrieval is naturally derived from the nonlinearities and uncertainties of planning and modeling of real world applications, which have been recognized and researched in some interesting work [13–18]. We also need to consider such nonlinearities and uncertainties in business process management. Traditional workflow representations mainly focus on activities/tasks and control oriented flow [19, 20]. Dataflow is not paid more attention to because traditional workflows are executed on a closed environment within a specific corporation rather than an open one. Some methods consider the semantic information but ignore the constraint which can reflect the quality of services.

Many workflow systems have emerged based on the recent research within semantic community [21–24]. Bergmann and Gil proposed a novel representation and retrieval method based on graph [20]. They extended a traditional workflow into a new representation whose nodes had three types: task node, dataflow node, and semantic node represented by a RDF file. However, specific applications or services cannot be provided by one single corporation due to the open, distributed, and heterogeneous execution environment. They need the cooperation of different kinds of service providers. Most of them did not consider constraint information which reflects requirements of specific users

[25]. Some constraints such as time and cost, are important to reflect the quality of services provided by workflows. They cannot be defined and formulated in advance. Internal dependencies between semantic information residing in tasks/services such as hierarchical relationship and primary and secondary relationship must also be considered. Seamless integration of semantic and structure information is necessary to represent, execute, and reuse workflows because of the open, heterogeneous, and distributed environment on the Web.

Graph matching plays a key role to measure the similarity of two workflow models. It is a popular and mutual research topic. There are two classes of graph matching. One is the exact graph matching which includes graph isomorphism and subgraph isomorphism. The other is the inexact graph matching which includes attributed graph matching and attributed subgraph matching. Time complexity is a difficult problem in real-life applications that should be considered. From the perspective of implementing algorithms, graph matching can be also classified into three classes: graph isomorphism, feature extraction, and iterative methods. Graph isomorphism is a common approach introduced in [26]. Feature extraction [27] uses the idea that certain properties might be shared by similar graphs. This method has been widely used in the applications of character recognition, fingerprint images. In the iterative method, it is assumed that the similarity of two nodes depends on the similarity of their adjacent nodes. After multiple iterations, the similarity between two graphs will be obtained. Key words instead of data and semantic information play an important role in traditional workflow retrieval. The seamless combination of structure and semantic similarities is an urgent challenge for workflow retrieval. Bergmann and Gil presented a new similarity model which could seem as an enhancement of the well-known local/global approach [28]. However, it seems that they had not considered the constraints between semantic information of nodes and data types. In essence, these approaches are control flow based rather than dataflow based. They cannot seamlessly combine semantic and structure information together for representation and similarity comparison of constrained data oriented workflows. In our paper, a holistic approach based on matrix norm is proposed to measure the similarity of two workflow models. The time complexity can be also acceptable for workflow retrieval. Furthermore, the increasing speed of execution time is no faster than the growing speed of graph size. Seamless integration of semantic and structure similarities leads to a high retrieval accuracy and efficiency.

3. Overview of Process for Retrieving CDWs

In the section, we give an overview of the whole procedure of our approach using a formal graph based representation model called CDW for data oriented workflows retrieval with constraints, where a CDW model seamlessly integrates the structural and semantic information, as well as the cost constraints, and so forth. We attempt to effectively and efficiently retrieve CDWs by measuring and comparing the

similarities of both the semantic and structural information between CDWs. We argue that combination of structural and semantic information within CDWs will be greatly helpful to find more suitable workflows that satisfy both the functional requirements and diverse constraints from users. The whole procedure for retrieving CDWs can be mainly divided into five steps.

Step 1. A repository SCDW of candidate constrained data oriented workflows should be constructed beforehand, denoted by $SCDW = \{CDW_1, CDW_2, CDW_3, \dots, CDW_n\}$. Each workflow in the repository can be represented by a formal representation model called CDW, which will be introduced in Section 4. Similarly, the request workflow that users require can be also represented by the CDW representation model, which is denoted by RCDW. Furthermore, we define a counter variable i , which will be used for traversing all the candidate CDWs in SCDW.

Step 2. We traverse each candidate CDW_i in SCDW and compute the semantic similarity between RCDW and CDW_i . The semantic information residing in task nodes is represented by the RDF ontology language. Each task node corresponds to a RDF file. During the semantic similarity computation for RCDW and CDW_i , the semantic similarity between task nodes within RCDW and CDW_i can be reduced to matching their similarity between their RDF files. Meanwhile, the data types of input data and output data between task nodes are also matched. For two task nodes, respectively, from RCDW and CDW_i , both the similarity for matching their RDF files and the similarity for matching their data types will be considered to compute the semantic similarity between them.

Step 3. We need to determine the identicalness between task nodes within RCDW and CDW_i . A similarity threshold value is set, which is 0.7 in this paper. That is, if the semantic similarity obtained in Step 2 between two task nodes is not less than the threshold value (i.e., 0.7), then we regard the two nodes as identical. The reason why we do that is the fact that task nodes in CDWs are very likely to be semantically heterogeneous in an open environment such as polysemy and toponymy.

Step 4. We further compute the similarity between RCDW and CDW_i by comparing their structures. Structural similarity between RCDW and CDW_i is made based on normalized matrices proposed in our previous work [7]. The normalized matrices for RCDW and CDW_i are, respectively, constructed. Then, a distance based metric is used to compute their structural similarity, that is, the distance between their normalized matrices.

Step 5. If $i \leq n$, then go to Step 2 for comparing the next CDW with RCDW. We select the candidate CDW_k as the one that is the most similar to RCDW, where k satisfies the following condition: $\text{sim}(RCDW, CDW_k) = \min\{\text{sim}(RCDW, CDW_i) \text{ for each } 1 \leq i \leq n\}$.

4. Basic Notations

4.1. Formal Representation for CDWs

Definition 1 (constrained data oriented workflow, CDW). A constrained data oriented workflow (CDW) is denoted by $CDW = (V, E, \alpha, \beta, \gamma, \lambda, \eta)$, which is a directed labeled graph.

- (i) V is a set of nodes representing tasks/services in actual applications.
- (ii) $E \subseteq V \times V$ is a set of ordered pairs of nodes, called directed edges.
- (iii) $\alpha : V \rightarrow L_V$ is a mapping function which assigns each node a label, where L_V is a set of label names for tasks.
- (iv) $\beta : V \rightarrow RF$ is a mapping function which maps each node to a RDF file, where RF is a set of RDF files that illustrate the semantic information of task nodes.
- (v) $\gamma : V \rightarrow 2^{DT}$ is a mapping function which maps each node to a subset of data types. It represents the input information which can be processed by task nodes, where DT is the set of data types.
- (vi) $\lambda : V \rightarrow 2^{DT}$ is a mapping function which maps each node to a subset of data types. It represents the output information provided by task nodes.
- (vii) $\eta : E \rightarrow L_E$ is a mapping function which assigns each edge a constraint label in set L_E .

In the representation of CDW, nodes represent different tasks or functional activities in an actual application. A service can be provided by a specific task node. Some of them integrated together can accomplish a complex and comprehensive service such as scientific computation.

Semantic information is represented using the RDF language and is defined as a property of node, which is a key feature of semantic representation of task nodes. We establish the correspondence relations between nodes and RDF files using a mapping function instead of extending each task node with a RDF graph. This method can largely eliminate the degree of redundancy of a graph and clearly reflect the main workflow execution.

Data oriented workflows are often designed based on open, distributed, and heterogeneous environment to accomplish complex data processing. Therefore, it should have a high degree of modularity. True users care much about the input and output parameters that the task node could process and provide instead of data processing details. The representation for dataflow mainly emphasizes the data types that are used to communicate information between task nodes.

Now, we will give an example of CDWs to intuitively illustrate a data oriented workflow with constraint shown in Figure 1. In this figure, there are three task nodes which provide the services of function analysis, price analysis, and audit, respectively. Task nodes are represented by rectangles. Two control flows are represented by solid arrows as two edges with the constraints [1, 2] and [0, 3]. Solid line ovals are

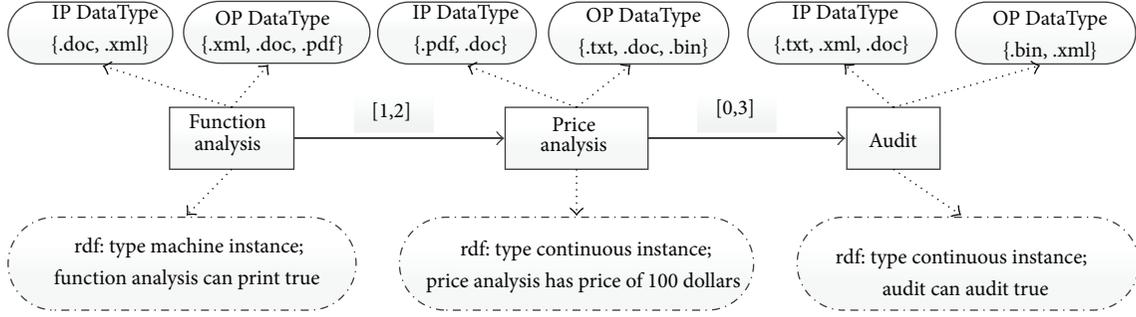


FIGURE 1: An example of CDW.

used to illustrate the input and output data types for each task node. Dotted lines are used to represent the relations between task nodes and data types properties. Dotted line ovals are used to represent corresponding RDF files of task nodes.

4.2. Definitions Related to Constraints

Definition 2 (constraint). Let $CDW = (V, E, \alpha, \beta, \gamma, \lambda, \eta)$ be a CDW. A constraint of edge $\langle v_i, v_j \rangle$ is represented by an interval of the form $[a, b]$, where $\eta(\langle v_i, v_j \rangle) = [a, b]$, and a and b are real numbers and $a \leq b$.

In actual applications, specific constraints can be determined in the domain of $[a, b]$. It reflects the variable and dynamic features of our representation for constrained data oriented workflows.

Definition 3 (intersection of constraints). Let $CDW = (V, E, \alpha, \beta, \gamma, \lambda, \eta)$ and $CDW' = (V', E', \alpha', \beta', \gamma', \lambda', \eta')$ be two CDWs. Let $\eta(\langle v_1, v_2 \rangle) = [a, b]$ and $\eta'(\langle v_3, v_4 \rangle) = [c, d]$ be two constraints for edges $\langle v_1, v_2 \rangle \in E$ and $\langle v_3, v_4 \rangle \in E'$, respectively. The intersection of constraints between the two edges can be computed by the following:

$$[a, b] \cap [c, d] = \begin{cases} [\max(a, c), \min(b, d)], & \text{if } a \leq d \wedge c \leq b, \\ \emptyset, & \text{otherwise.} \end{cases} \quad (1)$$

Definition 4 (union of constraints). Let $CDW = (V, E, \alpha, \beta, \gamma, \lambda, \eta)$ and $CDW' = (V', E', \alpha', \beta', \gamma', \lambda', \eta')$ be two CDWs. Let $\eta(\langle v_1, v_2 \rangle) = [a, b]$ and $\eta'(\langle v_3, v_4 \rangle) = [c, d]$ be two constraints for edges $\langle v_1, v_2 \rangle \in E$ and $\langle v_3, v_4 \rangle \in E'$, respectively. The union of constraints between the two edges can be computed by the following:

$$[a, b] \cup [c, d] = [\min(a, c), \max(b, d)]. \quad (2)$$

Definition 5 (duration of constraint). Let $CDW = (V, E, \alpha, \beta, \gamma, \lambda, \eta)$ be a constrained data oriented workflow and $\langle v_i, v_j \rangle \in E$. Let $\eta(\langle v_i, v_j \rangle) = [a, b]$ be the constraint for edge $\langle v_i, v_j \rangle$. The duration of this constraint is denoted by $\text{Dur}([a, b]) = b - a$.

Definition 6 (summation of constraints). Let $CDW = (V, E, \alpha, \beta, \gamma, \lambda, \eta)$ and $CDW' = (V', E', \alpha', \beta', \gamma', \lambda', \eta')$ be two

CDWs. Let $\eta(\langle v_1, v_2 \rangle) = [a, b]$ and $\eta'(\langle v_3, v_4 \rangle) = [c, d]$ be two constraints for edges $\langle v_1, v_2 \rangle \in E$ and $\langle v_3, v_4 \rangle \in E'$, respectively. The summation between their constraints can be denoted by $[a, b] + [c, d] = [a + c, b + d]$.

5. Semantic Similarity Computation

In our representation of CDW, all semantic information is represented as properties of nodes. Four mapping functions $\alpha, \beta, \gamma,$ and λ are to, respectively, associate a node v with its node name, its input set of data types, its output set of data types, and a RDF file RT_v . A RDF file represents the service function of a task node in the semantic level while a name is only a label in actual applications. The similarity between RDF files of two nodes can be used to measure the actual functional similarity in an open and heterogeneous environment. For example, we can measure the similarity between two nodes with different names that have the same function. It is also important for measuring the similarity between two task nodes to consider their data types that includes the semantic information of service data. As a consequence, the semantic similarity includes two closely related parts: the similarity for RDF files and the similarity for data types.

5.1. Similarity for RDF Files. In this paper, we use RDF to describe the semantic information of a task node, which is regarded as a property in the task node by function β . The similarity computation of RDF files can be reduced to the comparison of their corresponding RDF graphs. The matching for RDF graphs has been studied for decades, and many good methods have been proposed. Each RDF graph is composed by a set of statements (triples). Each statement consists of three elements: subject, property, and object. The similarity of two RDF graphs can be divided into three steps.

(1) We use a three-dimension vector \vec{ssv} to represent the similarity between two statements. Let statement = (s, p, o) and statement' = (s', p', o') be any two statements. The elements $s, p, o, s', p',$ and o' are the label strings of elements in statements. Vector $\vec{ssv} = (ssv_1, ssv_2, ssv_3)^T$, where $ssv_1 = \text{SimLD}(s, s'), ssv_2 = \text{SimLD}(p, p'),$ and $ssv_3 = \text{SimLD}(o, o')$.

The notation $\text{SimLD}(s_1, s_2)$ is the similarity of label strings s_1 and s_2 by using the Levenshtein distance; that is,

$$\text{SimLD}(s_1, s_2) = 1 - \frac{\text{LevD}(s_1, s_2)}{\max(|s_1|, |s_2|)}, \quad (3)$$

where $|s_1|$ and $|s_2|$, respectively, represent the length of strings s_1 and s_2 . Obviously, the value of $\text{SimLD}(s_1, s_2)$ is between 0 and 1. The notation $\text{LevD}(s_1, s_2)$ refers to the Levenshtein distance between strings s_1 and s_2 .

(2) We use a matrix ST as an auxiliary matrix for further computation of two RDF graphs: $\text{RDF}_1 = \{st_1, st_2, \dots, st_m\}$ and $\text{RDF}_2 = \{st'_1, st'_2, \dots, st'_n\}$. STU is the union of RDF_1 and RDF_2 . All elements in STU are contained in the row and column of ST . For any $1 \leq i, j \leq m+n$, the matrix $ST_{(\text{RDF}_1, \text{RDF}_2)}(i, j)$ between RDF_1 and RDF_2 is represented as follows:

$$ST_{(\text{RDF}_1, \text{RDF}_2)}(i, j) = \sqrt{\text{ssv}_{ij}^T \times \text{ssv}_{ij}}, \quad (4)$$

where ssv_{ij} is the statement similarity vector between the statement i in row and the statement j in column in the matrix ST .

(3) The similarity between RDF_1 and RDF_2 can be computed by the notation $\text{SimRG}(\text{RDF}_1, \text{RDF}_2)$, where

$$\text{SimRG}(\text{RDF}_1, \text{RDF}_2) = \sqrt{\text{tr} \left[ST_{(\text{RDF}_1, \text{RDF}_2)} \times ST_{(\text{RDF}_1, \text{RDF}_2)}^T \right]}. \quad (5)$$

Definition 7 (similarity for RDF files). Let $\text{CDW} = (V, E, \alpha, \beta, \gamma, \lambda, \eta)$ and $\text{CDW}' = (V', E', \alpha', \beta', \gamma', \lambda', \eta')$ be two CDWs. For any two nodes $v \in V$ and $v' \in V'$, the similarity between their RDF files can be computed according to the following equation:

$$\begin{aligned} \text{SimRF}(v, v') &= \text{SimRG}(\beta(v), \beta'(v')) \\ &= \sqrt{\text{tr} \left(ST_{(\beta(v), \beta'(v'))} \times ST_{(\beta(v), \beta'(v'))}^T \right)}, \end{aligned} \quad (6)$$

where $\beta(v)$ and $\beta'(v')$ are the RDF graphs for nodes v and v' , respectively, according to Definition 1.

5.2. Similarity for Input and Output Data Types. Data types, including input and output data types, indicate that information can be operated by specific task nodes. In this paper, they are represented by file names or information format such as .doc, .pdf, and .bin. Different task nodes may deal with different kinds of data types of services. Therefore, the compatibility of input and output data types of different nodes could be another factor to measure the services similarity provided by task nodes. We use $\text{ComIP}(v, v')$ and $\text{ComOP}(v, v')$ to represent the input data types compatibility and output compatibility, respectively, for nodes v and v' . $\text{SimDT}(v, v')$ is the data types similarity between nodes v and v' considering both input and output data types similarities. The definitions are as follows.

Definition 8 (compatibility for input and output data types). Let $\text{CDW} = (V, E, \alpha, \beta, \gamma, \lambda, \eta)$ and $\text{CDW}' = (V', E', \alpha', \beta', \gamma', \lambda', \eta')$ be two CDWs. For any two task nodes $v \in V$ and $v' \in V'$, the compatibility of input data types and output data types between v and v' can be computed by the following formula:

$$\begin{aligned} \text{ComIP}(v, v') &= \begin{cases} 1, & \text{if } \gamma(v) \cap \gamma(v') \neq \emptyset; \\ 0, & \text{otherwise,} \end{cases} \\ \text{ComOP}(v, v') &= \begin{cases} 1, & \text{if } \lambda(v) \cap \lambda(v') \neq \emptyset; \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \quad (7)$$

Definition 9 (similarity for input and output data types). Let $\text{CDW} = (V, E, \alpha, \beta, \gamma, \lambda, \eta)$ and $\text{CDW}' = (V', E', \alpha', \beta', \gamma', \lambda', \eta')$ be two CDWs. For any two task nodes $v \in V$ and $v' \in V'$, the similarity of data types between v and v' can be computed by the following equation:

$$\text{SimDT}(v, v') = \frac{[\text{ComIP}(v, v') + \text{ComOP}(v, v')]}{2}. \quad (8)$$

6. Similarity Computation between CDWs Based on Distance Metric

6.1. Identicalness Measure of Two Task Nodes. Semantic information similarity between two task nodes, which is measured by RDF files and data types similarities, can be used to indicate the identicalness of services provided by them. In this paper, we use an index $\text{IM}(v, v')$ to measure the identicalness of two task nodes from different constrained data oriented workflows. We set the threshold of IM 0.7. This index will be used to determine whether it is suitable to assign two task nodes the same name from two workflows for comparison as follows:

$$\text{IM}(v, v') = \text{SimRF}(v, v') \times \text{SimDT}(v, v'), \quad (9)$$

where $\text{SimRF}(v, v')$ and $\text{SimDT}(v, v')$ are, respectively, the RDF files similarity and data types similarity according to Definitions 7 and 9.

Identifying the identical task nodes between two CDWs is a preprocessing for comparing their similarity of the future. In this paper, the similarity computation of CDWs depends on their normalized matrices. We need to determine the task nodes related to the normalized matrices. Let $\text{CDW} = (V, E, \alpha, \beta, \gamma, \lambda, \eta)$ and $\text{CDW}' = (V', E', \alpha', \beta', \gamma', \lambda', \eta')$ be two CDWs.

Our preprocessing can be formally represented as follows.

- (1) For any two nodes $v \in V$ and $v' \in V'$, we use the notation $\text{Ident}(v, v')$ to represent that v and v' are identical. $\text{Ident}(v, v')$ if and only if $\text{IM}(v, v') > 0.7$ and $\text{IM}(v, v') = \max\{\text{IM}(v, v''), \text{IM}(v''', v')\}$ for all $v'' \in V'$ and $v''' \in V$. If $\text{Ident}(v, v')$, then v and v' will be assigned the same name label.
- (2) The set of task nodes related to normalized matrices is denoted by $\text{STN} = \{v_1, v_2, \dots, v_k\}$, where $v_i \in V \cup V'$ for all $1 \leq i \leq k$, and, for any $v_i, v_j \in V \cup V'$ and $i \neq j$, v_i and v_j are not identical.

6.2. Normalized Matrices and Distance Metric

Definition 10 (normalized matrices for two CDWs). Let $CDW = (V, E, \alpha, \beta, \gamma, \lambda, \eta)$ and $CDW' = (V', E', \alpha', \beta', \gamma', \lambda', \eta')$ be two CDWs. NM and NM' are, respectively, the two normalized matrices for CDW and CDW' . n is the cardinality of the set STN after preprocessing. Let $STN = \{v_1, v_2, \dots, v_k\}$, where $v_i \in V \cup V'$ for all $1 \leq i \leq k$, and, for any $v_i, v_j \in V \cup V'$ and $i \neq j$, v_i and v_j are not identical. The normalized matrices NM and NM' are computed by the formulas as follows:

$$NM(i, j) = \begin{cases} \frac{1}{n}, & \text{if } (v_i, v_j) \in E \setminus E', \\ \frac{\text{Dur}(\eta(v_i, v_j)) \cap \text{Dur}(\eta'(v_i, v_j))}{\text{Dur}(\eta(v_i, v_j))} \times \frac{1}{n}, & \text{if } (v_i, v_j) \in E \cap E', \\ 0, & \text{otherwise,} \end{cases} \quad (10)$$

$$NM'(i, j) = \begin{cases} \frac{1}{n}, & \text{if } (v_i, v_j) \in E' \setminus E, \\ \frac{\text{Dur}(\eta(v_i, v_j)) \cap \text{Dur}(\eta'(v_i, v_j))}{\text{Dur}(\eta'(v_i, v_j))} \times \frac{1}{n}, & \text{if } (v_i, v_j) \in E \cap E', \\ 0, & \text{otherwise.} \end{cases} \quad (11)$$

Definition 11 (distance metric [28]). Based on Definition 10, one uses a distance metric proposed in [28] for measuring the similarity between two constrained data oriented workflows CDW and CDW' whose normalized matrices are, respectively, NM and NM' . Their distance metric between NM and NM' is defined as follows:

$$DM(NM, NM') = \sqrt{\text{tr}[(NM - NM')^T \times (NM - NM')]} \quad (12)$$

The notation $\text{tr}[\bullet]$ means the trace of matrix \bullet , which is the sum of elements in the main diagonal of a matrix.

The distance metric DM satisfies all the three properties of distance measure as follows:

- (1) $DM(NM, NM') \geq 0$ if and only if $NM = NM'$;
- (2) $DM(NM, NM') = DM(NM', NM)$;
- (3) $(DM(NM, NM'') + DM(NM'', NM')) \geq DM(NM, NM')$.

What is worth noting is that $DM(NM, NM')$, *de facto*, is the dissimilarity of NM and NM' . The larger the DM value between two CDWs is, the more dissimilar they are.

7. Experiments and Evaluation

7.1. Evaluation Indices, Workflow Repository, and Benchmark Methods. In this paper, we will evaluate the effectiveness and efficiency of our method by comparing it with the three methods. We use the two indices: degree of richness

TABLE 1: List of indices of all methods.

Approach	ExT	DRR
Our method	ExT_CDW	DRR_CDW
Method in [29]	ExT_MCS	DRR_MCS
Method in [7]	ExT_CW	DRR_CW
Method in [28]	ExT_SSW	DRR_SSW

of retrieval (DRR) and time complexity. Evaluation of time complexity is mainly done by computing the execution time of workflow retrieval. So this index is shortened as ExT. Another index DRR is defined as follows:

$$DRR = \frac{SW_R}{SW_C}, \quad (13)$$

where SW_R is the number of the workflows satisfying requirements of users in the retrieved results. SW_C is the number of the workflows satisfying requirements of users in the workflow repository.

We extended the functionality of workflow modeling tool [7]. In the extended tool, workflows with constraints can be graphically constructed. In addition, it can store ontology based semantic information. Each task node can be also associated with a set of data types describing the input and output parameters of the task node and a RDF file describing the semantic information of the task node. A workflow repository is implemented by a file directory in which many CDWs are stored as .xml files, and semantic information of nodes is stored as .rdf files. We manually constructed ten groups of CDW models with different sizes (i.e., the node number of a CDW) and depths (i.e., the node number of path from the start node to the last node in a CDW) as the testing data set for our experiments and evaluation.

The benchmark methods to compare are as follows. Bunke and Shearer [29] presented a distance metric for comparing the similarity of graphs based on maximal common subgraphs. This approach can be used for measuring the similarity among workflow structures. The second approach fully considers the constraints residing in workflows, such as cost. Ma et al. [7] proposed an approach to compare workflows with time constraints, where a time constraint is represented by an interval. The third approach concentrates on the comparison among semantic workflows, which was proposed by Bergmann and Gil [28]. This method considers comparing the semantic information in workflows.

In the following, we will compare our method with the other three methods mentioned above according to the two indices ExT and DRR. For simplification, Table 1 is the list of indices of all methods.

7.2. Experimental Evaluation

7.2.1. Comparing Degrees of Retrieval Richness for Evaluating Effectiveness. Figure 2 shows the degree of retrieval richness of four methods. From this figure, we can find some facts. Generally speaking, the larger the size of data set becomes, the more suitably the candidate workflows may be contained. Therefore, the value of DRR will rise. From this figure, we can

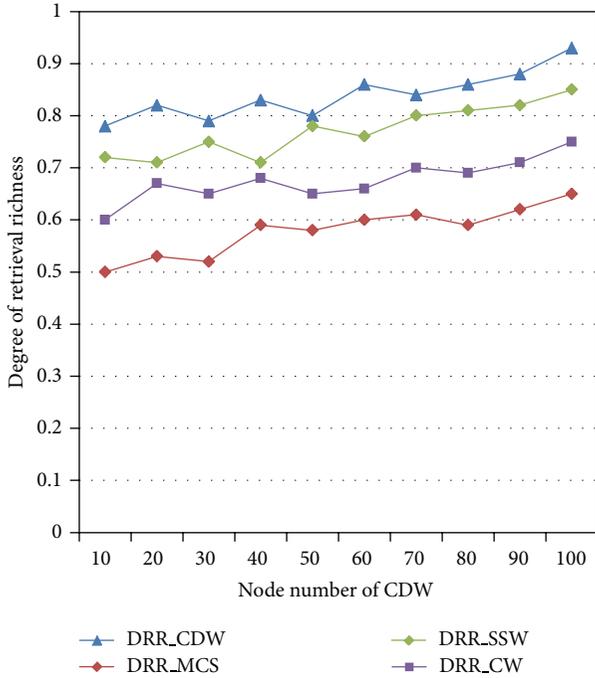


FIGURE 2: Comparison of degrees of retrieval richness.

find that the retrieval richness of our method is larger than others. It indicates that our method will find more suitable workflows with less irrelevant ones. DDR indicates that our approach has a stronger ability to discover information in a deeper level and will have a better retrieval effect obviously.

Higher value of DRR reflects that more suitable workflows have been discovered. One possible reason is that our method considers semantic and structural similarities for constrained data oriented workflows retrieval, which could discover deep relations such as nodes with different names but the same quality of service. This is quite important because users may choose cheaper services that are sufficient for their requirements. From the experimental results, we can conclude that our method with a higher degree of richness has an advantage over others to find more suitable information for specific users.

7.2.2. Comparing Time Complexity Based on Graph Depth for Evaluating Efficiency. Figure 3 is the comparison result of execution time between our method and that proposed in [28]. From the experimental results, we can have the following observation. As the size of graph grows larger, the execution time of both methods increases. For example, execution time of comparison between workflows with 70 nodes is longer than that with 30 nodes. In general, the execution time of our method is lower than that proposed in [28]. Graph depth has a larger effect on the execution of approach in [28], but less on our method. From the experimental results, we can find with the growing depths of the same size of graphs that the execution time increases faster than ours.

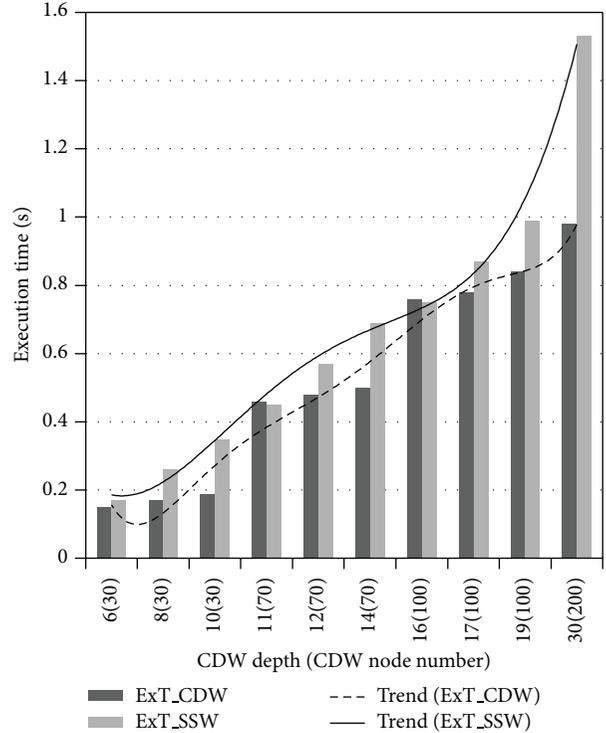


FIGURE 3: Execution time comparison based on depth.

There are many factors that will bring about these experimental results. First, both methods adopt graph to represent workflow models. When workflow models became complex and large, more structural and semantic information needed to be dealt with through the whole comparison process which demands more execution time obviously. Second, Bergmann and Gil in [28] used a case reasoning approach to measure the similarity of semantic information of different nodes. This method depends heavily on the depth of graph. Time complexity of this method will grow faster when the depth of graph increases. It straightly influences the whole similarity measure time. In our method, we seamlessly combine semantic and structural similarities into normalized matrices. The whole time complexity largely depends on the matrix computation. Usually time complexity of matrix computation can be affordable in many cases. For example, we have three workflow groups with 100 nodes in graphs but different graph depths: 16, 17, and 19, respectively. From the experimental results, ExT_SSW grows faster than ExT_CDW.

7.2.3. Comparing Total Time Complexity for Evaluating Performance. Figure 4 is the execution time comparison between our method and the other two. From this figure, we can find that all methods cost more time with the growing of graph sizes. The execution time of our method increases a little faster than the other two methods. However, the extra time cost will be tolerant and affordable if we further take a look at Figure 4 because extra time cost will contribute to a better degree of richness of retrieval.

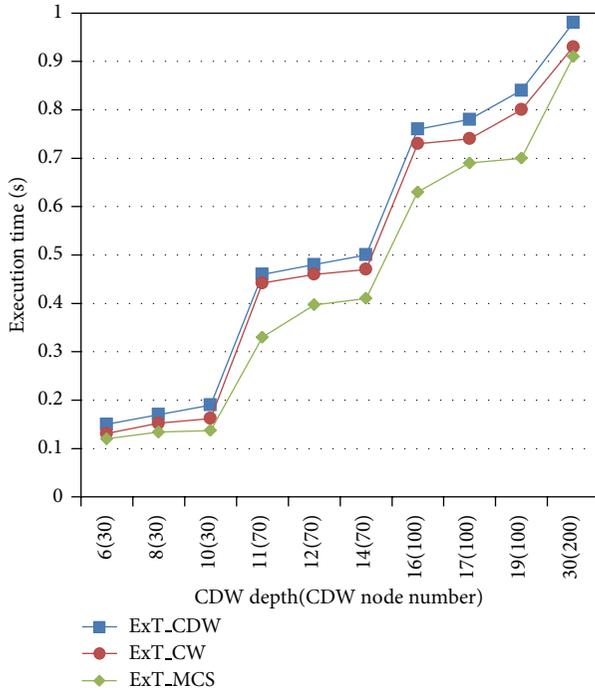


FIGURE 4: Execution time among the three methods.

Task nodes and relations between them are represented in a graph by Bunke and Shearer in [29]. The authors proposed a distance measure by maximal common subgraph. The formula is as follows: $D_MCS(W_1, W_2) = 1 - (|V| / \max\{|V_{W_1}|, |V_{W_2}|\})$. $D_MCS(W_1, W_2)$ computes the distance of W_1 and W_2 , where $|V|$ is the number of maximal common subgraph between workflows W_1 and W_2 . Time complexity mainly depends on the algorithm of subgraph isomorphism. The time complexity of subgraph isomorphism with unlabeled nodes in a graph has been proven to be an NP-complete problem. However, if each node has a unique label, the time complexity will decrease. Semantic information was not considered in the workflow models which would decrease the execution time.

In brief, we evaluate the effectiveness and efficiency of our approach by comparing it with the existing approaches of workflow retrieval. The experimental results show that (1) our method has more degree of semantic richness, which means that our approach can find out more accurate candidate workflows and has a higher retrieval precision; (2) the execution time that our method needs is no more than the other methods and is even lower than them. So our method outperforms the existing approaches of workflow retrieval and has higher performance.

8. Conclusion

In this paper, we proposed a graph based representation to describe constrained data oriented workflows, which seamlessly combines semantic and structure similarities to gain a better retrieval effect. It can mine out more and more deep

information that is crucial to reflect the quality of services provided by workflows. It is convenient for workflows reuse for different domain scientists with specific requirements. The similarity comparison is based on matrix norm to measure the distance of two normalized matrices for corresponding workflow models. The experimental evaluation shows that our method outperforms the existing approaches of workflow retrieval. The method proposed in this paper can be widely used in workflows retrieval, reuse, matching, and so on.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is partially supported by the National Natural Science Foundation of China (61001197, 61372182) and the Fundamental Research Funds for the Central Universities.

References

- [1] C. Berkley, S. Bowers, M. B. Jones et al., "Incorporating semantics in scientific workflow authoring," in *Proceedings of the 17th International Conference on Scientific and Statistical Database Management (SSDBM '05)*, Santa Barbara, Calif, USA, June 2005.
- [2] Y. Gil, E. Deelman, M. Ellisman et al., "Examining the challenges of scientific workflows," *Computer*, vol. 40, no. 12, pp. 24–32, 2007.
- [3] M. Zohrevandi and R. A. Bazzi, "The bounded data reuse problem in scientific workflows," in *Proceedings of the IEEE 27th International Symposium on Parallel & Distributed Processing (IPDPS '13)*, pp. 1051–1062, 2013.
- [4] Y. Gil, V. Ratnakar, and C. Fritz, "Assisting scientists with complex data analysis tasks through semantic workflows," in *Proceedings of the AAAI Fall Symposium on Proactive Assistant Agents*, pp. 14–19, November 2010.
- [5] K. K. Castillo-Villar, N. R. Smith, and J. F. Herbert-Acero, "Design and optimization of capacitated supply chain networks including quality measures," *Mathematical Problems in Engineering*, vol. 2014, Article ID 218913, 17 pages, 2014.
- [6] L. Longyi and Z. Yansheng, "Study of supply chain workflow based on grid," in *Proceedings of the International Conference on Management and Service Science (MASS '09)*, September 2009.
- [7] Y. Ma, X. Zhang, and K. Lu, "A graph distance based metric for data oriented workflow retrieval with variable time constraints," *Expert Systems with Applications*, vol. 41, no. 4, pp. 1377–1388, 2014.
- [8] R. Bergmann and Y. Gil, "Retrieval of semantic workflows with knowledge intensive similarity measures," in *Case-Based Reasoning Research and Development*, vol. 6880 of *Lecture Notes in Computer Science*, pp. 17–31, 2011.
- [9] D. Chiu, T. Hall, F. Kabir, and G. Agrawal, "An approach towards automatic workflow composition through information retrieval," in *Proceedings of the 15th Symposium on International Database Engineering & Applications (IDEAS '11)*, pp. 170–178, 2011.

- [10] P.-F. Tsai, "A label correcting algorithm for partial disassembly sequences in the production planning for end-of-life products," *Mathematical Problems in Engineering*, vol. 2012, Article ID 569429, 13 pages, 2012.
- [11] B. Cantalupo, L. Giammarino, N. Matskanis et al., "Semantic workflow representation and samples," 2005, http://eprints.soton.ac.uk/268554/1/2005-_18554.pdf.
- [12] N. Russell, A. H. M. Ter Hofstede, D. Edmond, and W. M. P. van der Aalst, "Workflow data patterns: identification, representation and tool support," in *Conceptual Modeling—ER 2005: 24th International Conference on Conceptual Modeling, Klagenfurt, Austria, October 24-28, 2005. Proceedings*, vol. 3716 of *Lecture Notes in Computer Science*, pp. 353–368, 2005.
- [13] M. Chadli, H. R. Karimi, and P. Shi, "On stability and stabilization of singular uncertain Takagi-Sugeno fuzzy systems," *Journal of the Franklin Institute*, vol. 351, no. 3, pp. 1453–1463, 2014.
- [14] M. Chadli and T. M. Guerra, "LMI solution for robust static output feedback control of discrete takagi-sugeno fuzzy models," *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 6, pp. 1160–1165, 2012.
- [15] S. Aouaouda, M. Chadli, P. Shi, and H. R. Karimi, "Discrete-time H- / Hoosensor fault detection observer design for nonlinear systems with parameter uncertainty," *International Journal of Robust and Nonlinear Control*, 2014.
- [16] M. Chadli, A. Abdo, and S. X. Ding, " H_2/H_∞ fault detection filter design for discrete-time Takagi-Sugeno fuzzy system," *Automatica A*, vol. 49, no. 7, pp. 1996–2005, 2013.
- [17] M. Chadli and H. R. Karimi, "Robust observer design for unknown inputs Takagi-Sugeno models," *IEEE Transactions on Fuzzy Systems*, vol. 21, no. 1, pp. 158–164, 2013.
- [18] M. Chadli, S. Aouaouda, H. R. Karimi, and P. Shi, "Robust fault tolerant tracking controller design for a VTOL aircraft," *Journal of the Franklin Institute*, vol. 350, no. 9, pp. 2627–2645, 2013.
- [19] R. Ikeda, S. Salihoglu, and J. Widom, "Provenance-based refresh in data-oriented workflows," in *Proceedings of the 20th ACM Conference on Information and Knowledge Management (CIKM '11)*, pp. 1659–1668, October 2011.
- [20] M. Hutchins, H. Foster, T. Goradia, and T. Ostrand, "Experiments on the effectiveness of dataflow- and controlflow-based test adequacy criteria," in *Proceedings of the 16th International Conference on Software Engineering*, pp. 191–200, May 1994.
- [21] E. Deelman, G. Singh, M.-H. Su et al., "Pegasus: a framework for mapping complex scientific workflows onto distributed systems," *Scientific Programming*, vol. 13, no. 3, pp. 219–237, 2005.
- [22] I. Foster, J. Voeckler, M. Wilde, and Y. Zhao, "Chimera: a virtual data system for representing, querying and automating data derivation," in *Proceedings of the 14th International Conference on Scientific and Statistical Database Management (SSDBM '02)*, pp. 37–46, 2002.
- [23] Y. L. Simmhan, B. D. Plale, Gannon, and S. Marru, "Performance evaluation of the karma provenance framework for scientific workflows," in *Provenance and Annotation of Data: International Provenance and Annotation Workshop, IPAW 2006, Chicago, IL, USA, May 3-5, 2006, Revised Selected Papers*, L. Moreau and I. T. Foster, Eds., vol. 4145, pp. 222–236, Springer, 2006.
- [24] "VDS—The GriPhyN Virtual Data System," <http://www.ci.uchicago.edu/wiki/bin/view/VDS/VDSWeb/WebMain>.
- [25] Y. Gil, E. Deelman, M. Ellisman et al., "Examining the challenges of scientific workflows," *Computer*, vol. 40, no. 12, pp. 24–32, 2007.
- [26] S. Fortin, "The graph isomorphism problem," Tech. Rep. 96-20, University of Alberta, Edmonton, Alberta, Canada, 1996.
- [27] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, *Feature Extraction: Foundations and Applications*, Springer, 2006.
- [28] R. Bergmann, G. Müller, and D. Wittkowsky, "Workflow clustering using semantic similarity measures," in *KI 2013: Advances in Artificial Intelligence*, vol. 8077 of *Lecture Notes in Computer Science*, pp. 13–24, 2013.
- [29] H. Bunke and K. Shearer, "A graph distance metric based on the maximal common subgraph," *Pattern Recognition Letters*, vol. 19, no. 3-4, pp. 255–259, 1998.

Research Article

Optimizing Gear Shifting Strategy for Off-Road Vehicle with Dynamic Programming

Xinxin Zhao,^{1,2} Wenming Zhang,¹ Yali Feng,¹ and Yaodong Yang¹

¹ School of Mechanical Engineering, University of Science and Technology Beijing, Beijing 100083, China

² Department of System Design Engineering, University of Waterloo, Waterloo, ON, Canada N2L 3G1

Correspondence should be addressed to Yali Feng; ylfeng126@126.com

Received 4 April 2014; Revised 28 May 2014; Accepted 6 June 2014; Published 22 June 2014

Academic Editor: Hamid Reza Karimi

Copyright © 2014 Xinxin Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Gear shifting strategy of vehicle is important aid for the acquisition of dynamic performance and high economy. A dynamic programming (DP) algorithm is used to optimize the gear shifting schedule for off-road vehicle by using an objective function that weighs fuel use and trip time. The optimization is accomplished through discrete dynamic programming and a trade-off between trip time and fuel consumption is analyzed. By using concave and convex surface road as road profile, an optimal gear shifting strategy is used to control the longitudinal behavior of the vehicle. Simulation results show that the trip time can be reduced by powerful gear shifting strategy and fuel consumption can achieve high economy with economical gear shifting strategy in different initial conditions and route cases.

1. Instruction

The notable differences between off-road vehicle and passenger car include special functions, huge weight capacity, and rough driving area, which may lead to gear shifting frequently. The automatic transmission had been installed in construction vehicle since the mid of last century. Mining trucks with mechanical powertrain system often equip automatic transmissions, which could enhance the production efficiency. To achieve high economy, gear shifting strategy of automatic transmission is a challenging problem for researchers. For a heavy vehicle weighing more than 15 tons, about 1/3 of the life cycle cost comes from the cost of fuel [1]. As a special vehicle, mining truck weighting around 50 tons, the average fuel consumption is around 100 L/100 km. For this type of vehicle, the gear shifting strategy improving fuel efficiency in every trip has good benefit in the whole life cycle.

The drive cycle of mining truck is often regulated in advance, which locates between source area and storage area. It leads to collecting elevation and slope data easily. With known road data, it is available to use dynamic programming approach to optimize the gear shifting strategy to reduce the fuel consumption and trip time. As a numerical method

for solving multistage decision-making problems, dynamic programming has been applied to optimize fuel and electricity costs associated with two supervisory control strategies for a series of plug-in hybrid electric vehicle and control strategy for heavy hybrid electric truck [2, 3]. Besides, it is proposed for design of the optimal gear shift strategy to study quantitatively an optimal trade-off between the fuel economy and the drivability [4]. At present, some researchers have developed economical gear shifting strategy applying for heavy duty truck. The problem that which kind of speed profile will minimize fuel consumption of a land vehicle is stated. Based on different guide way characteristics, the optimal speed profile is analyzed for minimizing fuel consumption [5]. Lattemann and his colleagues used predictive cruise control allowing the vehicle speed to vary around the cruise control by setting speed within a defined speed band in an effort to reduce fuel consumption for trucks [6]. The efficiency of engine, the powertrain system, and the pump are considered and set up the optimal economy gear shifting strategy for construction vehicles [7, 8]. Shi et al. used a modified recursive least square method to estimate the vehicle mass and generalized slope and made a self-adaptive gear shifting strategy of uphill and downhill

slopes without considering fuel consumption [9]. For costing lots of fuel consumption, it should be considered into gear shifting strategy. The method of modeling and calculating fuel consumption for heavy-duty vehicle was illustrated, and it had been verified by experiments [10]. Therefore, it is plausible to calculate the fuel consumption based on the accurate powertrain system model and real road profile. For optimization, economical strategy for a heavy truck is conducted by finding how to drive a truck over various road topographies. While the results show that the optimal solution is to keep constant speed for level road and in small gradients [11], Delgado et al. got the results which show that the use of average velocity and average positive acceleration was suitable for the translation of fuel consumption [12].

In this work, an optimal control algorithm based on dynamic programming is formulated to find gear shifting strategy for a mining truck weighting 50 tons. Comparing different step lengths, the proper length had been found with considering accuracy and enumeration load. With a traditional powertrain model and known road profile ahead, we minimized trip time and fuel consumption in different initial states on simple road profile separately. The results show there is a trade-off between trip time and fuel consumption in all of cases and when it achieves the highest fuel economy with sacrificing the trip time. Taking into account all characteristics of powerful strategy and economical strategy, the trip time becomes a constraint instead of a part of objective function. The optimal gear shifting strategy presents that it spends less fuel consumption than powerful strategy and takes fewer time than economical strategy. The optimal algorithm could be applied to find an economical gear shifting strategy with desired average speed. In addition, two special road profiles are substituted for simple road and the optimal results accord with the results derived from simple road profile. Combing a two-dimensional dynamic programming algorithm with weighted cost function and known route information will be discussed and compared in trip time and fuel consumption.

2. Modeling for Mining Truck Powertrain System

2.1. Powertrain System Configuration and Dynamics Analysis.

Powertrain system of mining truck comprises an engine, an automatic transmission, a drive shaft, a retarder, and a wheel-side reducer. The power produced by engine transfers from engine to wheel through transmission system to drive the vehicle [13]. Figure 1 shows the configuration and torque pass line of driveline for a 50-ton mining truck. Table 1 shows the main parameter of the truck.

The truck equips one of the M11 series diesel engines produced by Cummins, and the output power is related to the mass of fuel in every stroke cycle u_f and engine angle speed ω_e . Based on experimental data, the output torque can be expressed by a polynomial as follows:

$$T_e(\omega_e, u_f) = a_1\omega_e + a_2u_f + a_3, \quad (1)$$

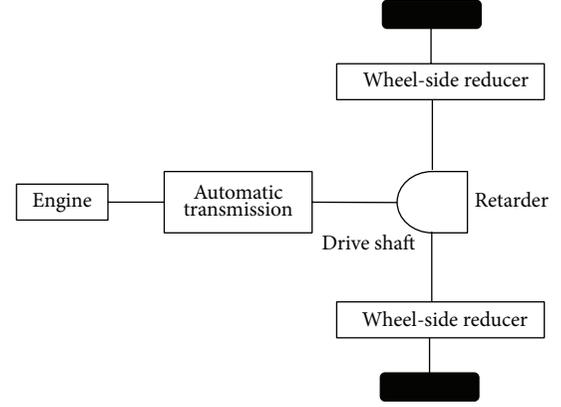


FIGURE 1: The scheme of the powertrain system for a 50-ton mining truck.

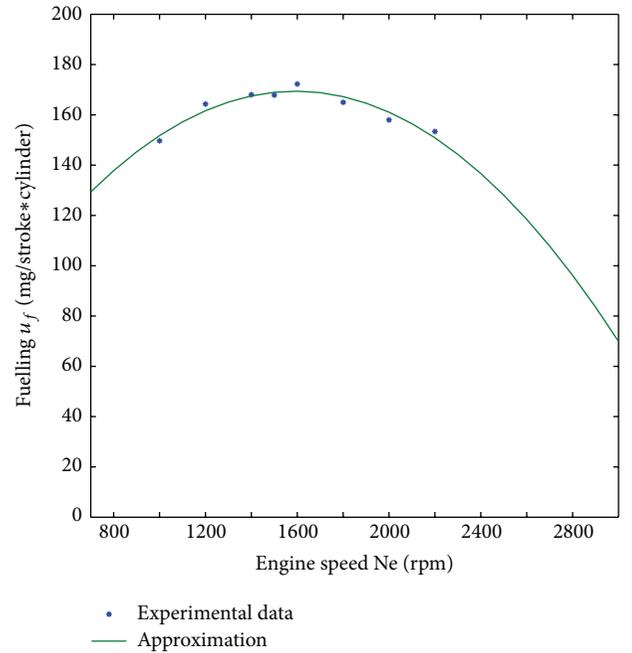


FIGURE 2: Fueling bounds and experimental data.

where the maximum value of fueling depends on a quadratic polynomial of the engine angle speed, like

$$u_{f,max} = a_4\omega_e^2 + a_5\omega_e + a_6. \quad (2)$$

In the above equation, the coefficients from a_1 to a_6 are derived from the experimental data of this engine. From Figure 2, it shows that the fueling bounds appear to be well approximated by using the polynomial (2).

Newton's second law of the motion gives the relation between the engine acceleration of angle speed $\dot{\omega}_e$, the inertia of the rotating parts J_e , the output torque of engine T_e , the input torque T_t to converter, and loss torque $T_{fric,e}$. Consider

$$J_e\dot{\omega}_e = T_e - T_{fric,e} - T_t. \quad (3)$$

TABLE 1: Main parameters of the vehicle.

Item	Description or quantity
Engine	Diesel; displacement: 8.9 L; maximum power: 242 kw; maximum torque: 1424 Nm
Transmission	Five-speed AT;
Vehicle mass	25400 kg

TABLE 2: Resistance.

Symbol	Explanation	Equation
F_{wind}	Air drag	$\frac{1}{2}c_w A_a \rho_a v^2$
F_R	Rolling resistance	$mgc_r \cos \alpha$
F_{gra}	The gravitational force	$mg \sin \alpha$

When a gear is engaged, the torque at wheel T_w can be determined by the lumped rotating inertia J_l , the gear ratio i_g , and the lumped reduction ratio from retarder and wheel-side reducer. Ignoring the brake torque, denote T_w and T_t as the following equations:

$$\begin{aligned} T_w &= ii_g \eta T_t, \\ J_l \dot{\omega}_w &= T_w - T_L - F_L r_w, \end{aligned} \quad (4)$$

where η is the transmission efficiency of the driveline system and $\dot{\omega}_w$, T_L , F_L , and r_w represent the angular acceleration, braking torque, friction, and radius of wheel, respectively.

Based on the longitudinal dynamics of vehicle, the friction of wheel can be described by Newton's second law. The equation shows the following, where m is the mass of vehicle and v is velocity of vehicle. In the longitudinal direction, the main resisting forces are considered to be air drag, rolling resistance, and gravitational force, which are illustrated in Table 2. Consider

$$F_L = m \frac{dv}{dt} + F_{\text{wind}} + F_R + F_{\text{gra}}. \quad (5)$$

Combining (1) to (5), the acceleration of vehicle can be represented by parameters that are mentioned in this section [14]. Consider

$$\begin{aligned} \frac{dv}{dt}(v, u_f, \alpha) &= \frac{r_w}{J_l + m r_w^2 + \eta i^2 J_e} \\ &\cdot (ii_g \eta T_e(v, u_f) \\ &- r_w (F_{\text{wind}}(v) + F_R(\alpha) + F_{\text{gra}}(\alpha))). \end{aligned} \quad (6)$$

2.2. Fuel Consumption. The advanced internal combustion engine produces power and emission from fuel and air in every stroke. The fuel consumption m_f (g/s) is identified depending on fueling u_f (g/stroke * cylinder) and engine speed ω_e (rad/s) [15]. Consider

$$\frac{dm_f}{dt}(\omega_e, u_f) = \frac{n_{\text{cyl}}}{2\pi n_r} \omega_e u_f = \frac{n_{\text{cyl}} ii_g}{2\pi n_r r_w} v u_f, \quad (7)$$

where n_{cyl} is the number of cylinders and n_r is the crank rotation number in every stroke.

2.3. Reformulation Model by Displacement. For optimization by numerical method, the model should be transformed from being parameterized by time to being parameterized by position. Since the states including velocity and fuel consumption are discrete in regard to time, they need to be represented by position. Assuming a short distance p , it is the function of time t and the function of displacements [16, 17]. Then, p can be represented as follows:

$$\frac{dp}{ds} = \frac{dp}{dt} \frac{dt}{ds} = \frac{dp}{dt} \frac{1}{v} \implies \Delta t = \frac{\Delta p}{v}. \quad (8)$$

When the speed $v > 0$, the short distance is assumed to be p_k . Therefore, (6) and (7) will be transferred to the next two equations:

$$\begin{aligned} v_{k+1} &= v_k + \Delta t_k \cdot \frac{dv}{dt} = v_k + \frac{p_k}{v_k} \cdot \frac{dv}{dt}, \\ m_{f,k+1} &= m_{f,k} + \Delta t_k \cdot \frac{dm_f}{dt} = m_{f,k} + \frac{p_k}{v_k} \cdot \frac{dm_f}{dt}. \end{aligned} \quad (9)$$

In this problem, the velocity of vehicle, fuel consumption, and gear number in current state can be described by their value of last state. The state vector is assumed to be $x = [v, g_k]^T$. By ignoring the braking torque, the control trajectory is the gear shifting state $u = [u_g]^T$.

3. Optimizing Gear Shifting Strategy by Dynamic Programming

Mining truck is different from passenger car in several aspects, such as driving condition. The trajectory of off-road vehicle is usually fixed from mining area to intermediate of transport. This route can be divided into small steps and the distance of each part is equal to p_k . The gear number for the next step is decided by dynamic programming algorithm to minimize cost function for the next step. Following the same way, the gear shifting strategy will be acquired for the whole route.

3.1. Constraints of State. The roads of mining truck may decline sharply which results in the actual speed just achieving 80% of the maximum speed. The speed should be less than the maximum speed of current gear ($v_{g,\text{max}}$) and larger than the minimum speed of current gear ($v_{g,\text{min}}$) [18]. For different gear, the range of speed is shown as follows:

$$0 < v_{\text{min}} \leq v(g) \leq 0.8v_{\text{max}}, \quad 0 < v_{g,\text{min}} \leq v(g) \leq v_{g,\text{max}}. \quad (10)$$

There are five gears for the automatic transmission equipping in mining truck, and the gear number should be continuously assuming no gear skipping. Therefore, gear

number needs to satisfy the following constraints, when the current gear is g_k and the next gear is g_{k+1} . Consider

$$g_{k+1} \in \begin{cases} (1, 2), & \text{if } g_k = 1, \\ (4, 5), & \text{if } g_k = 5, \\ (g_k, g_{k+1}, g_{k+2}), & \text{otherwise.} \end{cases} \quad (11)$$

Because of the fuel consumption calculated by real engine speed and torque, there are no other constraints.

3.2. Constraints of Control. The transmission control unit (TCU) sends gear shifting signals to transmission. The range of control policy also needs to be considered with the current gear number. When the gear is the highest gear, it is impossible to upshift any more. While the gear is the lowest gear, it cannot continue to downshift. Otherwise, the control policy can be any plausible value of $-1, 0$, and 1 . Consider

$$u_{g,k} \in \begin{cases} (0, 1), & \text{if } g_k = 1, \\ (-1, 0), & \text{if } g_k = 5, \\ (-1, 0, 1), & \text{otherwise.} \end{cases} \quad (12)$$

3.3. Objective Function. Considering fuel consumption with ignoring trip time, a truck will have a powerful performance. Therefore, the objective function included not only fuel consumption costing on a known route but also trip time. Moreover, penalty coefficients in the cost function are used to adjust a trade-off between powerful and economical performances of vehicle. The cost function taking into account both trip time and fuel consumption is shown as follows:

$$J = \xi_k(x_k, u_k) = \sum_{k=0}^{N-1} \alpha \cdot m_{f,k}(x_k) + \sum_{k=0}^{N-1} \beta \cdot t_k(x_k), \quad (13)$$

$$k = 0, 1, 2, \dots, N-1,$$

where α and β are the penalty coefficients for fuel consumption and trip time, respectively, and N is the number of steps dividing the whole journey.

3.4. Application of the Dynamic Programming Algorithm. For a given system, dynamic programming can be used to find the optimal control input that minimizes a chosen cost function. However, all dynamic programming algorithms are based on decision processes. It means that a dynamic system with continuous inputs and states has to be approximated by a discrete-value system. There are several issues need to be considered for implementing the algorithm, including grid selection, interpolation, and assuming parameters of algorithm. For dynamic programming approach, the control variable needs to be discrete firstly. Plausible values of gear control $u = [u_g]^T$ are given by the constraint of gear number. For system state $x = [v, g_k]^T$ in stage k , all plausible next states can be described by (14). When the state is shown as follows,

TABLE 3: Parameters for dynamic programming.

Parameter	Explanation	Value
p_k	Step length	5 m
N	Number of step	20
$p_k \cdot N$	Distance	100 m
α	Coefficient of economical performance	1 or 0
β	Coefficient of powerful performance	1 or 0

the cost-to-go function is written as (15), where x_{k+1}^n are state vectors at stage $k+1$. Consider

$$x_{k+1} = \begin{bmatrix} v_{k+1} \\ g_{k+1} \end{bmatrix} = \begin{bmatrix} v_k \\ g_k \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \cdot [u_g] + \begin{bmatrix} dv \\ ds \\ 0 \end{bmatrix} \cdot [ds], \quad (14)$$

$$\hat{J}_{k+1}(x_{k+1}) = \sum_{n=m}^{m+1} \psi^n J(x_{k+1}^n). \quad (15)$$

Based on the cost-to-go function and state values, the algorithm is illustrated as the following procedures of dynamic programming [19]:

- (i) find the $J_N(f(x, u)) = \min\{J(x_{k+1}^n)\}$;
- (ii) let $k = N - 1$;
- (iii) let $J_k(f(x, u)) = \min_{u \in U}\{\xi_k(x, u) + J_{k+1}(f(x, u))\}$;
- (iv) repeat the last step for $k = N - a, a = 2, 3, \dots, N$;
- (v) the minimum cost based on the optimal control is J_0 , where the controls are coming from the initial state.

To find the optimal gear shifting schedule for making mining truck powerful and economical, (13) is used as the cost function with considering constraints of states and controls. The states are discrete into a matrix whose size is 9×5 . Because velocity is continuous and valuable, the interval of discretization is assumed to be 5 km/h. Allowable controls are applied in every step, and the optimal control is recorded in every step, which makes cost become the minimum one. Repeating mentioned process, the minimum cost will be found when the step goes to the first step. It means that the controls are the best gear shifting strategy combining powerful and economical performance for mining truck. The parameters of approach are presented in Table 3.

4. Discussion

The problem optimizing gear shifting strategy to make vehicle performance powerful and economical will be transferred to a general problem that is searching an array control policy to minimize the cost function with specific constraints. In this section, the results are described with adjusting different step lengths, initial states, and coefficients of powerful and economical performance. How those mentioned parameters influence results is conducted in this section.

4.1. Optimizing Gear Shifting Strategy in Different Step Lengths. Assume the route ahead of mining truck is 100 m straight without any slope. The powerful coefficient is set to

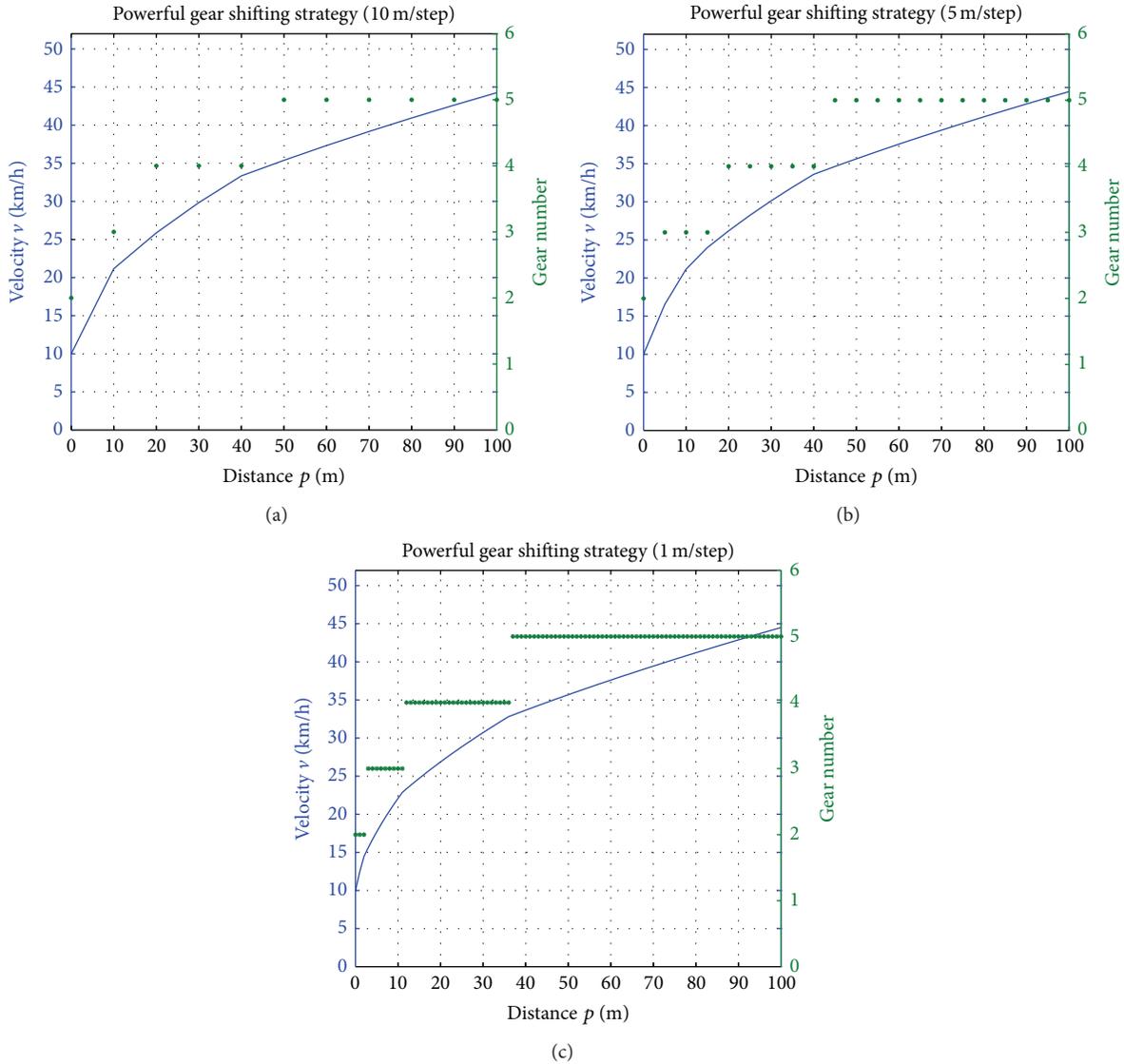


FIGURE 3: The performance of vehicle in different step lengths.

be 1 while economical coefficient is set to be 0. It means that the objective function just considered trip time and the optimal results are the powerful gear shifting strategy. The simulations in same initial state (10 km/h, 2nd gear) are conducted with different step lengths, such as 10 m, 5 m, and 1 m. Comparing the results in different step lengths, the optimal gear shifting strategy based on minimizing trip time can be seen in Figure 3.

The final states of different step lengths are almost the same, while the trip time of those conditions almost spends the same time: 11.48 s for 10 m/step, 11.41 s for 5 m/step, and 11.21 s for 1 m/step. The moments achieving the fifth gear for different conditions are different. For step length of 10 m, the gear increases by the fifth gear after the vehicle runs to 50 m, while the distances are 40 m and 37 m for 5 m/step and 1 m/step, respectively. Therefore, the optimal control is influenced by step length in the same horizon. However, step length is not as small as possible for a numerical approach.

The smaller step length is applied, the more times iteration is used. It results in increasing the computational complexity. For actual problem, the step length should consider the response performance of vehicle system. If the frequency of gear shifting exceeds the intrinsic system value, it deteriorates the comfort performance and components abrasion. As can be seen from Figure 3, the results of 5 m/step state the main performance index of vehicle gear shifting without calculation burden and this length is adapted in the following work.

The computational complexity is determined by the dimensions and the number of quantization levels used for the state and control spaces. The step length determines the minimum interval for gear shifting [20]. In different step lengths, the required computation time becomes

$$T = zNN_xN_u, \tag{16}$$

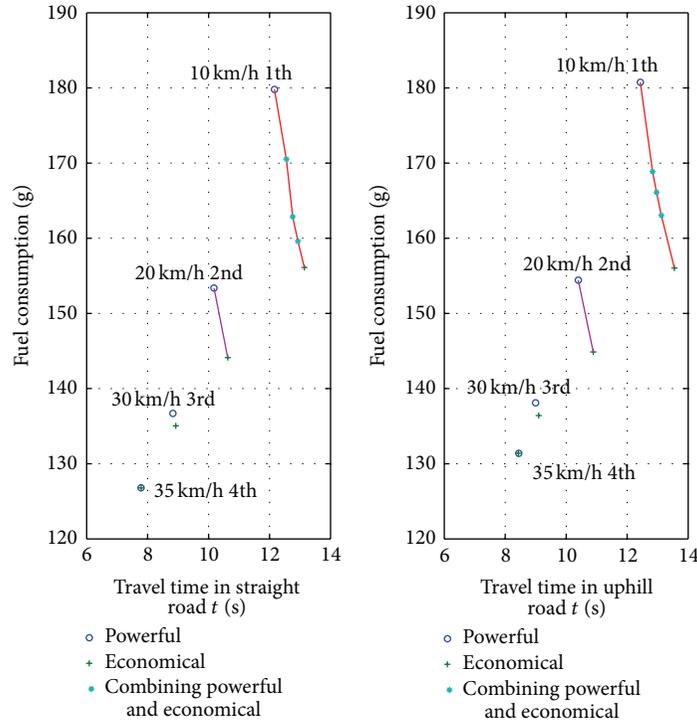


FIGURE 4: The trade-off between fuel consumption and trip time in different conditions.

where z is a constant and N_x and N_u are the total number of state grid points and discrete controls, respectively. The constant is mainly dependent on the capacity and speed of the available computer hardware.

4.2. The Relationship between Fuel Consumption and Trip Time. The best shifting schedule is to take the lowest fuel consumption and the shortest trip time in the same time. However, there is a coupling relationship between trip time and fuel consumption. According to the conservation of energy, it takes more fuel consumption to obtain more power when the trip time is too short, but when we need to save the fuel, the trip time should be increased to produce the same amount of power. Therefore, how much of the fuel consumption increases when reducing the trip time is the key to choose which kind of shifting schedule for this system.

To solve this problem, the trip time is changed to become a constraint by modifying the previous model of dynamic programming. The result shows that the optimized shifting schedule balances the power and the economy. Using the 10 km/h at the first gear as the initial state, the results are showed in Figure 4. The quantitative relationship between trip time and fuel consumption is set up, which has been improved and is acceptable. In Figure 4 the left figure shows the results in 100 m straight road, and right one states the results in 100 m uphill road with a 2.5 (rad) slope.

Choosing four kinds of states as initial states, the model is conducted in straight road and uphill road, respectively. As showed in Figure 4, when initial state is 35 km/h at the fourth gear, the results are in the area of high speed with low

TABLE 4: Main parameters on convex road.

Specific strategy	Trip time/s	Fuel consumption/L
Economical gear shifting strategy	62.15	0.7522
Powerful gear shifting strategy	53.23	0.7527

fuel consumption, which indicates that the result of powerful shifting schedule is the same as that economical shifting schedule. The gear shifting schedule by using the dynamic programming is beneficial for finding optimal strategy with fixed real road profile.

4.3. Comparison of Performances in Different Road Profiles

4.3.1. For Convex Road Surface. Examples of convex road are seen in Figure 5, which is 340 m. It gives the variation of trip time, fuel consumption, velocity, and gear versus the displacement. The blue stars represent the gear shifting schedule for minimizing fuel consumption. Gear increases to the fourth gear before 100 m while gear in optimizing trip time case keeps on low gear at that moment. The reason is that vehicle at low gear is easy to achieve high acceleration. The trip time taking by minimizing fuel consumption case is longer than another case, and velocity has achieved the same value at the end of route.

The specific value of trip time and fuel consumption is shown in Table 4, where the data in the row of economical

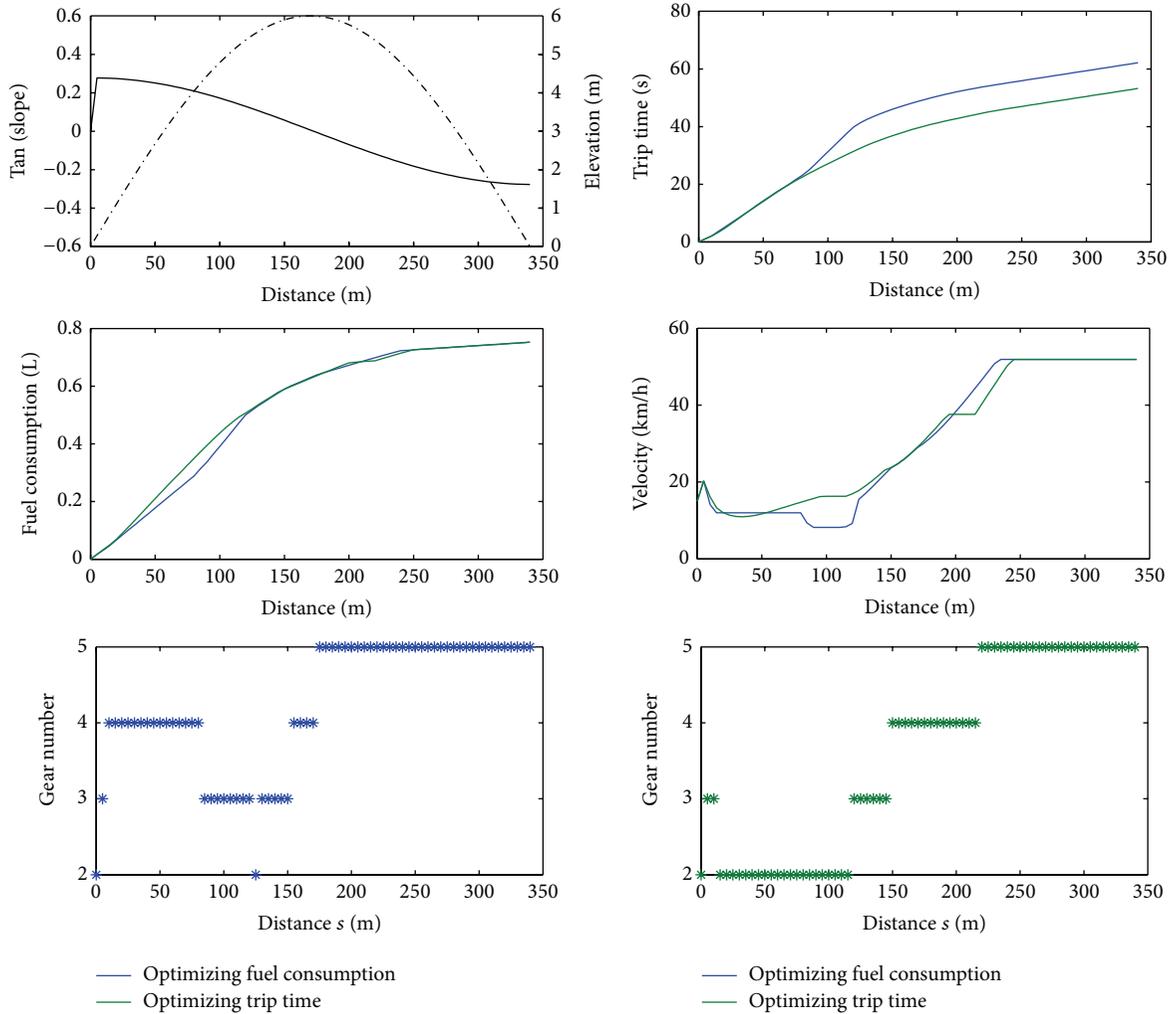


FIGURE 5: Vehicle performances on convex road.

gear shifting strategy means the results by minimizing fuel consumption. Simulations on 340 m convex route show that the trip time for a mining truck was reduced by 14.3% with an insignificant change in trip time with powerful gear shifting strategy. Although there are few differences for trip time and fuel consumption, it will become a large value when route extends to hundreds of miles.

4.3.2. *For Concave Road Surface.* In Figure 6 an example of concave road is shown. Simulations on the entire route are made with varying horizon lengths. The gear shifting strategies in different cost functions have similar trends before 300 m, while gear is decreased from fifth to third gear at the end of route when minimizing trip time. It leads to the fact that the velocity for optimizing fuel consumption is less than the velocity in another case. To see results clearly, the specific values were shown in Table 5.

Based on all of the results, the gear shifting strategy with different cost functions has correct response for the minimized parameter, and it validates that the algorithm of dynamic programming is reasonable. It can be applied for the

TABLE 5: Main parameters on concave road.

Specific strategy	Trip time/s	Fuel consumption/L
Economical gear shifting strategy	30.56	0.3883
Powerful gear shifting strategy	29.58	0.3998

real road profile for looking for the economical gear shifting strategy and powerful gear shifting strategy.

5. Conclusion

The paper established a shifting schedule based on the dynamic programming (DP) theory for off-road vehicle, with making fuel consumption and trip time as the objective function and gear changing as the controlling variable. Comparing and selecting the appropriate step length, it ensures the simulation to be accurate and efficient.

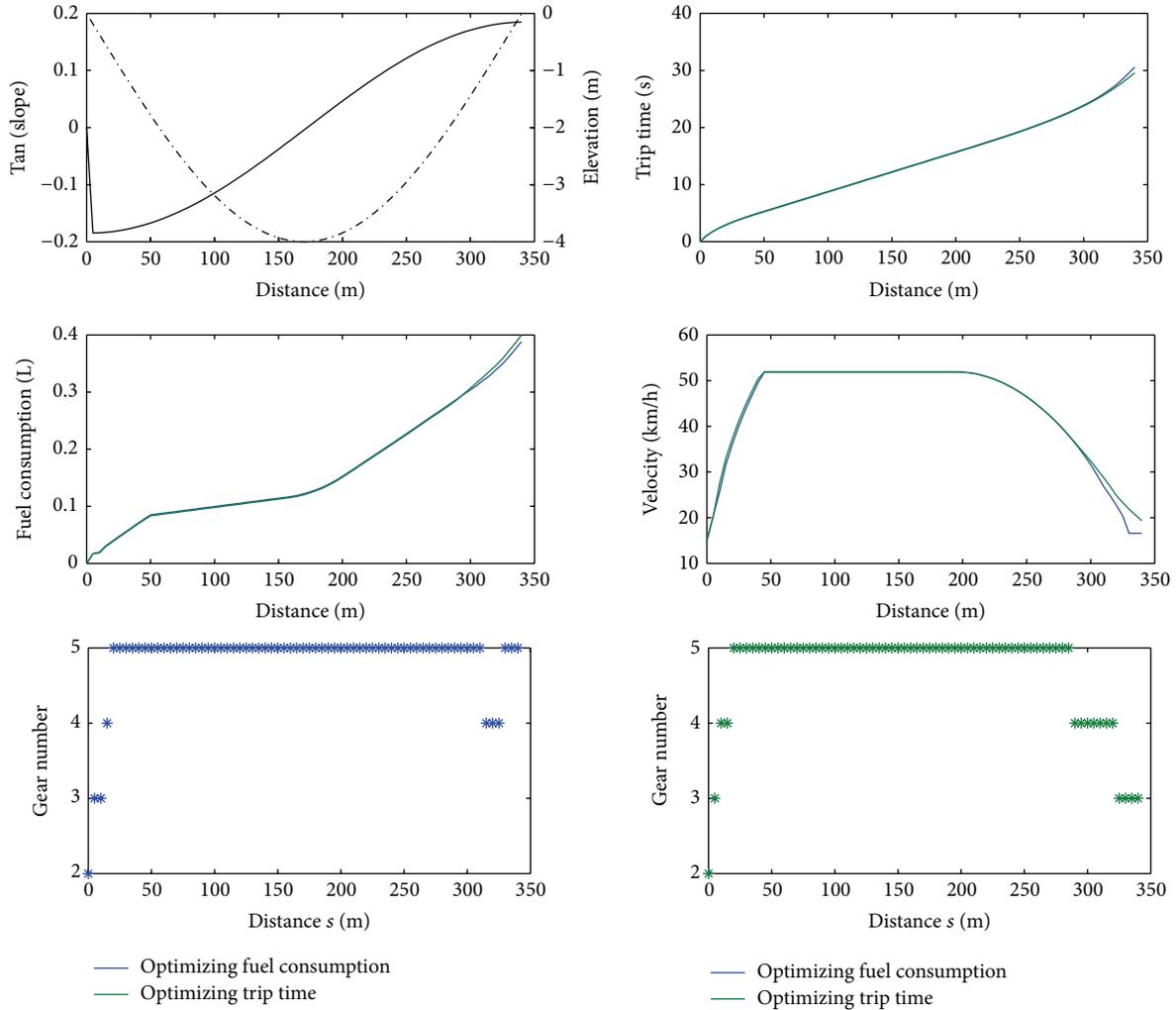


FIGURE 6: Vehicle performances on concave road.

Adjusting the penalty coefficients in the objective function is to optimize the powerful and economical strategy. The trip time is considered to be a constraint when making economical shifting strategy that balances the power and economy for mining truck. For the actual condition, the powerful shifting schedule has contributed to get the relationship between increment of the fuel consumption and decrement of the trip time. For the future work, the road profile should be alerted to the actual road with sorted of conditions and desired velocity penalty function can be added into objective function. Based on real road profile, the optimization will be conducted for minimizing fuel consumption for mining truck by using this method. The optimal control strategy derived from dynamic programming can be applied to the transmission control unit for saving fuel and time.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to gratefully acknowledge the China Scholarship Council for their support of this research. The authors also like to thank Professor Nasser Lashgarian Azad from the University of Waterloo for his good advises.

References

- [1] E. Hellström, M. Ivarsson, J. Åslund, and L. Nielsen, "Look-ahead control for heavy trucks to minimize trip time and fuel consumption," *Control Engineering Practice*, vol. 17, no. 2, pp. 245–254, 2009.
- [2] Y. Zou, D.-G. Li, and X.-S. Hu, "Optimal sizing and control strategy design for heavy hybrid electric truck," *Mathematical Problems in Engineering*, vol. 2012, Article ID 404073, 15 pages, 2012.
- [3] R. M. Patil, Z. Filipi, and H. K. Fathy, "Comparison of supervisory control strategies for series plug-in hybrid electric vehicle powertrains through dynamic programming," *IEEE*

- Transactions on Control Systems Technology*, vol. 22, no. 2, pp. 502–509, 2013.
- [4] V. D. Ngo, J. C. Navarrete, T. Hofman, M. Steinbuch, and A. Serrarens, “Optimal gear shift strategies for fuel economy and driveability,” *Proceedings of the Institution of Mechanical Engineers D: Journal of Automobile Engineering*, vol. 227, no. 10, pp. 1398–1413, 2013.
- [5] D. J. Chang and E. K. Morlok, “Vehicle speed profiles to minimize work and fuel consumption,” *Journal of Transportation Engineering*, vol. 131, no. 3, pp. 173–182, 2005.
- [6] F. Lattemann, K. Neiss, S. Terwen, and T. Connolly, “The predictive cruise control—a system to reduce fuel consumption of heavy duty trucks,” SAE Paper 2004-01-2616.
- [7] G. J. Cui, *Study on Three-Parameter Optical Shift Schedule and Control Method of Construction Vehicle*, Jilin University, Jilin, China, 2009.
- [8] Z. Miao, *Research on saving energy shift schedule with four-parameter and the control system of construction vehicle*, Jilin University, Jilin, China, 2006.
- [9] J. Shi, T. Lu, X. Li, and J. Zhang, “Self-adaptive slope gearshift strategy for automatic transmission vehicles,” *Transactions of the Chinese Society of Agricultural Machinery*, vol. 42, no. 4, pp. 1–7, 2011.
- [10] T. Sandberg, *Heavy Truck Modeling for Fuel Consumption Simulations and Measurements*, Linköping Studies in Science and Technology, Linköping, Sweden, 2001.
- [11] A. Fröberg, E. Hellström, and L. Nielsen, “Explicit fuel optimal speed profiles for heavy trucks on a set of topographic road profiles,” SAE Paper 2006-01-1071, 2006.
- [12] O. F. Delgado, N. N. Clark, and G. J. Thompson, “Heavy duty truck fuel consumption prediction based on driving cycle properties,” *International Journal of Sustainable Transportation*, vol. 6, no. 6, pp. 338–361, 2012.
- [13] A. Fröberg and L. Nielsen, “Efficient drive cycle simulation,” *IEEE Transactions on Vehicular Technology*, vol. 57, no. 3, pp. 1442–1453, 2008.
- [14] U. Kiencke and L. Nielsen, *Automotive Control System: For Engine, Driveline, and Vehicle*, Springer, Berlin, Germany, 2nd edition, 2005.
- [15] E. Hellström, A. Fröberg, and L. Nielsen, “A real-time fuel-optimal cruise controller for heavy trucks using road topography information,” SAE Paper 2006-01-0008.
- [16] M. Ivarsson, J. Åslund, and L. Nielsen, “Look-ahead control—consequences of a non-linear fuel map on truck fuel consumption,” *Proceedings of the Institution of Mechanical Engineers D: Journal of Automobile Engineering*, vol. 223, no. 10, pp. 1223–1238, 2009.
- [17] R. Franke, P. Terwiesch, and M. Meyer, “An algorithm for the optimal control of the driving of trains,” in *Proceedings of the 39th IEEE Conference on Decision and Control*, vol. 3, pp. 2123–2128, Sydney, Australia, December 2000.
- [18] B. Saerens, J. Vandersteen, T. Persoons, J. Swevers, M. Diehl, and E. V. Bulck, “Minimization of the fuel consumption of a gasoline engine using dynamic optimization,” *Applied Energy*, vol. 86, no. 9, pp. 1582–1588, 2009.
- [19] D. E. Kirk, *Optimal Control Theory: An Introduction*, New York, NY, USA, Dover, 2004.
- [20] E. Hellström, *Look-Ahead Control of Heavy Trucks Utilizing Road Topography*, Linköping University, Linköping, Sweden, 2007.

Research Article

A Generalized Minimum Cost Flow Model for Multiple Emergency Flow Routing

Jianxun Cui, Shi An, and Meng Zhao

School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin 150090, China

Correspondence should be addressed to Shi An; anshi@hit.edu.cn

Received 29 March 2014; Accepted 27 May 2014; Published 22 June 2014

Academic Editor: Michael Lütjen

Copyright © 2014 Jianxun Cui et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

During real-life disasters, that is, earthquakes, floods, terrorist attacks, and other unexpected events, emergency evacuation and rescue are two primary operations that can save the lives and property of the affected population. It is unavoidable that evacuation flow and rescue flow will conflict with each other on the same spatial road network and within the same time window. Therefore, we propose a novel generalized minimum cost flow model to optimize the distribution pattern of these two types of flow on the same network by introducing the conflict cost. The travel time on each link is assumed to be subject to a bureau of public road (BPR) function rather than a fixed cost. Additionally, we integrate contraflow operations into this model to redesign the network shared by those two types of flow. A nonconvex mixed-integer nonlinear programming model with bilinear, fractional, and power components is constructed, and GAMS/BARON is used to solve this programming model. A case study is conducted in the downtown area of Harbin city in China to verify the efficiency of proposed model, and several helpful findings and managerial insights are also presented.

1. Introduction

Unfortunately, real-life situations such as floods, hurricanes, chemical accidents, nuclear accidents, terrorist attacks, and other events may occur and threaten the lives and the health of human beings [1]. Evacuation and emergency rescue are two major activities in disaster response. Evacuation involves relocating the threatened populations to safer areas (i.e., shelters) as soon as possible, whereas emergency rescue operations aim to dispatch various special vehicles (i.e., police cars, fire trucks, ambulances, etc.) to save lives, mitigate emergency situations, or conduct other special operations. These two types of flow will inevitably conflict with each other during the emergency response phase because they share the same spatial road network during the same time window. Usually, emergency vehicles have a higher priority to obtain the right-of-way when they share the same link or intersection with evacuation vehicles. Evacuation vehicles and other public vehicles must be stopped to reduce response times and enhance traffic safety when a rescue vehicle approaches. As many researchers have claimed, when evacuation is implemented, the evacuation traffic demand will

surge into the capacitated road network over a short period of time and consequently may cause significant transportation problems. Traffic delays during evacuation may range from inconvenient to catastrophic. Clearly, traffic delays will further be aggravated if evacuation flow and rescue flow conflict with each other in the same spatial and temporal space, especially when evacuation vehicles on the road may be often required to yield the right-of-way to rescue vehicles that use their warning devices. Therefore, emergency managers face a serious and difficult problem, namely, how to establish collaboration between evacuation flow and rescue flow in a shared and capacitated road network to mitigate negative influence due to conflicts. An obvious solution is to completely separate these two types of flow using selected management policies, such as setting up special emergency lanes for rescue vehicles to pass. The major drawback of this effort may be that the network capacity will not be fully explored because mixing of two flows is absolutely excluded. Thus, emergency managers must not only reduce conflict between two flows but also waste as little network capacity as possible.

The other issue commonly faced by emergency managers is how to implement contraflow design (also referred to

as lane reversal operations) during emergency response to fully explore the capacity of the current network. Contraflow design commonly refers to the shift of the normal driving directions of a subset or all danger-bound lanes for use by safety-bound evacuation traffic. Such control is based on the observation that danger-bound traffic is usually light, whereas evacuation traffic always oversaturates the safety-bound capacity. In nature, contraflow design constitutes a network redesign problem. If evacuation flow and rescue flow exist together in the same network, we can reshape the network to better serve these two flows.

Based on the previous discussion, this paper aims to present a generalized minimum cost flow model for collaboration between evacuation flow and rescue flow that also accounts for optimal contraflow design in a complex road network.

The paper is organized as follows. We first review related prior work, and the subsequent section presents the model formulation. The Baron solver is used to solve the nonconvex mixed-integer nonlinear programming model, and a local network in downtown Harbin City, Heilongjiang, China, is adopted to implement the case study. The paper concludes with a discussion of the results and areas for further research.

2. Literature Review

Over the past several decades, a wide range of network flow models has been introduced to describe different versions of routing and evacuation problems. The maximum dynamic flow problem introduced by Ford and Fulkerson [2] can be easily interpreted in the evacuation context as the evacuation of as many people as possible from a danger zone (source node) into a safe zone (sink node). The earliest arrival flow problem presented by Gale [3] is an extension of the maximum dynamic flow problem. The earliest arrival flow translated into the evacuation context means that the maximal amount of evacuees enters the safe area in each time period. To represent the evolution of a building evacuation process over time, Chalmet et al. [4] constructed a dynamic network flow model by expanding the network into a time-space network. The objective is to minimize the time until the last evacuee exits, which is known as the quickest flow problem. Following the same line of inquiry, Hamacher and Tufekci [5] extended the quickest flow problem to take into account different priority levels for different components of the evacuation network. Choi et al. [6] formulated three dynamic network flow problems for building evacuation (i.e., maximum flow, minimum cost, and the quickest flow problems) and introduced additional constraints to define link capacity as a function of the incoming flow rate. By formulating evacuation routing as a minimum cost flow problem, Dunn and Newton [7] proposed two algorithms for finding the set of path flows that minimize the total travel distance through a capacity-constrained network. Cova and Johnson [8] proposed the concept of lane-based routing to reduce intersection delays by temporarily transforming intersections into uninterrupted flow facilities using proper turning restrictions. As an extension of the minimal cost

flow problem, the model minimizes the total travel distance while preventing flow conflicts and restricting merge points at intersections. Miller-Hooks and Patterson [9] proposed the time-dependent quickest flow problem in time-varying capacitated evacuation networks in which link travel times and capacities vary with time. Network flow is modeled with flow conservation constraints at each node as well as link capacity constraints. As an extension of the time-dependent quickest flow problem, Opananon [10] addressed the stochastic nature of the evacuation network for a large building and formulated two network flow problems to generate the optimal path flows. The minimal cost problem seeks to minimize the total travel time if both link capacities and travel time are random variables with time-varying probability mass functions. In contrast, the safest escape problem aims to maximize the minimum path probability of successful arrivals at destinations on a network with a deterministic travel time and stochastic time-varying link capacities. Recently, Bretschneider and Kimms [1] presented a basic mathematical model for evacuation problems based on time-expanded network flow model that minimizes the evacuation time and prohibits conflicts within intersections.

Besides using network flow model to produce origin-destination routes and schedules of evacuees on each route, traffic assignment-simulation approaches are also employed to model evacuation problems. These approaches often use static/dynamic, deterministic/stochastic traffic assignment models to get evacuation flow pattern on road network (e.g., [11, 12]) and then use traffic simulation tools, such as DYNASMART [13] and DynaMIT [14], to conduct stochastic simulation of traffic movements based on origin-destination traffic demands and use queuing methods to account for road capacity constraints. However, it may take a long time to complete the simulation process for a large transportation network.

Lane reversal, which is also known as a contraflow lane, is another common topic during evacuation modeling. The effectiveness, feasibility, and safety issues of implementing lane reversal have been extensively discussed in, for example, MacDorman [15], Glickman [16], Hemphill and Surti [17], and Caudill and Kuo [18]. In evacuation cases, it has been suggested that the traffic direction of the inbound lanes of eligible roadway segments may be reversed in the case of an overwhelming flow of outbound traffic to increase the outbound capacity. Since the late 1990s, lane reversal has been widely used for hurricane evacuations in the states of the U.S. located on the Atlantic and Gulf Coasts [19]. The results from both evacuation practices [19, 20] and numerical studies [21–24] show that lane reversal has great potential to enhance evacuation performance and reduce traffic delays. It is worth noting that Xie and Turnquist [25] presented a lane-based evacuation bilevel programming model that integrates lane reversal and crossing elimination strategies in which network redesign and evacuation flow modeling are perfectly integrated. However, this research only aims to minimize the total evacuation time without considering that there may be multiple emergency flows that conflict with each other.

In summary, prior studies primarily formulate the evacuation networks in terms of facilities with limited capacity

in which traffic can travel through links with known travel times if they do not exceed the link capacity. These problems typically involve two types of network flow constraints, namely, flow conservation constraints at every node and capacity constraints for each link. However, certain traffic phenomena, for example, congestion-caused delays, are not captured in such models. Additionally, evacuation flow does not independently move during the emergency response phase and usually coincides with emergency rescue flow in the same spatial and temporal framework. Ensuring that all flows collaborate together could be a more practical step for the emergency management field. Few researches have been conducted to investigate multiple flow conflict during evacuation flow modeling and integrating contraflow lane design together, except the research presented by Xie and Turnquist [26], which integrated reversing lanes, eliminating intersection crossings, and reserving lanes for use by emergency vehicles together into one model. Although reserving a specific lane for emergency vehicles' exclusive use would improve the rescue efficiency, it may cause some extent of waste of road capacity, since the reserving lanes cannot be used for evacuation even though rescue flow is relatively low. In our research, instead of separating evacuation flow and rescue flow absolutely, we define conflict cost to account for the result of two flows mixing, which can be regarded as a general extension of Xie and Turnquist research [26].

3. Model Formulation

In this section, we model multiple emergency flow routing (MEFR). It is assumed that, in a specified emergency scenario, evacuation and rescue operations are both necessary, and contraflow design also can be implemented to explore further capacity of original road network in an emergency area. The emergency manager seeks to maximize the efficiency of the emergency response. The mathematical notation used in our models is introduced in Notation summary.

Before developing the model formulation, we first present selected illustrations of road network representation. In a specified emergency response network, several evacuation origins and destinations may exist as well as several rescue origins and destinations. The original network can be augmented with "virtual links" leading from each real-world origin/destination point to one common virtual origin/destination point [27]. All these virtual links are assumed to have infinite capacity and zero cost (i.e., zero travel time) so as not to influence the flow routing.

The MEFR problem described in this section is based on the minimum cost flow problem [28]. The min-cost flow problem occupies a central position among the network optimization models because it encompasses a broad class of applications [29]. The objective is to minimize the cost of transporting all supply (source nodes) to meet all demand (destination nodes) in a capacitated network. The MEFR model is an extension of min-cost flow problem. The objective of MEFR is to minimize the four components of the cost, that is, evacuation-flow time cost, rescue-flow time cost, conflict cost, and lane reversal cost.

During emergency operations, evacuation flow and rescue flow may conflict with each other, and we define this conflict as follows.

Conflicting Flows. If evacuation flow and emergency flow coexist on a same road link, they are treated as conflicting flows.

The inspiration for this definition originates from events that often occur during emergency response. If a rescue flow mixes with the evacuation flow, significant traffic delay may result. For example, if an emergency vehicle is required to rapidly pass through a link, evacuation flow on that link must be stopped or slowed to avoid conflict. It is usually true that rescue flow has a higher level of priority than evacuation flow, whereas evacuation flow possesses a relatively higher demand. It can be observed that this type of conflict (two types of flow coexisting on the same links) can cause large traffic delays or even serious disorder. The resulting cost due to emergency flow conflict should be proportional to the size of the coexisting evacuation and rescue flow on a specified link $(i, j) \in A$, which is presented as follows:

$$H = \theta \sum_{(i,j) \in A} x_{ij} y_{ij}, \quad (1)$$

where H is the total conflict cost resulting from two types of flow. For special cases, when $x_{ij} = 0$ (that is no evacuation flow on link (i, j)) or $y_{ij} = 0$ (that is no rescue flow on link (i, j)), the conflict cost on link (i, j) will be zero.

There are also two components of time cost during emergency response: the time cost of evacuation flow and the time cost of rescue flow, which are determined by formulas (2):

$$\begin{aligned} E &= \lambda_1 \sum_{(i,j) \in A} t_{ij} x_{ij}, \\ P &= \lambda_2 \sum_{(i,j) \in A} t_{ij} y_{ij}, \end{aligned} \quad (2)$$

where E is the total evacuation time cost and P is the total rescue time cost. In a classical min-cost flow problem, the travel times of the links are assumed to be constant, but, for road transportation problems, this assumption should be relaxed because the travel speed of the link will decrease together with the increase in link traffic flow. To account for this traffic phenomenon, we introduce the BPR function [30] to describe the link travel time.

The final component of cost is the lane reversal cost. It should be noted that during contraflow lane design, two links are required for use in operations, one for each direction. These two links are referred to as a "pair" of links, and operators will not reverse them concurrently [31]. Normally, each road will contain two links serving two different directions, and, naturally, they will be considered as paired links. If a given road is one way (i.e., only one link), we can imagine that there is also a virtual link serving the opposite direction with zero capacity, and these consequently will become paired links. Contraflow lane operations can also incur costs because human resources and certain essential devices are needed to

manage the contraflow lane to avoid confusing drivers. This cost should be proportional to the number of reversal lanes and is described as follows:

$$M = \frac{\delta \sum_{(i,j) \in A} |\text{NL}_{ij} - \gamma_{ij}|}{2}, \quad (3)$$

where M is the total line reversal cost and $\delta|\text{NL}_{ij} - \gamma_{ij}|$ is the lane reversal cost shared by link (i, j) and its pair link (j, i) .

Based on the discussion of the four components of emergency cost, a programming model is presented to describe the MEFR problem by extending the minimal cost flow as follows (MEFR):

$$\begin{aligned} \min_{x,y,\gamma} z = & \lambda_1 \sum_{(i,j) \in A} t_{ij} x_{ij} + \lambda_2 \sum_{(i,j) \in A} t_{ij} y_{ij} + \theta \sum_{(i,j) \in A} x_{ij} y_{ij} \\ & + \frac{\delta \sum_{(i,j) \in A} |\text{NL}_{ij} - \gamma_{ij}|}{2} \end{aligned} \quad (4)$$

s.t.

$$\sum_{j \in \xi(i)} x_{ij} - \sum_{k \in \zeta(i)} x_{ki} = m_i, \quad \forall i \in V \quad (5)$$

$$\sum_{j \in \xi(i)} y_{ij} - \sum_{k \in \zeta(i)} y_{ki} = n_i, \quad \forall i \in V \quad (6)$$

$$m_i = \begin{cases} +d_1, & \text{if } i \text{ is the virtual evacuation origin node} \\ 0, & \text{if } i \text{ is an intermediate transfer node} \\ -d_1, & \text{if } i \text{ is the virtual evacuation sink node} \end{cases} \quad (7)$$

$$n_i = \begin{cases} +d_2, & \text{if } i \text{ is the virtual rescue origin node} \\ 0, & \text{if } i \text{ is an intermediate transfer node} \\ -d_2, & \text{if } i \text{ is the virtual rescue sink node} \end{cases} \quad (8)$$

$$\gamma_{ij} + \gamma_{ji} = \text{NL}_{ij} + \text{NL}_{ji}, \quad \forall (i, j), (j, i) \in A \quad (9)$$

$$x_{ij} + y_{ij} \leq \bar{\mu}_{ij}, \quad \forall (i, j) \in A \quad (10)$$

$$t_{ij} = t_{ij}^0 \times \left[1 + \alpha \left(\frac{x_{ij} + y_{ij}}{\bar{\mu}_{ij}} \right)^\beta \right] \quad (11)$$

$$\bar{\mu}_{ij} = \frac{(\gamma_{ij} \times \mu_{ij})}{\text{NL}_{ij}} \quad (12)$$

$$x_{ij} = S_j^{e,o}, \quad i = O_e^*, \quad \forall j \in O_e \quad (13)$$

$$x_{ij} = R_j^{e,d}, \quad i = D_e^*, \quad \forall j \in D_e \quad (14)$$

$$y_{ij} = S_j^{r,o}, \quad i = O_r^*, \quad \forall j \in O_r \quad (15)$$

$$y_{ij} = R_j^{r,d}, \quad i = D_r^*, \quad \forall j \in D_r \quad (16)$$

$$x_{ij} \geq 0, \quad y_{ij} \geq 0; \quad \forall (i, j) \in A \quad (17)$$

$$\gamma_{ij} \in \text{Integer}; \quad \forall (i, j) \in A. \quad (18)$$



FIGURE 1: Road network for the case study.

In MEFR, x_{ij} , y_{ij} , and γ_{ij} for each $(i, j) \in A$ are decision variables. The objective of (1) is to minimize the total cost composed of the four components of cost in which t_{ij} is defined in Constraint (11). Constraint (5) is the evacuation flow conservation at each node, and Constraint (6) is the rescue flow conservation at each node in which m_i (evacuation flow at node i) and n_i (rescue flow at node i) are defined by Constraints (7) and (8). Constraint (9) guarantees that the summation of the lane numbers of two arbitrary “paired” links after contraflow design is equal to the summation of lane numbers before contraflow design. Constraint (10) illustrates that the summation of two flows on each link is always not greater than the link capacity after contraflow design in which $\bar{\mu}_{ij}$ is defined by Constraint (12). Constraints (13)–(16) define the flow of virtual links to guarantee that each evacuation/rescue origin can send a specified amount of flow, and each evacuation/rescue destination can receive a specified amount of flow. Constraint (17) states that the evacuation flow and rescue flow on each link are nonnegative. Constraint (18) means that the number of lanes on each link after the contraflow design must be an integer.

The MEFR is relatively difficult to solve because it is a nonconvex mixed-integer nonlinear programming (nonconvex MINLP) model with bilinear, fractional, and power components. The MINLP problems are difficult to solve because they combine all the difficulties of both of their subclasses: the combinatorial nature of mixed-integer programs (MIP) and the difficulty of solving nonconvex (and even convex) nonlinear programs (NLP). Because subclasses MIP and NLP are among the class of theoretically difficult problems (NP-complete), it is not surprising that solving the MINLP can be a challenging and daring venture [32]. In this paper, the branch-and-reduce optimization navigator (BARON), that is, a GAMS solver for the global solution of nonlinear (NLP) and mixed-integer nonlinear programs (MINLP), is used to solve the MEFR model.

4. Case Study

This section presents a case study of the MEFR model. The study area consists of the downtown area of Nangang

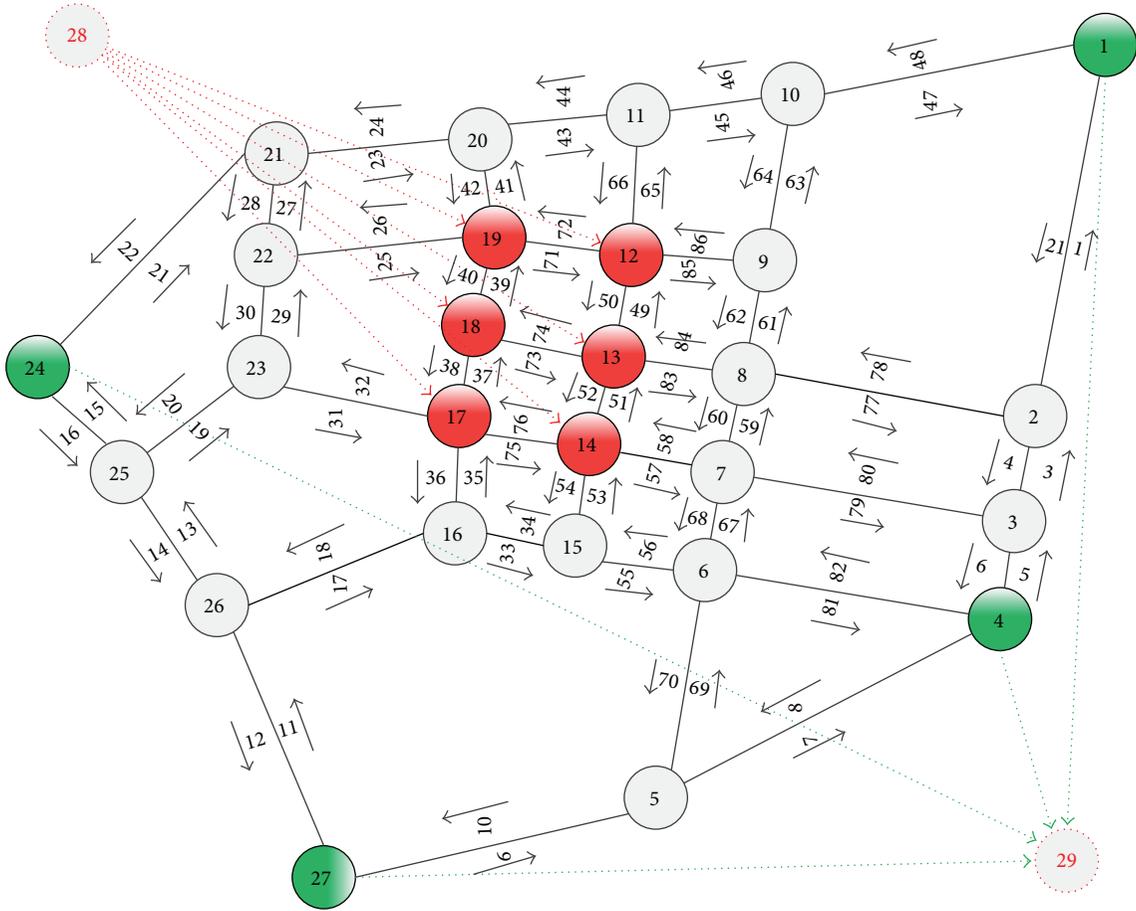


FIGURE 2: Network representation with virtual nodes and links.

District, Harbin City, China, which contains 27 intersections and 86 links (see Figure 1). This area includes a large-scale comprehensive international exhibition and sports center, the Heilongjiang Province TV Station, Wanda Plaza (the city’s comprehensive business center), large movie centers, shopping malls, hotels, hospitals, and many residential areas. Additionally, the density of roads is quite large in this area, including many main roads. The area is capable of generating many thousands of vehicle trips during a daytime evacuation and is therefore suitable for testing our model.

As stated in Section 3, we must transform the real-world road network into an arc-node network that integrates the virtual evacuation/rescue origin and the evacuation/rescue destination (see Figure 2). In Figure 2, the nodes shown in red (i.e., 12, 13, 14, 17, 18, and 19) are the hypothetical real-world evacuation/rescue origins, whereas the green nodes are the hypothetical real-world evacuation/rescue destinations (i.e., 1, 4, 24, and 27). In addition, the virtual origin node (i.e., 28) and the virtual destination node (i.e., 29) are introduced to generate and receive evacuation/rescue flows. All roads are bidirectional, and two links are paired for potential use in contraflow operations. The link IDs are depicted in Figure 2, and the characteristics of each link are presented in Table 1.

The evacuation flow demand and rescue flow demand in this hypothetical emergency scenario are presented in

Table 2 in which a positive number indicates origin flow and a negative number denotes sink flow.

We conducted four groups of experiments to test our model. In the first two groups, contraflow lane design and flow conflict are not considered together, that is, only contraflow lane or flow conflict is investigated independently in our model, whereas these flows are combined together in the other two groups. We aim to explore how the variations of the unit cost of lane reversal operations and flow conflict affect the four components of cost and total cost.

Group 1. In this group of tests, we do not consider flow conflict, which means that we do not emphasize the separation of evacuation flow and rescue flow. We aim to investigate how the unit cost of lane reversal operations affects the evacuation cost, rescue cost, lane reversal cost, and total cost. We define the values of the parameters as follows. The unit evacuation cost $\lambda_1 = 1$, the unit rescue cost $\lambda_2 = 3$, and the unit lane reversal cost δ increases 200 times from 0.1 in step sizes of 0.4. The results are shown in Figure 3.

We observe that, after the unit lane reversal cost reaches slightly more than 40, the lane reversal cost stabilizes at zero, and the evacuation cost, rescue cost, and total cost also stabilize at certain values. This result indicates that when the lane reversal cost becomes too large (i.e., larger than 40), we

TABLE 1: Characteristics of links.

Link ID	Number of lanes	t_0 (h)	Capacity (pch/h)	Link ID	Number of lanes	t_0 (h)	Capacity (pch/h)
1	3	0.022	3503.04	44	3	0.009	3503.04
2	3	0.022	3503.04	45	3	0.009	3503.04
3	3	0.005	3503.04	46	3	0.009	3503.04
4	3	0.005	3503.04	47	3	0.018	3503.04
5	3	0.005	3503.04	48	3	0.018	3503.04
6	3	0.005	3503.04	49	3	0.005	3503.04
7	3	0.024	3503.04	50	3	0.005	3503.04
8	3	0.024	3503.04	51	3	0.005	3503.04
9	5	0.016	4723.2	52	3	0.005	3503.04
10	5	0.016	4723.2	53	3	0.005	3503.04
11	5	0.018	5038.08	54	3	0.005	3503.04
12	5	0.018	5038.08	55	2	0.009	2634.66
13	5	0.009	5038.08	56	2	0.009	2634.66
14	5	0.009	5038.08	57	2	0.009	2634.66
15	5	0.008	5038.08	58	2	0.009	2634.66
16	5	0.008	5038.08	59	3	0.005	3503.04
17	2	0.018	2634.66	60	3	0.005	3503.04
18	2	0.018	2634.66	61	3	0.005	3503.04
19	2	0.012	2634.66	62	3	0.005	3503.04
20	2	0.012	2634.66	63	3	0.011	3503.04
21	3	0.021	3721.98	64	3	0.011	3503.04
22	3	0.021	3721.98	65	3	0.008	3503.04
23	3	0.011	3503.04	66	3	0.008	3503.04
24	3	0.011	3503.04	67	3	0.006	3503.04
25	2	0.015	2634.66	68	3	0.006	3503.04
26	2	0.015	2634.66	69	3	0.013	3503.04
27	2	0.006	2634.66	70	3	0.013	3503.04
28	2	0.006	2634.66	71	2	0.009	2634.66
29	2	0.008	2634.66	72	2	0.009	2634.66
30	2	0.008	2634.66	73	2	0.01	2634.66
31	2	0.014	2634.66	74	2	0.01	2634.66
32	2	0.014	2634.66	75	2	0.009	2634.66
33	2	0.007	2634.66	76	2	0.009	2634.66
34	2	0.007	2634.66	77	2	0.017	2479.68
35	3	0.006	3503.04	78	2	0.017	2479.68
36	3	0.006	3503.04	79	2	0.02	2634.66
37	3	0.005	3503.04	80	2	0.02	2634.66
38	3	0.005	3503.04	81	3	0.017	3503.04
39	3	0.006	3503.04	82	3	0.017	3503.04
40	3	0.006	3503.04	83	2	0.009	2634.66
41	3	0.006	3503.04	84	2	0.009	2634.66
42	3	0.006	3503.04	85	2	0.009	2634.66
43	3	0.009	3503.04	86	2	0.009	2634.66

TABLE 2: Evacuation and rescue flow demand (Veh/h).

	Origins						Destinations			
	12	13	14	17	18	19	1	4	24	27
Evacuation flow	2500	2400	2200	2100	3400	3200	-3900	-6000	-3500	-2400
Rescue flow	1600	500	400	190	210	200	-600	-700	-900	-900

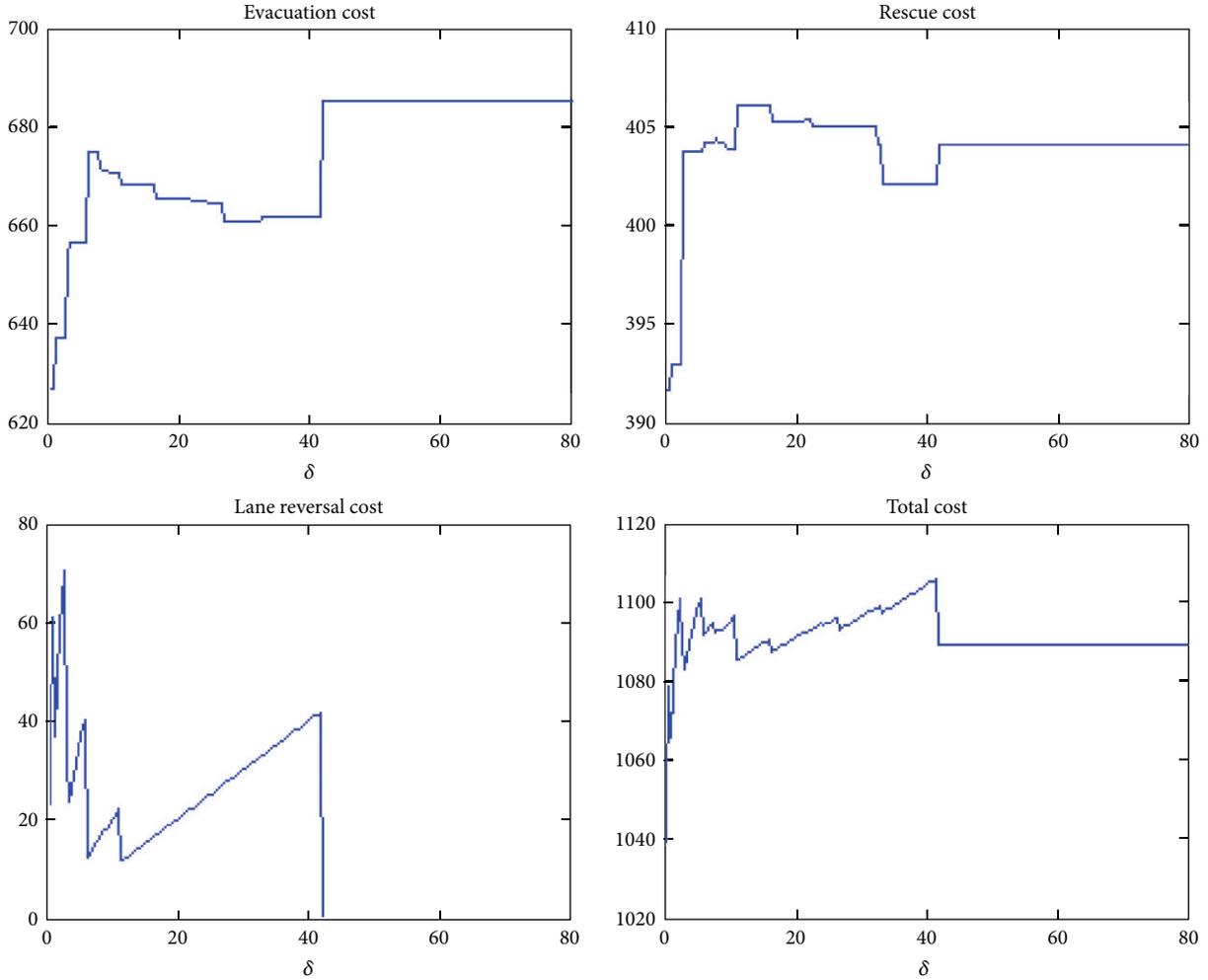


FIGURE 3: Results from tests of Group 1.

cannot use contraflow operation to obtain benefits because the reduction of the evacuation cost and rescue cost due to lane reversal operations will not produce an increase of the lane reversal cost. This finding is helpful for emergency managers to control lane reversal costs to a certain extent and obtain benefits from contraflow lane operations.

Group 2. In this group of tests, we do not consider contraflow lane operations, which means that we cannot change the network structure. We aim to investigate how the unit cost of flow conflict affects the evacuation cost, rescue cost, conflict cost, and total cost. The unit evacuation cost λ_1 and unit rescue cost λ_2 take on the same values as in Group 1. The unit flow conflict cost θ increases 200 times from 0 in step sizes of 0.000003. The results are shown in Figure 4.

In Figure 4, we observe the following three findings. (1) We find several intervals in which the flow conflict cost increases linearly while the evacuation cost and rescue cost do not change. This result indicates that in these intervals, the linear increase of θ does not further separate the evacuation flow and rescue flow, and only the flow conflict cost increases

linearly. This phenomenon also can be verified by the “total cost” curve in these intervals in which total cost linearly increases due to the linear increase of the conflict cost and the stabilization of the evacuation and rescue costs. (2) Additionally, at the beginning stage of varying unit conflict cost, an interesting phenomenon occurs in that the conflict cost decreases and increases alternately. This result may indicate that when the unit flow conflict cost θ increases, the flow conflict cost also increases because evacuation flow and rescue flow do not change routes. When θ increases to a certain extent (we refer to this as a “critical point”), the evacuation flow and rescue flow separate further, and the flow conflict cost decreases. Until θ is increased to the next critical point, the evacuation cost/rescue cost will remain constant, and the flow conflict cost linearly increases. (3) From the “conflict cost” and “total cost” curves, it can be observed that after certain periods of disturbances of the flow conflict cost and total cost, they increase linearly together. This result may indicate that, at the current evacuation and rescue demand, no matter how large θ becomes, we cannot obtain a solution that will separate the two types of emergency

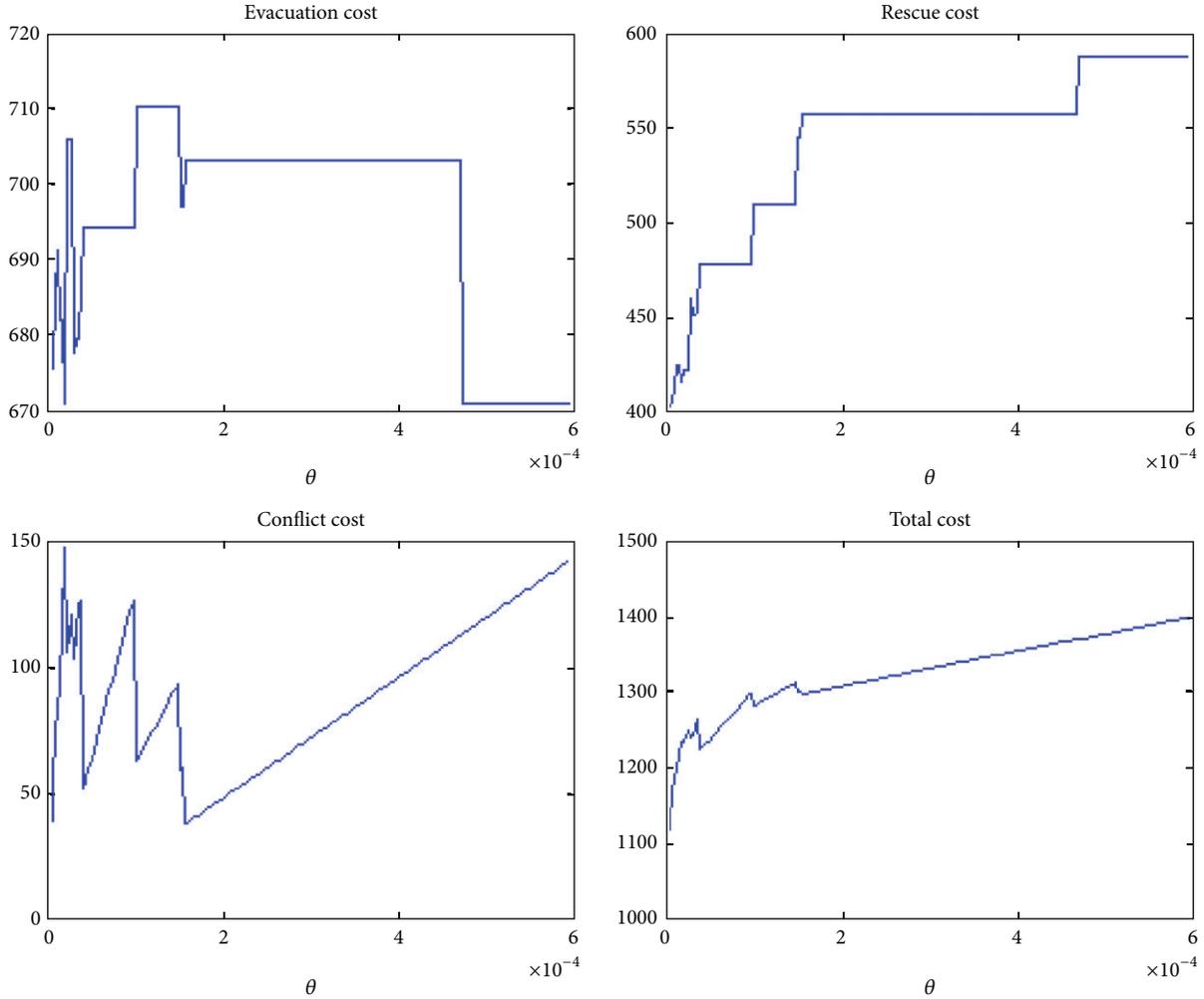


FIGURE 4: Results from tests of Group 2.

flow completely. This observation also inspires us to use lane reversal operations to improve this situation.

Group 3. In this group of tests, we consider contraflow lane operation and flow conflict together. We aim to investigate how the unit cost of flow conflict affects the evacuation cost, rescue cost, conflict cost, contraflow lane operation cost, and total cost. The unit evacuation cost λ_1 and unit rescue cost λ_2 take on the same values as in Group 1, whereas the unit lane reversal cost δ is equal to 0.4. The unit flow conflict cost θ increases 41 times from 0 in step sizes of 0.000003. The results are shown in Figure 5.

From the “evacuation and rescue cost” curve, we find that the evacuation cost and rescue cost vary nearly symmetrically. This result is true because, together with the increase in θ , the two types of flow will be separated increasingly further, and one of them will be “pushed out” from the original routes with less travel time to other routes with more travel time to avoid conflict. This observation leads to the fact that if the evacuation cost increases, the rescue cost will decrease and vice versa. From the “conflict cost” curve, it can be observed

that the conflict cost stabilizes at zero after θ increases to a certain extent. This phenomenon is rather interesting because it is completely different than the result from Group 2 in which the conflict cost does not stabilize but increases linearly together with the increase of θ . This observation indicates that the conflict between evacuation flow and rescue flow could be fully avoided using network redesign.

Group 4. In this group of tests, we consider the contraflow lane operation and flow conflict together. We aim to investigate how the unit cost of lane reversal affects the evacuation cost, rescue cost, conflict cost, contraflow lane cost, and total cost. The unit evacuation cost λ_1 and unit rescue cost λ_2 take on the same values as in Group 1, whereas the unit conflict cost θ is equal to 0.000003. The unit lane reversal cost δ increases 41 times from 0.1 in step sizes of 0.4. The results are shown in Figure 6.

From the “evacuation and rescue cost” curve, we find that the evacuation cost and rescue cost vary nearly with the same trend, which is quite different from the result in Group 3. This observation is reasonable because in Group 3, the variation

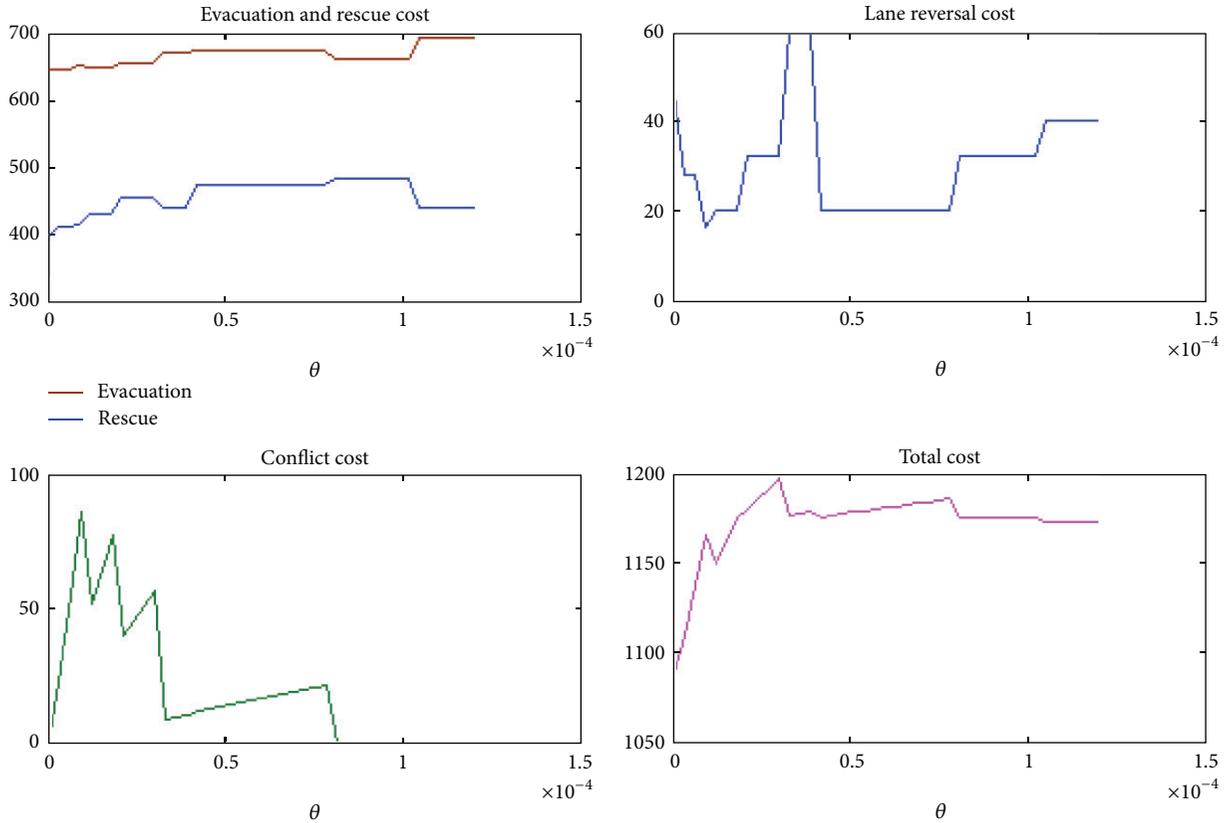


FIGURE 5: Results from tests of Group 3.

of θ will result in the separation of two flows such that the corresponding costs will change in different directions, but in Group 4, the increase and decrease of number of lane reversals together with the increase of δ will change the costs related to the two flows in the same direction. It also can be observed that, after δ becomes slightly larger than 10, the lane reversal cost stabilizes at zero, and the conflict cost, evacuation cost, rescue cost, and total cost stabilize together. The emergency manager can also use this finding to control lane reversal unit cost to maintain the efficiency of contraflow operations.

Selected Managerial Discussions. In emergency management practice, multiple emergency flows should be considered together and collaborate together in the same temporal and spatial framework. Identifying how best to utilize the current road network and route the different types of flow under emergencies in an efficient manner can be a challenging problem. The critical question related to this situation is how to determine the unit values of the four components of cost because the different values that they take on can lead to obviously different decision. For example, if the decision-makers emphasize no conflict between multiple flows, this means that θ will be relatively large, the capacity of the links will likely not be fully explored because various types of flow will independently occupy certain links to avoid conflict among them, and certain links may still contain much capacity. If θ

takes on a relatively small value, which occurs if managers do not pay much attention to conflict, this will lead to an absolutely different routing plan and contraflow operation strategy. Therefore, in practice, decision-makers should pay careful attention to deciding on the values of various unit costs to obtain a valuable and practical emergency solution.

Another interesting point is that the model presented in this research can be used in emergency flow cooperation scenarios but also can be used in other fields. A typical example is the assumption that a pipe network will be used for transportation of multiple materials. Certain materials are not allowed to conflict with each other or chemical reactivity will occur and cause damage, whereas other materials are allowed to conflict to a certain extent, but this will lead to an additional cost for separating the materials. The problem of how to decide on a plan to use the current pipe network to transport these materials is quite similar to the problem presented in this research.

5. Conclusions

This work studies a multiple emergency flow collaboration and optimal contraflow design (MEFR) problem in an urban highway network. A nonconvex mixed-integer nonlinear programming model is developed to describe this problem, which integrates four components of emergency cost:

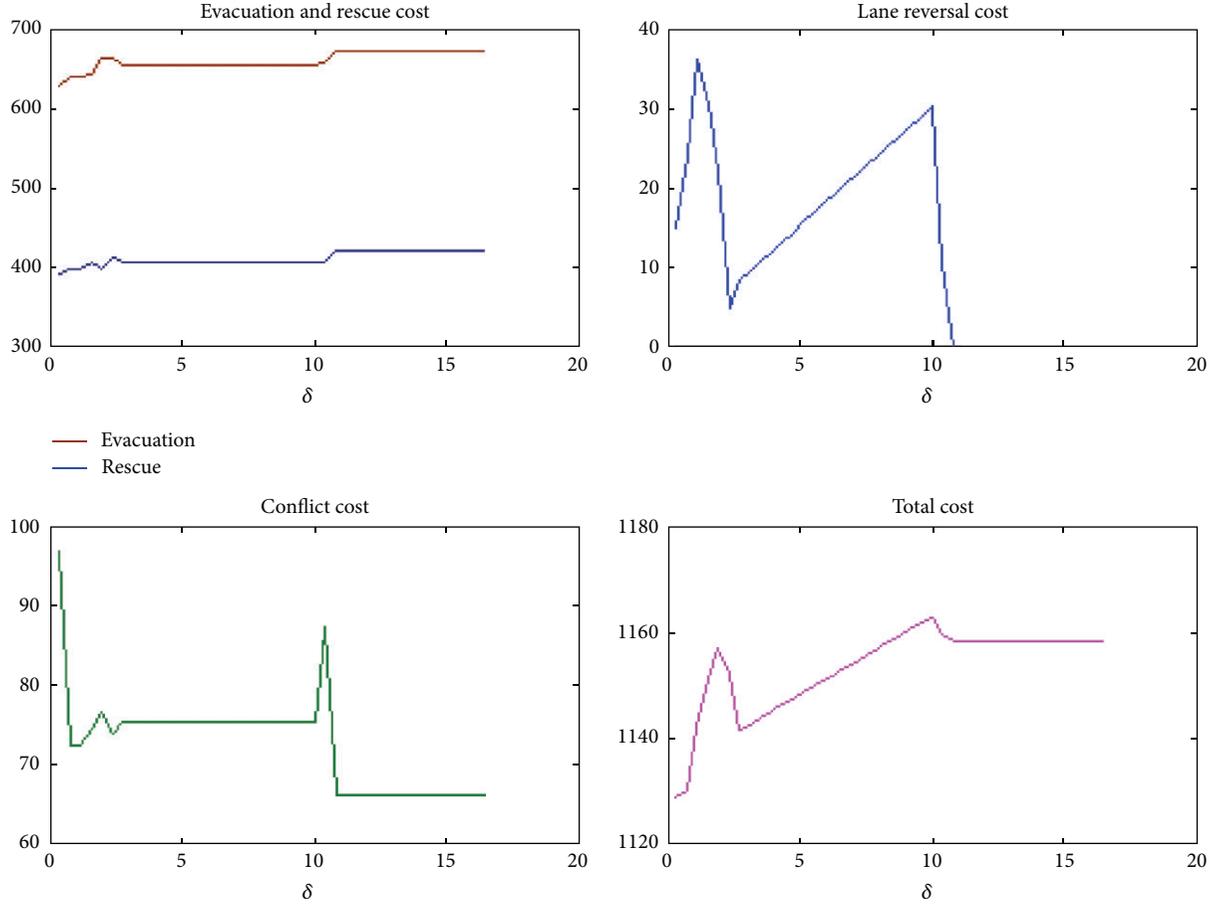


FIGURE 6: Results from tests of Group 4.

evacuation time cost, rescue time cost, flow conflict cost, and contraflow lane operation cost. Because this model is relatively difficult to solve, we use the GAMS/BARON solver to obtain solutions. Case studies in the downtown area of Harbin City in China are conducted to illustrate how our model operates. Various managerial insights are also drawn from the results as follows: (i) decision-makers should pay careful attention to determination of the values of various unit costs because these decisions will lead to completely different emergency strategies; (ii) decision-makers should properly control the lane reversal cost to guarantee that the operation of contraflow will produce a positive benefit; and (iii) different values of unit conflict cost can be used to adjust the separation level of multiple emergency flows.

In future work, an effective algorithm should be developed to solve the nonconvex MINLP model presented in this research. The solution of this model is relatively time consuming with an average cost of 15 minutes required to solve the model with BARON. Certain heuristic algorithms should be developed to achieve good balance between solution precision and solution efficiency.

Notation Summary

$G = (A, V)$:	A directed network with link set A and node set V
O_e^* :	Virtual evacuation origin node
D_e^* :	Virtual evacuation destination node
O_e :	Real-world evacuation origin node set;
D_e :	Real-world evacuation destination node set
O_r^* :	Virtual rescue origin node
D_r^* :	Virtual rescue destination node
O_r :	Real-world rescue origin node set
D_r :	Real-world rescue destination node set
x_{ij} :	Evacuation flow on link $(i, j) \in A$ (decision variable)
y_{ij} :	Rescue flow on link $(i, j) \in A$ (decision variable)
γ_{ij} :	Number of lanes of link $(i, j) \in A$ after optimal contraflow design (decision variable)
NL_{ij} :	Number of lanes of link $(i, j) \in A$ before optimal contraflow design

t_{ij}^0 :	The free flow travel time of link $(i, j) \in A$
t_{ij} :	Travel time of link $(i, j) \in A$
μ_{ij} :	Capacity of link $(i, j) \in A$ before optimal contraflow design
$\bar{\mu}_{ij}$:	Capacity of link $(i, j) \in A$ after optimal contraflow design
$\xi(i), \zeta(i)$:	The downstream nodes set and upstream nodes set of node $i \in V$
$S_j^{e,o}$:	Sending flow demand of evacuation origin node $j \in O_e$
$R_j^{e,d}$:	Receiving flow demand of evacuation destination node $j \in D_e$
$S_j^{r,o}$:	Sending flow demand of rescue origin node $j \in O_r$
$R_j^{r,d}$:	Receiving flow demand of rescue destination node $j \in D_r$
λ_1 :	Generalized cost of unit evacuation flow travel time
λ_2 :	Generalized cost of unit rescue flow travel time
θ :	Generalized cost of the conflict between unit rescue flow and unit evacuation flow
δ :	Generalized cost of the one lane reversal
d_1 :	Total evacuation demand flow; that is, $d_1 = \sum_{j \in O_e} S_j^{e,o} = \sum_{j \in D_e} R_j^{e,d}$
d_2 :	Total rescue demand flow; that is, $d_2 = \sum_{j \in O_r} S_j^{r,o} = \sum_{j \in D_r} R_j^{r,d}$
α, β :	Coefficients of the BPR function.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (Project no. 71203045), Heilongjiang Natural Science Foundation (Project no. E201318), and the Fundamental Research Funds for the Central Universities (Grant no. HIT.KISTP.201421). This work was performed at the Key Laboratory of Advanced Materials & Intelligent Control Technology on Transportation Safety, Ministry of Communications, China.

References

- [1] S. Bretschneider and A. Kimms, "A basic mathematical model for evacuation problems in urban areas," *Transportation Research A: Policy and Practice*, vol. 45, no. 6, pp. 523–539, 2011.
- [2] J. Ford and D. R. Fulkerson, "Maximal flow through a network," *Canadian Journal of Mathematics*, vol. 8, no. 3, pp. 399–404, 1956.
- [3] D. Gale, "Transient flows in networks," *Michigan Mathematical Journal*, vol. 6, no. 1, pp. 59–63, 1959.
- [4] L. G. Chalmet, R. L. Francis, and P. B. Saunders, "Network models for building evacuation," *Fire Technology*, vol. 18, no. 1, pp. 90–113, 1982.
- [5] H. W. Hamacher and S. Tufekci, "On the use of lexicographic min cost flows in evacuation modeling," *Naval Research Logistics*, vol. 34, pp. 487–503, 1987.
- [6] W. Choi, H. W. Hamacher, and S. Tufekci, "Modeling of building evacuation problems by network flows with side constraints," *European Journal of Operational Research*, vol. 35, no. 1, pp. 98–110, 1988.
- [7] C. E. Dunn and D. Newton, "Optimal routes in GIS and emergency planning applications," *Area*, vol. 24, no. 3, pp. 259–267, 1992.
- [8] T. J. Cova and J. P. Johnson, "A network flow model for lane-based evacuation routing," *Transportation Research A: Policy and Practice*, vol. 37, no. 7, pp. 579–604, 2003.
- [9] E. Miller-Hooks and S. S. Patterson, "On solving quickest time problems in time-dependent, dynamic networks," *Journal of Mathematical Modelling and Algorithms*, vol. 3, no. 1, pp. 39–71, 2004.
- [10] S. Opanason, *On finding paths and flows in multicriteria, stochastic and time-varying networks [Ph.D. dissertation]*, University of Maryland, 2004.
- [11] G. Ayfadopoulou, I. Stamos, E. Mitsakis, and J. Grau, "Dynamic traffic assignment based evacuation planning for CBD areas," *Procedia—Social and Behavioral Sciences*, vol. 48, pp. 1078–1087, 2012.
- [12] C. Xie, D.-Y. Lin, and S. Travis Waller, "A dynamic evacuation network optimization problem with lane reversal and crossing elimination strategies," *Transportation Research E: Logistics and Transportation Review*, vol. 46, no. 3, pp. 295–316, 2010.
- [13] H. S. Mahmassani, H. Sbayti, and X. Zhou, *DYNASmart-P Version 1.0 User's Guide*, Maryland Transportation Initiative, University of Maryland, 2004.
- [14] M. Ben-Akiva, D. McFadden, K. Train, J. Walker, C. Bhat, and M. Bierlaire, *Development of a Deployable Real-Time Dynamic Traffic Assignment System: DynaMIT And DynaMIT-P User's Guide*, Intelligent Transportation Systems Program, Massachusetts Institute of Technology, 2002.
- [15] L. C. MacDorman, "Case study in sensitivity of highway economic factors," *Highway Research Record*, vol. 100, pp. 2–19, 1965.
- [16] T. S. Glickman, *Optimal periodic control of reversible traffic operations [Ph.D. dissertation]*, Johns Hopkins University, 1970.
- [17] J. Hemphill and V. H. Surti, "A feasibility study of a reversible-lane facility for a Denver street corridor," *Transportation Research Record*, vol. 514, pp. 29–32, 1974.
- [18] R. J. Caudill and N. M. Kuo, "Development of an interactive planning model for contraflow lane evaluation," *Transportation Research Record*, pp. 47–54, 1983.
- [19] E. Urbina and B. Wolshon, "National review of hurricane evacuation plans and policies: a comparison and contrast of state practices," *Transportation Research A: Policy and Practice*, vol. 37, no. 3, pp. 257–275, 2003.
- [20] P. B. Wolshon and L. Lambert, *Convertible Roadways and Lanes: A Synthesis of Highway Practice*, vol. 340 of NCHRP Synthesis, National Cooperative Highway Research Program, Transportation Research Board, Washington, DC, USA, 2004.
- [21] H. Tuydes and A. Ziliaskopoulos, "Tabu-based heuristic approach for optimization of network evacuation contraflow," *Transportation Research Record*, no. 1964, pp. 157–168, 2006.
- [22] S. Kim, S. Shekhar, and M. Min, "Contraflow transportation network reconfiguration for evacuation route planning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 8, pp. 1115–1129, 2008.

- [23] G. L. Hamza-Lup, K. A. Hua, and R. Peng, "Leveraging e-transportation in real-time traffic evacuation management," *Electronic Commerce Research and Applications*, vol. 6, no. 4, pp. 413–424, 2007.
- [24] Q. Meng, H. L. Khoo, and R. L. Cheu, "Microscopic traffic simulation model-based optimization approach for the contraflow lane configuration problem," *Journal of Transportation Engineering*, vol. 134, no. 1, pp. 41–49, 2008.
- [25] C. Xie and M. A. Turnquist, "Lane-based evacuation network optimization: an integrated Lagrangian relaxation and tabu search approach," *Transportation Research C: Emerging Technologies*, vol. 19, no. 1, pp. 40–63, 2011.
- [26] C. Xie and M. A. Turnquist, "Integrated evacuation network optimization and emergency vehicle assignment," *Transportation Research Record*, no. 2091, pp. 79–90, 2009.
- [27] F. Yuan, *A proposed framework for simultaneous optimization of evacuation traffic distribution and assignment [Ph.D. dissertation]*, University of Tennessee, 2005.
- [28] L. R. Ford and D. R. Fulkerson, *Flows in Network*, Princeton University Press, Princeton, NJ, USA, 1962.
- [29] F. S. Hillier and G. J. Lieberman, *Introduction to Operations Research*, McGraw-Hill, New York, NY, USA, 1990.
- [30] S. C. Dafermos, "Sparrow the traffic assignment problem for a general network," *Journal of Research of the National Bureau of Standards B*, vol. 73, pp. 91–118, 1969.
- [31] Y. Liu, *An integrated optimal control system for emergency evacuation [Ph.D. dissertation]*, University of Maryland, 2007.
- [32] R. Östermark, "Concurrent processing of mixed-integer non-linear programming problems," *Kybernetes*, vol. 38, no. 6, pp. 966–989, 2009.

Research Article

Effects of Integration on the Cost Reduction in Distribution Network Design for Perishable Products

Z. Firoozi,¹ N. Ismail,¹ S. Ariafar,² S. H. Tang,¹ M. K. M. A. Ariffin,¹ and A. Memariani³

¹ Department of Mechanical and Manufacturing Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

² Department of Industrial Engineering, Faculty of Engineering, Shahid Bahonar University of Kerman, Kerman 7618891167, Iran

³ Department of Mathematics and Computer Science, University of Economic Sciences, Tehran 1593656311, Iran

Correspondence should be addressed to Z. Firoozi; zhr_firoozi@yahoo.com

Received 4 April 2014; Accepted 27 May 2014; Published 19 June 2014

Academic Editor: Michael Lütjen

Copyright © 2014 Z. Firoozi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Perishable products, which include medical and pharmaceutical items as well as food products, are quite common in commerce and industries. Developing efficient network designs for storage and distribution of perishable products plays a prominent role in the cost and quality of these products. This paper aims to investigate and analyze the impact of applying an integrated approach for network design of perishable products. For this purpose, the problem has been formulated as a mixed integer nonlinear mathematical model that integrates inventory control and facility location decisions. To solve the integrated model, a memetic algorithm (MA) is developed in this study. For verification of the proposed algorithm, its results are compared with the results of an adapted Lagrangian relaxation heuristic algorithm from the literature. Moreover, sensitivity analysis of the main parameters of the model is conducted to compare the results of the integrated approach with a decoupled method. The results show that as the products become more perishable, application of an integrated method becomes more reasonable in comparison with the decoupled one.

1. Introduction

Perishable products are very common in industries, commerce, and our daily life [1]. Major categories of perishable products include food products, medicines, pharmaceutical items, and many other plants and industrial goods. These products are only usable during their lifetime; when their lifetime is over, they must be discarded [2]. Nahmias [3] has classified perishable products based on their lifetime into two groups: (1) fixed-lifetime items that have a predetermined expiry date and (2) random-lifetime items for which there is no specified expiry date.

The most challenging nature of perishable products is their limited lifetimes that must be considered when deciding on the inventory control policies of these products [4–6]. However, traditional distribution network design models considered that products can be stored indefinitely in the stocking points (warehouses, distribution centers) of distribution networks. In other words, lifetime is not taken into

account in previous distribution network design modeling when deciding on inventory control policy of the network. Recently, perishable inventory control has gained much attention in distribution network design literature. Studies by Leśniewski and Bartoszewicz [7], Su et al. [8], Drezner and Scott [9], Firoozi et al. [10], and Coelho and Laporte [11] are some recent researches in this area.

Distribution network design (DND) problem, which is also known as supply chain network design, is one of the most promising areas in logistics and supply chain management [12, 13]. In a distribution network, items are produced at the manufacturers and shipped to the warehouses and then to the retailers to finally meet the customers' demand. Distribution network design plays a key role in the cost and quality of every final product [14], and this role becomes more critical when the products are perishable. There are a large number of distribution networks in the real world dealing with distribution and storage of perishable products [5, 6, 15, 16]. However, decisions to be made for designing a network are

highly interrelated [17, 18]. An obvious instance of this fact is the mutual effect of facility location and inventory control decisions of a distribution network, so that any change in facility location decisions, that is, the number and location of facilities, may influence the transportation and replenishment cost and, so, affects the optimal inventory policy. On the other hand, inventory policy determines the frequency of orders and, so, has some effects on the transportation cost. Due to the mentioned interrelationship, a huge amount of potential cost saving will be lost if relevant decisions are not optimized simultaneously [19].

In spite of the reasons counted above, meaning (1) existence of large numbers of distribution networks dealing with perishable products, (2) the difference between inventory control modeling of perishable and nonperishable items, and (3) the vast interrelation that exists between network design decisions, still, most of the network design models suffer from incorporation of perishable inventory control into other decisions of a distribution network. The model, which is developed in this study, therefore, aims to investigate the potential benefits derived from the integration of facility location and inventory control decisions of a distribution network that is responsible for distribution of perishable products. Hence, an integrated and a decoupled model are developed for the network design in this study, where the integrated model optimizes network design decisions simultaneously and the decoupled method optimizes decisions sequentially. Sensitivity analysis is conducted to show if the value of integration could be affected by the lifetime of products.

2. Literature Review

A distribution network normally consists of suppliers, retailers, and distribution centers (DCs) or warehouses. In order to take the advantages of risk pooling, each DC receives demands from several retailers and places order to the supplier. The inventory is kept by the DCs to meet the demands of retailers. The objective of distribution network design models is to determine the optimal number and location of DCs, allocation of retailers to DCs, and the ordering cycle and frequency of orders of DCs so that the total cost is minimized [20].

Daskin et al. [21] introduced one of the most well-known inventory-location models known as the location model with risk pooling (LMRP). The model integrated inventory and safety stock decisions with uncapacitated facility location model (UFLP). A Lagrangian relaxation heuristic was developed to solve the model. It was assumed in this model that the mean-to-variance ratios were identical for all retailers. In addition to that, the model assumed an identical lead time between supplier and DCs. The LMRP model was also studied by Shen et al. [22], but using a set partitioning approach for solving the model.

LMRP became the basis of many consecutive network design models. Several capacitated versions of LMRP were developed by Miranda and Garrido [23], Ozsen et al. [17], and Miranda and Garrido [24]. Shen [25] extended LMRP by considering multiple products for the model. Shu et al. [26]

removed the assumption of identical mean-to-variance ratio for demands of retailers.

Sourirajan et al. [27, 28] developed the LMRP by removing the assumption of identical lead times between supplier and distribution centers (DCs). Qi and Shen [29], Shavandi and Bozorgi [30], and Atamtürk et al. [31] studied the effects of uncertainty on network design decisions. Gebennini et al. [32] developed a dynamic version of LMRP, and Melo et al. [33] studied the redesigning of a distribution network.

Despite the existence of a large number of distribution networks that are dealing with the distribution of perishable products, integration of perishable inventory models with other network design decisions has not been considered in the distribution network design studies. Therefore, the aim of this study is to evaluate the effects of integrating network design decisions of a distribution network that is dealing with perishable products. The results of this study also help to answer the question as to whether the value of integration is affected by the length of the lifetime of products.

3. Problem Definition and Modeling

The distribution network considered in this study consists of one supplier, a set of retailers, and a set of distribution centers. In order to take the advantages of risk pooling, distribution centers order products from the supplier and store the inventory of products to satisfy the stochastic demands of retailers. In other words, the inventories of the retailers, which are assigned to a DC, are aggregated in that DC. This way results in a decrease in safety stock inventory of the network [34]. The products are perishable and have a limited lifetime. The objective is to estimate the possible cost saving that could be achieved by applying an integrated approach for network design instead of a decoupled approach. For this purpose, an integrated and a decoupled approach are developed and are compared with each other. The decoupled approach first determines the number and location of DCs and allocation of retailer to DCs, and then finds the optimal inventory policy of the DCs. However, the integrated approach optimizes these decisions simultaneously. The remainder of this section describes the integrated and the decoupled approaches. Moreover, the notation used to model the problem is listed at the end of the paper.

Effect of Lifetime on the Inventory Policy. Traditional EOQ inventory models compute the ordering cycle by the formula Q/D , where Q is the order quantity and D is the total mean demand. However, if products have a lifetime less than the ordering cycle, this formula cannot be used because, in that case, some products may meet their lifetime, while they are still keeping in the DCs. To prevent this situation the ordering cycle is required to be restricted in such a way that it does not exceed the products lifetime. So, if products lifetime starts as soon as they left the supplier to the DCs, then when products are delivered to DCs, they have lost a part of their lifetime equal to supplier-DC lead time. On the other hand, according to Figure 1, that shows the profile of inventory versus time in EOQ (Q, r) policy, the longest time that a product stays in a DC (or warehouse) equals the ordering cycle plus a time

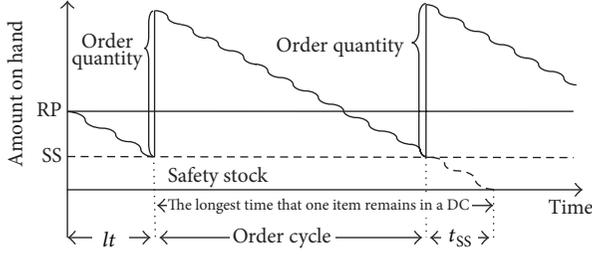


FIGURE 1: Profile of inventory level over time.

shown by t_{SS} , where t_{SS} is the period of time that takes for safety stock to be totally replaced by fresh inventory. The value of t_{SS} is computed by SS_i/D_i . Therefore, we can write

$$\frac{Q_i}{D_i} + \frac{SS_i}{D_i} \leq pt_i - lt_i. \quad (1)$$

On the left-hand side of inequality (1), the first term equals the ordering cycle, and the second term is t_{SS} . The terms appearing on the right-hand side of this inequality are, respectively, lifetime and lead time. Inequality (1) can be rewritten as follows:

$$Q_i \leq (pt_i - lt_i) D_i - SS_i. \quad (2)$$

Constraint 2, that restricts the order quantity, will be considered in both integrated and decoupled approaches.

3.1. Cost Components. The following section describes the cost components that occur in the network design, comprising holding inventory and safety stock cost, ordering cost, transportation cost, and fixed installation cost of DCs.

Inventory Holding Cost. The total cost of holding inventory in a DC is determined by (3), where h_i is the unit inventory holding cost of DC $_i$ and the terms $Q_i/2$ and $Z_\alpha \sqrt{lt_i} \sqrt{V_i}$, respectively, calculates the average working inventory and safety stock of a DC:

$$\frac{h_i Q_i}{2} + Z_\alpha h_i \sqrt{lt_i} \sqrt{V_i}. \quad (3)$$

Ordering Cost. The annual cost of placing orders by DC $_i$ to the supplier is computed by

$$\frac{O_i D_i}{Q_i}, \quad \forall i \in I, \quad (4)$$

where D_i/Q_i determined the total orders placed by DC $_i$ in a year.

Transportation Cost. The total fixed and variable transportation cost from the supplier to DCs and from DCs to retailers is computed by (5). The first term in this equation is the fixed transportation cost and the second term is the variable transportation cost:

$$\frac{A_i D_i}{Q_i} + TD_i (dis_{ij} + t_{DC-su}), \quad \forall i \in I. \quad (5)$$

Fixed Setup Cost. The annual setup cost of DC $_i$ is calculated by

$$F_i x_i, \quad \forall i \in I, \quad (6)$$

where x_i is a binary variable that is equal to 1 if DC $_i$ is established; otherwise, it is equal to 0.

3.2. Integrated Approach. The integrated model optimizes location allocation and inventory decisions simultaneously. The objective function and constraints of this model are provided in the following:

$$\begin{aligned} \text{Min} \quad & \sum_i h_i \cdot \left(\frac{Q_i}{2} + Z_\alpha \sqrt{lt_i} \sqrt{\sum_j (v_j y_{ij})} \right) \\ & + \sum_i \sum_j (O_i + A_i) \frac{(d_j y_{ij})}{Q_i} + \sum_i F_i x_i \end{aligned} \quad (7)$$

$$+ \sum_i \sum_j T (dis_{ij} + t_{DC-su}) d_j y_{ij}$$

$$\text{s.t.} \quad \sum_i y_{ij} = 1, \quad \forall j \in J \quad (8)$$

$$x_i \geq y_{ij}, \quad \forall i \in I, \forall j \in J \quad (9)$$

$$\begin{aligned} Q_i \leq & (pt_i - lt_i) \sum_j (d_j y_{ij}) - Z_\alpha \sqrt{lt_i} \sqrt{\sum_j (v_j y_{ij})}, \\ & \forall i \in I. \end{aligned} \quad (10)$$

$$Q_i \geq 0, \quad x_i, y_{ij} \in \{1, 0\}, \quad \forall i \in I, \forall j \in J. \quad (11)$$

Since the integrated model optimizes the facility location and inventory decisions simultaneously, the mean and variance of demand of a DC are not known in advance. Therefore, in the integrated model, mean and variance of demand of DC $_i$ are, respectively, written in the form of $D_i = \sum_j (d_j y_{ij})$ and $V_i = \sum_j (v_j y_{ij})$, where d_j and v_j are, respectively, mean and variance of demand of retailer $_j$. Then, binary variables of y_{ij} determine whether retailer $_j$ should be assigned to DC $_i$ or not. Moreover, in this model the first term is the holding inventory cost. The second term is the fixed ordering and shipment cost. The third term is the DC setup cost, and the last term is the variable transportation cost. Constraint set (8) specifies that each retailer can only be assigned to one DC. Constraint set (9) guarantees that retailers are only assigned to open DCs. Constraint set (10) makes sure that products are not kept in a DC for a time longer than their lifetime, and constraint set (11) specifies that x_i, y_{ij} are binary variables and Q_i is a nonnegative value.

3.3. Decoupled Approach. The decoupled approach is a two-stage procedure in which the first stage applies the classical UFLP model to determine the configuration of the network taking into account the transportation and DCs' installation

cost. Then the acquired configuration is given to the second stage where the order quantities and ordering cycles of DCs are determined. This two-stage approach is presented in the following and is solved by Lingo 12.0 software:

first stage:

$$\begin{aligned}
 \text{Min} \quad & \sum_i F_i x_i + \sum_i \sum_j T (dis_{ij} + t_{DC-su}) d_j y_{ij} \\
 \text{s.t.} \quad & \sum_i y_{ij} = 1, \quad \forall j \in J \\
 & x_i \geq y_{ij}, \quad \forall i \in I, \forall j \in J \\
 & x_i, y_{ij} = \{1, 0\}, \quad \forall i \in I, \forall j \in J,
 \end{aligned} \tag{12}$$

second stage:

$$\begin{aligned}
 \text{Min} \quad & \sum_i h_i \cdot \left(\frac{Q_i}{2} + Z_\alpha \sqrt{lt_i} \sqrt{V_i} \right) + \sum_i \sum_j (O_i + A_i) \frac{D_i}{Q_i} \\
 \text{s.t.} \quad & Q_i \leq (pt_i - lt_i) \sum_j (d_j y_{ij}) - Z_\alpha \sqrt{lt_i} \sqrt{\sum_j (v_j y_{ij})}, \\
 & \forall i \in I \\
 & Q_i \geq 0, \quad \forall i \in I.
 \end{aligned} \tag{13}$$

4. Solution Method for Integrated Approach

In this study, a memetic algorithm (MA) is developed to solve the integrated nonlinear mixed model described in Section 3.2. The efficiency of the developed algorithm is measured by comparing its results in terms of the total cost with a Lagrangian relaxation algorithm developed by Firoozi et al. [10]. The following sections describe the proposed memetic algorithm (MA).

4.1. Memetic Algorithm. Memetic algorithm (MA) is a metaheuristic algorithm that is a hybrid of an evolutionary framework (such as genetic algorithm) and local search algorithms [35]. MA has been applied successfully to solve various optimization problems. The memetic algorithm developed in this study is a hybrid of a genetic algorithm (GA) and a local research. GA is a stochastic metaheuristic that is inspired by the principles of genetics and natural selection [36]. It initially generates a population of chromosomes each representing a possible solution to the problem. Each chromosome consists of a number of genes that encode representations of a part of the solution. A number of chromosomes from the current population (called parents) are selected based on some selection rules and undergo mutation and crossover to make offspring. Other than offspring produced by crossover and mutation, elitism strategy selects the best fitted chromosomes in terms of the fitness function to survive in the next generation (new population). This strategy is to protect the search from losing the best found solutions so far. The combination

R ₁	R ₂	R ₃	R ₄	R ₅	R ₆
5	5	4	1	4	4

FIGURE 2: Chromosome representation.

of crossover, mutation, and elitism improves the population over generations until the fittest member of the population represents the optimal or a near optimal solution [28]. Size of the population remains constant over the generations. The following subsection describes the structure of chromosomes and the way of producing them.

4.1.1. Chromosome Representation. In this paper, an integer vector encoding scheme is applied for chromosome representation. This encoding is similar to the one considered by Diabat et al. [37] in which the length of a chromosome equals the number of retailers. The i th gene of the chromosome shows which DC supplies the i th retailer. Figure 2 displays a possible chromosome for a distribution network consisting of 5 potential distribution centers and 6 retailers. According to this figure the first and second retailers are supplied by distribution center number five; the third, fifth, and the last retailers are supplied by distribution center number four; and the retailer number four is supplied by distribution center number one.

A very important advantage of this kind of encoding is that it always ensures that the single sourcing assignment, that is, constraint (8), is satisfied. More importantly the length of chromosome (number of genes) is equal to the number of retailers. Hence, the time complexity of the algorithm will be less in comparison to other encoding that has $m \times n$ genes, where n is the number of DCs and m is the number of retailers.

The only concern about this encoding is that the produced chromosome might be infeasible with regard to constraint set (10), meaning there is at least one open distribution center in a generated solution (chromosome) with an order quantity greater than the right-hand side of constraint set (10). To settle this problem, the algorithm first obtains the value of order quantity, Q , for all DCs that are open in a given chromosome. This is done by taking the derivative of objective function (7) with respect to Q_i and solving it for Q_i , where Q_i is the order quantity of DC _{i} :

$$Q_i = \sqrt{\frac{2(A_i + O_i)D_i}{h_i}}. \tag{14}$$

If the order quantity of DC _{i} obtained by (14) satisfies constraint set (10), then this Q_i is considered as the optimal order quantity for DC _{i} . However, if there is at least one DC whose order quantity violates constraint (10), the condition would be different. In that case, since objective function (7) is convex, the optimal value of Q_i will occur at the border of the

interval. So, the optimal order quantity is obtained by setting constraint (10) equal to zero, as shown in the following:

$$Q_i = (pt_i - lt_i) \sum_j (d_j y_{ij}) - Z_\alpha \sqrt{lt_i} \sqrt{\sum_j (v_j y_{ij})}. \quad (15)$$

Another issue that arises here is that (15) may result in a negative order quantity for a DC. In such a case, the infeasible chromosome would be ignored, and the algorithm keeps running until the desired number of chromosomes that satisfy the nonnegativity condition of the order quantity for all DCs is generated. These two feasibility conditions are checked for every new chromosome generated during the search process of the algorithm.

4.1.2. Elitism. During the search process, the algorithm always keeps the best found solutions and stores them into a so-called elite population. The number of individuals in the elite population is one of the input parameters that must be tuned. Each time that an iteration is executed the elite population is updated.

4.1.3. Crossover and Mutation Operators. Crossover and mutation operators protect the search from being trapped into local optimal solutions [38]. In the crossover, two parents are combined and two offspring are produced. This operator chooses a similar position at random along the two parents' chromosomes and swaps the portions located after the positions. A sample for this operator is displayed in Figure 3. In this algorithm, parents for crossover operation are selected by tournament selection method. This method has a parameter of tournament size. A tournament size of n means " n " chromosomes are randomly selected from the current population, and the best of them in terms of the cost function is selected as one of the parents required for crossover operation. In contrast to crossover, in mutation only one parent is required and one offspring is generated. This operator chooses a number of genes at random along the chromosome and exchanges the value of selected genes for another feasible value (a number between 1 and the number of DCs). In this algorithm, parents for mutation are randomly chosen from the elite population. Figure 4 displays how mutation generates a new chromosome.

4.1.4. Improvement Algorithms. A five-step improvement heuristic is embedded into the GA to enhance the search. The improvement heuristic is executed on the elite and the mutation population. The steps of this algorithm are selected randomly for each individual and each step is run on an individual until no more improvement could be achieved. In each step of this algorithm, if a better solution is attained in terms of the total cost, the old solution is replaced with the

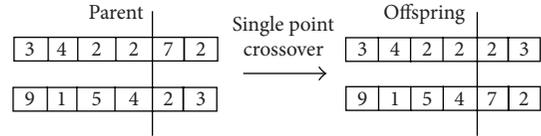


FIGURE 3: A sample of crossover operator.

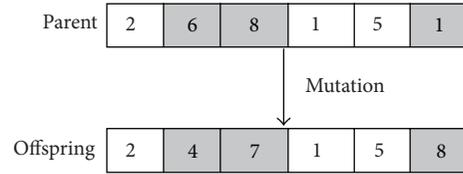


FIGURE 4: A sample of mutation operator.

better one. Otherwise, the previous solution does not change. The algorithm steps are as follows.

Step 1. This step assigns a retailer to other DCs other than the one that is currently assigned to.

Step 2. This step considers two DCs and exchanges their retailers.

Step 3. This step considers one retailer from a DC and one retailer from another DC and then swaps the retailers' assignment.

Step 4. This step considers a DC and assigns all of its retailers to another DC (either an open or a close DC).

Step 5. This step considers a DC and assigns all of its retailers to other randomly selected DCs (not all retailers to one DC).

4.2. MA Procedure. This section described how MA developed in this paper works. This algorithm randomly generates the initial population. The initial population is considered as the current population for the algorithm. Then the elite population that consists of the best members of the current population is made. The mutation population is generated afterward by randomly selecting individuals from the elite population and mutating the genes of selected individuals. Parents for crossover are selected from the current population by tournament selection method. Then an improvement local search heuristic described in Section 4.1.4 modifies the elite and mutation population, and a new population will be made by three operators (crossover, mutation, and elitism). The current population then is replaced with the new one. The algorithm keeps running until the maximum number of iterations is achieved. The stages performed by GA and MA are displayed in Figure 5.

5. Validation of the Proposed Algorithm

This section conducts numerical experiments to investigate the performance of the developed MA. A total number of 36 test problems are generated by varying the parameters

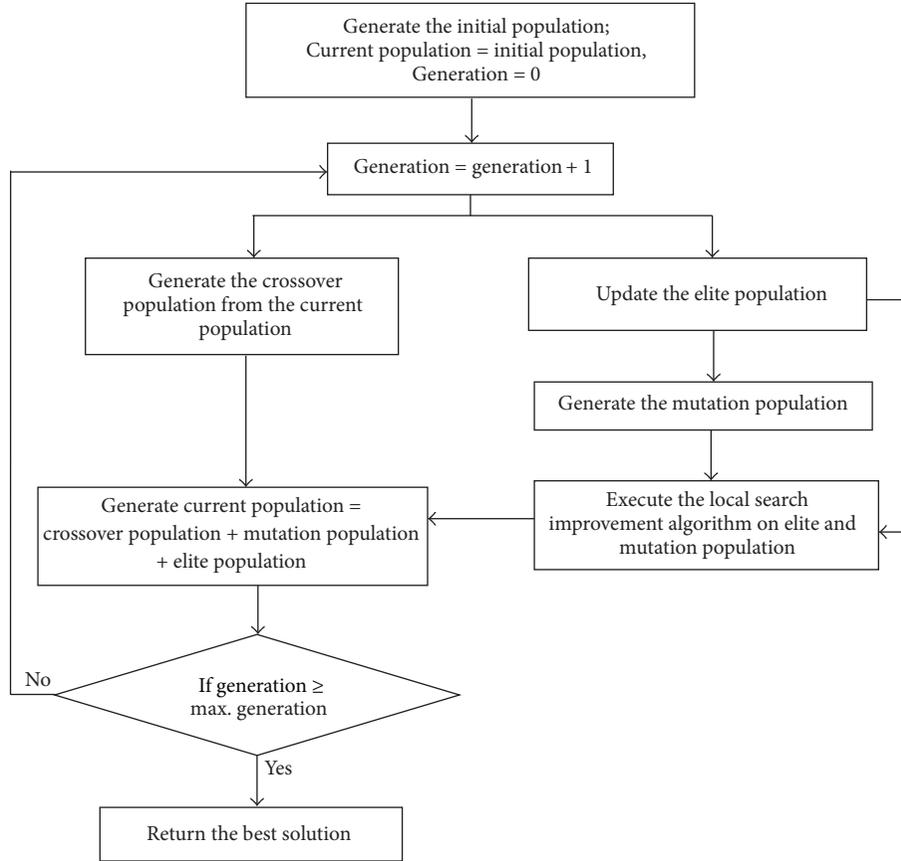


FIGURE 5: MA procedure.

TABLE 1: Memetic algorithm parameters.

Data set	Population size	Crossover percentage	Mutation percentage	Elitism percentage	Number of generations
15-node data set	10	20%	40%	40%	30
49-node data set	15	26.3%	20.3%	53.3%	60

of 15-node and 49-node data sets from Daskin [39]. These test problems are generated by considering three amounts for the inventory cost and fixed ordering cost and two amounts for products' lifetime. The results obtained by MA are compared with the results of a Lagrangian relaxation algorithm developed by Firoozi et al. [10]. Parameters of the memetic algorithm and Lagrangian relaxation are shown in Tables 1 and 2, respectively. The means and variances of retailers' demands are selected to be the same as the demand parameters of Daskin [39]. Distances between retailers are calculated using the great circle distance formula, based on the longitude and latitude of retailers' locations. Fixed installation costs are set to the fixed installation costs, as considered by Daskin [39], but divided by 10. Variable transportation costs are set to 0.4 units of cost and lead time is set to 1 day. Tables 3 and 4 compare the results of the two methods. As the results show, the gap between the results of two methods is zero for all the generated test problems.

6. Results and Discussion

This section is to investigate the average cost saving obtained by applying the integrated approach instead of a decoupled approach. The amount of cost saving or the value of integration is computed by the following formula:

$$\begin{aligned}
 &\text{Value of integration} \\
 &= \frac{\text{results of sequential approach} - \text{results of integrated approach}}{\text{results of integrated approach}} \\
 &\quad * 100.
 \end{aligned} \tag{16}$$

To find the average value of integration, sensitivity analysis is conducted on four main parameters of the problem including inventory holding cost, ordering cost, products' lifetime, and variances of demands. Sensitivity analysis investigates how parameters of the problem influence the value of integration.

TABLE 2: Lagrangian relaxation parameters.

Parameters	15-node	49-node
Maximum iterations	500	1500
Number of nonimproving iterations before halving step size	30	30
Initial value of step size	2	2
Minimum value of step size	10^{-7}	10^{-7}
Initial value of Lagrangian multiplier	$10(\bar{d} + \bar{f})$	$10(\bar{d} + \bar{f})$
Maximum optimality gap $((UB - LB) / UB) * 100\%$	0.1%	0.1%

\bar{d}, \bar{f} are, respectively, the average demands of retailer and the average fixed installation costs of the DCs.

TABLE 3: Comparison of results for 15-node data set.

Input data			MA result	Lagrangian relaxation results	Gap
h	a	pt			
1	1	3	1274300	1274300	0
1	10	3	1307100	1307100	0
1	100	3	1561000	1561000	0
10	1	3	1314300	1314300	0
10	10	3	1411400	1411400	0
10	100	3	1739100	1739100	0
100	1	3	1504200	1504200	0
100	10	3	1811300	1811300	0
100	100	3	2782500	2782500	0
1	1	11	1274300	1274300	0
1	10	11	1305000	1305000	0
1	100	11	1402300	1402300	0
10	1	11	1314300	1314300	0
10	10	11	1411400	1411400	0
10	100	11	1718500	1718500	0
100	1	11	1504200	1504200	0
100	10	11	1811300	1811300	0
100	100	11	2782500	2782500	0

TABLE 4: Comparison of results for 49-node data set.

Input data			MA result	Lagrangian relaxation results	Gap
h	a	pt			
1	1	3	412400	412400	0
1	10	3	506490	506490	0
1	100	3	1367300	1367300	0
10	1	3	490510	490510	0
10	10	3	686850	686850	0
10	100	3	1627700	1627700	0
100	1	3	863870	863870	0
100	10	3	1467900	1467900	0
100	100	3	3431300	3431300	0
1	1	11	411870	411870	0
1	10	11	472400	472400	0
1	100	11	694980	694980	0
10	1	11	490500	490500	0
10	10	11	681490	681490	0
10	100	11	1286800	1286800	0
100	1	11	863870	863870	0
100	10	11	1467800	1467800	0
100	100	11	3377800	3377800	0

The base case to perform the tests on is a 49-node data set from Daskin [39]. Four mentioned parameters of this case are designed carefully over a wide range to make a total of 1715 different test problems. These total cases are generated by varying the variance of demand, ordering cost, product lifetime and inventory holding cost. For this purpose, the variances of demands are altered from 0.25 to 1.75 times of their initial values, in steps of 0.25. The ordering cost is changed from 0 to 600 units of cost in steps of 100. The product lifetime is varied from 3 to 7 days, in steps of 1. Finally, the inventory holding cost is changed from 2 to 152 units of cost in steps of 25. For each parameter the average value of integration is calculated and analyzed.

In this section, demand parameters are selected from Daskin [39]. Distances between retailers are calculated using the great circle distance formula based on the longitude and latitude of locations provided by Daskin [39]. Fixed setup costs are selected from Daskin [39] but divided by 10. Total fixed transportation and ordering cost is set to 500 units of cost. Variable transportation cost is set to 50 units of cost.

Inventory holding cost is set to 5 units of cost, and lead time is set to 1 day.

Figure 6 shows the average value of integration versus changes in the products lifetime and holding inventory cost. This graph consists of seven curves; each one is corresponding to a different level of holding inventory cost. Each point in this graph is the average of 49 test problems that are rendered by keeping holding inventory cost and product lifetime by a constant amount and varying the other two parameters. It is observed from this figure that as the lifetime of products becomes shorter, that is, as the products become more perishable, the value of integration increases dramatically. Additionally, it is evident from Figure 6 that the value of integration increases as the holding inventory cost increases. The same conclusion can be derived when fixed ordering cost is changing instead of inventory holding cost as shown in Figure 7.

The average cost reduction observed by varying the variances of demands and products' lifetime are plotted in Figure 8. As this figure shows, the value of integration gets

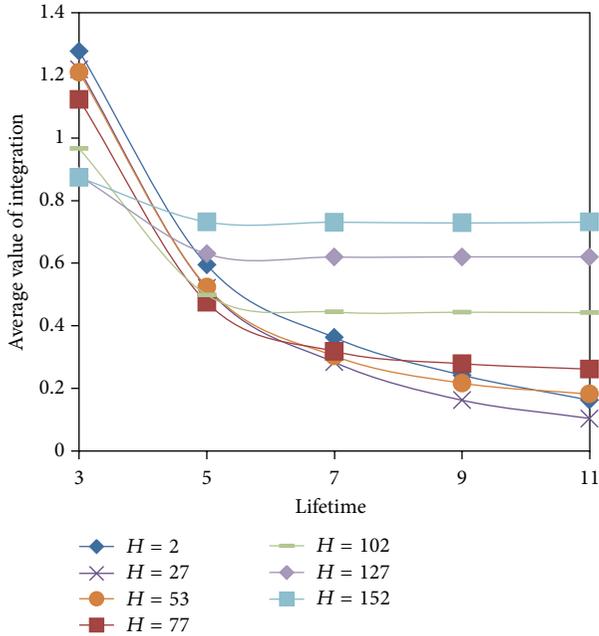


FIGURE 6: Value of integration versus change in lifetime and holding inventory cost.

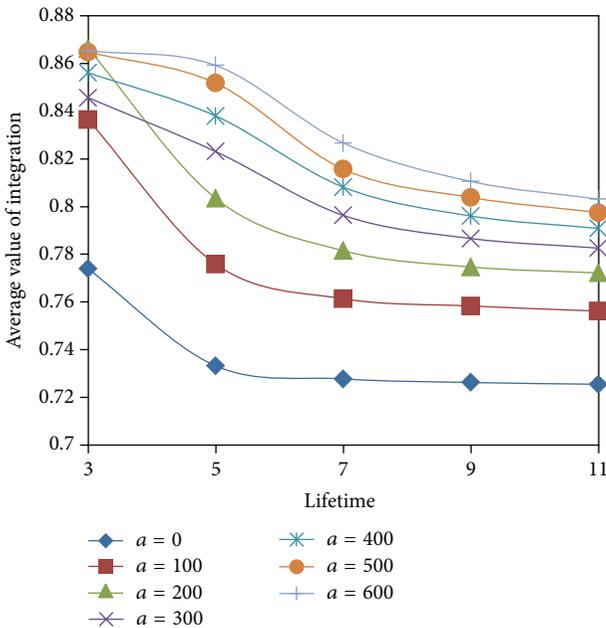


FIGURE 7: Value of integration versus change in lifetime and fixed ordering cost.

higher as the variances of demand increase, although this rise is not very noticeable. However, it is very evident from the figure that the average cost saving is directly dependent on the products' lifetime in such a manner that the integration value increases as the products lifetime decreases.

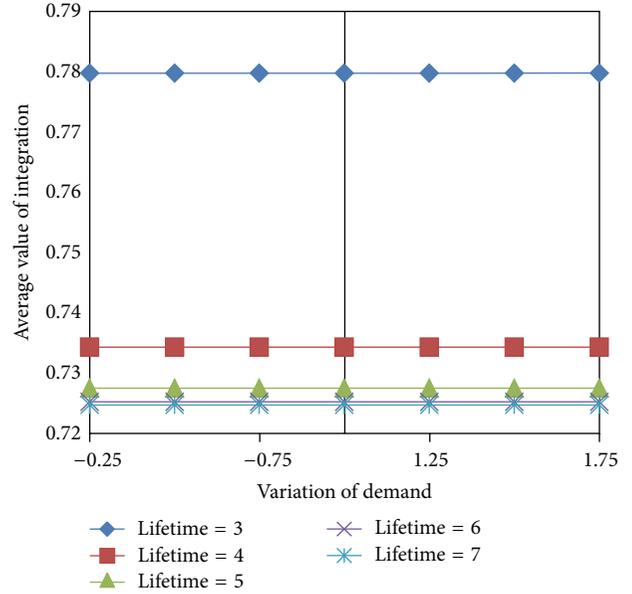


FIGURE 8: Value of integration versus changes in lifetime and variance of demand.

7. Conclusion

The impact of an integrated approach for optimizing network design decisions has been investigated in this study. For this purpose, the network design problem of perishable products was formulated as an integrated and a decoupled model. The integrated model optimized inventory control and location allocation decisions of the network simultaneously. However, the decoupled model optimized location allocation decisions first and inventory control decisions afterward. On key parameters of the problem, sensitivity analysis was performed to obtain the value of the integration. It was evident from the comparison of the results that the average cost saving obtained by integrated approach increased as the lifetime decreased. In other words, the more perishable the product was, the more valuable the integrated approach was. Furthermore, the performance of the developed MA was compared with a Lagrangian relaxation from the literature.

The models that are developed in this study are for fixed-lifetime perishable items, that is, items with known expiry dates, like processed food, dairy products, or many industrial items. For future works, it would be interesting to extend the results of this study to random-lifetime perishable items for which there is no specified expiry date.

Notation

Sets

J : Set of retailers

I : Set of candidate DC locations.

Indices

i : Index for DCs
 j : Index for retailers.

Input Parameters

F_i : Annual fixed setup cost for DC $_i$
 T : Transportation cost per unit of product per unit of distance
 t_{DC-su} : Per item transportation cost from the supplier to a DC
 A_i : Per shipment transportation cost from supplier to DC $_i$
 h_i : Inventory holding cost at DC $_i$ per unit of product per year
 O_i : Fixed ordering cost per order placed by DC $_i$ to the supplier
 d_j : Annual mean demand of retailer $_j$
 D_i : Annual mean demand of DC $_i$
 v_j : Variance of annual demand for retailer $_j$
 V_i : Variance of annual demand for DC $_i$
 dis_{ij} : Distance between DC $_i$ and retailer $_j$
 lt_i : Lead time in terms of year from the supplier to DC $_i$
 pt_i : Lifetime of product at DC $_i$
 α : Level of service that has to be achieved at the retailers
 Z_α : Standard normal deviation such that $P(z \leq z_\alpha) = \alpha$.

Decision Variables

Q_i : Order quantity of DC $_i$
 y_{ij} : Binary variable, taking the value 1 if retailer $_j$ is assigned to DC $_i$ and 0 otherwise
 x_i : Binary variable, taking the value 1 if DC $_i$ is open and 0 otherwise.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] H. Dong and G. Jiang, "Limit distribution of inventory level of perishable inventory model," *Mathematical Problems in Engineering*, vol. 2011, Article ID 329531, 9 pages, 2011.
- [2] P. Amorim, H. Meyr, C. Almeder, and B. Almada-Lobo, "Managing perishability in production-distribution planning: a discussion and review," *Flexible Services and Manufacturing Journal*, vol. 25, no. 3, pp. 389–413, 2013.
- [3] S. Nahmias, "Perishable inventory theory: a review," *Operations Research*, vol. 30, no. 4, pp. 680–708, 1982.
- [4] C. Kouki, E. Sahin, Z. Jemaï, and Y. Dallery, "Assessing the impact of perishability and the use of time temperature technologies on inventory management," *International Journal of Production Economics*, vol. 143, no. 1, pp. 72–85, 2013.
- [5] S. Minner and S. Transchel, "Periodic review inventory-control for perishable products under service-level constraints," *OR Spectrum*, vol. 32, no. 4, pp. 979–996, 2010.
- [6] J. Jia and Q. Hu, "Dynamic ordering and pricing for a perishable goods supply chain," *Computers and Industrial Engineering*, vol. 60, no. 2, pp. 302–309, 2011.
- [7] P. Leśniewski and A. Bartoszewicz, "LQ optimal sliding mode control of periodic review perishable inventories with transportation losses," *Mathematical Problems in Engineering*, vol. 2013, Article ID 325274, 9 pages, 2013.
- [8] J. Su, J. Wu, and C. Liu, "Research on coordination of fresh produce supply chain in big market sales environment," *The Scientific World Journal*, vol. 2014, Article ID 873980, 12 pages, 2014.
- [9] Z. Drezner and C. H. Scott, "Location of a distribution center for a perishable product," *Mathematical Methods of Operations Research*, vol. 78, no. 3, pp. 301–314, 2013.
- [10] Z. Firoozi, N. Ismail, S. Ariaifar, S. H. Tang, M. K. A. M. Ariffin, and A. Memariani, "Distribution network design for fixed lifetime perishable products: a model and solution approach," *Journal of Applied Mathematics*, vol. 2013, Article ID 891409, 13 pages, 2013.
- [11] L. C. Coelho and G. Laporte, "Optimal joint replenishment, delivery and inventory management policies for perishable products," *Computers & Operations Research*, vol. 47, pp. 42–52, 2014.
- [12] P. A. Miranda and R. A. Garrido, "Inventory service-level optimization within distribution network design problem," *International Journal of Production Economics*, vol. 122, no. 1, pp. 276–285, 2009.
- [13] M. Faccio, A. Persona, F. Sgarbossa, and G. Zanin, "Multi-stage supply network design in case of reverse flows: a closed-loop approach," *International Journal of Operational Research*, vol. 12, no. 2, pp. 157–191, 2011.
- [14] F. Dabbene, P. Gay, and N. Sacco, "Optimisation of fresh-food supply chains in uncertain environments—part II: a case study," *Biosystems Engineering*, vol. 99, no. 3, pp. 360–371, 2008.
- [15] C. I. Hsu, S. F. Hung, and H. C. Li, "Vehicle routing problem with time-windows for perishable food delivery," *Journal of Food Engineering*, vol. 80, no. 2, pp. 465–475, 2007.
- [16] R. A. C. M. Broekmeulen and K. H. van Donselaar, "A heuristic to manage perishable inventory with batch ordering, positive lead-times, and time-varying demand," *Computers and Operations Research*, vol. 36, no. 11, pp. 3013–3018, 2009.
- [17] L. Ozsen, C. R. Coullard, and M. S. Daskin, "Capacitated warehouse location model with risk pooling," *Naval Research Logistics*, vol. 55, no. 4, pp. 295–312, 2008.
- [18] Z. Firoozi, S. H. Tang, S. Ariaifar, and M. K. A. M. Ariffin, "An optimization approach for a joint location inventory model considering quantity discount policy," *Arabian Journal for Science and Engineering*, vol. 38, no. 4, pp. 983–991, 2013.
- [19] P. A. Miranda and R. A. Garrido, "Valid inequalities for Lagrangian relaxation in an inventory location problem with stochastic capacity," *Transportation Research E: Logistics and Transportation Review*, vol. 44, no. 1, pp. 47–65, 2008.
- [20] Z. J. M. Shen, "Integrated supply chain design models: a survey and future research directions," *Journal of Industrial and Management Optimization*, vol. 3, no. 1, pp. 1–27, 2007.
- [21] M. S. Daskin, C. R. Coullard, and Z. J. M. Shen, "An inventory-location model: formulation, solution algorithm and computational results," *Annals of Operations Research*, vol. 110, pp. 83–106, 2002.

- [22] Z. J. M. Shen, C. Coullard, and M. S. Daskin, "A joint location-inventory model," *Transportation Science*, vol. 37, no. 1, pp. 40–55, 2003.
- [23] P. A. Miranda and R. A. Garrido, "Incorporating inventory control decisions into a strategic distribution network design model with stochastic demand," *Transportation Research E: Logistics and Transportation Review*, vol. 40, no. 3, pp. 183–207, 2004.
- [24] P. A. Miranda and R. A. Garrido, "A simultaneous inventory control and facility location model with stochastic capacity constraints," *Networks and Spatial Economics*, vol. 6, no. 1, pp. 39–53, 2006.
- [25] Z. J. M. Shen, "A multi-commodity supply chain design problem," *IIE Transactions*, vol. 37, no. 8, pp. 753–762, 2005.
- [26] J. Shu, C. P. Teo, and Z. J. M. Shen, "Stochastic transportation-inventory network design problem," *Operations Research*, vol. 53, no. 1, pp. 48–60, 2005.
- [27] K. Sourirajan, L. Ozsen, and R. Uzsoy, "A single-product network design model with lead time and safety stock considerations," *IIE Transactions*, vol. 39, no. 5, pp. 411–424, 2007.
- [28] K. Sourirajan, L. Ozsen, and R. Uzsoy, "A genetic algorithm for a single product network design model with lead time and safety stock considerations," *European Journal of Operational Research*, vol. 197, no. 2, pp. 599–608, 2009.
- [29] L. Qi and Z. J. M. Shen, "A supply chain design model with unreliable supply," *Naval Research Logistics*, vol. 54, no. 8, pp. 829–844, 2007.
- [30] H. Shavandi and B. Bozorgi, "Developing a location-inventory model under fuzzy environment," *International Journal of Advanced Manufacturing Technology*, vol. 63, no. 1–4, pp. 191–200, 2012.
- [31] A. Atamtürk, G. Berenguer, and Z. J. M. Shen, "A conic integer programming approach to stochastic joint location-inventory problems," *Operations Research*, vol. 60, no. 2, pp. 366–381, 2012.
- [32] E. Gebennini, R. Gamberini, and R. Manzini, "An integrated production-distribution model for the dynamic location and allocation problem with safety stock optimization," *International Journal of Production Economics*, vol. 122, no. 1, pp. 286–304, 2009.
- [33] M. T. Melo, S. Nickel, and F. Saldanha-Da-Gama, "A tabu search heuristic for redesigning a multi-echelon supply chain network over a planning horizon," *International Journal of Production Economics*, vol. 136, no. 1, pp. 218–230, 2012.
- [34] S. Gaur and A. R. Ravindran, "A bi-criteria model for the inventory aggregation problem under risk pooling," *Computers and Industrial Engineering*, vol. 51, no. 3, pp. 482–501, 2006.
- [35] F. Neri and C. Cotta, "Memetic algorithms and memetic computing optimization: a literature review," *Swarm and Evolutionary Computation*, vol. 2, pp. 1–14, 2012.
- [36] R. L. Haupt and S. E. Haupt, *Practical Genetic Algorithms*, Wiley-Interscience, New York, NY, USA, 2004.
- [37] A. Diabat, T. Aouam, and L. Ozsen, "An evolutionary programming approach for solving the capacitated facility location problem with risk pooling," *International Journal of Applied Decision Sciences*, vol. 2, no. 4, pp. 389–405, 2009.
- [38] J. Fathali, "A genetic algorithm for the p -median problem with pos/neg weights," *Applied Mathematics and Computation*, vol. 183, no. 2, pp. 1071–1083, 2006.
- [39] M. Daskin, *Network and Discrete Location: Models, Algorithms and Applications*, vol. 48, Palgrave Macmillan, Basingstoke, UK, 1997.

Research Article

A Dynamic Optimization Strategy for the Operation of Large Scale Seawater Reverses Osmosis System

Aipeng Jiang, Jian Wang, Wen Cheng, Changxin Xing, and Shu Jiangzhou

School of Automation, Hangzhou Dianzi University, Hangzhou, Zhejiang 310018, China

Correspondence should be addressed to Aipeng Jiang; jiangaipeng@gmail.com

Received 29 January 2014; Accepted 21 May 2014; Published 18 June 2014

Academic Editor: Michael Lütjen

Copyright © 2014 Aipeng Jiang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this work, an efficient strategy was proposed for efficient solution of the dynamic model of SWRO system. Since the dynamic model is formulated by a set of differential-algebraic equations, simultaneous strategies based on collocations on finite element were used to transform the DAOP into large scale nonlinear programming problem named Opt2. Then, simulation of RO process and storage tanks was carried element by element and step by step with fixed control variables. All the obtained values of these variables then were used as the initial value for the optimal solution of SWRO system. Finally, in order to accelerate the computing efficiency and at the same time to keep enough accuracy for the solution of Opt2, a simple but efficient finite element refinement rule was used to reduce the scale of Opt2. The proposed strategy was applied to a large scale SWRO system with 8 RO plants and 4 storage tanks as case study. Computing result shows that the proposed strategy is quite effective for optimal operation of the large scale SWRO system; the optimal problem can be successfully solved within decades of iterations and several minutes when load and other operating parameters fluctuate.

1. Introduction

The shortage of freshwater resources is expected to worsen with the growth of population and industrialization, as well as climate change [1, 2]. Seawater desalination is one of the most promising approaches to get freshwater resources in the world and is considered as the most important strategy to develop new freshwater for the coastal counties [3, 4]. With the process of low-cost, high rejection membrane technique and high efficiency energy recovery device, seawater reverse osmosis (SWRO) technique is becoming the most popular and attractive seawater technique for its economy and convenience [5–7].

Recently, more attention was paid to optimal operation and energy management to further reduce the energy consumption of SWRO system [8]. Modeling methods based on first principle and data-driven are used for practical control and fault diagnosis [9]. And with development of advanced control techniques such as predictive control, sliding-mode control, and optimal control [10–13], more potential benefit can be expected.

Kim et al. comprehensively studied the RO process especially on the minimization of product cost with system engineering method [14]. Sassi and Mujtaba and Palacin et al. evaluated the optimal operation of SWRO system through minimizing the specific energy consumption [15, 16]; with the consideration of thermodynamic restriction, Zhu studied the energy cost optimization problem. Among which factors such as stages and energy recovery efficiency were discussed to get the optimal operation condition [17]. After a comprehensive first-principle based mathematical model has been developed and validated by plant data, Li studied the optimal plant operation of brackish water reverse osmosis (BWRO) desalination to reduce specific energy consumption (SEC). His computing results show that about 16% reduction of SEC can be achieved by optimizing operating condition [18]. Since all the cost spent on SWRO system includes not only operational cost but also capital cost, Geraldes studied the optimization problem through minimizing the objective function including all these costs; but to make solution of the problem with DAEs easier and simple, Geraldes used differential technology to discretize the equations and then

obtained the optimal results with low accuracy [19]. Since the operating environments change frequently, Sassi and Mujtaba studied the optimal operation problem which is the variation of load and feed temperature. In his study, storage tank was used as the buffering unit between water production and water supply, and water level of storage tank can be freely adjusted to add the flexibility of operation [20].

The researches above are of great significance to improve the economical performance of desalination, but most of them put emphasis on the optimal operation of steady-state process. As we know, in the actual process many operating parameters such as feed temperature, electricity price, and water supply requirement change dynamically over time [21]; the performance of equipment and reverse osmosis membrane also constantly change over time, so dynamic optimization to improve the operation will be more meaningful. However, since the dynamical optimization problem includes a set of differential and algebraic equations (DAEs), its efficient and stable solution is fairly tough when inequality constraints enforced on the bounds. Direct methods such as variational principle based method often failed because of their disadvantage in dealing with active inequalities [22, 23].

Though with development of computing technique, simultaneous approach with collocation on finite element is more suitable for this kind of problems, and large scale solvers such as sparse SQP and IPOPT become more powerful to solve large scale nonlinear problems, there are still many works to do to make the solution more efficient and stable, especially when the optimal problem is not convex and with the characteristic of strong nonlinear. But the real time optimization need the optimal operation problem be solved efficiently. In this paper, to solve optimization problem of RO process system with network structure more efficiently, based on the well-developed dynamic models of RO process and water storage process, an efficient optimization strategy to optimize the dynamic operation of SWRO system is proposed, which will be helpful to the reduction of energy cost and to the realization of real time optimization.

2. Dynamic Model of SWRO System

According to the flowsheet of SWRO system, RO unit is the key part for freshwater production. After the pretreatment, the seawater was pumped into RO modules by high pressure pump, from which the freshwater and the salt were separated through the work of solution-diffusion; then with simple posttreatment, the freshwater was pumped to terminal user [19] to meet the needs of daily life. In the RO module, the pressure, flow rate, and concentration change along the channel (shown in Figure 1). These variables satisfy the following equations [19]:

$$\frac{dV}{dz} = -\frac{2Jv}{h_{sp}}$$

$$z = 0, \quad V = V_f; \quad z = z_f; \quad V = V_r;$$

$$\frac{dP_d}{dz} = -\lambda \frac{\rho}{d_e} \frac{V^2}{2}$$

$$z = 0, \quad P_d = 0; \quad z = z_f; \quad P_r = P_f - P_d; \quad (1)$$

$$\frac{dC_b}{dz} = \frac{2Jv}{h_{sp}V} (C_b - C_p)$$

$$z = 0, \quad C_b = V_f; \quad z = z_f, \quad C_b = C_r;$$

here, Jv denotes solvent flux, and V represent the axial velocity in feed channel. C_b is the bulk concentration of RO module, and P_d is the pressure drop along the RO module. The parameters and variables from the above equations can be obtained from the solution-diffusion mass transport relations and principle of energy conservation; all the equations involved are listed as follows [19, 24–27].

RO process model equations are

$$Q_p = Q_f - Q_r$$

$$Q_f C_f = Q_r C_r + Q_p C_p \quad (2)$$

$$Jv = A_w (P_f - P_d - P_p - \Delta\pi).$$

See [24]

$$Js = B_s (C_m - C_{sp})$$

$$P_b = P_f - P_d$$

$$\Delta P = (P_b - P_p) \quad (3)$$

$$A_w = A_{w0} \exp\left(\alpha_1 \frac{T - 273}{273} - \alpha_2 (P_f - P_d)\right).$$

See [19]

$$B_s = B_{s0} \exp\left(\beta_1 \frac{T - 273}{273}\right)$$

$$\Delta\pi = RT (C_m - C_p)$$

$$\phi = \frac{C_m - C_p}{C_b - C_p} = \exp\left(\frac{Jv}{k_c}\right) \quad (4)$$

$$Sh = \frac{k_c d_e}{D_{AB}} = 0.065 Re^{0.875} Sc^{0.25}.$$

See [27]

$$Re = \frac{\rho V d_e}{\mu}$$

$$Sc = \frac{\mu}{(\rho D_{AB})}$$

$$Js = Jv * C_p \quad (5)$$

$$\lambda = 6.23 K_\lambda Re^{-0.3}$$

$$D_{AB} = 6.76 \times 10^{-6} \exp\left(0.155 \times 10^{-3} C_b - \frac{2513}{273 + T}\right).$$

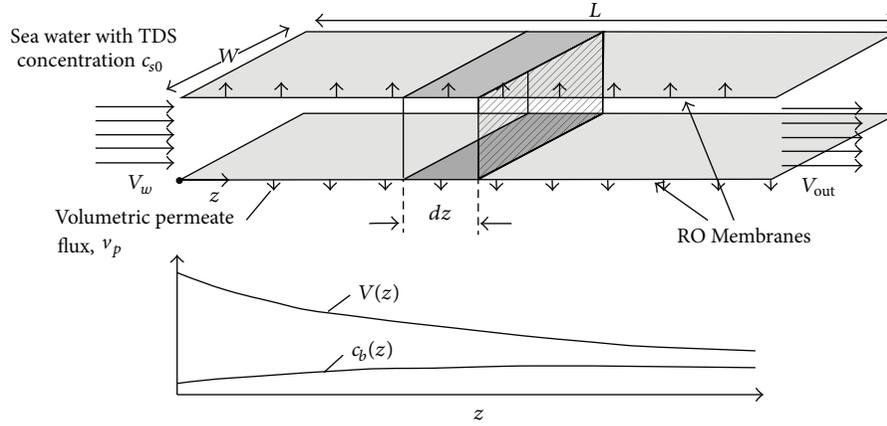


FIGURE 1: Scheme of the rectangular channel model of spiral wound module.

Consider (Masaaki Sekino, 1995)

$$\rho = 498.4M + \sqrt{248400M^2 + 752.4MC_b}$$

$$M = 1.0069 - 2.757 \times 10^{-4}T$$

$$\mu = 1.234 \times 10^{-6} \cdot \exp\left(0.00212C_b + \frac{1965}{273.15 + T}\right)$$

$$Q_r = Q_f \times \frac{V_r}{V_f}$$

$$R_{ec} = \frac{Q_p}{Q_f}$$

$$Sp = \frac{C_p}{C_f} \times 100\%$$

$$Ry = \left(1 - \frac{C_p}{C_f}\right) \times 100\%.$$

As to the large scale SWRO system with multiple RO units and multiple storage tanks, the storage tanks perform as the buffering units between water production of RO units and water supply. Since there are pipes to connect each RO unit with storage tanks, the feed water can come from each RO unit; the dynamic process can be formulated as

$$\frac{dH_{t,jj}(t)}{dt} = \frac{\left(\sum_{ii=1}^{NPT} Q_{pk}(ii, jj) - Q_{out,jj}\right)}{S_{t,jj}}$$

$$\frac{dC_{t,jj}(t)}{dt}$$

$$= \frac{1}{S_{t,jj}H_{t,jj}} \left(\sum_{ii=1}^{NPT} Q_{pk}(ii, jj) C_{p,ii} - \sum_{ii=1}^{NPT} Q_{pk}(ii, jj) C_{t,jj} \right)$$

$$Q_{sup} = \sum_{jj=1}^{NTK} Q_{out,jj}$$

$$C_{sup} = \sum_{jj=1}^{NTK} \frac{Q_{out,jj} C_{t,jj}}{Q_{sup}}$$

(6)

and equations

$$Q_{p,ii} = \sum_{jj=1}^{NTK} Q_{pk}(ii, jj)$$

(8)

$$\sum_{j=1}^{NTK} \sum_{i=1}^{NPT} Q_{pk}(ii, jj) = \sum_i^{NPT} Q_p(ii).$$

Bound constraint is as follows:

$$H_{t,jj,lo} < H_{t,jj} < H_{t,jj,up}. \quad (9)$$

Initial condition is as follows:

$$t = 0; \quad H_{t,jj} = H_{t,jj}(0), \quad (10)$$

$$C_{t,out,jj} = C_{t,out,jj}(0).$$

Here NPT denotes the number of RO units, NTK denotes the number of storage tanks, $Q_{pk}(ii, jj)$ denotes the flow rate from the ii th RO units to the jj th storage tanks, $Q_{out,jj}$ denotes the load requirement of freshwater of the jj th storage tank, Q_{sup} denotes the total flow rate of freshwater supply, and the C_{sup} denotes the concentration of the supplied freshwater.

3. Formulation of Dynamic Optimization Problem of SWRO System

Energy cost is the largest portion of the SWRO system's operation cost; generally it accounts for over 50% of the operation cost. So minimizing the energy cost can be set as the objective function. It is affected by the electricity price and the specific energy consumption (SEC) which is defined as [28]

$$\text{SEC} = \frac{P_f Q_f / \varepsilon_p - P_r Q_r \varepsilon_{ef}}{Q_p}. \quad (11)$$

With the consideration of daily fluctuation of feed temperature, electricity price, and the load requirement, the daily objective function can be denoted by

$$\min_{P_{f,ii}, Q_{f,ii}, H_{t,ii}} \int_0^{24} E_P \sum_{ii=1}^{NPT} (Q_{p,ii} \text{SEC}_{ii}) dt. \quad (12)$$

Since the control variables cannot be adjusted too frequently and the change of electricity price is based on an hour, it is reasonable to adjust the control variables one hour by one hour, and so (12) can be renewed in another form, and the optimization problem named Opt1 can be denoted by

Objective Function.

$$\min_{P_{f,ii}, Q_{f,ii}, H_{t,ii}} F_{\text{obj}} = \sum_{ii=1}^{24} \left[E_{P,ii} \sum_{ii=1}^{NTP} (Q_{p,ii} \text{SEC}_{ii}) \right]. \quad (13)$$

Equality Constraints.

- RO process model ((1)–(6)).
- Dynamic model of storage tanks ((7)–(8)).
- Specific energy cost (11).

Inequality Constraints.

- Water quality: $C_{\text{sup}} \leq C_{wq, \text{limit}}$.
- Scaling resistant: $C_{r,ii} \leq C_{r,ii, \text{limit}}$.
- Concentration polarization: $\phi \leq 1.2$.
- Pressure constraints: $P_{f,ii, \text{lo}} \leq P_{f,ii} \leq P_{f,ii, \text{up}}$.
- Flowrate of RO units: $V_{f,ii, \text{lo}} \leq V_{f,ii} \leq V_{f,ii, \text{up}}$.
- Water level $H_{t,jj, \text{lo}} < H_{t,jj} < H_{t,jj, \text{up}}$.
- Initial conditions of RO at each hour are as follows:

$$\begin{aligned} z = 0, \quad V_{ii} = V_{f,ii} &= \frac{Q_{f,ii}}{n_l W h_{\text{sp}}}; \\ z = L, \quad V_{ii} = V_{r,ii} &= \frac{Q_{r,ii}}{n_l W h_{\text{sp}}} \end{aligned} \quad (14)$$

$$\begin{aligned} z = 0, \quad P_b = P_f; \quad z = L, \quad P_b = P_r \\ z = 0, \quad C_b = C_f; \quad z = L, \quad C_b = C_r. \end{aligned}$$

Initial condition of storage tanks $t = 0$, (10).

4. The Strategy to Solve Opt2

4.1. Simultaneous Approach Based on Collocation on Finite Element. Since the Opt1 includes differential equations as well as strongly nonlinear algebraic equations, it belongs to a kind of differential-algebraic optimization problems (DAOPs). The Opt1 problem of SWRO system with NPT RO plants and NTK storage tanks will have $(3 \times NPT + 2 \times NTK)$ differential equations and more nonlinear algebraic equations, since there are also constraints to satisfy the water quality and the equipment's safety; it is really hard to get the solution efficiently through direct ways such as variational principle. Here we use the simultaneous approach to fully discretize all the control variables and status variables, which is quite suitable and has many advantages for this kind of problems. We use the following monomial basis representation for the differential profiles, which is popular for Runge-Kutta discretization [29]:

$$w(z) = w_{i-1} + h_i \sum_{q=1}^K \Omega_q \left(\frac{z - z_{i-1}}{h_i} \right) \frac{dw}{dz_{i,q}}. \quad (15)$$

Here w_{i-1} is the value of the differential variable at the beginning of element i , h_i is the length of element i , $dw/dt_{i,q}$ denotes the value of its first derivative in element i at the collocation point q , and Ω_q is a polynomial of order K , satisfying

$$\begin{aligned} \Omega_q(0) &= 0 \quad q = 1, \dots, K \\ \Omega'_q(\rho_r) &= \delta_{q,r} \quad q = 1, \dots, K; \end{aligned} \quad (16)$$

here ρ_r is the location of the r th collocation point within each element. Continuity of the differential profile is enforced by

$$w_i = w_{i-1} + h_i \sum_{q=1}^K \Omega_q(1) \frac{dw}{dz_{i,q}}. \quad (17)$$

In addition, the control and algebraic profile are approximated using a Lagrange basis representation which takes the form:

$$\begin{aligned} y(z) &= \sum_{q=1}^K \Psi_q \left(\frac{z - z_{i-1}}{h_i} \right) y_{i,q}, \\ u(z) &= \sum_{q=1}^K \Psi_q \left(\frac{z - z_{i-1}}{h_i} \right) u_{i,q}. \end{aligned} \quad (18)$$

Here $y_{i,q}$ and $u_{i,q}$ represent the values of the algebraic and control variables, respectively, in element i at collocation point q , z is the value satisfying $z_{i-1} \leq z \leq z_i$, and Ψ_q is the Lagrange polynomial of degree K .

Through the simultaneous method mentioned above, the Opt1 in the form of DAEs is discretized into the NLP problem denoted by (19a)–(19i); here the problem is named Opt2.

$$\min_{w(z_i^j), dw/dz_i^j, u(z_i^j), y(z_i^j), p_v} \varphi(w(z_i^j), u(z_i^j), y(z_i^j), p_v) \quad (19a)$$

$$\frac{dw}{dz_{i,j}} - F(w(z_i^j), u(z_i^j), y_i^j, z_i^j, p_v) = 0 \quad (19b)$$

$$H(w(z_i^j), u(z_i^j), y_i^j, z_i^j, p_v) = 0 \quad (19c)$$

$$w_{i,j} = w_{i-1} + h_i \sum_{j=1}^K \Omega_j \left(\frac{z - z_{i-1}}{h_i} \right) \frac{dw}{dz_{i,j}} \quad (19d)$$

$$w_i^0 - h_i \sum_{j=1}^k \Omega_j (1) \frac{dw}{dz_{i,j}} = 0 \quad (19e)$$

$$w_1^0 = w^0 \quad (19f)$$

$$w^L \leq w_i^j \leq w^U \quad (19g)$$

$$u^L \leq u_i^j \leq u^U \quad (19h)$$

$$y^L \leq y_i^j \leq y^U. \quad (19i)$$

Here $z_0 \leq z_i^j \leq z_f$, $i = 1, \dots, ne$ $j = 1, \dots, K$.

Since all the differential variables and continuous variables are fully discretized into algebraic variables along the spacious or time horizon, and large number of finite elements is generally selected to ensure the accuracy of the discretization, all the factors led to the great increase of the size of discretized problem of Opt2. As the result, though with the development of computing technique such as sparse matrix and automatic differentiation, large scale solvers such as IPOPT and CONOPT can solve problems with millions of variables, and its efficient and stable solution is a big challenge. Special work need to be paid attention to to reduce the computing time and to keep the stability of the problem.

There are still two key factors influencing the efficient solution of the discretized problem of Opt2: the first is to configure the number of finite element. Too large number of finite elements will lead to rapid increase in the size of variables and equations, which will greatly increase the computing time and memory occupation. The second problem is to obtain initial values for all the discretized variables, which is very critical to the stability of Opt2 formulated in (19a)–(19i). If the initial values are not carefully given, time-consuming and even failure of convergence will happen during the solution of Opt2. Based on the above reasons, techniques for assignment of initial values and suitable division of finite element are carefully designed as below.

4.2. Assignment of Initial Value through Simulation on Finite Element Individually. If all the control variables are fixed as constant values, all the variables of Opt2 can be obtained by simulation of the whole process model. According to

the flowsheet of SWRO system, the permeate flow rate and concentration of the permeate water are the input parameters for the dynamic process of storage tanks, so variables of storage tanks can only be obtained after the simulation of RO process.

The principle of collocation on finite element is shown in Figure 2. Though all the variables are discretized, to keep the continuity, the final values of the first finite element are equal to the initial value of the second finite element. So to get all the values of RO process, the control variables are fixed as constant values, and then the simulation is carried out from one finite element to the next along the space or time horizon. After the simulation of RO process was finished, the values of Q_p , C_p , and so on were used as input parameters and the simulations of storage tanks are then carried out in the same way of RO process.

After all the values of the discretized variables of Opt2 were obtained, we relax the control variables and enforce the constraints again. Then those obtained values were set as the initial value for Opt2, and nonlinear solver of IPOPT is used for the solution. The whole process of the technique to get good initial values is shown in Figure 3.

4.3. Method for Suitable Division of Finite Element. To balance the relationship of the discrete accuracy and computing efficiency, suitable ways to determine the number of finite element are necessary. Truong et al. improved an adaptive mesh refinement method [30] to improve the computing efficiency; Lan and Taylor designed a moving grid computing strategy [31], though the computational efficiency is not high, the method is more stable. Binder proposed a mesh refinement strategy, making it superior to the average grid distribution strategy [32]. Biegler et al. proposed a dynamic optimization method [33–35] based on moving finite element technique; the finite element is automatically adjusted through error analysis. In this work, with the consideration of the special structure of SWRO system, and the relationship between the change of objective function value and the number of finite element which affect discrete state and control variables, a simple but effective method to get suitable division of finite element is designed as follows.

- (1) Set the accuracy of the objective function which is defined as $\sigma = [1 - \text{abs}(F_{\text{obj}0}/F_{\text{obj}1})] \leq \gamma$; γ is a small positive value such as $1.0E - 7$.
- (2) Set the number of finite element as $N_{\text{ef}1}$ and the corresponding collocation point by conventional way.
- (3) Solve the discretized problem of Opt2, and get the value of objective function $F_{\text{obj}0}$.
- (4) Add the number of finite element to $N_{\text{ef}1}'$.
- (5) Resolve the discretized problem of Opt2, and then obtain another value of objective function $F_{\text{obj}1}$.
- (6) Compute the value of σ .
- (7) If the required accuracy is satisfied, then stop; otherwise, let $F_{\text{obj}0} = F_{\text{obj}1}$, and go to step (4).

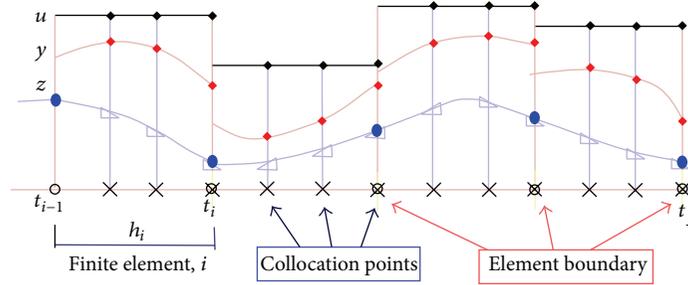


FIGURE 2: Schematic of collocation on finite elements.

Since IPOPT is one of the best solvers to solve large scale dynamical problems, here it was used for the solution of Opt2 quickly.

5. Case Study

The large scale SWRO system has 8 RO plants and 4 storage tanks, and the RO membranes are the products of Dow Chemical Company named SW30HR. Each RO plant has decades of pressure vessels in parallel, and several RO membranes in the pressure vessel connected in series are used to separate salt and pure water. The storage tanks are used as the buffering units between water supply and water production. The electricity price and other information for the system are listed in Tables 1-2 and shown in Figure 4. Here NPT rays and NMOD represent the number of pressure vessel and the number of membranes in pressure vessels. FF denotes the fouling factor, and ε_p and ε_{pf} denote the efficiency of high-pressure pumps and energy recovery unit.

The optimization problem of this system was discretized by simultaneous approach based on collocation on finite element. Since the control variables are not allowed to adjust frequently and the electricity price changes at different hour, 24 finite elements and 3 collocation points are selected to discretize the dynamic variables of storage tanks. For the RO process, 3 collocation points are used and the number of finite element is determined by the proposed technique. To get good initial values, the developed method in chapter 3.2 was used for the solution of Opt2. To comprehensively understand the effect of finite element on the solution, information of problem Opt2 with different finite element of RO process is listed in Table 3. Table 3 shows that when the number of finite element of RO increased to 40, the discretized variables and the equations of Opt2 reached 495371 and 494411, respectively; the size of the problem is quite large.

Nonlinear solver IPOPT [36, 37] which is based on interior point method was used for the solution of Opt2, and computing results with different finite element of RO process were listed in Table 4. The results show that the problem of Opt2 can be successfully solved when the finite element increases from 2 to 40, and the solution converged to the best point within decades of iterations. The computing time of simulation and optimization increased much more greatly when the finite element increases. In particular, it can be seen

TABLE 1: Designed operating parameters of RO unit.

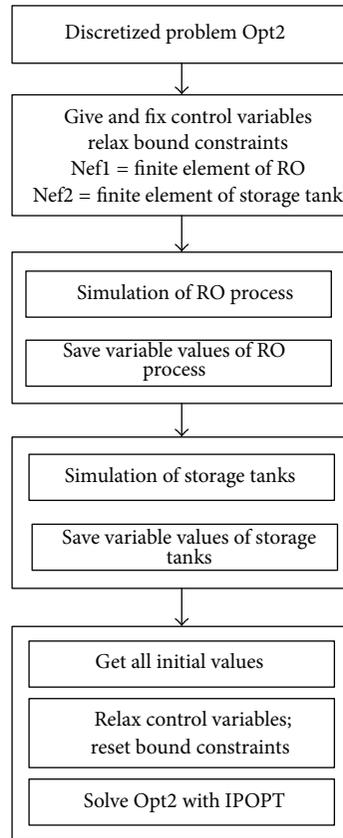
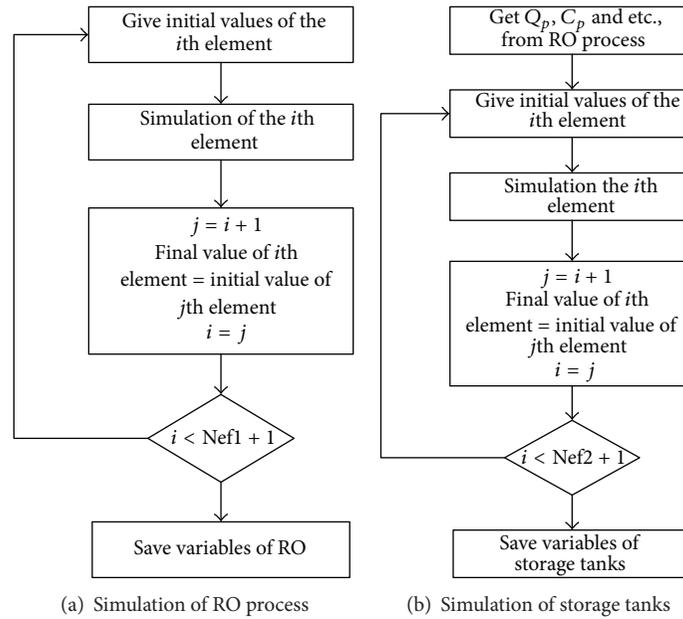
Feed concentration (kg/m^3)	30
Feed temperature ($^{\circ}\text{C}$)	25
Feed pressure (Bar)	59
Feed PH	5-8
Kinematic viscosity (kg/m^3)	$1.02e-6$ (at 0°C)
Mechanical efficiency of HP	0.85
Energy efficiency of PX	0.9
Mechanical efficiency of η_{vd}	0.95
Single water quality limit (kg/m^3)	0.9
Water quality limit (kg/m^3)	0.9
Spiral wound modules	SW30HR series

TABLE 2: Structure information about RO units.

Number of plant	NMOD	NPTrays	FF	ε_p	ε_{pf}
1	8	25	1.0	0.85	0.9
2	8	30	1.0	0.85	0.9
3	7	35	1.0	0.85	0.9
4	7	55	1.0	0.85	0.9
5	7	40	1.0	0.85	0.9
6	7	50	1.0	0.85	0.9
7	6	30	1.0	0.85	0.9
8	6	32	1.0	0.85	0.9

that the optimization time increases almost in the form of exponential. When 40 finite elements and 3 collocation points were selected, the optimization time will reach 3519.953 CPU seconds.

It can also be seen that the change of objective function became much smaller with the increase of finite elements. Compared with the objective function of 40 finite elements, the relative errors were obtained and showed in Figure 5. As the finite elements increased to 10, the relative errors become fairly small, and the required accuracy can be reached at this time. Compared with the computing time with 40 finite elements, more than 94% computing time was saved. To show the effectiveness of the above method, initial values for Opt2 were given by reasonable assumption near the set point, the optimal solution failed to converge when the finite element is 40; and even when the finite element is 10, more than 1187



(c) The whole process of the strategy

FIGURE 3: Initial value strategy based on step by step simulation.

TABLE 3: Information of the Opt2 through discretization.

Item number	Nefl	Ncp	Total variables	Equalities	Nonzero equality Jacobian	Nonzero Lagrangian Hessian
1	2	3	100378	40260	28427	27467
2	5	3	65291	64331	227674	86916
3	10	3	126731	125771	439834	164676
4	20	3	249611	248651	864154	320196
5	40	3	495371	494411	1712794	631236

TABLE 4: Computing results with different finite elements.

Number	Nefl	Sim. time (s)	Opt. time (s)	Total time (s)	Iteration	F_{obj} (10^4 CNY)	Relative errors ($\times 10^5$)
1	2	29.374	13.907	43.281	32	3.5913297	0.66349
2	5	60.407	46.64	107.047	33	3.5913061	0.00635
3	10	103.328	162.453	265.781	56	3.59130588	0.00022
4	20	209.14	868.328	1077.468	66	3.591305841	-0.00086
5	40	609.172	3519.953	4129.125	121	3.591305872	0

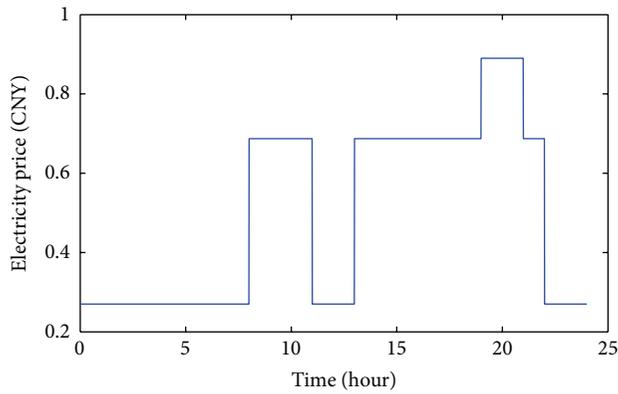


FIGURE 4: Daily profile of electricity price.

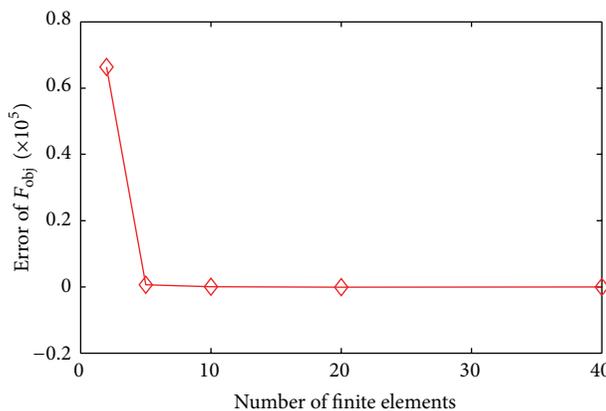


FIGURE 5: Relationship between errors of objective function values and number of finite element.

iterations and more than 4680 seconds were spent to get the optimal solution.

As 10 finite elements of RO unit were selected, the solution of problem of Opt2 was resolved under default

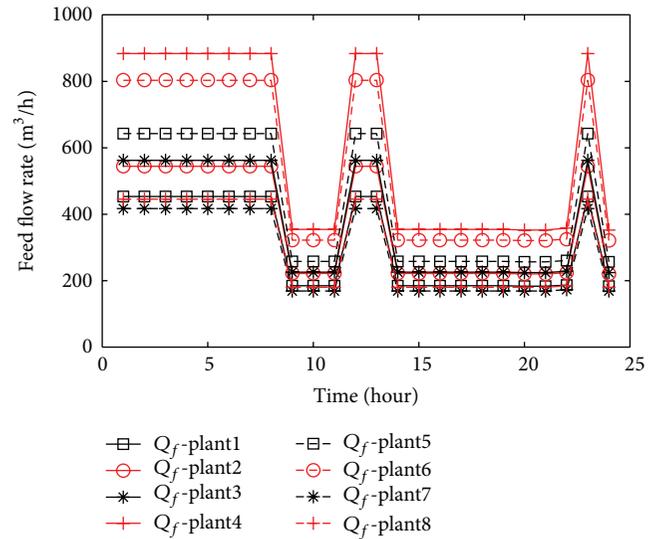


FIGURE 6: Profiles of feed flow rate of each RO plant.

operating conditions. The solution led to the profiles of control variables-feed flow rates and feed pressures along time horizon, from which the best control trajectory as well as the various status variables are achieved. Some of these results can be seen from Figures 6, 7, 8, and 9.

Since the feed temperature changes frequently and has significant effect on performance of RO membranes, it is necessary to optimize the process quickly through adjusting the control variables to the aimed value. To validate the performance of the proposed strategy with large fluctuation of feed temperature, increase the feed temperature from 21°C to 29°C and keep other parameters as default, and then solve the Opt2 individually. Summary of the computing results can be seen from Table 5. From which it can be seen that temperature has significant effect on the energy cost. The optimal energy cost reduces quickly as the temperature increases, when feed temperature increases from 21°C to

TABLE 5: Comparison of computing results with different feed temperatures.

T ($^{\circ}\text{C}$)	Simulation time (s)	Optimization time (s)	Total time (s)	Iter.	F_{obj} (10^4 CNY)	Bias (%)
21	79.608	129.985	209.593	43	3.8654	7.63
23	86.235	132.297	218.532	43	3.7166	3.49
25	103.328	162.453	265.781	56	3.5913	0.00
27	97.141	120.875	218.016	39	3.4915	-2.78
29	109.032	123.031	232.063	42	3.4202	-4.76

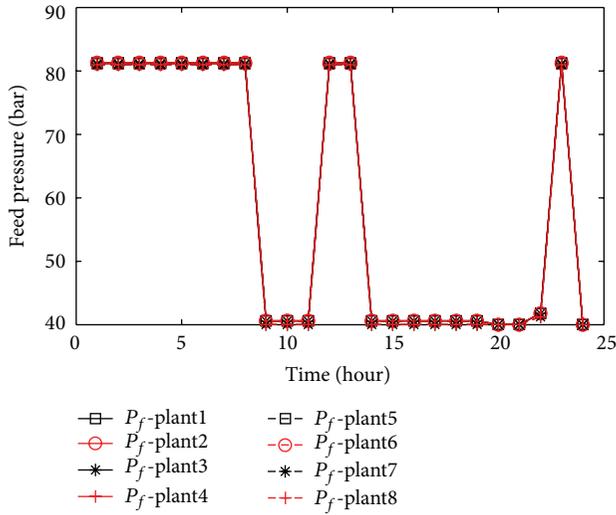


FIGURE 7: Profiles of feed pressure of each RO plant.

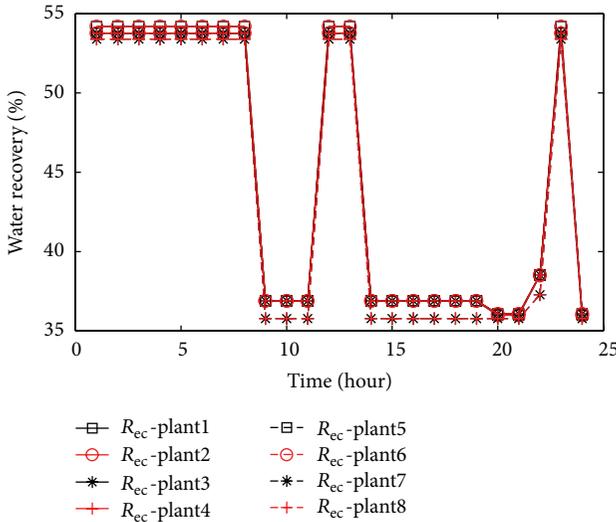


FIGURE 8: Profiles of water recovery of each RO plant.

29 $^{\circ}\text{C}$, the optimal energy cost reduces from 3.8654 Wan CNY to 3.4202 Wan CNY. It can also be seen that the change of feed temperature has little effect on the computing performance and convergence of the proposed method. All the solution can be successfully achieved within decades of iterations and within several minutes. The results show that the proposed strategy for the SWRO system not only is quite effective

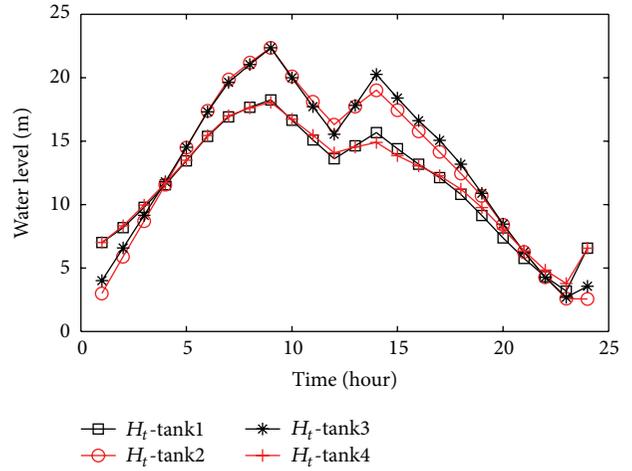


FIGURE 9: Profiles of water level of each storage tank.

but also has wide temperature adaptability. All the results show the good potential to the on-line optimization of actual SWRO system.

6. Conclusions

It is of significance to reduce the energy cost of seawater reverse osmosis (SWRO) system through the system engineering approach, especially when the system is operated in dynamical environments due to the changes of load and other operating parameters.

In this work, a dynamic optimization strategy for optimizing operation of large scale SWRO system was proposed. Based on first principle, the dynamic models of RO units and storage tanks which are developed by were formulated by differential-algebraic equations, and the objective function is formulated to minimize the energy cost. The optimization problem in the form of DAOPs was fully discretized into large scale NLP through simultaneous method, with which all the dynamic variables were approximated by polynomial equations with collocation point on each finite element. To solve the large scale NLP problem named Opt2 efficiently, a method was developed to get the initial value of all variables through simulation element by element and step by step. Then a simple but effective rule to determine the number of finite element was used to reduce the scale and at the same time to keep enough accuracy of Opt2.

Case study of a large scale SWRO system which includes 8 RO plants and 4 storage tanks was studied to validate

the proposed strategy. Computing results demonstrate that the method is quite efficient for the solution of dynamic optimization of SWRO system. The problem can be solved in decade's iterations within several minutes. The minimized objective function and accordingly the best daily profiles of control variables can be achieved. The proposed strategy also performs well with large fluctuation of feed temperature, which can fully satisfy the requirement of on-line real time optimization.

Appendix

Nomenclature

A_w :	Membrane water permeability ($\text{m}\cdot\text{s}^{-1}\cdot\text{Pa}^{-1}$)
A_{w0} :	Intrinsic membrane water permeability ($\text{m}\cdot\text{s}^{-1}\cdot\text{Pa}^{-1}$)
B_s :	Membrane TDS permeability (m/s)
B_{s0} :	Intrinsic membrane TDS permeability (m/s)
C_b :	Bulk concentration along feed channel (kg/m^3)
C_f :	Feed salt concentration (kg/m^3)
C_m :	Salt concentration of membrane surface (kg/m^3)
$C_{p,ii}$:	Permeate concentration of i th RO unit (kg/m^3)
C_r :	Brine concentration (kg/m^3)
$C_{t,jj}$:	Salt concentration of output freshwater from j th tanks (kg/m^3)
D_{AB} :	Dynamic viscosity (m^2/s)
d_e :	Hydraulic diameter of the feed spacer channel (m)
E_p :	Electricity price at different hours (CNY/kw·h)
$H_{t,jj}$:	Water level of j th storage tank (m)
h_i :	The length of the i th finite element
h_{sp} :	Height of the feed spacer channel
J_v :	Solvent flux
J_s :	Solute flux
K_λ :	Empirical parameter
k_c :	Mass transfer coefficient (m/s)
L :	Length of the RO module (m)
NPT :	The number of RO units
NTK :	The number of storage tanks
n_i :	Number of leaf in RO module
P_b :	Pressure along feed channel (bar)
P_d :	Pressure drop along RO spiral wound module (bar)
P_f :	Feed pressure (bar)
P_p :	Pressure in permeate side (bar)
P_r :	Brine pressure (bar)
Q_f :	Feed flow rate (m^3/h)
$Q_{\text{out},jj}$:	Output flow rate of the j th tank (m^3/h)
$Q_{p,ii}$:	Permeate flow rate from i th RO unit (m^3/h)

$Q_{pk(ii,jj)}$:	Flow rate from i th RO unit to the j th tank (m^3/h)
Q_r :	Brine flow rate (m^3/h)
Q_{sup} :	Total water supply flowrate (m^3/h)
R :	Gas law constant
Re :	Reynolds number (dimensionless)
R_{ec} :	Water recovery ratio (%)
Sc :	Schmidt number (dimensionless)
SEC :	Specific energy consumption ($\text{kw}\cdot\text{h}/\text{m}^3$)
Sh :	Sherwood number (dimensionless)
Sp :	Salt passage coefficient (%)
S_r :	Area of storage tank (m^2)
T :	Operational temperature (K)
V :	Axial velocity in feed channel (m/s)
V_f :	Input axial velocity in membrane (m/s)
V_r :	Output axial velocity in membrane (m/s)
W :	Width of the RO module (m)
Ry :	Salt rejection coefficient (%)
$y_{i,q}$:	Algebraic variables in element i at collo- cation point q
$u_{i,q}$:	Control variables in element i at colloca- tion point q
$w(t)$:	Differential variable
w_{i-1} :	Differential variable at the beginning of element i .

Greek Symbols

$\alpha_1, \alpha_2, \beta_1$:	Constant parameters
$\Delta\pi$:	Pressure loss of osmosis pressure (bar)
ρ :	Density of permeate water (kg/m^3)
μ :	Kinematic viscosity ($\text{kg}\cdot\text{m}^{-1}\cdot\text{s}^{-1}$)
λ :	Friction factor
ε_p :	Mechanical efficiency of high pressure pump
ε_{pf} :	Energy recovery efficiency
$\Omega_q(t)$:	Polynomial of order K
$\psi_q(t)$:	Lagrange polynomial of degree K .

Subscripts

b :	Bulk
f :	Module feed channel
m :	Membrane surface
p :	Permeate side
s :	Salt.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by National Natural Science foundation of China (nos. 60904058 and 61374142), National Key Technology Research and Development Program

(2009BAB47B06), and Natural Science Foundation of Zhejiang (LY12F03001). The authors would like to extend their sincere thanks to Mr. Yang Chen from Carnegie Mellon University for many helpful discussions.

References

- [1] M. N. A. Hawlader, J. C. Ho, and C. K. Teng, "Desalination of seawater: an experiment with RO membranes," *Desalination*, vol. 132, no. 1–3, pp. 275–280, 2000.
- [2] S. Gray, R. Semiat, M. Duke, A. Rahardianto, and Y. Cohen, "Seawater use and desalination technology," in *Treatise on Water Science*, W. Peter, Ed., pp. 73–109, Elsevier, Oxford, UK, 2011.
- [3] J. Wang, D. Jiang, C. Chen, and M. Ma, "21st century China water problems and solutions," *Territory & Natural Resources Study (China)*, no. 2, pp. 8–10, 1999.
- [4] S. Wang, *Desalination Project*, Chemical Industry Press, Beijing, China, 2003.
- [5] R. Semiat, "Desalination: present and future," *Water International*, vol. 25, no. 1, pp. 54–65, 2000.
- [6] L. F. Greenlee, D. F. Lawler, B. D. Freeman, B. Marrot, and P. Moulin, "Reverse osmosis desalination: water sources, technology, and today's challenges," *Water Research*, vol. 43, no. 9, pp. 2317–2348, 2009.
- [7] T. I. Yun, C. J. Gabelich, M. R. Cox, A. A. Mofidi, and R. Lesan, "Reducing costs for large-scale desalting plants using large-diameter, reverse osmosis membranes," *Desalination*, vol. 189, no. 1–3, pp. 141–154, 2006.
- [8] A. Scrivani, "Energy management and DSM techniques for a PV-diesel powered sea water reverse osmosis desalination plant in Ginostra, Sicily," *Desalination*, vol. 183, no. 1–3, pp. 63–72, 2005.
- [9] S. Yin, S. X. Ding, A. Haghani, H. Hao, and P. Zhang, "A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process," *Journal of Process Control*, vol. 22, no. 9, pp. 1567–1581, 2012.
- [10] H. Zhang and J. M. Wang, "Combined feedback-feedforward tracking control for networked control systems with probabilistic delays," *Journal of the Franklin Institute*, vol. 351, no. 6, pp. 3477–3489, 2014.
- [11] H. Zhang, H. R. Karimi, X. J. Zhang, and J. M. Wang, "Advanced control and optimization with applications to complex automotive systems," *Mathematical Problems in Engineering*, vol. 2014, Article ID 183580, 3 pages, 2014.
- [12] H. Y. Li, X. J. Jing, H. K. Lam, and P. Shi, "Fuzzy sampled-data control for uncertain vehicle suspension systems," *IEEE Transactions on Cybernetics*, 2013.
- [13] H. Y. Li, J. Y. Yu, C. Hilton, and H. H. Liu, "Adaptive sliding-mode control for nonlinear active suspension vehicle systems using T-S fuzzy approach," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 8, pp. 3328–3338, 2013.
- [14] Y. M. Kim, S. J. Kim, Y. S. Kim, S. Lee, I. S. Kim, and J. H. Kim, "Overview of systems engineering approaches for a large-scale seawater desalination plant with a reverse osmosis network," *Desalination*, vol. 238, no. 1–3, pp. 312–332, 2009.
- [15] K. M. Sassi and I. M. Mujtaba, "Simulation and optimization of full scale reverse osmosis desalination plant," *Computer Aided Chemical Engineering C*, vol. 28, pp. 895–900, 2010.
- [16] L. Palacin, C. de Prada, F. Tadeo, and J. Salazar, "Operation of medium-size reverse osmosis plants with optimal energy consumption," in *Proceedings of the 9th International Symposium on Dynamics and Control of Process Systems (DYCOPS '10)*, pp. 841–846, Leuven: Belgium, July 2010.
- [17] A. Zhu, P. D. Christofides, and Y. Cohen, "Effect of thermodynamic restriction on energy cost Optimization of RO membrane water desalination," *Industrial and Engineering Chemistry Research*, vol. 48, no. 13, pp. 6010–6021, 2009.
- [18] M. H. Li, "Optimal plant operation of brackish water reverse osmosis (BWRO) desalination," *Desalination*, vol. 293, pp. 61–68, 2012.
- [19] V. Geraldes, N. E. Pereira, and M. N. De Pinho, "Simulation and optimization of medium-sized seawater reverse osmosis processes with spiral-wound modules," *Industrial and Engineering Chemistry Research*, vol. 44, no. 6, pp. 1897–1905, 2005.
- [20] K. M. Sassi and I. M. Mujtaba, "Effective design of reverse osmosis based desalination process considering wide range of salinity and seawater temperature," *Desalination*, vol. 306, pp. 8–16, 2012.
- [21] K. M. Sassi and I. M. Mujtaba, "Optimal operation of RO system with daily variation of freshwater demand and seawater temperature," *Computers and Chemical Engineering*, 2013.
- [22] L. T. Biegler, A. M. Cervantes, and A. Wächter, "Advances in simultaneous strategies for dynamic process optimization," *Chemical Engineering Science*, vol. 57, no. 4, pp. 575–593, 2002.
- [23] L. T. Biegler, "An overview of simultaneous strategies for dynamic optimization," *Chemical Engineering and Processing: Process Intensification*, vol. 46, no. 11, pp. 1043–1053, 2007.
- [24] S. Kimura and S. Sourirajan, "Analysis of data in reverse osmosis with porous cellulose acetate membranes used," *AIChE Journal*, vol. 13, no. 3, pp. 497–503, 1967.
- [25] J. Marriott and E. Sørensen, "A general approach to modelling membrane modules," *Chemical Engineering Science*, vol. 58, no. 22, pp. 4975–4990, 2003.
- [26] T. Kaghazchi, M. Mehri, M. T. Ravanchi, and A. Kargari, "A mathematical modeling of two industrial seawater desalination plants in the Persian Gulf region," *Desalination*, vol. 252, no. 1–3, pp. 135–142, 2010.
- [27] A. Abbas, "Simulation and analysis of an industrial water desalination plant," *Chemical Engineering and Processing: Process Intensification*, vol. 44, no. 9, pp. 999–1004, 2005.
- [28] H. Oh, T. Hwang, and S. Lee, "A simplified simulation model of RO systems for seawater desalination," *Desalination*, vol. 238, no. 1–3, pp. 128–139, 2009.
- [29] L. T. Biegler, *Nonlinear Programming: Concepts, Algorithms and Applications to Chemical Processes*, Society for Industrial and Applied Mathematics, Cambridge University Press, Pittsburgh, Pa, USA, 2010.
- [30] A. H. Truong, C. A. Oldfield, and D. W. Zingg, "Mesh movement for a discrete-adjoint Newton-Krylov algorithm for aerodynamic optimization," *AIAA Journal*, vol. 46, no. 7, pp. 1695–1704, 2008.
- [31] Z. Lan and V. E. Taylor, "Dynamic load balancing for adaptive mesh refinement application," in *Proceedings of the International Conference on Parallel and Distributed Computing and Systems (IASTED-PDCS '01)*, Anaheim, Calif, USA, 2001.
- [32] T. Binder, L. Blank, W. Dahmen, and W. Marquardt, "Adaptive multiscale method for real-time moving horizon optimization," in *Proceedings of the Americal Control Conference*, pp. 4234–4238, Chicago, Ill, USA, June 2000.
- [33] J. E. Cuthrell and L. T. Biegler, "Simultaneous optimization and solution methods for batch reactor control profiles," *Computers and Chemical Engineering*, vol. 13, no. 1–2, pp. 49–62, 1989.

- [34] W. F. Chen, K. X. Wang, Z. J. Shao, and L. T. Biegler, "Moving finite elements for dynamic optimization with direct transcription formulations," in *Control and Optimization with Differential-Algebraic Constraints*, pp. 241–259, SIAM, Philadelphia, Pa, USA, 2012.
- [35] V. M. Zavala, C. D. Laird, and L. T. Biegler, "A fast moving horizon estimation algorithm based on nonlinear programming sensitivity," *Journal of Process Control*, vol. 18, no. 9, pp. 876–884, 2008.
- [36] L. T. Biegler and V. M. Zavala, "Large-scale nonlinear programming using IPOPT: an integrating framework for enterprise-wide dynamic optimization," *Computers and Chemical Engineering*, vol. 33, no. 3, pp. 575–582, 2009.
- [37] A. Waechter, *An interior point algorithm for large-scale nonlinear optimization with applications in process engineering [Ph.D. thesis]*, Carnegie Mellon University, Pittsburgh, Pa, USA, 2002.

Research Article

A Study of Network Violator Interception Based on a Reliable Game Model

Shi An, Jianxun Cui, Meng Zhao, and Jian Wang

School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin 150090, China

Correspondence should be addressed to Jianxun Cui; cuijianxun@hit.edu.cn

Received 29 March 2014; Revised 28 May 2014; Accepted 28 May 2014; Published 15 June 2014

Academic Editor: Hamid Reza Karimi

Copyright © 2014 Shi An et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study focuses on planning interceptor locations in a general transportation network to maximize the expected benefits from catching violators mixing in public traveler flow. Two reliability-related characteristics are also integrated into the planning model to make it more practical. One is that each interceptor (e.g., a sensor or a checkpoint) has a failure probability. The second is the existence of a “game” between the interceptor planner and violators. A nonlinear nonconvex binary integer programming model is presented. We develop a simulated annealing (SA) algorithm to solve this model, and numerical experiments are conducted to illustrate the computational efficiency of the proposed algorithm. We also analyze the sensitivity of the disruption probability of interceptors to optimal objective function values and discuss how to determine the values of these parameters in a violator route choice model.

1. Introduction

There currently exist many types of violators in daily life networks (such as ground transportation networks, commercial airline networks, and communication networks), any of whom may pose a serious threat to public safety. A typical example is that of the drunk drivers who frequently emerge in an urban road network [1–4]. This is a difficult, real, and common problem troubling many local governments in many countries. Car accidents caused by drunk driving claim the lives of more than 15,000 people in the US every year and injure many more. More intuitively, as concluded by authorities [5], (1) one-third of all Americans will be involved in drunk driving accidents at some point in their lifetimes; (2) someone is injured in an alcohol-related car accident approximately every 60 seconds; (3) someone is killed in an alcohol-related car accident approximately every 40 minutes; and (4) drunk driving car accidents cost the US an average of \$114 billion annually. This phenomenon also exists in China, where drunk driving and consequent traffic accidents are even more prevalent and serious. Eliminating drunk driving has long been a major focus of traffic safety professionals at the federal, state, and local levels. Terrorists, another type of violators, aim to attack airline networks

[6–9]. They mix in with the public flow of passengers and watch for an opportunity to launch an attack. As further examples, viruses and illegal data may exist within the public information flow in a local intranet or a world-wide internet. Viruses, in particular, may steal or violate a public user’s private information or cause disruption to public internet users. Means of intercepting and eliminating illegal viruses is an important subject in the field of network security.

A common problem faced by decision makers who want to protect networks is how to catch violators mixing in the public flow as soon as possible by setting up interceptors (which could be DUI checkpoints for catching drunk drivers, or a virtual network police force that screens for viruses and illegal information). A network cannot have unlimited numbers of interceptors due to budgetary or logistical limits (e.g., a DUI checkpoint for drunk driving at one location will cause delays because vehicles need to stop and be checked, limiting the number of checkpoints that can be established). Furthermore, an unlimited number of interceptors is not necessary. One can always attempt to find an optimal interceptor layout with less than a given number of interceptors to obtain the largest expected benefit. This interceptor location problem (ILP) is very similar to “network interdiction” problems. The deterministic network interdiction problem has been

studied by Zenklusen [10], Granata et al. [11], Rad and Kakhki [12], and Yates and Sanjeevi [13] with military applications and with applications to the interdiction of illegal drugs and precursor chemicals. In this problem, given a capacitated network, a source \mathbf{s} and a sink \mathbf{t} , the aim is to interdict arcs in the network to minimize the maximum flow from \mathbf{s} to \mathbf{t} , subject to constraints on the number of prohibited arcs. The stochastic version of this problem studied by Cormican et al. [14] and Ramirez-Marquez and Rocco S. [15] assumes that interdictions on arcs do not always succeed but may be either completely unsuccessful, partially successful, or completely successful.

All network interdiction problems are rooted in the well-known max-flow-min cut theorem. These problems consider violators a type of commodity and attempt to break arcs to leave as few as possible arc-disjoint paths remaining for violators to use. However, for the ILP, we treat violators as a minority mixing in with public travelers, who are the majority. We attempt to discriminate violators from innocent public travelers as soon as possible, while they share the same network at the same time. ILP is not a max-flow-min problem in nature, as the purpose of network interdiction problems is to intercept as many violators as possible regardless of how long the violator will travel in the network or how many public travelers will be threatened by the violator. Additionally, in practice, we implement a “game” between interceptor planners and violators. After planners instantiate a specific interceptor layout, violators will adjust their route choices as much as possible to avoid being caught by interceptors. Thus, we expect that a robust solution to the interceptor layout will achieve equilibrium between planners and violators.

For the ILP, which is different from network interdiction research, we propose a reliable interceptor location problem (RILP). We assume that, in a general network, two types of flow exist: one is the public traveler flow, composed of innocent travelers, and the other is the flow of violators, who are mixed in with the public traveler flow and pose a threat to public travelers. We aim to effectively identify violators from the public flow by setting up interceptors along routes of flow. The installed interceptors are assumed to have a given probability of failure at capturing violators as they pass by. Additionally, “game” behaviors between planners and violators are integrated into the proposed model. To the best of our knowledge, our attempt is the first attempt to study ILP in a reliable and gamified manner.

This paper is organized as follows. Section 2 presents the model formulation. Section 3 discusses the algorithms for this proposed model. Section 4 presents some numerical tests to investigate the computational performance of the proposed model and algorithms. Section 5 concludes the paper and discusses future research directions.

2. Model Formulation

To formulate RILP, a network should first be classified into a series of origin-destination (OD) pairs between which daily traffic demand (including violators) is generated. We use w to represent the set of all OD pairs, φ_r to represent the set

of available routes between a specified OD pair $r \in w$ and φ to represent the set of all possible routes between OD pairs. Let ϕ_i be the set of candidate locations along route $i \in \varphi_r$ for interceptors and let ϕ be the set of all candidate interceptor locations. D_r represents the expected number of violators generated in OD pair $r \in w$ during any specified time interval.

For any OD flow, an interceptor will check passengers if and only if the flow passes the interceptor. Any violators in the flow will be caught. In this case, we say that the flow is covered by the interceptor or that we have *flow coverage*. The benefit obtained from setting up an interceptor along a specific route, which we denote by *path coverage*, depends on not only the inspected passenger volume but also the lengths of the covered OD paths. For simplicity, this study adopts a vehicle-mile coverage measure [16] such that the path coverage benefit for an OD path is proportional to both its traffic volume and the covered length.

The existence of violators will pose a safety threat to the traveling public. The magnitude of the threat is dependent on the public traveler flow encountered and the length of the path traveled by the violators before their interception. For a specific route i , setting up an interceptor at candidate location $j \in \phi_i$ results in a corresponding potential benefit b_{ij} stemming from the public’s protection from violators on that route. Therefore, in a network with link set A , the *path coverage* benefit b_{ij} can be defined as follows:

$$b_{ij} = \sum_{a \in A} \delta_{ij}^a f_a L_a, \quad \forall i \in \varphi, \forall j \in \phi_i, \quad (1)$$

where L_a and f_a are, respectively, the length and the public traveler flow on link $a \in A$ and parameter δ_{ij}^a is defined as follows:

$$\delta_{ij}^a = \begin{cases} 1, & \text{link } a \text{ is on path } i \text{ and downstream of location } j \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

In practice, due to the performance limit of interceptors, violators may not always be caught when they pass by interceptors. Even when interceptors are fully competent to discriminate violators from the public flow, we should not assume that all passing passenger flows are checked by interceptors because it is not practical to stop the flow to check every passenger (i.e., checking for drunk drivers at a DUI checkpoint will cause traffic delays if the sampling rate is high). This is similar to the concept of “sensor failure” that is widely studied in the literature [17–21]. In this paper, we define this failure as “interceptor failure.” Complementary to the introduction of interceptor failure, the expected intercepting benefit from setting up an interceptor at candidate location j along route i is related to the “head level” of that interceptor. Supposing that there are n_i interceptors installed on route i . We see that once the locations with installations on i are given (i.e., $\{j_1^i, j_2^i, \dots, j_{n_i}^i\}$ ordered from upstream to downstream), their head levels are defined as follows.

Definition 1 (head level). An interceptor at location j_s^i is the level- s head interceptor along route i .

The primal decision variables $x = \{x_j\}$ determine where to install DUI sites, where

$$x_j = \begin{cases} 1, & \text{a interceptor is installed at candidate location } j \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

Given x , the auxiliary variables $h = \{h_{ij}^m\}$ decide how interceptors are assigned to paths, where

$$h_{ij}^m = \begin{cases} 1, & \text{a interceptor is installed at candidate location } j \\ & \text{along route } i \text{ and assigned to head level } m \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

Assume that each interceptor fails independently with an identical probability $0 \leq p < 1$. The *path coverage* benefit from setting up interceptors at location j along route i is b_{ij} . As interceptor j functions to catch violators with probability $(1-p)p^{m-1}$, its expected contribution to catching violators is $(1-p)p^{m-1}h_{ij}^m b_{ij}$.

Additionally, each route $i \in \varphi_r$ has a probability α_i of being selected by violators, such that $\sum_{i \in \varphi_r} \alpha_i = 1, \forall r \in w$. This route selecting probability should not be defined statically but should instead be related to the layout of interceptors. For example, locating a DUI checkpoint at a certain location one night may induce drunk drivers to avoid this location later. To account for this “game” between interceptor planners and violators, we assume that violators will select a route between a specific OD pair according to a Logit model, in which the utility of violators to select a specific route is determined by two aspects: the numbers of installed interceptors along all possible OD routes and the travel lengths of these routes. We assume that violators always try to find a route that possesses as few interceptors as possible and a short travel distance. We define α_i as follows:

$$\alpha_i = \frac{\exp(\beta_1 p^{n_i} - \beta_2 L_i)}{\sum_{k \in \varphi_r} \exp(\beta_1 p^{n_k} - \beta_2 L_k)}, \quad \forall r \in w, \forall i \in \varphi_r, \quad (5)$$

where $n_i = \sum_{j \in \phi_i} \sum_{m=1}^{R_i} h_{ij}^m$ is the number of installed interceptors along route i , L_i is the travel length of route i , and β_1, β_2 are two coefficients corresponding to interceptor number and travel length, respectively, in the utility function.

Taking into account the discussion above, the expected intercepting benefits across an exponential number of failure scenarios and the game behavior between planners and

violators can be consolidated into a compact expression, and RILP can be written as follows:

$$(RILP) \quad \max_{x,h} \sum_{r \in w} \sum_{i \in \varphi_r} \sum_{j \in \phi_i} \sum_{m=1}^{R_i} D_r \alpha_i b_{ij} h_{ij}^m p^{m-1} (1-p) \quad (6)$$

$$\text{s.t.} \quad \sum_{j \in \phi} x_j \leq N \quad (7)$$

$$\sum_{m=1}^{R_i} h_{ij}^m = x_j; \quad \forall i \in \varphi, \forall j \in \phi_i \quad (8)$$

$$\sum_{j \in \phi_i} h_{ij}^1 \leq 1; \quad \forall i \in \varphi \quad (9a)$$

$$\sum_{j \in \phi_i} h_{ij}^m \leq \sum_{j \in \phi_i} h_{ij}^{m-1}; \quad \forall i \in \varphi, \forall m \in [2, R_i] \quad (9b)$$

$$R_i = \min(|\phi_i|, N); \quad \forall i \in \varphi \quad (10)$$

$$x_j, h_{ij}^m \in \{0, 1\}; \quad \forall i \in \varphi, \forall j \in \phi_i, \quad (11)$$

$$\forall m \in [1, R_i].$$

Objective function (6) aims to maximize the total expected benefit from all interceptors. Constraint (7) enforces the budget limit, which is that no more than N interceptors are allowed to be installed. Constraint (8) ensures that each installed interceptor is assigned to each of its corresponding paths at one and only one head level. Constraint (9a) indicates that no more than one head interceptor is assigned to each path at each level. Constraint (9b) implies that, for each path i , all the implemented head assignment levels $\{m \mid \sum_{j \in \phi_i} h_{ij}^m = 1\}$ start from 1 and form a consecutive sequence. Constraint (10) is the definition of the head level number along a specific route. Constraint (11) defines the binary variables.

3. Solution Algorithms

Solving the model of RILP is relatively difficult, as it is a non-linear, nonconvex binary integer programming model, the objective function of which contains exponential, fractional forms of decision variables. We will introduce simulated annealing (SA) to solve this model.

SA is a generic probabilistic metaheuristic for the global optimization problem of locating a good approximation to the global optimum of a given function in a large search space. SA is often used when the search space is discrete (e.g., all tours that visit a given set of cities). For certain problems, simulated annealing may be more efficient than exhaustive enumeration, provided that the goal is merely to find an acceptably good solution in a fixed amount of time, rather than to find the best possible solution. The algorithm is well

Step 1. Initialization.

(1.1) Randomly select N candidate locations to initialize an initial feasible solution x^0 . For an arbitrary route i , determine h^0 according to interceptor layout x^0 , then compute objective function value $z(x^0)$ according to (6). Let $x = x^0$.

(1.2) Set Markov length ML, initial temperature T_0 , error ε , and let $T = T_0$.

(1.3) Set the outer iteration counter $n = 1$.

Step 2. For the given T , perform the following:

(2.1) Set the inner iteration counter to $k = 1$.

(2.2) Randomly select an element from x (selected candidate locations) and $\phi - x$ (unselected candidate locations), respectively, and exchange them to produce a new solution \hat{x} .

(2.3) Compute the objective function value $z(\hat{x})$ according to (6).

(2.4) Set $\Delta z = z(\hat{x}) - z(x)$, if $\Delta z > 0$, $x = \hat{x}$; else let $x = \hat{x}$ with probability $p(\Delta z) = \exp(\Delta z/T)$.

(2.5) Check whether $k = ML$. If $k = ML$, go to Step 3; else $k = k + 1$, return to (2.2).

Step 3. Determine whether to stop

(3.1) Check whether $T < \varepsilon$. If $T < \varepsilon$, stop; go to (3.2).

(3.2) Let $T = \eta T$, $n = n + 1$, and return to Step 2. (where η decides the speed of annealing)

ALGORITHM 1: SA on RILP.

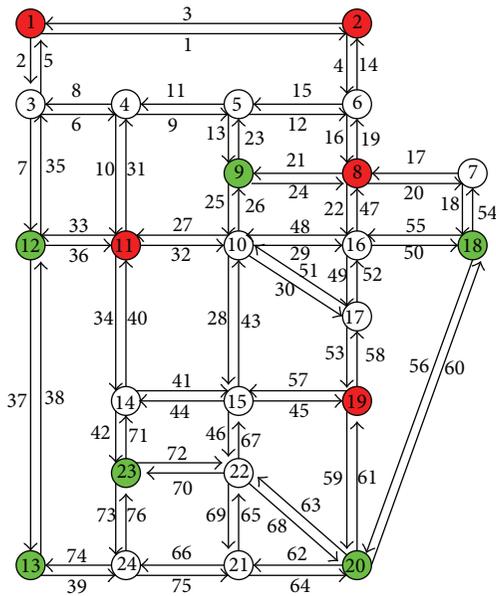


FIGURE 1: Sioux Falls network.

described by Liu [22]. A brief summary of the algorithm is given below.

(1) *Initialization.* An initial solution is generated randomly from the feasible region. The initial temperature should be high enough to allow all candidate solutions to be accepted.

(2) *Markov Length.* The iteration number M is used in each temperature. This number should be set appropriately high such that the objective function values reach a Boltzmann distribution.

(3) *Cooling Schedule.* The cooling schedule is the rate at which the temperature is reduced. In this paper, 0.8 is used for the first 12th temperature reductions. A cooling schedule of 0.8 means the temperature of the next stage is 0.8 times the

current temperature. A cooling schedule of 0.5 is used after the 12th temperature reduction.

(4) *Step Size.* Step size at each move should be decreased along with the reduction in temperature. Feasible solutions at lower temperatures are close to the optimal solution. When the temperature is low, a stochastic search tends to be a deterministic search. If the step size is too large, at low temperature, some feasible solutions will be rejected, thereby wasting computation time.

(5) *Neighboring Solutions.* Neighboring solutions are the set of feasible solutions that can be generated from the current solution. Each feasible solution can be directly reached from the current solution by a move and the resulting neighboring solution.

(6) *Stopping Criteria.* The algorithm stops when the number of temperature transitions reaches a prespecified number, or when the temperature is reduced to a threshold or when the neighboring solution was not improved after a given period of time.

The algorithm of SA on RILP is summarized in Algorithm 1.

4. Case Study

In this section, we consider a special type of violator, drunk drivers who blend in with public travelers in an urban road network and aim to find a reliable layout of interceptors (DUI checkpoints) to catch them and consequently protect public innocent travelers. Several sets of numerical experiments are performed to test the computational performance of the proposed model and solution algorithm. All algorithms are coded in MATLAB R2010b and run on a desktop computer with 2.70 GHz CPU and 4.00 GB memory.

The proposed models are applied to the Sioux Falls network, which includes 24 nodes and 76 links, as shown in Figure 1. Every node represents a zone. The nodes with

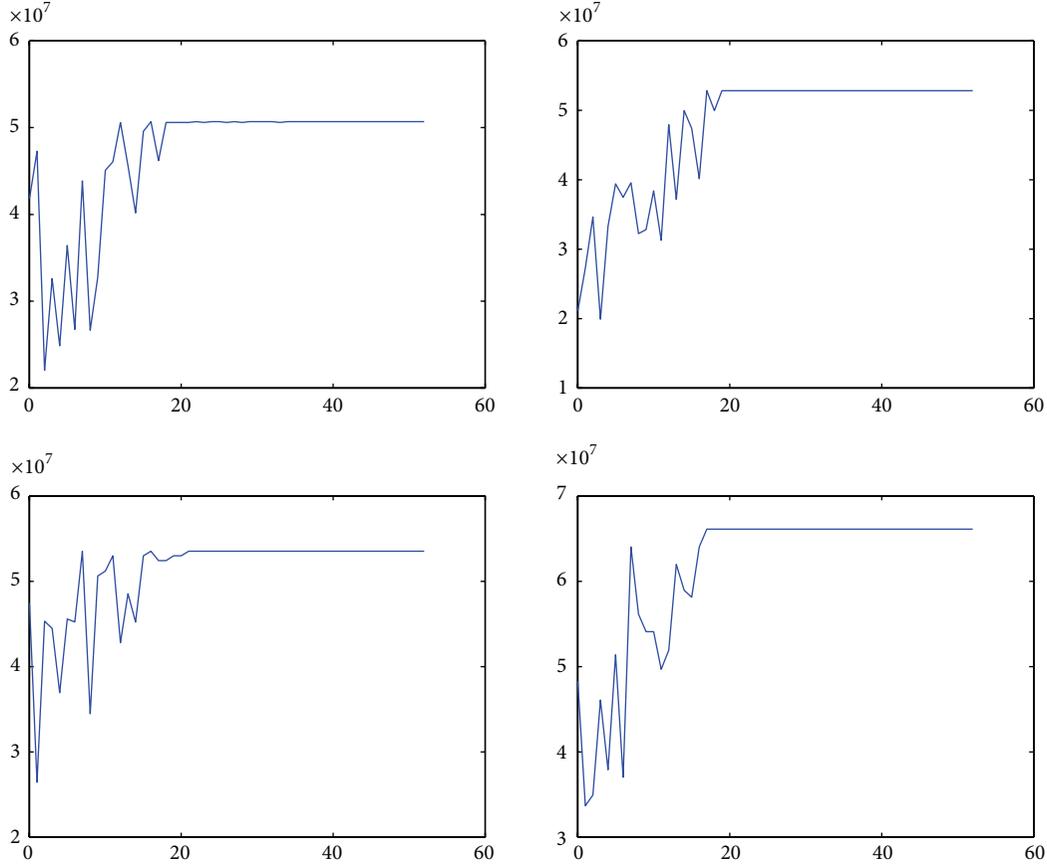


FIGURE 2: Objective function values generated by SA with $N = 5$, $T_0 = 500$, Markov length = 100, $\eta = 0.8$, $\varepsilon = 0.005$.

red/green colors are assumed to be the origins/destinations of drunk drivers. All OD trip data and network data, including the link lengths, link free flow travel times, link capacity, link characteristics (i.e., BPR function), and link indices, are downloaded from Bar-Gera [23]. Using network data and OD trip data, the traffic assignment of user equilibrium (UE) criteria [24] is implemented to obtain the link flow $f_a, \forall a \in A$. We then use f_a to calculate b_{ij} with formula (1).

The difficulty of implementing RILP lies in the exponential number of routes between each OD pair. This will be exacerbated for larger networks. However, it can reasonably be assumed that not every route will be used by drunk drivers. Based on this fact, we select a feasible and reasonable number of routes between each OD pair for drunk drivers. We only use those routes whose probabilities of being taken by drunk drivers are no less than a certain small value (e.g., 0.1). The probability that each route between a given OD pair is taken by drunk drivers is computed by the following Logit-like formula:

$$p_i = \frac{\exp(\sum_{m \in \varphi_r} t_m / t_i)}{\sum_{n \in \varphi_r} \exp(\sum_{m \in \varphi_r} t_m / t_n)}, \quad \forall r \in w, \forall i \in \varphi_r, \quad (12)$$

where t_i , t_m , and t_n are the travel times of routes i , m , and n between OD pair r , respectively, which can be obtained by the implementation of UE. The loop-less k -shortest path

algorithm presented by Yen [25] can find an arbitrary number of ranked shortest paths between each OD pair. We use that algorithm to successively find and add a new route to the set of k paths until the selected probability, computed by formula (12), is smaller than a predefined small value.

In our example, there are a total of 30 OD pairs (with 5 origins and 6 destinations) for generating and absorbing drunk drivers. The “demand” of drunk drivers between each OD pair is randomized between 1 and 100 in Excel 2010. The total number of routes generated by the k -shortest path algorithm is 99. In this example, all 24 nodes in the Sioux Falls network are assumed to be candidate locations for interceptors.

4.1. Solution Algorithm Performance. To analyze the performance of the SA algorithm in solving our RILP model, we run a series of instances for $p = 0.1$, $\beta_1 = 1000$, $\beta_2 = 2.5$, $N \in \{5, 10\}$, $T_0 = 500$, $Markov_Length = 100$, $\eta = 0.8$, and $\varepsilon = 0.005$. In each instance, the initial solution for the interceptor layout is randomized by the MATLAB function “Randperm”. The results are shown in Figures 2 and 3. As observed, the best objective function value always stabilizes after several iterations of SA in each run, and the near-optimal result is very similar across all runs for a specified N ($N = 5$ or $N = 10$). This indicates that the proposed SA algorithm for RILP is nonsensitive to the initial solution

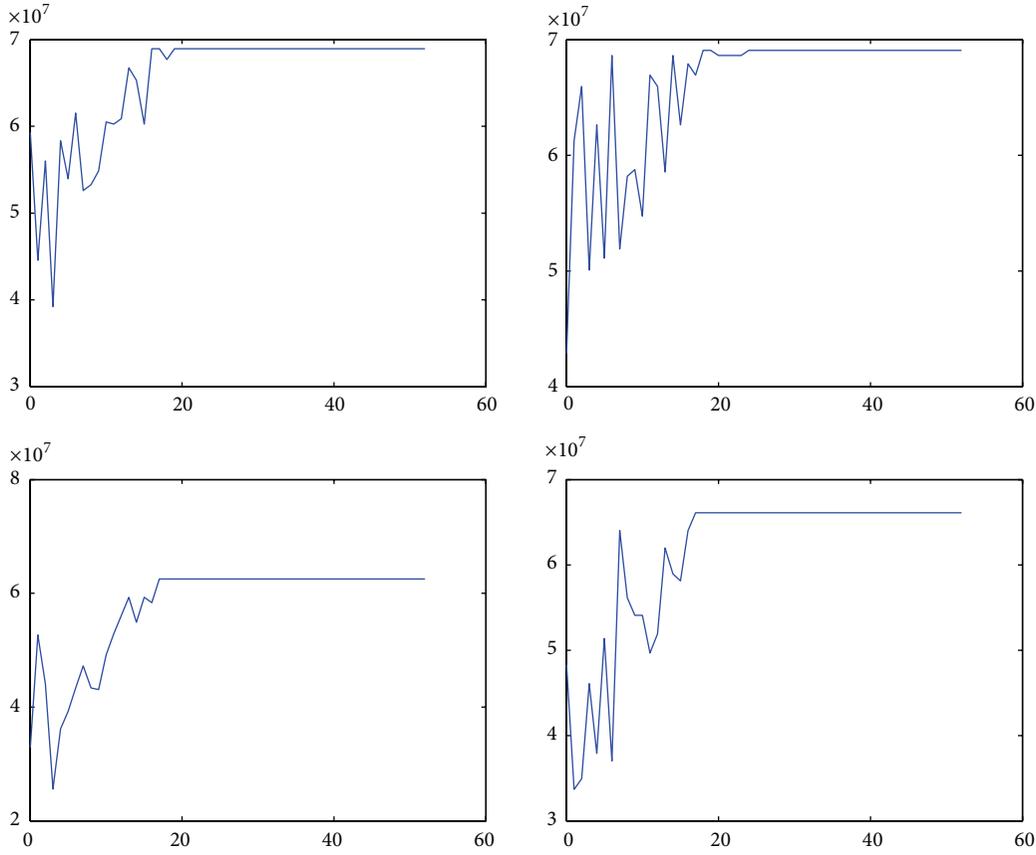


FIGURE 3: Objective function values generated by SA with $N = 10$, $T_0 = 500$, Markov length = 100, $\eta = 0.8$, $\epsilon = 0.005$.

and performs stably. Additionally, each run almost stabilizes at close to 20 iterations, and in each iteration 100 (because $Markov_length = 100$) computations are necessary to obtain objective function values. Thus, SA uses approximately 22×100 computations to approach an optimal solution, while an enumeration algorithm may cost $C_{24}^5 = 42504$ computations. Therefore, the proposed SA algorithm is more effective at solving RILP than enumeration.

4.2. Sensitivity Analysis of Disruption Probability p . In this set of experiments, we aim to investigate changes in objective values and interceptor deployment solutions under different disruption scenarios.

Using four different disruption probability (p) values, we use SA to solve RILP with N values from 1 to 24. For each N , we implement SA 4 times and use the average objective value as the final result. Figure 4 shows the benefits obtained for each p over varying numbers of installed interceptors. Overall, objective values approach stability when the number of installed interceptors is more than 5 under each disruption scenario. We call $N = 5$ the “critical point”. This is a reasonable result because we assume there are only 5 potential origins for violators, and if these 5 origins all contain installed interceptors, all of the first nodes of each violator route will be covered. Before the number of installed interceptors achieved at the “critical point”, the potential improvement of

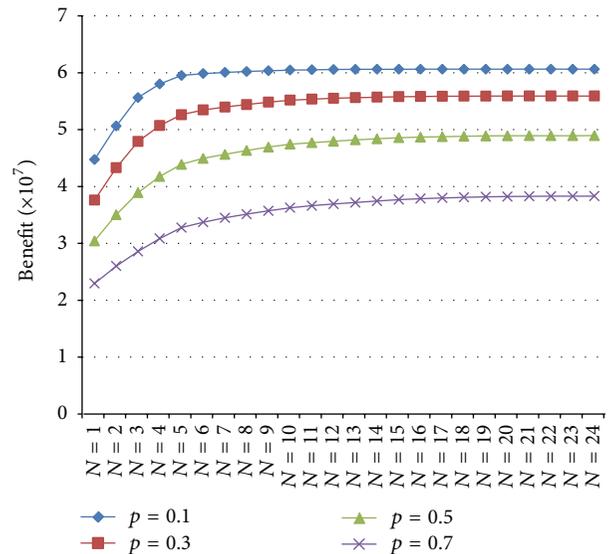


FIGURE 4: Objective values under different disruption probabilities p .

the objective value from adding a new interceptor will be much greater than that after the critical point. Under each disruption scenario, the curve of objective values increases very steeply before the “critical point”, then increases at a flatter trajectory after $N = 5$. We can see that the larger

TABLE 1: Interceptor deployment solutions under four disruption scenarios.

	Selected candidate interceptor
$N = 3$	
$p = 0.1$	8, 10, 15
$p = 0.3$	8, 10, 15
$p = 0.5$	9, 10, 15
$p = 0.7$	9, 10, 15
$N = 4$	
$p = 0.1$	8, 9, 10, 15
$p = 0.3$	8, 9, 10, 15
$p = 0.5$	8, 9, 10, 15
$p = 0.7$	9, 10, 15, 16
$N = 5$	
$p = 0.1$	8, 9, 10, 15, 17
$p = 0.3$	8, 9, 10, 15, 17
$p = 0.5$	8, 9, 10, 15, 17
$p = 0.7$	8, 9, 10, 15, 16

the disruption probability, the sharper the increase in the objective value before the “critical point” and the slower the increase after the “critical point”. Figure 4 also indicates that while N is determined, the decrease in objective values is not linear across different disruption probabilities, although we attempt to linearly increase the probability. The difference in objective values between any two neighbor disruption scenarios is exacerbated sequentially.

Table 1 presents interceptor solutions under four different disruption scenarios (with only $N = 3, 4, 5$ shown as examples). We can see that when the disruption probability varies, the “optimal” interceptor solution also changes.

4.3. Additional Discussions. In practice, β_1 represents the resultant benefit from a one unit increase in the probability of a drunk driver avoiding an interceptor, and β_2 represents the cost incurred by increasing the travel distance by one unit. One should be careful in determining the values of these two parameters because they will cause significant differences in how drunk drivers select routes and, consequently, differences in the optimal interceptor layout. If a government increases the amount of attention it pays to eliminating drunk driving, β_1 should be very large (e.g., in China, drunk driving leads to high fines and also may lead to suspended driver’s licenses and jail time).

5. Conclusion

This paper studies an interceptor location problem in an arc-node network. The interceptor identifies violators from network traffic and reduces the safety risk for the general public. We propose a reliable interceptor location model (RILP) to account for interceptor disruption and the “game” played between interceptor planners and violators. The RILP is a nonlinear, nonconvex binary integer programming model that is difficult to solve. We develop a simulated annealing (SA) algorithm to solve the RILP, and numerical case studies

are conducted to test the algorithm. Various managerial insights are also drawn from the numerical results. We find that the proposed algorithm is able to effectively and efficiently solve the RILP. Sensitivity analysis of the disruption probability shows that (i) the total benefit of the system decreases as the disruption probability increases and (ii) there exists a “critical point” for the number of installed interceptors. We further discussed the physical meanings of two parameters in a Logit model for the selection of routes by drunk drivers and how to determine these parameters.

In this study, we consider a heuristic algorithm, SA. Because of the uncertainty inherent in heuristic methods, future studies should aim to develop global optimal algorithms to solve this nonlinear programming model.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (Project no. 71203045), Heilongjiang Natural Science Foundation (Project no. E201318), and the Fundamental Research Funds for the Central Universities (Grant no. HIT.KISTP.201421). This work was performed at the Key Laboratory of Advanced Materials & Intelligent Control Technology on Transportation Safety, Ministry of Communications, China.

References

- [1] B. Hubicka, H. Laurell, and H. Bergman, “Criminal and alcohol problems among Swedish drunk drivers—predictors of DUI relapse,” *International Journal of Law and Psychiatry*, vol. 31, no. 6, pp. 471–478, 2008.
- [2] S. Pavanello, R. Snenghi, A. Nalesso, D. Sartore, S. D. Ferrara, and M. Montisci, “Alcohol drinking, mean corpuscular volume of erythrocytes, and alcohol metabolic genotypes in drunk drivers,” *Alcohol*, vol. 46, no. 1, pp. 61–68, 2012.
- [3] M. Portman, A. Penttilä, J. Haukka et al., “Profile of a drunk driver and risk factors for drunk driving. Findings in roadside testing in the province of Uusimaa in Finland 1990–2008,” *Forensic Science International*, vol. 231, no. 1–3, pp. 20–27, 2013.
- [4] S. Velmurugan, S. Padma, E. Madhu, S. Anuradha, and S. Gangopadhyay, “A study of factors influencing the severity of road crashes involving drunk drivers and non drunk drivers,” *Research in Transportation Economics*, vol. 38, no. 1, pp. 78–83, 2013.
- [5] National Highway Traffic Safety Administration (NHTSA), *Traffic Safety Facts: Alcohol-Impaired Driving*. US Department of Transportation, Washington, DC, USA, 2012, <http://www-nrd.nhtsa.dot.gov/Pubs/811606.pdf>.
- [6] C. F. Greer and K. D. Moreland, “United Airlines’ and American Airlines’ online crisis communication following the September 11 terrorist attacks,” *Public Relations Review*, vol. 29, no. 4, pp. 427–441, 2003.
- [7] V. S. Guzhva and N. Pagiavlas, “US Commercial airline performance after September 11, 2001: decomposing the effect of the

- terrorist attack from macroeconomic influences,” *Journal of Air Transport Management*, vol. 10, no. 5, pp. 327–332, 2004.
- [8] H. Ito and D. Lee, “Assessing the impact of the September 11 terrorist attacks on U.S. airline demand,” *Journal of Economics and Business*, vol. 57, no. 1, pp. 75–95, 2005.
- [9] K. V. Smith, C. Mansfield, and H. A. Klaiber, “Terrorist threats, information disclosures, and consumer sovereignty,” *Information Economics and Policy*, vol. 25, no. 4, pp. 225–234, 2013.
- [10] R. Zenklusen, “Network flow interdiction on planar graphs,” *Discrete Applied Mathematics*, vol. 158, no. 13, pp. 1441–1455, 2010.
- [11] D. Granata, G. Steeger, and S. Rebennack, “Network interdiction via a critical disruption path: branch-and-price algorithms,” *Computers & Operations Research*, vol. 40, no. 11, pp. 2689–2702, 2013.
- [12] M. A. Rad and H. T. Kakhki, “Maximum dynamic network flow interdiction problem: new formulation and solution procedures,” *Computers and Industrial Engineering*, vol. 65, no. 4, pp. 531–536, 2013.
- [13] J. Yates and S. Sanjeevi, “A length-based, multiple-resource formulation for shortest path network interdiction problems in the transportation sector,” *International Journal of Critical Infrastructure Protection*, vol. 6, no. 2, pp. 107–119, 2013.
- [14] K. J. Cormican, D. P. Morton, and R. K. Wood, “Stochastic network interdiction,” *Operations Research*, vol. 46, no. 2, pp. 184–197, 1998.
- [15] J. E. Ramirez-Marquez and C. M. Rocco S., “Stochastic network interdiction optimization via capacitated network reliability modeling and probabilistic solution discovery,” *Reliability Engineering and System Safety*, vol. 94, no. 5, pp. 913–921, 2009.
- [16] P. B. Mirchandani, M. Gentili, and Y. He, “Location of vehicle identification sensors to monitor travel-time performance,” *IET Intelligent Transport Systems*, vol. 3, no. 3, pp. 289–303, 2009.
- [17] X. Li and Y. F. Ouyang, “Reliable sensor deployment for network traffic surveillance,” *Transportation Research B: Methodological*, vol. 45, no. 1, pp. 218–231, 2011.
- [18] Y. F. Ouyang, X. Li, C. P. L. Barkan, A. Kawprasert, and Y. C. Lai, “Optimal locations of railroad wayside defect detection installations,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 24, no. 5, pp. 309–319, 2009.
- [19] F. O. Hounkpevi and E. E. Yaz, “Robust minimum variance linear state estimators for multiple sensors with different failure rates,” *Automatica*, vol. 43, no. 7, pp. 1274–1280, 2007.
- [20] L. V. Snyder and M. S. Daskin, “Reliability models for facility location: the expected failure cost case,” *Transportation Science*, vol. 39, no. 3, pp. 400–416, 2005.
- [21] X. Li and Y. Ouyang, “A continuum approximation approach to reliable facility location design under correlated probabilistic disruptions,” *Transportation Research B: Methodological*, vol. 44, no. 4, pp. 535–548, 2010.
- [22] C. Q. Liu, *Advanced Traffic Planning*, People's Traffic Press, 2001.
- [23] H. Bar-Gera, “Transportation network test problems,” 2001, <http://www.bgu.ac.il/~bargera/tntp/>.
- [24] Y. Sheffi, *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*, Prentice Hall, England Cliffs, NJ, USA, 1985.
- [25] J. Y. Yen, “Finding the k shortest loopless paths in a network,” *Management Science*, vol. 17, pp. 712–716, 1971.

Research Article

Unrecorded Accidents Detection on Highways Based on Temporal Data Mining

Shi An, Tao Zhang, Xinming Zhang, and Jian Wang

School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin 150090, China

Correspondence should be addressed to Xinming Zhang; 12b332001@hit.edu.cn

Received 3 April 2014; Revised 26 May 2014; Accepted 26 May 2014; Published 15 June 2014

Academic Editor: Hamid Reza Karimi

Copyright © 2014 Shi An et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Automatic traffic accident detection, especially not recorded by traffic police, is crucial to accident black spots identification and traffic safety. A new method of detecting traffic accidents is proposed based on temporal data mining, which can identify the unknown and unrecorded accidents by traffic police. Time series model was constructed using ternary numbers to reflect the state of traffic flow based on cell transmission model. In order to deal with the aftereffects of linear drift between time series and to reduce the computational cost, discrete Fourier transform was implemented to turn time series from time domain to frequency domain. The pattern of the time series when an accident happened could be recognized using the historical crash data. Then taking Euclidean distance as the similarity evaluation function, similarity data mining of the transformed time series was carried out. If the result was less than the given threshold, the two time series were similar and an accident happened probably. A numerical example was carried out and the results verified the effectiveness of the proposed method.

1. Introduction

Road accidents are regarded as one of the leading causes of death for people between the ages of 5 and 44 according to the World Health Organization [1]. More than that, traffic crashes result in serious economic losses on account of traffic congestion which in turn leads to a wide variety of adverse consequences such as traffic delays, supply chain interruptions, travel time unreliability, and increased noise pollution, as well as deterioration of air quality [2]. Thus, reducing or avoiding traffic collisions is of great significance to traffic safety. High collision concentration location (HCCL) [3] detection is an effective means to find out the accident black spots and take some necessary continuous improvement measures. Historical accident data is the necessary foundation of any research on this subject. One of the main problems of the accident data was considered as the heterogeneity [4] in the previous studies. A great many methods had been proposed to solve this problem, such as latent class clustering [5–8], Bayesian networks (BNs) [9–11], and continuous risk profile (CRP) [12, 13]. One of the commonalities between these methods is that historical accident data, more specifically, recorded historical accident data, were used as the basic data. However,

not all traffic crashes are known and recorded by traffic police. It is undeniable that some minor accidents often happened on highways and were settled privately for trivial losses. Usually, traffic accident black spots are identified mainly based on the traffic crash data recorded by traffic police departments [14]. This just helps to find out the collisions we know, while ones we do not know, which did happen, keep unconsidered yet. In part, these observations motivated our study.

Traffic accidents are contingent events and are difficult to detect if there is no alarm. Nonetheless, a traffic accident was bound to create an impact on traffic flow pattern and cause different levels of congestion [15, 16]. The traffic volume, traffic speed, and traffic density were changed by crashes, even minor accidents. Thus, capturing the change of these traffic flow parameters is very helpful for traffic accidents, especially unrecorded traffic accidents detection.

Given this, an automatic traffic accident detection method was proposed in this paper. Traffic flow data was used and simulated by cell transmission model (CTM). According to the different inflow between two cells, time series model was constructed to reflect traffic flow state. Time series pattern, when accidents happened, was established based on historical accident data. To overcome the defect

of Euclidean distance, which does not consider the linear drift in the time domain, and to reduce the computational cost, discrete Fourier transform was implemented to turn the time series from time domain to frequency domain. Leveraging the strengths of temporal data mining at finding time-varying patterns, any time series that were similar to the given pattern could be figured out, namely, the unknown and unrecorded accidents, by similarity search. The premier aim and contribution of this paper are to find out “the hidden accidents,” such as compounding in private, using temporal data mining method. A case study using the real highway traffic data in Harbin, China, was conducted for verification.

2. Material and Methods

2.1. Construction of Time Series Reflecting Traffic Flow State. For highway traffic flow, there was a unique state in each period. One of the ways to describe this was the metaphor of a screen capture for the traffic flow over consecutive periods of time. Every picture reflected the state of the traffic flow at a certain time. These pictures constituted a sequence over time. This was the consideration of building the time series model to describe the evolution of traffic flow in this paper. Traffic conditions estimation was achieved through dynamic traffic assignment (DTA) simulation that utilized temporal aspects of a transportation system. Different values of inflow between cells in CTM were typically expressed as ternary numbers (0, 1, and 2). A series of ternary numbers, generated by CTM, were introduced to illustrate the traffic flow state. Then, time series data were created by converting ternary numbers to decimal numbers. Thus, the traffic flow state could be reflected by time series data, and it was the basic work for unrecorded accidents mining.

2.1.1. Cell Transmission Model. To model the propagation of traffic flow and construct time series data in the section below, the spread of highway traffic flow was simulated by CTM in this paper. CTM was proposed by Daganzo [17, 18] and was considered as a proper method. It was believed that the relationship between traffic flow (q) and density (k) was of the form depicted figurally as follows:

$$q = \min \{vk, q_{\max}, \omega(k_j - k)\}, \quad (1)$$

where v , q_{\max} , ω , and k_j denoted the free-flow speed, the maximum flow (or capacity), the backward wave speed, and the maximum (or jam) density, respectively, as shown in Figure 1.

In Figure 1, if the density is less than k_1 , the traffic flow q is equal to vk ; if it is between k_1 and k_2 , then q reaches its maximum, q_{\max} ; if it is between k_2 and k_j , q is equal to $\omega(k_j - k)$; and q is 0 when the density reaches k_j .

Then, the continuous Lighthill-Whitham-Richards (LWR) equations [19, 20] for a single highway link were discretized through this method and could be approximated by a set of difference equations. The state of the system was updated over time. Thus, the discontinuous changes of traffic flow could be captured.

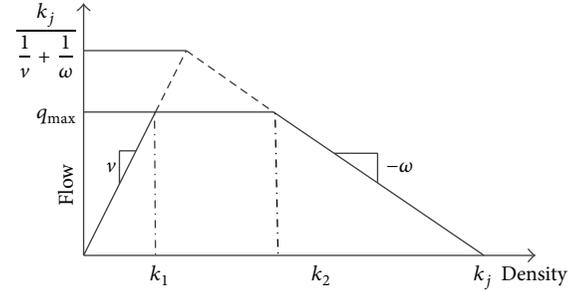


FIGURE 1: The relationship of traffic flow and density in CTM.

In CTM, a single road was divided into homogeneous sections (cells), i , whose lengths equaled the distance traveled by free-flowing traffic speed in one clock interval. The state of the system at instant t was then given by the number of vehicles contained in each cell, $n_i(t)$. The following parameters were defined for each cell.

$N_i(t)$ is the maximum number of vehicles that can be present in cell i at time t , and $Q_i(t)$ is the maximum number of vehicles that can flow into cell i when the clock advanced from t to $t + 1$.

These constants could vary with time (e.g., contingent traffic incidents or conscious traffic control measures), but this dependence was able to be ignored for simplicity of notation. The first constant $N_i(t)$ was defined to be the product of the cell’s length and its jam density, and the second one was the product of the time interval and the cell’s capacity.

If cells were numbered consecutively starting with the upstream end of the road from $i = 1$ to I , the recursive relationship of the CTM, as discussed by Daganzo [17, 18], could be expressed as

$$n_i(t + 1) = n_i(t) + y_i(t) - y_{i+1}(t), \quad (2a)$$

where $y_i(t)$ was the inflow to cell i in the time interval $(t, t + 1)$, given by

$$y_i(t) = \min \{n_{i-1}(t), Q_i(t), \delta [N_i(t) - n_i(t)]\}, \quad (2b)$$

where $\delta = \omega/v$.

The formulas (2a) and (2b) constituted the fundamental equations of CTM. Equation (2a) expressed the status updates of cells over time, while the latter gave the variations for updating.

2.1.2. Constructing Time Series of Traffic Flow. Time series data was a sequence of data evolving through time. There were two strengths of temporal data mining: data-based and pattern-based. The former was more likely to approach the truth; the latter was more likely to extract the features. Every traffic accident was considered to change the traffic flow state more or less. Features of this change were supposed to be extracted by temporal data mining and these features were able to be used to find out “the hidden accidents.”

Gao et al. used the NaSch traffic model to simulate the evolution of traffic flow [21, 22]. The state of traffic flow at

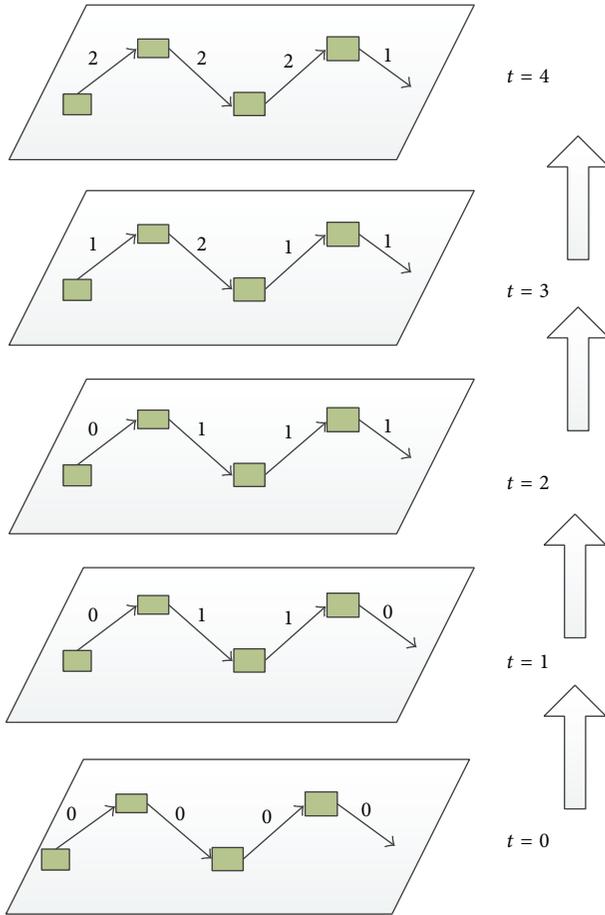


FIGURE 2: The traffic flow trends with the tick of a clock.

each time period was regarded as a node of a network. Then, a complex network was constructed, also known as a multiple-mode system model, which could describe the evolution of traffic flow. Zhao et al. studied state estimation of this kind of systems [23]. However, the temporal relations among these nodes were ignored in their research.

According to CTM, at each time step, the inflow $y_i(t)$ to cell i could be $n_{i-1}(t)$, $Q_i(t)$ or $\delta[N_i(t) - n_i(t)]$. When $y_i(t) = n_{i-1}(t)$, the state of the cell was represented by number 0; when $y_i(t) = Q_i(t)$, it was represented by number 1, otherwise represented by number 2. Then, the state of the system was expressed by a sequence of ternary numbers, for example, $\{0, 0, 1, 2, 1\}$.

The model employed was described as follows. For N cells, we assumed that, at time period $(t, t + 1)$, the state was represented by a set of ternary numbers, namely, $S_t = \{s_1, s_2, \dots, s_N\}$. Then, S_t is the precursor of S_{t+1} . Here, each ternary number s_i can be considered as an element, which can take three different states; that is, $s_i = 0, 1, 2$. Figure 2 depicted the evolution of traffic flow with the tick of a clock.

As shown in Figure 2, when the clock advanced from 0 to 4, the system (a single segment) state change could be discovered clearly using the “screen capturing” method. The state S_t was time-varying and was represented by a sequence of numbers. For convenience, a parameter M was introduced

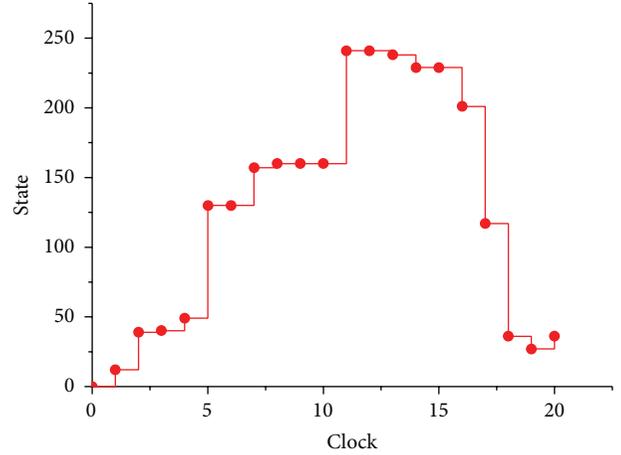


FIGURE 3: Time series data reflecting traffic flow trends of a five-cell road.

in this paper to represent the value of S_t and $M = \sum_{i=1}^N 3^i s_i$; thus, ternary numbers were converted to decimal numbers; for example, $\{0, 0, 1, 2, 1\}$ was converted to 16. Thus, time series data was created (see Figure 3).

Each single decimal number represented a system’s state. As shown in Figure 3, at time step 9, for instance, the value of the system state was 78; thus, the ternary numbers were $\{0, 2, 2, 2, 0\}$, which was the state of traffic flow.

2.2. Feature Extraction. Noise in the raw time series data could reduce accuracy and creditability of data mining. Linear drift was certainly an example. In many clustering analysis methods, K -means, for example, Euclidean distance, was frequently used as a similarity measure function. The unrecorded accidents detection method proposed in this paper was mostly based on similarity mining, which would be discussed later. Linear drift was the most important factor to influence the accuracy of the results.

For time series, $X = \{x_1, x_2, \dots, x_n\}$ and $Y = \{y_1, y_2, \dots, y_n\}$, if, at time t , $x_i - y_i = \epsilon$, $i = 1, \dots, n$, where ϵ was a constant. That was to say that the relationship between X and Y was linear drift in the time domain, as shown in Figure 4.

In this case, if Euclidean distance was used and ϵ was beyond the threshold, these two time series X and Y were not considered to be similar by mining algorithms. While they had similar shape and trend apparently, the judging result was inaccurate obviously. Aiming to prevent such errors and to realize data compression and reduce the computational cost, it was necessary to extract feature from the original time series data, using the image in feature space to replace the original one.

Discrete Fourier transform (DFT), which had unique merits in time series analysis, was an alternative way. For a given time series object, DFT could be used to turn it to frequency domain from time domain. According to Parseval theory, the time-domain energy function was equal to the frequency-domain energy function. And most of the energy in frequency domain concentrated on the first few

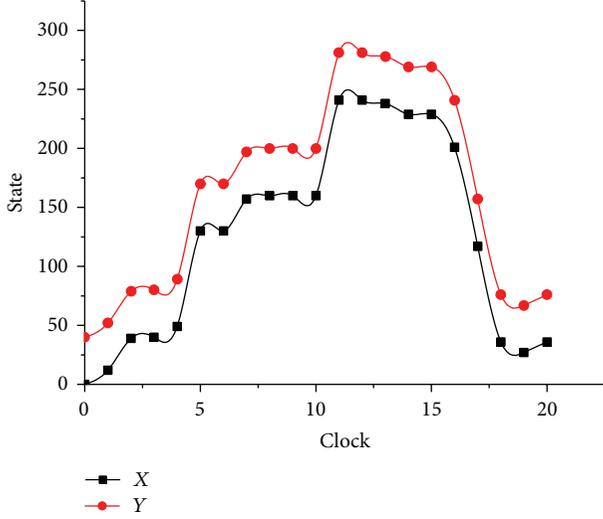


FIGURE 4: Linear drift between time series X and Y.

coefficients; hence, other coefficients could be omitted. Thus, the remaining coefficients were able to be seen as the features of the original time series.

For a given time series $X = \{x_t\}$, $t = 0, 1, \dots, n-1$, translating the series from time domain to frequency domain, the new sequence obtained by DFT was denoted by $\bar{X} = \{x'_f\}$, $f = 0, 1, \dots, n-1$, where

$$x'_f = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} x_t \exp\left(-\frac{j2\pi ft}{n}\right), \quad f = 0, 1, \dots, n-1. \quad (3)$$

Taking the data in Figure 3, for example, the result of DFT was shown in Figure 5.

The area, in which the frequency was greater than 0.0, was the real transformed data of the original finite time series. 17 elements were left out of the initial total of 21 ones. Data compression had been realized and as a result of the transformation from time domain to frequency domain, linear drift no longer existed as well as other noises in the time domain.

2.3. Unrecorded Accidents Detection Using Similarity Mining.

Similarity search is an important research field in temporal data mining. As mentioned above, each recorded accident could bring a piece of time series data, and all recorded historical accidents data over a period of time could explain the traffic flow trends when accidents happened. After data processing, using the above methods, which could be called data preprocessing, clustering analysis was supposed to be implemented in this paper. The classical K -means method would be able to meet the accuracy requirements due to the appropriate data preprocessing. Results of cluster were considered as “normal traffic flow trends” under accidents and the “hidden accidents” could be found out by similarity search.

The method of similarity measurement used in this paper was Euclidean distance. Set the time sequence of “normal

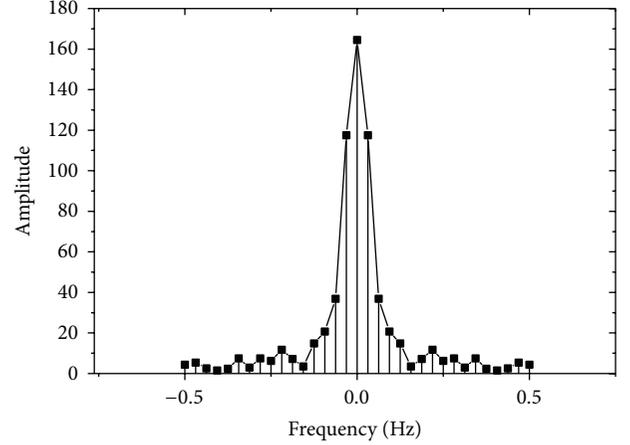


FIGURE 5: Discrete Fourier transform.

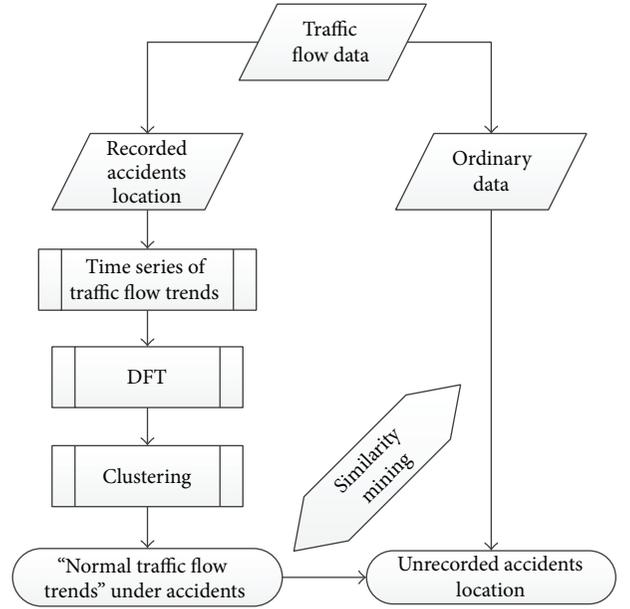


FIGURE 6: The procedure of unrecorded accidents detection.

traffic flow trends” of a single road segment which was $\{x_i\}$, whose length was n . The time series to be measured was denoted by $\{y_i\}$ with its length N , $N \geq n$ generally. In similarity search, subsequences of $\{y_i\}$, whose lengths were n , were the measuring object. These subsequences were denoted by $\{z_i\}$. It was known that the number of $\{z_i\}$ was J , $J = N - n + 1$. Then, the similarity metric function could be defined as follows:

$$\min_J \sum_{i=1}^n (x_i - K_J z_i^J)^2, \quad (4)$$

where K_J was the scaling factor. The calculation times were $N - n + 1$ obviously.

So far, for a single road segment, the procedure to detect unrecorded accidents was described in Figure 6.

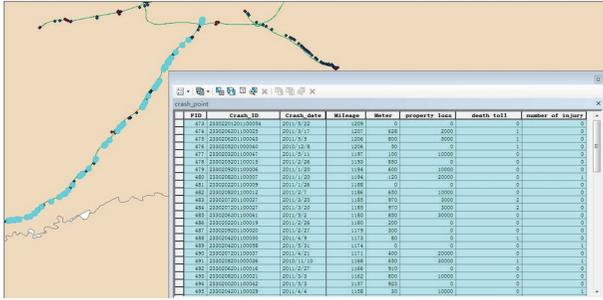


FIGURE 7: The study site and data.

3. Results and Discussions

3.1. The Study Site and Data Preparation. A case study was conducted based on the data extracted from records collected on Beijing-Harbin Expressway (G1) between Harbin and Lalinhe from January 2010 to July 2011. The traffic accidents dataset included a total of 73 crashes recorded on the 72-kilometer length road in total. The study site and the traffic accidents data were shown in Figure 7.

In one and a half years, for such a highway with the annual traffic reaching 8–10 million, “only” 73 crashes happened. Experience told us that it did not indicate how safe the highway was, but some minor accidents were not known or recorded. It was meaningful to detect the unrecorded collisions for traffic safety research.

As the limit of detection device, traffic flow data worked out using data collected by highway toll collection system. For detailed calculation process, readers could refer to Weng et al. [24]. The estimation accuracy could meet the requirements technically.

3.2. Numerical Experiment. In the numerical experiment, the time horizon was set within half an hour and the time interval was 30 seconds. The accidents happened at the eleventh clock interval, namely, 5 minutes after the start time of simulation. Thus, there were 60 points in one piece of time series data. The free-flow speed was 120 km/h; thus, the length of each cell was the product of the free-flow speed and the unit clock interval that was 120 km/h * 30 s = 1km. Based on the historical accident data, the accident location was set in the middle cell in CTM and five cells were brought in to represent a single highway segment where a crash happened, as shown in Figure 8.

The virtual cell in Figure 8 was on behalf of the demand generation. There were five inflow states, as shown by the five arrows in the figure above. Using the method shown in Figure 6, first of all, it was significant to identify the “normal traffic flow trends” under accidents. Time series data was conducted using the method mentioned above and DFT was implemented then for each crash record. K-means cluster analysis was adopted and the results were indicated in Figure 9.

As is shown in Figure 9, two clusters were generated by this method. In Figure 9(a), there were 34 accidents gathered together. The common point of them was that congestion formed and dissipated gradually in the simulation period. The

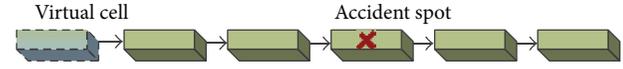


FIGURE 8: CTM for a single highway segment.

shapes of these time series data were similar to the normal distribution curve. According to the accident records of traffic police, most of these crashes were single-vehicle crashes and both-car accidents. In Figure 9(b), 33 crashes were involved and the congestion was not eliminated in study period. Two-thirds of them were crashes between two heavy trucks and one-third were multivehicle accidents. This kind of crashes could cause severe congestion and influence traffic seriously. Actually, another cluster existed in this case and only five accidents were involved. This cluster had less interference to traffic, but, for the records, they were collisions between vehicles and pedestrians. It was a serious threat to traffic safety but beyond the scope of this paper, so the result was not shown here.

For the unrecorded accidents, usually there were no casualties and little losses. The reason why they were not recorded was that these minor accidents were settled privately and even kept unknown by the traffic police. So, there was every reason to believe that these unrecorded accidents were single-vehicle crashes or both-car accidents. Thus, the cluster shown in Figure 9(a) was our target in the similarity search. For verification, one accident out of the 73 crashes was separated out, pretending to be unknown.

By similarity search, the “prepared” accident was found out and the time series data was shown in Figure 10. The traffic flow trend in Figure 10 was indeed similar to the cluster in Figure 9(a). At the first 10 clock intervals in both figures, the traffic was smooth because of the unsaturated traffic flow and no accidents. Then, congestion formed and dissipated gradually in a certain period. The result of similarity search proved the reliability of the proposed method.

4. Conclusions

Unrecorded accidents were significant to identify traffic accident-prone location. Based on the observation that the traffic volume, traffic speed, and traffic density were changed by crashes, even minor accidents, an automatic traffic accident identification method was proposed. As most of the current studies did not pay enough attention to the time factor when studying the relationship between traffic state and crashes on highways, this paper proposed a method to construct time series data using traffic flow data when accidents happened. To avoid the defect of not considering the linear drift in the time domain between two sequences, DFT was carried out to extract features from original time series data. Traffic flow trend could be well understood by clustering analysis. Then, through the method of similarity search, unrecorded accidents, which were believed to be single-vehicle crashes or both-car accidents, were found out. The case study using real data in Harbin showed the feasibility of the proposed method.

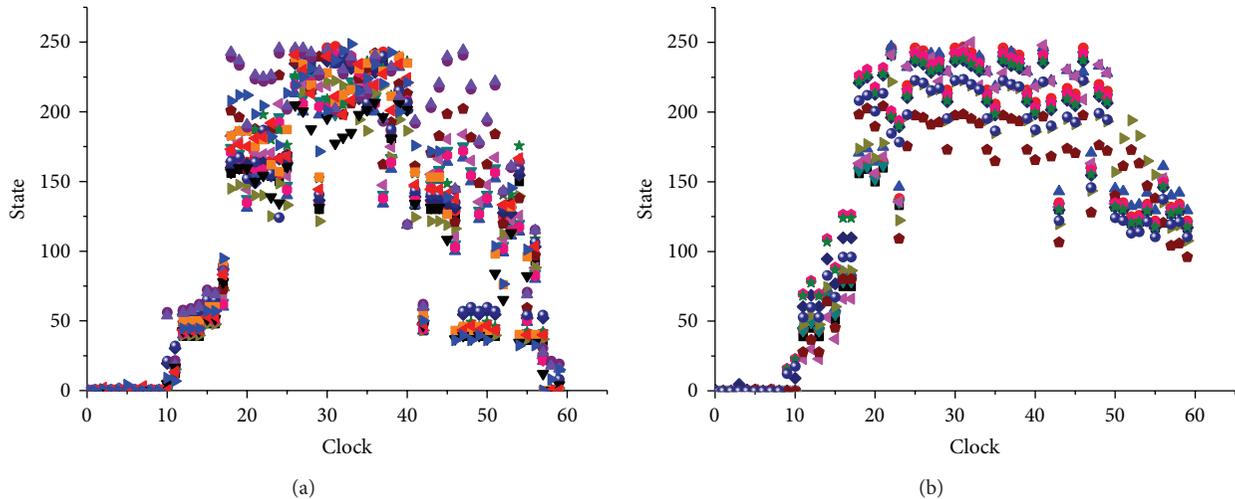


FIGURE 9: Results of cluster.

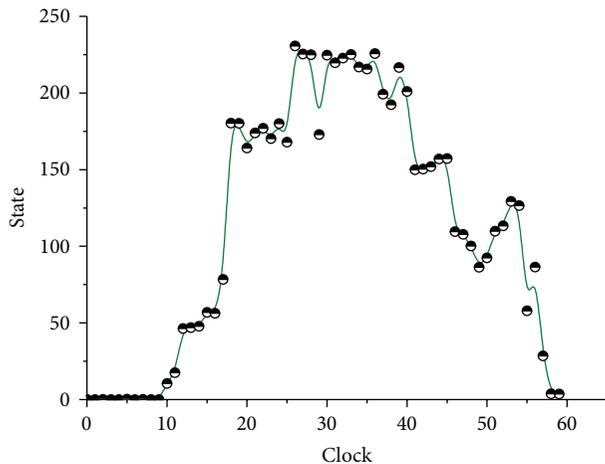


FIGURE 10: Result of similarity search.

For further research, data of car insurance would be valuable for data mining in this area or for verification of the automatic traffic accidents identification. And further study could focus on the traffic flow trends under accidents, such as the influence diffusion and the elimination of the crash influence.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the HTRDP (863) under Grant no. 2012AA112310 and Research Fund for the Doctoral Program of Higher Education of Ministry of Education of China (20112302110054).

References

- [1] E. Krug, "Decade of action for road safety 2011–2020," *Injury*, vol. 43, no. 1, pp. 6–7, 2012.
- [2] A. Gregoriades and K. C. Mouskos, "Black spots identification through a Bayesian Networks quantification of accident risk index," *Transportation Research C: Emerging Technologies*, vol. 28, pp. 28–43, 2013.
- [3] O. H. Kwon, M. J. Park, H. Yeo, and K. Chung, "Evaluating the performance of network screening methods for detecting high collision concentration locations on highways," *Accident Analysis and Prevention*, vol. 51, pp. 141–149, 2013.
- [4] P. T. Savolainen, F. L. Mannering, D. Lord, and M. A. Quddus, "The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives," *Accident Analysis and Prevention*, vol. 43, no. 5, pp. 1666–1676, 2011.
- [5] B. Depaire, G. Wets, and K. Vanhoof, "Traffic accident segmentation by means of latent class clustering," *Accident Analysis and Prevention*, vol. 40, no. 4, pp. 1257–1266, 2008.
- [6] J. M. Pardillo-Mayora, C. A. Dominguez-Lira, and R. Jurado-Piña, "Empirical calibration of a roadside hazardousness index for Spanish two-lane rural roads," *Accident Analysis and Prevention*, vol. 42, no. 6, pp. 2018–2023, 2010.
- [7] B.-J. Park and D. Lord, "Application of finite mixture models for vehicle crash data analysis," *Accident Analysis and Prevention*, vol. 41, no. 4, pp. 683–691, 2009.
- [8] B.-J. Park, D. Lord, and J. D. Hart, "Bias properties of Bayesian statistics in finite mixture of negative binomial regression models in crash data analysis," *Accident Analysis and Prevention*, vol. 42, no. 2, pp. 741–749, 2010.
- [9] M. Simoncic, "A Bayesian network model of two-car accidents," *Journal of Transportation and Statistics*, vol. 7, no. 2-3, pp. 13–25, 2005.
- [10] J. De Oña, R. O. Mujalli, and F. J. Calvo, "Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks," *Accident Analysis and Prevention*, vol. 43, no. 1, pp. 402–411, 2011.
- [11] R. O. Mujalli and J. De Oña, "A method for simplifying the analysis of traffic accidents injury severity on two-lane highways using Bayesian networks," *Journal of Safety Research*, vol. 42, no. 5, pp. 317–326, 2011.

- [12] K. Chung, K. Jang, S. Madanat, and S. Washington, "Proactive detection of high collision concentration locations on highways," *Transportation Research A: Policy and Practice*, vol. 45, no. 9, pp. 927–934, 2011.
- [13] K. Chung, D. R. Ragland, S. Madanat, and S. M. Oh, "The continuous risk profile approach for the identification of high collision concentration locations on congested highways," in *Proceedings of the 19th International Symposium on Transportation & Traffic Theory (ISTTT '09)*, pp. 463–480, 2009.
- [14] G. Vandenbulcke, I. Thomas, and L. Int Panis, "Predicting cycling accident risk in Brussels: a spatial case-control approach," *Accident Analysis and Prevention*, vol. 62, pp. 341–357, 2014.
- [15] Z. Zheng, "Empirical analysis on relationship between traffic conditions and crash occurrences," *Procedia-Social and Behavioral Sciences*, vol. 43, pp. 302–312, 2012.
- [16] A. Hamzehei, E. Chung, and M. Miska, "Pre-crash and non-crash traffic flow trends analysis on motorways," in *Proceedings of the Australasian Transport Research Forum*, 2013.
- [17] C. F. Daganzo, "The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation Research B: Methodological*, vol. 28, no. 4, pp. 269–287, 1994.
- [18] C. F. Daganzo, "The cell transmission model, part II: network traffic," *Transportation Research B: Methodological*, vol. 29, no. 2, pp. 79–93, 1995.
- [19] M. J. Lighthill and G. B. Whitham, "On kinematic waves I: flow movement in long rivers. II: a theory of traffic flow on long crowded roads," *Proceedings of the Royal Society of London A*, vol. 229, no. 1178, pp. 281–316, 1955.
- [20] P. I. Richards, "Shock waves on the highway," *Operations Research*, vol. 4, pp. 42–51, 1956.
- [21] G. Zi-You and L. Ke-Ping, "Evolution of traffic flow with scale-free topology," *Chinese Physics Letters*, vol. 22, no. 10, pp. 2711–2714, 2005.
- [22] X.-G. Li, Z.-Y. Gao, K.-P. Li, and X.-M. Zhao, "Relationship between microscopic dynamics in traffic flow and complexity in networks," *Physical Review E*, vol. 76, no. 1, Article ID 016110, 2007.
- [23] X. Zhao, H. Liu, J. Zhang, and H. Li, "Multiple-mode observer design for a class of switched linear systems," *IEEE Transactions on Automation Science and Engineering*, 2013.
- [24] J.-C. Weng, L.-L. Liu, and B. Du, "ETC data based traffic information mining techniques," *Journal of Transportation Systems Engineering and Information Technology*, vol. 10, no. 2, pp. 57–63, 2010.

Research Article

Research on Coordinated Robotic Motion Control Based on Fuzzy Decoupling Method in Fluidic Environments

Wei Zhang,¹ Haitian Chen,¹ Tao Chen,¹ Zheping Yan,¹ and Hongliang Ren²

¹ College of Automation, Harbin Engineering University, Harbin, Heilongjiang 150001, China

² Department of Bioengineering, National University of Singapore, Singapore 117575

Correspondence should be addressed to Tao Chen; chentao.7777@163.com

Received 7 January 2014; Revised 29 April 2014; Accepted 14 May 2014; Published 11 June 2014

Academic Editor: Michael Lütjen

Copyright © 2014 Wei Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The underwater recovery of autonomous underwater vehicles (AUV) is a process of 6-DOF motion control, which is related to characteristics with strong nonlinearity and coupling. In the recovery mission, the vehicle requires high level control accuracy. Considering an AUV called BSAV, this paper established a kinetic model to describe the motion of AUV in the horizontal plane, which consisted of nonlinear equations. On the basis of this model, the main coupling variables were analyzed during recovery. Aiming at the strong coupling problem between the heading control and sway motion, we designed a decoupling compensator based on the fuzzy theory and the decoupling theory. We analyzed to the rules of fuzzy compensation, the input and output membership functions of fuzzy compensator, through compose operation and clear operation of fuzzy reasoning, and obtained decoupling compensation quantity. Simulation results show that the fuzzy decoupling controller effectively reduces the overshoot of the system, and improves the control precision. Through the water tank experiments and analysis of experimental data, the effectiveness and feasibility of AUV recovery movement coordinated control based on fuzzy decoupling method are validated successful, and show that the fuzzy decoupling control method has a high practical value in the recovery mission.

1. Introduction

Autonomous underwater vehicles (AUV) are widely used in hydrographical surveys, security check near the coasts of ports, marine environmental monitoring, and salvage operations. The data collected by AUV is very important, which means that the vehicle needs to be recovered safely after missions. Voluminous literatures have been presented in AUV recovery area, which could be basically classified into 2 types:

- (1) surface recovery, which is conducted by a surface mother ship;
- (2) underwater recovery, which is carried out by using an underwater lifting platform for underwater docking recovery operation.

The latter 2nd type is AUV clients' favour for its convenience in recovery operation, where the simulation results and pool experiments have demonstrated the feasibility of this type.

Currently, there are some approaches for the AUV underwater recovery as follows.

- (1) Pole docking: this recovery method needs to equip the recovery mechanism which can capture the rope or the pole target. Typical representative is Odyssey IIB which was developed by Woods Hole Marine research institute and the MIT and Starbug MKIII AUV which was developed by Queensland University of Technology [1].
- (2) Funnel Docking Station: typical representatives have the FDS Funnel Docking Station which was developed by Eurodoker for REMUS [2] and a Funnel Docking Station that was developed by The Monterey Bay Aquarium Research Institute for the 54 cm diameter (21-in) AUV [3].
- (3) The torpedo tube recovery mode: the recovery method is mainly used in the military field, belonging to a kind of independent recovery mode. The first successful case of the torpedo tube recovery mode,

which was a LMRS [4] type AUV, about 5.88 m long, 51 cm diameter, was completed on “Hart Mesa” nuclear submarine in 2007.

- (4) The embedded equipped recovery mode: the mode is a deep water recovery technology as follows: build deep space station as a platform, build a boat of AUV, sail the boat back to the nonpressure hull of mother ship, or build a boat in the side of mother ship. Typical representative is the ALVIN [4], developed by the United States, which returns to a huge dry shelter cabin, such as DDS, on the back of the mother ship, during recovery. Another one is DSRV; it adopts the docking during recovery, which also belongs to such a mode of recovery.
- (5) The platform recovery mode: the recovery method uses underwater platform to achieve the task of the recovery of AUV. Typical representatives are the Marine-Bird [5], developed by Kawasaki Heavy Industries of Japan and FAU AUV [4] and developed by USA, which adopt Cable Latch Docking, both belonging to this recovery method.

The AUV control of recovery mission is working with each linked coordinate; it will decrease the regulating quality of the control system if the objects have coupling. The system even cannot operate in the condition of serious coupling. Therefore, the discussion of the decoupling problem is very meaningful work for both the control theory and engineering practice [3]. In spite of the theoretical research which has gained substantial achievements, it is not satisfactory when we compare the decoupling theory in engineering application with other branches such as the optimal control and the adaptive control in engineering application [6, 7].

Many scholars have researched fuzzy control with strong robustness a lot and used it in decoupling systems. From the perspective of reducing the inputs dimension of the decoupling controller, we distinguish in real-time the coupling relationship of the asymmetrical gas collector pressure by using dynamic coupling analysis method, and then fuzzy decoupling controller is employed to decouple and control the gas pressure of collector. Paper [8] outlined the Zadeh-MacFarlane-Jamshidi trio in their pursuit concerning the theory and application of fuzzy logic. These developments built a theoretical basis to apply single-context decision which makes a problem governed by the knowledge based on coupled fuzzy rules. The developed theorems establish an analytical equivalence to analyze the relationship between the decisions made from a coupled set of fuzzy rules and an uncoupled set of fuzzy rules concerning the same problem domain. These developments have been widely adopted in the field of supervisory control of an industrial fish cutting machine. Paper [9] designed a unique fuzzy self-tuning disturbance decoupling controller for a serial-parallel hybrid humanoid arm to complete the throwing trajectory-tracking mission. Paper [10] proposed a self-learning fuzzy decoupling controller. This method can online generate and modify the fuzzy rules through the method of self-learning. And this kind of controller was applied to the control of aircraft engine.

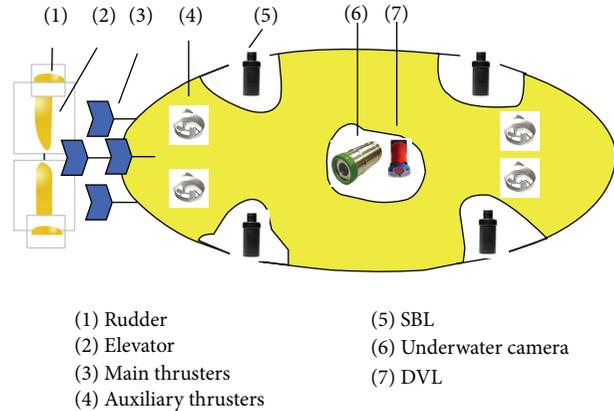


FIGURE 1: AUV structure layout.

The simulation results showed that the control effect was ideal.

2. AUV Recovery System

The overall layout of the BSAV AUV is shown in Figure 1. It is mainly composed of four main thrusters, four auxiliary thrusters, a rudder, and an elevator. The main thrusters and control planes are located in the stern of the AUV. The auxiliary thrusters, whose centerlines are designed with a certain angle to its water plane, are placed in the main section of the vehicle. These thrusters can provide sway and heave control forces at the same time. The SBL (short base line) is an underwater position system which can provide the vehicle's position information relative to the recovery platform where hydrophones are placed. The DVL (Doppler velocity log) is used to measure the AUV's velocity relative to the bottom of water. And the underwater camera is very helpful to obtain a more accurate position measurement with optical method when the vehicle is near the platform with guiding lights mounted. On the horizontal plane motion, it will produce coupling against sway motion and surge motion through the differential of the left and right main thruster to adjust the heading angle of AUV. The main problem is the coupling between the heading angle control and the sway motion control. And it is also the focus of this study. However, we will do a further research in a follow-up article on the coupling between the heading control and the surge motion control.

3. The Coordination Control Strategy in AUV Recovery

The AUV recovery control system is designed by the hierarchical structure of discrete event perception; it consists of three top-down layers, including layer for mission planning (mission layer), task transfer (task layer), and behavior realization (behavior layer). The mission layer is mainly responsible for the planning of movement coordination control strategy during the whole AUV recovery process, the state, and planning the next state. The task layer is mainly responsible for transferring the planning information

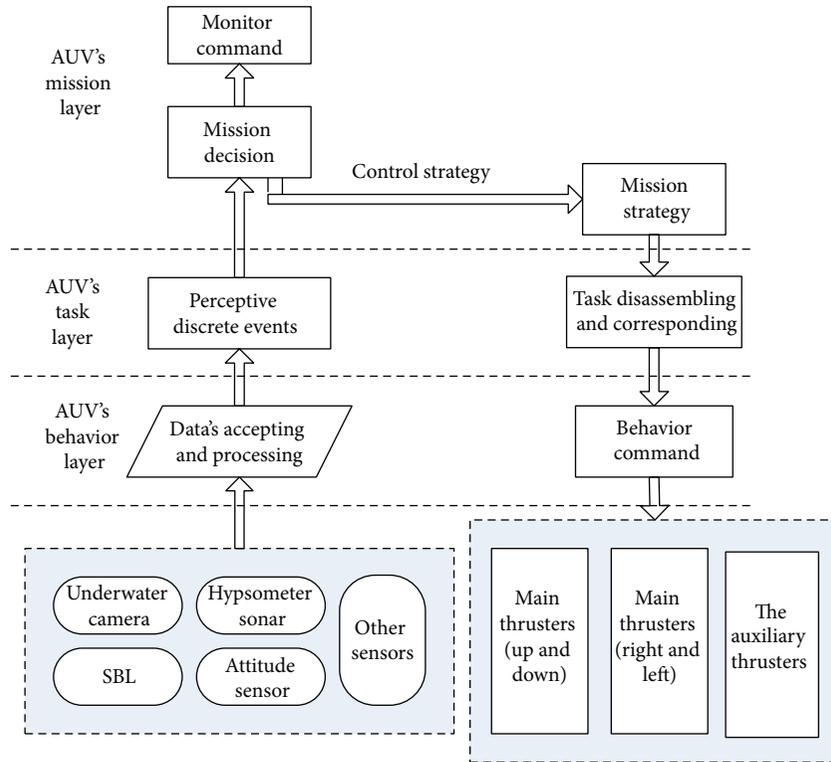


FIGURE 2: The AUV recovery control system.

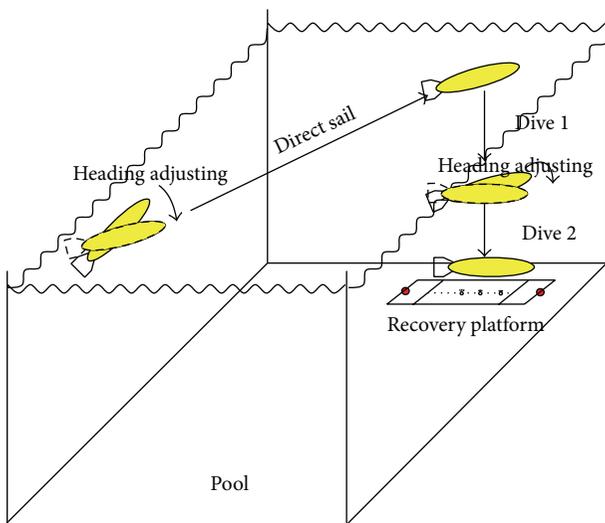


FIGURE 3: A diagram of AUV's recovery process.

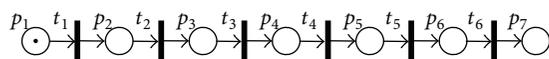


FIGURE 4: The Petri network structure of mission layer.

in various stages of recovery of AUV to the specific controller implementation. The behavior layer is the lowest layer system

structure; it is mainly responsible for that data acquisition, fusion, and controller action execution, as shown in Figure 2.

During the AUV recovery of the whole process, after starting recovery command, AUV in unmanned condition approached the recovery platform, finally seated to the recovery platform, and completed the underwater recovery mission. We can see that in Figure 3. For such a complex process control containing a lot of conflict and concurrent events, it is wise to introduce Petri net theory into the design of the mission layer and task layer of the control system.

Firstly, we design the mission layer. The entire recovery process of AUV consists of heading adjusting, then direct sailing, dive, heading adjusting, dive, and several discrete events. The transition between the various discrete events is continuous and dynamic. The Petri network structure is shown in Figure 4. Then Table 1 shows the transition meanings of the Petri net structure library of mission layer, in which we have the following:

region A: the horizontal circle is at the top of coordinate origin, and the circle's radius is 0.5 m;

region B: the horizontal circle is above the platform 3 m and the circle's radius is 0.1 m;

region C: the region is above the platform 3 m ($Z = 2$ m in depth).

The task layer of AUV control system is a transition layer between the mission layer and the behavior layer. It includes the processing of data information and the coordinating of control strategy.

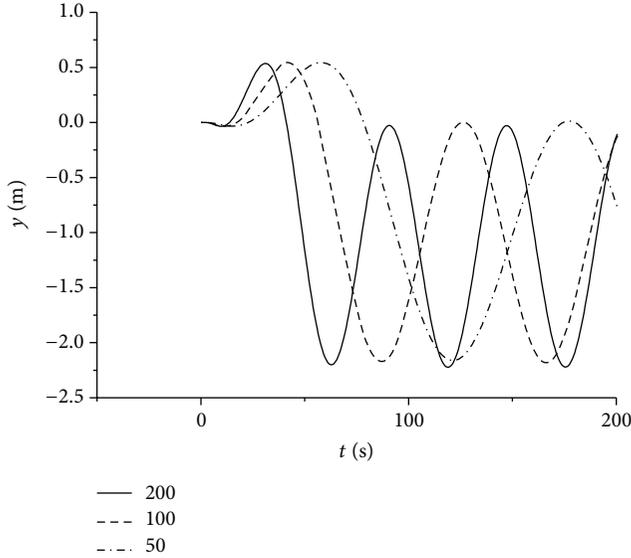


FIGURE 5: The effect of heading control to sway motion.

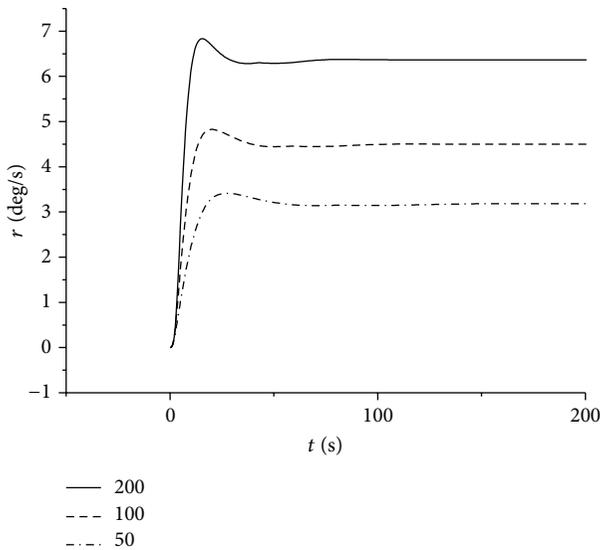


FIGURE 6: The change of heading angle velocity.

Coupling variables of the control of recovery of AUV mainly exist in the horizontal movement, and the movement, horizontal motion control is the key point of this research. Starting from the design of single degree of freedom controller, and the decoupling control of coupling variables, we can design the horizontal motion coordination controller, and finally we can get the recovery control system of AUV.

4. Coupling Analysis and Decoupling Controller Design

4.1. Coupling Analysis. Due to the complication and intelligent development of the modern control system, along

TABLE 1: The transition meanings of the Petri net structure library of mission layer.

The meaning of Library	The meaning of transition
P_1 standby state	t_1 starting system
P_2 the ready for heading adjusting	t_2 heading adjusting to the target value
P_3 the ready for direct sailing	t_3 sailing to A region
P_4 sailing at A region and ready for dive	t_4 dive to B region
P_5 sailing at C region with position deviations	t_5 heading adjusting
P_6 sailing at C region and ready for docking	t_6 seated to the recovery platform
P_7 the success of recovery of AUV	

with the increasing number of variables, the system becomes more difficult to control because of the coupling between variables [11, 12]. There are multiple couplings among the recovery motion. The main problem is the coupling between the heading angle control and the sway motion control. And it is also the focus of this study. Now through a case to illustrate the coupling between them and the reasons, then provide the basics for decoupling controller design.

Without the controller, adjusting the AUV heading can affect the sway motion. Through the differential of the left and right main thruster to adjust the heading, that is, equal the thruster thrust and opposite in direction to adjust the heading. Figures 5 and 6 show the effect on sway motion and longitudinal motion as adjusting heading while the left and right thruster thrust are 200 N, 100 N, and 50 N (the left thruster is positive and the right thruster is negative).

First, we can analyze AUV structure and hydrodynamic characteristics. When AUV began to adjust heading, the torque was almost equal in size and relatively balanced, so it had little effect from the surge and sway motion. We can see from Figure 5 that the curve is relatively flat in the first 15 seconds. But as the rotation continues, it has a relative motion between AUV bow and stern around the barycenter and the surrounding water. When the speed reaches a certain level, the resistance that rudder suffered significantly increased. However, streamlined bow has a relatively small resistance so that the moment balance is broken. AUV is clockwise (overlooking) when the heading angle increases, but AUV deviates to the right relative of the original barycenter because of the reaction of the stern strong resistance. With the heading angle increasing, the value of sway motion begins to decrease when it exceeds 90 degrees. So the barycenter presents an elliptical track along with the unceasing changing of the heading angle. The track is shown in Figures 7, 8, and 9. It also has effects from surge motion, but it is the main thruster that the surge motion controller has enough power and energy to control the effects of this interference. However, the auxiliary thruster has a problem that the thrust is insufficient compared to the main thruster. So it is necessary to add decoupling to suppress its effects.

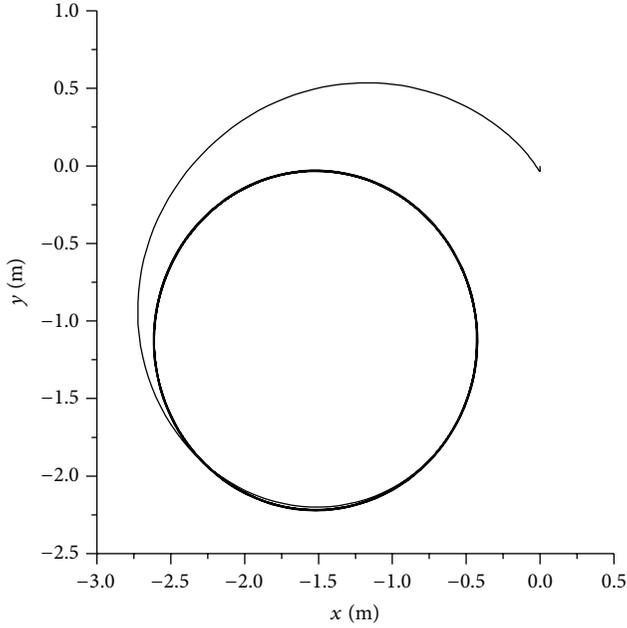


FIGURE 7: The track of the horizontal plane while adjusting heading with 200 N.

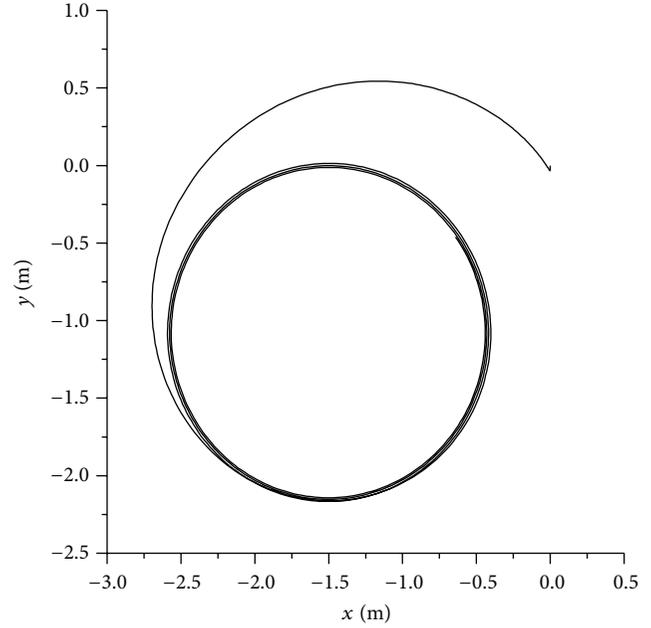


FIGURE 9: The track of the horizontal plane while adjusting heading with 50 N.

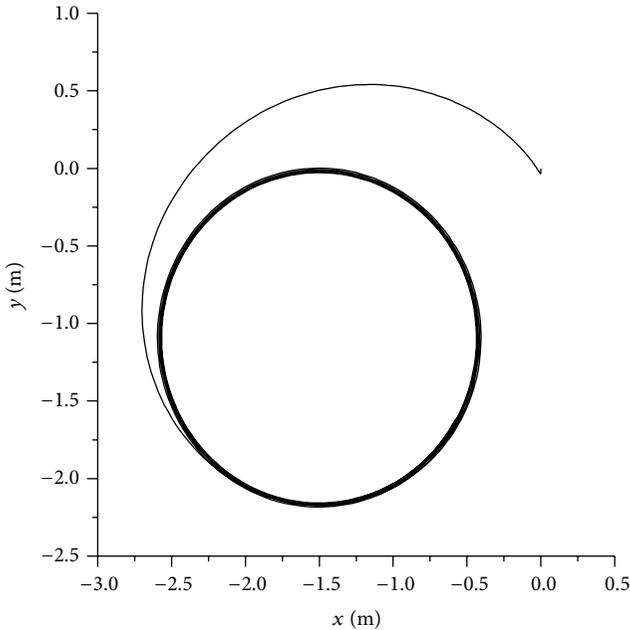


FIGURE 8: The track of the horizontal plane while adjusting heading with 100 N.

4.2. Design of Decoupling Control System. In control theory for multivariable systems, fuzzy decoupling is a kind of theory that combines decoupling theory with fuzzy control theory, which has a low demand to the mathematical model of system. Analysis of the coupling between variables can achieve decoupling control, weaken the influence between each other from a certain extent, coordinate output control quantity of each variable in order to improve the control

efficiency of control system, and also have a great significance to solve the problem of multivariable coupling.

In coupling analysis, the simulation results in coupling of sway motion control and heading control in the AUV recovery mission which shows that when the speed reaches a certain degree and the torque balance is broken, the greater the heading control force, the greater the impact from motion. So this paper designed a decoupling compensator and we can see the structure of fuzzy decoupling controller shown in Figure 10. The u_1 and u_2 are output control quantity of the controller, and they can be as two inputs of the fuzzy compensator after appropriate adjustment through adjusting parameters k_1 and k_2 . For fuzzy compensator, a two-input and one-output fuzzy controller that is in common use, compute an output u_3 by two inputs that have been adjusted by adjusting parameter, and then add u_3 that have been adjusted by k_3 and k_4 to the original output of the controller to constitute control quantity to send to the actuator. Among that, a_1 and a_2 can be +1 or -1, which is to determine whether compensation quantity on the variable is positive or negative compensation.

The design of fuzzy compensator is critical in decoupling control system design. Firstly, we should fuzz the accurate quantity when we design a fuzzy compensator. The process of fuzz is to map the accurate input to the domain of language value. For two-input and one-output fuzzy compensator which is more commonly used, we concern, under normal circumstances, the deviation E and the deviation change rate EC of controlled quantity as input language variable and the control quantity to the object as output language variable of fuzzy compensator.

The membership function is shown in Figures 11, 12, and 13. In fuzzy compensation rule table shown in Table 2, Y_{prop}

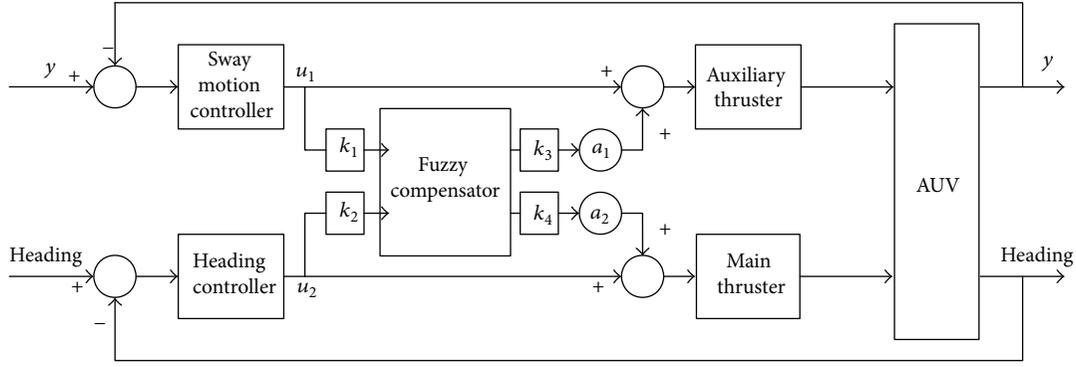


FIGURE 10: The structure of decoupling controller.

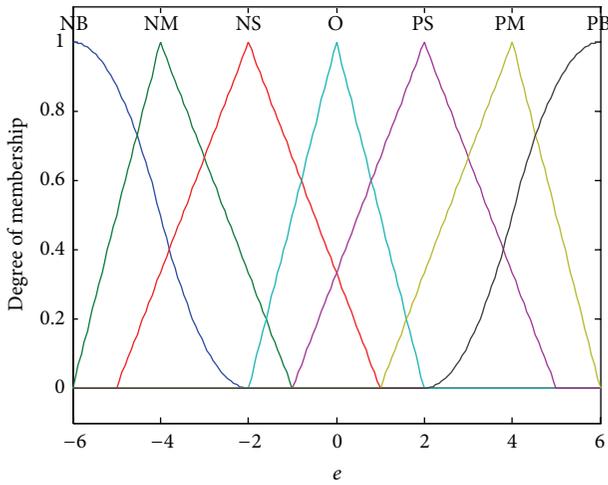


FIGURE 11: Membership function of deviation.

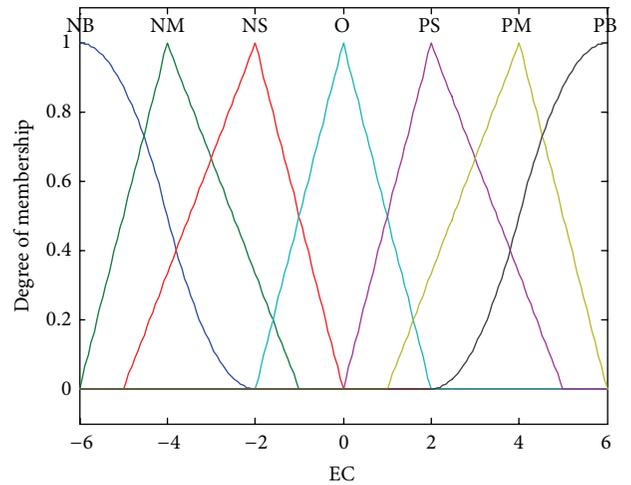


FIGURE 12: Membership function of deviation change rate.

TABLE 2: Compensation rules.

U	N_{prop}						
	NB	NM	NS	O	PS	PM	PB
Y_{prop}							
NB	NS	NS	NS	O	PS	PS	PS
NM	NM	NM	NS	O	PS	PM	PM
NS	NB	NM	NS	O	PS	PM	PB
O	O	O	O	O	O	O	O
PS	PB	PM	PS	O	NS	NM	NB
PM	PM	PM	PS	O	NS	NM	NM
PB	PS	PS	PS	O	NS	NS	NS

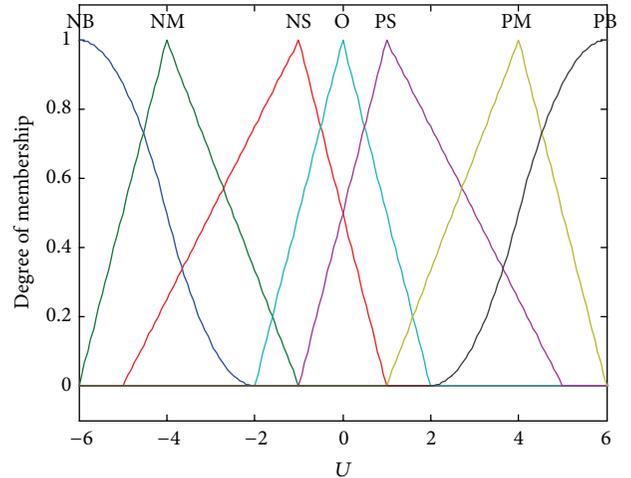


FIGURE 13: Membership function of the output.

and N_{prop} are two input language variables, namely, sway thrust and transgenic bow thrust moment and U is the output language variable.

According to the rules of fuzzy compensation, by the input and output membership functions of fuzzy compensator, through compose operation and clear operation of fuzzy reasoning, obtained decoupling compensation quantity and made fuzzy decoupling compensation table, as shown

in Table 3. In practical engineering, we can get the corresponding compensation value through interpolation to query fuzzy decoupling compensation table after the fuzzification operation of the output of heading controller and sway motion controller. The compensation value multiplied by a

TABLE 3: The nondimensional hydrodynamic coefficients of the AUV.

U	N_{prop}												
	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
Y_{prop}													
-6	-2.13	-2.20	-2.13	-2.25	-2.13	-1.09	0.00	1.09	2.13	2.25	2.13	2.20	2.13
-5	-2.79	-2.79	-2.79	-2.85	-2.20	-1.09	0.00	1.09	2.20	2.85	2.79	2.79	2.79
-4	-3.16	-3.11	-3.16	-3.00	-2.13	-1.09	0.00	1.09	2.13	3.00	3.16	3.11	3.16
-3	-3.97	-3.97	-3.90	-3.00	-2.25	-1.07	0.00	1.07	2.25	3.00	3.90	3.97	3.97
-2	-4.85	-4.41	-3.91	-3.00	-2.13	-1.09	0.00	1.09	2.13	3.00	3.91	4.41	4.85
-1	-2.85	-2.94	-2.64	-2.41	-1.65	-1.09	0.00	1.09	1.65	2.41	2.64	2.94	2.85
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	2.85	2.94	2.64	2.41	1.65	1.09	0.00	-1.09	-1.65	-2.41	-2.64	-2.94	-2.85
2	4.85	4.41	3.91	3.00	2.13	1.09	0.00	-1.09	-2.13	-3.00	-3.91	-4.41	-4.85
3	3.97	3.97	3.90	3.00	2.25	1.07	0.00	-1.07	-2.25	-3.00	-3.90	-3.97	-3.97
4	3.16	3.11	3.16	3.00	2.13	1.09	0.00	-1.09	-2.13	-3.00	-3.16	-3.11	-3.16
5	2.79	2.79	2.79	2.85	2.20	1.09	0.00	-1.09	-2.20	-2.85	-2.79	-2.79	-2.79
6	2.13	2.20	2.13	2.25	2.13	1.09	0.00	-1.09	-2.13	-2.25	-2.13	-2.20	-2.13

certain factor, and then adds to the output of sway motion and heading controller, so can form the final control quantity of the actuator.

5. Simulation

Simulation Case 1. Two simulations were carried out here to demonstrate the effect of the proposed controller. In the first one, a heading command was issued from 0 degrees to 60 degrees using the left and right thrusters, where the vehicle was initially static at the surface. For comparison, another simulation was conducted under the same condition using the decoupling compensator.

The hydrodynamic coefficient of model and some adjusting parameters of fuzzy decoupling compensator are set as follows: $k_1 = 3, k_2 = 1, k_3 = 0.15, k_4 = 0.1, a_1 = 1,$ and $a_2 = 1$. The quantitative factor of sway thrust Y_{prop} and heading control moment N_{prop} is set as follows: $K_Y = 0.02, K_N = 0.05,$ and the scale factor $K = 1$. Figure 14 shows the deviation of effects on sway motion before and after decoupling compensation. Two inputs and one output of compensator were shown in Figures 15 and 16. Figure 17 shows the compensation force that sways motion control obtained while adjusting heading.

From the simulation result, we can see that heading control has a relatively large effect on sway motion before decoupling. It is dangerous to high-accuracy motion like AUV recovery. After decoupling, although there is a certain deviation on the opposite direction, the overshoot is reduced by half overall.

6. Water Tank Experiment

The preliminary verification of AUV recovery experiment based on the coordinated control strategy and fuzzy decoupling compensator design was carried out. The details are as follows.

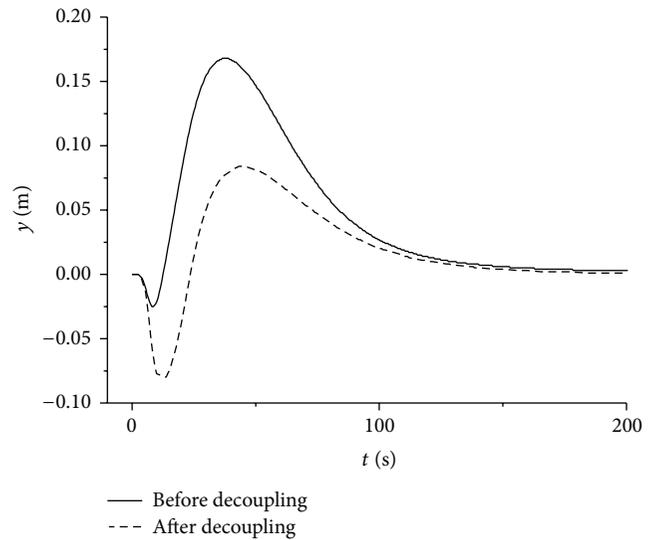


FIGURE 14: The comparison of transverse movement before and after decoupling compensation.

6.1. Introduction of Water Tank Experiment's Guiding Device. Guiding device for the water tank experiment is mainly composed of monocular imaging system, linear light source array, and underwater recovery guide positioning system of AUV, which is based on the short baseline.

6.1.1. Monocular Imaging System. Monocular imaging system is composed of cameras, capture cards, and processing host. The underwater experiment employs black and white camera Tornado with low illumination. Image acquisition card uses the DaHeng DH-CG320 image capture card that supports PCI04-Plus bus mode, and the card can be used to collect high-quality colour/black and white video signal in real time and transfer it to the memory for real-time storage via PCI04-Plus bus. Visual processing host is AUV visual

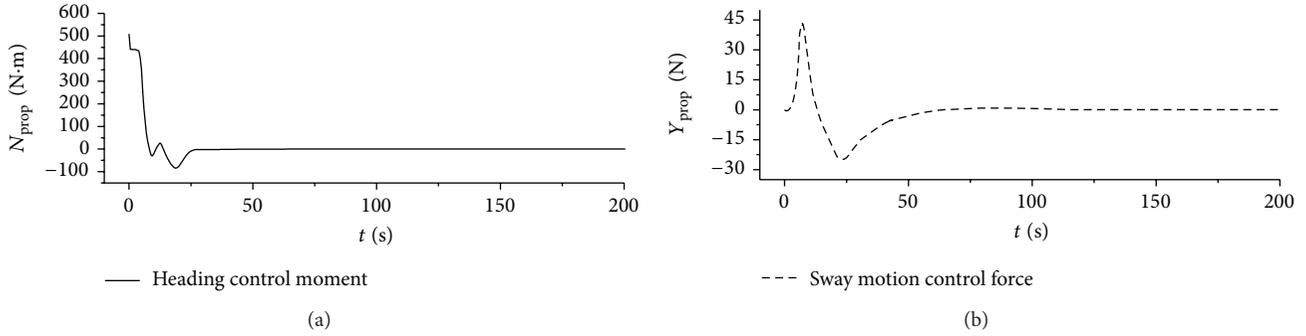


FIGURE 15: Two inputs of the compensator.

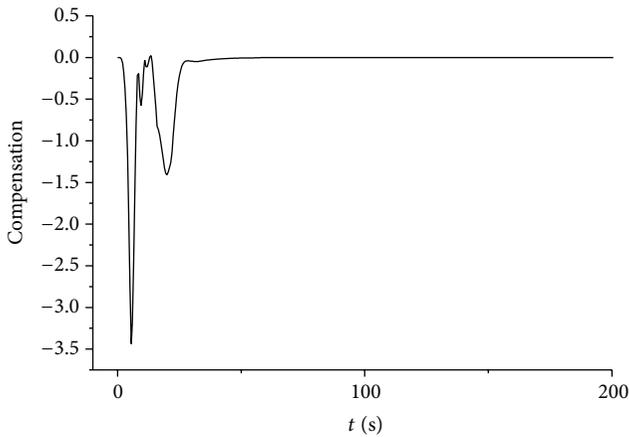


FIGURE 16: Output of the compensator.

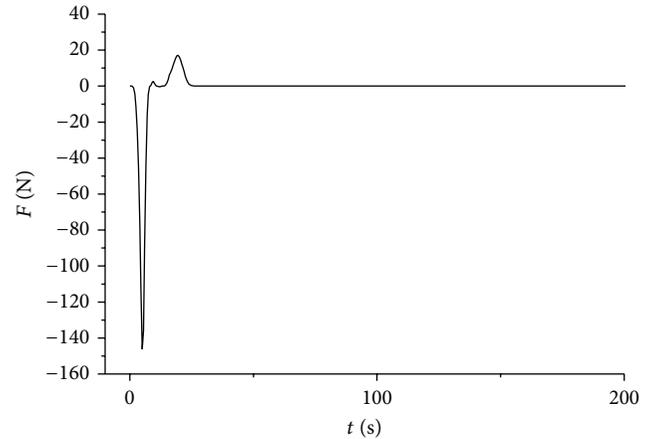


FIGURE 17: The compensation force of sway control.

brain responsible for image acquisition, image processing, positioning solver, and so forth, adopting PCI04 + bus based on embedded systems.

6.1.2. Linear Light Source Array. A linear light source array for monocular vision on the wire position is designed in this underwater experiment. Where the reference light source is a heart-shaped light source that is made by transforming a rectangular surface light source, which uses the U.S. AI company's SL6404 (Figure 18). Other eight light sources all use blue astigmatism LED point light source assembled add waterproof cover. All light sources are arranged on the longitudinal center line at the bottom of the recovery dock tank. Heart-shaped vertex points to the heading of the dock tank, the distance between the adjacent sources is 350 mm, and the entire lighting system is 2800 mm long.

6.2. The Composition and Installation of the AUV Underwater Recovery Guide Positioning System Based on the Short Baseline. The short baseline positioning sonar system adopted by the system basically has the following several parts: short baseline sonar array (four transducers, Figure 19(a)), a beacon host, and two beacons (transducers F1, F2).

Transponder beacon consists of beacon host and F1, F2. F1 is hairstyle and F2 is emission type, radiating device faces

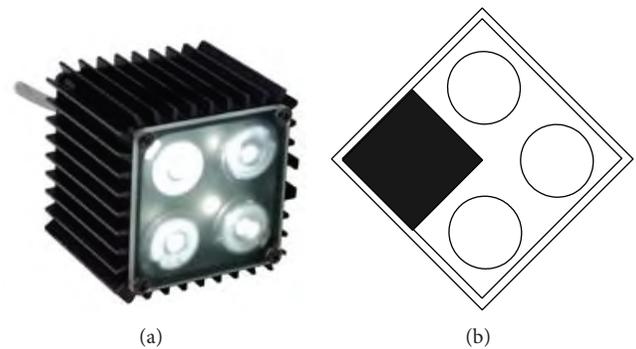


FIGURE 18: The reference light SL6404.

to the surface of the water when used. The beacon host is installed as shown in Figure 19(b), and transducers F1 are installed as shown in Figure 19(c).

6.3. The Analysis of Water Tank Experimental Data. The trajectory of the recovery process AUV obtained from each projection plane, such as the x - y , x - z , and three-dimensional trajectories, is shown as follows in Figure 20.

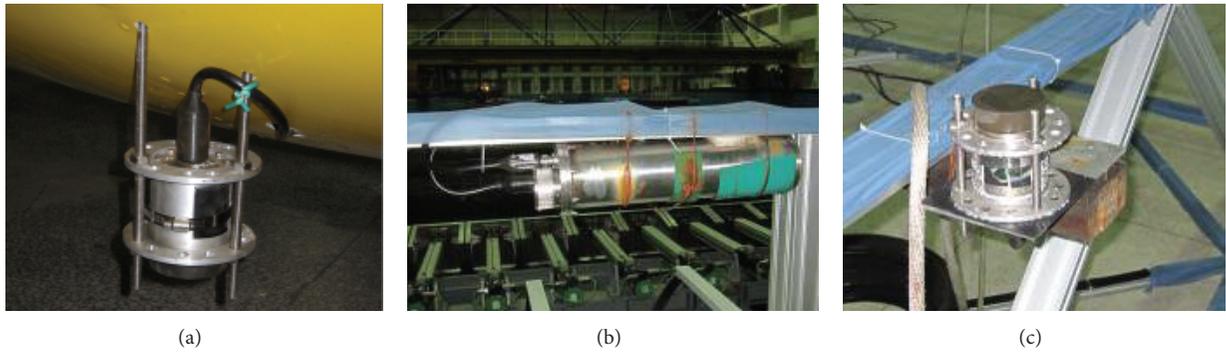


FIGURE 19: Short baseline physical installation diagram.

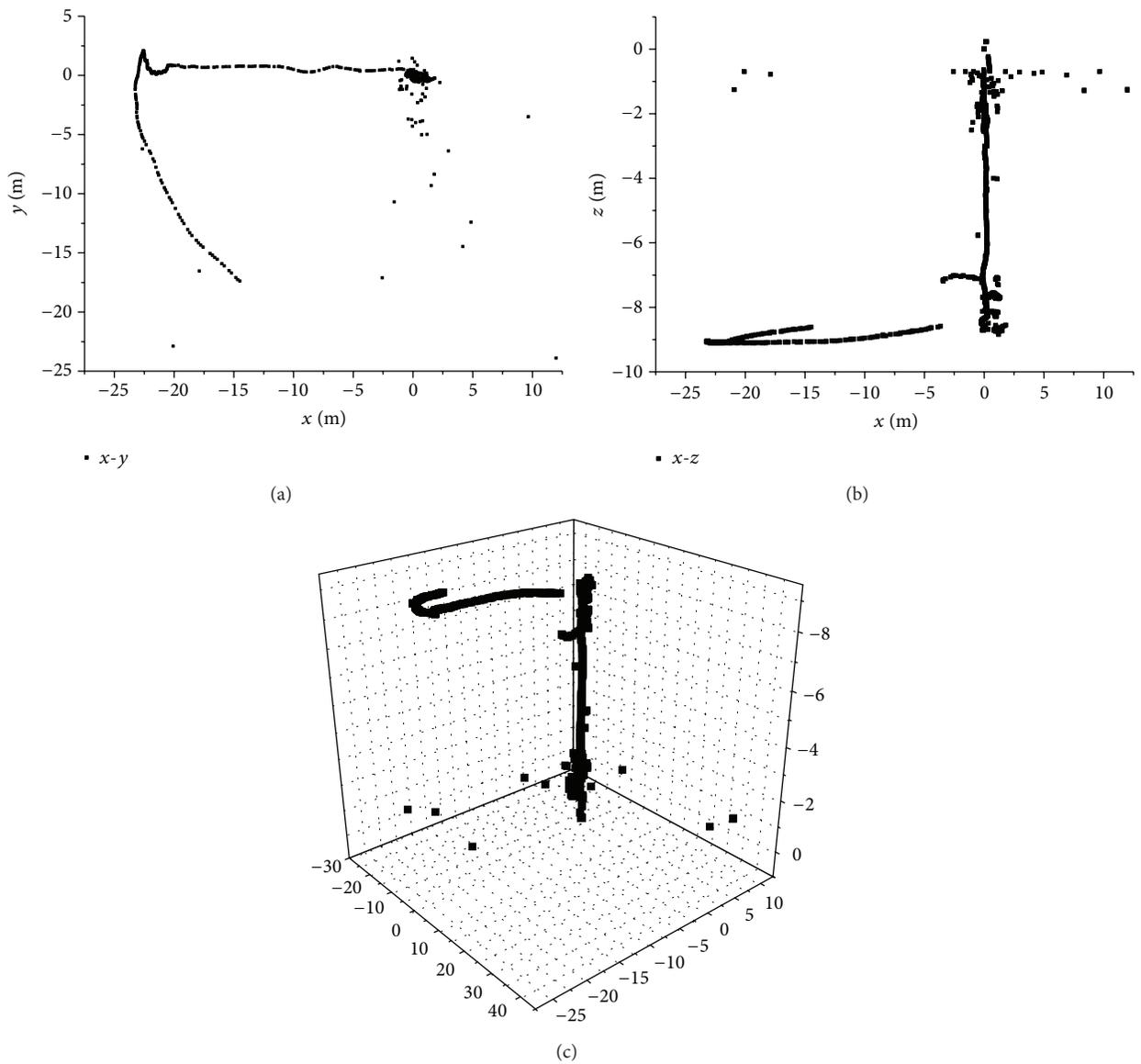


FIGURE 20: The track graph of water tank experiment.

As to the analysis of experimental data obtained by water tank, since this type of AUV has better suppression and self-recovery performance to roll, shielding the x -axis torque output of the controller did not take the initiative to suppress the roll. Since the data is based on body-fixed frame, the starting point of y is small only means that heading direction is almost aligned with the recovery platform at this moment. Also in the height of the curve, jump section shows that depth sonar's surveying sound waves hit the recovery platform frame.

Since the data is based on body-fixed frame, this is rather distinct with the actual trajectory in the water tank, which only is used to verify the effectiveness of the controller here. In addition, according to the entire recovery process which is divided into several stages, it needs to analyze actions according to the characteristics of the various stages. In the vertical plane, although the height value of space has more jump error, due to the fact that the speed is faster, there are only a few points that have the jump error and do not use the height value in the process. Until the coordinate stabilizes, it has no jump error during the whole process. Therefore, it did not affect the control process.

Through the water tank experiments and analysis of experimental data, the effectiveness and feasibility of AUV recovery movement coordinated control based on fuzzy decoupling method are validated successfully. However, further trials need to be carried out in order to achieve the parameter optimization.

7. Conclusion

In this paper the coupling relationship between sway movement and heading control in AUV recovery process was studied. Based on BSAV- AUV platform, this paper proposed an underwater recovery strategy and analyzed the coupling variable and coupling reason in recovery motion control through simulation. It has a high demand to sway and surge motion control in AUV recovery process. But the actuator (auxiliary thruster) of sway motion controller has a problem which is insufficient thrust compared with the actuator (main thruster) of surge motion controller. Therefore, this paper mainly focuses on the influence that the heading control acts on sway motion. And according to the coupling mechanism formulate fuzzy decoupling rules and design fuzzy decoupling compensator. The results of simulation and water tank experiments have shown that the sway movement was eliminated to a permitted range with the help of a fuzzy decoupling controller, which was considered to be effective in the AUV recovery motion control. However, other aspects of tuning parameters need to be further improved and optimized and still need further experimental verification.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research work is supported by the Youth Project of National Natural Science Foundation of China (51309067/E091002), the National Natural Science Foundation of China (51109043/E091002), and the Fundamental Research Funds for the Central Universities (HEUCFX41402).

References

- [1] H. Singh, J. G. Bellingham, F. Hover et al., "Docking for an autonomous ocean sampling network," *IEEE Journal of Oceanic Engineering*, vol. 26, no. 4, pp. 498–514, 2001.
- [2] R. Stokey, B. Allen, T. Austin et al., "Enabling technologies for REMUS docking: an integral component of an autonomous ocean-sampling network," *IEEE Journal of Oceanic Engineering*, vol. 26, no. 4, pp. 487–497, 2001.
- [3] R. S. McEwen, B. W. Hobson, L. McBride, and J. G. Bellingham, "Docking control system for a 54-cm-diameter (21-in) AUV," *IEEE Journal of Oceanic Engineering*, vol. 33, no. 4, pp. 550–562, 2008.
- [4] P. Guang, "Current situation and development trend of AUV recovery y technology," *Torpedo Technology*, vol. 16, no. 6, pp. 10–14, 2008.
- [5] T. Hardy and G. Barlow, "Unmanned underwater vehicle (UUV) deployment and retrieval considerations for submarines," in *Proceedings of the 9th International Naval Engineering Conference (INEC '08)*, Hamburg, Germany, April 2008.
- [6] B.-H. Jun, J.-Y. Park, F.-Y. Lee et al., "Development of the AUV "ISiMP" and a free running test in an ocean engineering basin," *IEEE Journal of Oceanic Engineering*, vol. 36, no. 1, pp. 2–14, 2009.
- [7] P. Sotiropoulos, D. Grosset, G. Giannopoulos, and F. Casadei, "AUV docking system for existing underwater control panel," in *Proceedings of the OCEANS*, pp. 1–5, Bremen, Germany, May 2009.
- [8] C. W. de Silva, "Zadeh-MacFarlane-Jamshidi theorems on decoupling of a fuzzy rule base," *Scientia Iranica D*, vol. 18, no. 3, pp. 611–616, 2011.
- [9] Y. Wang, R. Shi, and H. Wang, "Dynamic modeling and fuzzy self-tuning disturbance decoupling control for a 3-DOF serial-parallel hybrid humanoid arm," *Advances in Mechanical Engineering*, vol. 2013, Article ID 286074, 14 pages, 2013.
- [10] X.-Y. Ren and S.-Q. Fan, "Aero-engine multivariable learning fuzzy decouple control," *Journal of Propulsion Technology*, vol. 25, no. 6, pp. 535–538, 2004.
- [11] X. Bian, J. Zhou, and Z. Yan, "Adaptive neural network control system of path following for AUVs," in *Proceedings of the OCEANS*, 2011.
- [12] H. Li, "Design of multivariable fuzzy-neural network decoupling controller," *Control and Decision*, vol. 21, no. 5, pp. 593–596, 2006.

Research Article

Multiobjective Order Assignment Optimization in a Global Multiple-Factory Environment

Rong-Chang Chen and Pei-Hsuan Hung

*Department of Distribution Management, National Taichung University of Science and Technology,
No. 129, Section 3, Sanmin Road, Taichung 404, Taiwan*

Correspondence should be addressed to Rong-Chang Chen; rcchens@nutc.edu.tw

Received 31 January 2014; Revised 15 April 2014; Accepted 18 April 2014; Published 1 June 2014

Academic Editor: Hamid Reza Karimi

Copyright © 2014 R.-C. Chen and P.-H. Hung. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In response to radically increasing competition, many manufacturers who produce time-sensitive products have expanded their production plants to worldwide sites. Given this environment, how to aggregate customer orders from around the globe and assign them quickly to the most appropriate plants is currently a crucial issue. This study proposes an effective method to solve the order assignment problem of companies with multiple plants distributed worldwide. A multiobjective genetic algorithm (MOGA) is used to find solutions. To validate the effectiveness of the proposed approach, this study employs some real data, provided by a famous garment company in Taiwan, as a base to perform some experiments. In addition, the influences of orders with a wide range of quantities demanded are discussed. The results show that feasible solutions can be obtained effectively and efficiently. Moreover, if managers aim at lower total costs, they can divide a big customer order into more small manufacturing ones.

1. Introduction

To best satisfy the requirements of customers and quickly respond to changes in the market environment, some manufacturers who produce time-sensitive products, such as fashion garments or high-tech goods, have established many manufacturing sites in other countries. Global manufacturing has become a competitive strategy for many manufacturers due to the cheaper labor and raw material costs overseas. The reduction of these costs, however, is generally accompanied by longer production times. How to assign customer orders rapidly to the most suitable sites while lowering costs and shortening production times has thus become an urgent issue for global companies.

The issue mentioned above is concerned with order assignment [1, 2] of companies with factories spreading over the globe. Companies have to consider a number of influencing factors [3]. The first factor is related to customers' requirements, which are generally diverse in quantity. Some customers have significant changes in the order quantities which range from tens of sample products to hundreds of thousands [4]. If companies can respond to a wide range of

quantities demanded, then their responsiveness [5] can be enhanced. Next, the production capacity of each plant should be considered. Due to the variation in production capacity at different plants, the processing time of an order is dependent on the plant to which it is assigned. The production capability of each site should also be considered. Production capability at each site significantly influences the fulfillment of orders. For instance, in some African countries, product quantities are limited by lower skill levels, whereas plants in some Asian countries have high-level skills to produce a variety of products. Some products requiring higher skill levels have to be produced at specific sites [4]. Therefore, using a good scheme to select proper sites is very important for global companies. The last issue that a company should consider is its objectives. What a company pursues today has become multiple objectives instead of single ones. For manufacturers who produce time-sensitive products such as fashion apparel or high-tech products that typically have a short life cycle, the shortest total production time undoubtedly is the most important objective. The main objectives for a company are to pursue the lowest cost and the shortest total production time, in order to achieve the global optimization of their

supply chain. Multiple objectives make the order assignment problem more difficult to deal with.

Multiobjective order assignment for companies with multiple plants worldwide is very complex. It belongs to the combinatorial optimization problem, which is known to be NP-hard [6]. One effective scheme is required. A company expects to obtain practical plans soon to respond to fast changing business environment. Therefore, a heuristic algorithm such as genetic algorithm (GA) [7–11] which can provide feasible solutions in a short time is more suitable than an exact algorithm that can get optimal solutions but that also requires more computation time.

There have been a number of studies related to this issue [12–19]. However, some [13–16] focus little concern on globalized industry. On the other hand, some studies focus on single objective problems [12, 17, 18]. As mentioned before, in a now extremely competitive environment, the capability to achieve multiple objectives is a necessity for a company. To address the described issue, Chen et al. [19] considered a two-objective optimization in a global multiple-factory environment, but they did not take production time into consideration. This paper proposes an effective mechanism and uses genetic algorithm to find a solution. To prove the validity of the proposed method, real data from a famous company in Taiwan are used as the basis for some experiments. Plant-related factors such as production capacity, manufacturing costs, material costs, and transportation costs are considered. Subsequently, MOGA (multiobjective genetic algorithm) is employed to assign orders to optimally satisfy the objectives of a company, that is, the lowest total cost and the shortest production time. Some possible solutions are generated by using the genetic algorithm. One clear advantage of the genetic algorithm is that, by its very nature, we are able to produce a number of feasible solutions, thus facilitating discussion on the merits of various decisions and supporting multiobjective decision making [20]. Therefore, the decision maker can choose the solution that best satisfies customers' requirements and also achieves the company's objectives.

The rest of this paper is organized as follows. Section 2 describes the problem. Section 3 contains the model and the method of solution. In Section 4, results and discussion are presented. Section 5 presents concluding remarks and suggestions for further studies.

2. The Problem

2.1. Problem Description. The order assignment issue can be briefly described as below. A company has a logistics center and many sites for production, as illustrated in Figure 1. Here, the number of production sites is represented as m . The logistics center aggregates customer orders demanded from anywhere in the world and divide them into several manufacturing orders (MO), the total number of which is expressed as n . Next, the company intends to assign n MOs to m production sites and once production is completed, the plant will deliver it to the place designated by customers. Main factors that a company in pursuit of "the lowest total cost" and "the shortest production time" should consider

are customer demands, production capacity, production skill level, due date, manufacturing cost, material cost, and delivery cost. At the same time, companies would like to respond to a wide range of order quantities demanded, in order to gain the largest advantage under a competitive environment.

2.2. Modeling

2.2.1. Assumptions. To simplify the problem, some assumptions are made.

- (1) A customer order can be divided into several MOs. Each MO produces only one product and can be fulfilled only at a manufacturing site.
- (2) Each product is composed of some kinds of raw materials. The related information on raw material is known.
- (3) Each kind of raw material is provided by only one supplier.
- (4) The production capacity of a manufacturing site is stable and constant. That is, the production capacity is known and fixed, but different manufacturing sites may have different capacities.
- (5) The production type is MTO (make to order). Thus, there is no inventory. The quantity demanded for each product is known; there is no stock and no production in advance.
- (6) Due dates of orders are known and fixed.
- (7) The material cost, manufacturing cost, delivery cost, and order delay cost are known. The mode of transportation is by sea or by land, so the transportation cost can be estimated. The inbound tax on raw material and inbound transportation cost are included in the material cost. Similarly, the manufacturing cost includes the labor cost and the manufacturing overhead. The delivery cost includes the outbound transportation cost and tariff.
- (8) The transportation modes from a raw material supplier to a manufacturing site and from a manufacturing site to the destination are known. Therefore, the transportation cost is known if the origin and the destination are given.
- (9) Lack of material is negligible.

2.2.2. Formulation. For easy description, some symbols are defined as shown in abbreviation section.

The first objective that global companies take into account is the total cost, which can be expressed as follows:

Total cost = material cost (includes direct material cost, inbound tax, and inbound transportation cost) + manufacturing cost (includes labor cost and manufacturing overhead)

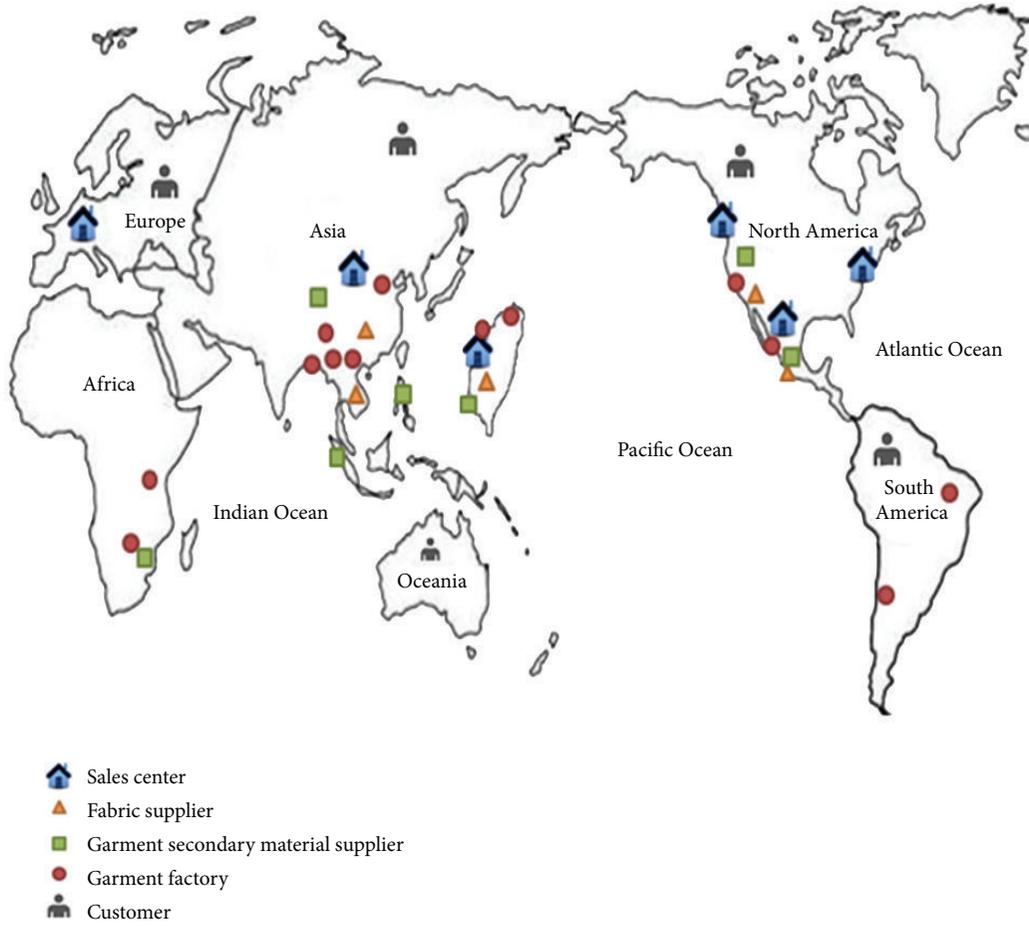


FIGURE 1: A schematic diagram of the deployment of a global garment company [4].

+ delivery cost (includes outbound transportation cost and tariff) + order delay cost:

$$\min Z_1 = \sum_i \sum_j c_j^{RM} Q_{ij}^{RM} + \sum_i \sum_k x_{ik} c_k^{MANU} \frac{Q_i^{FG}}{S_k} + \sum_i \sum_k \sum_h x_{ik} z_{kh} c_{kh}^{DELIV} Q_{ih}^{FG} + \sum_i c_i^{DELAY} D_i. \quad (1)$$

Note that in (1), the final term is the cost caused by order delay. This cost arises in some industries such as the garment sector. In practice, some companies will be asked to pay a penalty if an order is delayed. For example, the penalty rate is about 3%–5% of the transaction amount for some garment companies. This term is related to due dates. Some might argue: why not using due dates (such as on-time delivery rate) as an objective? One advantage of using the delay cost over using the due date as an objective is that the former provides a more accurate calculation on delayed days. As an example shown in Table 1 and Figure 2, suppose that the number of MOs is 10 and the due date for each MO is 5 days. Two different solutions (*E* and *F*) are obtained. The on-time delivery rates for solutions *E* and *F* are 0.9 and 0.8, respectively, indicating that solution *E* is better than solution

TABLE 1: The comparison between due dates and delayed days based objectives.

Solution	On-time delivery rate	Total delayed days
<i>E</i> (upper chart in Figure 2) based on due dates	90%	5
<i>F</i> (lower chart in Figure 2) based on delayed days	80%	3

F. However, the total delayed days for solution *E* is 5 and for solution *F* is 3, conversely showing that solution *F* is better than solution *E*. For manufacturers who produce time-sensitive products such as fashion garments or high-tech goods, an objective based on the number of delayed days might be more suitable than one based on due dates.

Another important objective for global companies to consider is total production time. If an order can be finished within a shorter time, then the lead time can be reduced and the company has more time to deliver the products. For

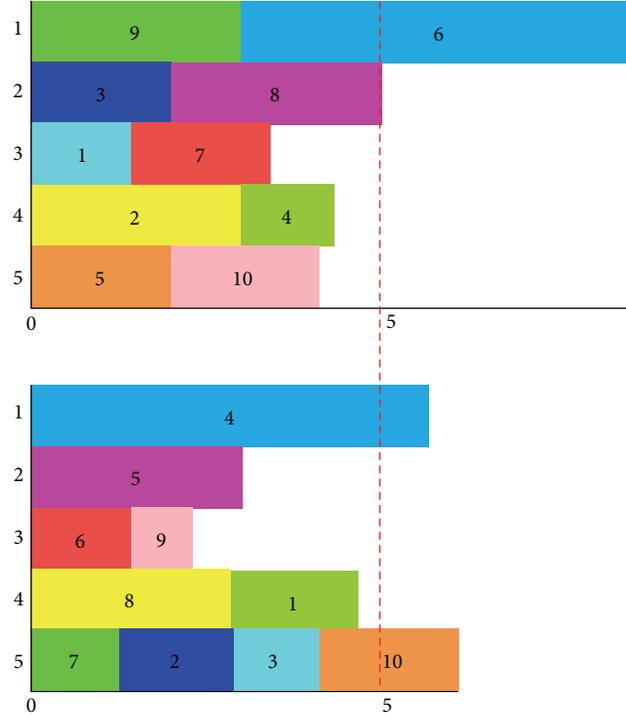


FIGURE 2: Illustration of two different solutions for comparison. The dashed line indicates the due date.

time-sensitive products, the achievement of this objective is a necessity:

$$\min Z_2 = \sum_i \sum_k x_{ik} \frac{Q_i^{\text{FG}}}{S_k}. \quad (2)$$

The constraints to be considered are

$$\sum_k x_{ik} = 1, \quad \forall i \quad (3)$$

$$\sum_i x_{ik} S_i \leq S_k, \quad \forall k \quad (4)$$

$$x_{ik} = \begin{cases} 1 & \text{if MO } i \text{ is assigned to site } k, \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

Equation (3) requires each MO to be assigned to only one manufacturing site, while (4) requires that the necessary capacity to finish products must not be greater than the capacity of each manufacturing site. x_{ik} is the decision variable. As indicated in (5), x_{ik} is either 0 or 1.

3. The GA Structure

This paper employs MOGA as a tool to find solutions. Detailed description of the GA structure is introduced in the following.

(1) *Encoding*. Each chromosome represents a solution, which is comprised by some genes. To implement GA, chromosomes must be encoded first. A simple and efficient encoding

MO	1	2	3	4	5	6	...	n
Site	2	5	3	4	1	5	...	1

FIGURE 3: Representation of a chromosome.

method is used in this paper. The sequence of genes from left to right indicates the index number of MO. The value of a gene (alleles) stands for the index number of a manufacturing site. For example, MO “1” is assigned to site “2” and MO “2” is assigned to site “5,” as illustrated in Figure 3.

(2) *Initial Solutions and Calculation of Fitness Values*. The initial population is generated at random. In general, the larger the population is, the wider the search range will be. A larger population is likely to obtain better fitness values. However, the computation time for a larger population is increased.

The evaluation of a solution is performed by calculation of the fitness value of a chromosome. In this paper, the calculation is based on the total cost and the total production time. Better fitness values are retained and the best ones are chosen as the Pareto solutions, which are nondominated by other chromosomes. An efficient scheme proposed by Fonseca and Fleming is called FFGA [21], as indicated in Figure 4. By this scheme, the evolutionary time can be reduced. Let G_i represent a chromosome at generation t which is dominated

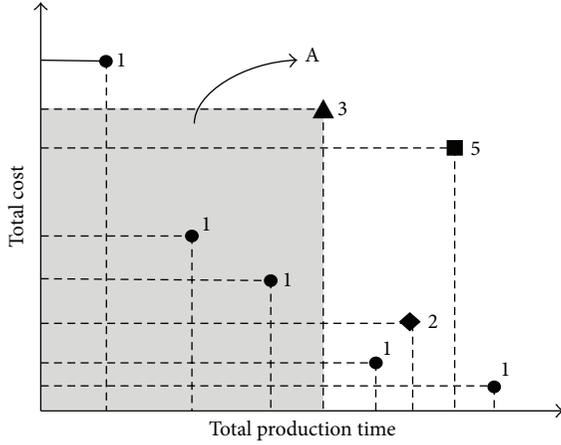


FIGURE 4: Illustration of the FFGA algorithm.

by $p_i^{(t)}$ individuals in the current generation. Then the rank of the individual at generation t can be expressed as [21]

$$\text{rank}(G_i, t) = 1 + p_i^{(t)}. \quad (6)$$

FFGA sorts the fitness values of chromosomes in the population. Each chromosome is ranked by the number of chromosomes which dominate it according to the fitness values. The number plus one is set to be the rank of a chromosome. The more a chromosome is dominated by others, the larger value its rank is. In contrast, if a chromosome does not have others which dominate it, according to (6), it is assigned rank 1. The one which is not dominated by anyone is designated as rank “1” and is collected into the Pareto set; the one dominated by two chromosomes is designated as rank “3” (see chromosome A in Figure 4). This makes the ranking process of chromosomes more efficient and shortens the evolutionary time.

(3) *Selection and Reproduction.* The modified Roulette wheel selection method is applied to select chromosomes. First, all the chromosomes’ fitness values are calculated and then sorted. The larger the fitness value is, the lower the priority is. For example, if there are 20 chromosomes and their fitness values are 1 to 20, respectively, then the chromosome with a fitness value of 1 will have the probability of $20/(1 + 2 + 3 + \dots + 20)$, the chromosome with a fitness value of 2 will have the probability of $19/(1 + 2 + 3 + \dots + 20)$, and the like. In other words, superior parent chromosomes have more opportunities to be selected into next generation.

(4) *Crossover.* This study employs partially matched crossover to avoid repeating. As Figure 5 shows, two points of crossover are chosen at random, which will divide parent generations into three parts, respectively. The child chromosome will grab the first and third parts of parent chromosome 1 and the second part of parent chromosome 2. At the same time, the program checks whether the values of genes from parent chromosomes are repeated or not. If they are repeated, skip

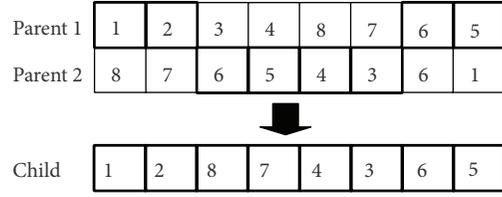


FIGURE 5: Illustration of partially matched crossover.

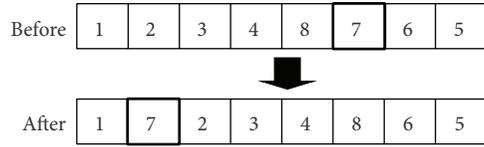


FIGURE 6: Illustration of insertion mutation.

the repeated gene value and fill in a new value. This step will not end until all crossovers have been done.

(5) *Mutation.* This study adopts the insertion mutation method to mutate. A point is randomly chosen as an insertion point and a gene is chosen to insert. As Figure 6 shows, the sixth gene is chosen and the second position is an insertion point. The rest of the genes from and after the second gene are moved to the right. The insertion mutation method can avoid the repetition of a number, which represents an unfeasible solution.

(6) *Creating Next Generation.* To create the next generation, the tournament selection method is used. A gene pool is set up to store chromosomes. Two chromosomes are then chosen at random and compared with each other. If the fitness value of a chromosome is superior to that of a counterpart chromosome, then the chromosome is stored into the pool and used to generate the child chromosome.

(7) *Termination.* A number of ways can be used to terminate the program. The termination condition in this study is the achievement of a preassigned number of evolution generations set by the user. If the generation number is reached, the program will stop running and the solutions will be output. To obtain better solutions, multiple trials are suggested if the initial population is generated randomly.

4. Results and Discussion

This paper employs MOGA to find solutions. In addition, the Brutal-force method (BFM, or exhaustion method), which can obtain the optimal solution, is used to evaluate the performance of the MOGA program. The comparison of these two methods is shown in Section 4.1. Influences of genetic parameters are discussed in Section 4.2. Next, the influences of plant amount and order amount on the order assignment are introduced. Section 4.3 discusses the results concerning orders with wide ranges of quantities demanded. Finally, the reassignment of orders is addressed.

The program was run in a PC with an operating system of Windows XP Professional SP2. The CPU is Intel Core 2 at 2.6 GHz and the RAM is 1 GB.

4.1. Data Input and Validation of MOGA. To evaluate the effectiveness of the proposed approach, real data from a famous garment company in Taiwan are used as a base to perform some experiments. The base case is set as follows: five garment plants and eight end products. The plants are in Taiwan (A and B plants), China, Vietnam, and the USA. Products are classified into four types: shirt, skirt, pants, and overcoat. Each one has two styles. Therefore, there are eight types of finished products, which are assigned numbers from one to eight, respectively. Input data of the numerical examples for the base case are shown in Tables 2, 3 and 4.

The data concerning MOs include its number, quantity demanded, due date, daily cost for delay, the arrival destination, and the index number of finished products. The plant-related data include production capacities and manufacturing costs. Each plant has a monthly capacity and a daily capacity.

For easy description, we designate the number of generation as N_g , the population size as Pop , the crossover rate as R_c , the mutation rate as R_m , and the coefficient of variation as C_v . Each case was run 30 times to evaluate the GA program.

BFM is an exact algorithm that can obtain optimal solutions. If the result from MOGA is near or equal to that from BFM, then the correctness of MOGA developed in this study can be supported. On the other hand, computation time is one of the performance indices that can evaluate the efficiency of the approach. The shorter the computation time is, the faster a decision can be made.

As for genetic parameters, the generation numbers N_g are changed from 50 to 500; the population sizes Pop are set to be 200 and 500. The crossover rate R_c is set to be 0.5 and the mutation rate R_m is 0.05 at the base case. The results are summarized in Table 5. From experimental results we may find that MOGA can get good workable solutions. The accuracy for each case is over 96%, even to 100%. The accuracy is defined as (based on total cost):

$$\text{Accuracy} = \left\{ 1 - \left| \frac{\text{MOGA solution} - \text{BFM solution}}{\text{BFM solution}} \right| \right\} \times 100\%. \quad (7)$$

The comparison of the computation time between MOGA and BFM is shown in Figure 7. As the number of MO increases from 8 to 13, the computation times of MOGA and BFM differ increasingly. When the number of MOs is over 11, the difference rapidly increases. Though BFM can get optimal results, the required computation time exceeds 10 hours, as the MO number is over 13. When the number of MOs becomes larger, the computation time may be over a hundred days. In practice, managers hope to get a reasonable result in a short time. BFM cannot satisfy their needs. On the contrary, by using MOGA as an analytical tool, they can get good results within seconds. It is very efficient, making the order assignment mechanism more practical.

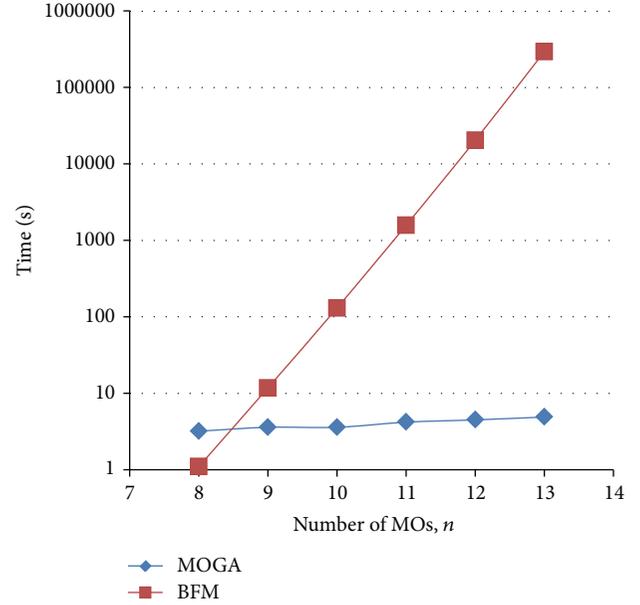


FIGURE 7: Comparison of the computation times between MOGA and BFM.

4.2. Influences of Genetic Parameters. There are four important genetic parameters when using MOGA: generation number, population size, crossover rate, and mutation rate. At the base case, there are ten MOs, eight types of products, and five plants. The crossover rate and the mutation rate are set to be 0.9 and 0.01, respectively. The number of generation is set to be 50, 100, 200, 300, and 500. The number of population is set to be 200 and 500. The MOGA program randomly creates an initial population. The results, therefore, are different by random nature. To investigate the performance of the MOGA program, each case was run with 30 trials.

When the generation number and population size equal $500 * 200$ and $500 * 500$, it is easier to obtain good results by MOGA, as shown in Table 6. Generally, the coefficient of variation falls between 0.01 and 0.04, indicating that the MOGA program is stable. However, from Figure 8 we may see that the computation time for the latter is about triple that of the former. Thus, this study uses $500 * 200$ to perform the following tests.

To understand the influences of the crossover rate and mutation rate, the generation number is fixed at 500 and the population size is fixed at 200. Crossover rates are then changed from 0.5 to 0.9 with an increase of 0.1 each time; mutation rates are changed from 0.01 to 0.05. When the mutation rate is 0.05, the results are quite stable; when the crossover rate is 0.5 and the mutation rate is 0.05, coefficients of variation C_v are between 0.006 and 0.028. As compared to the result from BFM, the accuracy is 98.07% to 99.98%, showing that these parameter values can give quite good results.

4.3. Influences of Changes in Plant Number and Order Quantity. As Figure 9 illustrates, the total cost increases when the total production time decreases. However, the average cost

TABLE 2: Input data related to manufacturing orders for the base case.

MO	Demand (piece)	Due date (day)	Delayed cost (NTD/day)	Destination	Product number
1	13,500	1	10,000	2	7
2	28,800	5	20,000	1	5
3	30,000	2	50,000	4	5
4	23,100	2	38,500	2	6
5	13,500	3	22,500	2	2
6	20,100	2	33,500	5	6
7	16,800	4	28,000	3	1
8	7,500	1	12,500	3	3
9	10,500	5	17,500	3	2
10	19,800	3	33,000	2	1

TABLE 3: Input data related to manufacturing sites for the base case.

Site	Capacity (piece/day)	Capacity (piece/month)
1	7,000	210,000
2	7,000	210,000
3	10,000	300,000
4	16,000	480,000
5	10,000	300,000

TABLE 4: Sites eliminated because of unsuitable production skill levels.

MO	Eliminated sites
1	3, 4, 5
2	1, 2, 3
3	5
4-10	None

per plant decreases as the total production time decreases. When the plant number is changed, there is no great change in the coefficient of variation, which falls between 0.0018 and 0.0146. This indicates that changes in plant number do not affect the stability of the GA program.

If the plant number increases, what influences will have on order assignment is worth discussing. To address this, some experiments are performed. The experiments performed here consider the cost of a new plant. Figure 9 shows the variation of the optimal total cost and total production time with the number of plants. With an increase in plant number, though the extra cost for a new plant is added, the total production time and total cost decrease. The computation time increases linearly with the number of plants.

After a company receives orders from all over the world, their logistics center will gather together all the customer orders and then assign them. In practice, the range of order quantity changes a lot and the amount of different customers' demands also varies. Wal-Mart, as an example, may have up to ten thousand, even hundreds of thousands of products demanded. However, only several sample products are needed for factory owners. When the quantity of an order demanded is over hundred thousand, to enhance the

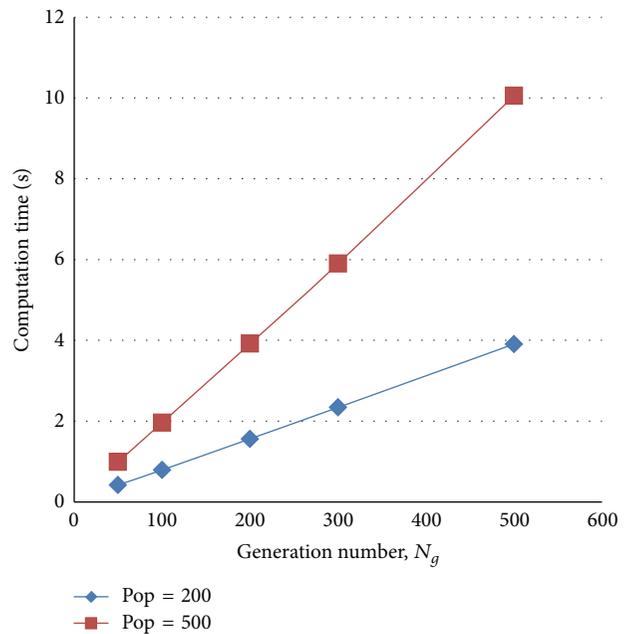


FIGURE 8: Variations of computation times with different generation numbers and population numbers.

capacity utilization rate or to balance the production loads between plants, an additional manual operation for order assignment is considered. To utilize the production capacity properly, a company may divide a big order into several smaller ones. This process is called division. If there is no process of division, it will cause some plants to be left unused and some other plants to be insufficient in production capacity. Consequently, companies should consider the capacity utilization (the percentage of the capacity that was used to the available capacity) and on-time delivery rate (the percentage of the number of on-time delivery orders to the total number of orders) as well when aiming at the shortest total production time and minimal total cost.

To investigate the influences of such a wide range of quantities demanded on the order assignment, the quantities of orders are set between 80 and 200,000 and divided on average into 10, 15, 20, 25, and 30 smaller orders. The

TABLE 5: Comparison of MOGA results with those by BFM.

$N_g * \text{Pop}$	Total Production time (day)	Total cost by MOGA	C_v	Total cost by BFM	Accuracy (%)	Average computation time (sec)
50 * 200	27	33,768,890	0.016	33,410,860	98.93	0.418
	28	33,046,345	0.020	32,566,915	98.53	
	29	33,061,395	0.013	32,519,485	98.33	
	30	33,264,065	0.007	32,408,030	97.36	
100 * 200	27	33,668,865	0.022	33,410,860	99.23	0.790
	28	33,064,945	0.014	32,566,915	98.47	
	29	32,979,000	0.013	32,519,485	98.59	
	30	33,176,050	0.008	32,408,030	97.63	
200 * 200	27	33,890,290	0.018	33,410,860	98.57	1.561
	28	33,393,335	0.018	32,566,915	97.46	
	29	33,075,020	0.013	32,519,485	98.29	
	30	33,075,060	0.009	32,408,030	97.94	
300 * 200	27	33,668,865	0.018	33,410,860	99.23	2.338
	28	32,834,345	0.017	32,566,915	99.18	
	29	32,760,385	0.018	32,519,485	99.26	
	30	33,143,460	0.016	32,408,030	97.73	
500 * 200	27	33,654,765	0.014	33,410,860	99.27	3.907
	28	32,566,915	0.023	32,566,915	100.00	
	29	33,012,990	0.011	32,519,485	98.48	
	30	33,564,380	0.005	32,408,030	96.43	
50 * 500	27	33,659,265	0.019	33,410,860	99.26	0.991
	28	33,085,945	0.018	32,566,915	98.41	
	29	32,579,885	0.016	32,519,485	99.81	
	30	32,945,955	0.016	32,408,030	98.34	
100 * 500	27	33,900,290	0.017	33,410,860	98.54	1.961
	28	33,096,345	0.022	32,566,915	98.37	
	29	32,559,485	0.015	32,519,485	99.88	
	30	32,923,960	0.009	32,408,030	98.41	
200 * 500	27	33,653,865	0.024	33,410,860	99.27	3.916
	28	33,045,945	0.018	32,566,915	98.53	
	29	32,539,485	0.027	32,519,485	99.94	
	30	33,030,060	0.018	32,408,030	98.08	
300 * 500	27	33,648,865	0.018	33,410,860	99.29	5.898
	28	32,983,840	0.017	32,566,915	98.72	
	29	32,997,990	0.014	32,519,485	98.53	
	30	33,114,460	0.011	32,408,030	97.82	
500 * 500	27	33,644,265	0.015	33,410,860	99.30	10.052
	28	33,085,945	0.020	32,566,915	98.41	
	29	33,179,515	0.014	32,519,485	97.97	
	30	32,822,455	0.016	32,408,030	98.72	

influences of the division number (C) on the production time, the production cost, the capacity utilization, and the on-time delivery rate are examined. Results from the experiments show that although the division process will lead to an increase in the total production time, it helps enhance the capacity utilization. In addition, no matter whether an order

is divided or not, the on-time delivery rates are ranging from 88% to 96%.

For the division numbers $C = 25$ and $C = 30$, the results are crossed, as shown in Figure 10. On the right side of the crossed point, the higher the division number is, the smaller the total cost will be. If managers aim at decreasing the total

TABLE 6: Variations of solutions with different crossover rates and different mutation rates.

$R_c * R_m$	Production time (day)	Total cost by MOGA	C_v	Total cost by BFM	Accuracy (%)	Average computation time (sec)
0.9 * 0.01	27	33,644,265	0.021	33,410,860	99.30	3.901
	28	32,833,315	0.023	32,566,915	99.18	
	29	32,800,385	0.014	32,519,485	99.14	
	30	33,087,465	0.014	32,408,030	97.90	
0.9 * 0.02	27	33,425,860	0.022	33,410,860	99.96	3.899
	28	32,973,440	0.019	32,566,915	98.75	
	29	32,539,485	0.017	32,519,485	99.94	
	30	32,885,040	0.009	32,408,030	98.53	
0.9 * 0.03	27	33,774,290	0.012	33,410,860	98.91	3.903
	28	32,828,945	0.024	32,566,915	99.20	
	29	32,997,990	0.016	32,519,485	98.53	
	30	33,161,450	0.010	32,408,030	97.68	
0.9 * 0.04	27	33,768,890	0.018	33,410,860	98.93	3.899
	28	33,056,345	0.020	32,566,915	98.50	
	29	32,979,000	0.016	32,519,485	98.59	
	30	33,114,460	0.011	32,408,030	97.82	
0.9 * 0.05	27	33,677,260	0.019	33,410,860	99.20	3.899
	28	32,983,840	0.017	32,566,915	98.72	
	29	32,539,485	0.014	32,519,485	99.94	
	30	33,170,355	0.008	32,408,030	97.65	
0.8 * 0.01	27	33,668,865	0.020	33,410,860	99.23	3.463
	28	33,778,930	0.017	32,566,915	96.28	
	29	32,965,995	0.015	32,519,485	98.63	
	30	32,896,955	0.014	32,408,030	98.49	
0.8 * 0.02	27	33,768,890	0.019	33,410,860	98.93	3.461
	28	33,060,345	0.018	32,566,915	98.48	
	29	33,061,395	0.014	32,519,485	98.33	
	30	33,087,465	0.012	32,408,030	97.90	
0.8 * 0.03	27	33,659,265	0.025	33,410,860	99.26	3.459
	28	33,219,365	0.011	32,566,915	98.00	
	29	33,075,020	0.011	32,519,485	98.29	
	30	32,669,940	0.014	32,408,030	99.19	
0.8 * 0.04	27	33,677,260	0.026	33,410,860	99.20	3.460
	28	32,983,840	0.024	32,566,915	98.72	
	29	32,941,010	0.018	32,519,485	98.70	
	30	33,392,075	0.007	32,408,030	96.96	
0.8 * 0.05	27	33,416,260	0.028	33,410,860	99.98	3.460
	28	32,838,715	0.014	32,566,915	99.17	
	29	33,176,000	0.013	32,519,485	97.98	
	30	33,384,470	0.009	32,408,030	96.99	
0.7 * 0.01	27	33,768,890	0.020	33,410,860	98.93	3.036
	28	33,056,345	0.019	32,566,915	98.50	
	29	32,898,985	0.013	32,519,485	98.83	
	30	33,173,460	0.009	32,408,030	97.64	
0.7 * 0.02	27	33,659,265	0.018	33,410,860	99.26	3.031
	28	33,045,945	0.020	32,566,915	98.53	
	29	32,941,010	0.014	32,519,485	98.70	
	30	33,161,550	0.009	32,408,030	97.67	

TABLE 6: Continued.

$R_c * R_m$	Production time (day)	Total cost by MOGA	C_v	Total cost by BFM	Accuracy (%)	Average computation time (sec)
0.7 * 0.03	27	33,644,265	0.016	33,410,860	99.30	3.031
	28	33,203,965	0.020	32,566,915	98.04	
	29	33,022,990	0.015	32,519,485	98.45	
	30	33,087,065	0.012	32,408,030	97.90	
0.7 * 0.04	27	33,683,865	0.021	33,410,860	99.18	3.029
	28	33,045,945	0.015	32,566,915	98.53	
	29	33,055,020	0.014	32,519,485	98.35	
	30	33,333,480	0.007	32,408,030	97.14	
0.7 * 0.05	27	33,644,265	0.014	33,410,860	99.30	3.030
	28	33,203,965	0.017	32,566,915	98.04	
	29	32,779,985	0.016	32,519,485	99.20	
	30	33,170,355	0.009	32,408,030	97.65	
0.6 * 0.01	27	33,659,265	0.013	33,410,860	99.26	2.605
	28	32,838,715	0.012	32,566,915	99.17	
	29	32,975,995	0.012	32,519,485	98.60	
	30	33,466,585	0.007	32,408,030	96.73	
0.6 * 0.02	27	33,904,890	0.021	33,410,860	98.52	2.604
	28	33,085,945	0.018	32,566,915	98.41	
	29	32,559,485	0.019	32,519,485	99.88	
	30	32,858,030	0.014	32,408,030	98.61	
0.6 * 0.03	27	33,649,365	0.018	33,410,860	99.29	2.603
	28	32,833,315	0.020	32,566,915	99.18	
	29	33,075,020	0.015	32,519,485	98.29	
	30	33,156,050	0.009	32,408,030	97.69	
0.6 * 0.04	27	33,668,865	0.011	33,410,860	99.23	2.605
	28	33,096,345	0.016	32,566,915	98.37	
	29	32,779,985	0.015	32,519,485	99.20	
	30	33,395,860	0.006	32,408,030	96.95	
0.6 * 0.05	27	33,649,365	0.021	33,410,860	99.29	2.605
	28	33,126,935	0.019	32,566,915	98.28	
	29	33,135,420	0.017	32,519,485	98.11	
	30	32,878,960	0.012	32,408,030	98.55	
0.5 * 0.01	27	33,410,860	0.021	33,410,860	100.00	2.175
	28	32,973,440	0.021	32,566,915	98.75	
	29	32,974,905	0.016	32,519,485	98.60	
	30	32,854,560	0.013	32,408,030	98.62	
0.5 * 0.02	27	33,644,265	0.011	33,410,860	99.30	2.178
	28	33,085,945	0.021	32,566,915	98.41	
	29	32,519,485	0.016	32,519,485	100.00	
	30	33,191,450	0.012	32,408,030	97.58	
0.5 * 0.03	27	33,697,660	0.024	33,410,860	99.14	2.178
	28	32,833,315	0.018	32,566,915	99.18	
	29	33,171,910	0.015	32,519,485	97.99	
	30	32,838,960	0.012	32,408,030	98.67	
0.5 * 0.04	27	33,644,265	0.020	33,410,860	99.30	2.179
	28	33,096,345	0.018	32,566,915	98.37	
	29	33,075,020	0.014	32,519,485	98.29	
	30	33,065,060	0.010	32,408,030	97.97	

TABLE 6: Continued.

$R_c * R_m$	Production time (day)	Total cost by MOGA	C_v	Total cost by BFM	Accuracy (%)	Average computation time (sec)
0.5 * 0.05	27	33,682,660	0.018	33,410,860	99.19	2.18
	28	32,572,315	0.022	32,566,915	99.98	
	29	32,760,385	0.013	32,519,485	99.26	
	30	33,033,580	0.008	32,408,030	98.07	

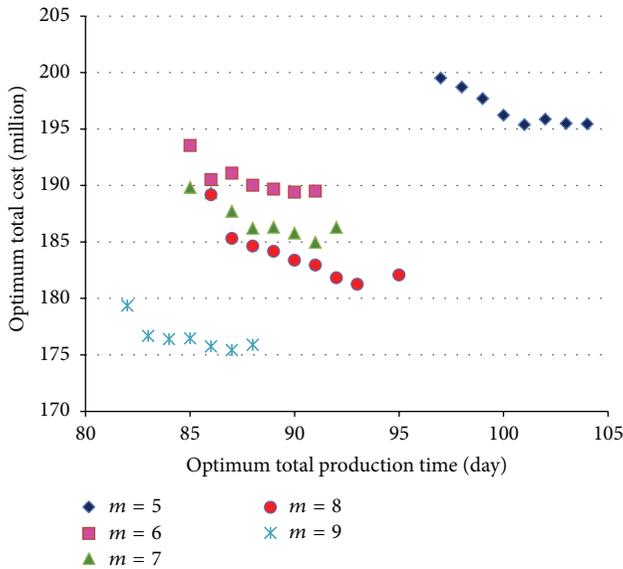


FIGURE 9: Variations of the optimal total costs and total production times with the number of plants.

cost, this way can be chosen. On the contrary, if they hope to get the minimum production time, the way in the left side can be chosen: the smaller the division number is, the shorter the production time is. If there is no division process, although it can enable the total cost to be lower, it prevents the capacity to be fully used and thus results in a low utilization rate. The division process increases the cost but enhances capacity utilization.

Some good and interesting questions might come up as a result of order division process. The first one is “should the same products produced at different manufacturing sites be gathered again before delivery?” In practice, not a few global companies prefer to deliver products directly from the manufacturing sites to the destinations. Since the contexts discussed in this paper occur in the environment of business-to-business (B2B) commerce, the destinations are usually fixed in a period of time. Direct shipment from different sites can not only reduce the transportation costs, but also shorten the delivery time. Another question is concerned with the additional shipping cost caused by the order division process. Since the division process occurs only on orders with huge amounts, the influence is very likely to be small. Even if the shipping cost is increased, minimizing the total cost can benefit the company as a whole.

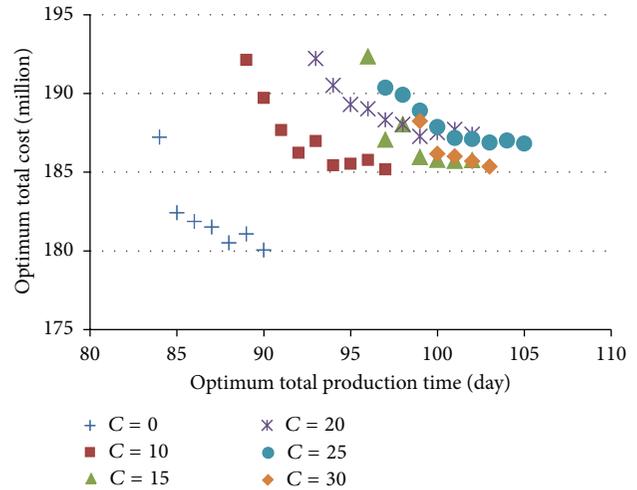


FIGURE 10: Changes in the minimum total cost and shortest total production time of each division number ($C = 0$, nondivision; $C = 10$, division number is 10, ..., $C = 30$, division number is 30).

4.4. Assignment and Reassignment. As time proceeds, new orders will be released. How to deal with new incoming orders is an important issue for managers. In this paper, we reassign the new incoming orders on a rolling time basis [12], as illustrated in Figure 11. At time t_1 , new orders are released. At this time, some previous orders are completed, while orders are in process. The unfinished orders will be kept in process and the rest will be rescheduled (reassigned).

For easy reassignment, a visualized system is developed. The results of assignment and reassignment are shown in Figures 12(a) and 12(b).

The visualized system developed by this study is easy to use. It is very convenient for users to change, cancel, or move the orders. If the managers are not satisfied with the assignment result, they can draw the mouse cursor onto the selected orders and move them to the desired locations. Managers can also cancel an order easily, as shown in Figure 13.

5. Conclusions

This paper employs multiobjective genetic algorithm (MOGA) as an analytical tool to investigate the order assignment issue for global companies with multiple plants around the world. MOGA is used to deal with order

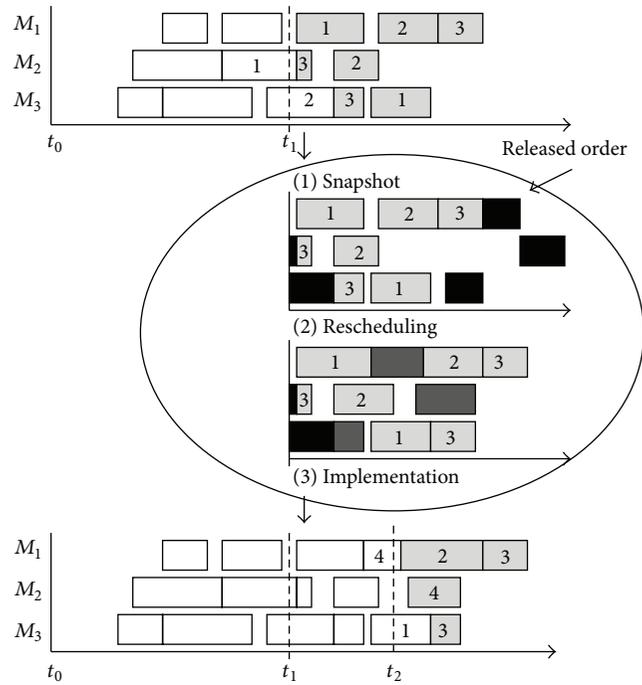
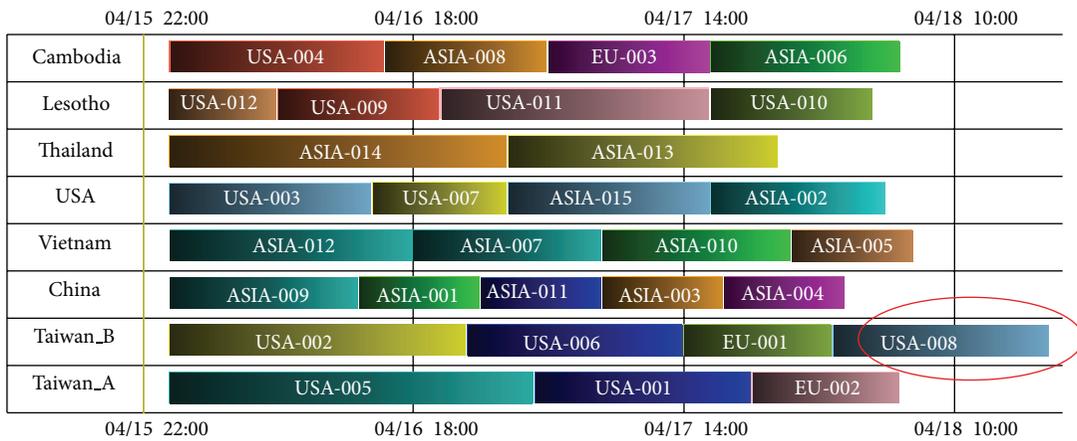
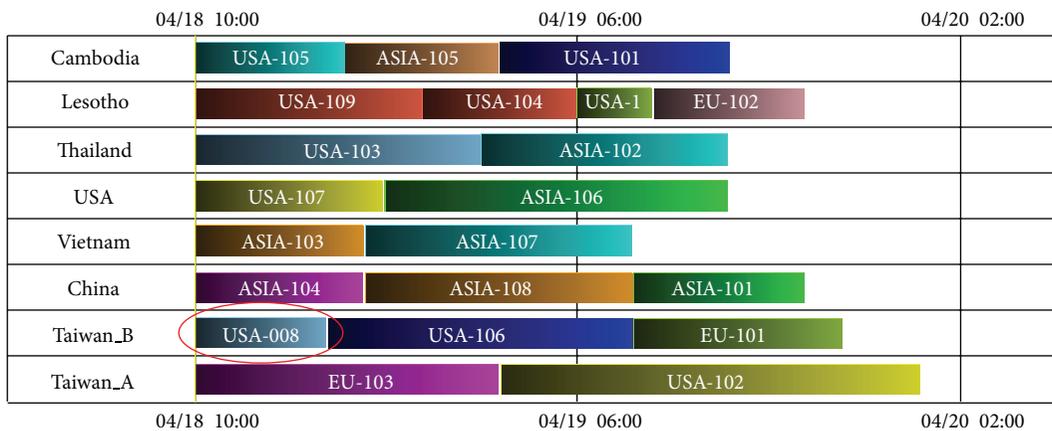


FIGURE 11: Orders are assigned and reassigned on a rolling time basis [12].



(a)



(b)

FIGURE 12: (a) The assignment result. The MO USA-008 is unfinished at time 04/18 10:00. (b) The reassignment result. Note that the MO USA-008 is fixed and the new coming MOs are reassigned.

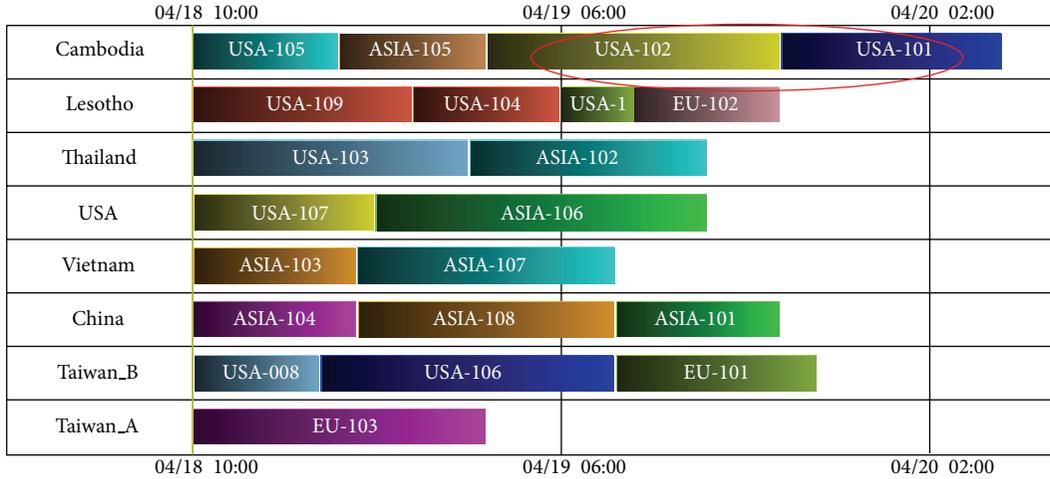


FIGURE 13: The system makes it easy to move or change orders. MO USA-102 was moved from the site in Taiwan A to the site in Cambodia, as compared with Figure 12(b).

assignment under the objectives of the lowest total cost and shortest total production time. To investigate the effectiveness of the proposed approach, this study uses a famous garment company’s real data as a base to perform some experiments. This study also discusses the influences caused by a wide range of order amount demanded. Results from this study show that good workable solutions are obtainable efficiently. When it comes to orders with a wide range of changes in quantity, if managers consider the lowest total cost, they can choose solutions with larger division numbers; on the contrary, if they hope the production time is the shortest, they can choose the ones with smaller division numbers.

There are several suggestions for further studies in the future. This study makes some assumptions for the reduction of complexity. In reality, many uncertain factors, such as cancellation of order, rush order, material shortage and political factors, will cause certain plants to be unable to produce. One possible solution is to introduce autonomous control, that is, to allow some parts of a large network to make their own decisions based on local situations and available information [22, 23]. Studies may also combine MOGA with other methods. In addition, real data in other industries can be collected to perform more experiments to help further understand the differences between different industry sectors.

Abbreviations

Superscripts

- DELAY: Delayed delivery
- DELIV: Delivery
- FG: Finished goods
- MANU: Manufacturing or production
- RM: Raw material.

Subscripts

- h : h th destination
- i : i th MO
- j : j th material
- k : k th Manufacturing site
- c_i^{DELAY} : The penalty cost per day of the i th order caused by delayed delivery
- c_k^{MANU} : The unit manufacturing cost of site k
- c_j^{RM} : The unit material cost of raw material j
- c_{kh}^{DELIV} : The unit delivery cost from site k to destination h
- C : The number of smaller orders; the larger order was divided into groups of C
- D_i : The delayed days of the i th order
- Q_i : The quantity of the i th MO
- Q_i^{FG} : The quantity of finished goods of i th MO
- Q_{ih}^{FG} : The quantity of finished goods of i th MO delivered to destination h
- Q_{ij}^{RM} : The quantity of the i th raw material of i th MO,
- m : The number of factories
- n : The number of manufacturing orders
- S_i : The required capacity of MO i
- S_k : The capacity of the k th site
- z_{kh} : 0-1 variable; $z_{kh} = 1$ if site k delivers products to destination h ; otherwise, $z_{kh} = 0$.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to express their sincere thanks to three anonymous reviewers who provided valuable input to substantially improve this paper. This work was supported partly by the National Science Council under Grants 102-2221-E-025-013 and 98-2622-E-025-002-CC3. Thanks are also extended to Dr. Chih-Chiang Lin, Ms. Nancy D. Wall, Ms. Mei-Ching Wu, Ms. Yu-Ying Chou, Ms. Nuo-Jhen Ma, and Ms. Mei-Hui Wu for their help during the course of this paper.

References

- [1] Z.-L. Chen and G. Pundoor, "Order assignment and scheduling in a supply chain," *Operations Research*, vol. 54, no. 3, pp. 555–572, 2006.
- [2] Z. X. Guo, W. K. Wong, Z. Li, and P. Ren, "Modeling and Pareto optimization of multi-objective order scheduling problems in production planning," *Computers and Industrial Engineering*, vol. 64, no. 4, pp. 972–986, 2013.
- [3] P. P. Dornier, R. Ernst, M. Fender, and P. Kouvelis, *Global Operations and Logistics: Text and Cases*, John Wiley & Sons, New York, NY, USA, 2008.
- [4] R. C. Chen and T. T. Hu, "A decision-making mechanism considering carbon footprint and cost to fulfill orders for multi-site global companies," to appear in *International Journal of Shipping and Transport Logistics*.
- [5] S. Chopra and P. Meindl, *Supply Chain Management: Strategy, Planning, and Operation*, Prentice Hall, Upper Saddle River, NJ, USA, 2012.
- [6] B. Błażewicz, W. Domschke, and E. Pesch, "The job shop scheduling problem: conventional and new solution techniques," *European Journal of Operational Research*, vol. 93, no. 1, pp. 1–33, 1996.
- [7] J. H. Holland, *Adaptation in Natural and Artificial Systems*, MIT Press, Cambridge, Mass, USA, 1975.
- [8] D. E. Goldberg, *Genetic Algorithm in Search, Optimization, and Machine Learning*, Addison-Wesley, Boston, Mass, USA, 1989.
- [9] M. Gen, R. Cheng, and L. Lin, *Network Models and Optimization: Multi-Objective Genetic Algorithm Approach*, Springer, Heidelberg, Germany, 2008.
- [10] R. C. Chen, "Grouping optimization based on social relationships," *Mathematical Problems in Engineering*, vol. 2012, Article ID 170563, 19 pages, 2012.
- [11] R. C. Chen, M. R. Huang, R. G. Chung, and C. J. Hsu, "Allocation of short-term jobs to unemployed citizens amid the global economic downturn using genetic algorithm," *Expert Systems with Applications*, vol. 38, no. 6, pp. 7535–7543, 2011.
- [12] C. Bierwirth and D. C. Mattfeld, "Production scheduling and rescheduling with genetic algorithms," *Evolutionary Computation*, vol. 7, no. 1, pp. 1–17, 1999.
- [13] R. Bhatnagar, P. Chandra, and S. K. Goyal, "Models for multi-plant coordination," *European Journal of Operational Research*, vol. 67, no. 2, pp. 141–160, 1993.
- [14] K. K. Yang and C. C. Sum, "A comparison of job shop dispatching rules using a total cost criterion," *International Journal of Production Research*, vol. 32, no. 4, pp. 807–820, 1994.
- [15] E. Zitzler and L. Thiele, "Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach," *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 4, pp. 257–271, 1999.
- [16] D. Özgen, S. Önüt, B. Gülsün, U. R. Tuzkaya, and G. Tuzkaya, "A two-phase possibilistic linear programming methodology for multi-objective supplier evaluation and order allocation problems," *Information Sciences*, vol. 178, no. 2, pp. 485–500, 2008.
- [17] R. C. Chen, S. S. Li, C. C. Lin, and C. C. Feng, "A GA-based global decision support system for garment production," in *Proceedings of the IEEE International Conference on Neural Networks and Brain (ICNNB '05)*, vol. 2, pp. 805–809, Beijing, China, October 2005.
- [18] R. C. Chen, S. S. Li, C. C. Lin, and T. S. Chen, "Application of self-adaptive genetic algorithm on allocating international demand to global production facilities," in *Proceedings of the 6th World Congress on Intelligent Control and Automation (WCICA '06)*, pp. 7152–7156, Dalian, China, June 2006.
- [19] R. C. Chen, C. H. Chen, and P. H. Hung, "Product miles as an objective to assign orders in a globally multiple plant environment," *Advanced Science Letters*, vol. 19, no. 8, pp. 2231–2235, 2013.
- [20] P. R. Harper, V. de Senna, I. T. Vieira, and A. K. Shahani, "A genetic algorithm for the project assignment problem," *Computers and Operations Research*, vol. 32, no. 5, pp. 1255–1265, 2005.
- [21] C. M. Fonseca and P. J. Fleming, "Genetic algorithms for multi-objective optimization: formulation, discussion and generalization," in *Proceedings of the 5th International Conference on Genetic Algorithms (ICGA '93)*, pp. 416–423, Morgan Kaufmann Publishers, Urbana-Champaign, Ill, USA, 1993.
- [22] H. R. Karimi, N. A. Duffie, and S. Dashkovskiy, "Local capacity H_∞ control for production networks of autonomous work systems with time-varying delays," *IEEE Transactions on Automation Science and Engineering*, vol. 7, no. 4, pp. 849–857, 2010.
- [23] S. Dashkovskiy, M. Görge, and L. Naujok, "Autonomous control methods in logistics—a mathematical perspective," *Applied Mathematical Modelling*, vol. 36, no. 7, pp. 2947–2960, 2012.

Research Article

Application of Stochastic Regression for the Configuration of Microrotary Swaging Processes

Daniel Rippel,¹ Eric Moumi,² Michael Lütjen,¹ Bernd Scholz-Reiter,¹ and Bernd Kuhfuß²

¹ BIBA - Bremer Institut für Produktion und Logistik GmbH, University of Bremen, Hochschulring 20, 28359 Bremen, Germany

² Bremen Institute for Mechanical Engineering, University of Bremen, Badgasteiner Straße 1, 28359 Bremen, Germany

Correspondence should be addressed to Daniel Rippel; rip@biba.uni-bremen.de

Received 31 January 2014; Accepted 5 March 2014; Published 1 June 2014

Academic Editor: Hamid Reza Karimi

Copyright © 2014 Daniel Rippel et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In micromanufacturing, a precise adjustment of manufacturing, handling, and quality control processes constitutes an essential factor for success. The continuing miniaturization of workpieces and production devices results in ever decreasing tolerances, whereas machines and processes become increasingly more specialized. Thereby, the so-called size effects render the direct application of knowledge from the area of macromanufacturing impossible. In this context, this paper describes the application of the μ -ProPIAn method for the configuration of an infeed rotary swaging process for microcomponents. At this, the cause-effect relationships between relevant process parameters are analyzed using stochastic regression models, in order to determine cost-efficient process configurations for the manufacturing of bulk and tubular microcomponents.

1. Introduction

During the last years, the demand of microcomponents increased strongly. Thereby, single components became increasingly smaller, while providing more functionality and consisting of more complex geometries [1]. In addition, increasing production rates and high demands regarding the components quality render the manufacturing of microcomponents a complex task. In this context, several authors define a microcomponent as being smaller than one millimeter in at least two geometrical dimensions [2]. One approach to achieving high throughput rates at comparably low costs is the application of cold forming techniques for micro-manufacturing [3]. Different cold forming processes can be combined to achieve highly flexible manufacturing facilities with comparably low spatial requirements, for example, by applying desktop factories [4]. Such production systems face the challenge of producing vast amounts of high quality components while remaining cost efficient. Thereby, micro-manufacturing is characterized by very low tolerances, the occurrence of so-called size effects [5], and a high degree of specialized manufacturing technologies.

As a result, the adjustment of a diversity of processes in micromanufacturing poses several challenges to the process designer. On the one hand, a careful selection of available process technologies is required. The suitability of certain process technologies strongly depends on the manufacturing context (e.g., materials, tools, preceding, or succeeding processes). Due to the high specialization, it might even be possible that processes have to be adapted or developed for a certain task. On the other hand, the adjustment of selected processes in a process chain has to be performed carefully. This includes manufacturing, handling, and quality control processes. Thereby, the design of process chains in micro-manufacturing involves the configuration of relevant process parameters [6]. Such parameters cover material properties, machine settings, or specific tool configurations. Due to the low tolerances, slight changes to a single parameter can affect the overall process chain, causing high costs for setups or reconfigurations. Consequently, designers require additional support in the selection and adjustment of manufacturing, handling, and quality inspection processes in the area of micromanufacturing [7].

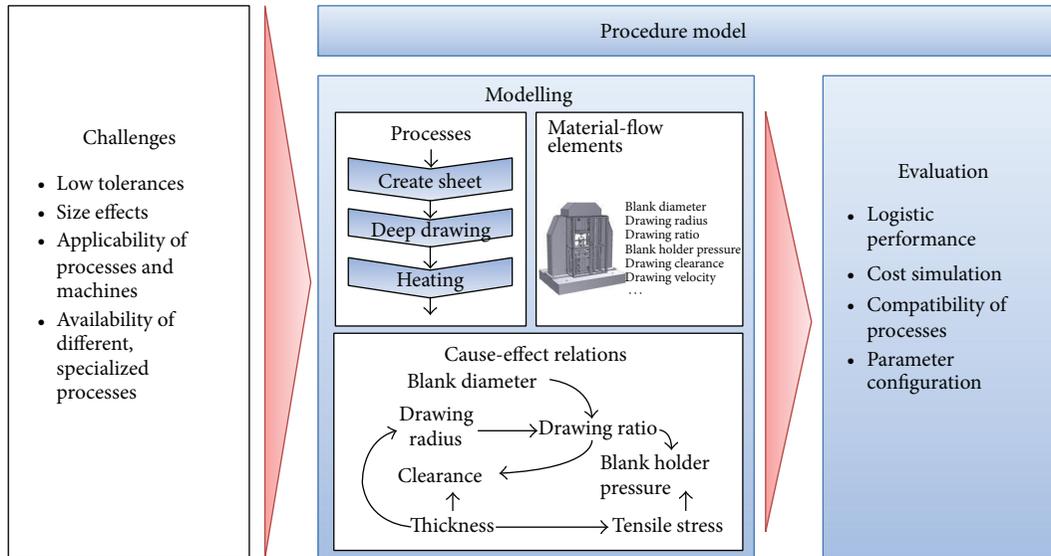


FIGURE 1: Components of the μ -ProPlAn framework (after [11]).

In order to support process designers with the design and configuration of process chains in micromanufacturing, the μ -ProPlAn (microprocess planning and analysis) framework (Figure 1) is being developed. The framework comprises methods to model and evaluate process chains in the context of micromanufacturing, covering all aspects from the process flow over material-flows to the modeling of so-called cause effect networks (Figure 1). These cause-effect networks describe correlations between technical and/or logistic parameters, within the microprocess chain. Moreover, the μ -ProPlAn framework provides different methods to assess the quality of the designed process chains in terms of logistic performance and technical feasibility by computer-aided evaluations.

This paper demonstrates the application of μ -ProPlAn for the configuration of a rotary swaging process in a micromanufacturing scenario. In contrast to classical process design methods, μ -ProPlAn enables an integrated design and evaluation of the (logistic) process flow with respect to the manufacturing processes configurations. Thereby, statistical methods are applied to facilitate material-flow simulations by approximating the single manufacturing processes behaviors and results, based on experimental- or production-related data. The focus of this paper's case study is set on the production of high quality components, while avoiding costly setup operations by reconfiguring related process parameters. The next section introduces the area of micro-cold forming and describes particular challenges encountered in this area. Section 3 outlines the μ -ProPlAn framework and its components. Afterwards, Section 4 describes the use-case scenario and the corresponding process models. Afterwards, it provides the simulation study conducted upon the use-case scenario and presents the results. The paper closes with a conclusion and outlines further work in this area.

2. Micro-Cold Forming

Cold forming processes provide a suitable approach to manufacturing high quality microcomponents, while maintaining high manufacturing accuracies and high throughputs. Additionally, workpieces usually become hardened during the cold forming process, resulting in products that are more robust. Compared to other manufacturing approaches, cold forming processes lead to a reduction of waste materials and energy consumption, leading to an environmentally friendly production [12].

Although cold forming processes are well known and widely used in macromanufacturing for mass production, they cannot be applied directly to micromanufacturing. The downscaling of those cold forming processes, and thus of the workpieces, tools, and machines, is only possible up to a certain degree. Thereafter, the impact of so-called size effects impedes a further downscaling of the cold forming process.

Vollertsen et al. define size effects as “deviations from intensive or proportional extrapolated extensive values of a process, which occur when scaling the geometrical dimensions” [2]. In this context, they define intensive values as parameters, which are not expected to change due to a change of an object's mass (e.g., its temperature or its density). In contrast, extensive values are expected to vary with a different mass (e.g., the object's inertia force or its heat content). Generally, size effects occur due to the inability to scale all relevant process parameters equally [2]. As an example, the downscaling of a metal sheet's thickness can result in a changing density due to local defects, although the density is considered an intensive variable. In addition to these effects, technical limitations further facilitate the occurrence of size effects. For example, the downscaling of mechanical grippers is limited by technical factors. For very small workpieces, the gripper's Van-der-Waals forces will eventually overcome

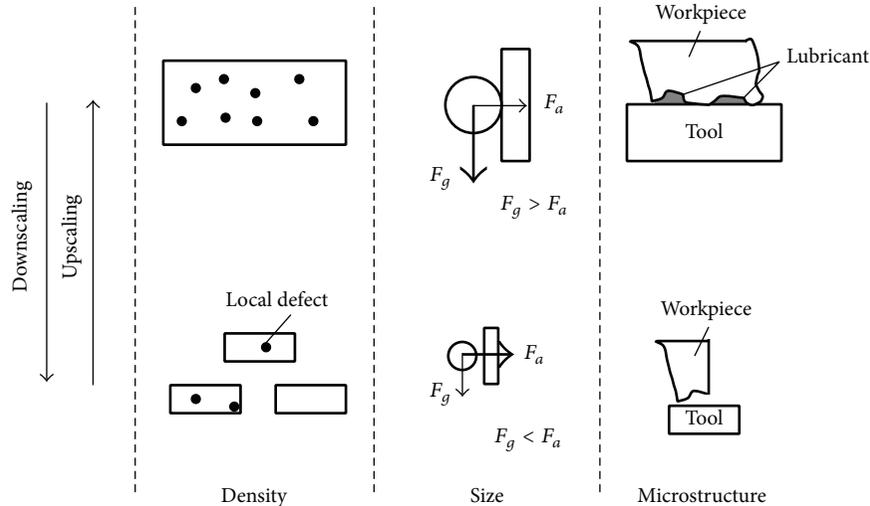


FIGURE 2: Categories of size effects [2].

the gravitational force at a certain point. Consequently, the gripper will not be able to release the workpiece without aid. Basically, Vollertsen and Walther define three distinct categories of size effects (Figure 2) [2].

- (i) Density size effects occur, when the density of a material is held constant while scaling down its geometrical dimensions. For instance, local defects become more serious with a continuing miniaturization. Thereby, the distribution of local defects within a material can lead to more delimited sets of good and bad parts.
- (ii) Shape size effects occur due to the increasing ratio of an object's total surface area, compared to its volume. An example of this category is provided by the described imbalance of the adhesive force in relation to the gravitational force.
- (iii) Microstructure size effects occur because microstructural features (e.g., the grain size or the surface roughness) cannot be scaled down the same way as the geometrical size of an object.

The occurrence of size effects requires a precise planning of technical parameters throughout the overall process chain. Technical parameters between, as well as within, each manufacturing, handling, and quality-inspection process have to be regarded and adjusted to each other. Moreover, as new processes and technologies for micromanufacturing emerge quickly, interdependencies between those parameters cannot be described precisely or are unknown in several cases. As a result, the μ -ProPIAn framework incorporates different methods to determine correlations between production relevant parameters by the application of stochastic methods from the fields of mathematics and artificial intelligence. These methods, on the one hand, enable a detailed analysis of cause-effect relations, while, on the other hand, they provide a suitable approach to determine parameter configurations based on production or experimental data.

3. Microprocess Planning and Analysis (μ -ProPIAn)

As depicted in Figure 1, μ -ProPIAn consists of three basic parts. First, the modelling notation enables a clear and detailed modelling of the production system, covering manufacturing, handling, and quality inspection aspects throughout the process level, the material-flow level down to the configuration level. As a second part, a simultaneous engineering procedure model guides the modelling process. This procedure model aims at iteratively creating and adapting the model in parallel to the product development, enabling the detection of problems or inconsistencies early in the development phase. The third part finally covers the evaluation of the models. This includes the analysis of cause-effect relations during model creation, as well as the evaluation of the modelled production system concerning its technical feasibility and its logistic performance.

3.1. Modelling Notation. The modelling notation relies on the notation of process chains for the description of the overall process (Figure 3) [6, 13]. Thereby, a process chain consists of a sequence of manufacturing, handling, and quality inspection processes. Each process again consists of one or more operations, relating to manufacturing, handling, or quality inspection tasks. The distinction between processes and operations enables the integration of several tasks in one process. For example, in-process quality inspections or simple transportation tasks can be represented as operations within a single process, if all of these tasks are conducted by a single workstation. In contrast, more complex tasks, for example, the separation of workpieces, can be represented as independent processes. In particular, complex quality inspection tasks often require several handling and inspection operations as a part of the overall quality inspection process to be conducted.

In order to adjust processes and operations to each other, the concept of process chains applies the so-called

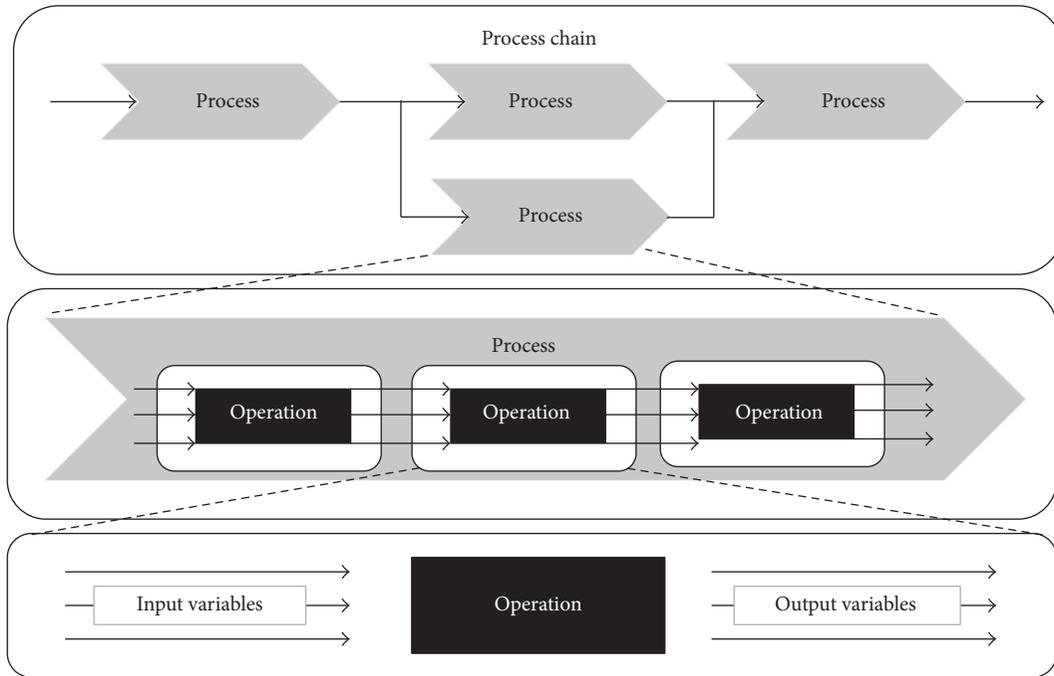


FIGURE 3: Structure of process chains (after [13]).

technological interfaces [13]. Therefore, each operation and, respectively, each process are assigned a set of technological input and output variables, which are modified during the operation. These variables usually constitute attributes of the workpiece, for example, geometrical features, its hardness, or surface roughness. Thereby, input variables describe the required state of the workpiece before the operation, while output variables describe the operations result.

While technological interfaces provide a powerful tool for the configuration of well-known processes in macromanufacturing, they are often inadequate for the configuration of processes in the microdomain. In micromanufacturing, the relationship between process parameters, related to machine or workstation settings, and the effects they have on the workpieces' properties are often unknown or underspecified. In addition, due to the very small tolerances inherent to the microdomain, even small changes to a process parameter can have a strong influence on the resulting workpiece's properties. Consequently, μ -ProPIAn extends the classical concept of process chains by the so-called cause-effect networks. These networks consist of all relevant process and workpiece parameters and depict the cause-effect relations between each parameter. In order to minimize the modelling effort and to facilitate the reuse of cause-effect networks, they are modelled as a part of the respective material-flow element, for example, the machine, workpiece, or tool, and later on combined to cause-effect networks for an operation by assigning the material-flow elements used in the respective operation. Thus, each operation at least consists of an input and an output workpiece, usually constituting the operation's technological interfaces, as well as a machine and a tool, used to conduct the operation. Additionally, personnel and operating

supplies can be assigned in order to enable a comprehensive logistic evaluation. Figure 4 shows a simplified cause-effect network for the operation "Rotary Swaging," at which several technological and logistic parameters were left out to increase the readability. This network is constructed from the cause-effect networks modelled for the rotary swaging machine (left), the respective workpiece (center), and the tool (right) used in this operation. Therefore, the single networks are connected via cause-effect relations between their parameters. In addition to the technological parameters, each cause-effect network contains a set of default logistic parameters. For example a tool requires the definition of its durability and its investment costs; operations require the specification of their duration and rejection rates, while machines provide parameters related to maintenance, malfunctions, and costs inflicted while processing or while being idle (for a complete list see [11]).

In order to enable the configuration of each operation and to enable a comprehensive technological and logistic evaluation of the modelled production system, the value of each parameter should be estimated based on its input parameters. Therefore, μ -ProPIAn requires each parameter's cause-effect relations to be described quantitatively. In cases, in which the relationships are known, a mathematical function $y = f(x_1, x_2, \dots, x_n)$ can be defined for each parameter. In case of unknown cause-effect relations between the predictors and the dependent parameter, μ -ProPIAn offers a set of methods to create an analytical regression model or to learn abstract prediction models, based on experimental or production-related data. A summary of these methods is provided in Table 1, together with a short description of each method. In addition to this quantitative description, each parameter

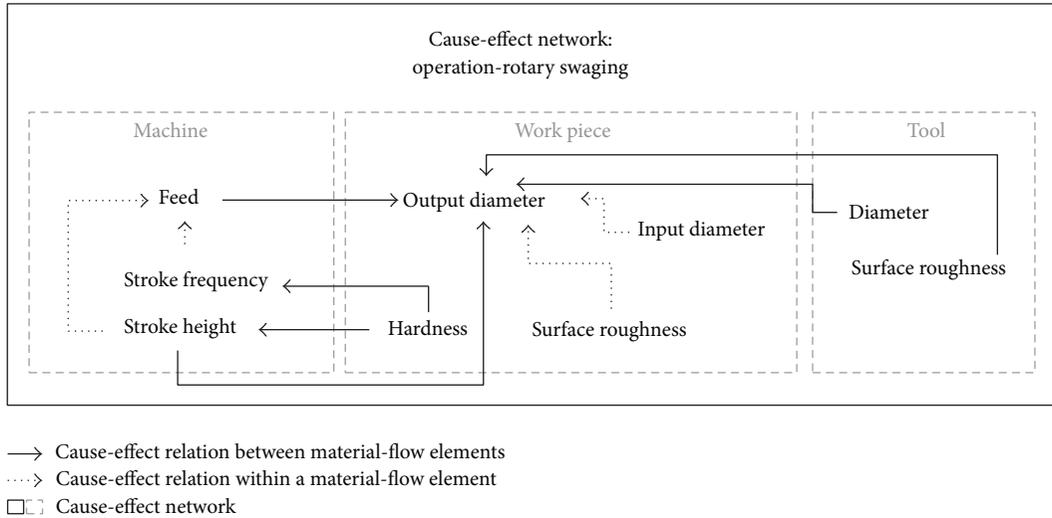


FIGURE 4: Simplified cause-effect network for the operation “Rotary Swaging.”

is assigned an interval of valid values. For example, some machines can only use tools with a specific range of sizes or operate using specific feed rates.

3.2. *Simultaneous Engineering Procedure Model.* The μ -ProPIAn procedure model aims to support the development of new process chains in the context of a running production. Thereby, it relies on the existence of premodelled material-flow elements, although it is easily possible to add additional machines, tools, and so forth during the development. The procedure model utilizes concepts of “Simultaneous Engineering” [6]. In general, “Simultaneous Engineering” describes an approach, in which the different phases of new product development, from the first basic idea to the moment when the new product finally goes into production, are carried out in parallel [14]. By parallelizing the development of the product and of the manufacturing processes, all characteristics of the product’s life cycle can be taken into account early. Ignoring such issues may lead to high costs in later stages of the product development [15]. Due to the high specialization of micromanufacturing processes, available manufacturing and handling technologies might not be suitable for a provided task. Accordingly, new technologies or tools have to be developed or the product design has to be adapted. The application of Simultaneous Engineering techniques enables an early detection of such problems, which would usually emerge at later stages of the production planning [16].

The procedure model consists of three phases (Figure 5). During the first phase, the overall process chain is developed and configured. Thereby, the process designer includes required manufacturing, handling, and quality inspection processes and defines the processes’ technological interfaces, based on the product structure. In case of changes to the product structure, this phase can be repeated to adapt the model. The second phase covers the specification and configuration of the required resources. Thereby,

the process designer selects suitable technologies and machines for each operation and combines their respective cause-effect networks. The last phase covers the model analysis. As a first step, he evaluates the cause-effect networks to ensure the technical feasibility of the process chain concerning the product’s and the respective machines’ tolerances. In a second step, valid process chains can be analyzed by means of a material-flow simulation, in order to determine and compare their logistic performances, for example, throughput times, due date adherence, workloads, or total costs. Thereby, it is possible to compare different process chains, for example, by varying used technologies, machines, or configurations, in order to determine the optimal setup and configuration.

3.3. *Model Evaluation.* The evaluation of the modelled production system is split in two parts: first, the evaluation of a process chain’s configuration, and thus, its technological feasibility and second, the evaluation of its logistic performance. In order to validate the technological feasibility, all parameter values are propagated throughout the complete process chain using the quantitative cause-effect relations. Thereby, each parameter is calculated based on its input parameters and checked against the specified tolerances with respect to the technological interfaces and to the machines capabilities. If a parameter exceeds these tolerances, the configuration has to be adapted or, in the worst case, a more suitable technology or machine has to be selected.

After defining one or more technically feasible process chains, these models can be evaluated with reference to their logistic performance. Therefore, the process chain model is converted into a material-flow simulation, whereby the cause-effect networks’ parameters are directly conveyed into the simulation. Due to the integration of technological and logistic parameters within the cause-effect networks, changes to technological parameters can directly influence their logistic counterpart. For example, an increasing feed rate usually decreases the duration, while increasing the rejection

TABLE 1: Regression and learning methods included in the μ -ProPIAn software prototype.

Method	Description
Linear regression	As fundamental multivariate regression function, μ -ProPIAn offers the option to perform a simple linear regression on the provided data. As a result, a function of the form $y = \sum_{i=1}^n (a_i \cdot x_i) + b$ is determined. Thereby, the least squares method is applied to minimize the distance between the regression model and the original data points. Additionally, μ -ProPIAn offers additional functionality to linearize, for example, exponential or logarithmic data to enable linear regressions.
Polynomial regression	In case of univariate cause-effect relations, a polynomial regression can be conducted to achieve functions of the form $y = a_0 + a_1x + a_2x^2 + a_3x^3 + \dots + a_nx^n + \varepsilon$. This method again uses the least squares method.
Tree-/rule-based regressions (e.g., [8])	In contrast to the methods above, tree- or rule-based approaches do not result in a single analytic function. In general, they divide the search space into smaller segments, for which a regression can be performed. Usually, both methods use linear regressions for each segment.
Locally weighted linear regression (LWL) (e.g., [9])	LWL constitutes an abstract prediction model, which performs a locally weighted linear regression each time a prediction is requested. Thereby, a kernel function is used to weight adjacent data points and finally a linear regression is performed. This method particularly excels in interpolating missing data in between available data points.
Support vector regression(e.g., [10])	A support vector machine is usually used as classifier. Thereby, it learns a model, which separates a set of data points in one or more classes, maximizing the distance between each data point and the classification curve. The same method can be used for regression, particularly if the provided data contains strong variances.
Neuronal networks	A neural network usually consists of a number of layers, each containing a defined number of nodes. The nodes of each layer are interconnected. During the supervised training of the network, these connections' weights are adapted to recreate the desired output on the last layer. Thereby, the quality of the prediction strongly depends on a suitable network structure (i.e., number of layers/nodes, selected activation functions, etc.)

rate of an operation. The material-flow simulation provides several statistics regarding the underlying production system. General statistics refer to the throughput times, the due-date adherence, the single devices utilization, the work-in-progress, and the costs incurred. For each machine, more specific statistics regarding costs, down times, rejection rates, and so forth are recorded to enable a comprehensive evaluation of a single process chain as well as a comparison between different process chain models.

4. Use-Case Scenario

This section describes the application of μ -ProPIAn for the configuration of a microrotary swaging process. Although μ -ProPIAn was originally designed to adjust several processes along a process chain, the same methods can be applied for the optimization of a single process within a given process chain.

4.1. Rotary Swaging Process. Rotary swaging is a well-known cold forming process, for example, in the automotive industry. Specific characteristics of this process support lightweight and stress optimized designs of tubular components like steering columns or drive shafts [17]. The workpiece forming takes place in the swaging head in incremental steps by radial oscillating dies. Two process types are established: infeed swaging and plunge swaging that differs in the direction of the feed motion (Figure 6(a)).

An ongoing demand for miniaturization and functional integration of mechatronic systems, for example, in medical applications (minimally invasive diagnosis and surgery) or automotive industry (miniature pumps and valves), increasingly brings rotary swaging into the focus of interest, particularly on the microscale. Since 2007, research is performed to adapt the rotary swaging process to microcomponents, with a diameter between 0.3 and 1.0 mm (c.f. Figure 6(b)) [18, 19]. Actually, the prediction of the resulting workpiece's

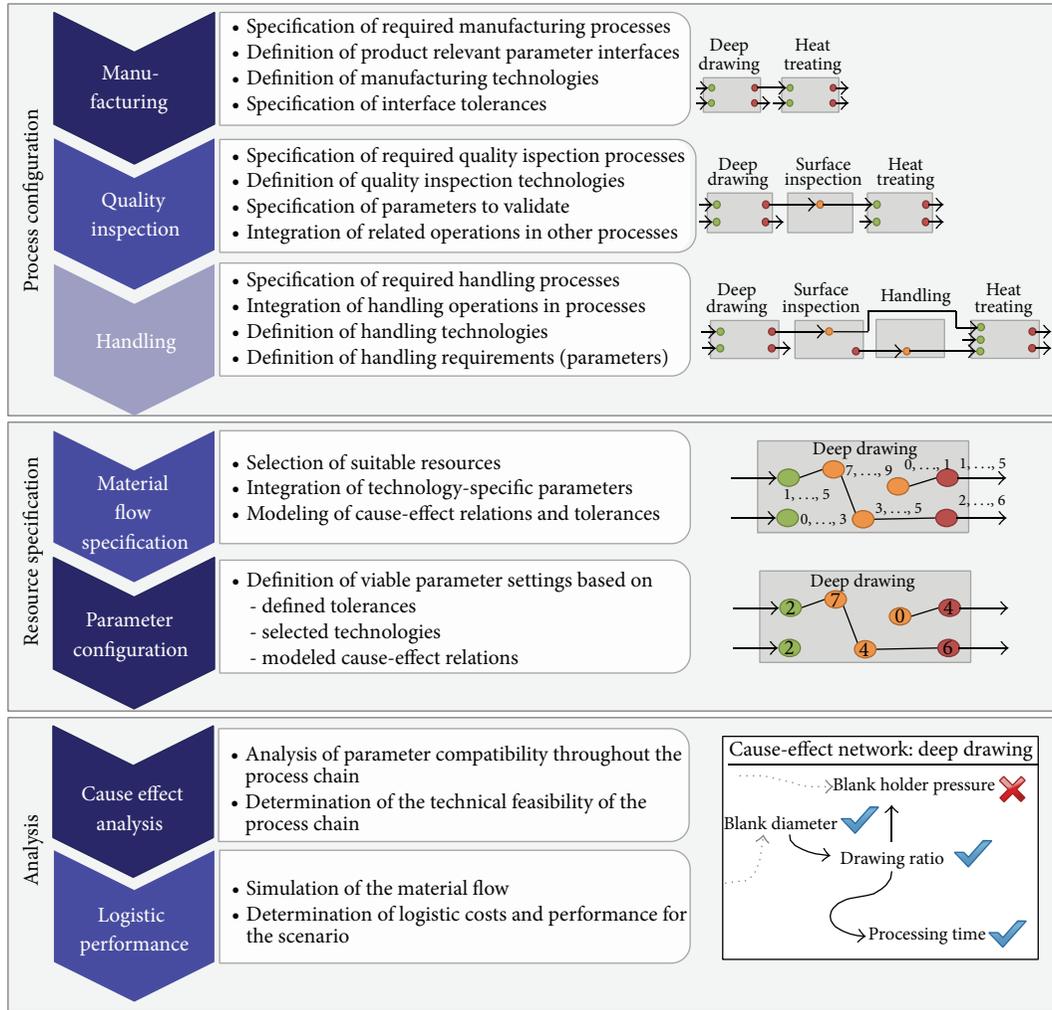


FIGURE 5: μ -ProPIAn procedure model.

quality, for example, the final diameter with respect to the process parameters, is insufficient, even in the macroscale [20]. Material properties (e.g., its hardness), the die closing pressure, the dies' oscillation frequency, and the feed rate mainly affect the remaining spring-back, which leads to a difference between the closed dies' diameter and the finished workpiece's diameter.

4.2. Model Definition and Simulation Study. In general, the process chain for this use-case scenario consists of four processes. First, the raw material, a steel wire with a diameter of 1 mm, is heat treated to increase its degree of deformation and to achieve a consistent hardness. Then, the described rotary swaging process is applied to decrease the diameter. Afterwards, a laser-based free from heating process is applied to create a mass accumulation at the end of the wire. Finally, plunge rotary swaging or another cold forming process is used to calibrate the mass accumulation into its final shape. As the primary focus of the experiment was the determination of the most cost-efficient setup in manufacturing steel

wires with a defined diameter and mechanical properties (second process), the remaining processes were considered constant within the model. Thus, they were assigned a fixed duration and a fixed rejection rate independent of the microrotary swaging process' configuration.

In order to determine the most cost-efficient configuration for the microrotary swaging process, a set of 55 experiments was conducted using the described device. For these experiments, each of the three primary machining parameters (feed rate, oscillation frequency/stroke frequency, and die closing pressure) was modified, while the other two were held constant. As a result, the diameter of the resulting wire was measured seven times along the wire, with a distance of 1 cm between each measurement. Each experiment was repeated four times to determine process variances. For all experiments, the remaining process parameters were held constant, while the same material and tools were used to reduce their effects on the process and to simplify the cause-effect network. In case of the experimental setup, the die closing pressure is regulated using an additional, adjustable wedge, inflicting pressure on the dies (compare Figure 6(a)).

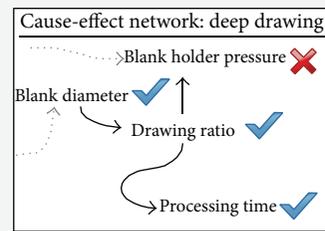


TABLE 2: Parameter definition.

Parameter		Unit	Quantitative function	Valid values
Die closing pressure	P	[Position]	—	[69; 73]
Oscillation frequency	F	[Hz]	—	[30; 120]
Feed rate	V_f	$[\mu\text{m/s}]$	—	[1000; 4000]
Initial diameter	D_0	$[\mu\text{m}]$	—	[350; 1000]
Final diameter	D_1	$[\mu\text{m}]$	$D_1 = D_T - 35.20195 \cdot P + 9.747 \cdot \ln(V_f) - 0.36423 \cdot F + 2.472e03$	[300; 1000]
Final diameter variance	D_v	$[\mu\text{m}]$	Random value in between 0 and 2.5	[0; 1000]
Tool-diameter	D_T	$[\mu\text{m}]$	—	[300; 1000]
Output-length	L	$[\mu\text{m}]$	—	[1000; ∞]
Tolerance	T	[%]	—	[0; 100]
Operation-duration	O_T	[Time Units]	$O_T = L/V_f$	[0; ∞]
Operation-rejection rate	O_R	[%]	$O_R = \max(2, (100/D_1) \cdot (D_v - T))$	[2; 100]

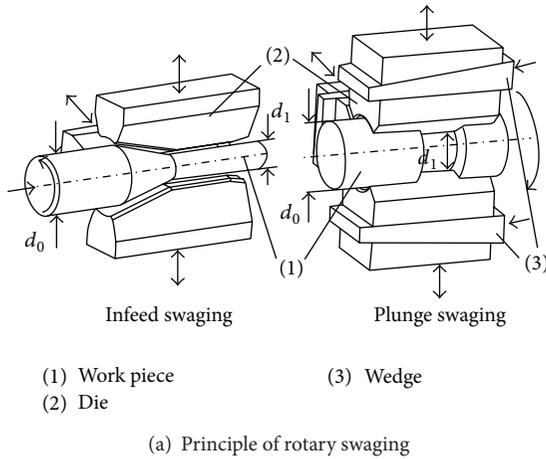


FIGURE 6: Infeed rotary swaging.

The material-flow elements (workpieces, the rotary swaging device, and the tools) were created and assigned all relevant parameters. For the rotary swaging operation, a cause-effect network was constructed as sketched in Figure 7. Thereby, the final diameter is assumed to depend on the tool diameter as well as on the three machining parameters. For the rotary swaging operation, the duration depends on the feed rate and the desired length of the final wire, while its rejection rate depends on a specified tolerance and the maximal variance of the final diameter. The variance itself is assumed to depend on the machining parameters.

In order to quantify the cause-effect network, the data was first preexamined. In a first step, the single relations between the machining parameters and the final diameter were investigated, to enable a suitable selection for the learning/regression algorithm. Therefore, a small subset of experiments was taken from the training data, in which only one of the respective parameters varied. As depicted in Figures 8(a)–8(c) the machining parameters showed a linear, respectively, logarithmic influence. Consequently, the final diameter was quantified using a generalized adaptive linear regression model of the form $D_1 = D_T + c_1 \cdot P + c_2 \cdot \ln(V_f) + c_3 \cdot F + c_4$. As a result of the regression, the coefficients were determined as provided in Table 2. Thereby, a residual standard error of 3.287 remained while the multiple R -squared value of 0.992 (adjusted R -squared: 0.9915) was achieved.

Another result of this preliminary investigation showed a clear process window for each of these parameters (summarized in Table 2). Exceeding these windows generally leads either to strong, visible deformations of the final workpiece or to insufficient deformation results. In a second step, the diameters variance was characterized as the standard deviation over the distinct measurements of the final diameter for each experiment. Unfortunately, no covariance between the variance and any other parameter within the cause-effect network could be determined. Due to the consistent variance between 0 and 2.5 μm for all final diameters between 370 μm and 570 μm (Figure 8(d)), the variance was treated as a random variable. Nevertheless, some experiments resulted in higher variances, assumingly due to material defects. In order to include these outliers in the model, a minimum rejection rate of 2% was assumed. Table 2 summarizes the parameters together with their valid ranges and quantitative functions.

4.3. Simulation Model and Results. In general, for a given set of material-flow elements (machine, tool, workers, etc.), the balancing of the process duration and its rejection rate (based on the quality of the resulting workpieces) implicates a major task in the process configuration, in order to balance the incurring process and material costs. Nevertheless, the quantification and the respective preevaluation of the data on the one hand demonstrated that the process behaved very

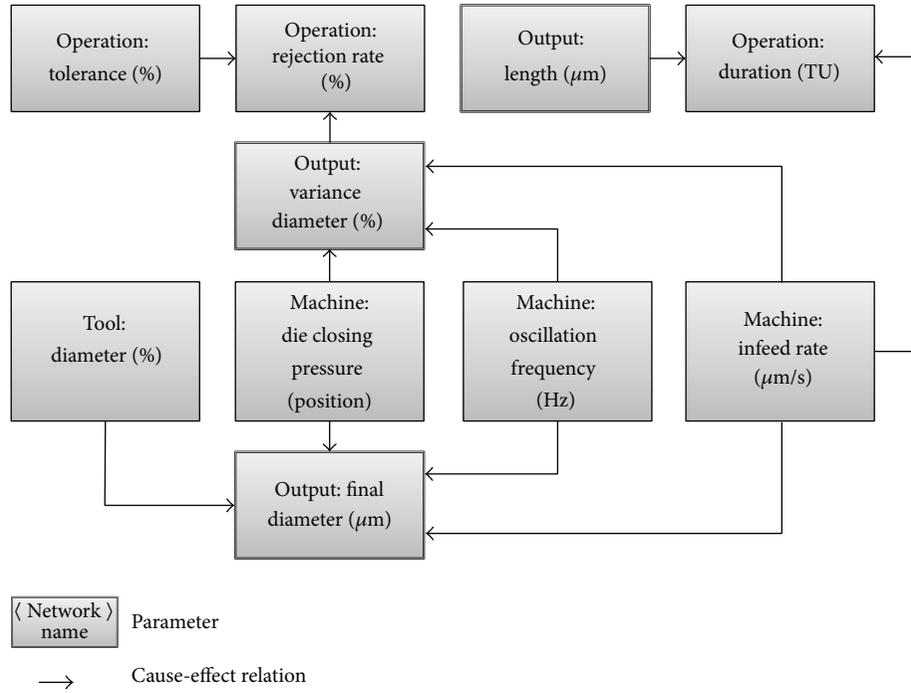


FIGURE 7: Cause-effect network for the rotary swaging operation.

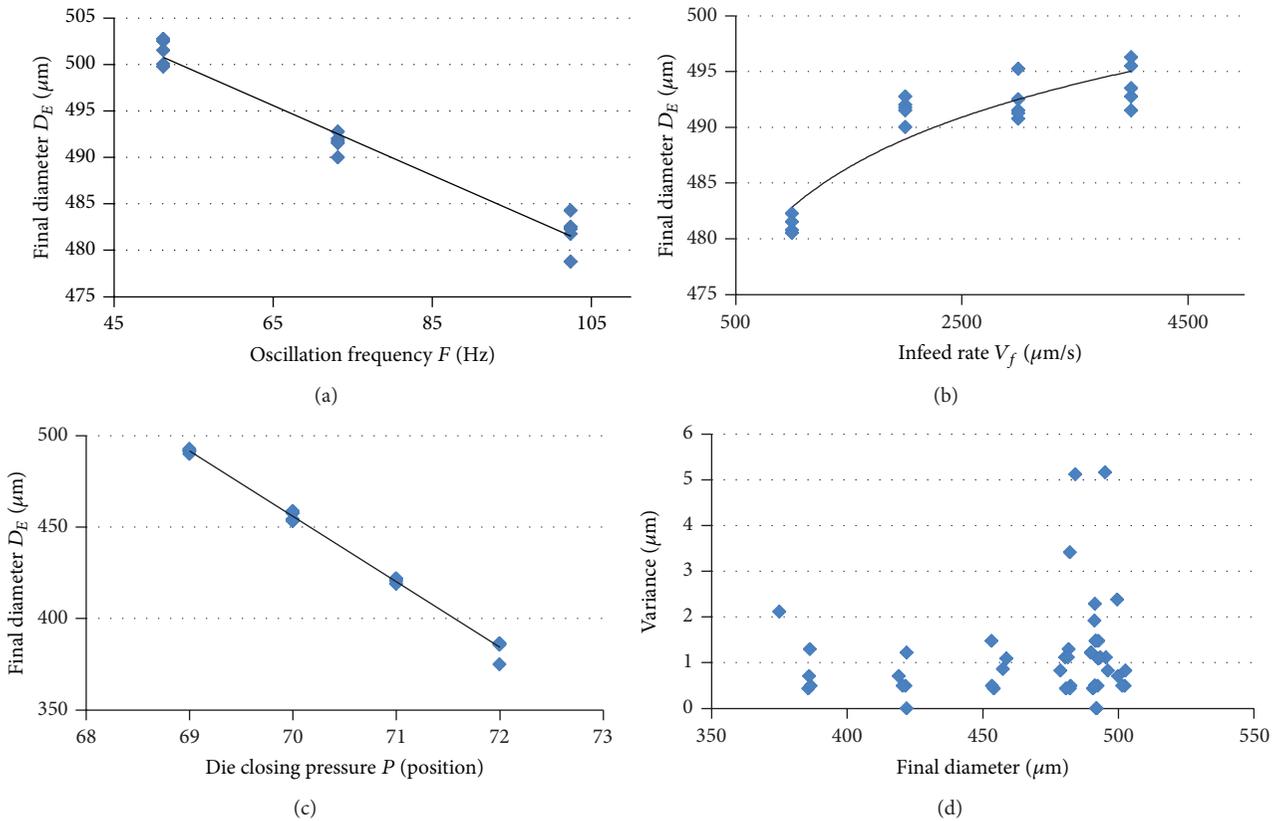


FIGURE 8: Scatterplots of selected experiments relating to the influence of single machining parameters (a)–(c) and to the distribution of the final diameters variance (d).

TABLE 3: Parameter configurations used in the second simulation scenario.

Final diameter [μm]	Tool diameter [μm]	Die closing pressure [position]	Feed rate [$\mu\text{m/s}$]	Oscillation frequency [Hz]	Process duration [sec]
300	300	72.0	3000	36,587	33
350	300	70.0	2000	87,808	50
400	300	69.0	1350	36,587	74
350	400	73.0	2100	73,173	48
400	400	71.5	2000	80,491	50
450	400	70.0	2000	87,808	50
500	400	69.0	2000	43,904	50
450	500	73.0	2100	73,173	48
500	500	71.5	2050	80,491	49
550	500	70.0	2000	80,491	50
600	500	69.0	1350	36,587	74

TABLE 4: Simulation results.

	Scenario 1	Scenario 2
Simulation statistics		
Simulated time elapsed (time units)	76838.0	50358.8
Number of jobs finished	1000.0	1000.0
WIP	5048.0	2974.1
Machine statistics		
Average queue length (Jobs)	2581,7653	1492,5153
Maximum queue length (Jobs)	3104,7659	1568,2296
Average utilization (incl. setup) (%)	100,0000	100,0000
Average utilization (incl. setup) (time units)	76838,0000	50358,8000
Average utilization (excl. setup) (time units)	50000,0000	49938,8000
Average utilization (excl. setup) (%)	65,0720	99,1660

stable, as long as all parameters remained within their respective windows. In contrast to other processes investigated using μ -ProPIAn (e.g., [21]), an increasing feed rate does not lead to a significant decrease of the final workpieces quality. On the other hand, the experiments showed a very strong influence of the machining parameters on the final workpiece diameter. Thereby, a broad range of final diameters could be achieved using the same tool (in terms of the die opening diameter) while altering the parameters. Actually, manufacturing takes place using a tool with a diameter of about $\pm 50 \mu\text{m}$ difference to the final diameter, while producing several test components to calibrate the process parameters. Using the regression function provided in Table 2, a more precise adjustment of the parameters could be achieved. Moreover, by exploiting the wide range of final diameters achievable with a single tool, setup times can be reduced strongly. Therefore, this subsection describes a simulation study conducted to compare the classical approach with an approach based on the application of the regression model with the aim to reduce setup times.

The simulation model covers the process chain described earlier, whereby the other three processes retain a constant duration and rejection rate. The infeed rotary swaging process

assumes the same machine used in the experiments in order to derive real-world information about setup times and process behavior. For this device, it is possible to change the machining parameters in-process. Once modified, the machine directly operates using the new settings. The exchange of a tool requires about 7 minutes including the de- and reconstruction of the outer casings. To ease the simulation, all wires (jobs) are presumed to have a pre-determined, consistent length of 10 cm ($100.000 \mu\text{m}$), while the diameter varies between $300 \mu\text{m}$ and $600 \mu\text{m}$ in steps of $50 \mu\text{m}$ depending on the job. Setup—thus a change of parameters and/or the exchange of a tool—is only performed after a job is completed. The simulation uses a dynamic job source for each type of job (diameter). Consequently, jobs are generated in a random sequence and amount for all available diameters. Thereby, the random generators use a fixed seed to ensure the same random sequence of orders for all simulation scenarios. In general, the simulation assumes a buffered production and applies priority rules for shop floor control. Priority rules enable each machine to select the most appropriate job from their buffers. For this simulation, the priority rule first selects jobs with a minimum setup time. If several jobs require the same setup time, the most urgent jobs are selected and executed.

For the classic simulation scenario, an individual tool is used for each type of job. Thus, to manufacture a wire with a diameter of $300\ \mu\text{m}$ a tool with a diameter of $300\ \mu\text{m}$ is used, resulting in 7 tools in total. Thereby, the simulation assumes that the parameter settings for each configuration are known and no further experiments are required to restart manufacturing after the setup. For the simulation scenario that uses the parameter settings from the regression model, only three different tools are required to cover the complete range of products. Table 3 summarizes the parameter settings and tool selections used in this simulation scenario.

Each simulation scenario was conducted ten times, whereas the simulation covered the fixed production of 1000 workpieces. The approach relying on parameter reconfiguration was thereby capable of reducing the required setup time by 34%. Table 4 summarizes the main results of both simulation scenarios.

5. Conclusion and Future Work

The application of regression models provides a suitable approach to describing the interrelations between different parameters in a manufacturing context. Particularly for domains such as micromanufacturing, in which the specific effects of parameter changes are unknown or barely describable, such models can provide a deeper understanding of the processes based on experimental data. As demonstrated, a more detailed understanding of these effects can be used in the development of process control strategies, for example, to avoid expensive setup operations or to reduce process times in general. Future work in this use-case will deal with the inclusion of additional data and parameters to the cause-effect networks (e.g., material properties) to cover a wider range of manufacturing scenarios. In addition, more elaborate production controllers or production control strategies will be developed based on the extended cause-effect networks. In general, future work will focus on extending μ -ProPIAn by methods for the automatic deduction of suitable process configurations along a process chain. This will enable a more precise adjustment of single processes. Moreover, future work will investigate possible methods to combine discrete information derived from single experiments with continuous sensor data gained during production. By combining these different types of values, more detailed cause-effect networks can be created and the acquisition of data, particularly in industrial applications, can be eased.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The authors gratefully acknowledge the financial support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for Subproject C4 “Simultaneous Engineering” and Subproject A4 “Material Displacement” within the CRC 747 (Collaborative Research Center).

References

- [1] J. P. Wulfsberg, T. Redlich, and P. Kohrs, “Square foot manufacturing: a new production concept for micro manufacturing,” *Production Engineering—Research and Development*, vol. 4, no. 1, pp. 75–83, 2010.
- [2] F. Vollertsen, H. S. Niehoff, and Z. Hu, “State of the art in micro forming,” in *Proceedings of the 1st International Conference on New Forming Technology*, Harbin Institute of Technology Press, Harbin, China, 2004.
- [3] Y. Qin, “Micro-forming and miniature manufacturing systems—development needs and perspectives,” *Journal of Materials Processing Technology*, vol. 177, no. 1–3, pp. 8–18, 2006.
- [4] O. Klemd, “Desktop factory—new approaches for lean micro assembly,” in *Proceedings of the IEEE International Symposium on Assembly and Manufacturing (ISAM '07)*, pp. 161–165, Ann Arbor, Mich, USA, July 2007.
- [5] F. Vollertsen and R. Walther, “Energy balance in laser-based free form heading,” *CIRP Annals—Manufacturing Technology*, vol. 57, no. 1, pp. 291–294, 2008.
- [6] B. Scholz-Reiter, M. Lütjen, and N. Brenner, “Technologieinduzierte Wirkungszusammenhänge in der Mikroproduktion—Entwicklung eines Modellierungskonzepts,” in *22. HAB-Forschungssseminar “Digital Engineering—Herausforderung für die Arbeits- und Betriebsorganisation”*, M. Schenk, Ed., pp. 81–102, GITO, Magdeburg, Germany, 2009.
- [7] B. Scholz-Reiter, N. Brenner, and A. Kirchheim, “Integrated micro process chains,” in *Advances in Production Management Systems*, B. Vellespir and A. Thècle, Eds., pp. 27–32, Springer, Berlin, Germany, 2010.
- [8] G. Holmes, M. Hall, and E. Frank, “Generating rule sets from model trees,” in *Proceedings of the 12th Australian Joint Conference on Artificial Intelligence*, pp. 1–12, Springer, Berlin, Germany, 1999.
- [9] E. Frank, M. Hall, and B. Pfahringer, “Locally weighted naive bayes,” in *Proceedings of the 19th Conference in Uncertainty in Artificial Intelligence*, pp. 249–256, Morgan Kaufmann, 2003.
- [10] J. Platt, “Fast training of support vector machines using sequential minimal optimization,” in *Advances in Kernel Methods—Support Vector Learning*, B. Schoelkopf, C. Burges, and A. Smola, Eds., MIT Press, 1998.
- [11] B. Scholz-Reiter and D. Rippel, “Eine Methode zum Design von Mikroprozessketten,” *Industrie Management*, vol. 29, no. 2, pp. 15–19, 2013.
- [12] E. P. DeGarmo, J. T. Black, and R. A. Kohser, *Materials and Processes in Manufacturing*, John Wiley & Sons, 9th edition, 2003.
- [13] B. Denkena and H. K. Tönshoff, “Prozessauslegung und -integration in die Prozesskette,” in *Spanen—Grundlagen*, B. Denkena and H. K. Tönshoff, Eds., pp. 339–362, Springer, Berlin, Germany, 2011.
- [14] R. G. Schroeder and B. B. Flynn, *High Performance Manufacturing: Global Perspectives*, John Wiley & Sons, New York, NY, USA, 2002.
- [15] T. T. Pullan, M. Bhasi, and G. Madhu, “Application of object-oriented framework on manufacturing domain,” *Journal of Manufacturing Technology Management*, vol. 22, no. 7, pp. 906–928, 2011.
- [16] G. Reinhart and J. F. Meis, “Requirements management as a success factor for simultaneous engineering—enabling manufacturing competitiveness and economic sustainability,” in

Enabling Manufacturing Competitiveness and Economic Sustainability, H. A. ElMaraghy, Ed., pp. 221–226, Springer, Berlin, Germany, 2012.

- [17] B. Kuhfuss, “Hollow drive shafts—a contribution towards weight and cost reduction in automotive construction,” *Automotive Technology International*, no. 4, pp. 106–110, 1998.
- [18] B. Kuhfuss, E. Mouri, and V. Piwek, “Micro rotary swaging: process limitations and attempts to their extension,” *Microsystem Technologies*, vol. 14, no. 12, pp. 1995–2000, 2008.
- [19] B. Kuhfuss and E. Mouri, “Influence of the feedrate on work quality in micro rotary swaging,” in *Proceedings of the 3rd International Conference on Micromanufacturing (ICOMM '08)*, 2008.
- [20] P. Groche, D. Fritsche, E. A. Tekkaya, J. M. Allwood, G. Hirt, and R. Neugebauer, “Incremental bulk metal forming,” *CIRP Annals—Manufacturing Technology*, vol. 56, no. 2, pp. 635–656, 2007.
- [21] D. Rippel, M. Lütjen, and B. Scholz-Reiter, “A framework for the quality-oriented design of micro manufacturing process chains,” *Journal of Manufacturing Technology Management*, vol. 25, no. 7, 2014.

Research Article

An Uncertain Programming for the Integrated Planning of Production and Transportation

Deyi Mou and Xiaoding Chang

Institute of Mathematics for Applications, Civil Aviation University of China, Tianjin 300300, China

Correspondence should be addressed to Deyi Mou; deyimou@hotmail.com

Received 17 March 2014; Revised 13 May 2014; Accepted 13 May 2014; Published 29 May 2014

Academic Editor: Michael Lütjen

Copyright © 2014 D. Mou and X. Chang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The goal of this paper is to tackle joint decisions in assigning production and organizing transportation for single product in a production-transportation network system with multiple manufacturers and multiple demands. In order to meet practical situation, assume that the variant costs and the amounts of the consumption of raw materials that every manufacturer produces per unit product are all uncertain variables in manufacturing processes; meanwhile, the demands that each destination needs are random variables in the transportation problem. Then, a joint optimization model of production and transportation is developed, in which the uncertain chance constraint and the stochastic chance constraint are applied in the manufacturing processes and the transporting processes, respectively, and transformed into a deterministic form by taking expected value on objective function and confidence level on the constraint functions. Finally, a practical example points out the effectiveness of our model.

1. Introduction

Production problem and transportation problem, which were solved separately in the past, are two important problems for the manufacturing firms. But in order to pursue the maximization of benefits, the firms have to consider the whole process of production and transportation which contains not only how to assign the production but also how to design the transportation plan. Hence, many researches have been made to coordinate production and transportation network, and more effort is now being made to do this work. Glover et al. [1] presented the production, distribution, and inventory planning system (PDI) in the form of supply chain management (SCM). In the network, by introducing the supply policy node and supply policy arcs to each of the supply sites and demand policy arcs from each demand site to a demand policy node, the Agrico PDI network model was established to minimize the costs. After that, the integrated models in production and transportation ([2–5]) and production and inventory [6–8], as well as transportation and inventory were investigated one after another. All of these models are directly or indirectly linked with inventory and inventory cost is a significant portion of the total cost.

Additionally, these models were also formulated into different versions of single or multiple periods, single or multiple products [9], and so on.

On the other hand, in almost all the works mentioned above, the conventional mathematical programming methods are used to solve problems in SCM and the goals and relevant inputs are generally assumed to be deterministic/crisp. However, in real-world problems in supply chains, environmental coefficients and/or parameters, involving available supply, resources, capacities, demand, and related operating costs, are often uncertain owing to information being incomplete and unavailable over the planning horizon. Therefore, conventional deterministic mathematical programming methods cannot solve all programming problems in uncertain environments. Holmberg and Tuy [10] developed a stochastic programming to deal with the integrated decisions of production and transportation modeled by the convex nonlinear production costs and stochastic demand. They reported computational tests that indicated that quite large problems can be solved efficiently. Liang [11] presented a possibilistic linear programming (PLP) method for solving the integrated manufacturing/distribution planning decision problems with multiple imprecise goals in supply chains

under an uncertain environment. An industrial case was used to demonstrate the feasibility of applying the proposed method to a real problem.

Because of the lack of the information about the raw materials, changes in sales, weather and road conditions, and so forth, there are unpredicted things which will make the amounts of the supply of the raw materials, the costs of production and transportation, and the demands be uncertain values. There are some limitations when the traditional stochastic models and possibilistic methods above deal with such problems at these situations. In such cases, we have to invite some experts to evaluate their degree of belief that each event will occur. However, humans tend to overweigh unlikely events [12]; thus, the degree of belief may have a much larger range than the real frequency. In this situation, if we insist on dealing with the degree of belief using the probability theory, some counterintuitive results will be obtained. In order to deal with the uncertain situations, Liu [13] presented the uncertainty theory. After that, Liu [14, 15] made some researches about uncertainty theory and proposed the uncertain programming and then applied the uncertainty theory to model human uncertainty [16] and refined the uncertainty theory in 2013 [17]. More and more researchers have contributed to this area. Liu and Ha [18] researched the expected value of function of uncertain variables. Rong [19] proposed two new uncertain programming models of inventory with uncertain cost. So the uncertain programming began to be regarded as a new technique of mathematic programming and be applied widely. Based on the uncertain programming, the transportation problem [20, 21], the facility location problem [22], and the vehicle routing problem [23] were all discussed in the simplest basic form. Until now, no studies are found on the integrated decisions by exploiting uncertainty theory about production, transportation, or inventory. In other ways of application of the uncertainty theory, uncertain hypothesis testing [24] and uncertain optimal control [25] were also studied, respectively.

Motivated by the above-described applications in the production problem and transportation problem, in this paper we study an integrated optimization model of production and distribution operations which applies the uncertain programming from uncertainty theory into the joint model. Therefore, optimizing the trade-off between the production cost and the transportation cost is the goal of our systems. The problem is to find an optimal operation of production and transportation at the level of detailed scheduling such that the raw material gets used adequately and the demands of transportation are satisfied completely. As we know, when the production and transportation happen in reality, the different production plans need to arrange different transportation solutions; inversely, the different transportation demands also need the production plans to be adjusted accordingly. The joint decision problem addressed here attempts to make the total cost of production and transportation be the least.

The rest of this paper is organized as follows. In Section 2, some preliminaries from the uncertainty theory are provided for use in the next few sections, and, in Section 3, the problem formulation and some basic assumptions about this paper are given. Then, Section 4 will develop the uncertain production

model and the stochastic transportation model, respectively. Subsequently, the joint optimization model of production and transportation for a single product is obtained. In Section 5, one numerical example with multiple manufacturers and multiple destinations of demand is presented. At the end of this paper, the conclusions are proposed in Section 6.

2. Preliminaries

In this section, we will introduce some basic knowledge on uncertainty theory which will be used in the next few sections.

Definition 1 (see [17]). An uncertain variable ξ is a measurable function from an uncertainty space $(\Gamma, \mathcal{L}, \mathcal{M})$ to the set of real numbers; that is, for any Borel set B of real numbers, the set

$$\{\xi \in B\} = \{\gamma \in \Gamma \mid \xi(\gamma) \in B\} \quad (1)$$

is an event.

For a sequence of uncertain variables $\xi_1, \xi_2, \dots, \xi_n$ and a measurable function f , Liu [17] proved that

$$\xi = f(\xi_1, \xi_2, \dots, \xi_n), \quad (2)$$

defined as $\xi(\gamma) = f(\xi_1(\gamma), \xi_2(\gamma), \dots, \xi_n(\gamma))$, $\forall \gamma \in \Gamma$, is also an uncertain variable.

In order to research an uncertain variable more deeply, a concept of uncertainty distribution is given as follows.

Definition 2 (see [17]). The uncertainty distribution Φ of an uncertain variable ξ is defined by

$$\Phi(x) = \mathcal{M}\{\xi \leq x\}, \quad (3)$$

for any real number x .

To calculate the uncertain measure from an uncertainty distribution, Liu presented the measure inversion theorem.

Theorem 3 (see [17]). *Let ξ be an uncertain variable with continuous uncertainty distribution Φ . Then, for any real number x , one has*

$$\mathcal{M}\{\xi \leq x\} = \Phi(x), \quad \mathcal{M}\{\xi \geq x\} = 1 - \Phi(x). \quad (4)$$

Theorem 4 (see [17]). *Let $\xi_1, \xi_2, \dots, \xi_n$ be independent uncertain variables with regular uncertainty distributions $\Phi_1, \Phi_2, \dots, \Phi_n$, respectively. If $f(x_1, x_2, \dots, x_n)$ is strictly increasing with respect to x_1, x_2, \dots, x_m and strictly decreasing with respect to $x_{m+1}, x_{m+2}, \dots, x_n$, then*

$$\xi = f(\xi_1, \xi_2, \dots, \xi_n) \quad (5)$$

is an uncertain variable with inverse uncertainty distribution

$$\Phi^{-1}(\alpha) = f(\Phi_1^{-1}(\alpha), \dots, \Phi_m^{-1}(\alpha), \Phi_{m+1}^{-1}(1-\alpha), \dots, \Phi_n^{-1}(1-\alpha)). \quad (6)$$

Definition 5 (see [17]). Let ξ be an uncertain variable. Then the expected value of ξ is defined by

$$E[\xi] = \int_0^{\infty} \mathcal{M}\{\xi \geq r\} dr - \int_{-\infty}^0 \mathcal{M}\{\xi \leq r\} dr, \quad (7)$$

provided that at least one of the two integrals is finite.

Theorem 6 (see [17]). *Let ξ be an uncertain variable with regular uncertainty distribution Φ ; if the expected value exists, then*

$$E[\xi] = \int_0^1 \Phi^{-1}(\alpha) d\alpha. \quad (8)$$

For $\xi = f(\xi_1, \xi_2, \dots, \xi_n)$, we have

$$E[\xi] = \int_0^1 f(\Phi_1^{-1}(\alpha), \dots, \Phi_m^{-1}(\alpha), \Phi_{m+1}^{-1}(1-\alpha), \dots, \Phi_n^{-1}(1-\alpha)) d\alpha. \quad (9)$$

At the same time, Liu proved the linearity of expected value operator, that is,

$$E[a\xi + b\eta] = aE[\xi] + bE[\eta], \quad (10)$$

for any two independent uncertain variables ξ, η with finite expected values and any two real numbers a, b .

3. Problem Formulation

For a manufacturing enterprise, there are plants each producing multiple parts and multiple assemblies that serve different assembly plants in a year; meanwhile, each assembly plant demands multiple parts from many different manufacturers. Hence, the two processes of production and transportation are integrated into a combined production-transportation network consisting of multiple manufacturers and multiple demands. The basic objective for an effective supply-demand system is the integration of production and transportation functions, which needs the simultaneous consideration of cost of consumption in production and cost-savings opportunity in transportation. In production, in order to minimize production cost which contains the setup cost and processing cost of each manufacturer, they must finish the task that they are asked according to their own producing capacities. After that, the transportation will come as follows. The produced final goods are packed in a cargo, from which unit delivery cost is incurred, and are then delivered to different destinations via transportation fleets such as trucks, railroads, or aircrafts. Therefore, a reasonable transportation program based on demands must be drawn up so as to reduce unnecessary deliveries and lower delivery cost.

In this study, a single-product problem is considered for uncertain demands under the integrated production and transportation operations. The joint decision problem addressed here is, based on the delivery amounts from each manufacturer to individual destinations to meet their respective total demands at a minimum total cost in the

system, to determine production quantity to be assigned to each manufacturer. Then, an uncertain and stochastic programming model about production and transportation is developed under the assumptions that production costs and the amounts of the consumption of raw materials are uncertain variables and the delivery demands are random variables.

4. Joint Decision Model Development

4.1. Uncertain Production Model. Assume that there are sets of m manufacturers and n destinations, where each of the manufacturers and destinations is indicated by the subscripts i and j , respectively. The manufacturers produce a certain product P for the destinations, where the P is processed by kinds of l ($l = 1, 2, \dots, k$) raw materials h_1, h_2, \dots, h_k . In order to describe the production model, the following notations are applied to serve a single product:

- c_{i0} : the setup cost at manufacturer i ;
- η_i : uncertain variable, the unit variant cost for producing the product P at manufacturer i ;
- ξ_{ih_l} : uncertain variable, the amounts of the consumption of raw materials h_l ($l = 1, 2, \dots, k$) for producing per unit product P at manufacturer i ;
- φ_i : the uncertain distribution of the uncertain variable η_i ;
- Φ_{ih_l} : the uncertain distribution of the uncertain variable ξ_{ih_l} ;
- S_{h_l} : the amounts of the supply of raw materials h_l ;
- y_i : decision variable, the yield for producing the product P at manufacturer i ;
- V_i : production capacity for producing the product P at manufacturer i ;
- α_{h_l} : the predetermined confidence level.

Based on the above assumptions, not only do we minimize the production cost, but also we should make best use of the raw materials in the whole producing process. So we have proposed the uncertain production-producing model as follows:

$$\begin{aligned} \text{Min} \quad & Z_1 = \sum_{i=1}^m c_{i0} + E \left[\sum_{i=1}^m \eta_i y_i \right] \\ \text{subject to} \quad & \\ & \mathcal{M} \left\{ \sum_{i=1}^m \xi_{ih_l} y_i \geq S_{h_l} \right\} \geq \alpha_{h_l}, \quad (l = 1, 2, \dots, k) \\ & 0 \leq y_i \leq V_i, \quad (i = 1, 2, \dots, m). \end{aligned} \quad (11)$$

In this model, the objective is to minimize the expected production cost, and the major constraint is an uncertain chance constraint which will make raw materials be fully used. In order to solve this model, we will have the following corollary according to the uncertainty theory that was introduced in the preliminaries.

Corollary 7. Assume the constraint function $g(x, \xi_1, \xi_2, \dots, \xi_n)$ is strictly increasing with respect to $\xi_1, \xi_2, \dots, \xi_k$ and strictly decreasing with respect to $\xi_{k+1}, \xi_{k+2}, \dots, \xi_n$. If $\xi_1, \xi_2, \dots, \xi_n$ are independent uncertain variables with uncertainty distributions $\Phi_1, \Phi_2, \dots, \Phi_n$, respectively, then the chance constrain

$$\mathcal{M}\{g(x, \xi_1, \xi_2, \dots, \xi_n) \geq 0\} \geq \alpha \quad (12)$$

holds if and only if

$$g(x, \Phi_1^{-1}(\alpha), \dots, \Phi_k^{-1}(\alpha), \Phi_{k+1}^{-1}(1-\alpha), \dots, \Phi_n^{-1}(1-\alpha)) \geq 0. \quad (13)$$

Proof. It follows from Theorem 4 that the inverse uncertainty distribution of $g(x, \xi_1, \xi_2, \dots, \xi_n)$ is

$$\Psi^{-1}(x, \alpha) = g(x, \Phi_1^{-1}(\alpha), \dots, \Phi_k^{-1}(\alpha), \Phi_{k+1}^{-1}(1-\alpha), \dots, \Phi_n^{-1}(1-\alpha)). \quad (14)$$

Thus $\mathcal{M}\{g(x, \xi_1, \xi_2, \dots, \xi_n) \geq 0\} \geq \alpha$ holds if and only if $\Psi^{-1}(x, \alpha) \geq 0$. The theorem is thus verified. \square

Theorem 8. The uncertain chance constraint

$$\mathcal{M}\left\{\sum_{i=1}^m \xi_{ih_i} y_i \geq S_{h_i}\right\} \geq \alpha_{h_i}, \quad (15)$$

in model (11), is equivalent to

$$\sum_{i=1}^m \Phi_{ih_i}^{-1}(\alpha_{h_i}) y_i \geq S_{h_i}. \quad (16)$$

Proof. It is obvious that the function

$$g(y, \xi) = \sum_{i=1}^m \xi_{ih_i} y_i \quad (17)$$

is strictly increasing with respect to $\xi_{1h_i}, \xi_{2h_i}, \dots, \xi_{mh_i}$. And ξ_{ih_i} are independent uncertain variables with uncertainty distributions Φ_{ih_i} , respectively. Therefore

$$\mathcal{M}\left\{\sum_{i=1}^m \xi_{ih_i} y_i \geq S_{h_i}\right\} = \mathcal{M}\{g(y, \xi) - S_{h_i} \geq 0\}. \quad (18)$$

It follows from Corollary 7 that we have immediately that

$$\sum_{i=1}^m \Phi_{ih_i}^{-1}(\alpha_{h_i}) y_i - S_{h_i} \geq 0. \quad (19)$$

The theorem is proved. \square

Then, the uncertain production-producing model is expressed in the following deterministic model with the goal of minimizing the production cost:

$$\text{Min } Z_1 = \sum_{i=1}^m c_{i0} + \sum_{i=1}^m y_i \int_0^1 \varphi_i^{-1}(\alpha) d\alpha$$

subject to

$$\sum_{i=1}^m \Phi_{ih_i}^{-1}(\alpha_{h_i}) y_i \geq S_{h_i}, \quad (l = 1, 2, \dots, k)$$

$$0 \leq y_i \leq V_i, \quad (i = 1, 2, \dots, m).$$

(20)

4.2. Transportation Model with Stochastic Demands. Assume that the demands at the n destinations are stochastic and are supplied by the above-described sets of m manufacturers. Additionally, there are some parameters which will be used in the transportation problem as follows:

c_{ij} : the transportation cost of delivering unit value of product from manufacturer i to destination j ;

x_{ij} : decision variable, the number of units shipped from manufacturer i to destination j ;

b_j : independent random variable, the demand at destination j which follows the normal distribution $N(\mu_j, \sigma_j^2)$;

β_j : the predetermined confidence level ($j = 1, 2, \dots, n$);

$\Phi(\theta)$: the standard normal distribution function.

Thus, on the basis of the typical transportation problem, the transportation model with stochastic demands is formulated into the following form:

$$\text{Min } Z_2 = \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij}$$

subject to

$$\text{Pr}\left\{\sum_{i=1}^m x_{ij} \geq b_j\right\} \geq \beta_j, \quad (j = 1, 2, \dots, n) \quad (21)$$

$$\sum_{j=1}^n x_{ij} \leq V_i, \quad (i = 1, 2, \dots, m)$$

$$x_{ij} \geq 0, \quad (i = 1, 2, \dots, m, j = 1, 2, \dots, n).$$

Based on the possibility theory, we also get the theorem below.

Theorem 9. Assume that b_j ($j = 1, 2, \dots, n$) are random variables which follow the normal distribution of $N(\mu_j, \sigma_j^2)$. Then, the chance constraint

$$\text{Pr}\left\{\sum_{i=1}^m x_{ij} \geq b_j\right\} \geq \beta_j \quad (22)$$

in model (21) is equivalent to

$$\sum_{i=1}^m x_{ij} \geq \mu_j + \sigma_j \Phi^{-1}(\beta_j). \quad (23)$$

Proof. Since b_j are random variables, $y(x) = \sum_{i=1}^m x_{ij} - b_j$ are also random variables. Then we have

$$E(y(x)) = \sum_{i=1}^m x_{ij} - \mu_j, \quad V(y(x)) = \sigma_j^2, \quad (24)$$

where $E(\cdot)$ and $V(\cdot)$ denote expected value and variance, respectively.

On the other hand, since the inequalities $\sum_{i=1}^m x_{ij} \geq b_j$ are equivalent to

$$\frac{\sum_{i=1}^m x_{ij} - b_j - [\sum_{i=1}^m x_{ij} - E(b_j)]}{\sigma(b_j)} \leq \frac{\sum_{i=1}^m x_{ij} - E(b_j)}{\sigma(b_j)},$$

$$\eta = -\frac{\sum_{i=1}^m x_{ij} - b_j - [\sum_{i=1}^m x_{ij} - E(b_j)]}{\sigma(b_j)}, \quad (25)$$

are random variables which follow standard normal distribution of $N(0, 1)$ with possibility distribution function $\Phi(\eta) = (1/\sqrt{2\pi}) \int_{-\infty}^{\eta} e^{-t^2/2} dt$. Thus, we have

$$\Pr \left\{ \eta \leq \frac{\sum_{i=1}^m x_{ij} - E(b_j)}{\sigma(b_j)} \right\} \geq \beta_j. \quad (26)$$

According to the definition of possibility distribution function, we also have

$$\Phi \left\{ \frac{\sum_{i=1}^m x_{ij} - E(b_j)}{\sigma(b_j)} \right\} \geq \beta_j. \quad (27)$$

So, we have

$$\sum_{i=1}^m x_{ij} \geq \mu_j + \sigma_j \Phi^{-1}(\beta_j). \quad (28)$$

The theorem is proved. \square

According to Theorem 9, the improved transportation model with stochastic demands will be transferred into the following model:

$$\text{Min } Z_2 = \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij}$$

subject to

$$\sum_{i=1}^m x_{ij} \geq \mu_j + \sigma_j \Phi^{-1}(\beta_j), \quad (j = 1, 2, \dots, n)$$

$$\sum_{j=1}^n x_{ij} \leq V_i, \quad (i = 1, 2, \dots, m)$$

$$x_{ij} \geq 0, \quad (i = 1, 2, \dots, m, j = 1, 2, \dots, n). \quad (29)$$

4.3. The Joint Optimization Model of Production and Transportation. In actual life, the production and transportation are usually combining closely. So in order to reduce the total cost of the production and transportation, we have to consider the two great problems comprehensively rather than one by one. Firstly, we must organize the production effectively and arrange the transportation reasonably for reaching the aim of saving the cost. Thus, we could get the number of units that every manufacturer produces to be transported completely. That is,

$$y_i = \sum_{j=1}^n x_{ij}. \quad (30)$$

Upon this, we can effectively adjust the production of the product according to the changes of the demands and then reach the goal of reducing the cost of each of the counterparts and optimizing the production-transportation network.

Then, combining models (11) and (21), we can get the joint optimization model of production and transportation under uncertain environment:

$$\text{Min } Z = Z_1 + Z_2 = \sum_{i=1}^m c_{i0} + E \left[\sum_{i=1}^m \eta_i \gamma_i \right] + \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij}$$

subject to

$$\mathcal{M} \left\{ \sum_{i=1}^m \xi_{ih_l} \gamma_i \geq S_{h_l} \right\} \geq \alpha_{h_l}, \quad (l = 1, 2, \dots, k)$$

$$\Pr \left\{ \sum_{i=1}^m x_{ij} \geq b_j \right\} \geq \beta_j, \quad (j = 1, 2, \dots, n)$$

$$0 \leq y_i = \sum_{j=1}^n x_{ij} \leq V_i, \quad (i = 1, 2, \dots, m)$$

$$x_{ij} \geq 0, \quad (i = 1, 2, \dots, m, j = 1, 2, \dots, n). \quad (31)$$

Combining models (20) and (29), we can get the equivalent model of model (31) immediately:

$$\text{Min } Z = \sum_{i=1}^m c_{i0} + \sum_{i=1}^m \gamma_i \int_0^1 \varphi_i^{-1}(\alpha) d\alpha + \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij}$$

subject to

$$\sum_{i=1}^m \Phi_{ih_l}^{-1}(\alpha_{h_l}) \gamma_i \geq S_{h_l}, \quad (l = 1, 2, \dots, k)$$

$$\sum_{i=1}^m x_{ij} \geq \mu_j + \sigma_j \Phi^{-1}(\beta_j), \quad (j = 1, 2, \dots, n)$$

$$0 \leq y_i = \sum_{j=1}^n x_{ij} \leq V_i, \quad (i = 1, 2, \dots, m)$$

$$x_{ij} \geq 0, \quad (i = 1, 2, \dots, m, j = 1, 2, \dots, n). \quad (32)$$

This model is a typical linear programming which can be solved easily by the simplex method.

5. Numerical Experiment

To evaluate and test the performance of the above uncertain and stochastic model, this section presents an example with three manufacturers and four destinations to illustrate the application of the model.

Assume that the production of P , which is made up of three raw materials h_1, h_2, h_3 , is assigned to be produced by three manufacturers A_1, A_2, A_3 , and then the production of P will be delivered to four destinations B_1, B_2, B_3, B_4 . According to the expertise experience, the amounts of the consumption of raw materials h_1, h_2, h_3 for producing per unit product P follow linear uncertain distribution $L(a_{ih_i}, b_{jh_i}), i = 1, 2, 3, j = 1, 2, 3$, respectively. The unit variant cost for producing the product P and the demand at destination j follow the zigzag uncertain distribution $Z(a_i, b_i, c_i), i = 1, 2, 3$, and the normal random distribution $N(\mu_j, \sigma_j^2), j = 1, 2, 3, 4$, respectively. Tables 1, 2, and 3 give the value of the parameters and others.

It follows from model (31) that we have the following model about this problem:

$$\text{Min } Z = \sum_{i=1}^3 c_{i0} + E \left[\sum_{i=1}^3 \eta_i y_i \right] + \sum_{i=1}^3 \sum_{j=1}^4 c_{ij} x_{ij}$$

subject to

$$\mathcal{M} \left\{ \sum_{i=1}^3 \xi_{ih_i} y_i \geq S_{h_i} \right\} \geq \alpha_{h_i}, \quad (l = 1, 2, 3) \quad (33)$$

$$\text{Pr} \left\{ \sum_{i=1}^3 x_{ij} \geq b_j \right\} \geq \beta_j, \quad (j = 1, 2, 3, 4)$$

$$0 \leq y_i = \sum_{j=1}^4 x_{ij} \leq V_i, \quad (i = 1, 2, 3)$$

$$x_{ij} \geq 0, \quad (i = 1, 2, 3, j = 1, 2, 3, 4).$$

Note that the confidence levels are $\alpha_{h_l} = 0.9, l = 1, 2, 3$, $\beta_j = 0.95$, and $j = 1, 2, 3, 4$. Since the inverse uncertainty distribution of linear uncertain variable $L(a, b)$ is

$$\Phi^{-1}(\alpha) = (1 - \alpha)a + \alpha b, \quad (34)$$

the inverse uncertainty distribution of zigzag uncertain variable $Z(a, b, c)$ is

$$\Phi^{-1}(\alpha) = (2 - 2\alpha)b + (2\alpha - 1)c, \quad \text{if } \alpha \geq 0.5. \quad (35)$$

TABLE 1: Parameters of linear distribution of the consumption and the supply of the raw materials.

	h_1	h_2	h_3
A_1	$L(2, 4)$	$L(3, 5)$	$L(4, 6)$
A_2	$L(0, 2)$	$L(1, 3)$	$L(2, 4)$
A_3	$L(1, 3)$	$L(2, 4)$	$L(3, 5)$
S_{h_i}	60	80	100

TABLE 2: Parameters of the setup cost, the capacity, and the unit variant cost at the manufacturers.

	c_{i0}	V_i	η_i
A_1	35	15	$Z(90, 100, 110)$
A_2	30	25	$Z(50, 60, 70)$
A_3	35	10	$Z(70, 80, 90)$

TABLE 3: Parameters of the unit delivery cost and the normal distribution of demands at the destinations.

	B_1	B_2	B_3	B_4
A_1	7	5	12	9
A_2	6	3	6	5
A_3	9	5	8	5
(μ_j, σ_j^2)	$(5, 2^2)$	$(10, 2^2)$	$(8, 2^2)$	$(6, 2^2)$

For the standard normal distribution, we have that $\Phi^{-1}(0.95) = 1.645$. Thus, model (33) has the form as follows:

$$\text{Min } Z = \sum_{i=1}^3 c_{i0} + \sum_{i=1}^3 y_i \int_0^1 \varphi_i^{-1}(\alpha) d\alpha + \sum_{i=1}^3 \sum_{j=1}^4 c_{ij} x_{ij}$$

subject to

$$\sum_{i=1}^3 (0.1a_{ih_i} + 0.9b_{ih_i}) y_i \geq S_{h_i}, \quad (l = 1, 2, 3) \quad (36)$$

$$\sum_{i=1}^3 x_{ij} \geq \mu_j + 1.645\sigma_j, \quad (j = 1, 2, 3)$$

$$0 \leq y_i = \sum_{j=1}^4 x_{ij} \leq V_i, \quad (i = 1, 2, 3)$$

$$x_{ij} \geq 0, \quad (i = 1, 2, 3, j = 1, 2, 3, 4).$$

By means of the simplex method, the optimal production plan, transportation scheme, and the demands are listed in Table 4.

The production cost is 302.0000, the transportation cost is 21.2600, and the total cost is 333.2600.

6. Conclusions

This paper considers the whole process of production and transportation and establishes the joint optimal model for a single product with multiple manufacturers and multiple destinations. By introducing uncertainty theory, we presented

TABLE 4: The results of the transportation scheme, the yields of the manufacturers, and the demands of the destinations.

x_{ij}	B_1	B_2	B_3	B_4	y_i
A_1	7.2000	0.0000	0.0000	0.0000	7.2000
A_2	1.1000	12.6000	11.3000	0.0000	25.0000
A_3	0.0000	0.7000	0.0000	9.3000	10.0000
Demands	8.3000	13.3000	11.3000	9.3000	

uncertain production model, and in the transportation process, we assumed that the demands are random variables and proposed the stochastic transportation model. Based on these above, the uncertain and stochastic model was developed by taking expected value on objective function and confidence level on the constraint functions. In the end of this paper, a numerical example was given and showed the effectiveness of the method by which we establish the model. Then, the production plan and transportation scheme were obtained by the simplex method.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This work has been financed by the Fundamental Research Funds for the Central Universities (no. 3122013K008).

References

- [1] Glover, F. G. Jones, D. Karney, D. Klingman, and J. Mote, "An integrated production, distribution, and inventory planning system," *Interfaces*, vol. 9, pp. 21–35, 1979.
- [2] D. E. Blumenfeld, L. D. Burns, and C. F. Daganzo, "Synchronizing production and transportation schedules," *Transportation Research B: Methodological*, vol. 25, no. 1, pp. 23–37, 1991.
- [3] Z. L. Chen and G. L. Vairaktarakis, "Integrated scheduling of production and distribution operations," *Management Science*, vol. 51, no. 4, pp. 614–628, 2005.
- [4] B. Scholz-Reiter, E. M. Frazzon, and T. Makuschewitz, "Integrating manufacturing and logistic systems along global supply chains," *CIRP Journal of Manufacturing Science and Technology*, vol. 2, no. 3, pp. 216–223, 2010.
- [5] E. M. Frazzon, J. Pintarelli, T. Makuschewitz, and B. Scholz-Reiter, "Simulation-based analysis of integrated production and transport logistics," in *Dynamics in Logistics*, Lecture Notes in Logistics, pp. 499–508, Springer, Berlin, Germany, 2013.
- [6] M. A. Cohen and H. L. Lee, "Strategic analysis of integrated production-distribution systems: models and methods," *Operations Research*, vol. 36, no. 2, pp. 216–228, 1988.
- [7] R. M. Hill, "The single-vendor single-buyer integrated production-inventory model with a generalised policy," *European Journal of Operational Research*, vol. 97, no. 3, pp. 493–499, 1997.
- [8] S. K. Goyal, "On improving the single-vendor single-buyer integrated production inventory model with a generalized policy," *European Journal of Operational Research*, vol. 125, no. 2, pp. 429–430, 2000.
- [9] K. L. Yung, J. Tang, A. W. H. Ip, and D. Wang, "Heuristics for joint decisions in production, transportation, and order quantity," *Transportation Science*, vol. 40, no. 1, pp. 99–116, 2006.
- [10] K. Holmberg and H. Tuy, "A production-transportation problem with stochastic demand and concave production costs," *Mathematical Programming*, vol. 85, no. 1, pp. 157–179, 1999.
- [11] T.-F. Liang, "Integrated manufacturing/distribution planning decisions with multiple imprecise goals in an uncertain environment," *Quality and Quantity*, vol. 46, no. 1, pp. 137–153, 2012.
- [12] D. Kahnema and A. Tversky, "Prospect theory: an analysis of decision under risk," *Econometrica*, vol. 47, pp. 263–292, 1979.
- [13] B. Liu, *Uncertainty Theory*, Springer, Berlin, Germany, 2nd edition, 2007.
- [14] B. Liu, "Some research problems in uncertainty theory," *Journal of Uncertain Systems*, vol. 3, no. 1, pp. 3–10, 2009.
- [15] B. Liu, *Theory and Practice of Uncertain Programming*, Springer, Berlin, Germany, 2nd edition, 2009.
- [16] B. Liu, *Uncertainty Theory: A Branch of Mathematics for Modeling Human Uncertainty*, Springer, Berlin, Germany, 2010.
- [17] B. Liu, *Uncertainty Theory*, Springer, Berlin, Germany, 4th edition, 2013.
- [18] Y. Liu and M. Ha, "Expected value of function of uncertain variables," *Journal of Uncertain Systems*, vol. 4, no. 3, pp. 181–186, 2010.
- [19] L. Rong, "Two new uncertainty programming models of inventory with uncertain costs," *Journal of Information and Computational Science*, vol. 8, no. 2, pp. 280–288, 2011.
- [20] Y. H. Sheng and K. Yao, "Fixed charge transportation problem and its uncertain programming model," *Industrial Engineering and Management Systems*, vol. 11, no. 2, pp. 183–187, 2012.
- [21] D. Y. Mou, W. L. Zhao, and X. D. Chang, "A transportation problem with uncertain truck times and unit costs," *Industrial Engineering and Management Systems*, vol. 12, no. 1, pp. 30–35, 2013.
- [22] Y. Gao, "Uncertain models for single facility location problems on networks," *Applied Mathematical Modelling*, vol. 36, no. 6, pp. 2592–2599, 2012.
- [23] G. Meng and X. Zhang, "Optimization uncertain measure model for uncertain vehicle routing problem," *Information*, vol. 16, no. 2, pp. 1201–1206, 2013.
- [24] X. Wang, Z. Gao, and H. Guo, "Uncertain hypothesis testing for two experts' empirical data," *Mathematical and Computer Modelling*, vol. 55, no. 3-4, pp. 1478–1482, 2012.
- [25] Y. Zhu, "Uncertain optimal control with application to a portfolio selection model," *Cybernetics and Systems*, vol. 41, no. 7, pp. 535–547, 2010.

Research Article

Application of Fuzzy Optimization to Production-Distribution Planning in Supply Chain Management

S. Ariafar,^{1,2} S. Ahmed,¹ I. A. Choudhury,¹ and M. A. Bakar¹

¹ *Manufacturing System Integration (MSI), Department of Mechanical Engineering, Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia*

² *Department of Industrial Engineering, Faculty of Engineering, Shahid Bahonar University of Kerman, Kerman 7618891167, Iran*

Correspondence should be addressed to S. Ahmed; ahmed@um.edu.my

Received 31 January 2014; Accepted 7 April 2014; Published 28 May 2014

Academic Editor: Michael Lütjen

Copyright © 2014 S. Ariafar et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A production-distribution model has been developed that not only allocates the limited available resources and equipment to produce the products over the time periods, but also determines the economical distributors for dispatching the products to the distribution centers or retailers. The model minimizes production, inventory holding, backordering, and transportation cost while considering the time value of money. Since uncertainty is an inevitable issue of any real-world production system, then to provide a realistic model, the concept of fuzzy sets has been applied in the proposed mathematical modeling. To illustrate and show the feasibility and validity of the model, a real case analysis, which is pertaining to a mineral water bottling production factory, has been used. The case has been solved using a three-step solution approach developed in this study. The results show the feasibility and validity of the mathematical model, and also the solution procedure.

1. Introduction

Nowadays, with the globalization and the evolution of newer global marketplace over the time, manufacturing companies have been forced to find methods to design and operate efficient supply chains to meet customer demands and maximize the profit of uncertain business environment [1, 2]. A typical supply chain may consist of a number of suppliers and a plant, where processing of the relevant materials adds value to what have been received from suppliers and delivers the products to one or more distributors to dispatch them to several distribution centers (DC) or retailers. In such a supply chain, transportation cost may vary from one distributor to another due to variations in skills, labor cost, or different vehicle types, and equipment that is being used to do the transport. Hence, a significant interest has arisen in production-distribution planning decision (PDPD) of supply chains.

Production planning is a process, which involves allocating available equipment and resources over a time period to perform a series of required tasks in order to manufacture finished products, according to a specified schedule [3, 4], while distribution planning includes methods for determining the ways to get the materials and products from the delivery points to the consuming points in supply chains [5]. In general, optimization of supply chain decision in isolation keeps the firms away from gaining maximum possible effectiveness [6–8]. Therefore, great efforts have been made to optimize production and distribution planning problems simultaneously. Several researchers have investigated the advantages of integrated production-distribution planning in different manufacturing environments. For instance, Park indicated the effectiveness of integrating production and distribution planning in multiplant, multi-item, multiperiod, and multiretailer environment [9]. Safaei et al. demonstrated

the overall cost reduction resulting from the integration in a multiproduct, multisite, multiperiod production-distribution network [10]. Amorim et al. studied the effects of concurrent optimization of production and distribution decisions for perishable products [11].

A comprehensive review on integrated production-distribution models and techniques is conducted by Fahimnia et al. According to the review, some prominent characteristics that influence the complexity of models are as follows: single stage or multistage supply network, single or multiple objective functions, single or multiple products, single or multiple plants or distribution centers, and single or multiperiod [5]. This section is not intended to review or classify the literature but to direct readers to some of the most relevant and most recent studies in the area of integrated production and distribution planning. Reviewing these and other related studies reveals that most of the integrated production-distribution models in the literature are deterministic. However, one of the realities of current production systems is that the goals and other relevant inputs such as available resources, market demand, and production rates are not deterministic. Hence, conventional mathematical models that assume the input parameters are deterministic and crisp cannot handle the uncertainty of real manufacturing systems. In such a situation, fuzzy and also stochastic mathematical modeling can be applied to cope with the decision making under uncertainties of manufacturing environments [12–14].

Most recently, fuzzy programming as one of the methods that is able to take into account uncertainty of manufacturing systems has been applied in integrated production-distribution planning problems. Lin and Liang developed a fuzzy multiobjective linear programming (FMOLP) model for an aggregate production planning problem [15]. Then, Wang and Liang developed a possibilistic linear programming (PLP) approach to solve a multiproduct and multiperiod aggregate production planning [16]. Aliev et al. developed a multiproduct and multiperiod mathematical model in a supply chain to integrate the production and distribution decisions simultaneously. They developed a genetic algorithm-based method to solve the model [17]. However, although there are more studies that have considered fuzzy aggregate production planning in a supply chain [6, 18, 19], to the best of the authors' knowledge there is not any previous study that considers the production planning and distributor selection in a supply chain simultaneously.

The aim of this study is to develop a production-distribution mathematical model that not only determines the production planning of the system, but also selects the best contractors for dispatching the products to the distribution centers or retailers. Moreover, since some data from real-world manufacturing environments are unobtainable or imprecise, in order to provide a more realistic mathematical model, fuzzy set theory has been applied in this study. The remainder of this paper is organized as follows. Section 2 is dedicated to the problem description. In Section 3, the mathematical model will be presented. Then, Section 4 provides a discussion regarding the solution procedure. In Section 5, a real case from a mineral water bottling factory will be illustrated to show the feasibility and validity of the proposed

mathematical model. Finally, in Section 6 this study will be concluded.

2. Problem Description

This study assumes that there is a mineral water bottling factory that produces different kinds of bottled water. The main product of this factory, which is 1.5-liter bottles of mineral water, is dispatched to several distribution centers by several companies as distributors to satisfy imprecise demand over a medium planning horizon. On the other hand, the factory capacity and also labor levels are imprecise and fuzzy due to unobtainable or incomplete available records. This study aims to develop a fuzzy mathematical model that not only determines the optimal production plan for the system, but also selects the economical contract that provides the minimum transportation cost for dispatching the products to the distribution centers of the supply chain in an uncertain environment.

The assumptions of the proposed fuzzy mathematical model are as follows.

- (1) Triangular distribution pattern is applied to represent all fuzzy numbers.
- (2) The objective function is considered to be fuzzy with imprecise aspiration level.
- (3) The nonincreasing continuous linear membership function is used to specify the decision maker's satisfaction degree.
- (4) The minimum operator is applied to aggregate the fuzzy sets.
- (5) The demand of each distribution center over a special time period might be satisfied or backordered, although the backorders have to be fulfilled next period until the end of the planning horizon.
- (6) Each distribution center for each time period has a minimum requirement of supply that should be satisfied.

Assumption 1 addresses the application of triangular fuzzy numbers to represent imprecise data. Simplicity and effectiveness of triangular fuzzy numbers enhance the computational efficiency and also facilitate data acquisition [19–22]. Assumption 2 is to state the fuzziness of the objective function. Assumption 3 is used to mention that according to the satisfaction degree of the decision maker (DM) linear membership function in comparison with other membership functions is preferred. Assumption 4 is to state that the concept of fuzzy decision-making of Bellman and Zadeh, together with the minimum operator of Zimmermann is used to aggregate all fuzzy sets in this study [23, 24]. Assumption 5 shows that a portion of demand is allowed to be backordered, but it has to be fulfilled in the next time periods of the running of manufacturing system. Finally, assumption 6 is to ensure that each distribution center in each period at least will receive a minimum amount of products to supply to the retailers.

3. Mathematical Model Development

The main aim of this study is to provide a fuzzy mathematical model that not only determines the aggregate production plans in a midterm period, but also selects the economical contracts that dispatch the products of supply chain to distribution centers or retailers. The notations that are used in the proposed mathematical model are as follows.

3.1. Index Sets

- j : Index for destinations $j = 1, \dots, J$
- n : Index for time periods $n = 1, \dots, N$
- k : Index for distributors $k = 1, \dots, K$.

3.2. Decision Variables

- Q_n : Production volume in period n (boxes)
- I_n : Inventory level in period n (boxes)
- B_n : Backorder level in period n (boxes)
- T_{jkn} : Number of transported products to destination j by distributor k in period n (boxes)

$$Y_k = \begin{cases} 1 & \text{If distributor } k \text{ dispatches the products to the} \\ & \text{DCs or retailers.} \\ 0 & \text{Otherwise.} \end{cases} \quad (1)$$

3.3. Parameters

- CP_n : Production cost per box in period n (\$/box)
- CI_n : Inventory-holding cost per box in period n (\$/box)
- CB_n : Backordering cost per box in period n (\$/box)
- CT_{jkn} : Per-box transportation cost to destination j in period n by distributor k (\$/box)
- FC_{jkn} : Fixed transportation cost to destination j in period n by distributor k (\$)
- D_n : Demand of product in period n (box)
- R_{jn} : Minimum supply of product for destination j in period n (box)
- MARR: Minimum attractive rate of return as the escalating factor for the costs (%)
- U : Required machine-hour to produce one product (machine-hour/box)
- F_n : Maximum factory capacity in period n (machine-hour)
- V : Required man-hour to produce one product (man-hour/box)

M_n : Maximum manpower available in period n (man-hour)

S : Required warehouse volume per box of product (m^2 /box)

AW_n : Maximum available warehouse volume in period n (m^2).

3.4. Fuzzy Mathematical Model

$$\begin{aligned} \text{Min } z \cong & \sum_{n=1}^N \left(CP_n * Q_n + CI_n * I_n + CB_n * B_n \right. \\ & \left. + \sum_{k=1}^K \sum_{j=1}^J Y_k \cdot (CT_{jkn} * T_{jkn} + FC_{jkn}) \right) \quad (2) \\ & * (1 + \text{MARR})^n \end{aligned}$$

subject to

$$I_{n-1} + Q_n - I_n = \sum_{k=1}^K \sum_{j=1}^J (Y_k * T_{jkn}) \quad \forall n, \quad (3)$$

$$I_{n-1} - B_{n-1} + Q_n - I_n + B_n = \tilde{D}_n \quad \forall n, \quad (4)$$

$$Q_n \geq B_{n-1} \quad \forall n, \quad (5)$$

$$\sum_{k=1}^K (Y_k * T_{jkn}) \geq R_{jn} \quad \forall j, \forall n, \quad (6)$$

$$U * Q_n \leq \tilde{F}_n \quad \forall n, \quad (7)$$

$$V * Q_n \leq \tilde{M}_n \quad \forall n, \quad (8)$$

$$\sum_{k=1}^K \sum_{j=1}^J Y_k \cdot (S * T_{jkn}) \leq AW_n \quad \forall n, \quad (9)$$

$$\sum_{k=1}^K Y_k = 1, \quad (10)$$

$$Q_n, O_n, I_n, B_n, T_{jkn} \geq 0 \quad \forall n, Y_k = \{0, 1\}. \quad (11)$$

The first part of the fuzzy mathematical model, in (2), is the objective function. It minimizes the total cost of the system which includes production, inventory, backorder, and transportation cost. Equation (3) indicates that the amounts of products that can be transported in each time period are those products that have been produced in the same period plus the amount of products that have been stocked from the previous period minus the amount of products that will be maintained in the warehouse for the next time period. Based on the fifth assumption of this study, which declares that the demand for each time period might be satisfied or backordered, (4) has been developed. Moreover, based on the same assumption, (5) implies that the amount of products, which are produced in each period, at least should satisfy

the backorder of the previous time period. Equation (6) is based on the sixth assumption. It tries to satisfy the minimum supply requirement for each distribution center in every time period. Equations (7), (8), and (9) are orderly related to the normal limitation in the capacity of machine-hour, man-hour, and the available warehouse volume of the factory. Equation (10) shows that the decision maker only looks for one contractor to dispatch the products to distribution centers or retailer. Finally, set of (11) shows the nonnegativity and kinds of decision variables that have been used in the proposed fuzzy mathematical model.

4. Solution Procedure

The fuzzy mathematical model that has been developed in this study is flexible in the value of the objective function and also has vagueness in some constraints. Hence, the fuzzy model has to be converted into an equivalent crisp one in order to be solved by ordinary methods. For this purpose, a three-step procedure has been proposed as following.

- (1) In this study, it has been assumed that a triangular distribution pattern is adopted for all the fuzzy data. Hence, based on the level of α -cut, all the imprecise constraints are converted into crisp ones using the weighted averaging method.
- (2) The fuzzy objective function, based on the decision's maker satisfaction degree, is treated.
- (3) The auxiliary variable L is introduced at first. After that the original mathematical model is transformed into an equivalent ordinary mathematical model by using minimum operator. Then, the ordinary linear programming (LP) model is solved by conventional methods.

4.1. Treatment of Fuzzy Constraints. In solving fuzzy mathematical models, if there is any vagueness in the constraints, it should be treated. In this study, the volume of demand, the factory capacity level, and also the level of manpower available in each of the planning periods, because of the vague nature of these data, have been considered to be fuzzy with triangular fuzzy distribution. The reason of using triangular fuzzy numbers for the data is its flexibility and simplicity in performing the fuzzy arithmetic operations [19, 22]. These fuzzy numbers, as it can be seen from the proposed mathematical model, are devoted to the right-hand side of the constraints (4), (7), and (8). In this study, in order to convert these fuzzy numbers into crisp ones, the weighted average method has been used [19, 25]. This method, in addition to the level of α -cut, which is the minimal acceptable possibility level of occurrence for the data, uses the three prominent components of every triangular distribution pattern that are most pessimistic (p), most likely (m), and most optimistic value (o) for the set of available data. Figure 1 shows the distribution of demand (D) in each period, which has the triangular distribution pattern.

In this study, in order to convert the fuzzy demand into a crisp number, the weighted average method has been used.

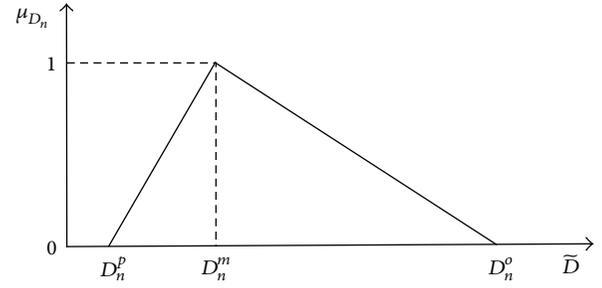


FIGURE 1: The triangular distribution of fuzzy number \bar{D}_n .

After defuzzification, the corresponding expression for the fuzzy equation (4) is as follows:

$$\begin{aligned} I_{n-1} - B_{n-1} + Q_n - I_n + B_n \\ = w_1 \cdot D_{n,\alpha}^p + w_2 \cdot D_{n,\alpha}^m + w_3 \cdot D_{n,\alpha}^o \quad \forall n. \end{aligned} \quad (12)$$

In (12), w_1 , w_2 , and w_3 are the corresponding weights for the most pessimistic, the most likely, and the most optimistic values for demand, while the summation of these three weights is equal to one ($w_1 + w_2 + w_3 = 1$). In addition, based on the concept of the most likely values, which has been proposed by Lai and Hwang, this study has set w_1 , w_2 , and w_3 as ($w_2 = 4/6$) and ($w_1 = w_3 = 1/6$) [26]. Similarly, two other fuzzy constraints of (7) and (8) can be converted into crisp expressions by using the weighted average method as follows:

$$\begin{aligned} U * Q_n &\leq w_1 \cdot F_{n,\alpha}^p + w_2 \cdot F_{n,\alpha}^m + w_3 \cdot F_{n,\alpha}^o \quad \forall n, \\ V * Q_n &\leq w_1 \cdot M_{n,\alpha}^p + w_2 \cdot M_{n,\alpha}^m + w_3 \cdot M_{n,\alpha}^o \quad \forall n. \end{aligned} \quad (13)$$

4.2. Treating the Objective Function. When the objective function of a mathematical model is fuzzy, it should be transformed to an auxiliary crisp one in order to be proceeded for solving the model. The most important part of this treatment is to find an appropriate membership function for this conversion. In this study, it has been assumed that according to the satisfaction degree of the decision maker, linear membership function is the most suitable function for this transformation. The process of defuzzification in this study is based on the concept of fuzzy decision making of Bellman and Zadeh together with the fuzzy programming method of Zimmermann [23, 24]. In this regard, at first the negative ideal solution (NIS) and positive ideal solution (PIS) for the fuzzy objective function will be introduced as follows:

$$Z_{\text{PIS}} = \text{Min}z \quad Z_{\text{NIS}} = \text{Max}z. \quad (14)$$

Then, the membership function, which according to the satisfaction degree of the decision maker is supposed to be nonincreasing continuous linear, will be defined for the fuzzy

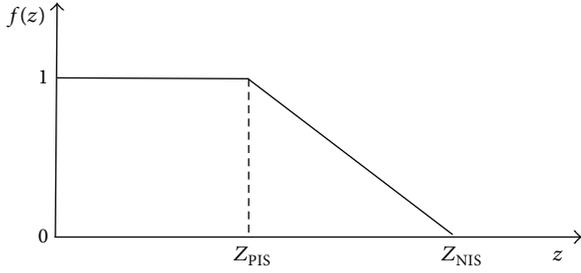


FIGURE 2: A schematic for a nonincreasing continuous membership function.

objective function. The formulation that has been used in this study for treating the objective function is as follows:

$$f(z) = \begin{cases} 1 & Z \leq Z_{PIS} \\ \frac{Z_{NIS} - Z}{Z_{NIS} - Z_{PIS}} & Z_{PIS} < Z < Z_{NIS} \\ 0 & Z_{NIS} \leq Z. \end{cases} \quad (15)$$

Figure 2 depicts a schematic diagram of a nonincreasing linear membership function.

4.3. *Developing an Auxiliary Mathematical Model.* In solving fuzzy programming models, after the defuzzification of constraints and objective functions, in order to solve the fuzzy mathematical model by ordinary methods, it should be converted into an auxiliary crisp model. In this study, the minimum operator has been used. The auxiliary mathematical model of this study can be constructed as follows:

$$\begin{aligned} & \text{Max } L \\ & \text{s.t. } L \leq f(z), \\ & \text{Equations (3), ((5) - (6)), ((9) - (10)), ((12) - (13)),} \\ & Q_n, O_n, I_n, B_n, T_{jkn}, \quad L \geq 0 \quad \forall n, \quad Y_k = \{0, 1\}. \end{aligned} \quad (16)$$

5. Model Implementation

In this section, the developed fuzzy mathematical model will be applied for a mineral water bottling factory that is located in Bardsir, Kerman, Iran. The main product of this factory is 1.5-liter bottles of mineral water that are distributed to three distribution centers which are located in Kerman, Sirjan, and Rafsanjan. The production planner of the factory has to provide an overall perspective regarding the amount of production, inventory, and also the backorder of their main product for a midterm scheduling period, as an aggregate production plan. Moreover, this factory may select among two alternative contractors for dispatching their products to three mentioned distribution centers and each contractor utilizes different types of vehicles and equipment. Therefore, the transportation costs and consequently transportation fee that are offered by these two companies are different. Hence,

TABLE 1: Imprecise forecasted data.

Period	Demand (D) (box)	Factory capacity (F) (machine-hour)	Manpower available (M) (man-hour)
June	(11500, 12400, 14500)	(177, 193, 197)	(1130, 1155, 1180)
July	(10500, 12000, 13500)	(175, 185, 195)	(970, 980, 1110)
August	(12000, 12500, 15400)	(182, 196, 198)	(1170, 1180, 1190)

TABLE 2: Minimum supply for each destination (box).

Period	Rafsanjan	Sirjan	Kerman
June	2000	2500	5000
July	1500	2000	6000
August	1000	2000	6500

TABLE 3: Costs data (\$/box).

Period	Production cost	Inventory-holding cost	Backordering cost
June	1.60	0.10	0.50
July	1.80	0.08	0.40
August	1.25	0.09	0.35

TABLE 4: Per-box transportation cost for each destination (\$/box).

Period	Rafsanjan		Sirjan		Kerman	
	$k = 1$	$k = 2$	$k = 1$	$k = 2$	$k = 1$	$k = 2$
June	0.094	0.112	0.106	0.142	0.064	0.082
July	0.096	0.118	0.102	0.144	0.060	0.080
August	0.092	0.112	0.105	0.156	0.058	0.088

TABLE 5: Fixed transportation cost for each destination (\$).

Period	Rafsanjan		Sirjan		Kerman	
	$k = 1$	$k = 2$	$k = 1$	$k = 2$	$k = 1$	$k = 2$
June	110	50	180	80	100	30
July	120	55	175	75	80	25
August	130	60	160	70	90	35

the production planner also has to select the best distributor for dispatching their product based on the transportation cost.

5.1. *Data Description.* The data related to the water bottling factory is as follows.

- (i) The main product of the factory is 1.5-liter bottles of mineral water, and each dozen of bottles is packed in one box.
- (ii) The planning horizon of the factory is three months, June, July, and August.

TABLE 6: Results for the mineral water bottling production company (box).

Period	Q	I	B	Rafsanjan		Sirjan		Kerman	
				k = 1	k = 2	k = 1	k = 2	k = 1	k = 2
June	12800	800	0	2000	—	2500	—	8000	—
July	10900	0	300	1500	—	2000	—	8200	—
August	13000	0	0	1000	—	2000	—	10000	—
Objective function				L = 100%				Z = 70170.19	

TABLE 7: Sensitivity analysis for the α -cut level.

α level	Period	Q	I	B	Rafsanjan		Sirjan		Kerman	
					k = 1	k = 2	k = 1	k = 2	k = 1	k = 2
$\alpha = 0.3$	June	12786	766	0	2000	—	2500	—	8020	—
	July	10988	0	246	1500	—	2000	—	8254	—
	August	12986	0	0	1000	—	2000	—	9986	—
Objective function				L = 100%				Z = 70283.98		
$\alpha = 0.7$	June	12826	866	0	2000	—	2500	—	7960	—
	July	10728	0	406	1500	—	2000	—	8094	—
	August	13026	0	0	1000	—	2000	—	10026	—
Objective function				L = 100%				Z = 69943.24		

- (iii) The boxes of 1.5-liter bottles of mineral water are distributed to three distribution centers.
- (iv) There are two candidates as the distributors for the products of the factory.
- (v) The initial inventory for the main product of the factory is 500 boxes while the inventory for the end period (August) is zero.
- (vi) The initial backorder and also the end backordering volume in the third period (August) are equal to zero.
- (vii) According to the previous time studies, production of each box of the main product utilizes 0.015 machine-hours of the capacity of the factory.
- (viii) The required manpower to produce one box of the product equals 0.09 man-hours.
- (ix) The dimension of each box of the main product is $31 \times 27 \times 37 \text{ cm}^3$. Hence, it has been considered that every 32 boxes of the product can be stored in 1 m^3 of the space of the warehouse. The dimension of the warehouse equals $15 \times 20 \text{ m}^2$ and the boxes are stored to 1.5 m height. Hence, the volume of the warehouse equals 450 m^3 .
- (x) Minimum attractive rate of return as the escalating factor for the costs is equal to 8%.
- (xi) The value for α level, which is the minimal possibility level for accepting the membership, for all the fuzzy and imprecise numbers is considered to be 0.5.

5.2. *Computational Results and Discussions.* In this section, on the basis of the solution procedure that has been developed in this study, at first, the fuzzy constraints are treated using ((12)-(13)). Then, it is the turn to treat the fuzzy objective function, by introducing the PIS, NIS, and related linear membership function using (15). Finally, the auxiliary mathematical model is developed based on (16). This model can be solved by using conventional methods. In this study, Lingo Optimization Software has been applied to solve the model. The amounts of Z_{PIS} and Z_{NIS} , after solving the ordinary mathematical programming model at $\alpha = 0.5$, are equal to $Z_{\text{PIS}} = 70170.19$ and $Z_{\text{NIS}} = 73842.91$. The results of the mathematical model for the case have been summarized in Table 6.

After solving the model, in order to provide more information on the results, the effects of variation of α -cut on the results of the problem have been investigated. For this purpose, sensitivity analysis of the parameter of α -cut has been conducted by solving the mathematical model for $\alpha = 0.3, 0.7$. The results are presented in Table 7.

According to the results and sensitivity analysis, several managerial implications will be risen as follows.

- (i) The mathematical model not only provides an aggregate production planning for the production system, but also selects the best contractors for transporting the products to the distribution centers or retailers.
- (ii) By application of fuzzy set theory, the model is capable of handling the imprecise nature of data, and providing a better imitation of the real-world uncertain environments.

Other relevant data are shown in Tables 1, 2, 3, 4, and 5.

- (iii) The feasibility of the results implies the validity of the proposed mathematical model and also the solution approach.
- (iv) As from the results it can be observed that the value of L (the level of the satisfaction of the decision maker) is equal to 100%. It is an indicator for the full satisfaction of the decision maker with the results.
- (v) As the results of sensitivity analysis indicate, the degree of satisfaction of the decision maker, which is shown by the value of L , does not have any changes against the variation of the α -cut level, whereas the optimal solution of the problem changes.

6. Conclusion

In today's competitive and uncertain business environment, designing and operating efficient supply chains are crucial for every industry to meet customer demands and maximize the profit. Separate optimization of supply chain keeps the firms away from gaining maximum possible effectiveness. Hence, a mathematical model for a production-distribution problem was developed in this study that not only allocates the limited resources to the production of the products, but also determines the best contractors for dispatching the products to the distribution centers or retailers. Additionally, in order to make the model more realistic, because some data of real-world production systems are imprecise or unobtainable, the fuzzy set theory was applied to the mathematical model. For illustration and verification of the proposed fuzzy mathematical model, data from a mineral water bottling factory was applied to the model and solved by a three-step solution approach that also was developed in this study. Solving the real case by the solution method and finding feasible solutions for the model show the feasibility and validity of the proposed fuzzy mathematical model and also the developed solution procedure. The results show the full satisfaction of the decision maker with the results. Moreover, it is shown that, even by varying the level of α -cut, there is not any change in the level of satisfaction of the decision maker.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The authors thankfully acknowledge the financial support from the University of Malaya Research Grant UMRG Project RG139/12AET for this research.

References

- [1] J. Wang and Y.-F. Shu, "Fuzzy decision modeling for supply chain management," *Fuzzy Sets and Systems*, vol. 150, no. 1, pp. 107–127, 2005.
- [2] C. H. Chiu, T. M. Choi, H. T. Yeung, and Y. Zhao, "Sales rebate contracts in fashion supply chains," *Mathematical Problems in Engineering*, vol. 2012, Article ID 908408, 19 pages, 2012.
- [3] A. Baykasoglu and T. Gocken, "Multi-objective aggregate production planning with fuzzy parameters," *Advances in Engineering Software*, vol. 41, no. 9, pp. 1124–1131, 2010.
- [4] S. M. J. Mirzapour Al-e-hashem, A. Baboli, and Z. Sazvar, "A stochastic aggregate production planning model in a green supply chain: considering flexible lead times, nonlinear purchase and shortage cost functions," *European Journal of Operational Research*, vol. 230, no. 1, pp. 26–41, 2013.
- [5] B. Fahimnia, R. Z. Farahani, R. Marian, and L. Luong, "A review and critique on integrated production-distribution planning models and techniques," *Journal of Manufacturing Systems*, vol. 32, pp. 1–19, 2013.
- [6] T.-F. Liang, "Fuzzy multi-objective production/distribution planning decisions with multi-product and multi-time period in a supply chain," *Computers and Industrial Engineering*, vol. 55, no. 3, pp. 676–694, 2008.
- [7] Z. Firoozi, N. Ismail, Sh. Ariafar, S. H. Tang, M. K. A. M. Ariffin, and A. Memariani, "Distribution network design for fixed lifetime perishable products: a model and solution approach," *Journal of Applied Mathematics*, vol. 2013, Article ID 891409, 13 pages, 2013.
- [8] Z. Firoozi, S. H. Tang, Sh. Ariafar, and M. K. A. M. Ariffin, "An optimization approach for a joint location inventory model considering quantity discount policy," *Arabian Journal for Science and Engineering*, vol. 38, no. 4, pp. 983–991, 2013.
- [9] Y. B. Park, "An integrated approach for production and distribution planning in supply chain management," *International Journal of Production Research*, vol. 43, no. 6, pp. 1205–1224, 2005.
- [10] A. S. Safaei, S. M. Moattar Husseini, R. Z.-Farahani, F. Jolai, and S. H. Ghodsypour, "Integrated multi-site production-distribution planning in supply chain by hybrid modelling," *International Journal of Production Research*, vol. 48, no. 14, pp. 4043–4069, 2010.
- [11] P. Amorim, H.-O. Günther, and B. Almada-Lobo, "Multi-objective integrated production and distribution planning of perishable products," *International Journal of Production Economics*, vol. 138, no. 1, pp. 89–101, 2012.
- [12] H. Selim, C. Araz, and I. Ozkarahan, "Collaborative production-distribution planning in supply chain: a fuzzy goal programming approach," *Transportation Research Part E: Logistics and Transportation Review*, vol. 44, no. 3, pp. 396–419, 2008.
- [13] S. M. J. Mirzapour Al-E-Hashem, A. Baboli, S. J. Sadjadi, and M. B. Aryanezhad, "A multiobjective stochastic production-distribution planning problem in an uncertain environment considering risk and workers productivity," *Mathematical Problems in Engineering*, vol. 2011, Article ID 406398, 14 pages, 2011.
- [14] S. Ariafar, N. Ismail, S. H. Tang, M. K. A. M. Ariffin, and Z. Firoozi, "The reconfiguration issue of stochastic facility layout design in cellular manufacturing systems," *International Journal of Services and Operations Management*, vol. 11, no. 3, pp. 255–266, 2012.
- [15] T.-M. Lin and T.-F. Liang, "Aggregate production planning with multiple fuzzy goals," *Journal of the Chinese Institute of Industrial Engineers*, vol. 19, no. 4, pp. 39–47, 2002.
- [16] R.-C. Wang and T.-F. Liang, "Applying possibilistic linear programming to aggregate production planning," *International Journal of Production Economics*, vol. 98, no. 3, pp. 328–341, 2005.

- [17] R. A. Aliev, B. Fazlollahi, B. G. Guirimov, and R. R. Aliev, "Fuzzy-genetic approach to aggregate production-distribution planning in supply chain management," *Information Sciences*, vol. 177, no. 20, pp. 4241–4255, 2007.
- [18] S. M. J. Mirzapour Al-E-Hashem, H. Malekly, and M. B. Aryanezhad, "A multi-objective robust optimization model for multi-product multi-site aggregate production planning in a supply chain under uncertainty," *International Journal of Production Economics*, vol. 134, no. 1, pp. 28–42, 2011.
- [19] T.-F. Liang, "Imprecise multi-objective production/distribution planning decisions using possibilistic programming method," *Journal of Intelligent & Fuzzy Systems*, vol. 25, no. 1, pp. 219–230, 2013.
- [20] H.-J. Zimmermann, *Fuzzy Set Theory-And Its Applications*, Springer, 2001.
- [21] K. Demirli and A. D. Yimer, "Fuzzy scheduling of a build-to-order supply chain," *International Journal of Production Research*, vol. 46, no. 14, pp. 3931–3958, 2008.
- [22] K. Taghizadeh, M. Bagherpour, and I. Mahdavi, "Application of fuzzy multi-objective linear programming model in a multi-period multi-product production planning problem," *International Journal of Computational Intelligence Systems*, vol. 4, no. 2, pp. 228–243, 2011.
- [23] R. E. Bellman and L. A. Zadeh, "Decision-making in a fuzzy environment," *Management Science*, vol. 17, pp. B141–B164, 1970.
- [24] H.-J. Zimmermann, "Fuzzy programming and linear programming with several objective functions," *Fuzzy Sets and Systems*, vol. 1, no. 1, pp. 45–55, 1978.
- [25] S. A. Torabi and E. Hassini, "An interactive possibilistic programming approach for multiple objective supply chain master planning," *Fuzzy Sets and Systems*, vol. 159, no. 2, pp. 193–214, 2008.
- [26] Y.-J. Lai and C. L. Hwang, "A new approach to some possibilistic linear programming problems," *Fuzzy Sets and Systems*, vol. 49, no. 2, pp. 121–133, 1992.

Research Article

An Integrated Model for Production and Distribution Planning of Perishable Products with Inventory and Routing Considerations

S. M. Seyedhosseini and S. M. Ghoreyshi

Department of Industrial Engineering, Iran University of Science & Technology, P.O. Box 1684613114, Narmak, Tehran, Iran

Correspondence should be addressed to S. M. Ghoreyshi; ghoreyshi@iust.ac.ir

Received 28 January 2014; Revised 22 April 2014; Accepted 25 April 2014; Published 12 May 2014

Academic Editor: Michael Freitag

Copyright © 2014 S. M. Seyedhosseini and S. M. Ghoreyshi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In many conventional supply chains, production planning and distribution planning are treated separately. However, it is now demonstrated that they are mutually related problems that must be tackled in an integrated way. Hence, in this paper a new integrated production and distribution planning model for perishable products is formulated. The proposed model considers a supply chain network consisting of a production facility and multiple distribution centers. The facility produces a single perishable product that is storable only for predetermined periods. A homogenous fleet of vehicles is responsible for delivering the product from facility to distribution centers. The decisions to be made are the production quantities, the distribution centers that must be visited, and the quantities to be delivered to them. The objective is to minimize the total cost, where the trip minimization is considered simultaneously. As the proposed formulation is computationally complex, a heuristic method is developed to tackle the problem. In the developed method, the problem is divided into production submodel and distribution submodel. The production submodel is solved using LINGO, and a particle swarm heuristic is developed to tackle distribution submodel. Efficiency of the algorithm is proved through a number of randomly generated test problems.

1. Introduction

During the past decades, managers have faced big global changes in the business environment as results of advances in technology, globalization of markets, and new situations in economy and politics. With the increase of the number of world-class competitors, organizations have been under pressure to improve their interorganizational processes. In such situations, managers understood that such changes not only are not enough in long term, but also they must enter in management of their supply, distribution, and after sales companies. With such an approach, the terms “supply chain” and “supply chain management” were created [1].

A supply chain includes all facilities, works, and activities that are engaged in production and delivering a product or service from suppliers (and their suppliers) to customers (and their customers). It includes but not limited to planning and management of material requisition, manufacturing, warehousing, distribution, and after sales services. The supply

chain management should coordinate such activities so that customers can have high quality products with minimum possible cost [2].

In any supply chain, two main areas that are very important to the managers are production and distribution planning. Production planning tackles decision of how to transform raw materials into finished products respecting to meet demands on time with minimum cost. Determining lot sizes, that is, calculating the quantity to be produced for each item at each time, is an important decision in tactical production planning. Considerations such as planning horizon, number of levels, number of products, capacity or resource constraints, deterioration of items, demand type, setup structure, and inventory shortage may complicate the lot sizing problem [3]. For detailed information regarding classifications and characteristics of lot sizing problem, see, for instance, Ben-Daya et al. [4], Robinson et al. [5], and Buschkühl et al. [6].

On the other hand, distribution planning tackles decision of how to deliver the finished products to the customers respecting to meet their demands on time with minimum cost. Routing problem is among one of the model types in the literature that are applied to the distribution planning. In such problems, distribution of a product(s) from a central location to multiple geographically dispersed customers using definite/indefinite number of capacitated vehicles is considered. The decisions to be made are (1) when to deliver to each customer, (2) how much to deliver to each customer at each time, and (3) how to serve customers using the vehicles [7].

In a conventional supply chain, independent manufacturers, wholesalers, and retailers were separate business entities seeking to maximize their own profits. However, it is now demonstrated that the production and distribution decisions are mutually related problems that need to be dealt with in an integrated way. Integrated production and distribution planning in the context of supply chain management has been under attention of many researchers over the past few years. There might be two primary reasons behind this trend: (1) positive effect of integrated production and distribution planning on profitability of supply chains and (2) positive effect of integrated production and distribution planning on reducing lead times and offering quicker responses to market changes [8]. In this paper, we focus on the integrated production and distribution planning of perishable products. The organization of this paper is as follows. In the next section, related literature is reviewed and then in Section 3 the studied problem is thoroughly defined and formulated. As the proposed model is hard to solve specially in large problem sizes, a heuristic method is developed which is efficient in light of both quality of the solutions and the running time. Finally, concluding remarks and future research remarks are given in Section 6.

2. Literature Review

An integrated production and distribution system often includes facility(s) producing the product(s) and number of distribution centers warehousing products. An integrated production and distribution planning problem is the problem of simultaneously finding the decision variables from different functions that have traditionally been optimized independently. Some of these variables are as follows:

- (i) quantity of product(s) produced in facility(s) at each planning period,
- (ii) inventory amount of product(s) temporarily stored in facility(s) at the end of each planning period,
- (iii) quantity of product(s) shipped from facility(s) to distribution centers in each planning period,
- (iv) inventory of finished products stored in distribution centers at the end of each planning period.

Due to the number of decision variables to be determined, the integrated production and distribution planning problem is so complex that optimal values are very hard to obtain. In addition, considerations such as complex structure of

the network, geographical span of the supply chain, and involvement of different entities with conflicting objectives can further complicate the problem [9]. It is worthy to note that integrated production and distribution planning problems arise in environments where vendor managed inventory (VMI) is implemented. In vendor managed inventory (VMI) environments, the supplier/manufacturer manages inventories at customers to ensure that they do not face any shortages.

The integrated production and distribution planning problem has received a lot of attention among researchers particularly in the recent years. Sarmiento and Nagi [10], Chen [11], and Fahimnia et al. [8] provided comprehensive reviews on the general subject. Below we mention some more recent researches. Bard and Nananukul [12] formulated an integrated lot sizing and inventory routing problem as a mixed integer program with the objective of maximizing the net. They developed a two-step procedure that first estimated daily delivery quantities and then solved a vehicle routing problem for each day of the planning horizon. Boudia and Prins [13] studied an NP-hard multiperiod production—distribution problem to minimize sum of three costs: production setups, inventories, and distribution. This problem then was solved by a memetic algorithm with population management (MAPM). Bard and Nananukul [14] presented a model that included a single production facility and a set of customers with time varying demand. They developed a procedure centering on reactive tabu search for solving the problem. Bilgen and Günther [15] presented an integrated production scheduling and truck routing model for a supply chain of fruit juice. They considered different transportation modes. Bard and Nananukul [16] investigated their previously developed model aimed at minimizing costs. They also developed a hybrid methodology that combined exact and heuristic procedures within a branch-and-price framework. Bolduc et al. [17] proposed a tabu search heuristic for the split delivery vehicle routing problem with production and demand calendars. Jolai et al. [18] considered a supply chain network consisting of a manufacturer, with multiple plants, products, distribution centers, retailers, and customers. They developed a multiobjective linear programming model for integrating production—distribution decisions. They also proposed three metaheuristics to tackle the problem. Toptal et al. [19] considered a joint production and transportation planning problem where two vehicle types were available for outbound shipments. They also presented formulations for three different solution approaches. Nananukul [20] introduced an enhanced clustering model for the latter problem. An algorithm based on a reactive tabu search for solving the clustering problem was also proposed. Liu and Papageorgiou [21] addressed production, distribution, and capacity planning of global supply chains considering cost, responsiveness, and customer service level simultaneously. They developed a multiobjective mixed-integer linear program (MILP) with total cost, total flow time, and total lost sales as key objectives.

Considering perishable products, production/inventory and distribution decisions are often studied separately. Many researchers focused on extending economic production quantity (EPQ) and economic order quantity (EOQ) models for such products. The reader may refer to Nahimias [22],

Raafat [23], Goyal and Giri [24], and Bakker et al. [25] for comprehensive reviews.

On the other hand, some researchers focused on extending distribution planning models with concentration on vehicle routing problem for perishable products. Tarantilis and Kiranoudis [26] proposed a fast and robust algorithm to find effective delivery schedules for one of the biggest dairy companies in Greece. They formulated the milk delivery problem as a heterogeneous fixed fleet vehicle routing problem (VRP). Chen and Vairaktarakis [27] studied an integrated scheduling model of production and distribution operations in food catering service industries. In their model, a set of customer orders was first processed in a facility and then delivered to the customers directly. The problem was to find a joint schedule of production and distribution such that customer service level and total distribution cost objectives were optimized. Hsu et al. [28] developed a stochastic VRP with time windows formulation for delivery of perishable food products. The goal of their model was to find delivery paths, workloads, and departure times. Similarly, researches like Osvald and Stirn [29], Chen et al. [30], and Gong and Fu [31] applied VRP with time windows to distribution planning of perishable products. Osvald and Stirn [29] formulated a vehicle routing problem with time windows and time-dependent travel times (VRPTWTD) where the travel times between two locations depend on both the distance and on the time of the day. Their model considered the impact of the perishability as part of the overall distribution costs and a tabu search heuristic was used to solve the problem. Chen et al. [30] proposed a nonlinear mathematical model for production scheduling and vehicle routing with time windows for perishable food products. They assumed stochastic demands at retailers. Gong and Fu [31] proposed a multiobjective vehicle routing problem with time window model that includes fixed vehicle cost, operation cost, shelf life loss, and default cost. They applied an ant colony optimization with ABC customer classification (ABC-ACO) to solve the problem.

To the best of our knowledge, no research has ever tackled the integrated problem of production and distribution planning of perishable products with inventory and routing considerations. Therefore, in this paper we deal with an integrated model for production and distribution planning of perishable products.

3. Problem Statement

In this section, the studied problem is defined with details and the proposed model is formulated. We are given a single production facility that produces a single perishable product. The perishable product has a fixed lifetime that is measured by a number of planning periods that can be stored either in the production facility or in distribution centers. The planning horizon is divided into T equal discrete time interval that each one is a planning period. The production capacity in each planning period is limited to p_{\max} . If we have production at the facility in planning period t , then a considerable setup cost $fp_c(t = 1, \dots, T)$ is incurred. A limited amount of inventory can be temporarily stored in

the production facility with unit holding cost of h_0 . There is a set of n distribution centers geographically dispersed around the production facility that the products are to be delivered to them. Each distribution center i ($i = 1, 2, \dots, n$) has a nonnegative and deterministic demand $d_{i,t}$ in planning period t of the planning horizon that must be fully met; that is, shortages are not allowed. A limited amount of inventory can be stored in distribution center i with unit holding cost of h_i . A fleet of K capacitated homogeneous vehicles is responsible for shipping of the product from the facility to the distribution centers. The company is not owner of fleet and incurred costs are calculated based of number of trips vehicles make, not the distance they travel. Therefore, we aim to minimize the number of vehicles' trips. In addition, two other rules must be applied to deliveries; each vehicle can make at most one delivery per planning period and each distribution center can be visited at most once per planning period.

Our objective is to construct a production-distribution plan by integrating production, inventory, and routing decisions that minimizes total costs while ensuring that all demands are met and there is no shortage. The model must determine, for each planning period, whether there must be production or not, the amount to be produced, the distribution centers that must be visited, and the quantities to be delivered to them. Considering the setup cost, variable production cost, and inventory holding costs over the planning horizon, the model must decide when to overproduce and when to hold items as inventory.

In standard models, the inventory levels are only restricted by the physical storage capacities, while in our model, the perishability dominates the physical storage capacity. We defined an upper bound for inventory levels at production facility and distribution centers. Considering sl as the shelf life of the perishable product, the inventory upper bound for production facility in planning period t is equal to the sum of all distribution centers' demands in that period and next sl planning periods. Similarly, the inventory upper bound for distribution center i in planning period t is equal to the sum of its demands in that period and next sl planning periods. These upper bounds are our main contribution of the formulation.

In the development of the model, see Nomenclature section.

The formulated model is as follows.

Integrated Production and Distribution Model. Consider

$$\begin{aligned} \text{Min} \quad & \sum_t fp_c \cdot Z_t + \sum_t vp_c \cdot P_t + \sum_t h_0 \cdot I_{0,t} + \sum_{i,t} h_i \cdot I_{i,t} \\ & + \sum_{k,t} ft c_k \cdot Y_{k,t} \end{aligned} \quad (1)$$

$$I_{0,t} = I_{0,t-1} + P_t - \sum_i \sum_k W_{i,k,t} \quad \forall t \in T \quad (2)$$

$$I_{i,t} = I_{i,t-1} + \sum_k W_{i,k,t} - d_{i,t} \quad \forall i \in N, t \in T \quad (3)$$

$$I_{0,t} \leq u_{0,t} \quad \forall t \in T \quad (4)$$

$$I_{i,t} \leq u_{i,t} \quad \forall i \in N, t \in T \quad (5)$$

$$P_t \leq p_{\max} \cdot Z_t \quad \forall t \in T \quad (6)$$

$$W_{i,k,t} \leq q \cdot X_{i,k,t} \quad \forall i \in N, k \in K, t \in T \quad (7)$$

$$\sum_i W_{i,k,t} \leq q \quad \forall k \in K, t \in T \quad (8)$$

$$\sum_k X_{i,k,t} \leq 1 \quad \forall i \in N, t \in T \quad (9)$$

$$\sum_i X_{i,k,t} \leq 1 \quad \forall k \in K, t \in T \quad (10)$$

$$\sum_i X_{i,k,t} \leq q \cdot Y_{k,t} \quad \forall k \in K, t \in T \quad (11)$$

$$Z_t, X_{i,k,t}, Y_{k,t} \in \{0, 1\} \quad \forall i \in N, k \in K, t \in T \quad (12)$$

$$P_t, I_{0,t}, I_{i,t}, W_{i,k,t} \geq 0 \quad \forall i \in N, k \in K, t \in T. \quad (13)$$

In the above model, objective function (1) minimizes the sum of all costs. Note that the term $\sum_{k,t} \text{ftc}_k \cdot Y_{k,t}$ aims at minimizing the total number of vehicles trips in planning horizon. Constraints (2) are inventory balance equations at production facility that relate its inventory levels to the production quantities and deliveries to distribution centers. Similarly, constraints (3) are inventory balance equations at distribution centers. Constraints (4) and (5) guarantee that inventory levels at production facility or distribution centers are never greater than the total demands in next sl following periods. These constraints guarantee that perishable products will never be thrown away. Production quantities in planning period t are limited to the production capacity using constraints (6). Constraints (7) demonstrate that if there is a delivery to a distribution center in a planning period, then the pertaining binary variable must be 1. Vehicle capacity is respected by constraints (8). Constraints (9) guarantee that each distribution center is visited mostly once per planning period. Similarly, constraints (10) guarantee that each vehicle is used mostly once per planning period. Constraints (11) demonstrate number of vehicles' trips during planning horizon. Finally, relations (12) and (13) present decision variables types.

The size of proposed model is determined largely by constraints (7) and the number of binary variables $X_{i,k,t}$, which both grow at a rate proportional to $O(NKT)$. A typical problem with 20 vehicles, 50 distribution centers, and a 10-day planning horizon contains roughly 12000 constraints, 10000 binary variables, and 10000 continuous variables. Initial attempts to solve instances of this size with LINGO

commercial optimizer were not encouraging. This led to development of a more efficient algorithm presented in the next section.

4. Solution Method

In this section, a two-phase algorithm is presented to solve the proposed model. We face two types of decision variables in the model, the variables pertaining to production decisions ($Z_t, P_t, I_{0,t}$), and the variables pertaining to distribution decisions ($X_{i,k,t}, Y_{k,t}, W_{i,k,t}, I_{i,t}$). A natural way to solve an integrated model is to decompose the problem and consider each part dependently. In this way, we can separate the integrated model into two dependent submodels, the production submodel and the distribution submodel. In proposed method, the production submodel is solved first and the pertaining variables are determined. Then, the results are fed back into the distribution submodel and the pertaining variables are determined. In the next step, the production submodel is solved again considering the previous step. The algorithm continues this iterative solve-feedback-solve procedure until a stopping criterion is met. The details of the proposed method are given in Algorithm 1.

The algorithm consists of two phases, decomposition phase and integration phase. In the decomposition phase, the first iteration of the algorithm, the optimal lot sizes are found considering the sum of distribution centers' demands in each planning period as demand of that period. The production submodel is demonstrated in relations (14) to (19). It is a variation of lot sizing model and it is easy to solve by any commercial optimizer such as LINGO. In this submodel, the following additional parameter is used.

Additional Parameters. D_t is the sum of distribution centers' demands in planning period t ($D_t = \sum_i d_{i,t}; \forall t \in T$).

Production Submodel. Consider

$$\text{Min} \quad \sum_t \text{fp}c_t \cdot Z_t + \sum_t \text{vp}c_t \cdot P_t + \sum_t h_0 \cdot I_{0,t} \quad (14)$$

$$I_{0,t} = I_{0,t-1} + P_t - D_t \quad \forall t \in T \quad (15)$$

$$I_{0,t} \leq u_{0,t} \quad \forall t \in T \quad (16)$$

$$P_t \leq p_{\max} \cdot Z_t \quad \forall t \in T \quad (17)$$

$$Z_t \in \{0, 1\} \quad \forall t \in T \quad (18)$$

$$P_t, I_{0,t} \geq 0 \quad \forall t \in T. \quad (19)$$

Then, the algorithm proceeds with finding the quantity of products to be delivered to each distribution center and the distribution centers that each vehicle must visit in each planning period. This is done through the distribution submodel. Relations (20) to (30) demonstrate the distribution submodel. The production quantities (P_t) from production submodel work as input parameter to this submodel.

```

! Decomposition phase
Solve production sub model using LINGO
Solve distribution sub model using Particle Swarm Heuristic
Save the solution as CURRENT SOLUTION and BEST SOLUTION
Calculate CURRENT COST and set BEST COST = CURRENT COST
! End of decomposition phase
! Integration phase
While the stopping criterion is not met
  Update ARTIFICIAL DEMAND
  Run Perturbation Mechanism
  Solve production sub model using LINGO considering ARTIFICIAL SETUP COST and
  ARTIFICIAL HOLDING COST
  Solve distribution sub model using Particle Swarm Heuristic
  Save the solution as CURRENT SOLUTION
  Calculate the CURRENT COST
  If (CURRENT COST < BEST COST)
    Set BEST SOLUTION = CURRENT SOLUTION
    Set BEST COST = CURRENT COST
  End If
End While
! End of integration phase
Return BEST SOLUTION and BEST COST

```

ALGORITHM 1: Detailed steps of proposed method.

Distribution Submodel. Consider

$$\text{Min} \quad \sum_{i,t} h_i \cdot I_{i,t} + \sum_{k,t} ft c_k \cdot Y_{k,t} \quad (20)$$

$$I_{0,t} = I_{0,t-1} + P_t - \sum_i \sum_k W_{i,k,t} \quad \forall t \in T \quad (21)$$

$$I_{i,t} = I_{i,t-1} + \sum_k W_{i,k,t} - d_{i,t} \quad \forall i \in N, t \in T \quad (22)$$

$$I_{i,t} \leq u_{i,t} \quad \forall i \in N, t \in T \quad (23)$$

$$W_{i,k,t} \leq q \cdot X_{i,k,t} \quad \forall i \in N, k \in K, t \in T \quad (24)$$

$$\sum_i W_{i,k,t} \leq q \quad \forall k \in K, t \in T \quad (25)$$

$$\sum_k X_{i,k,t} \leq 1 \quad \forall i \in N, t \in T \quad (26)$$

$$\sum_i X_{i,k,t} \leq 1 \quad \forall k \in K, t \in T \quad (27)$$

$$\sum_i X_{i,k,t} \leq q \cdot Y_{k,t} \quad \forall k \in K \quad (28)$$

$$X_{i,k,t}, Y_{k,t} \in \{0, 1\} \quad \forall i \in N, k \in K, t \in T \quad (29)$$

$$I_{0,t}, I_{i,t}, W_{i,k,t} \geq 0 \quad \forall i \in N, k \in K, t \in T. \quad (30)$$

During initial test, we found that LINGO is not able to find optimal or good quality solution for the above submodel in a reasonable time. Therefore, we used LINGO to obtain

initial feasible solution and then a particle swarm based heuristic to improve it.

Particle swarm optimization (PSO) simulates the social behavior of natural organisms such as bird flocking and fish schooling to find a place with enough food. Indeed, in those swarms, a coordinated behavior using local movements emerges without any central control. In PSO, a swarm consists of number of particles. Each particle is a candidate solution to the problem. A particle has its own position and velocity. Optimization takes advantage of the cooperation between the particles. The success of some particles will influence the behavior of the others. Each particle successively adjusts its position according to the following two factors: the best position visited by itself and the best position visited by the whole swarm [32]. Originally, PSO has been successfully designed for continuous optimization problems in [33, 34]. However, Kennedy and Eberhart [35] firstly introduced discrete version of PSO.

The details of proposed heuristic are given in Algorithm 2.

In particle swarm heuristic, how to encode the problem to set of particles is of great importance. Consider the problem in which N distribution centers are to be served by K vehicles in T planning periods; we have to map a 2-dimensional array of $(N \times K, T)$ for each particle. The first dimension includes K sections, where each section has N binary points. The second dimension includes T binary points. If a value is equal to 1, it represents that the corresponding distribution center is served by the relevant vehicle in relevant planning period. An example of encoding structure for problem with five distribution centers, two vehicles, and five planning periods is given in Table 1.

```

Run LINGO to find an initial feasible solution
Set the initial feasible solution as Localbest Solution and Globalbest Solution
Generate particles and set all equal to initial feasible solution
Set all the particles equal to initial feasible solution
While the stopping criterion is not met
  For each particle  $X$ , calculate  $V = V + \alpha \cdot (X_{localbest} - X) + \beta \cdot (X_{globalbest} - X)$ .
  For each bit  $x^j$  in particle  $X$ , If ( $rand < Sigmoid(v^j)$ ) then  $x^j = 1$ , else  $x^j = 0$ 
  Check feasibility for each particle and repair the particle
  Calculate fitness function for each particle
  Update Localbest Solution and Globalbest Solution
End while
Run LINGO using Globalbest Solution to obtain real valued variables

```

ALGORITHM 2: Details of particle swarm heuristic.

TABLE 1: An example of encoding structure.

	Served by vehicle 1					Served by vehicle 2				
	DC-1	DC-2	DC-3	DC-4	DC-5	DC-1	DC-2	DC-3	DC-4	DC-5
Period 1	1	1	0	0	0	0	0	1	1	1
Period 2	1	0	1	0	0	0	1	0	1	0
Period 3	0	0	0	0	1	0	1	1	0	0
Period 4	1	0	0	1	0	0	0	0	0	0
Period 5	0	0	0	0	0	1	0	0	0	0

Based on the formulation, the following rules must be respected in each solution: (1) each distribution center must be served at most once per planning period, (2) each vehicle can make at most one delivery per planning period, and (3) the perishability of the product must be respected; that is, time between two consecutive delivery to a distribution center must be at most equal to the shelf life of the product. For instance, if the shelf life of the product is 2 days, a delivery on Monday and another on Thursday is not allowed, because it results in shortage. During the algorithm run, if any of the above rules was violated, the solution becomes infeasible and must be repaired. For rules 1 and 2, if the value of more than one position in the corresponding positions in any particle is 1, we randomly select one position and set its value to 1 and the others to 0. For rule 3, consider that values corresponding to planning periods a and b are 1 in a particle. Without loss of generality, let us assume that $a < a + sl < b$. We set the corresponding value to planning period $b = 0$ and set the value corresponding to planning period $a + sl = 1$.

During algorithm run, each particle must be measured according to a fitness function. Because we aim at minimizing the total number of vehicles trips, the term $\sum_{i,k,t} ftc_k \cdot X_{i,k,t}$ is used as fitness function. Finally, when the binary variables $X_{i,k,t}$ and $Y_{k,t}$ are determined via above procedure, we use them as input to distribution subproblem in LINGO to find real variables.

It is clear that approach of decomposing an integrated problem into two submodels and then solving each submodel separately does not necessarily lead to a good quality solution for the integrated problem. So during integration phase that

```

If (NUMBER OF SETUPS > T1)
  ARTIFICIAL SETUP COST = 2 * SETUP COST
  ARTIFICIAL HOLDING COST = HOLDING COST/2
End If
If (NUMBER OF SETUPS < T2)
  ARTIFICIAL SETUP COST = SETUP COST/2
  ARTIFICIAL HOLDING COST = 2 * HOLDING COST
End If

```

ALGORITHM 3: Details of perturbation mechanism.

starts from iteration two, new lot sizes and new delivery quantities are found considering the quantities determined in the previous iteration. In a more specific word, new lot sizes are found considering the amount of products delivered to distribution centers in the last distribution plan as demand, instead of the sum of demands of all distribution centers. We call these new demands *ARTIFICIAL DEMANDS*. Similarly, new delivery quantities are found using the later lot sizes. This iterative process continues until stopping criterion is met, that is, when the outputs of submodels become unchanged.

During the preliminary tests, we often observed a fast convergence of the algorithm in a few iterations. This shows that the algorithm is trapped in the local optimal point. To escape from such point, we implemented a perturbation mechanism in the production submodel. After the computation of such a perturbed lot sizes, the algorithm proceeds with the distribution submodel as usual. Details of perturbation mechanism are given in Algorithm 3.

TABLE 2: Number and size of test problems.

Problem number	Small size	Large size	
	Size ($N \times T \times K$)	Problem number	Size ($N \times T \times K$)
1	$4 \times 5 \times 2$	9	$30 \times 10 \times 10$
2	$4 \times 10 \times 2$	10	$30 \times 15 \times 10$
3	$6 \times 5 \times 3$	11	$40 \times 10 \times 15$
4	$6 \times 10 \times 3$	12	$40 \times 15 \times 15$
5	$8 \times 5 \times 4$	13	$50 \times 10 \times 20$
6	$8 \times 10 \times 4$	14	$50 \times 15 \times 20$
7	$10 \times 5 \times 5$	15	$60 \times 10 \times 25$
8	$10 \times 10 \times 5$	16	$60 \times 15 \times 25$

5. Computational Results

In this section, the efficiency of the proposed algorithm is discussed with respect to time performance and quality of results. We implemented the proposed algorithm in MATLAB 2010 in which DLL of LINGO 8.0 is contained in the code. All test runs of the algorithm are performed on a Celeron(R) CPU 2.40 GHz PC with 512 MB of RAM.

5.1. Test Problems. To the best of our knowledge, no public test problems are available for our problem. So in order to evaluate the algorithm efficiency, we randomly generated test problems. Sizes of test problems, including number of distribution centers, number of planning periods, and number of available vehicles, are given in Table 2. We classified test problems into two groups, small size and large size.

We generated model parameters as follows.

- (i) Demand quantities were generated from a uniform distribution with lower bound = 10 and upper bound = 20.
- (ii) Variable production costs were generated from a uniform distribution with lower bound = 50 and upper bound = 100.
- (iii) Fixed production costs were generated from a uniform distribution with lower bound = 500 and upper bound = 1000.
- (iv) Inventory holding costs in production facility and distribution centers were generated from a uniform distribution with lower bound = 5 and upper bound = 10.
- (v) Fixed transportation costs were generated from a uniform distribution with lower bound = 200 and upper bound = 300.
- (vi) Shelf life of the product was set to 2 or 3 periods randomly.
- (vii) Vehicle capacity was set to $(1.5 \sum_{i,t} d_{i,t})/N \cdot T$.
- (viii) Production capacity was set to $(3.5 \sum_{i,t} d_{i,t})/N \cdot T$.

5.2. Lower Bound Computation. A simple way to validate the quality of solutions provided by proposed method is to

TABLE 3: Algorithm parameters.

Parameter	Value
T1	$[T/3]$
T2	$[2T/3]$
Number of PSO particles	50
α	0.35
β	0.65
Number of PSO iterations	100

compare them to lower bounds on the optimal solution. In order to calculate a lower bound, it is necessary to solve a relaxation of the proposed model. We formulated such a relaxation in relations (31) to (39). In this formulation, the following additional parameter is used.

Additional Parameters. vtc_k is the variable transportation cost of one item using vehicle k which is equal to ftc_k/q .

The main difference between the original model and the relaxed model is that the binary variables $X_{i,k,t}$ and $Y_{k,t}$ and pertaining constraints are removed, which makes the relaxed model easy to solve.

Relaxed Model. Consider

$$\begin{aligned} \text{Min} \quad & \sum_t fpc_t \cdot Z_t + \sum_t vpc_t \cdot P_t + \sum_t h_0 \cdot I_{0,t} + \sum_{i,t} h_i \cdot I_{i,t} \\ & + \sum_{i,k,t} vtc_k \cdot W_{i,k,t} \end{aligned} \quad (31)$$

$$I_{0,t} = I_{0,t-1} + P_t - \sum_i \sum_k W_{i,k,t} \quad \forall t \in T \quad (32)$$

$$I_{i,t} = I_{i,t-1} + \sum_k W_{i,k,t} - d_{i,t} \quad \forall i \in N, t \in T \quad (33)$$

$$I_{0,t} \leq u_{0,t} \quad \forall t \in T \quad (34)$$

$$I_{i,t} \leq u_{i,t} \quad \forall i \in N, t \in T \quad (35)$$

$$P_t \leq p_{\max} \cdot Z_t \quad \forall t \in T \quad (36)$$

$$\sum_i W_{i,k,t} \leq q \quad \forall k \in K, t \in T \quad (37)$$

$$Z_t \in \{0, 1\} \quad \forall t \in T \quad (38)$$

$$P_t, I_{0,t}, I_{i,t}, W_{i,k,t} \geq 0 \quad \forall i \in N, k \in K, t \in T. \quad (39)$$

It is worthy to note that any feasible solution for the original model is also feasible to the relaxed model. In addition, the objective function of the relaxed model is similar to original model, except for the term $\sum_{i,k,t} vtc_k \cdot W_{i,k,t}$ that we substituted to the term $\sum_{k,t} ftc_k \cdot Y_{k,t}$ in the original model and always underestimate it.

5.3. Results. To run the proposed method, the algorithm parameters were set according to Table 3.

TABLE 4: Computational results for small size test problems.

Number	Z^{Alg}	CPU ^{Alg}	Z^{LINGO}	Gap
1	38369	6.3	38369	0.00%
2	76739	14.7	75709	1.34%
3	57554	17.0	56239	2.29%
4	115108	36.9	113258	1.61%
5	76739	32.5	75315	1.86%
6	153478	70.3	149429	2.64%
7	95924	48.9	93993	2.01%
8	191004	118.6	184744	3.28%

A summary of the results for small size test problems is reported in Table 4. In columns 2–4 of the table, objective function obtained by the proposed method (Z^{Alg}), runtime of the proposed method in seconds, and optimal solution obtained by LINGO (Z^{LINGO}) are reported, respectively. In the final column, the gap between Z^{Alg} and Z^{LINGO} that is equal to $((Z^{\text{Alg}} - Z^{\text{LINGO}})/Z^{\text{Alg}}) \times 100$ is presented.

We observe in Table 4 that the proposed method can provide good solutions in short runtime. The gaps between Z^{Alg} and Z^{LINGO} are less than 3% in most cases. The worst performance of the method also can obtain a solution gap that is less than 4%.

The proposed method was run 5 times for each large test problem and a summary of the results is reported in Table 5. In columns 2–5 of the table, the average and worst objective functions obtained by the proposed method (Z^{Alg}) and the average and worst runtimes of the proposed method in seconds are reported. In columns 6–8, the solution obtained by relaxed model (Z^{LB}) and the average and worst gap between Z^{Alg} and Z^{LB} that is equal to $((Z^{\text{Alg}} - Z^{\text{LB}})/Z^{\text{LB}}) \times 100$ are reported, respectively. Finally in columns 9–10, the best solution obtained by LINGO within 2 hours run (Z^{LINGO}) and the gap between Z^{Alg} and Z^{LB} which is equal to $((Z^{\text{LINGO}} - Z^{\text{LB}})/Z^{\text{LB}}) \times 100$ are presented, respectively.

The gaps are computed according to the lower bound not to the developed method. In this way, they represent the maximum deviation from the optimal solution.

It is worthy to note that we run LINGO for the test problem number 9 for a night and did not see any significant improvement in results to when we run it for 2 hours. Therefore, we set the LINGO for other test problems for 2 hours.

We observe in Table 4 that the proposed method can provide good solutions in short runtime. The integrality gaps of the proposed method are less than 8.5% in average and less than 12% in worst. This means that the maximum deviation from optimal solution is very reasonable according to the problems complexity. The deviations between the average and worst gaps are about 3% that means the algorithm is very robust to produce results. Comparing the proposed method and LINGO, we see in the table that gap 1 is about half of gap 2 that means the proposed method can provide better solutions than LINGO. In light of the run time, proposed method can provide solutions in less time than LINGO.

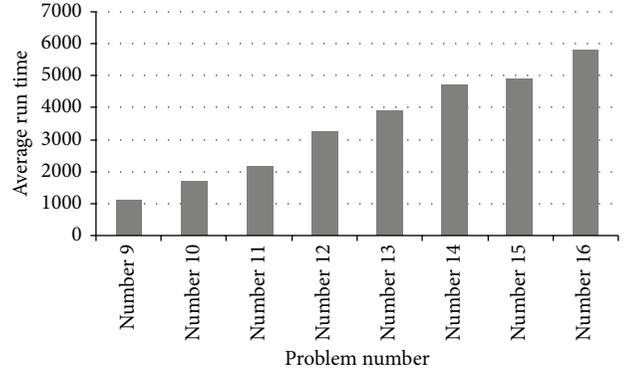


FIGURE 1: Average run time of the proposed method.

As stated in the final paragraph of the Section 3, complexity of the proposed model depends on multiplication value of three parameters; number of distribution centers, vehicle, and planning periods. As shown in Figure 1, with increase of this value, run time of the proposed method also increases.

6. Conclusion and Future Research Remarks

In this paper, a new formulation for integrating production planning and distribution planning of perishable products in a supply chain is presented. Some assumptions about the problem are as follows: the supply chain network includes a single production facility and multiple distribution centers. There is also fleet of homogenous vehicles that are responsible for transporting the product from facility to distribution centers. The company is not owner of the fleet and pertaining incurred costs are fixed and measured based on the number of trip vehicles make not the distance they travel. In each planning period, each vehicle can make one trip at most and each distribution center can be visited at most once. The products are perishable; that is, they are storable only for predetermined periods. The production capacity is limited and shortage in distribution centers is not allowed. The decisions to be made in each planning period are the production quantities, the distribution centers to be visited, and the delivery quantities. Based on the above assumptions, a mixed integer linear program (MILP) is formulated.

Respecting the computational complexity of the model, a heuristic method is developed to tackle it. In the proposed method, LINGO and particle swarm optimization heuristic are combined. In order to validate the performance of proposed method, a number of randomly generated test problems are used. The results are compared with the LINGO and lower bound of the optimal solution. The proposed method is efficient in light of solution quality and time performance. In addition, it is able to provide robust results. In addition, it is shown that time complexity of the proposed method depends on linear multiplication of three parameters, number of distribution centers, vehicle, and planning periods.

Although the studied problem fits into the real world application, it may imposes some limitations into the decision

TABLE 5: Computational results (large size test problems).

Number	Z^{Alg}		CPU Alg		Z^{LB}	Gap 1		Z^{LINGO}	Gap 2
	Avg.	Worst	Avg.	Worst		Avg.	Worst		
9	556028	572704	1097	1141	526941	5.52%	8.68%	590796	12.12%
10	834042	860841	1691	1773	769475	8.39%	11.87%	895742	16.41%
11	741370	765830	2175	2244	692744	7.02%	10.55%	799822	15.46%
12	1112056	1147772	3252	3374	1054671	5.44%	8.83%	1180939	11.97%
13	926713	955584	3919	4039	863965	7.26%	10.60%	984028	13.90%
14	1390069	1426843	4702	4924	1293635	7.45%	10.30%	1476745	14.15%
15	1112056	1143883	4892	5088	1030104	7.96%	11.05%	1184692	15.01%
16	1668083	1711109	5791	5965	1549744	7.64%	10.41%	1792531	15.67%

making. For example, the problem can be developed when the company owns the vehicle fleet and so the distance each vehicle travels in each trip is important. In this situation, a customized inventory routing model must be developed with respect to problem definition. In addition, other real world assumptions such as inventory transshipments between distribution centers or split and delivery of the inventories can be added to the model. It should be noted that such assumptions complicate the model, which may result in more sophisticated solution methods.

Nomenclature

Indices

- $t = 1, 2, \dots, T$: Set of planning periods
- $i = 0, 1, \dots, N$: Set of production facility and distribution centers, where 0 corresponds to production facility
- $k = 1, 2, \dots, K$: Set of vehicles.

Parameters

- $d_{i,t}$: Demand of distribution center i in planning period t
- fp_c : Fixed production cost in planning period t
- vp_c : Variable production cost in planning period t
- p_{max} : Production capacity
- h_0 : Inventory holding cost at production facility
- h_i : Inventory holding cost at distribution center i
- sl: Shelf life of perishable products which is measured in number of planning periods that the product can be stored
- $u_{0,t}$: Upper bound of inventory level at production facility in planning period t , which is equal to $\sum_i \sum_{t \leq \tau \leq t+sl} d_{i,\tau}$
- $u_{i,t}$: Upper bound inventory level at distribution center i in planning period t , which is equal to $\sum_{t \leq \tau \leq t+sl} d_{i,\tau}$
- q : Vehicle capacity
- ft_{c_k} : Fixed transportation cost of using vehicle k .

Variables

- Z_t : 1 if there is production on planning period t , 0 otherwise
- P_t : Production quantity in planning period t
- $I_{0,t}$: Inventory level at production facility in planning period t
- $X_{i,k,t}$: 1 if distribution i is served by vehicle k in planning period t , 0 otherwise
- $Y_{k,t}$: 1 if vehicle k is used in planning period t to serve distribution centers, 0 otherwise
- $W_{i,k,t}$: Amount delivered to distribution center i in planning period t by vehicle k
- $I_{i,t}$: Inventory level at distribution center i in planning period t .

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] D. Blanchard, *Supply Chain Management Best Practices*, John Wiley & Sons, New York, NY, USA, 2010.
- [2] S. Chopra and P. Meindl, *Supply Chain Management. Strategy, Planning & Operation*, Gabler, Wiesbaden, Germany, 2007.
- [3] B. Karimi, S. M. T. Fatemi Ghomi, and J. M. Wilson, "The capacitated lot sizing problem: a review of models and algorithms," *Omega*, vol. 31, no. 5, pp. 365–378, 2003.
- [4] M. Ben-Daya, M. Darwish, and K. Ertogral, "The joint economic lot sizing problem: review and extensions," *European Journal of Operational Research*, vol. 185, no. 2, pp. 726–742, 2008.
- [5] P. Robinson, A. Narayanan, and F. Sahin, "Coordinated deterministic dynamic demand lot-sizing problem: a review of models and algorithms," *Omega*, vol. 37, no. 1, pp. 3–15, 2009.
- [6] L. Buschkühl, F. Sahling, S. Helber, and H. Tempelmeier, "Dynamic capacitated lot-sizing problems: a classification and review of solution approaches," *OR Spectrum. Quantitative Approaches in Management*, vol. 32, no. 2, pp. 231–261, 2010.
- [7] L. Bertazzi, M. Savelsbergh, and M. G. Speranza, "Inventory routing," in *The Vehicle Routing Problem: Latest Advances and New Challenges*, Springer, Berlin, Germany, 2008.
- [8] B. Fahimnia, R. Zanjirani Farahani, R. Marian, and L. Luong, "A review and critique on integrated production-distribution

- planning models and techniques," *Journal of Manufacturing Systems*, vol. 32, no. 1, pp. 1–19, 2013.
- [9] A. Pandey, M. Masin, and V. Prabhu, "Adaptive logistic controller for integrated design of distributed supply chains," *Journal of Manufacturing Systems*, vol. 26, no. 2, pp. 108–115, 2007.
- [10] A. M. Sarmiento and R. Nagi, "A review of integrated analysis of production-distribution systems," *IIE Transactions*, vol. 31, no. 11, pp. 1061–1074, 1999.
- [11] Z.-L. Chen, "Integrated production and outbound distribution scheduling: review and extensions," *Operations Research*, vol. 58, no. 1, pp. 130–148, 2010.
- [12] J. F. Bard and N. Nananukul, "Heuristics for a multiperiod inventory routing problem with production decisions," *Computers and Industrial Engineering*, vol. 57, no. 3, pp. 713–723, 2009.
- [13] M. Boudia and C. Prins, "A memetic algorithm with dynamic population management for an integrated production-distribution problem," *European Journal of Operational Research*, vol. 195, no. 3, pp. 703–715, 2009.
- [14] J. F. Bard and N. Nananukul, "The integrated production-inventory-distribution-routing problem," *Journal of Scheduling*, vol. 12, no. 3, pp. 257–280, 2009.
- [15] B. Bilgen and H.-O. Günther, "Integrated production and distribution planning in the fast moving consumer goods industry: a block planning application," *OR Spectrum. Quantitative Approaches in Management*, vol. 32, no. 4, pp. 927–955, 2010.
- [16] J. F. Bard and N. Nananukul, "A branch-and-price algorithm for an integrated production and inventory routing problem," *Computers & Operations Research*, vol. 37, no. 12, pp. 2202–2217, 2010.
- [17] M.-C. Bolduc, G. Laporte, J. Renaud, and F. F. Boctor, "A tabu search heuristic for the split delivery vehicle routing problem with production and demand calendars," *European Journal of Operational Research*, vol. 202, no. 1, pp. 122–130, 2010.
- [18] F. Jolai, J. Razmi, and N. K. M. Rostami, "A fuzzy goal programming and meta heuristic algorithms for solving integrated production: distribution planning problem," *Central European Journal of Operations Research*, vol. 19, no. 4, pp. 547–569, 2011.
- [19] A. Toptal, U. Koc, and I. Sabuncuoglu, "A joint production and transportation planning problem with heterogeneous vehicles," *Journal of the Operational Research Society*, vol. 65, no. 1, pp. 180–196, 2013.
- [20] N. Nananukul, "Clustering model and algorithm for production inventory and distribution problem," *Applied Mathematical Modelling*, vol. 37, no. 24, pp. 9846–9857, 2013.
- [21] S. Liu and L. G. Papageorgiou, "Multi objective optimisation of production, distribution and capacity planning of global supply chains in the process industry," *Omega*, vol. 41, no. 2, pp. 369–382, 2013.
- [22] S. Nahmias, "Perishable inventory theory: a review," *Operations Research*, vol. 30, no. 4, pp. 680–708, 1982.
- [23] F. Raafat, "Survey of literature on continuously deteriorating inventory models," *Journal of the Operational Research Society*, vol. 42, no. 1, pp. 27–37, 1991.
- [24] S. K. Goyal and B. C. Giri, "Recent trends in modeling of deteriorating inventory," *European Journal of Operational Research*, vol. 134, no. 1, pp. 1–16, 2001.
- [25] M. Bakker, J. Riezebos, and R. H. Teunter, "Review of inventory systems with deterioration since 2001," *European Journal of Operational Research*, vol. 221, no. 2, pp. 275–284, 2012.
- [26] C. D. Tarantilis and C. T. Kiranoudis, "A meta-heuristic algorithm for the efficient distribution of perishable foods," *Journal of Food Engineering*, vol. 50, no. 1, pp. 1–9, 2001.
- [27] Z.-L. Chen and G. L. Vairaktarakis, "Integrated scheduling of production and distribution operations," *Management Science*, vol. 51, no. 4, pp. 614–628, 2005.
- [28] C.-I. Hsu, S.-F. Hung, and H.-C. Li, "Vehicle routing problem with time-windows for perishable food delivery," *Journal of Food Engineering*, vol. 80, no. 2, pp. 465–475, 2007.
- [29] A. Osvald and L. Z. Stirn, "A vehicle routing algorithm for the distribution of fresh vegetables and similar perishable food," *Journal of Food Engineering*, vol. 85, no. 2, pp. 285–295, 2008.
- [30] H.-K. Chen, C.-F. Hsueh, and M.-S. Chang, "Production scheduling and vehicle routing with time windows for perishable food products," *Computers & Operations Research*, vol. 36, no. 7, pp. 2311–2319, 2009.
- [31] W. Gong and Z. Fu, "ABC-ACO for perishable food vehicle routing problem with time windows," in *Proceedings of the International Conference on Computational and Information Sciences (ICCCIS '10)*, pp. 1261–1264, December 2010.
- [32] E. G. Talbi, *Metaheuristics: From Design to Implementation*, John Wiley & Sons, New York, NY, USA, 2009.
- [33] R. Eberhart and J. Kennedy, "New optimizer using particle swarm theory," in *Proceedings of the 6th International Symposium on Micro Machine and Human Science*, pp. 39–43, October 1995.
- [34] J. Kennedy and R. Eberhart, "A discrete binary version of the particle swarm algorithm," in *Proceeding of the IEEE International Conference on Systems, Man, and Cybernetics, Computational Cybernetics and Simulation*, vol. 5, pp. 4104–4108, Orlando, Fla, USA, October 1997.
- [35] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceeding of the IEEE International Conference on Neural Network*, pp. 1942–1948, December 1995.

Research Article

Study on Application of T-S Fuzzy Observer in Speed Switching Control of AUVs Driven by States

Xun Zhang,¹ Xinglin Hu,¹ Yang Zhao,² and Zhaodong Tang¹

¹ College of Automation, Harbin Engineering University, Harbin 150001, China

² China Ship Research and Development Academy, Beijing 100192, China

Correspondence should be addressed to Xun Zhang; zhangxun2008@hrbeu.edu.cn

Received 31 January 2014; Accepted 31 March 2014; Published 12 May 2014

Academic Editor: Michael Lütjen

Copyright © 2014 Xun Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Considering the inherent strongly nonlinear and coupling performance of autonomous underwater vehicles (AUVs), the speed switching control method for AUV driven by states is presented. By using T-S fuzzy observer to estimate the states of AUV, the speed control strategies in lever plane, vertical plane, and speed kept are established, respectively. Then the adaptive switching law is introduced to switch the speed control strategies designed in real time. In the simulation, acoustic Doppler current profile/side scan sonar (ADCP/SSS) observation case is employed to demonstrate the effectiveness of the proposed method. The results show that the efficiency of AUV was improved, the trajectory tracking error was reduced, and the steady-state ability was enhanced.

1. Introduction

Many typical tasks of AUV require accurate speed control, such as path tracking and attitude control [1]. Hence, the suitable speed control strategies are important to the whole control system of AUV. However, the highly nonlinear, highly coupling, and underactuated characteristics of the dynamic model of AUV make it difficult to design the speed controller. Considering that the complex task of AUV usually is divided into lever plane task and vertical plane task and executed independently, the conventional dynamic model of AUV can be decoupled into control of speed, yaw, and depth independently [2]. Based on the linearization or decoupling of usual dynamic model of AUV, a large class of methods have been introduced to the speed control of AUV.

As the AUV dynamic model can be linearized at the domain of the working point, the PID speed controller is designed to keep the speed of AUV at a given value [3]. However, the general PID controller's parameters are fixed and difficult to set; then the adaptive PID control method is introduced in [4]; the parameters of the controller will be updated according to the speed error. Due to the uncertain disturbance of AUVs dynamic model, such as the current

disturbance [5], unmodeled dynamics [6, 7], and measurement noise [8], we cannot get the precise dynamic model of AUV. Then, the control methods which do not rely on the exact model are introduced in [9]; in this paper, wave disturbances in both vertical and horizontal planes were considered. A nonlinear, Lyapunov-based, adaptive output feedback control law is designed to depth tracking and attitude control of AUV. In [10], the robust adaptive speed controller is designed for a flexible hypersonic vehicle model which is nonlinear and multivariable and includes uncertain parameters. With a view to the unknown current and unmodeled dynamics, the radial basis function (RBF) neural net is concerned to estimate the uncertain parameters of the vehicle's model [11], and then the speed controller is designed. However, the vehicle dynamics change by different scientific missions and the acoustic motion estimation and control method are researched in [12]. In order to estimate the states of the unknown inputs, the observer was designed for T-S fuzzy models in [13]. The stability of speed controller is also important to the whole control system. Reference [14] analyse the stability of PD speed controller and S-surface speed controller, which were designed based on the Lyapunov's direct.

The methods mentioned above all depend on the decoupling dynamic model of AUV. Through the analysis of AUV dynamic model, it is concluded that when AUV is sailing in lever plane or diving in vertical plane, the coupling torque developed by vertical steering rudder or lever steering rudder can be approximately ignored, but the coupling force or torque developed by speed still exists, and with the increase of the AUV speed, the coupling will be higher [15]. In this paper, considering the coupling between speed and states, a speed switching control method driven by AUV states is proposed. Through the estimation of AUV states by T-S fuzzy observer, the speed control strategy is designed in lever plane, vertical plane, and speed kept, respectively. Then the adaptive switching law is given to switch the speed control strategies established in real time.

The structure of the paper is as follows. Section 2 introduces the nonlinear dynamic model of AUV, and the nonlinear coefficient matrix is stressed. In Section 3, the T-S fuzzy observer is designed, which is based on the dynamical model of AUV. By the estimation of AUV states, Section 4 establishes the speed control strategies of AUV. ADCP/SSS observation case is considered in Section 5 to evaluate the efficiency of the speed control method proposed in this paper. Finally, we draw conclusion in Section 6.

2. AUV Model

The vehicle we studied in this paper is equipped with three actuators: a main thruster for propulsion, a vertical steering rudder, and a lever steering rudder. Since the controllable dimensions are less than model dimensions, the controller designed for such AUV is underactuated. According to [16], the 6-DOF dynamical model of AUV described by matrix and vector is put forward as follows:

$$\begin{aligned} T\dot{x} &= A(x)x + M\tau_d + D(x)\tau + \alpha, \\ y &= Cx, \end{aligned} \quad (1)$$

where $x = [u, v, w, q, r]^T$ is the linear velocity and angular velocity of AUV in the body-fixed reference frame; $\tau = [X_{\text{prop}}, \delta_r, \delta_s]^T$ is the actuator output, which includes propeller thrust, rudder angle, and stern angle, respectively; $A \in R^{5 \times 5}$ is coefficient matrix; $D \in R^{5 \times 3}$ is control matrix; are nonlinear matrix; $M \in R^{5 \times 5}$ is the interference gain matrix; $a \in R^{5 \times 1}$ and $C \in R^{5 \times 5}$ are constant matrices. $y = [\theta, \psi, X, Y, Z]$ is the pitch, yaw, and generalized position of the AUV in the earth-fixed reference frame.

According to the standard motion equation of AUV, the nonlinear matrix $A(x)$ and $D(x)$ in (1) can be described as

$$\begin{aligned} A(x) &= \begin{bmatrix} X_{uu}u & X_wv & X_{ww}w & (-m + X_{wq})w + X_{qq}q & (m + X_{vr})v + X_{rr}r \\ 0 & Y_v & Y_{vw}v & Y_{vq}v & -mu + Y_r + Y_{wr}w + Y_{qr}q \\ 0 & Z_wv & Z_w & mu + Z_q & Z_{vr}v + Z_{rr}r \\ 0 & M_vv & M_w + M_{|w|}(v^2 + w^2)^{1/2} & M_{q|q}q + M_{wq}(v^2 + w^2)^{1/2} & (J_y - J_z)q + J_{xy}q + J_{zx}r + M_{rr}r + M_{vr}v \\ 0 & N_v & N_{vw}v + N_{|v|}(v^2 + w^2)^{1/2} & -J_{xy}q - J_{zx}r + N_{vq}v & N_r + N_{wr}w + N_{q}q + N_{r|r}|r| + N_{|v|r}(v^2 + w^2)^{1/2} \end{bmatrix}, \\ D(x) &= \begin{bmatrix} a_T & X_{\delta_r \delta_r} & X_{\delta_s \delta_s} \\ 0 & Y_{\delta_r} Y_{|r| \delta_r} |r| & 0 \\ 0 & 0 & Z_{\delta_s} + Z_{|q| \delta_s} |q| \\ 0 & K_{\delta_r} & 0 \\ 0 & N_{\delta_r} + N_{|r| \delta_r} |r| & 0 \end{bmatrix}, \end{aligned} \quad (2)$$

where $X_{[\cdot]}, Y_{[\cdot]}, Z_{[\cdot]}, M_{[\cdot]}, N_{[\cdot]}, J_{[\cdot]}$, are dimensional hydrodynamic coefficients of AUV dynamical model.

3. AUV State Estimation Based on T-S Fuzzy Observer

The nonlinear parts in (2) can be described by set as follows:

$$S = \left\{ u, v, w, q, r, (v^2 + w^2)^{1/2}, \delta_r, \delta_s \right\}. \quad (3)$$

To construct the T-S fuzzy observer based on AUV's dynamical model, (1) must be rewritten as T-S fuzzy form. So, the neighborhood nonlinear approximation principle of fuzzy inference system [17] is introduced; then (1) can be rewritten as

R_i : If u is $N_i^u \dots$ and $(v^2 + w^2)^{1/2}$ is N_i^{vw} ; then

$$\begin{aligned} \dot{x} &= A_i x + M_i \tau_d + D_i \tau + a, \\ y &= Cx, \end{aligned} \quad (4)$$

where $i = 1, 2, \dots, r$, r is the number of rules; N is the fuzzy set; $A_i \in R^{5 \times 5}$, $M_i \in R^{5 \times 5}$, $D_i \in R^{5 \times 5}$, and $C = I$ are the constant matrices which were linearized by (2). Then, the approximate T-S fuzzy model based on (4) can be designed as follows:

$$\begin{aligned} \dot{x} &= \sum_{i=1}^n w_i(z) (A_i x + M_i \tau_d + D_i \tau + a), \\ y &= Cx, \end{aligned} \quad (5)$$

where $z = [u, v, w, q, r, (v^2 + w^2)^{1/2}, \delta_r, \delta_s]^T$, and $w(z)$ is the weight unitary function; according to [13, 18], it can be calculated as

$$w_i(z) = \frac{h_i(z)}{\sum_{i=1}^n h_i(z)}, \quad i = 1, \dots, n, \quad (6)$$

$$\begin{aligned} h_i(z) &= h_{m1}(u) h_{m2}(v) h_{m3}(w) h_{m4}(q) h_{m5}(r) \\ &\times h_{m6}\left((v^2 + w^2)^{1/2}\right) h_{m7}(\delta_r) h_{m8}(\delta_s), \end{aligned} \quad (7)$$

where $m_1, m_2, m_3, m_4, m_5, m_6 \in \{1, 2\}$; $h_m(x)$ can be calculated as

$$\begin{aligned} h_1(x) &= \frac{x_{\max} - x}{x_{\max} - u_{\min}}, \\ h_2(x) &= \frac{x - x_{\min}}{x_{\max} - u_{\min}}. \end{aligned} \quad (8)$$

Together with (6), we can get that the total number of rule is $n = 2^8 = 256$.

Considering the external disturbances of AUV which mainly come from sea currents and the unmodeled dynamics of model, we can define sea currents disturbance as a and unmodeled dynamics of AUV as A_δ and B_δ . Then, the uncertain input of AUV can be described as

$$\tau_d = A_\delta x + B_\delta u + a, \quad (9)$$

where $A_\delta \in R^{5 \times 5}$, $B_\delta \in R^{5 \times 3}$, and $a \in R^{5 \times 1}$ are unknown. Then (4) can be rewritten as

$$\begin{aligned} \dot{x} &= T^{-1} \sum_{i=1}^n w_i(z) (A_i x + M_i (A_{\delta i} x + B_{\delta i} u + a) + D_i \tau), \\ y &= Cx. \end{aligned} \quad (10)$$

Then, we can design T-S fuzzy observer based on AUV T-S fuzzy model (10) as follows:

$$\begin{aligned} \hat{\dot{x}} &= \sum_{i=1}^m w_i(z) (A_i \hat{x} + D_i \tau + L_i (y - \hat{y}) \\ &\quad + M_i (\hat{A}_{\delta i} \hat{x} + \hat{B}_{\delta i} u + \hat{a}_i)), \\ \hat{y} &= C\hat{x}, \\ \hat{A}_{\delta i} &= w_i(z) M_i^T P C' e_y \hat{x}^T, \end{aligned}$$

$$\begin{aligned} \hat{B}_{\delta i} &= w_i(z) M_i^T P C' e_y u^T, \\ \hat{a}_i &= w_i(z) M_i^T P C' e_y, \end{aligned} \quad (11)$$

where C' donates the Moore-Penrose pseudoinverse of output matrix C ; L_i , $i = 1, 2, \dots, n$ are gain matrices for each rule; $P = P^T > 0$, L_i , Λ_i^k , $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$, $k = 1, 2, \dots, p$ can be calculated by LMI as

$$\begin{aligned} H &\left(P \begin{pmatrix} A_i - L_i C_j + A_j - L_j C_i & M_i + M_j & 0 & \dots & 0 \\ -\Lambda_i^1 C_j - \Lambda_j^1 C_i & 0 & 2I & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\Lambda_i^{p-1} C_j - \Lambda_j^{p-1} C_i & 0 & 0 & \dots & 2I \\ -\Lambda_i^p C_j - \Lambda_j^p C_i & 0 & 0 & \dots & 0 \end{pmatrix} \right) \\ &< 0, \quad \forall j \geq i : \exists z : w_i(z) w_j(z) \neq 0. \end{aligned} \quad (12)$$

The basic sufficient stability conditions for this observer were derived in [19].

4. Speed Control Strategies Driven by States

According to the 6-DOF dynamic model of AUV, we know that the speed of AUV is strong coupling with the states of AUV. So if the AUV speed is static, then the coupling force and moment will be different depending on the difference of AUV states, which can influence the precision of actual operation of AUV [20]. Our objective is to design a strategy of AUV speed, which is driven by AUV states, so as to decrease or eliminate the influence derived by the changing of AUV states.

4.1. Lever Variable Speed Control Strategy. When AUV is sailing in the horizontal plane, the states involved are AUV speed, sway velocity, and yaw velocity. According to the experiments, we know that when AUV yaws in the horizontal plane, the sway velocity should be as small as possible and the lateral error will be close to zero. In order to reduce the lateral displacement and heeling angle, the speed of AUV should be decreased. On the other hand, if AUV is sailing direct or the velocity of yaw is small, we should consider accelerating the AUV speed so as to work more efficiently.

Assume that at any time t the speed of AUV is $\hat{u}(t)$ which is estimated by T-S fuzzy observer; then AUV speed at time $t + 1$ can be defined as

$$u_{\text{com}}(t + 1) = u_{\text{com}}(t) + k_1 f_1(\hat{v}, \hat{r}), \quad (13)$$

where \hat{v} , \hat{r} are estimated sway velocity and yaw velocity; k_1 is the gain of speed control; f_1 is a function which is driven by AUV states. It is assumed that $|r| \leq m$, and the index of yaw is defined as $k = m/2$. Namely, when $|r| < k$, the AUV speed should be accelerated. And f_1 is designed as follows:

$$f_1 = \frac{\tan [(-|\hat{r}| + k) * n]}{14.26}, \quad n = \frac{3}{m}. \quad (14)$$

And Figure 1 described the physical meanings of function f_1 .

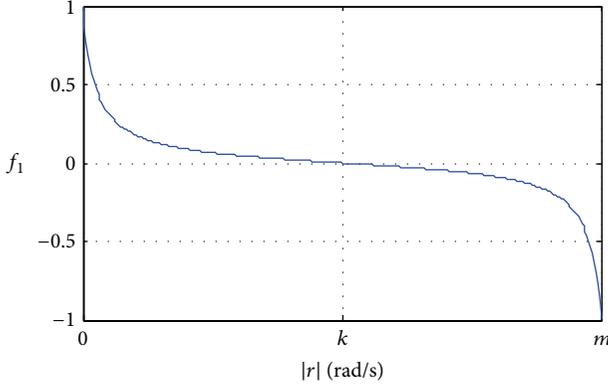
FIGURE 1: Physical meanings of f_1 .

Figure 1 illustrates that when $|r|$ is approaching k , f_1 will tend to 0; then the speed acceleration or deceleration is nearly 0; when $|r|$ is approaching 0, it means that AUV is sailing almost directly, then f_1 should quickly rise to 1 so that the speed of AUV can be kept in a high value; when $|r|$ is approaching m , it means that AUV is surging; then f_1 should decrease to -1 quickly so that AUV speed can be kept in a low value.

Combined with (13) and (14), the speed error system in lever plane can be calculated as

$$\begin{aligned} e(t) &= u_{\text{com}}(t+1) - \hat{u}(t) \\ &= u_{\text{com}}(t) + k_1 f_1(\hat{v}, \hat{r}) - \hat{u}(t). \end{aligned} \quad (15)$$

To follow the speed desired, PID controller is introduced, and the instruction of thruster can be described as

$$T_{\text{com}}(t+1) = k_p e(t) + k_I \int e(t) dt + k_D \frac{de(t)}{dt}. \quad (16)$$

4.2. Vertical Variable Speed Control Strategy. When AUV is diving in the vertical plane, the states involved are AUV speed u , heave velocity w , and pitch velocity q . The most important indexes when AUV is diving are heaving time and overshoot. To ensure the stability of heave, the range of pitch is set to $|\theta| \leq 25^\circ$. According to the 6-DOF model of AUV, we know that when the lever rudder δ_s is fixed, the pitch will be controlled only by AUV speed; therefore, we should control the range of AUV speed so that the pitch can be kept in the allowed range. Hence, at the beginning of the diving stage, u should be kept at a high value so that it can reach maximum pitch with little time; when the desired pitch or deep has been reached, the low speed is designed so that it can decrease the overshoot and keep the pitch stability.

To judge whether the current pitch has reached the bound, the estimation pitch $\hat{\theta}$ is introduced as one of the speed control system states. According to the analysis above, the vertical variable speed control strategy which is driven by AUV states is designed as follows:

$$u_{\text{com}}(t+1) = u_{\text{com}}(t) + k_2 f_2(\hat{w}, \hat{q}, \hat{\theta}), \quad (17)$$

where f_2 can be expressed as

$$\begin{aligned} f_2 &= \begin{cases} \frac{1}{1 + |(\hat{q} - c)/4|^{2 \text{sign}(\hat{w} - \varepsilon_2)}}, & |\Delta\theta| > \varepsilon_1, \\ 0, & |\Delta\theta| \leq \varepsilon_1, \end{cases} \\ \Delta\theta &= \begin{cases} \hat{\theta} - \theta_{\min}, & \hat{w} < -\varepsilon_2, \\ \theta_{\max} - \hat{\theta}, & \hat{w} > \varepsilon_2, \\ 0, & \text{otherwise,} \end{cases} \end{aligned} \quad (18)$$

where θ_{\min} and θ_{\max} , respectively, represent the minimum and maximum value of pitch when AUV is diving in vertical plane; c , ε_1 , and ε_2 are constants. Generally, $\theta_{\min} = -25^\circ$ and $\theta_{\max} = 25^\circ$.

Combined with (17)-(18), the speed error in vertical plane can be calculated as

$$\begin{aligned} e(t) &= u_{\text{com}}(t+1) - \hat{u}(t) \\ &= u_{\text{com}}(t) + k_2 f_2(\hat{w}, \hat{q}, \hat{\theta}) - \hat{u}(t). \end{aligned} \quad (19)$$

Similarly, the instruction of thruster when AUV is diving can be calculated by (16).

4.3. Speed Kept Control Strategy. Considering the bound of the thruster, the task needs, and the security restricts of AUV, the maximum and minimum value of AUV speed should be set. When the speed of AUV accelerates to the maximum value or decreases to the minimum value, the speed should be kept at this value. It is assumed that

$$u_{\text{com}}(t+1) = u_0, \quad (20)$$

where $u_0 = u_{\max}$ and $u_0 = u_{\min}$ are constants. Then the error of AUV speed can be calculated as

$$e(t) = u_{\text{com}}(t+1) - \hat{u}(t) = u_0 - \hat{u}(t). \quad (21)$$

The calculation of thruster instruction is similar to (16).

4.4. Switching Law Design. To realize the stability switching of the speed control strategies designed at a previous section, in this section, our objective is to design a suitable switching law. The switching signal value can be calculated according to the states of AUV, and then the instruction of thruster can be calculated based on the current speed error. Assume that the expression of switching system is as follows:

$$u_{\text{com}}(t+1) = A_{\sigma(t)} \chi(t), \quad (22)$$

where $\chi(t) = [\hat{u}(t), \hat{w}(t), \hat{q}(t), \hat{r}(t), \hat{\theta}(t)]^T$ are the states of switching system; $\sigma(t) \in M$ is switching signal; $M = \{1, \dots, n\}$ is the set of switching signal; $A_i \in R^{n \times n}$, $i \in M$ is coefficient matrix. We can know from the previous section that $M = \{1, 2, 3\}$.

Assume that w_i is the weight of subsystem A_i in the whole system. According to [21], we can know that if each subsystem A_i ($i = 1, 2, 3$) is Hurwitz matrix and average matrix A_0 is

$$A_0 = \sum_{i=1}^3 w_i A_i, \quad (23)$$

then A_0 is Hurwitz matrix too. Solve the Lyapunov function as follows:

$$A_0^T P + P A_0 = -I_n, \quad (24)$$

where $P = P^T > 0$. Define

$$Q_i = A_i^T P + P A_i, \quad i = 1, \dots, n; \quad (25)$$

then, for any initial state $x(t_0) = x_0$, the first switching sequence can be defined as

$$\sigma(t_0) = \arg \min \{x_0^T Q_1 x_0, \dots, x_0^T Q_n x_0\}, \quad (26)$$

where $\arg \min$ represents the minimum order value of indicators; if there are multiple minimum indicators, then it is the minimum number of subsystem. By recursion, we can define other switching sequences as

$$t_{k+1} = \inf \{t > t_k : x^T(t) Q_{\sigma(t_k)} x(t) > 0\},$$

$$\sigma(t_{k+1}) = \arg \min_{i=1, \dots, n} \{x^T(t_{k+1}) Q_i x(t_{k+1})\}, \quad (27)$$

$$k = 0, 1, \dots$$

From (27), we obtain that the switching law designed calculated switching signal based on the variable system states $x(t)$, and the minimum value of $x^T(t_{k+1}) Q_i x(t_{k+1})$, $i = 1, \dots, 6$ represents the maximum weight of speed control subsystem.

5. Simulation Results

In this section, the proposed speed control method designed on a simulation case was illustrated. The case under consideration is ADCP/SSS observation, which is one of the typical tasks of AUV. The simulation environment is MATLAB and with a full nonlinear model of AUV. Set the initial states and position in fixed frames as zero; the maximum of propulsion thrust is 2000 N; the largest rudder angle is 30° ; $u_{\max} = 2.5$ m/s; and $u_{\min} = 1.8$ m/s. To prevent the heave angle which is too large and affect the ability of AUV, the maximum pitch is set to be 25° . The unvarying current is set to be heading x -axis with 0.25 m/s.

Conventional control system of AUV contains three forward channels, namely, the control of AUV speed, the control of AUV heading, and the control of AUV diving. The actuators are propeller, horizontal rudder, and vertical rudder, respectively. Combined with the T-S fuzzy observer designed, the AUV speed control system is constructed as in Figure 2, where \hat{x}' is the vector of AUV states estimated by T-S observer; u_d is the desired value which is calculated by the speed switching control strategies designed; u_w is the current speed, which is set to be heading east and with 0.25 m/s.

Assume that the depth of sea area is 200 m. Consider that the best range of ADCP is 30–40 m, so the desired depth of observation is 170 m. After diving to the specified depth, the AUV is ordered to keep this depth and sail along comb-shaped track. Considering the range of SSS, the comb

TABLE 1: Programming points coordinates (m).

X	0	100	300	300	450	450	600	600
Y	0	0	0	1000	1000	0	0	1000
Z	0	0	170	170	170	170	170	170
X	750	750	900	900	1050	1050	1150	1350
Y	1000	0	0	1000	1000	0	0	0
Z	170	170	170	170	170	170	170	0

TABLE 2: Comparison of overlength (m).

Overshoot term	Speed kept by PID controller	Thrust kept	Speed switching control driven by AUV states
Diving overlength	12.33	10.73	0
Surging overlength	1.50	1.50	0.72
The maximum overlength along x -axis	34.96	20.31	10.28
The maximum overlength along y -axis	20.02	12.86	7.71

interval is set to be 150 m, and y -axis distance is 1 km. The programming points coordinates are shown in Table 1.

Three speed control methods are considered in the simulation, which are speed kept by PID controller, propulsion thrust kept, and speed switching control driven by AUV states, respectively. The simulation results are shown in Figures 3–10. Figure 3 shows the spatial trajectory tracking response under three speed control methods, where S is the start point and E is the end of the mission. Table 2 presents the overlength value under different speed control methods. Combining Figure 3 and Table 2, we can see that compared with other two speed control strategies, under the method proposed in this paper, the trajectory tracking overshoot is decreased; the tracking error is smaller, and the diving angle is kept at the maximum value.

The real AUV speed response is presented in Figure 4. As can be seen, when AUV is diving or surging, under the proposed speed control method, the AUV speed will decrease quickly as the feedback of AUV states. However, under the other two speed control methods, the AUV speed is kept at a stable value. Especially, under the speed kept by PID controller, as the feedback of the speed error, at the beginning of diving or surging, the propulsion thrust will increase. Combined with Figures 5, 6, and 7, we can clearly see that under the proposed method the swaying velocity and heel angle are smaller and under the speed kept by PID controller, swaying velocity and heel angle are higher.

From Figures 8 and 9, we can see that the yaw angle changes smoothly and with no overshoot, it can also be seen from Figure 1. The pitch response can be seen from Figure 10 that under the proposed method the pitch reaches maximum value quickly so as to dive more efficiently. Otherwise, From Figures 4–10, we can see that using the approach proposed in this paper can make AUV work more efficiently.

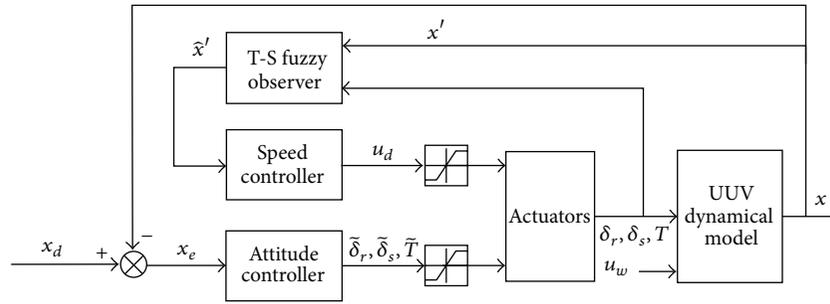
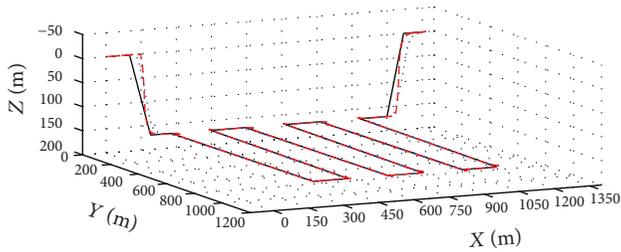
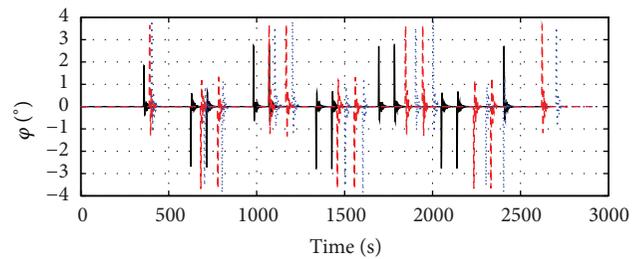


FIGURE 2: Structure of control system.



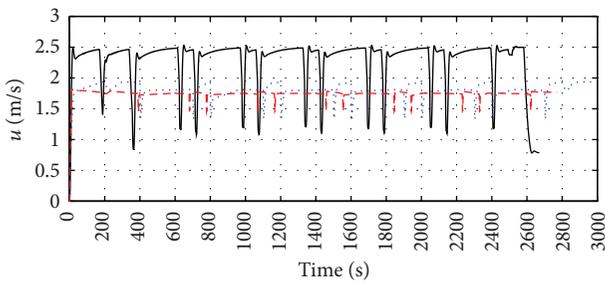
- Speed kept by PID controller
- Thrust kept
- Speed switching control driven by UUV states

FIGURE 3: Spatial trajectory tracking response.



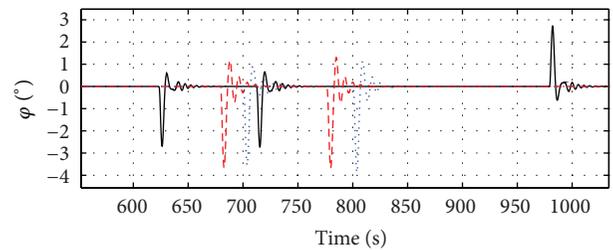
- Speed switching control driven by UUV states
- Speed kept by PID controller
- Thrust kept

FIGURE 6: Heel response of AUV.



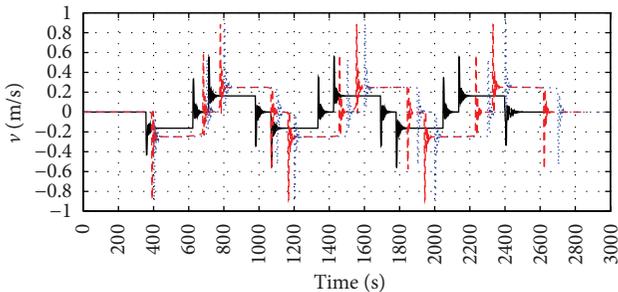
- Speed switching control driven by UUV states
- Speed kept by PID controller
- Thrust kept

FIGURE 4: AUV speed response.



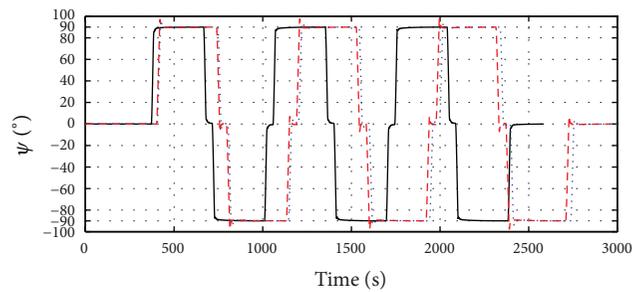
- Speed switching control driven by UUV states
- Speed kept by PID controller
- Thrust kept

FIGURE 7: Zoomout of heel.



- Speed switching control driven by UUV states
- Speed kept by PID controller
- Thrust kept

FIGURE 5: Swaying velocity response.



- Speed switching control driven by UUV states
- Speed kept by PID controller
- Thrust kept

FIGURE 8: Yaw response of AUV.

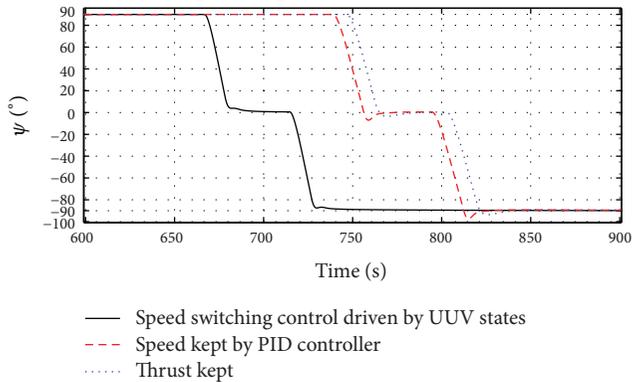


FIGURE 9: Zoomout of yaw.

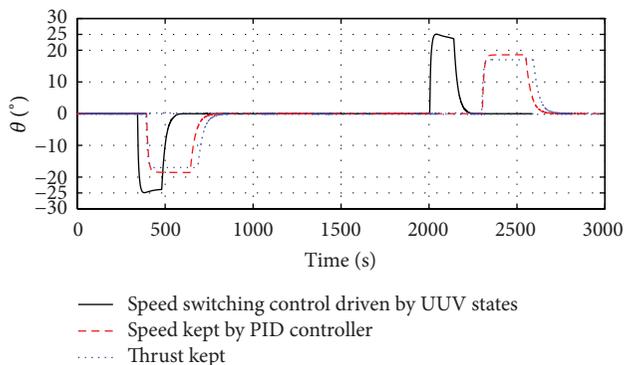


FIGURE 10: Trim response of AUV.

6. Conclusion

In this paper, a speed switching control method for AUV has been proposed, which is driven by AUV's states. Through designed T-S fuzzy observer to estimate the states of AUV, several speed control strategies were put forward, and an appropriate switching law is given to switch the speed control strategies designed in real time according to the states estimated. Finally, ADCP/SSS observation case was introduced to illustrate the effectiveness of the proposed speed control method. The varying currents and other unmodeled dynamics of AUV will be considered in future work.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities, under Grant HEUCFX041401 and the National Natural Science Foundation of China, under Grant 51309067/E091002. The authors also would like to thank the editor and two reviewers for their helpful comments.

References

- [1] Y.-R. Xu and K. Xiao, "Technology development of autonomous ocean vehicle," *Acta Automatica Sinica*, vol. 33, no. 5, pp. 518–526, 2007.
- [2] K. Ishaque, S. S. Abdullah, S. M. Ayob, and Z. Salam, "A simplified approach to design fuzzy logic controller for an underwater vehicle," *Ocean Engineering*, vol. 38, no. 1, pp. 271–284, 2011.
- [3] E. Cavallo, R. C. Michelini, and V. F. Filaretov, "Path guidance and attitude control of a vectored thruster AUV," in *Proceedings of the ASME 7th Biennial Conference on Engineering Systems Design and Analysis*, pp. 19–22, Manchester, UK, July 2004.
- [4] M. Caccia and G. Veruggio, "Guidance and control of a reconfigurable unmanned underwater vehicle," *Control Engineering Practice*, vol. 8, no. 1, pp. 21–37, 2000.
- [5] A. H. Izadparast and J. M. Niedzwecki, "Estimating the potential of ocean wave power resources," *Ocean Engineering*, vol. 38, no. 1, pp. 177–185, 2011.
- [6] M. Ishitsuka and K. Ishii, "Control of an underwater manipulator mounted for an AUV considering dynamic manipulability," *International Congress Series*, vol. 1291, pp. 269–272, 2006.
- [7] A. Alessandri, M. Caccia, and G. Veruggio, "Fault detection of actuator faults in unmanned underwater vehicles," *Control Engineering Practice*, vol. 7, no. 3, pp. 357–368, 1999.
- [8] S. McPhail, "Autosub6000: a deep diving long range AUV," *Journal of Bionic Engineering*, vol. 6, no. 1, pp. 55–62, 2009.
- [9] T. F. Curado, A. A. Pedro, and A. Pascoal, "Nonlinear adaptive control of an underwater towed vehicle," *Ocean Engineering*, vol. 37, no. 13, pp. 1193–1220, 2010.
- [10] F. Wang, B. Tian, W. Fan, Q. Zong, and J. Wang, "Adaptive high order sliding mode controller design for hypersonic vehicle with flexible body dynamics," *Mathematical Problems in Engineering*, vol. 2013, Article ID 749689, 11 pages, 2013.
- [11] J. Zhou, Z. Tang, and H. Zhang, "Spatial path following for AUVs using adaptive neural network controllers," *Mathematical Problems in Engineering*, vol. 2013, Article ID 357685, 9 pages, 2013.
- [12] M. Caccia, G. Casalino, R. Cristi, and G. Veruggio, "Acoustic motion estimation and control for an unmanned underwater vehicle in a structured environment," *Control Engineering Practice*, vol. 6, no. 5, pp. 661–670, 1998.
- [13] C. Mohammed and R. K. Hamid, "Robust observer design for unknown inputs Takagi-Sugeno models," *IEEE Transactions on Fuzzy Systems*, vol. 21, no. 1, pp. 158–164, 2013.
- [14] L. I. Ye, J. P. Yong, and W. A. N. Lei, "Stability analysis on speed control system of autonomous underwater vehicle," *China Ocean Engineering*, vol. 23, no. 2, pp. 345–354, 2009.
- [15] J. Petrich and D. J. Stilwell, "Robust control for an autonomous underwater vehicle that suppresses pitch and yaw coupling," *Ocean Engineering*, vol. 38, no. 1, pp. 197–204, 2011.
- [16] P. Encarnação and A. Pascoal, "3D path following for autonomous underwater vehicle," in *Proceedings of the 39th IEEE Conference on Decision and Control*, pp. 2977–2982, Sydney, Australia, December 2000.
- [17] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 15, no. 1, pp. 116–132, 1985.
- [18] S. B. Nagesh, Z. Lendek, and A. A. Khalate, "Adaptive fuzzy observer and robust controller for a 2-DOF robot arm," in *IEEE International Conference on Fuzzy Systems*, pp. 1–7, Brisbane, Australia, June 2012.

- [19] Z. Lendek, J. Lauber, T. M. Guerra, R. Babuška, and B. de Schutter, "Adaptive observers for TS fuzzy systems with unknown polynomial inputs," *Fuzzy Sets and Systems*, vol. 161, no. 15, pp. 2043–2065, 2010.
- [20] E. Sebastián and M. A. Sotelo, "Adaptive fuzzy sliding mode controller for the kinematic variables of an underwater vehicle," *Journal of Intelligent and Robotic Systems*, vol. 49, no. 2, pp. 189–215, 2007.
- [21] Z. Sun, "Stabilizing switching design for switched linear systems: a state-feedback path-wise switching approach," *Automatica*, vol. 45, no. 7, pp. 1708–1714, 2009.

Research Article

A Study of Bending Mode Algorithm of Adaptive Front-Lighting System Based on Driver Preview Behavior

Zhenhai Gao^{1,2} and Yang Li¹

¹ State Key Laboratory of Automobile Simulation and Control, Jilin University, Changchun 130022, China

² State Key Laboratory of Vehicle NVH and Safety Technology, Changan Automobile Holding, Chongqing 401120, China

Correspondence should be addressed to Yang Li; 89271667@qq.com

Received 25 February 2014; Accepted 4 April 2014; Published 4 May 2014

Academic Editor: Hamid Reza Karimi

Copyright © 2014 Z. Gao and Y. Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The function of adaptive front-lighting system is to improve the lighting condition of the road ahead and driving safety at night. The current system seldom considers characteristics of the driver's preview behavior and eye movement. To solve this problem, an AFS algorithm modeling a driver's preview behavior was proposed. According to the vehicle's state, the driver's manipulating input, and the vehicle's future state change which resulted from the driver's input, a dynamic predictive algorithm of the vehicle's future track was established based on an optimal preview acceleration model. Then, an experiment on the change rule of the driver's preview distance with different speeds and different road curvatures was implemented with the eye tracker and the calibration method of the driver's preview time was established. On the basis of these above theories and experiments, the preview time was introduced to help predict the vehicle's future track and an AFS algorithm modeling the driver's preview behavior was built. Finally, a simulation analysis of the AFS algorithm was carried out. By analyzing the change process of the headlamp's lighting region while bend turning which was controlled by the algorithm, its control effect was verified to be precise.

1. Introduction

AFS (adaptive front-lighting system) is a front-lighting system that can change the light pattern and illumination area according to the vehicle's state such as the velocity, the steering wheel angle, and road environment to light the road ahead effectively so as to reduce accidents at night. According to AFS productions of international companies and research institutes' research findings and directions currently, the classic algorithm of AFS can be divided into the following four categories.

(1) *The AFS Algorithm Based on the ECER123 Regulation [1]*. The basic demand of this algorithm was to control the deflection angle of the headlamp at an allowed range of the regulation. Strambersky et al. (European Lighting Technology Centre of Visteon Company) used Ackerman steering geometry principle to calculate the headlamp's deflection angle in the regulation's allowed range [2]. The algorithm

based on the regulation needed to be improved in accuracy and trajectory prediction.

(2) *The AFS Algorithm Based on Brake Safety*. The algorithm used the stopping sight distance to calculate arc length and the stopping sight distance was larger than brake distance to make sure that the driver would observe the obstacle in front. Huang, Lifu and Qian, Yu, Deng, Mclaughlin et al., and Rong et al. had used similar methods in their studies [3–8]. The design based on brake safety could ensure the safety, but the headlamp's lighting location had a lag relative to the vehicle's location.

(3) *The AFS Algorithm Based on Early Lighting*. When a car was driven on the curve, the headlamp should be lighted on the location where the car would arrive at after a while. This kind of research solved the problem to some extent that the headlamp illumination direction had a lag compared to the driving direction [9, 10]. This kind of AFS can be divided into nonpredictive AFS and predictive AFS.

(a) *Nonpredictive AFS*. Koji Ishiguro and Yuji Yamada who were in Denso and Toyota in Japan studied the relationship between vehicles' speed and drivers' fixation distance during bend driving in the day time. They concluded that the fixation distance got larger while the speed increased. By studying drivers' reaction time in dangerous circumstances such as collision, they regarded $t = 3$ s as an appropriate time which they used in their AFS algorithm [11]. Nonpredictive AFS could improve the lighting condition of the road ahead to some extent, but it could only use the current steering wheel angle and velocity to estimate the vehicle's transient current trajectory and then calculate the headlamp's deflection angle. When the road's curvature changed a lot, the beam might deflect too slow in the corner and even deflected to a wrong direction at an S-curve.

(b) *Predictive AFS*. By GPS, electronic map, CCD which perceived the traffic environment information including vehicle locations, road types, lane numbers and road curvatures, and vehicles' state information which was used to predict cars' future track, the P-AFS controlled the headlamp in advance. Ibrahim of Visteon applied for the patent which was called "Predictive adaptive front lighting algorithm for branching road geometry," Patent number US 7558822 B2 [10]. They matched the GPS information with the database of electronic map to confirm the vehicle's current location and predict the road's information such as road type, lane numbers, and curvature and road branches and finally calculated the headlamp deflection angle by P-AFS algorithm. Kim et al. in Korea (2010) proposed a new method of P-AFS which was called curvature estimated swivel in their paper [12]. By using the existing devices such as LDWs to replace the GPS to predict road curvature, the headlamp's deflection angle was calculated. Predictive AFS solved the nonpredictive AFS's error problem to some extent when the road's curvature changed a lot, but it did not take the driver's visual characteristics into consideration.

(4) *The AFS Algorithm Based on the Driver's Vision in the Corner*. Many researchers have studied the driver's sight movement in a bend or crossing [13–15]. Although these researches aimed to study the driver's fixation behavior while not providing information for the development or design of AFS, they had some guiding significances for the design of AFS.

On the basis of these experiments, Young et al. used the regular pattern of heads' turning angle at daytime and night to guide the design of AFS [16], but the paper did not mention the specific algorithm of AFS and the precision could not be ensured because of the limit of measuring devices.

Considering the regulation and safety, the algorithm of AFS currently estimated the vehicles' future driving track according to vehicles' moving states, traffic environment, and drivers' wheel input. Then the AFS algorithm controlled the headlamp's deflection angle to illuminate the estimated driving track in advance. But most of these algorithms assumed that the speed and the wheel angle would not change and assumed the driving track as a circle. According to Zhang et al.'s study on vehicles' lateral moving characteristics [17–19],

the possible variation of vehicles' longitude and lateral speed was not considered which would also result in the change of driving track. In some circumstances such as in a curve entry, in a curve exit in an irregular road, or the driver's acceleration or deceleration to pass a varying curve, the algorithm could not predict the future driving track well and the effect of AFS was not very well.

Meanwhile, as the most important active safety device, the headlamp's main effect is to provide illumination for the road ahead. Hence, the AFS should satisfy the safety demand, ensure that the driver's concerned region is fully illuminated, and must not influence the driver's fixation behavior. Its illumination effect also has a direct relationship with the driver's eye comfort. Compared to a good environmental illumination condition such as in the daytime, it should not additionally increase the driver's visual fatigue. In this aspect, although some researches have used the statistic rule of drivers' visual field to control the headlamp's deflection, in the practical application, this method which was totally based on the statistic rule cannot ensure the safety.

Based on the considerations above, an AFS algorithm considering the driver's preview behavior was proposed. The vehicle's kinematics and dynamics characteristics were used comprehensively to calculate the vehicle's future track which was more reliable than the method purely based on the vehicle's state information the headlamp's deflection lag and direction error was avoided. Under the premise of ensuring safety, parameters of the driver's fixation behaviors were introduced to increase the measuring precision of the driver's visual statistical rule by using the eye tracker and the headlamp's illumination is more in compliance with the driver's fixation behavior. This paper proposes an original technology route; the work was first carried out in early 2010, and it has obtained national invention patents [20].

2. AFS Algorithm Considering the Driver's Preview Behavior

2.1. AFS Algorithm. By modeling the driver's preview behavior at a bend, the vehicle's future track in a period of time was predicted according to the vehicle's current state and the vehicle's steady state response characteristics to the driver's wheel, gas pedal, or brake pedal input. Simultaneously, the driver's fixation location on the future track was determined according to drivers' preview behavior rule (the rule of fixation location or preview spot) at real bends; then the headlamp's deflection angle was controlled and the location was illuminated effectively. The technical route of the algorithm was shown in Figure 1.

The algorithm simulated the driver's preview behavior at a large curvature bend and an integral algorithm to predict the vehicle's track was proposed. The algorithm was based on the hypothesis of steady preview and dynamic correction [21]. According to the vehicle's current state and its possible change which resulted from the driver's inputs, the vehicle's future track was predicted by the algorithm [22]. As Figure 2 showed, the vehicle's coordinate system at current time t was the reference coordinate system. The preview time t_p was

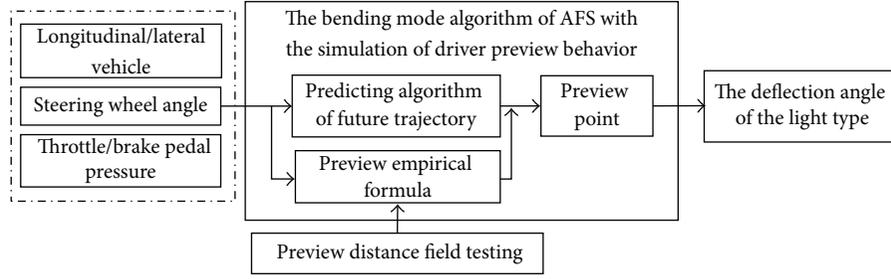


FIGURE 1: The specific technical route.

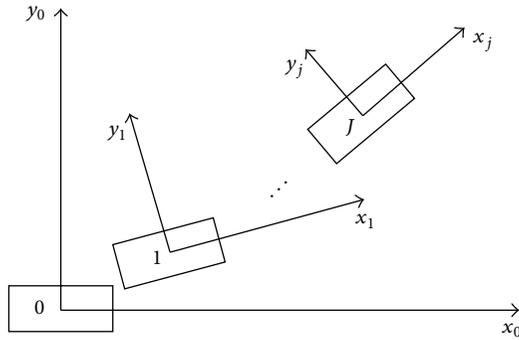


FIGURE 2: The coordinate system.

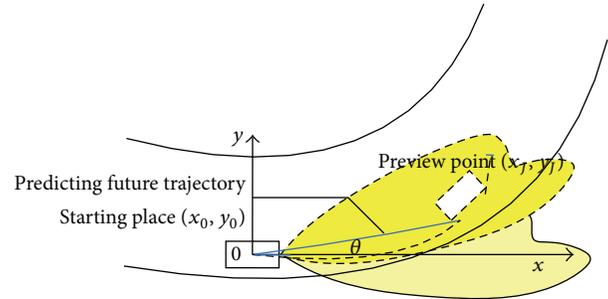


FIGURE 3: Adaptive front-lighting system deflection angle calculation mathematical model.

evenly divided into J pieces ($\Delta t_p = t_p/J$). During the preview time, the longitudinal and lateral acceleration which resulted from the driver's wheel or gas pedal input were the same.

Because every time piece is quite short, the mutual effect of the vehicle's longitudinal and lateral movement can be neglected. On the basis of the vehicle's state (location, velocity, and acceleration) at the initial moment of the time piece, the vehicle's state at the end of the time piece was calculated using the rigid body kinematics principle [23]. Then from the moment t , every time piece's vehicle state could be calculated. As (1) showed (at the time piece j , the initial moment is t_{j-1} and the end moment is t_j), the combination of the location at every time formed the vehicle's future track in t_p time:

$$\begin{aligned} \begin{bmatrix} \dot{x}_j \\ \dot{y}_j \end{bmatrix} &= \begin{bmatrix} \dot{x}_{j-1} \\ \dot{y}_{j-1} \end{bmatrix} + \begin{bmatrix} \ddot{x}_j \\ \ddot{y}_j \end{bmatrix} \times \Delta t_p, \\ \begin{bmatrix} x_j \\ y_j \end{bmatrix} &= \begin{bmatrix} x_{j-1} \\ y_{j-1} \end{bmatrix} + A_{j-1,0} \left(\begin{bmatrix} \dot{x}_{j-1} \\ \dot{y}_{j-1} \end{bmatrix} \times \Delta t_p + \frac{1}{2} \begin{bmatrix} \ddot{x}_j \\ \ddot{y}_j \end{bmatrix} \times \Delta t_p^2 \right). \end{aligned} \quad (1)$$

The coordinate transformation matrix is

$$A_{j0} = \begin{pmatrix} \cos \varphi_j & -\sin \varphi_j \\ \sin \varphi_j & \cos \varphi_j \end{pmatrix}. \quad (2)$$

φ is path angle which is calculated by

$$\varphi_j = \varphi_{j-1} + \frac{\ddot{y}_j}{\dot{x}_j} \cdot \Delta t_p. \quad (3)$$

$j = 1, 2, \dots, J$; (x_j, y_j) , (\dot{x}_j, \dot{y}_j) , and (\ddot{x}_j, \ddot{y}_j) were the vehicle's location, the vehicle's velocity, and the vehicle's steady acceleration at time t_j ; φ_j was the vehicle's course angle at time t_j . The vehicle's coordinate system at current time t (t_0) was the reference coordinate system. During the preview time, the longitudinal and lateral acceleration which resulted from the driver's wheel or gas pedal input were the same.

Then the vehicle's mass centre coordinate (x_j, y_j) after the preview time t_p can be calculated at the coordinate system of current time t . As Figure 3 showed, the point was the driver's preview location at current time t and the headlamp's deflection angle θ which was the angle between the vehicle's predicted location and the vehicle's longitudinal axis was calculated by

$$\theta = \arctan \frac{y_j - y_0}{x_j - x_0}. \quad (4)$$

Based on the above theory, the measurement of the preview distance for real drivers was conducted. By the advanced eye tracker, the driver's fixation behavior was recorded when the driver drove through the bend. The driver's gazing direction was analyzed and the average value of the preview distance was calculated. Then an empirical equation of the relationship between preview distance, preview time and velocity, and road curvature was proposed and the preview time for the AFS algorithm was modified.

TABLE 1: The experiment velocity and the curvature radius.

Radius of curvature (m)	Velocity (km/h)
20	10, 15, 20, 25, 30
30	20, 30, 40
40	20, 30, 40

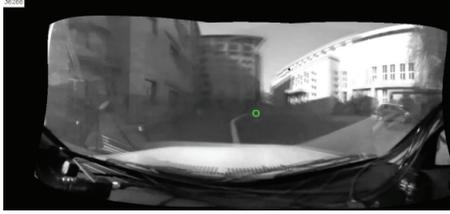


FIGURE 4: Driver's gaze point position comparison chart when "being close to the bend."

3. The Experiment Measurement and Change Rule Study on the Driver's Preview Distance

3.1. Experiment Measurement

3.1.1. *Experiment Equipment.* SmartEye Pro eye tracker (sample frequency: 60 Hz), A JETTA car made by FAW-Volkswagon, a 12 V spare battery, meter ruler, the adhesive tape to make the lane, and others.

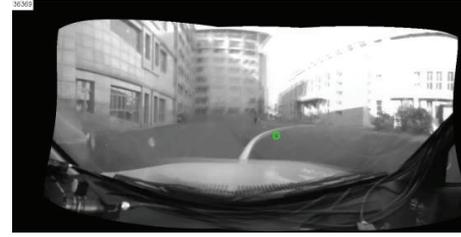
3.1.2. *Experiment Conditions.* Eight drivers with normal vision. Three quarter-circle lanes with curvature radius of 20 m, 30 m, and 40 m were used as the experiment road. With every lane, the experiment was done for three times repeatedly. Before the formal experiment, every driver did 10-minute exercise. The experiment velocity and the curvature radius were shown in Table 1.

3.1.3. *The Definition of Fixation Location.* The SmartEye Pro eye tracker recorded the driver's raw eye movement data. According to the thresholds of eye movement parameters set up by the users, the fixation and saccade behaviors were confirmed. Meanwhile, by the corresponding software, the video of the driving scene was replayed and the fixation location (the green circle in Figure 4) was spotted on the driving scene.

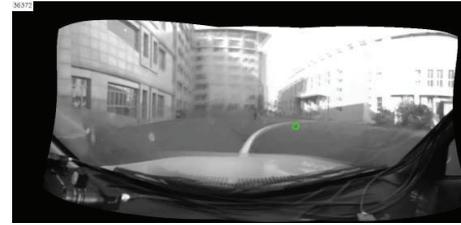
The threshold settings of eye movement were as follows:

- (1) the threshold of visual angle deviation: 2 deg.;
- (2) the eyeball movement velocity threshold during fixation: 15°/s, which was the highest velocity of eyeball movement during a fixation;
- (3) the eyeball movement velocity threshold during saccade: 35°/s, which was the lowest velocity of eyeball movement during a saccade;
- (4) the fixation duration threshold: 200 ms.

According to the road environment, the state change of the vehicle, the fixation location (visual angle), and its change



(a)



(b)

FIGURE 5: Driver's gaze point position comparison chart before and after into the corner: (a) is before into the corner, and (b) is after.

in the driving scene video, the entire driving process was divided into four sections: "straight lane," "being close to the bend," "entering the bend," and "being out of the bend."

"Straight lane": in this section, the entire bend was shown in the driver's visual field. Repeated experiments showed that the driver would sweep the bend and form a rough impression of the available track; then the fixation location would be a few meters away from the vehicle as shown in Figure 4.

In the section of "being close to the bend," because the vehicle's movement state had not changed, the nonpredictive AFS algorithm could not predict the bend ahead and activate the AFS bend lighting function.

"Entering the bend": as shown in Figure 5 (there were three frames intervals between two images which meant the time interval was 0.05 s). There was a saccade in almost all experiments when the driver just drove from the "being close to the bend" section to the "entering the bend" section. It was also shown in the driver's eye movement data that there was a saccade in the moment. This saccade behavior was defined as the boundary between the "being close to the bend" section and the "entering the bend" section. After this saccade, the lane area was defined as the "entering the bend" section.

In the "entering the bend" section, the driver needed to steer and the AFS needed to activate bend lighting function and control the headlamp's deflection according to some rules. So according to the purpose of this experiment and our algorithm's parameters' need, this section was the emphasis of our study.

"Being out of the bend": the driver would brake when he were going to be out of the bend. This section would not be analyzed in detail.

3.1.4. *The Preview Distance Calculation and the Analysis of Its Changing Rule.* The preview distance was calculated in the



FIGURE 6: The eye tracking system's installation location.

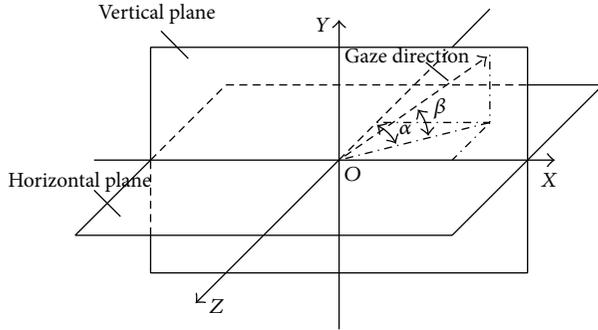


FIGURE 7: The world coordinate system and gaze direction angle.

“entering the entrance” section during the driving process. The fixation point in the front near the centre line of the road was defined as the preview point. The SmartEye Pro system was installed as shown in Figure 6. The definition of the SmartEye Pro world coordinate system was as follows: the connection line between number 2 camera and number 3 camera in SmartEye Pro system was the X-axis of the world coordinate, the midpoint of the connection line was the coordinate origin, and the Y-axis was in the vertical plane. The coordinate system was shown in Figure 7.

As shown above, the angle α between the gaze direction's projection in XOZ plane and the negative direction of the world coordinate's Z-axis (the positive direction of the vehicle coordinate's X direction) was the gaze direction's horizontal direction angle. The angle β between the gaze direction and the horizontal plane was the gaze direction's vertical angle.

The gaze direction's horizontal angle is

$$\alpha = \arctan \frac{\text{GazeDirection}.x}{-\text{GazeDirection}.z}. \quad (5)$$

The gaze direction's vertical angle is

$$\beta = \arcsin \text{GazeDirection}.z. \quad (6)$$

$\text{GazeDirection}.x(y, z)$ was the projection of the unit vector along the gaze direction on the X(Y, Z) axis of the world coordinate system.

According to the gaze direction and the gaze origin in the world coordinate system, the preview point's coordinate could be calculated in the world coordinate system as shown in Figure 8. The distance away from the driver which was the preview distance could be calculated by (7) (the vehicle's roll

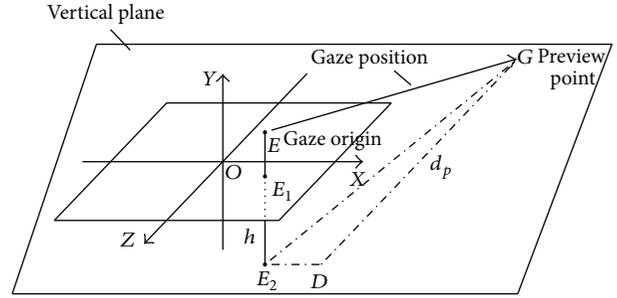


FIGURE 8: Preview distance calculation schematic.

TABLE 2: Preview distance data under 40 m bend radius.

Data operation	Bend radius (20 m)				
	10 km/h	15 km/h	20 km/h	25 km/h	30 km/h
Mean	9.6253	9.7248	9.8515	10.0431	10.1059
Variance	1.0292	0.9282	0.8827	0.8319	1.2074
Maximum	7.5492	7.7634	8.0599	8.0124	8.2775
Minimum	12.3558	12.0545	12.1912	12.4505	12.1334

TABLE 3: Preview distance data under 20 km/h velocity.

Data operation	Velocity (20 km/h)	
	30 m	40 m
Mean	9.4637	11.3121
Variance	1.8896	1.4886
Maximum	6.6242	7.6275
Minimum	12.7901	13.9346

and pitch movement were neglected). The detailed data of the preview distance was shown in Tables 2 and 3:

$$d_p = (\text{GazeOrigin}.y + h) \frac{\text{GazeDirection}.z}{\text{GazeDirection}.y}. \quad (7)$$

h is the vertical height of the world coordinate system's origin away from the ground; $\text{GazeOrigin}.y$ is the y coordinate of the gaze origin; $\text{GazeDirection}.z$ is the z coordinate of the unit vector along the gaze direction; $\text{GazeDirection}.y$ is the y coordinate of the unit vector along the gaze direction.

When the road curvature radius was 20 m, the preview distance increased slowly from 9,6253 m to 10.1059 m with the adding of the velocity. When the velocity was constantly 20 km/h, the driver's preview distance increased too while the road curvature radius increased from 30 m to 40 m. Noh et al. had used the preview time and the response of neuromuscular system as main human factors to study the preview distance at different velocities and different road curvatures whose conclusion was similar to ours [24].

The experiment's result was consistent with the previous test's conclusion. With the consideration of the characteristics of the preview distance's distribution and change rule, the preview distance-velocity regression analysis was done to get the preview distance-velocity empirical equation (9). At the speed of 20 km/h, according to the preview distance data with

the curvature radius of 30 m and 40 m, the preview distance-curvature radius empirical equation (10) was gotten by liner fitting. The fitting line is

$$y = 0.025x + 9.358. \quad (8)$$

y is the preview distance; x is velocity; R is the correlation coefficient, $R^2 = 0.98082$, and the observed value $r = 0.99036$. $n = 5$; with the significance level $\alpha = 0.05$, $r = 0.99036 > r_{0.05}(n-2) = 0.878$; the regression result was obvious and the hypothesis "the velocity change will affect the driver's preview distance" was confirmed. To be consistent with previous described parameters, the empirical equation ($r = 20$ m) was rewritten as follows:

$$d_p = 0.025v + 9.358, \quad r = 20 \text{ m}, \quad (9)$$

where v is velocity, km/h, r is curvature radius, and d_p is the preview distance.

At the speed of 20 Km/h, the empirical equation of the preview distance-curvature radius was as follows:

$$d_p = 0.185r + 1.529, \quad v = 20 \text{ km/h}. \quad (10)$$

According to the relationship between the preview time and the preview distance, $d_p = vt_p$, the empirical equation of the preview time was as follows:

$$t_p = 0.09 + \frac{33.689}{v}, \quad r = 20 \text{ m}, \quad (11)$$

$$t_p = 0.0333r + 0.2752, \quad v = 20 \text{ km/h}. \quad (12)$$

According to (11), the preview time calculated under the speed of 10 km/h, 15 km/h, 20 km/h, 25 km/h, and 30km/h was 3.489 s, 2.336 s, 1.774 s, 1.438 s, and 1.213 s, respectively; the average preview time is 1.70 s. According to (12), the average preview time is 1.44 s. Then (11) and (12) as empirical equations of the preview time simulating the driver's preview behavior were used to replace the fixed preview time t_p in the AFS algorithm to confirm the location of the driver's preview point where the vehicle would arrive after the preview time. Then the headlamp's deflection angle could be calculated.

4. The Simulation Analysis of the AFS Algorithm Simulating the Driver's Preview Behavior

4.1. The Headlamp's Light Distribution. The optical analysis software LucidShape was used to simulate the headlamp's light distribution and acquire the headlamp's isolux curve on the ground. In the following simulation process, the envelope area of the isolux curve at a certain illumination was regarded as the headlamp's irradiation area. According to the relative relationship between the area and the road, the control effect of the algorithm was verified by the headlamp's irradiation effect. With the reference of ECE regulation, a dipped headlight in accordance with the ECE regulation was chosen in our simulation and its optical structure was not changed. The dipped headlight's location settings were as follows.

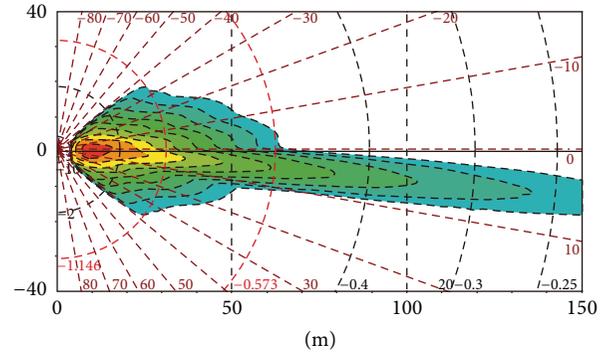


FIGURE 9: The simulation result of the headlamp's illumination on the road.

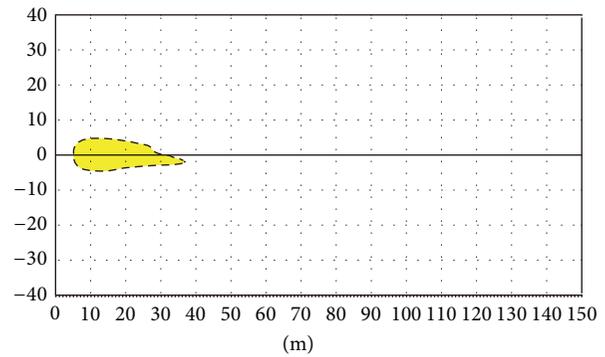


FIGURE 10: The envelope area of 32lx isolux curve.

- (1) The headlamps were installed in front of the vehicle and the headlamps were 1.2 m apart and were symmetrically on both sides of the vehicle's longitudinal axis.
- (2) The installation height of headlamps was 0.65 m.

The simulation result of the headlamp's illumination on the road was shown in Figure 9. The coordinate origin was the midpoint of the vehicle's forefront. The horizontal and vertical coordinates were horizontal and vertical locations relative to the vehicle (unit m); different colors represented different illuminations.

Meanwhile, because the road's curvature radius was relatively small in our simulation, to clearly verify our algorithm's irradiation effect, the envelope area of 32lx isolux curve was drawn in LabVIEW as the headlamp's irradiation area which was shown in Figure 10.

4.2. The Algorithm Verification. The vehicle model, driver model, and road model of CarSim were used in the simulation. The AFS algorithm and the real-time display of the headlamp's deflection angle, beam location, and irradiation effect were realized in LabVIEW. By the cosimulation of CarSim and LabVIEW, the bend driving simulating real environment was achieved. According to the calculated headlamp's deflection angle, irradiation area, and the vehicle's future track, the algorithm's control effect was verified.

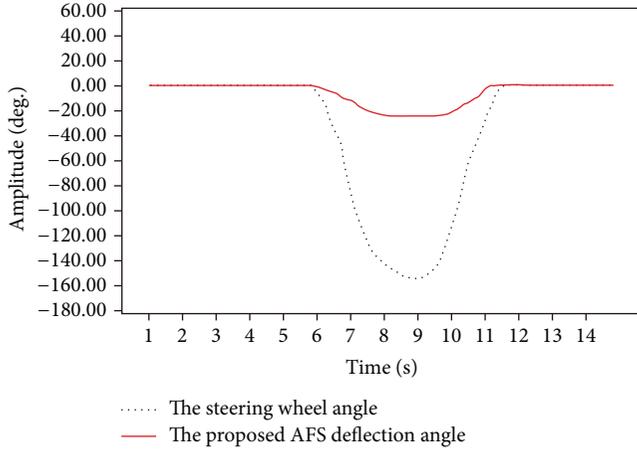


FIGURE 11: The headlamp's deflection angle and steering wheel angle.

4.2.1. Simulation Conditions and Parameters. The vehicle's speed was 30 km/h, the curvature radius of the bend was 20 m, and the preview time t_p (s) was as follows:

$$t_p = 0.09 + \frac{33.689}{v} = 1.213. \quad (13)$$

4.2.2. The Simulation Result Analysis. The headlamp's beam deflection angle of our algorithm was shown in Figure 11. The dotted line represented the wheel's steering angle and the solid line stood for the headlamp's deflection angle. It was shown that the deflection angle calculated by our algorithm could follow the wheel's steering angle very well. In other words, the algorithm could respond to the change of vehicle state quite well and controlled the headlamp's deflection angle. The control effect of the algorithm was shown in Figure 12. The dotted line represented the fixed headlamp beam.

It was shown in Figure 12 that the predicted track of our algorithm was close to the road's centre line which meant that the track prediction's accuracy was quite well. The vehicle's location on the predicted track after the preview time was used to guide the headlamp's deflection. The headlamp was controlled to deflect to the direction of the bend and the headlamp's irradiation area covered most areas of the front bend. When the vehicle was out of the bend, the algorithm could make the headlamp beam back on the nondeflection location to throw light on the front straight line.

4.3. The Contrast Analysis of the Algorithm's Control Effect

4.3.1. The Simulation Conditions

Condition 1. The initial speed was 20 km/h and the driver accelerated to 30 km/h in the time of 11 s–13 s and kept the velocity steady to pass through a half circle bend with the radius of 20 m. The driving strategy was to make sure that the

vehicle has no lateral offset relative to the lane's center line. The time parameter t_p rose to 1.774 s from 1.213 s as follows:

$$t_p = 0.09 + \frac{33.689}{v}. \quad (14)$$

Condition 2. The driver drove through two continuous quarter circles (the radiuses were 30 m and 40 m) at a constant speed of 20 km/h. The driver model used in the simulation was a model built-in CarSim. The time parameter t_p rose to 1.607 s from 1.274 s as follows:

$$t_p = 0.0333r + 0.2752. \quad (15)$$

The contrast algorithm was a nonpredictive algorithm proposed by Ishiguro and Yamada which was relatively mature [11]. The main point of this algorithm is to control the headlamp's deflection to illuminate the location that the vehicle would arrive after 3 s. The contrast simulation result was as follows.

4.3.2. The Simulation Result Analysis

Condition 1. The headlamp's deflection angle of the contrast algorithm and our algorithm was shown in Figure 15. The black dotted line was the wheel's steering angle curve, the solid line was the headlamp's deflection angle curve of our algorithm, and the dash dot line was the headlamp's deflection angle curve of the contrast algorithm.

It was shown in Figure 13 that both algorithms' headlamp deflection angle could follow the driver's steering angle. Relative to our algorithm, the contrast algorithm's headlamp deflection angle was about 20° bigger at the moment 16 s. The control effect of these two algorithms was shown in Figure 14 by the headlamps' irradiation area. The dotted line represented the fixed headlamp beam, the solid line represented the headlamp beam of our algorithm, and the long dashed line represented the headlamp beam of the contrast algorithm.

It was shown in Figure 14 that at the moment of $t = 10.19$ s, the vehicle just arrived at the bend; both the contrast algorithm and our algorithm could predict the vehicle's future track well and make the headlamp deflect to the direction of the bend. Their deflection angles were basically the same. Most areas of the front bend were covered by the central zone of the headlamp's irradiation region which meant that their illumination effects were basically the same.

At the moment of $t = 12.27$ s, the vehicle was in the bend. The headlamp's deflection angle of the contrast algorithm was bigger than our algorithm which meant that its headlamp beam was more close to the inner side of the bend. Although the headlamp had illuminated more regions of the front bend, the control target of the contrast algorithm was to illuminate the location that the vehicle would arrive after 3 s while the driver's preview time was 1.213–1.774 s according to the empirical equation, so the region the driver cared for was a region which was close to the vehicle (the purple line in front of the vehicle) and it was not in the centre

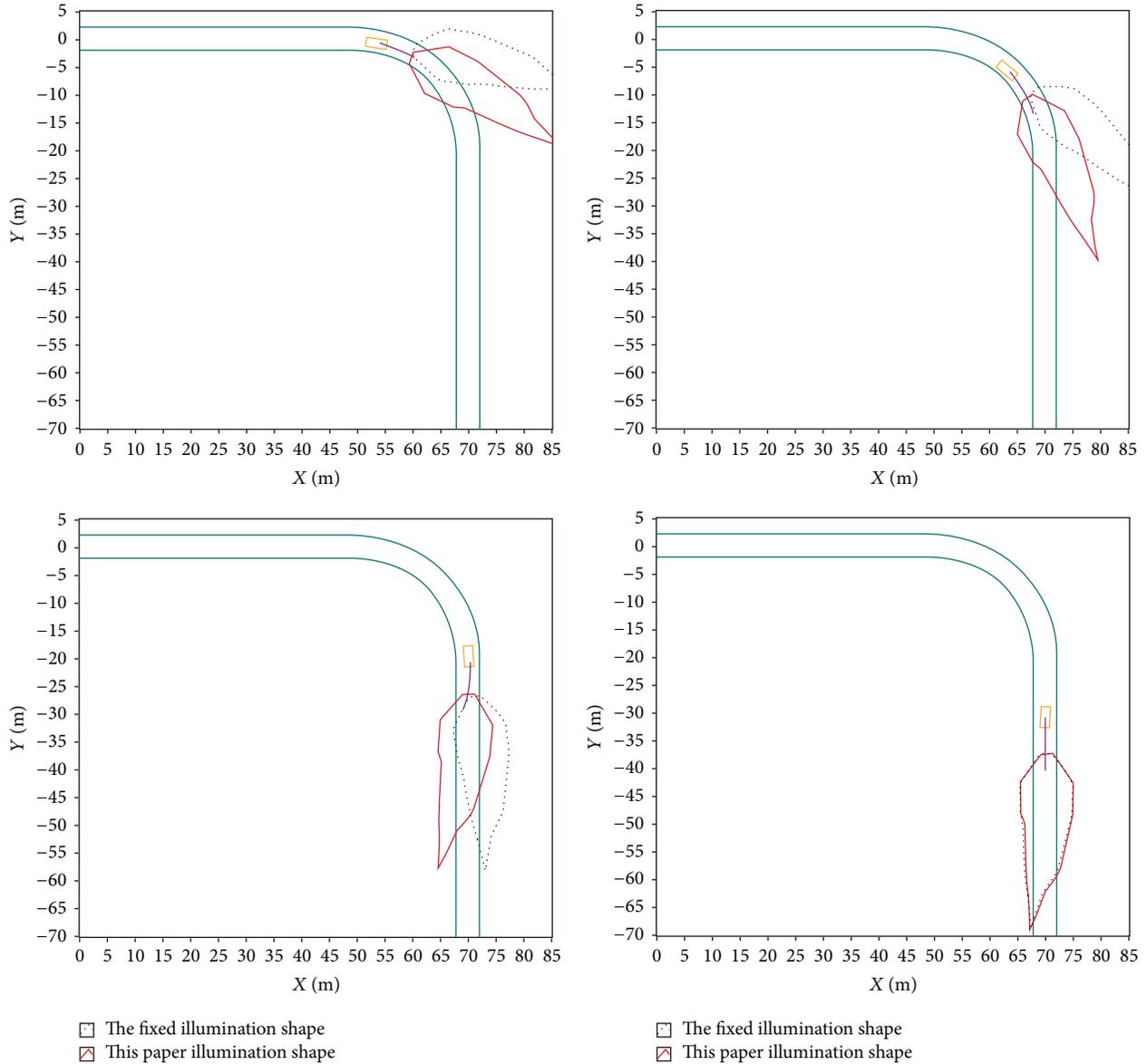


FIGURE 12: The front-lighting illumination area simulation schematic.

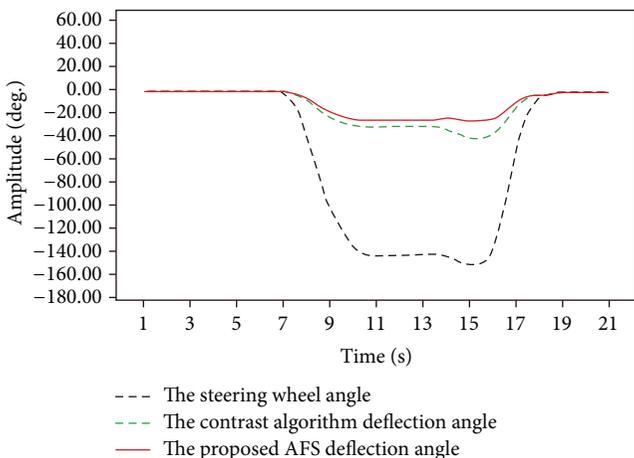


FIGURE 13: The headlamp's deflection angle and steering wheel angle (contrast algorithm and our algorithm diagram 1).

of the contrast algorithm's illumination area which meant that its illumination effect was not well. The future track's prediction and irradiation region of our algorithm were more precise.

At the moment of $t = 16.77$ s, the vehicle was going to be out of the bend. The headlamp's deflection angle of the contrast algorithm was bigger and its illumination effect was not ideal (relative to the fixed beam, its illumination effect could be worse). The future track's prediction of our algorithm was more precise and the headlamp's beam covered most of the front road which meant the illumination effect was relatively well.

At the moment of $t = 18.86$ s, the vehicle was out of the bend. Both the contrast algorithm and our algorithm could quit from the bend illumination mode and control the headlamp back to the nondeflection location to ensure the straight

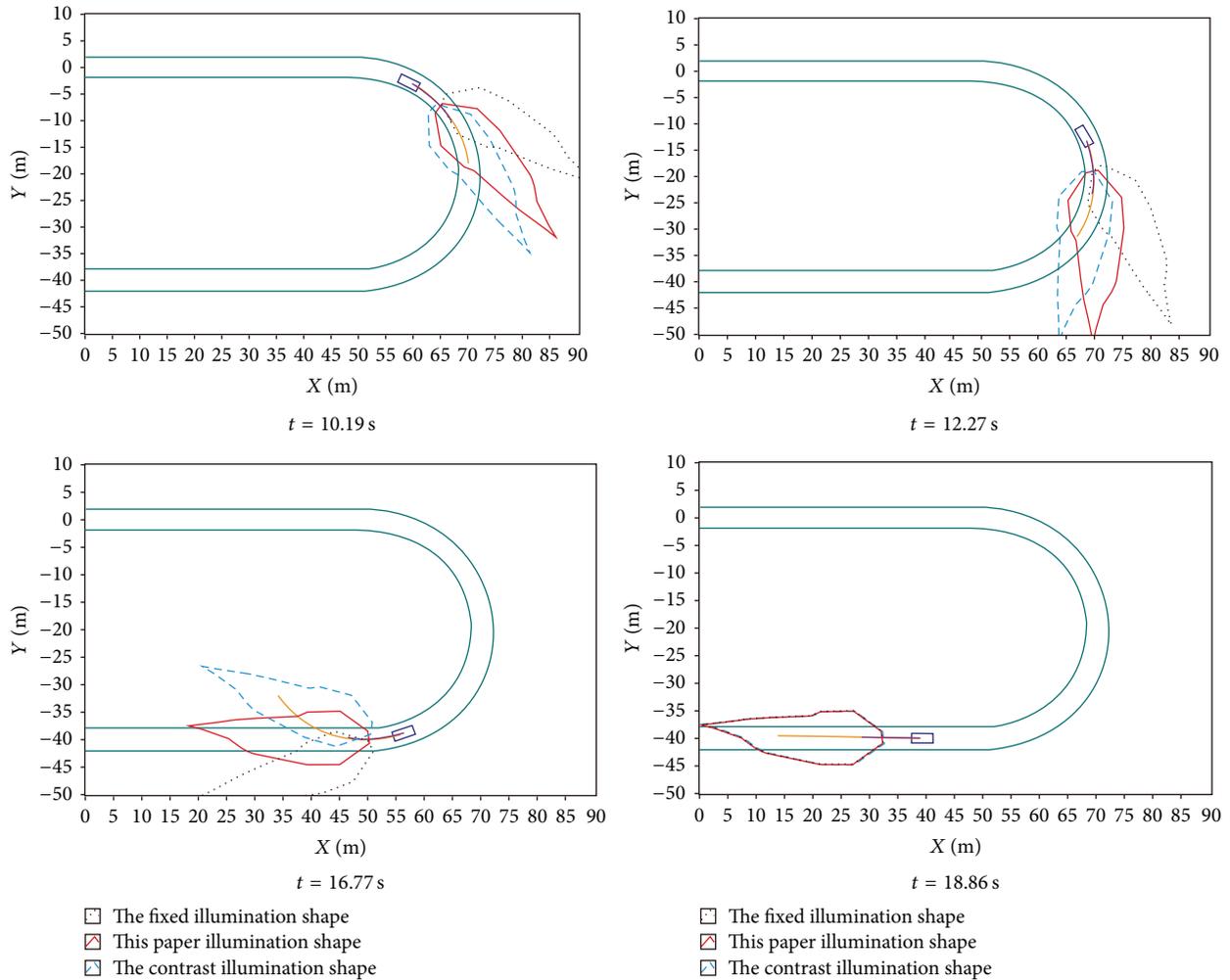


FIGURE 14: The front-lighting illumination area simulation schematic diagram 1.

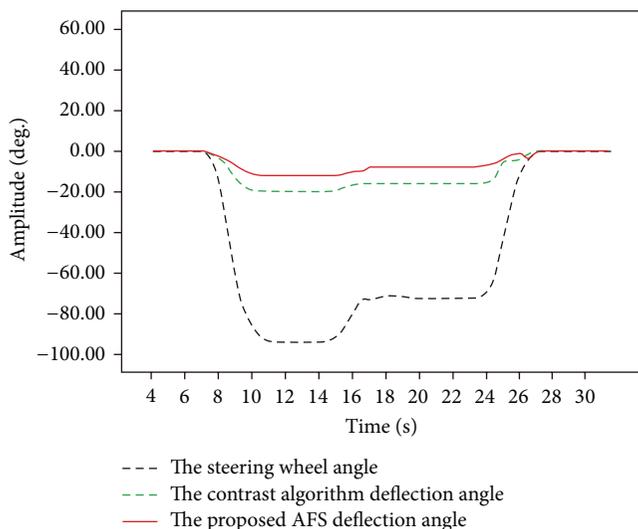


FIGURE 15: The headlamp's deflection angle and steering wheel angle (contrast algorithm and our algorithm diagram 2).

line's illumination which meant that their illumination effects were the same.

Condition 2. The headlamp's deflection angle of the contrast algorithm and our algorithm was shown in Figure 15.

Both algorithms' deflection angle could follow the input of the driver's steering angle. Relative to our algorithm, the contrast algorithm's deflection angle was about 20° bigger when the steering wheel began to deflect.

The specified control effect was shown in Figure 16 by the headlamp's illumination area. The explanation of the curves was similar to Figure 14 and the result was similar to Condition 1 which would not be analyzed again.

By the simulation result analysis of these two conditions, the conclusions were as follows.

- (1) The future track's prediction of our algorithm was more precise (the purple line in the image during the entire simulation was more close to the lane's center line), so the algorithm would control the headlamp's

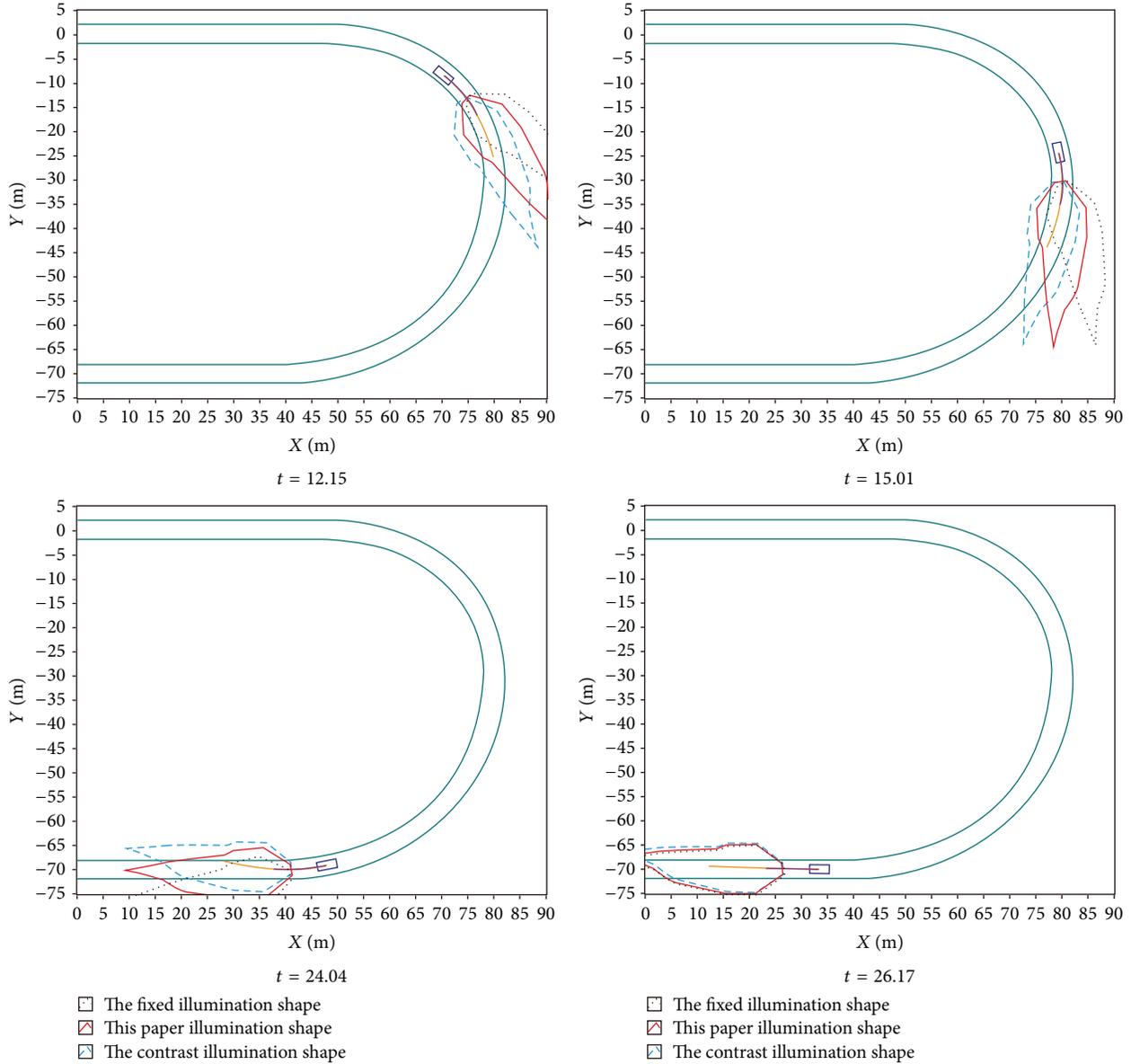


FIGURE 16: The front-lighting illumination area simulation schematic diagram 2.

deflection more precisely to ensure the front road's illumination effect well.

- (2) Relative to the contrast algorithm, when the vehicle was in the bend and was going to be out of the bend, the headlamp's deflection angle of our algorithm was smaller; the headlamp's beam could cover the front road more well and make sure that the region the driver was interested in was in the central zone of the headlamp's beam which meant better illumination effect.
- (3) When the vehicle was just in the bend and was out of the bend after a while, the headlamp's deflection angle of both algorithms was the same and their illumination effects were basically the same too.

In summary, the proposed algorithm could predict the vehicle's future track more precisely; on the basis of it, the headlamp's deflection angle was controlled to improve the front road's illumination condition effectively. Meanwhile, the proposed algorithm could make sure that the central zone of the headlamp's beam covered more regions which the driver was interested in.

5. Conclusion

The previous road experiment has found that the preview distance increased linearly with the velocity and the road curvature radius's increasing. Based on this rule, the precise preview distance was acquired by the eye tracker. The average preview time under the speed of 10–30 km/h is 1.70 s. Under

the curvature radius of (20–40) m, the average preview time is 1.44 s. An empirical model of the preview distance was proposed which accorded with the driver's visual characteristics.

Based on the dynamic predict algorithm of the vehicle's future track using the driver's best preview acceleration model, the algorithm's fixed preview time was modified by road experiments and a full AFS algorithm was developed based on the algorithm. The future track's prediction of the proposed algorithm was precise.

Finally, by the cosimulation of CarSim and LabVIEW, the headlamp's illumination area and the vehicle's future track of different time were analyzed in detail in the bend driving process. The illumination effect was analyzed in two aspects and was compared with a mature nonpredictive algorithm. The result showed that the control effect was obvious and the algorithm had a good application prospect.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This paper is supported by the National Natural Science Foundation of China (50975120), Specialized Research Fund for the Doctoral Program of Higher Education (20120061110028), Jilin Provincial Research Foundation for Technology Guide (20130413058GH), and Program for Chang Jiang Scholars and Innovative Research Team in University (no. IRT1017), China. Thanks are due to them.

References

- [1] ECE Regulation No. 123, Uniform Provisions Concerning the Approval of Adaptive Front-lighting System (AFS) for Motor Vehicles, 2007.
- [2] L. Strambersky, M. Cejnek, and J. Martoch, "Fast Track—Static and Dynamic Bending Analysis," SAE Paper 2004-01-0440, 2004.
- [3] R. Huang, *Automobile Adaptive Lighting System Development*, Dalian University of Technology, 2008.
- [4] L. Lifu and D. Qian, "Research on the mathematical model of bending mode of vehicle adaptive front-lighting system," *Automobile Technology*, no. 10, pp. 43–46, 2010.
- [5] X. Yu, *Study on Automobile Adaptive Lighting Control System Dalian*, University of Technology, 2009.
- [6] Z. Deng, "The develop of AFS-adaptive frontlighting-system," *China Light & Lighting*, no. 6, pp. 8–11, 2006.
- [7] S. McLaughlin, J. Hankey, C. A. Green, and M. Larsen, "Target detection distances and driver performance with swiveling HID headlamps," SAE Paper 2004-01-2258, 2004.
- [8] H. Rong, J. Gong, and J. Cao, "Research on key technology of AFS system auto electric parts," no. 5, pp. 15–18, 2008.
- [9] W. Pu, "The profile and trends of Adaptive Front Lighting System (AFS)," *Lamps and Lighting*, no. 2, pp. 22–24, 2009.
- [10] F. A.-K. Ibrahim, "Predictive Adaptive Front Lighting Algorithm for Branching Road Geometry: United States," Tech. Rep., US 7568822 B2, 2009.
- [11] K. Ishiguro and Y. Yamada, "Control technology for bending mode AFS," SAE Paper 2004-01-0441, 2004.
- [12] K.-H. Kim, D.-H. Yum, D.-K. Byeon, D.-Y. Kim, and D.-I. Lee, "Improving driver's visual field using estimation of curvature," in *International Conference on Control, Automation and Systems (ICCAS '10)*, pp. 728–731, KINTEX, October 2010.
- [13] M. F. Land, "Predictable eye-head coordination during driving," *Nature*, vol. 359, no. 6393, pp. 318–320, 1992.
- [14] M. F. Land and D. N. Lee, "Where we look when we steer," *Nature*, vol. 369, no. 6483, pp. 742–744, 1994.
- [15] M. Land and J. Horwood, "Which parts of the road guide steering?" *Nature*, vol. 377, no. 6547, pp. 339–340, 1995.
- [16] M. Young, C. Chang, and Y. Chuang, "A study of drivers' viewing field when driving on curvy roads," in *Proceedings of the the 3rd International Conference of Traffic & Transport Psychology*, Nottingham, UK, 2004.
- [17] H. Zhang and X. Zhang, "Robust gain-scheduling energy-to-peak control of vehicle lateral dynamics stabilisation," *Vehicle System Dynamics*, vol. 52, no. 3, pp. 309–340, 2013.
- [18] Z. Shuai and H. Zhang, "Lateral motion control for four-wheel-independent-drive electric vehicles using optimal torque allocation and dynamic message priority scheduling," *Control Engineering Practice*, vol. 24, pp. 55–56, 2014.
- [19] Z. Shuai and H. Zhang, "Combined AFS and DYC control of four-wheel-independent-drive electric vehicles over CAN network with time-varying delays," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 2, pp. 591–602, 2014.
- [20] Z. Gao and F. Gao, "Intelligent vchle head lamp system simulating preview action of driver," CN 201010175695, China, 2010.
- [21] X. Guan, Z. Gao, and K. Guo, "A hypothesis of steady preview and dynamic correction for driver model," *Chinese Automotive Engineering*, vol. 25, no. 3, pp. 19–22, 2003.
- [22] Z. Gao, X. Guan, Q. Li et al., "Driver optimal preview longitudinal acceleration model," *Automotive Engineering*, vol. 24, no. 5, pp. 434–437, 2002.
- [23] K. Guo, *Vehicle Handling Dynamics Theory*, Jiangsu Science and Technology Publishing House, Nanjing, China, 2011.
- [24] K. Noh, D. Jung, H. Choi et al., "Development of ergonomic driver model considering human factors," SAE Paper 2007-01-3584, 2007.

Research Article

Robust Production Planning in Fashion Apparel Industry under Demand Uncertainty via Conditional Value at Risk

Abderrahim Ait-Alla,¹ Michael Teucke,¹ Michael Lütjen,¹
Samaneh Beheshti-Kashi,¹ and Hamid Reza Karimi²

¹ Bremer Institut für Produktion und Logistik GmbH at the University of Bremen (BIBA), 28359 Bremen, Germany

² Department of Engineering, Faculty of Technology and Science, University of Agder, 4898 Grimstad, Norway

Correspondence should be addressed to Michael Teucke; tck@biba.uni-bremen.de

Received 31 January 2014; Accepted 25 March 2014; Published 29 April 2014

Academic Editor: Michael Freitag

Copyright © 2014 Abderrahim Ait-Alla et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a mathematical model for robust production planning. The model helps fashion apparel suppliers in making decisions concerning allocation of production orders to different production plants characterized by different lead times and production costs, and in proper time scheduling and sequencing of these production orders. The model aims at optimizing these decisions concerning objectives of minimal production costs and minimal tardiness. It considers several factors such as the stochastic nature of customer demand, differences in production and transport costs and transport times between production plants in different regions. Finally, the model is applied to a case study. The results of numerical computations are presented. The implications of the model results on different fashion related product types and delivery strategies, as well as the model's limitations and potentials for expansion, are discussed. Results indicate that the production planning model using conditional value at risk (CVaR) as the risk measure performs robustly and provides flexibility in decision analysis between different scenarios.

1. Introduction

This contribution deals with production planning problems of fashion apparel products. Fashion apparel products belong to the most important consumer goods. Global retail revenues amounted to \$1,032 billion in 2009 and are expected to grow to \$1,163 billion by 2016 [1].

Production planning for fashion apparel products has to cope with demand uncertainties. Accordingly, the uncertain nature of the customer demand has to be taken into consideration by generating the production plan and in particular the production quantities, in order to meet uncertain customer demand in the best way possible and maximize the profit, by minimizing production costs.

In this case, production planning, in particular the correct placement of production orders concerning place, or region, of production, as well as time scheduling and sequencing of production orders, is of high economic importance for fashion apparel suppliers. However, at the time of generating the production plan, the predicted customer

demands are largely uncertain. Therefore, it is crucial to produce a robust production plan, which can manage the risk resulting from the demand forecast. This risk trade-off can be achieved by constraining the objective function or problem limits with CVaR. Indeed, CVaR intends to protect against undesirable realization of uncertain parameters beyond the expected evaluation due to the uncertainty of system parameters [2].

Existing papers dealing with the robust optimization in fashion apparel do not take into account the risk of losing more than an acceptable level of profits due to write-offs caused by an overly optimistic demand forecast, or earning less than a desired target profit due to an overly pessimistic demand forecast. This paper is the first study to address this problem for the fashion apparel industry.

To deal with this uncertainty in customer demand in the apparel industry, we propose a risk-constrained profit-expected maximization model. This model considers the stochastic nature of customer demand and generates a production plan which indicates the quantities of each product

that should be produced, the start of the production of each product, and the facility in which the products have to be produced. The objective is to maximize the total profit of productions by means of CVaR.

Consideration of risk in the optimization problem has a crucial role in optimization under uncertainty, particularly when the optimization problem has to deal with the losses that might be incurred under conditions of unfavorable demand. To consider the risk of erroneous demand forecasts, a loss function $f(X, s)$ will be defined, where X represents a decision vector and s is a vector representing uncertainty related to the future values of a number of stochastic parameters of the problem. These stochastic parameters, presented by the vector s , are governed by probability distribution P_s . Once the loss function f is defined, we denote the distribution function of f by $\varphi(X, \beta) := P\{s \mid f(X, s) \leq \beta\}$, which corresponds to the probability that the value of the loss function, for each realisation of a scenario of the vector s , and for a fixed X , does not exceed the value β . In this context, for a specified confidence value $\alpha \in [0, 1[$, that could be equal to 0,90 or 0,95 in some applications, α -VaR denotes the probability that the expected value of the loss function exceeds β only in $(1 - \alpha) \cdot 100\%$ for all possible realisations of stochastic parameters, which can be seen as the worst-case scenario. Based on the definition of value at risk (VaR), conditional value at risk (CVaR) is defined as the mean of the tail distribution exceeding VaR [3].

The model is applied to a problem based on a case study in the apparel industry. Different factories located in different countries manufacture a number of product types and sell these products to customers in Europe. The orders consist of the type of products, delivery date, and quantity. Usually, the quantity demand of a product i is of stochastic nature and having a probability distribution $p(i)$. After collecting orders, the production plan has to be generated for the next season by considering the manufacturing capacity, production cost, and transport cost while satisfying a CVaR constraint in order to maximize the expected total profit.

In order to generate different possible demand realisations for the mathematical model, different scenarios of Ω are generated by means of Monte Carlo simulation, using the demand probability distribution of each product within the set of products.

The paper is structured as follows: the next chapter will provide some additional information on production planning and demand uncertainty in the apparel industry. The subsequent chapter studies some of the literature on robust production planning under demand uncertainty. After that, our model will be described and subsequently applied to a case study that includes three scenarios (a pessimistic scenario, a normal scenario, and an optimistic scenario) differing in customer demand. These three scenarios have been proposed in order to cover different potential realizations of customer demand, which can occur but are not known at the time of planning. Thereby, the customer demand can be estimated as low (pessimistic estimate), normal (normal estimate), or high (optimistic estimate), based on historic data and the recommendations of experts. The paper ends with a conclusion and an outlook on further research.

2. Production Planning and Demand Uncertainty in Apparel Industry

Manufacturing and finishing of ready-to-wear garments are a complex process involving often a large number of collaborating parties within a production network [4]. These include garment manufacturers and their raw material suppliers, procurement agencies, logistic service providers, and retailers [5].

Due to limited potential for automation and price importance in competition, production of fashion apparel products has been largely outsourced to low wage countries, often situated either in East Asia, the Near East, in particular in China, Turkey, Bangladesh, India, and Vietnam. Due to the absence of technical constraints and availability of subcontractors, production can be shifted with few constraints between different regions. However, the large majority of the products are still sold in Europe and North America, even though rising domestic demand in Asian countries, such as China, has started to increasingly compete for local production resources [6–8]. In this context, many European garments suppliers have reduced the depth of their own production, or dropped it, and have adapted their role towards planning and coordinating activities within apparel supply chains, integrating garment manufacturers and raw materials suppliers, logistic service providers, and retailers [9]. The wide geographic distribution of supply chains results in considerable lead times of apparel production, due to the time needed for procurement of raw materials, and the time needed for transport of products from the production plants to the customers [10, 11].

In addition to low product costs, quality and reliability of logistic services are crucial for economic success in the apparel branch. Fashion apparel products are characterized by rapid product obsolescence. The sales seasons for a period of a few months or even weeks are short in comparison to their long production lead and delivery times. This increases the importance of adherence to delivery dates. On average, 95% of stock keeping units change for every new sales season, at least twice or four times a year [5, 12]. The production cycle is characterized by fixed seasonal cycles with fixed dates for product offers, orders, and deliveries, which are repeated production volumes based on aggregation of the retail preorders. However, as in practice production and distribution lead times usually exceed the length of the delivery times expected by customers, production planning and procurement of raw materials have to start before the end of the preorder period, using forecasts of total demand based on the orders arrived so far. Products may also be produced in larger quantities than if only based on preorders, with the excess offered directly from the warehouses to called postorders as long as stocks last [13].

Once their sales season has ended, articles, which have not been sold so far, can be sold only for reduced prices and thus with reduced profit in the next season. Thus, at the end of a sales season, the stock levels of the products should ideally be close to zero. On the other hand, stocks should not run out before the end of the sales period. Reproduction of successful products during a sales season is not possible due to the long

delivery times [12]. For these reasons, demand forecasting has to be considered an important input of production planning and supply chain planning models. Generally, forecasting of potential sales permeates all aspects of business operations [14]. Customer demand for fashion products is volatile and may vary broadly for different variants of the same product. Demand fluctuations are difficult to predict at the time of production planning. Optimistic forecasts may lead to overstocks and higher production costs and cause markdowns. Pessimistic forecasts may lead to loss of opportunity revenues for potential sales that could not be realized and to customer dissatisfaction with later brand changes.

On the other hand, traditional statistical forecasting methods are difficult to apply [15]. Specifically developed demand forecasting models often use soft computing methods, such as artificial neural networks [16], fuzzy logic, and evolutionary procedures [12, 17]. In addition, the integration of preorder information as well as expert judgments has produced accurate forecasts [18]. Most of these models have not been integrated into standard business software and are difficult to use, with sometimes mixed quality of the forecasts. Order fulfilment and product delivery failures, such as stock-outs, or increased storage costs and product depreciation write-offs continue to affect profits in the apparel branch in a very unfavourable way. It is estimated that, due to the long lead times and demand forecasting difficulties mentioned in the second section of this chapter, up to 30%–40% of high-fashion articles cannot be sold during the short sales season before the life expectancy of the article expires [19]. The need to cope with this situation and the impetus to realize Quick Response concepts have resulted in vertical integration of apparel suppliers and retailers in a number of cases [20]. Vertically integrated suppliers generally achieve better reaction to market changes due to shortened reaction times and more closely integrated information flows. However, these advantages cannot easily be copied by most traditional suppliers, which are often small or medium sized enterprises. Still, it is estimated that they have to write off up to 15%–20% of their high-fashion articles [19].

3. Literature Review on Problem Related Production Planning

Production and distribution operations are two key functions in the supply chain. In general, the supply chain management procedures emphasize production scheduling more than the distribution scheduling, because distribution is more flexible. When integrated supply chain management of production and distribution is realized, the resources are used more efficiently [21]. In a comprehensive review of integrated scheduling approaches Mula et al. showed that over 44 papers have been published between 1989 and 2009 [22].

In general, master, tactical, and operational planning levels can be separated. The purpose of most models is to minimize the total cost of the supply chain. Some work at the master planning level, integrating procurement in addition to production and distribution scheduling, is described in [23]. Most of the models focus on tactical planning, considering

different demand levels, costs, and capacities, and use centralized scheduling approaches. Scholz-Reiter et al. showed a scalable graph-based scheduling approach, which is focused on the operational level and real-time scheduling [24]. A large variety of OR techniques have been applied, comprising mathematical optimisation, heuristics, and metaheuristics, but it could be observed that most of the papers focus on deterministic solutions [22].

The special conditions of production planning for fashion products have received some studies as well. The work in [25] described a networked production planning process in a fashion oriented apparel supply chain. The production planning process was found to be an important area of improvement for a network in a time-based logic; shortening the production planning period, in fact, significantly affects the weighted average delivery anticipation. It leads, however, to increase setup and transportation costs due to the greater number of jobs generated during a campaign. Scheduling of integrated production and distribution systems in dynamic environments holds a potential to improve efficiency but also poses a challenging planning task due to its computational complexity. Long lead times, external and internal perturbations in productive processes, unstable business environments, and contextual differences (e.g., institutional, economic, and cultural) emphasize the relevance to the argued integration [26]. Three types of aggregate production planning methods for the apparel industry have been proposed in [27]. These allow changing a production model seasonally according to actual demands, or maintaining the same production model for several seasons with production for inventory. The mentioned work on integrated production and distribution planning does not deal with stochastic input data, such as demands. The work in [28] studied empirically the impact of the subsidy policy on total factor productivity for the example of Chinese cotton production.

Modern, large, and widely distributed production networks are subject to many forms of disturbances. Karimi, Duffy, and Dashkovsky et al. performed research on how to better deal with such dynamic influences. One possible solution is to introduce autonomous control, that is, to allow some parts of a large network to make their own decisions based on local situation and available information. However, stability of the network and robustness with respect to external and internal disturbances and time delays in signals must be assured to guarantee a reasonable performance and vitality of the whole system. For this purpose their work proposed an approach for controller design for large scale autonomous work systems capable of coping with time delays and explains its implementation and advantages on a concrete example [29–31].

Robust optimization models try to formulate production planning problems in a way that cost, or wastage, effects of uncertainty or risks are minimized or expected profit is maximized. A robust optimization model for a multisite production medium-term planning problem is developed in [32], based on the problems facing a multinational lingerie company with production sites in East Asia. It generates a cost minimal production plan for an uncertain environment with associated probabilities of different economic growth

scenarios. The cost minimal production plan is less sensitive to changes in the noisy and uncertain data. The work in [33] dealt with a portfolio selection model in which the methodologies of robust optimization are used for the minimization of the conditional value at risk of a portfolio of shares. The work in [34] proposed a method for robust self-scheduling based on CVaR. The proposed method is based on a security-constrained optimal power flow (SCOPF) program that explicitly treats the trade-off between risk and reward. The work in [35] proposed a fuzzy mathematical programming model for supply chain planning which considers supply, demand, and process uncertainties. The model has been formulated as a fuzzy mixed-integer linear programming model where data are ill-known and modelled by triangular fuzzy numbers. The fuzzy model provides the decision maker with alternative decision plans for different degrees of satisfaction. This proposal is tested by using data from a real automobile supply chain. A supply chain design problem for a new market opportunity with uncertain demand in an agile manufacturing setting is considered in [36]. This model integrates the design of supply chain and production planning for the supply chain's members and develops a robust optimization formulation. A novel framework based on conditional value at risk theory has been applied in [37] to the problem of operational planning for large-scale industrial batch plants under demand due date and amount uncertainty. The objective of the proposed model is to provide a daily production profile that not only is a tight upper bound on the production capacity of the plant but also is immune to the various forms of demand uncertainty. In further work, The work in [38] applied a robust optimization framework, as well as CVAR theory to the multisite operational planning problem under multiple forms of system uncertainty. They considered different forms of system uncertainty such as demand due date, demand amount, and transportation time uncertainty in the model. Their objective is to ensure the maximization of customer satisfaction along with the minimization of resource misallocation.

A robust multiobjective mixed integer nonlinear programming model is proposed by [39] to deal with a multisite, multiperiod, multiproduct aggregate production planning (APP) problem under uncertainty, considering two conflicting objectives simultaneously, as well as the uncertain nature of the supply chain. Their proposed model is solved as a single-objective mixed integer programming model applying the LP-metrics method. A two-stage real world capacitated production system with lead time and setup decisions uncertain production costs and customer demand is studied in [40], using a robust optimization approach. A mixed-integer programming (MIP) model is developed to minimize total production costs.

4. Description of the Model

Our model maximizes expected profits in scenarios of stochastic customer demand. The model generates production plans containing the production quantities of each product of a given product program, the production start of

each product, and the production plants where the products should be produced.

Notation. A set I of n products has to be manufactured, under restriction of several time and resource constraints. The forecast demand of each product to be manufactured is uncertain and a probabilistic distribution is used to characterize this uncertainty. The manufacturing of different products has to take place in just one of m different facilities located in different countries.

A facility $j \in J$ has a fixed production cost CP_{ij} , to produce the product i , a fixed transport cost CT_j , and a production capacity CAP_{ij} of products which can be manufactured during one period.

A product $i \in I$ has a delivery due date DD_i , and if it is supplied earlier than the recommended DD_i that would cause extra holding costs CH_i , and if it is supplied later than the due date DD_i that would cause penalty costs PEC_i .

The objective is to find a feasible plan, which determines the quantity of each product to be produced and indicates the start of production of each product i , manufactured from facility j , to fulfil the stochastic demand and maximize the total production profit.

Parameters. In the following, the parameters for the model are defined.

T is set of planning periods, where $k = |T|$.

t is index of the planning period $t \in \{1, \dots, k\}$.

J is set of facilities, where $m = |J|$.

j is index of facility $j \in \{1, \dots, m\}$.

I is set of products, where $n = |I|$.

i is index of product $i \in \{1, \dots, n\}$.

Ω is set of all possible scenarios of stochastic variables $\Omega = \{1, 2, \dots, S\}$.

s is index of scenario res. Realisation of stochastic variables $s \in [1, S]$, where $S := |\Omega|$

D_i^s is the demand of product i under scenario s .

P_s is probability of scenario s .

SP_i is the unit selling price for product i .

CP_{ij} is the unit production cost for product i manufactured in facility j .

CT_j is transport cost from the facility j in terms of number of periods.

CH_i is the unit inventory holding cost for product i at the end of production.

SL_i is salvage value per unit for product i .

SR_i is shortage penalty per unit for product i .

TT_j is transport time from the facility j in terms of number of periods.

DD_i is delivery due date of product i in terms of number of periods.

PF_i is penalty cost incurred to each late delivery for product i when it is supplied later than the delivery date DD_i .

CAP_{ij} is production capacity (units) for product i during a period in facility j .

α is confidence level of risk parameter CVaR, where $\alpha \in [0, 1[$.

β is risk parameter.

λ is weight presented on solution variance.

μ is weight placed on model infeasibility which controls the trade-off between solution and model robustness.

Variables. Consider the following.

X_i is quantity of product i that shall be produced (decision variable).

Y_{ij} is binary variable, which indicates whether product i is produced in facility j (decision variable).

STP_i is start period of the production of product i (decision integer variable);

TP_i is time required in periods to produce X_i of product i ; $TP_i = \sum_{j=1}^m Y_{ij}(X_i/CAP_{ij})$

$ExpEndP_i$ is expected delivery date of product i in period unit; $ExpEndP_i = STP_i + TP_i + \sum_{j=1}^m Y_{ij} \cdot TT_j$

MIN_i^s is quantity of product i effectively sold under scenario s . $MIN_i^s = \min(X_i, D_i^s)$.

WAS_i^s is wastage quantity of product i under scenario s . $WAS_i^s = \max(0, X_i - D_i^s)$.

SHO_i^s is shortage quantity of product i under scenario s . $SHO_i^s = \max(0, D_i^s - X_i)$.

I_i is binary variable, which indicates whether the product i has to be held or not after its production and transportation:

$$I_i = \begin{cases} 1, & \text{if } ExpEndP_i < DD_i \\ 0, & \text{else.} \end{cases} \quad (1)$$

Objective Function. The objective function has the following components:

- (1) sales revenue,
- (2) production cost,
- (3) transport cost,
- (4) holding cost,
- (5) penalty cost due to a late delivery,
- (6) risk cost incurred for possible wastage cost,
- (7) risk cost incurred for possible shortage cost.

The above components have been considered due to the fact that they are the most important costs which can be mainly affected by the decision variable of the model.

The objective is to maximize the total profit revenue consisting of sales revenue gained by selling the products, production (PC), transport (TC), penalty, holding (HC), and other risk costs which can be incurred for possible wastage or shortage quantities:

$$SR^s \text{ (sales revenue)} = \sum_{i=1}^n MIN_i^s \cdot SP_i, \quad (2)$$

$$PC \text{ (production cost)} = \sum_{i=1}^n \sum_{j=1}^m Y_{ij} X_i CP_{ij}, \quad (3)$$

$$TC \text{ (transport cost)} = \sum_{i=1}^n \sum_{j=1}^m Y_{ij} X_i CT_j, \quad (4)$$

$$HC \text{ (holding cost)} = \sum_{i=1}^n \{I_i \cdot (DD_i - ExpEndP_i) \cdot X_i CH_i\}, \quad (5)$$

$$PEC \text{ (penalty cost)} = \sum_{i=1}^n \{(1 - I_i) (ExpEndP_i - DD_i) X_i PF_i\}, \quad (6)$$

$$WC^s \text{ (wastage cost)} = \sum_{i=1}^n WAS_i^s \cdot SL_i, \quad (7)$$

$$SC^s \text{ (shortage cost)} = \sum_{i=1}^n SHO_i^s \cdot SR_i. \quad (8)$$

Equation (2) represents the total sales revenue due to selling the set of products I . Equation (3) is the total production cost of manufacturing the set of products I . Equation (4) is the total transport cost of the set products I . Equation (5) corresponds to the total holding cost of units of products, which has to be stored in the warehouses for a determined holding period. Equation (6) represents the total penalty cost of supplying the set of products I later than the appropriate delivery date. Equation (7) represents the wastage cost to liquidate the overstocks set of products I . Equation (8) represents an artificial penalty for the demand dissatisfaction and subsequently to loss of opportunity revenue.

The following formulation of robust problem is defined according to Leung et al. [32].

We denote the profit function for each scenario s with

$$F_s = [SR^s + WC^s - SC^s - PC - TC - HC - PEC]. \quad (9)$$

The objective function, which represents the maximizing expected profit of the production planning problem with the demand uncertainty, is formulated as follows:

$$\begin{aligned} \max E [\text{profit}] \\ = \max \left\{ \sum_{s=1}^{\Omega} p_s F_s + \lambda \sum_{s=1}^{\Omega} p_s \left[F_s - \sum_{s'=1}^{\Omega} p_{s'} F_{s'} \right] \right. \\ \left. + \mu \sum_{s=1}^{\Omega} p_s \theta_s \right\}. \end{aligned} \quad (10)$$

The first term of (10) is the mean value of the total profit. The second term of (10) denotes the measure solution robustness of the model. The third term in (10) represents the model robustness and is used to penalize model infeasibility.

Constraints. First, we define the loss function as follows:

$$\forall s \in \Omega, \quad f(X, s) = \sum_{i=1}^n f(X_i, s) = WC^s + SC^s - \beta \cdot SR^s. \quad (11)$$

The objective function is subject to the following constraints:

$$\forall i \in I, \quad \text{ExpEnd}P_i \leq k, \quad (12)$$

$$\forall i \in I, \quad \sum_{j=1}^m Y_{ij} = 1, \quad (13)$$

$$\forall i \in I, \quad \text{STP}_i > 0, \quad (14)$$

$$\forall i, i' \in I, \quad \text{if } \exists \mu \in J \text{ so that } Y_{i\mu} = Y_{i'\mu} = 1, \quad (15)$$

$$\text{STP}_i < \text{STP}_{i'} \text{ then } \text{STP}_{i'} \geq \text{STP}_i + \text{TP}_i,$$

$$\forall s \in \Omega, \quad P[s \mid f(X, s) \leq 0] \leq \alpha. \quad (16)$$

Constraint (12) ensures that the expected delivery date for all products should not exceed the length of the planning horizon $k = |T|$. Constraint (13) guarantees that the production of a product i will take place only in one facility j . Constraint (14) indicates that the start of production of a product i is positive and integer. Constraint (15) ensures that the production of more than one product cannot be carried out in a parallel manner. Constraint (16) restricts the probability of the loss function to be negative, for $\alpha \cdot 100\%$, which means that the undesirable realization of uncertainties can restrict the loss function to be positive in only $(1 - \alpha) \cdot 100\%$. As an example, for $\alpha = 0.95$ the constraint (16) has to be satisfied for 95% for all possible realizations of demand D_i^s for all $i \in I$ and for all $s \in \Omega$.

5. Experimental Results for a Case Study with Three Scenarios

In order to cover different potential realizations of customer demand, three scenarios have been selected, including a “pessimistic” scenario characterized by assumed low customer demand, a “normal” scenario characterized by average customer demand, and an “optimistic” scenario characterized by relatively high customer demand. These three scenarios take into account the insecurity concerning customer demand for the products. Different values have been chosen based on data taken from a real case study.

In this section we analyze the behavior of the model by means of a case study based scenario. The input data parameters of the model are summarized in Tables 1 and 2.

Three different scenarios are considered as a basis for application of the model and its numerical results. The first scenario “pessimistic scenario” illustrates low demand

TABLE 1: Input parameter used to solve the model.

Parameters	Inputs
$k = T $	10
$m = J $	2
$n = I $	4
Transport cost from facility 1 (China)	300 (pro 1000 product units)
Transport cost from facility 2 (Turkey)	100 (pro 1000 product units)
Transport time from facility 1 (China)	3
Transport time from facility 2 (Turkey)	1
α	0.95
β	0.1

TABLE 2: Unit production parameters.

	Product 1	Product 2	Product 3	Product 4
Production				
Minimum value	10000	7500	5000	8000
Maximum value	40000	30750	20750	31600
Production cost				
Facility 1	1	1	2	2
Facility 2	3	2	3	5
Capacity				
Facility 1	20000	20000	20000	20000
Facility 2	10000	10000	150000	10000
Penalty factor PF_i	0.1	0.2	0.8	0.5
Delivery due date DD_i	6	6	6	6
Shortage cost unit	5	4	5	9
Salvage cost unit	0.5	0.5	1	1
Unit selling price	8	6	8	14

forecasts for the next season. The second scenario “normal scenario” represents the case when the customer demand is normal. The third scenario “optimistic scenario” assumes high customer demand. To characterize the uncertainties of demand forecast of different products, a distribution probability is applied. Figures 1 and 2 show an example of the demand distribution, respectively, of the products under pessimistic scenario and demand distribution of the product 1 for the three scenarios.

The risk to be considered from solving the model can occur only because of wastage or shortage possibility. Therefore, according to our definition of the loss function $f(X, s) = \sum_{i=1}^n f(X_i, s) = WC^s + SC^s - 0.1 \cdot SR^s$, the constraint (16) for this case study ensures that the probability that the risk costs incurring by, respectively, wastage and shortage costs do not exceed the 10% ($\beta = 0.1$) of the expected profit revenue in 95% ($\alpha = 0.95$) of all possible realization of stochastic parameters.

The model has been efficiently solved by means of frontline risk solver which supports the robust optimization through CVaR.

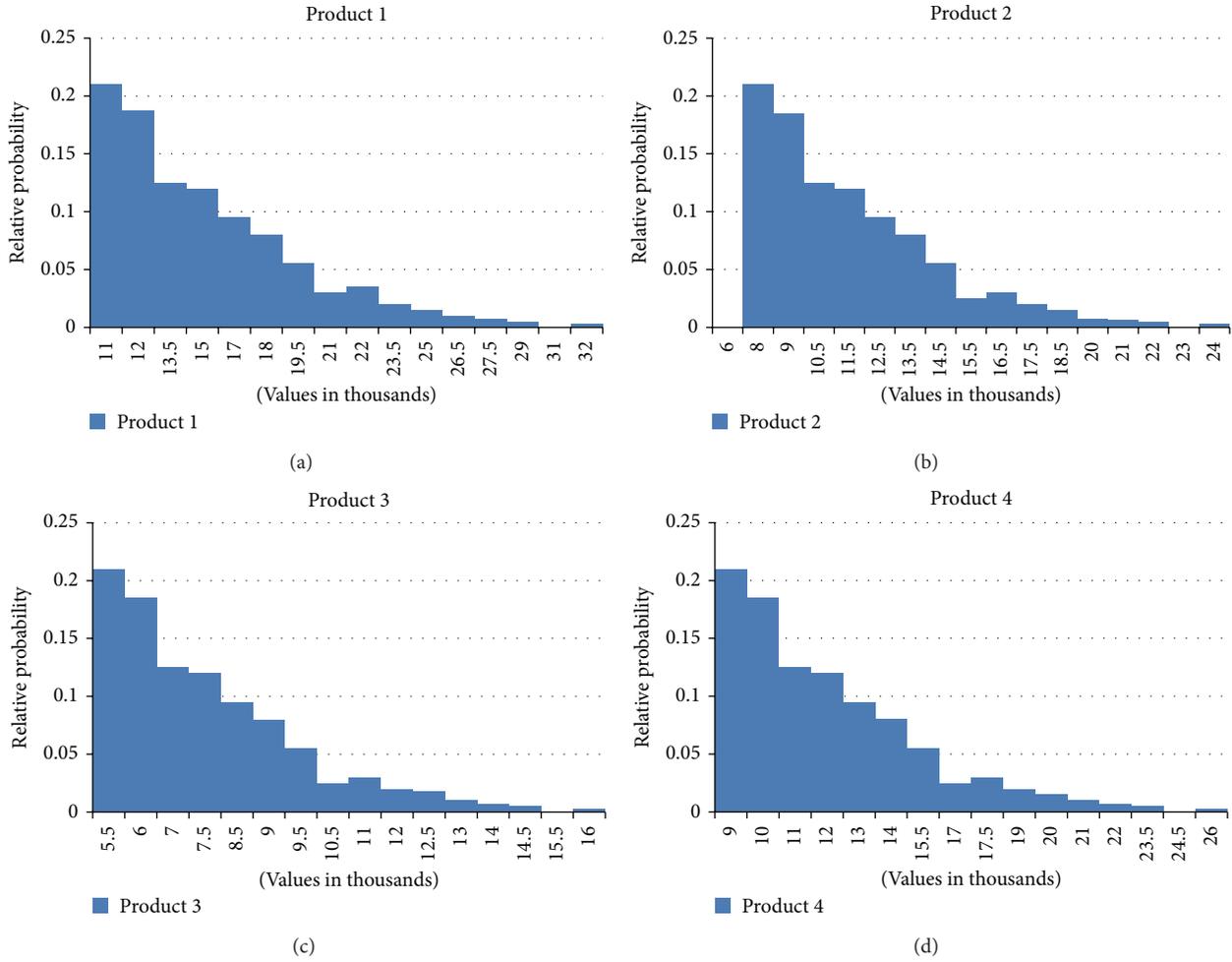


FIGURE 1: Demand distribution of the products in a pessimistic scenario.

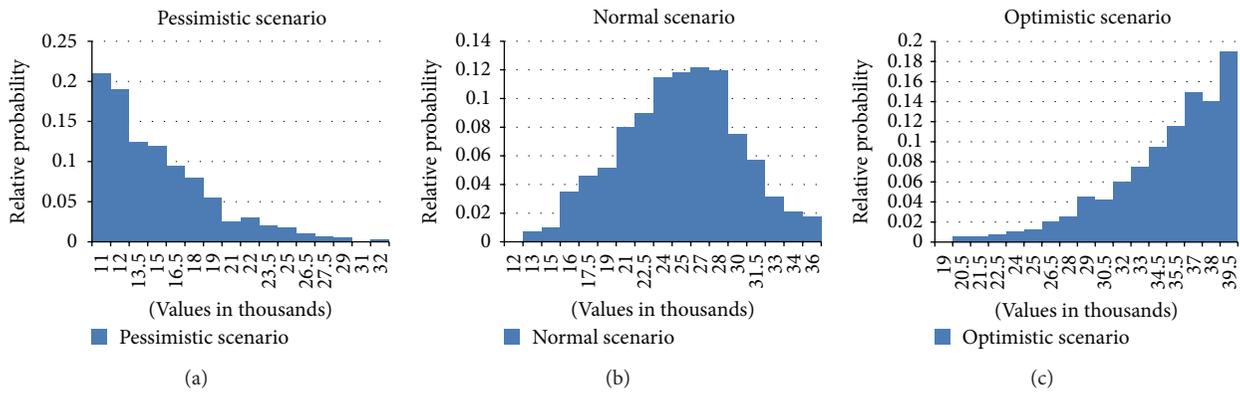


FIGURE 2: Demand distribution of product 1 in different scenarios.

5.1. *Pessimistic Scenario.* Table 3 lists the production plan in a pessimistic scenario. The expected average profit of the production for the next season is about 219000. Moreover, the solution can guarantee that the loss function will be negative in 96% cases of all possible realizations of stochastic parameters (Figure 3).

5.2. *Normal Scenario.* In the normal scenario, the demand of product 1 and product 4, respectively, has nearly doubled, whereas the demand of product 2 and product 3 has been slightly increased. This is due to the low penalty factor of product 1 and product 4 which can be incurred for each delayed delivery and for the capacity restriction of

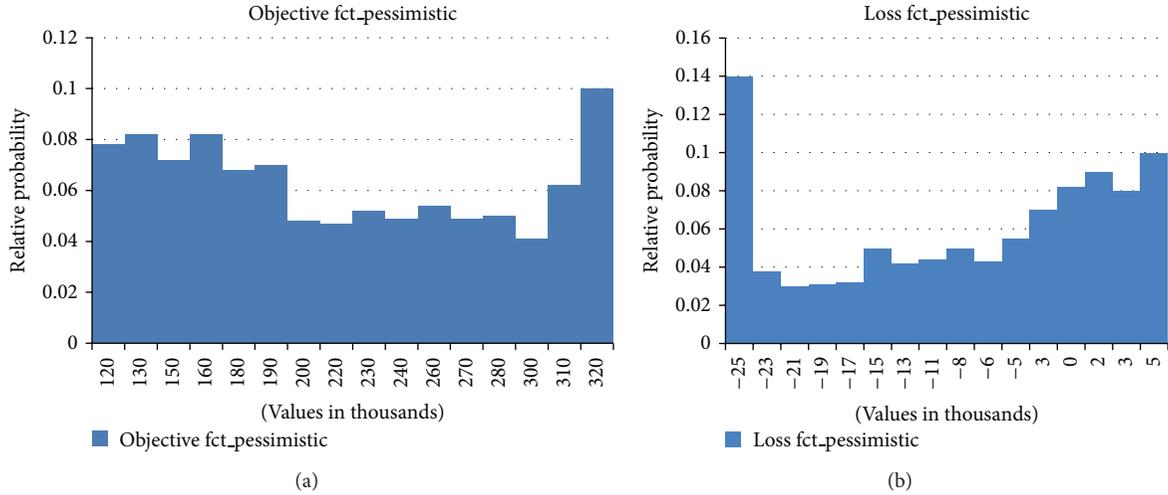


FIGURE 3: Profit and loss function distribution in a pessimistic scenario.

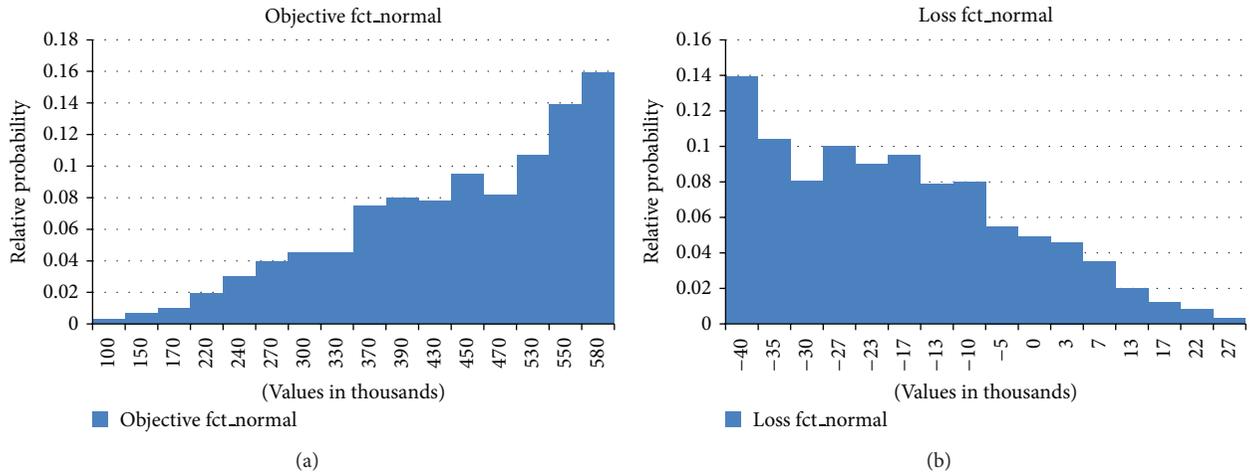


FIGURE 4: Profit and loss function distribution in a normal scenario.

TABLE 3: Production plan with the demand quantities in a pessimistic scenario.

	Product 1	Product 2	Product 3	Product 4
Demand next season	17969	16549	10169	14945
Start of production	3	2	1	1
Expected delivery date	6	6	5	4
Produced In				
Facility 1	0	1	1	1
Facility 2	1	0	0	0

TABLE 4: Production plan with the demand quantities in a normal scenario.

	Product 1	Product 2	Product 3	Product 4
Demand next season	32355	19682	14415	26284
Start of production	1	4	3	1
Expected delivery date	6	8	7	6
Produced in				
Facility 1	0	1	1	1
Facility 2	1	0	0	0

the facilities (Table 4). In this case, the expected average profit is about 450000, and the loss function is negative in 97.5% of all possible realizations of the stochastic parameters (Figure 4).

5.3. *Optimistic Scenario.* Since the optimistic scenario predicts a boom economic scenario, the production reaches its

maximum (Table 5). In this case, the average profit is about 665000, and the loss function is satisfied with 99.1% of all possible realizations of the stochastic parameters (Figure 5).

6. Conclusion and Outlook

In this paper we have presented a risk-constrained profit-expected maximization model for a textile industry scenario.

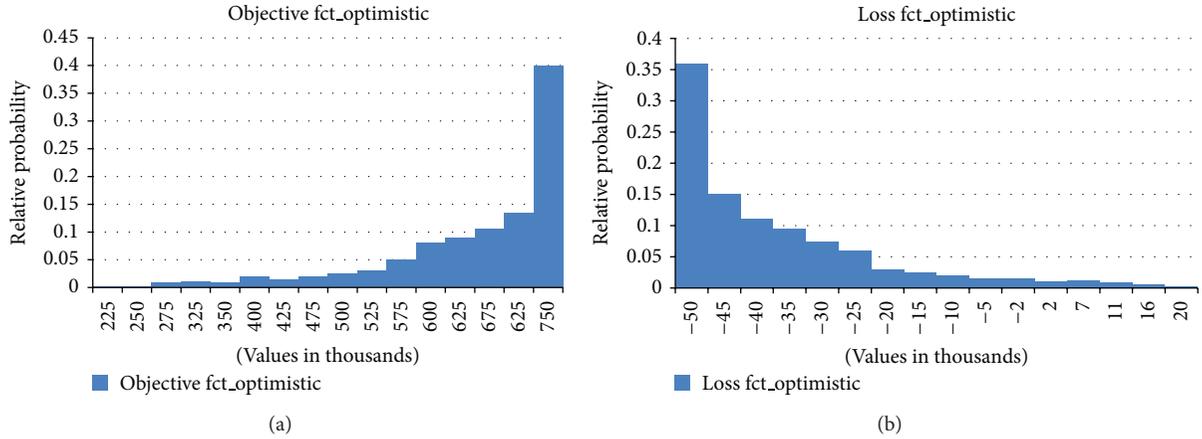


FIGURE 5: Profit and loss function distribution in an optimistic scenario.

TABLE 5: Production plan with the demand quantities in an optimistic scenario.

	Product 1	Product 2	Product 3	Product 4
Demand next season	37933	29312	19410	31600
Start of production	1	4	3	1
Expected delivery date	6	9	7	6
Produced in				
Facility 1	0	1	1	1
Facility 2	1	0	0	0

The forecast demand uncertainty has been explicitly considered in the model by means of the robust optimization and conditional value at risk theory by introducing and restricting a loss function. The robust optimization model originally proposed by [32] was adopted as the benchmark formulation of uncertainty considered in this paper. A case study has been used to demonstrate the viability of the proposed mathematical model. Results indicate that the production planning model using CVaR as the risk measure performs robustly and provides flexibility in decision analysis between different scenarios. The next step will be the further development of the profit maximization model to consider other types of uncertainties, for example, concerning the availability of the production plants, transportation means, and demand due dates.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] Datamonitor, Global Apparel Retail, 2010, <http://www.datamonitor.com/>.
 [2] R. T. Rockafellar and S. Uryasev, “Conditional value-at-risk for general loss distributions,” *Journal of Banking and Finance*, vol. 26, no. 7, pp. 1443–1471, 2002.

[3] S. Zhu and M. Fukushima, “Worst-case conditional value-at-risk with application to robust portfolio management,” *Operations Research*, vol. 57, no. 5, pp. 1155–1168, 2009.
 [4] A. Bruckner and S. Müller, *Supply Chain Management in der Bekleidungsindustrie*, Forschungsstelle der Bekleidungsindustrie, Cologne, Germany, 2003.
 [5] J. Fissahn, *Marktorientierte Beschaffung in der Bekleidungsindustrie*, Münster University, Münster, Germany edition, 2001.
 [6] A. Schneider, *Internationalisierungsstrategien in der deutschen Textil- und Bekleidungsindustrie-eine empirische Untersuchung [M.S. dissertation]*, RWTH Aachen University, Aachen, Germany, 2003.
 [7] G. Gereffi and O. Memedovic, *The Global Apparel Value Chain—What Prospects for Upgrading by Developing Countries?* UNIDO, Vienna, Austria, 2003.
 [8] M. Grömling and J. Matthes, *Globalisierung und Strukturwandel der Deutschen Textil- und Bekleidungsindustrie*, Dt. Inst.-Verlag, Colougne, Germany, 2003.
 [9] I. Langenhorst, *Shop-logistik in der bekleidungswirtschaft: eine analyse der anforderungen herstellereinitiiert shop-systeme an die logistikprozesse der bekleidungsindustrie [Ph.D. thesis]*, Münster University, 1999.
 [10] H.-C. Pfohl, M. Gomm, and X. Shen, *China: Textil- und Bekleidungs-Supply Chain Zwischen Deutschland und China*, Free Beratung, Korschenbroich, Germany, in H. Wolf-Kluthausen, *Jahrbuch der Logistik*, 2007.
 [11] M. Teucke and B. Scholz-Reiter, “Improving order allocation in fashion supply chains using radio frequency identification (RFID) technologies,” in *Fashion Supply Chain Management Using Radio Frequency Identification (RFID) Technologies*, W. K. Wong and Z. X. Guo, Eds., vol. 152 of *Woodhead Publishing Series in Textiles*, Woodhead Publishing, Cambridge, UK, 2014.
 [12] S. Thomassey, “Sales forecasts in clothing industry: the key success factor of the supply chain management,” *International Journal of Production Economics*, vol. 128, no. 2, pp. 470–483, 2010.
 [13] D. Ahlert and G. Dieckheuer, *Marktorientierte Beschaffung in der Bekleidungsindustrie*, FATM, Münster, Germany, 2001.
 [14] J. T. Mentzer, C. C. Bienstock, and K. B. Kahn, “Benchmarking sales forecasting management,” *Business Horizons*, vol. 42, no. 3, pp. 48–56, 1999.

- [15] J. Quick, M. Rinis, C. Schmidt, and B. Walber, *SupplyTex—Erfolgreiches Supply-Management in KMU Der Textil- Und Bekleidungsindustrie*, Forschungsinstitut für Rationalisierung (FIR) e.V. an der Universität Aachen, Aachen, Germany, 2010.
- [16] Z.-L. Sun, T.-M. Choi, K.-F. Au, and Y. Yu, “Sales forecasting using extreme learning machine with applications in fashion retailing,” *Decision Support Systems*, vol. 46, no. 1, pp. 411–419, 2008.
- [17] S. Thomassey, M. Happiette, and J.-M. Castelain, “A global forecasting support system adapted to textile distribution,” *International Journal of Production Economics*, vol. 96, no. 1, pp. 81–95, 2005.
- [18] J. Mostard, R. Teunter, and R. de Koster, “Forecasting demand for single-period products: a case study in the apparel industry,” *European Journal of Operational Research*, vol. 211, no. 1, pp. 139–147, 2011.
- [19] K. Hoyndorff, S. Hülsmann, D. Spee, and M. ten Hompel, *Fashion Logistics. Grundlagen Über Prozesse und IT Entlang der Supply Chain*, Munich, Huss-Verlag GmbH, 2010.
- [20] J. Richardson, “Vertical integration and rapid response in fashion apparel,” *Organization Science*, vol. 7, no. 4, pp. 400–412, 1996.
- [21] Z.-L. Chen, “Integrated production and outbound distribution scheduling: review and extensions,” *Operations Research*, vol. 58, no. 1, pp. 130–148, 2010.
- [22] J. Mula, D. Peidro, M. Díaz-Madroñero, and E. Vicens, “Mathematical programming models for supply chain production and transport planning,” *European Journal of Operational Research*, vol. 204, no. 3, pp. 377–390, 2010.
- [23] S. A. Torabi and E. Hassini, “An interactive possibilistic programming approach for multiple objective supply chain master planning,” *Fuzzy Sets and Systems*, vol. 159, no. 2, pp. 193–214, 2008.
- [24] B. Scholz-Reiter, J. Hartmann, T. Makuschewitz, and E. M. Frazzon, “A generic approach for the graph-based integrated production and intermodal transport scheduling with capacity restrictions,” in *Procedia CIRP*, vol. 7, pp. 109–114, 2013.
- [25] A. D. Toni and A. Meneghetti, “Production planning process for a network of firms in the textile-apparel industry,” *International Journal of Production Economics*, vol. 65, no. 1, pp. 17–32, 2000.
- [26] E. M. Frazzon, A. G. N. Novaes, B. Scholz-Reiter, J. Hartmann, and T. Makuschewitz, “An approach for the rescheduling of integrated production and transport systems,” in *Proceedings of the 4th World Conference Production & Operations Management. Serving the World, Production & Operations Management Society*, p. 10, Amsterdam, The Netherlands, 2012.
- [27] H. Ishikura, “Study on the production planning of apparel products: determining optimal production times and quantities,” *Computers and Industrial Engineering*, vol. 27, no. 1–4, pp. 19–22, 1994.
- [28] Y. W. Tan, J. B. Guan, and H. R. Karimi, “The Impact of the subsidy policy on total factor productivity: an empirical analysis of China’s cotton production,” *Mathematical Problems in Engineering*, vol. 2013, Article ID 248537, 8 pages, 2013.
- [29] H. R. Karimi, N. A. Duffie, and S. Dashkovskiy, “Local capacity H_{∞} control for production networks of autonomous work systems with time-varying delays,” *IEEE Transactions on Automation Science and Engineering*, vol. 7, no. 4, pp. 849–857, 2010.
- [30] S. Dashkovskiy, H. R. Karimi, and M. Kosmykov, “Stability analysis of logistics networks with time-delays,” *Journal of Production Planning & Control*, vol. 24, no. 7, pp. 567–574, 2013.
- [31] A. Mehrsai, H. R. Karimi, and B. Scholz-Reiter, “Application of learning pallets for real-time scheduling by the use of radial basis function network,” *Neurocomputing*, vol. 101, pp. 82–93, 2013.
- [32] S. C. H. Leung, S. O. S. Tsang, W. L. Ng, and Y. Wu, “A robust optimization model for multi-site production planning problem in an uncertain environment,” *European Journal of Operational Research*, vol. 181, no. 1, pp. 224–238, 2007.
- [33] A.-G. Quaranta and A. Zaffaroni, “Robust optimization of conditional value at risk and portfolio selection,” *Journal of Banking and Finance*, vol. 32, no. 10, pp. 2046–2056, 2008.
- [34] R. A. Jabr, “Robust self-scheduling under price uncertainty using conditional Value-at-Risk,” *IEEE Transactions on Power Systems*, vol. 20, no. 4, pp. 1852–1858, 2005.
- [35] D. Peidro, J. Mula, R. Poler, and J.-L. Verdegay, “Fuzzy optimization for supply chain planning under supply, demand and process uncertainties,” *Fuzzy Sets and Systems*, vol. 160, no. 18, pp. 2640–2657, 2009.
- [36] F. Pan and R. Nagi, “Robust supply chain design under uncertain demand in agile manufacturing,” *Computers and Operations Research*, vol. 37, no. 4, pp. 668–683, 2010.
- [37] P. M. Verderame and C. A. Floudas, “Operational planning of large-scale industrial batch plants under demand due date and amount uncertainty: II. Conditional value-at-risk framework,” *Industrial and Engineering Chemistry Research*, vol. 49, no. 1, pp. 260–275, 2010.
- [38] P. M. Verderame and C. A. Floudas, “Multisite planning under demand and transportation time uncertainty: Robust optimization and conditional value-at-risk frameworks,” *Industrial and Engineering Chemistry Research*, vol. 50, no. 9, pp. 4959–4982, 2011.
- [39] S. M. J. Mirzapour Al-E-Hashem, H. Malekly, and M. B. Aryanezhad, “A multi-objective robust optimization model for multi-product multi-site aggregate production planning in a supply chain under uncertainty,” *International Journal of Production Economics*, vol. 134, no. 1, pp. 28–42, 2011.
- [40] D. Rahmani, R. Ramezani, P. Fattahi, and M. Heydari, “A robust optimization model for multi-product two-stage capacitated production planning under uncertainty,” *Applied Mathematical Modelling*, vol. 37, no. 20–21, pp. 8957–8971, 2013.

Research Article

Adaptability of the Logistics System in National Economic Mobilization Based on Blocking Flow Theory

Xiangyuan Jing, Qingmei Tan, Liusan Wu, and Xiaohui Li

School of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

Correspondence should be addressed to Liusan Wu; wuhusheng563520@126.com

Received 20 December 2013; Revised 7 April 2014; Accepted 9 April 2014; Published 28 April 2014

Academic Editor: Hamid Reza Karimi

Copyright © 2014 Xiangyuan Jing et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the process of national economic mobilization, the logistics system usually suffers from negative impact and/or threats of such emergency events as wars and accidents, which implies that adaptability of the logistics system directly determines realization of economic mobilization. And where the real-time rescue operation is concerned, heavy traffic congestion is likely to cause a great loss of or damage to human beings and their properties. To deal with this situation, this article constructs a blocking-resistance optimum model and an optimum restructuring model based on blocking flow theories, of which both are illustrated by numerical cases and compared in characteristics and application. The design of these two models is expected to eliminate or alleviate the congestion situation occurring in the logistics system, thus effectively enhancing its adaptability in the national economic mobilization process.

1. Introduction

Mobilization of national economy refers to a series of activities of planning and organizing a nation's economy from the previously regular situation to an irregular state in order to satisfy emergency needs when wars or accidents occur [1]. In the whole operation, resources mobilization is given priority and thus the logistics activities play an essential role in the phases of resource preparation, distribution, and demobilization [2]. Since national economic mobilization is featured by quick response to wars and other urgent events, a logistics system that works mainly to meet emergency demand is defined as the national economic mobilization logistics (abbreviated as mobilization logistics). Specifically, the national economic mobilization logistics activities involve a whole procedure of spatially transferring mobilized resources from suppliers to end users through preparation, transportation, packaging, storage, and even consumption, and so forth [3].

Wars and emergency events, on the one hand, are the reasons why mobilization logistics comes into being and are the objects that the logistics system has to deal with, and, on the other hand, they actually form attacks and threats to the mobilization logistics [4]. As statistical data show,

in the rescue operation aftermath a natural disaster, an ineffective logistics distribution usually causes substantial losses of casualties and property, accounting for approximately 15% to 20% of the overall losses [4]. Once suffering from an impact or threat, the mobilization logistics system is needed to adjust its own internal components and structural functions in order to become adaptable to any change, which is regarded as the logistics system's adaptability [5]. By referring to system disturbance and monitoring theories, the authors define the mobilization logistics system adaptability as an ability of a logistics system to respond to disturbance and emergency events such as wars and accidents through restructure of its internal elements and functions when the nation's mobilization logistics system is disturbed [6].

National economic mobilization logistics is essentially a summit logistics. Accordingly, if any war or unexpected event occurs, logistics demand will soar and logistics network systems will probably be attacked. Under the circumstances, there will emerge blocks in logistics network nodes and links and even a blocking in the whole network [7]. To achieve a higher capability to control the network in real time, some methodologies have been directly assigned to distributed flow objects [8–11]. In this paper, the authors reckon the blocking phenomenon to be coincident with issues described

in blocking flow theories. Furthermore, considering potential losses incurred by serious traffic blocking, this paper attempts to construct a blocking-resistance optimum model and an optimum restructuring model using blocking flow theories, followed by a detailed analysis of their difference and usage. Such design is expected to eliminate or alleviate congestion situation occurring in the logistics system, thus effectively enhancing its adaptability in the national economic mobilization process.

2. Generalization of Blocking Flow

The concept of blocking flow originates from the maximum flow algorithm proposed by Dinitz in 1970, which was originally used to deal with the transition value in maximum flow algorithm; later on, it had been widely introduced in maximum flow algorithm. Since 1994, Professor Ning Xuanxi has been engaged in the blocking flow theory and application systematically and deeply, funded by the National Natural Science Foundation three times. He made great contributions to the blocking flow theory by having proposed relevant concepts and definitions, proved fundamental theorems, solved minimum flow problems in the network, constructed a basic frame of block theory, and made a series of progress in approaches and application [12].

2.1. Basic Concepts of Network Flow

- (1) Value of flow: the materials' quantity passes through the arc $i-j$ during a specific period of time, denoted by f_{ij} , a variable ready to be calculated in the network flow problem. The total flow value in a network is represented by $v(f)$.
- (2) Capacity: the maximum permissible flow of an arc is generally denoted by c_{ij} .
- (3) Feasible flow: the flow satisfying the conditions is as follows.
 - (a) Capacity constraint: As for each arc $v_i v_j \in A$ (A standing for the arcs set), there is $0 \leq f_{ij} \leq c_{ij}$, that is, the flow in arc $i-j$ being not more than its capacity.
 - (b) Equilibrium condition: The flow entering the initial vertex v_s is equal to the total sum of flow, and the one outflowing from the terminal vertex v_t is equal to total sum of flow, and the one entering an intermediate vertex i ($i \neq s, t$) is equal to that outflowing this vertex, shown as

$$\begin{aligned}
 \sum_{v_s v_j \in A} f_{sj} - \sum_{v_i v_s \in A} f_{is} &= v(f), \\
 \sum_{v_i v_j \in A} f_{tj} - \sum_{v_i v_t \in A} f_{it} &= -v(f), \\
 \sum_{v_i v_j \in A} f_{ij} - \sum_{v_j v_i \in A} f_{ji} &= 0.
 \end{aligned} \tag{1}$$

- (4) Saturated arc and unsaturated arc: The arc with the same flow value as the capacity (i.e., $f_{ij} = c_{ij}$) is called saturated arc, while that with less flow value than capacity (i.e., $f_{ij} < c_{ij}$) is called unsaturated arc.
- (5) Forward arc and backward arc: In the complete network, the arc with direction from the initial vertex v_s to the terminal vertex v_t is called forward arcs, and the one with direction from v_t to v_s is called backward arc.

2.2. Definitions of Blocking Flow

Definition 1. An augmenting directed path from source vertex s to sink vertex t is defined as a path along which flow value can be increased and each arc in the path has the direction consistent with the flow direction. The augmenting directed path in discussion exists without repeated vertexes or arcs.

Definition 2. A feasible flow is defined as the saturated flow of the network if there is no augmenting directed path.

Definition 3. The saturated flow of a network is defined as the blocking flow if the value of the saturated flow is less than the entrance flow capacity of the network.

Definition 4. The capacity potential at the vertex V_i refers to the difference between the total capacities of outgoing arcs and the total capacity of incoming arcs at V_i , shown as

$$\Phi_{vi} = \sum_j c_{ij} - \sum_j c_{ji}, \tag{2}$$

where c_{ji} denotes the capacity of the arc e_{ji} . Clearly, when $\Phi_A \geq 0$, there will be no structural blocking phenomenon at the vertex A (neither source vertex nor sink vertex); and, when $\Phi_A < 0$, structural blocking will probably occur at the vertex A . In this case, the vertex with negative capacity potential is called structurally blocking vertex.

Definition 5. If capacities of all vertexes in a network (excluding source and sink vertexes) are commonly zero, such a network is called a totally balanced one.

3. Blocking-Resistance Optimum Model for the Traffic Network

Since the traffic network design is generally required not only to meet the demand for traffic flow but also to alleviate blocking phenomena as much as possible, both parameters of maximum flow and minimum flow in a network are worth special notice. Traditionally, some designed networks were unable to meet the demand for the maximum flow in practical operation by ignoring the minimum flow. Accordingly, to avoid any blocking in a network, the proposed model amends the original capacity parameters in a permissible range and thus makes the capacity potential of all vertexes more than zero, with exception of the source and sink vertex.

3.1. *Model Design.* Take a traffic network $N = (V, A, s, t)$ into consideration, in which s denotes the source vertex, t the sink vertex, V the vertex set, and A the arc set. The original design is described as follows.

- (1) The originally designed capacity, and the permissible maximum and minimum capacity are already known.
- (2) The requirement concerning incoming flow is given.
- (3) The cost used to construct one unit capacity of an arc is learned.

Note that, in the proposed model, the fluctuation of capacities corresponds to the original ones comparatively. The cost occurring from a unit capacity fluctuation of an arc is supposed to be equal, denoted as $b(e_{ij})$ for the arc e_{ij} . The optimum objective is to achieve the least transformation cost [13].

The objective function is

$$\min \left(\sum \{b(e_{ij}) \cdot \Delta c(e_{ij}) \mid e_{ij} \in E\} \right), \quad (3)$$

where $\Delta c(e_{ij})$ indicates the increased capacity of the arc e_{ij} .

The constraint conditions are shown below.

- (1) As for the negative capacity potential, there is

$$\begin{aligned} \bar{\Phi}_{v_i} &= \sum_j \{\Delta c(e_{ij}) \mid e_{ij} \in E\} \\ &- \sum_j \{\Delta c(e_{ji}) \mid e_{ji} \in E\} + \Phi_{v_i} = 0, \quad \forall v_i \in V. \end{aligned} \quad (4)$$

- (2) As for the nonnegative capacity potential, there is

$$\begin{aligned} -\Phi_{v_i} &\leq \sum_j \{\Delta c(e_{ij}) \mid e_{ij} \in E\} \\ &- \sum_j \{\Delta c(e_{ji}) \mid e_{ji} \in E\} \leq 0, \quad \forall v_i \in V. \end{aligned} \quad (5)$$

- (3) The incoming flow is restricted by

$$\sum_j \bar{c}(e_{sj}) = F, \quad \forall v_i \in V. \quad (6)$$

- (4) In the network, each arc's flow value and capability are restricted by

$$\begin{aligned} x(e_{ij}) &\leq \bar{c}(e_{ij}), \quad \forall e_{ij} \in E, \\ mc_2(e_{ij}) &\leq \bar{c}(e_{ij}) \leq mc_1(e_{ij}), \quad \forall e_{ij} \in E, \end{aligned} \quad (7)$$

where $mc_1(e_{ij})$ and $mc_2(e_{ij})$ are the maximum and minimum capacities, respectively.

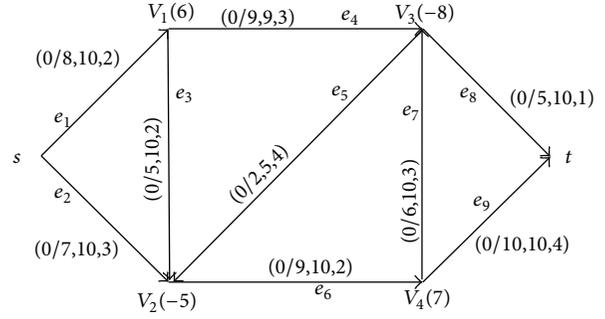


FIGURE 1: Numerical case of blocking-resistant optimum design.

3.2. *Numerical Case.* Figure 1 illustrates a traffic network related to the mobilization logistics system and six vertexes that represent six cities, respectively. As for each arc, there are four parameters: $x(e_i)/c(e_i)$, $mc_1(e_i)$, and $b(e_i)$, of which $c(e_i)$ indicates the original capacity of the arc e_i with $mc_2(e_i) = 2$ and $b(e_i)$ indicates the cost occurring from changing per unit capacity, and the numbers next to the vertex indicates the capacity potential of each vertex. Besides, the incoming flow is assumed to obey the rule of $F = 15$.

The blocking-resistant optimum design model is set up as follows.

The objective function is

$$\min \left(\sum \{b(e_i) \cdot \Delta c(e_i) \mid \forall e_i \in E\} \right). \quad (8)$$

Constraint conditions are

$$-6 \leq \Delta c(e_3) + \Delta c(e_4) - \Delta c(e_1) \leq 0, \quad (9)$$

$$-7 \leq \Delta c(e_7) + \Delta c(e_9) - \Delta c(e_6) \leq 0, \quad (10)$$

$$\Delta c(e_6) - \Delta c(e_2) - \Delta c(e_3) - \Delta c(e_5) = 0, \quad (11)$$

$$\Delta c(e_5) + \Delta c(e_8) - \Delta c(e_4) - \Delta c(e_7) = 0, \quad (12)$$

$$\bar{c}(e_1) + \bar{c}(e_2) = 15, \quad (13)$$

$$mc_2(e_i) \leq \Delta c(e_i) + c(e_i) \leq mc_1(e_i), \quad \forall e_i \in E. \quad (14)$$

Among them, formulas (9) and (10) separately correspond to the vertexes v_1 and v_4 who have positive capacity potentials, while formulas (11) and (12) separately correspond to the vertexes v_2 and v_3 who have negative capacity potentials. In addition, in formula (13), e_1 and e_2 indicate incoming arcs, and the incoming flow of network is specified as 15. Let $\Delta c(e_i) = x_i$; then the matrix in terms of the above-mentioned objective function and constraint conditions are illustrated as follows.

The objective function is $\min(Ax)$.

The constraint condition is $C \leq Bx \leq D$,

TABLE 1: Parameters comparison between the original design and the optimum design.

	Arcs capacity distribution	Maximum flow	Minimum saturated flow	Construction cost
Original	(8, 7, 5, 9, 2, 9, 6, 5, 10)	14	9	163
Optimum	(10, 5, 2, 8, 2, 9, 2, 8, 7)	15	15	131

where

$$A = [2 \ 3 \ 2 \ 3 \ 4 \ 2 \ 3 \ 1 \ 4],$$

$$X = [x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ x_6 \ x_7 \ x_8 \ x_9]^T,$$

$$B = \begin{bmatrix} -1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 1 \\ 0 & -1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & -1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$C = [-6 \ -7 \ 5 \ 8 \ 0 \ -6 \ -5 \ -3 \ -7 \ 0 \ -7 \ -4 \ -3 \ -8]^T,$$

$$D = [0 \ 0 \ 5 \ 8 \ 0 \ 2 \ 3 \ 5 \ 0 \ 3 \ 1 \ 4 \ 5 \ 0]^T. \quad (15)$$

The final objective value turns out to be -32 with the help of Lingo software. In other words, the transformation cost decreases by 32 units compared with the traditionally designed network. Table 1 shows the comparison of parameters between the newly designed network and the original one.

4. The Blocking-Resistant Optimum Transformation Design for the Traffic Network

This model is aimed at transforming the prevailing traffic system into a blocking-resistant one. Compared with the aforementioned one, this model appears to be different in existing arcs reconstruction, which requires expansion instead of reduction in capacity variation. As a result, structural blocks are diminished or reduced as many as possible, thus bringing about a totally balanced traffic network [14].

4.1. Model Construction. There are some conditions already known before the traffic network is reconstructed, including the following.

- (1) The cost occurring from expanding unit capacity of each path (i.e., the length of each arc), denoted as b_i , is dependent on facilities removal charges, path's length, and operation complexity.

- (2) The maximum capacity limit for each path is described as mc_{ij} .

- (3) In addition, each path is required to be expanded instead of narrowed; that is, $\Delta c_{ij} \geq 0$.

As for each arc, given the four parameters of $[x_{ij}, c_{ij}, mc_{ij}, b_{ij}]$, among which x_{ij} indicates the current flow volume, c_{ij} the original capacity, mc_{ij} the maximum capacity, and b_{ij} the cost paid to improve unit capacity (if $x_{ij} < c_{ij}$, $b_{ij} = 0$), blocking-resistant transformation is thus to be understood as a problem of how to enlarge every arc's capacity with the least cost. That is, the negative capacity potential vertexes will be increased as many as possible, while the positive ones will be reduced and not less than 0. After being transformed, the network will suffer less structural blocks with improvement in negative capacity potentials.

Let the added capacity value of each arc be

$$\Delta c(e_{ij}) = \bar{c}(e_{ij}) - c(e_{ij}). \quad (16)$$

The objective function is

$$\min \left(\sum \{b(e_{ij}) \cdot \Delta c(e_{ij}) \mid \Delta c(e_{ij}) > 0, e_{ij} \in E\} \right), \quad (17)$$

where $b(e_{ij})$ refers to the cost needed for improving unit capacity of arc e_{ij} .

The constraints are as follows.

- (1) For the negative capacity potential vertex, there is

$$\sum_j \{\Delta c(e_{ij}) \mid e_{ij} \in E\} - \sum_j \{\Delta c(e_{ji}) \mid e_{ji} \in E\} \geq 0. \quad (18)$$

- (2) For the nonnegative capacity potential vertex, there is

$$\begin{aligned} -\Phi_{v_i} &\leq \sum_j \{\Delta c(e_{ij}) \mid e_{ij} \in E\} \\ &\quad - \sum_j \{\Delta c(e_{ji}) \mid e_{ji} \in E\} \leq 0, \quad \forall v_i \in V. \end{aligned} \quad (19)$$

- (3) Furthermore, restrictions on flow volume and capacity for the network are

$$\begin{aligned} x(e) &\leq \bar{c}(e), \quad \forall e \in E, \\ 0 &\leq \bar{c}(e) \leq mc(e), \quad \forall e \in E. \end{aligned} \quad (20)$$

Note that the incoming flow is limited to being F and the maximum capacity of arc e is $mc(e)$.

TABLE 2: Comparison of arcs' capacity and saturated flow before and after transformation.

	The network before transformation	The network after transformation
Arcs' capacity distribution	[8, 7, 5, 9, 2, 9, 6, 5, 10]	[10, 7, 5, 9, 2, 10, 6, 9, 10]
Network's maximum flow	14	17
Network's minimum saturated flow	9	13
Minimum transformation cost	—	10

4.2. *Numerical Case.* Referring to Figure 1, the figures above the arrow line represent x_{ij}/c_{ij} , mc_{ij} , and b_{ij} , and those next to all nodes describe the capacity potentials. Besides, the nodes v_2, v_3 are negative capacity potential vertexes, while v_1, v_4 are positive capacity potential vertexes. Before being transformed, the maximum flow and minimum flow are $F = 14$ and $v = 9$, respectively. The optimum design model is set up as below.

The objective function is

$$\min \left(\sum \{b(e_i) \cdot \Delta c(e_i) \mid \forall e_i \in E\} \right). \quad (21)$$

The constraint conditions are

$$\begin{aligned} -6 &\leq \Delta c(e_3) + \Delta c(e_4) - \Delta c(e_1) \leq 0, \\ -7 &\leq \Delta c(e_7) + \Delta c(e_9) - \Delta c(e_6) \leq 0, \\ \Delta c(e_6) - \Delta c(e_2) - \Delta c(e_3) - \Delta c(e_5) &\geq 0, \\ \Delta c(e_5) + \Delta c(e_8) - \Delta c(e_4) - \Delta c(e_7) &\geq 0, \\ \tilde{c}(e_1) + \tilde{c}(e_2) &= 15, \end{aligned} \quad (22)$$

$$mc_2(e_i) \leq \Delta c(e_i) + c(e_i) \leq mc_1(e_i), \quad \forall e_i \in E.$$

We obtain the objective value of 10 using Lingo software. The optimized network is illustrated in Figure 2, and both the arc capacity and saturated flow before and after transformation are compared in Table 2.

5. Conclusion

In short, the traffic-network blocking-resistant optimum model differs from the blocking-resistant optimum transformation model in nature.

As far as the optimum design model is mentioned, the capacity parameters in original design need to be amended with the objective of preventing the network from structural blocks with the capacity potential being less than zero for each vertex, except for source and intermediate vertexes. As shown in the numerical case, once optimized, the capacity potential of vertexes v_2, v_3 with negative ones in original network becomes 0 and the maximum and minimum saturated flow are equal, thus resulting in a totally balanced network.

In the optimum transformation model, however, the objective is to diminish and reduce structural blocking vertexes as many as possible. Seen from Figure 2, through transformation, the vertex v_2 and vertex v_3 achieve the capacity potentials of -4 and -4 , compared with -5 and -8 without transformation, respectively. That is, the two vertexes

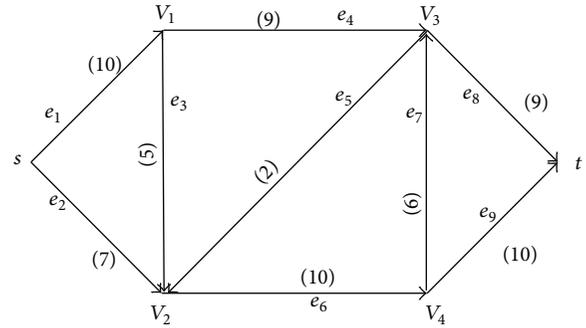


FIGURE 2: Transformed Network.

mentioned above still have negative capacity potentials, which maintain structural blocks in the network. In other words, the network is likely to suffer from blocking phenomena. More importantly, both the maximum and minimum saturated flow will increase simultaneously after experiencing transformation, and thus effectively improving the network.

In conclusion, the optimum model is suitable for the phase during which a traffic network is being designed, while the optimum transformation model is more adaptable for reconstruction of the already established traffic network. In mobilization logistics practice, the optimum transformation model will be commonly utilized for it helps prevent or diminish blocking phenomena by increasing arcs' capacity in accordance with actual conditions of traffic network transformation, which may lay a solid foundation for the research of mobilization logistics system adaptability.

In practice, travel time of saturated flows is one of the essential parameters employed in optimization and transformation designing process. Actually, the travel time in question seems ambiguous and uncertain owing to coefficient of environmental and human factors. Accordingly, to achieve a reasonable resource allocation, our further research will focus on a synthetic optimization combining reconstruction cost, travel time, and saturated flow distribution.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] Q. Zhu, *National Economic Mobilization*, Publishing House of Military Science, Beijing, China, 2nd edition, 2007.

- [2] J. Zhang, "Economic mobilization logistics problems," *Jiangsu Business*, no. 2, pp. 61–63, 2008.
- [3] J. Cui and L. Kong, "The logistics chain of mobilization of national economy," *Journal of Beijing Institute of Technology*, vol. 6, no. 4, pp. 16–19, 2004.
- [4] Q. Zhu, *China's Economic Mobilization Adaptability*, Publishing House of Military Science, Beijing, China, 2007.
- [5] L. Shang and Q. Tan, "The connotation and characteristics of economic mobilization logistics network," *China Circulation Economy*, no. 7, pp. 25–28, 2013.
- [6] J. Liu, *National economy mobilization logistics system construction and the key technology [M.S. thesis]*, Nanjing University of Aeronautics & Astronautics, 2013.
- [7] X. Jing and Q. Tan, "The evaluation of national economy mobilization logistics system's adaptability," *Economy of Technology*, vol. 4, no. 2, pp. 24–28, 2013.
- [8] A. Mehraei, H.-R. Karimi, K.-D. Thoben, and B. Scholz-Reiter, "Using metaheuristic and fuzzy system for the optimization of material pull in a push-pull flow logistics network," *Mathematical Problems in Engineering*, vol. 2013, Article ID 359074, 19 pages, 2013.
- [9] S. Dashkovskiy, H.-R. Karimi, and M. Kosmykov, "Stability analysis of logistics networks with time-delays," *Journal of Production Planning & Control*, vol. 24, no. 7, pp. 567–574, 2013.
- [10] S. Dashkovskiy, H.-R. Karimi, and M. Kosmykov, "A Lyapunov-Razumikhin approach for stability analysis of logistics networks with time-delays," *International Journal of Systems Science*, vol. 43, no. 5, pp. 845–853, 2012.
- [11] H.-R. Karimi, N. A. Duffie, and S. Dashkovskiy, "Local capacity H_∞ control for production networks of autonomous work systems with time-varying delays," *IEEE Transactions on Automation Science and Engineering*, vol. 7, no. 4, pp. 849–857, 2010.
- [12] X. Ning, *Flow Theory and Its Application*, Science Press, Beijing, China, 2009.
- [13] W. Wu and X. Ning, "Anti blockage transformation of an emergency evacuation network," *Systems Engineering*, vol. 21, no. 3, pp. 244–247, 2006.
- [14] W. Wu and X. Ning, "Simulation research on improving the flow capability of emergency networks," *Chinese Management Science*, vol. 14, no. 3, pp. 86–90, 2006.

Research Article

A Multiobjective Optimization Model of Production-Sourcing for Sustainable Supply Chain with Consideration of Social, Environmental, and Economic Factors

Zhixiang Chen¹ and Svenja Andresen²

¹ School of Business, Sun Yat-sen University, Guangzhou 510275, China

² WFI, Ingolstadt School of Management, Catholic University of Eichstätt-Ingolstadt, 85072 Eichstätt, Germany

Correspondence should be addressed to Zhixiang Chen; mnsczx@mail.sysu.edu.cn

Received 20 January 2014; Revised 31 March 2014; Accepted 4 April 2014; Published 27 April 2014

Academic Editor: Hamid Reza Karimi

Copyright © 2014 Z. Chen and S. Andresen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper incorporates the three pillars of sustainability—economic, environmental, and social dimensions—into a supply chain. A multiobjective programming model which jointly minimizes costs, emissions, and employee injuries in a supply chain is first constructed. Using the weighted-sum approach with weights setting by the analytic hierarchy process (AHP), the model is solved by normalization of the minima of the three objectives. A numerical example is conducted to test the model. The results show that it is indeed possible to integrate environmental and social metrics in supply chain system optimization. Multiobjective optimization can balance the social, environmental, and economic performance. This paper presents a new multidimension perspective for optimizing supply chain; it will inspire practitioners to change their decision ideas and improve supply chain sustainability.

1. Introduction

Sustainable development is becoming a world-wide hot topic in academics and practice. “If everyone used energy and resources the same way we do in the Western World, we would need three more earths at least. And we have only one.” [1]. The former Minister for Sustainable Development in Sweden addressed one of the major challenges we have to face in the 21st century—the scarcity of resources. More and more countries, institutions, and businesses recognize that the world population has to change its behavior to tackle problems that might easily become or already are constraints to international growth. In a report for the OECD, Strange and Bayley [2] draw up a great past in terms of economic growth with incomes having increased eightfold since 1820, but they also suggest that there are “clouds on the horizon.” Salim [3] states that “unsustainable development has degraded and polluted the environment in such a way that it acts now as the major constraint followed by social inequity that limits the implementation of perpetual growth.” The bad effect of pollution, also on economic wealth, has been

examined especially in the context of climate change. The Climate Vulnerability Monitor by the NGO DARA reveals the impact of global warming on every aspect of life [4]. It estimates that in 2010 the global economy has to bear total losses of around 1.2 trillion dollars due to the carbon emission and climate change. Different from other consequences, those losses are also significant in the industrialized world [4].

There are other driving factors, such as customer awareness, media interference or governmental regulations that push the development of sustainable solutions [5]. For example, in 2005, the Emissions Trading System had been launched in the European Union. It covers more than 11,000 power stations and industrial plants in 31 countries. This “cap-and-trade” approach gives high-emitting industry sectors the choice to either cut down on pollution or to bear the cost of buying emission allowances [6]. This is definitely an incentive for European companies to pursue greener production. Besides the environmental problem, corporate social responsibility is also a growing issue. In 2012, a TV news report about the working conditions in manufacturing sites for jeans in China triggered countrywide

discussions in Germany [7]. And this is just one media coverage amongst many. There are also many other events of media interfering sustainable development, such as, the German magazine “Wirtschaftswoche” denounced textile factories in Bangladesh [8], the US newspaper “The New York Times” wrote about the human costs of an iPad [9], and, even in China, the “China Daily” called for an improvement of migrant workers’ situation [10]. Nowadays, consumers thereby are becoming more and more concerned about the sustainability of companies’ activities.

Under this international background of sustainable development, sustainability management becomes increasingly important, and supply chain management surely cannot be excluded from this stream of sustainable development. On the contrary, supply chain managers are significantly involved in companies’ operations and bear more and more responsibilities in essential decision making processes, such as the selection of suppliers, transport modes, or production methods during sustainable manufacturing [11]. In academics, sustainable supply chain management (SSCM) has become a new hot topic for research; more and more authors are researching the problem of sustainable supply chain strategy and operations management.

While there is an abundance of supply chain models that integrates economic and environmental goals concerning green supply chain (e.g., [12–16]), there is only a few of them considering the social layer issues (e.g., [17, 18]). Carter and Rogers [19] also criticize that general management as well as the operations management literature often focuses on the ecological dimension of sustainability and neglects social aspects.

In the literature of the area of production and logistics system optimization, there are some authors who have studied the complex production-supply logistics systems; Ait-Alla et al. [20] study the robust production planning in an apparel industry under demand uncertainty via conditional value at risk. Karimi et al. [21] study a class of production networks of autonomous work systems with time varying delays in capacity changes. Mehraei et al. [22] address a pull-push flow strategy in a supply network and propose a novel solution for optimizing the pull side employing conwip system using metaheuristic and fuzzy system optimization method. Mehraei et al. [23] study production-logistics system using simulation method, in which a learning autonomous pallets’ concept is used in a discrete simulation model to analyze and compare several decentralized control strategies.

However, all those studies do not consider the social objectives. Using different previous research, in this paper, we take an iron and steel factory as case background, considering the above three sustainable factors, study the complex multiobjective optimization model of production-sourcing logistics system. In this model, several raw materials are supplied from different suppliers, and then the iron and steel factory produces products to meet customers demand. Through constructing a multiobjective programming model by concurrently considering social, environmental, and economic factors, we demonstrate the benefit of sustainable supply chain system optimization based on multiple factors instead of single economic objective.

This paper is organized as follows. First, a definition of sustainability in general and sustainable supply chain management in particular is developed based on the literature review. Then, based on existing research, an optimization model of sustainable supply chain system based on multi-objective programming approach is constructed and, thirdly, the corresponding solution method is developed. Fourthly, a numerical example is used to demonstrate the application of the model. At last, conclusion and future research directions are summarized.

2. Definition and Framework of Sustainable Supply Chain

(1) Sustainability. The idea of sustainability goes back to the mid-ages. In 1713, Hanns Carl von Carlowitz, head of the Royal Mining Office in the Kingdom of Saxony, in order to meet the challenge of a predicted shortage of timber—the key resource of the time—requested that the number of trees that are cut down may not exceed the amount of new trees growing [24]. This idea tells that the nature resource of the earth is limited, and human being should moderately utilize the nature resource.

In modern age, the discussion about sustainability was initiated in 1972 when the Club of Rome published its first report “Limits to growth” [5]. One of the milestones in the development of the sustainability definition was a UN (United Nations) report commonly known as Brundtland Report, and the original name of the report was “Our common future” [25]. It is notable to mention that the former Norwegian Prime Minister, Gro Harlem Brundtland, shaped not only the name of this memorandum but the whole definition of sustainability for years to come [25]. Even today, more than 30 years later, “development that meets the need of the present without compromising the ability of future generations to meet their needs” (World Commission on Environment and Development [25]) is still the basis for most definitions of sustainability.

One of the most famous literatures when it comes to sustainability is Elkington’s book “Cannibals with Forks: The Triple Bottom Line of the 21st Century” [26]. He compares sustainability to a fork with three prongs. The three prongs stand for the economic, environmental, and social performance of an organization. The idea of the triple bottom line promotes the simultaneous optimization of all three dimensions. Carter and Rogers have transformed this idea into a figure (see Figure 1) and conclude that “there are activities that organizations can engage in which not only positively affect the natural environment and society, but which also result in long-term economic benefits and competitive advantage for the firm” [19].

These three aspects of sustainability are sometimes also called 3 Ps: profit, planet, and people [5]. Dyllick and Hockerts [27] describe the triple bottom line as 3 different types of capital: economic, natural, and social capital. In order to reach corporate sustainability, they conclude that all three types of capital have to be optimized.

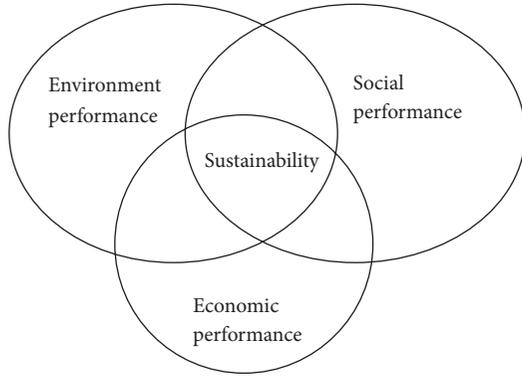


FIGURE 1: The triple bottom line of sustainability [19].

Nowadays, the notion of sustainability has been defined on a macroeconomic basis as well as in the sense of corporate sustainability [26, 27]. In this paper, we will base on the idea of sustainability to construct the optimization model of sustainable supply chain management.

(2) *Sustainable Supply Chain Management*. There is no consensus definition on the sustainable supply chain management. Carter and Rogers [19] define sustainable supply chain management (SSCM) as follows.

The strategic, transparent integration and achievement of an organization’s social, environmental, and economic goals in the systemic coordination of key inter-organizational business processes for improving the long term economic performance of the individual company and its supply chains.

Tang and Zhou [28] use the PPP (profit, people, and planet) approach to construct a sustainable supply chain model (see Figure 2). In Figure 2, the economic dimension is depicted by the blue arrows (i.e., the central part of the figure, from the box of “producers” to the box of “supply chain” and the box of “consumers”), the green arrow represents the environmental problems (i.e., from the box of “wastes and emissions” to the box of “natural resource” in the figure) and the red arrow stands for social facets (i.e., from the box of “producers” to the box of “consumers”).

All of the SSCM definitions have one thing in common: they emphasize the necessity to simultaneously consider all three dimensions within the supply chain; that is, while considering the economic performance of supply chain, improving social and environmental impact also has to be considered. Cetinkaya et al. [29] employ a three-dimension metrics system to describe the sustainability, that is, social, economic, and environmental dimensions, and each of the dimensions is broken down into other three subdimensions (see Figure 3). They state that an improvement in any of these subdimensions, without a negative effect on another subdimension, will ultimately lead to a more sustainable supply chain.

In order to make the sustainable supply chain quantifiable, in this paper, we will follow the approach of Cetinkaya et al. [29], as it is necessary to find and categorize metrics. According to this framework, we will construct a series of metrics of sustainable supply chain, and then under

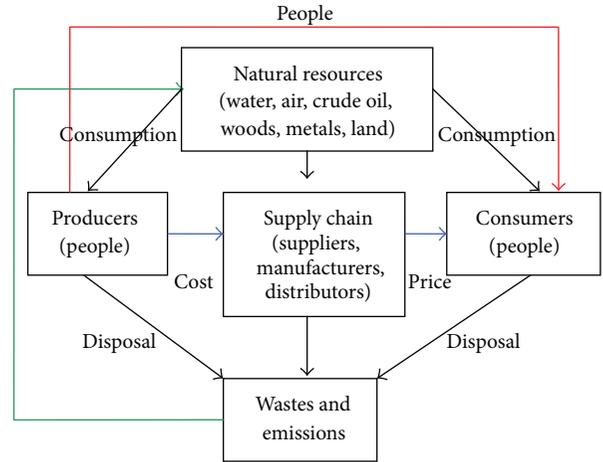


FIGURE 2: Supply chain integrated in the PPP ecosystem [28].

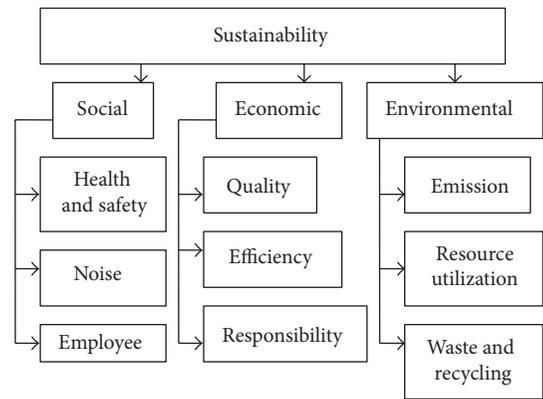


FIGURE 3: Metric dimensions and subdimensions of sustainability [29].

these three objective dimensions—social, economic, and environmental dimensions, to construct a multiobjective programming model.

3. Problem Definition and Modeling

Before applying the mathematical optimization model, we first define the assumptions of the supply chain and then formulate the optimization problem.

3.1. *Supply Chain Model and Problem Definition*. The optimization problem proposed in this paper is simulated from the production facility’s perspective with only one period of time under consideration. In the model, a production facility produces one type of output, which is shipped to a group of customers. The raw materials are sourced from a finite number of suppliers. The suppliers ship the raw materials via different transport modes to the production facility.

For the optimization of the supply chain, several assumptions are made as follows.

- (1) The supply is organized in a just-in-time manner, so that the effect of incoming items inventory can be neglected.

- (2) The ordered quantities of all raw materials will meet the needs of the production.
- (3) There is more than one type of raw materials sourced from several suppliers.
- (4) The production facility, suppliers, and transportation options have capacity constraints.
- (5) All suppliers together can fulfill the ordered quantities needed for production and meet the demand of customers.
- (6) The demand of the customers is known and the production facility is able to satisfy the demand.
- (7) Outgoing items inventory will remain insignificantly small and is therefore neglected.
- (8) Transportation costs depend on the distance from suppliers to production facility, as well as on the quantity shipped.
- (9) Production emissions depend on the chosen production method, as well as the produced quantity, while transportation emissions are dependent on the shipped quantity of the respective material and the distance from the suppliers to the production facility.

The proposed model gives answers to the following questions (established as decision variables).

- (i) How much quantity of the respective raw material should be ordered from which supplier?
- (ii) How much quantity of a raw material should be shipped using what transport mode from the respective supplier to the production facility?
- (iii) Which production method should be chosen?

Based on viewpoint of sustainable supply chain, the optimization model of supply chain includes three objectives: economical, environmental, and social objectives. Subsequently, the three dimensions should be optimized simultaneously. According to the subdimensions of a sustainable supply chain presented in Cetinkaya et al. [29], several metrics can be chosen to quantify the three aspects under consideration. Due to the limited space of the paper, therefore, for each dimension, only one metric will be selected to measure the supply chain's performance in the respective field. For the sake of computation convenience, the chosen indicators should be simple, understandable, easy to reproduce, comparable, complementary to regulatory programs, cost-effective in data collection, stackable and scalable, useful as a management tool, and protective of company information [30].

The first objective is to achieve economic sustainability. The three subdimensions of economic sustainability are quality, efficiency, and responsiveness. As for Cetinkaya et al.'s [29] statement, efficiency is "the most familiar dimension for all managers" (page 61). This is also reflected by Székely and Knirsch [30], who study the metrics used in 20 different companies' annual reports and sustainability reports. There are several measures for efficiency: utilization, productivity, cost reduction, and service level. For the purpose of this

paper and under the assumptions, minimizing costs will be chosen to represent the economic dimension. The cost of the production facility is composed of production costs, transportation costs, and material costs for the sourced raw materials.

The environmental dimension consists of the following components: emissions, natural resources utilization and waste, and recycling [29]. Tang and Zhou [28], as well as Dekker et al. [11], pinpoint that one of the most important ways to estimate environmental impact is the measurement of emissions and pollution. This is the fact due to "more stringent government regulations and increasing awareness of environmental protection among consumers and society" ([28], page 591). The two major sources of emission in supply chains are production operations and transportation. Therefore the environmental objective in the model will be assessed using the emissions of the production method and the transportation options. As shown by Daniel et al. [31], pollution can be classified into three categories: water, air, and soil pollution. If applicable, harm to all three areas should be considered.

Last but not least, probably the most difficult part is the social dimension. Health and safety, employees, and noise are the three subdimensions determined by Cetinkaya et al. [29]. One of the most commonly adapted metric in this field is the accident and illness number, or rather the accident/illness number in relation to working hours [30]. This measure will be taken as a basis and modified to accommodate the severity of different accidents and illnesses.

3.2. Mathematical Formulation of the Model. Having completed the general model description and problem introduction, we can now formulate the problem. At first, indices and variables will be defined, followed by the construction of the objective functions and the constraint functions.

3.2.1. Indices and Notations

Indices

- i ($1, \dots, I$): supplier;
- r ($1, \dots, R$): raw material;
- m ($1, \dots, M$): production method;
- t ($1, \dots, T$): transport mode;
- p ($1, \dots, P$): pollutant;
- n ($1, \dots, N$): severity class of injury.

Decision Variables

- Q_i^r : quantity of raw material r bought from supplier i ;
- PM_m : 1, if production method m is chosen, 0 otherwise;
- TQ_{it}^r : quantity of raw material r , shipped from supplier i , using transport mode t .

Cost Parameters

C_i^r : cost per unit of raw material r bought from supplier i ;

H_m : hours of work for one unit of product under production method m ;

L : labor cost per hour;

DP_m : depreciation per unit of product, if production method m is chosen;

UC_m : utility cost per unit of product, if production method m is chosen;

X : quantity of products produced in units (equals demand of product);

TC_{ti}^r : cost for transportation per kilometer of one unit of raw material r with transport mode t from supplier i ;

d_i : distance from supplier i to the production facility.

Pollution Parameters

A_t^{pr} : pollution through pollutant p per km and unit of raw material r , if transport mode t is chosen;

A_m^p : pollution through pollutant p per unit of product, if production method m is chosen.

Safety Parameters

I_m^n : number of injuries in severity class n , if production method m is chosen.

Constraint Parameters

PCA_m : capacity of production method m ;

TCA_{ti}^r : capacity of transportation mode t for raw material r , on route from supplier i to the production facility;

SCA_i^r : capacity of supplier i for raw material r ;

U_m^r : usage of resource r in production method m for one unit of product.

3.2.2. *Objective Functions*

(1) *Economic Objective.* As mentioned above, the economic objective will be included by minimizing the costs within the supply chain. The production costs are assumed to be mainly labor costs, depreciation costs (for the machinery), and utility costs. That is,

Minimize $F_1 =$ Material cost + Production cost + Transportation cost,

$$\begin{aligned} \min F_1 = & \sum_{r=1}^R \sum_{i=1}^I C_i^r Q_i^r \\ & + \sum_{m=1}^M (H_m X L + DP_m X + UC_m X) P M_m \quad (1) \\ & + \sum_{i=1}^I \sum_{r=1}^R \sum_{t=1}^T TC_{ti}^r d_i TQ_{ti}^r. \end{aligned}$$

(2) *Environmental Objective.* The environmental dimension is optimized by minimizing emissions within the supply chain, that is, pollution created by transportation and by production. That is,

Minimize $F_2 =$ Transportation emissions + Production emissions,

$$\min F_2 = \sum_{p=1}^P \sum_{r=1}^R \sum_{t=1}^T \sum_{i=1}^I A_t^{pr} TQ_{ti}^r d_i + \sum_{m=1}^M \sum_{p=1}^P A_m^p X P M_m. \quad (2)$$

(3) *Social Objective.* The foundation for the social objective function is the injury/illness incidence rate used by the American Bureau of Labor Statistics [32], which is formulated as

Incidence rate = (Number of injuries and illnesses * 200,000)/Employee hours worked.

The 200,000 hours in the formula represents the equivalent of 100 employees working 40 hours per week, 50 weeks per year, and provides the standard base for the incidence rates. The formula includes all nonfatal work-related injuries and illnesses that are recordable.

For the purpose of this paper, the formula will be amended to include the severity of injuries and illnesses. To achieve this, it is proposed that all injuries/illnesses will be categorized according to the days absent from work. The injuries/illnesses that are less severe will be given a low severity level, while the highest rank will be work-related fatalities. For example injuries/illnesses could be ranked as follows (see Table 1).

To represent the severity of the injuries/illnesses, an exponential function will be applied. The exponential function can show that the higher the class, the more unfavorable an incidence. In this way, the model can guide the decision maker (in our case the company) to employ production methods that will especially decrease the danger of severe injuries or illnesses. On the other hand, minor incidences become acceptable to a certain degree. Largely those small accidents (like cuts or tripping over) are most probably not even preventable.

In the example, cases that are in category 4 will therefore be included in the injury/illness incidence rate with factor 1, while incidents in lower classes are less weighted and occurrences in higher categories become more important (see Table 2).

TABLE 1: Severity classes of work-related injuries/illnesses.

n	Severity class description
1	Less than a week absent from work
2	Less than a month absent from work
3	Less than 3 months absent from work
4	Less than 6 months absent from work
5	Less than a year absent from work
6	Occupational disability (the person will not be able to work in the previous job again)
7	Fatalities

TABLE 2: The exponential severity function.

n	$n - (\sum n)/N$	$e^{n - (\sum n)/N}$
1	-3	0.049787
2	-2	0.135335
3	-1	0.367879
4	0	1
5	1	2.718282
6	2	7.389056
7	3	20.08554

According to this, the third objective function can be modeled, which includes the rephrased injury/illness incidence rate as follows:

$$\min F_3 = \sum_{m=1}^M \text{PM}_m \sum_{n=1}^N e^{n - \sum n/N} I_m^n \frac{200,000}{H_m X}. \quad (3)$$

3.2.3. *Constraints.* Consider that

$$\sum_{i=1}^I Q_i^r = \sum_{m=1}^M \text{PM}_m U_m^r X, \quad \forall r, \quad (4)$$

$$\sum_{m=1}^M \text{PM}_m \text{PCA}_m \geq X, \quad (5)$$

$$Q_i^r = \sum_{t=1}^T \text{TQ}_{ti}^r, \quad \forall r, i, \quad (6)$$

$$\sum_{t=1}^T \sum_{i=1}^I \text{TQ}_{ti}^r = \sum_{i=1}^I Q_i^r, \quad (7)$$

$$\text{TCA}_{ti}^r \geq \text{TQ}_{ti}^r, \quad \forall r, i, t, \quad (8)$$

$$\text{SCA}_i^r \geq Q_i^r, \quad \forall r, i, \quad (9)$$

$$\sum_{i=1}^I \text{SCA}_i^r \geq \sum_{i=1}^I Q_i^r, \quad \forall r, \quad (10)$$

$$\text{PM}_m = \{0, 1\}, \quad \forall m, \quad (11)$$

$$Q_i^r \geq 0, \quad \forall i, r, \quad (12)$$

$$\text{TQ}_{ti}^r \geq 0, \quad \forall r, i, t. \quad (13)$$

The first constraint (4) establishes the assumption that the quantities of a raw material ordered at all suppliers together have to meet the needs of the production. Besides, the capacity of the chosen production method has to be able to fulfill the customers' demands as constraint (5). Constraints (6) and (7) show that the shipped quantities have to equal the ordered quantities, individually for each supplier and as a sum. Constraints (8), (9), and (10) display the capacity limits of the suppliers and the transportation modes. Moreover, the production method variable can only take a value of 0 or 1, which is the constraint (11). The other two decision variables must be nonnegative, which is shown in constraints of (12) and (13).

4. Solution Method

Now that we have compiled all the objectives and constraints, the solution method can be studied more thoroughly. As introduced in the literature review section, the weighted sum model will be used to solve the multiobjective optimization problem. The weighted sum model has the following general form [33]:

$$\begin{aligned} \min \quad & \sum_{i=1}^k w_i f_i(x) \\ \text{with } & w_i \geq 0, \quad \forall i, \\ & \sum_{i=1}^k w_i = 1. \end{aligned} \quad (14)$$

For our model, this general form can be reformulated into

$$\begin{aligned} \min \quad & w_1 F_1 + w_2 F_2 + w_3 F_3 \\ \text{with } & w_i \geq 0, \quad \forall i, \\ & \sum_{i=1}^3 w_i = 1. \end{aligned} \quad (15)$$

If the objectives, like in our case, have a different magnitude, the objective functions have to be normalized. In this case, the general form changes to [33]

$$\begin{aligned} \min \quad & w_1 \theta_1 F_1 + w_2 \theta_2 F_2 + w_3 \theta_3 F_3 \\ \text{with } & w_i \geq 0, \quad \forall i, \\ & \sum_{i=1}^3 w_i = 1. \end{aligned} \quad (16)$$

w_i is the weight assigned by the decision maker and θ_i is the normalization factor. Thus, before the optimization problem can be solved, weights and normalization factors have to be computed.

4.1. Weights Setting. There are various methods to model a decision maker's preferences into weights. One of the most famous methods is the AHP method, which was first introduced by Saaty [34]. AHP stands for analytic hierarchy process; it is a method for scaling weights of the elements in each level of a hierarchy with respect to an element (i.e., a criterion or objective) of the next higher level [35].

In order to set the weights of objectives, in our case, we assume that

- (i) the economical performance (factor 1) is slightly more important than the environmental performance (factor 2) of the supply chain;
- (ii) the economical performance is moderately more important than the social performance (factor 3);
- (iii) the environmental performance is slightly more important than the social performance (factor 2).

Based on this assumption, the weights for the different objectives can be derived through calculating the eigenvector of the evaluation matrix (see Table 3) (the detail procedure of AHP is omitted).

According to the calculated weights, the optimization problem (17) can be established:

$$\min 0.5396\theta_1 F_1 + 0.2970\theta_2 F_2 + 0.1634\theta_3 F_3. \quad (17)$$

4.2. Normalization. Last but not least, the objective functions have to be normalized. Mausser [33] proposes three different methods of normalization:

- (i) normalization by the magnitude of the objective function at the initial point x_0 ;
- (ii) normalization by the minimum of the objective functions;
- (iii) normalization by the difference of the optimal values in the Nadir and Utopia points.

Although Mausser [33] states that the normalization using the Nadir and Utopia points might render the best results, for the purpose of this paper, the more simple method of applying the minimum of the objective function will be used. As neither of the objective functions' minima will equal zero, this method is acceptable. In this case, the normalization factors will be calculated by

$$\theta_i = \frac{1}{f_i(x^{[i]})}, \quad (18)$$

where $x^{[i]}$ solves $\min_x \{f_i(x) : x \in \Omega\}$.

Thus the solution function takes on the following form:

$$\begin{aligned} \min & 0.5396 \frac{1}{\min F_1} F_1 + 0.2970 \frac{1}{\min F_2} \\ & + 0.1634 \frac{1}{\min F_3} F_3. \end{aligned} \quad (19)$$

TABLE 3: Weights for the sustainable supply chain objectives.

Objectives	Weights (w_i)
Economical performance	0.5396
Environmental performance	0.2970
Social performance	0.1634
	$\sum w_i = 1$

5. Numerical Example

After having formulated and scalarized the multiobjective optimization problem, we can now continue with testing the effectiveness of the model using example. The simulated production facility produces steel, for which the two main inputs are iron ore and coking coal. Within the time period under consideration, the steel mill has a hypothetical output of 500,000 tons of steel. It will be assumed that the site is located in Central Europe. Potential suppliers for iron ore are located in India, Brazil, and Australia, while suppliers for coking coal can be found in Australia, USA, and Canada. Those countries are, respectively, the biggest exporting nations. The possible transportation modes comprise either a combination of ship and rail or a combination of ship and truck. Last but not least, the company can choose from three different production methods, with the assumption that the first one (PM1) is the cheapest, the second one is the safest (PM2), and the third one is the most environmental friendly (PM3).

5.1. Setting Parameters. Before solving the model, we need to set parameters for the model. These parameters comprise the cost parameters, pollutant emissions parameters, injury parameters, and constraint parameters.

5.1.1. Cost Parameters. At first, the parameters constituting costs have to be established. As demonstrated in the model, costs comprise material, transportation, and production costs.

(1) **Material Costs.** The estimations for the material costs are based on the development of prices on the commodity exchange for iron ore and coal over the last few months [36]. For different suppliers, located in different countries, the material prices may be different. Table 4 lists estimations of prices for iron ore and coking coal depending on suppliers, that is, C_i^r .

(2) **Transportation Costs.** In our model, the transportation cost parameter TC_{ii}^r denotes the cost for transportation per kilometer of one unit raw material r with transport mode t from supplier i . For simplification, we assume that transportation costs are only related to the transportation mode. Beresford et al. [37] have investigated the prices of multimodal supply chains of iron ore between Australia and China. The prices used for our simulation were derived from their study, taking into consideration the average distance traveled by ship and the average distance traveled by rail or truck. The result is listed in Table 5.

TABLE 4: Material prices for iron ore and coking coal (C_i^r).

Price in \$/ton	India	Brazil	Australia	USA	Canada
Iron ore	120	130	140	n/a*	n/a*
Coking coal	n/a*	n/a*	130	140	150

* Note: only the biggest exporting nations are taken into consideration.

TABLE 5: Transportation cost of transportation modes (TC_i^r).

Transportation mode	Ship + rail	Ship + truck
Cost in \$/ton/km	0.0038	0.0037

As regards the estimation of distances between the mine and the steel production facility (d_i), for simplification, the air-line distances are used, which are listed in Table 6.

(3) *Production Costs.* The production cost parameters are oriented at the steel production cost model spreadsheet provided by MetalMiner [38]. At first, the depreciation of machines and equipment has to be considered. The higher the initial cost of acquisition, the higher the cost for depreciation. Besides, utility costs arise in the production processes also have to be included. Table 7 lists the parameters of depreciation of machine for different production methods, that is, DP_m . Table 8 shows the utility cost per unit of product if production method m is chosen, that is, UC_m .

Additionally, labor costs contribute towards production costs. In a less automated production environment, more manual labor will be required (production method 1, i.e., PM1). Technical advanced production method on the other hand would reduce the labor hours required to produce a ton of product, through increased productivity (production method 3, i.e., PM3). The hourly wages will be estimated with \$20 based on German base rates for steel workers (IG Metall [39]). Table 9 shows the labor hours required to produce one unit product for three production methods.

5.1.2. *Pollution Parameters.* The second objective of the mathematical model is the minimization of pollution emissions, including emissions in the production and transportation.

(1) *Production Emissions.* The first part of supply chain pollutant emissions considered in the model is the production emissions. The base case for emissions in steel production is deduced from ThyssenKrupp's sustainability report [40] based on a German steel manufacturer. Table 10 presents the emissions for the three production methods.

(2) *Transport Emissions.* The second part of the supply chain significantly adding to total pollution is the transportation of iron ore and coking coal from the mines to the production site. Using average distances for sea and land haulage, the following data are concluded from Dekker et al. [11] estimation of average transportation emission. Table 11 shows the emission in different transportation methods.

5.1.3. *Injury Parameters.* The third part of the objective function is the social objective, which is modeled as injury/illness

TABLE 6: Supplier distance to production facility (d_i).

Supplier location	India	Brazil	Australia	USA	Canada
Distance in km	6700	10500	17500	7000	7100

TABLE 7: Depreciation of machines (DP_m).

Production method	PM1	PM2	PM3
Depreciation in \$/ton	15	25	25

TABLE 8: Utility costs in steel production (UC_m).

Production method	PM1	PM2	PM3
Utility cost in \$/ton	4	7	5

TABLE 9: Labor hours required to produce one ton of steel (H_m).

Production method	PM1	PM2	PM3
Labor hours in h/ton	0.5	0.48	0.49

TABLE 10: Steel production emissions (A_m^P).

Emission in g/ton	PM1	PM2	PM3
CO ₂	566513.52	550289.87	509198.9
NO _x	518.15	499.9	489.05
PM*	167.75	167.75	160.07

Note: * particulate matter.

TABLE 11: Emissions in different transportation modes (A_i^{Pr}).

Emission in g/ton/km	Ship + rail	Ship + truck
CO ₂	7.898	9.842
SO _x	0.2161	0.2114
NO _x	0.1357	0.1328

number incidence rate. Due to a lack of nominal figures in this field, the numbers themselves are pure assumptions. But certain guidance about the average injury incidence figure was provided by ThyssenKrupp's sustainability report [40] and the BLS figures (Bureau of Labor Statistics, [41]) for the steel industry. Table 12 specifies the injury classes and the respective number of injuries/illnesses for different production methods.

5.1.4. *Constraint Parameters.* We now set parameters for the constraints. Production capacity (PCA_m) is set at 1,000,000 tons for the assessed period. All other constraint parameters are presented in Tables 13, 14, and 15.

In the model, it is assumed that the transportation capacities of different suppliers for different raw materials are the same for the same transportation mode, as the density of iron ore and coking coal is very similar. Thus, we only list the transportation capacities for different transportation modes in Table 15.

5.2. *Model Solutions.* The model is mixed integer linear problem with integer variable PM_m , continuous variables Q_i^r

TABLE 12: Injuries according to severity (I_m^n).

Severity (n)	Production method (m)		
	PM1	PM2	PM3
1	20	17	18
2	6	5	5
3	3	2	2
4	2	1	2
5	1	1	1
6	1	0	1
7	1	0	0

Note: the severity classes from 1 to 7 have been defined in Table 1.

TABLE 13: Usage of iron ore and coal in production (U_m^r).

Usage in ton/ton	PM1	PM2	PM3
Iron ore	1.765	1.766	1.763
Coking coal	0.696	0.697	0.695

TABLE 14: Capacities of different suppliers (SCA_i^r).

Capacity in: 1000 ton	India	Brazil	Australia	USA	Canada
Iron ore	100	500	600	0	0
Coking coal	0	0	200	100	200

and TM_{ii}^r , the Solver type is Branch and Bound Solver of LINGO 12.0. Global optimization solution is obtained in 3 seconds CPU time at the computer configuration as: Intel Core i7-2600CPU, 3.4 GHz, 4.00 RAM, and 32 bit operation system. For solving the model, it has to be run independently for each objective to compute the minima for normalization (5.1, 5.2, and 5.3).

$$\begin{aligned} \min F_1 &= 236,728,200; \\ \min F_2 &= 370783.8 \text{ (emissions are scaled in tons);} \\ \min F_3 &= 4.980914. \end{aligned}$$

Using the proposed multiobjective optimization model, the production facility's optimal decision would be to choose production method PM2. Raw materials order quantities from difference suppliers are presented in Table 16 and the shipped quantities of transportation modes are displayed in Table 17.

Those solutions for the decision variables will lead to total costs of \$243,701,800 (cost per ton of steel \$487.4). Furthermore a total of 391,699.5 tons of pollutants are emitted, with production accounting for 275,478.8 tons and transportation for 116,220.8 tons. The injury and illness incidence rate will be equal to 4.879263.

5.3. Discussion of Results. First and foremost, the results of the numerical example show that the constructed model is indeed feasible and effective. This is the most important finding. Beyond that, one can draw even more conclusions from the solutions.

Firstly, it is revealed that a multiobjective optimization approach presents different results from the single objective

TABLE 15: Capacities of different transportation modes (TCA_{ii}^r).

Transportation mode	Ship + rail	Ship + truck
Capacity in ton	400,000	100,000

TABLE 16: Order quantities of iron ore and coking coal from the different suppliers.

Order quantities in ton	India	Brazil	Australia	USA	Canada
Iron ore	100,000	500,000	283,000	0	0
Coking coal	0	0	48,000	100,000	200,000

functions alone. For example, production method PM1 is chosen when the economic objective (F_1) is minimized independently. However, when the multiple objective optimization model is applied, production method PM2 is identified as the optimal production method. The same outcome can be found when examining the shipped quantities. If only costs are considered, the truck will be chosen much more frequently. If the environmental performance is added to the model, on the contrary, railway transportation becomes the favourable option (see Table 17).

Secondly, one can learn from the results that a multi-objective optimization model makes it possible for companies to accept reasonably higher costs, if other factors are taken into consideration. Hence, the minimum of costs is at \$236,728,200, but the final amount of money spent in the supply chain equals an amount of \$243,701,800. Those slightly higher costs are the price one has to pay for a safer and more ecological production and transportation. Still, if the costs would rise enormously, making the company economical uncompetitive, the bad economic performance would outweigh the other factors and the multiple objective optimization model would compute different results. This is the consequence of the simultaneous consideration of all three dimensions.

Lastly, the numerical example enlightens the importance of the social dimension. As mentioned before, corporate social responsibility has widely been neglected in operational research literature so far. Incorporating the employee safety into the multiple objective optimization model leads to a surprising result. Through the exponential function, production methods, such as PM1, that have a higher risk of fatalities or severe injuries, become unacceptable. Even production method PM3, which produces much less emissions, is unfavourable compared to PM2. Due to a better incidence rate on the one hand (compared to PM1 and PM3) and lower costs on the other hand (compared to PM3), PM2 is chosen in the multiple objective optimization model. Hence, clearly the constructed model adds information to a company's decisions and can therefore reach different results, which can lead to improved supply chains from a sustainability perspective.

6. Conclusion

Sustainable supply chain has become one important operations strategy for corporate competition, especially for

TABLE 17: Raw material quantities shipped via different transportation modes.

Shipped quantities in ton	Iron ore		Coking coal	
	Ship + rail	Ship + truck	Ship + rail	Ship + truck
India	100,000	0	0	0
Brazil	400,000	100,000	0	0
Australia	283,000	0	48,500	0
USA	0	0	100,000	0
Canada	0	0	200,000	0

large international corporations. Many firms have an annual report of CSR (corporate social responsibility). Under this background, it is necessary to incorporate CSR into supply chain optimization model. This paper addresses this problem, incorporates the three pillars of sustainability—economic, environmental, and social dimensions—into a supply chain optimization model. The purpose is to give supply chain managers a tool for decision making that is not only based on reducing costs or maximizing profit but also focuses on environmental protection and social responsibility. Research result shows that this idea is feasible.

Naturally, there are also certain deficiencies in the model, mainly resulting from the limited scope of the work. Some improvement works can be done to enhance the application value of the model. Firstly, more metrics for each of the dimensions might be considered to better reflect the reality. Also the model might be extended to accommodate more than one period and/or uncertainty in demand. From a mathematical viewpoint, testing different multiobjective solution methods might lead to better or more reliable results, as the weighted sum model yields problems, if nonconvex Pareto curves exist. Further enhancements could be made by using the difference of Nadir and Utopia point to normalize the objective functions.

Despite the described shortcomings, the existing simple model has already been proven to be effective. The program has shown that including different criteria might lead to changes in making choices. For example, when only considering costs, production method PM1 would have been chosen. But when the model also comprises the other two dimensions, production method PM2 is selected. Thus, including sustainability criteria into a company's decision making is not extremely arduous or unrealistic. Considering the pressure from environment and society, corporations should make amends to their existing decision support systems. Bearing in mind new governmental regulations, the cost of climate change or the increasing awareness of customers of environmental protection and social responsibility, corporations should give up a pure focus on profit maximization. In the long-run, sustainable supply chain models will certainly pay off, even though it might initially involve expensive investments. As introduced in the literature review, multiobjective optimization is able to process more information than a conventional single-objective method. Thus, using multicriteria decision model might eventually lead to improved choices. A multiobjective model like the one introduced in this paper can provide guidance to supply chain managers in a time when circumstances are changing quickly.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This paper is funded by NSFC (Grant no. 71372154).

References

- [1] M. Sahlin, "Advancing sustainable development in Sweden," in *Institutionalizing Sustainable Development*, chapter 1, pp. 15–17, OECD, Paris, France, 2007.
- [2] T. Strange and A. Bayley, *Sustainable Development: Linking Economy, Society, Environment*, OECD Insights, OECD Publishing, 2008.
- [3] E. Salim, "The paradigm of sustainable development," in *Institutionalizing Sustainable Development*, chapter 3, pp. 25–30, OECD, Paris, France, 2007.
- [4] DARA, *Climate Vulnerability Monitor*, 2nd edition, 2013, <https://s3.amazonaws.com/daraint/CVM2ndEd-ExecutiveSummary.pdf>.
- [5] Á. Halldórsson, H. Kotzab, and T. Skjøtt-Larsen, "Supply chain management on the crossroad to sustainability: a blessing or a curse?" *Logistics Research*, no. 2, pp. 83–94, 2009.
- [6] European Commission, *The EU Emissions Trading System (EU ETS)*, European Union, 2013, http://ec.europa.eu/clima/publications/docs/factsheet_ets_2013_en.pdf.
- [7] ARD, "Der Preis der Blue-Jeans," [TV report], 2012, http://mediathek.daserste.de/sendungen_a-z/799280_reportage-dokumentation/11398300_der-preis-der-blue-jeans.
- [8] F. Willershausen, "Die Modelüge—wie deutsche Firmen produzieren lassen," *Wirtschaftswoche*, 2012, <http://www.wiwo.de/unternehmen/industrie/textilindustrie-die-modeluege-wie-deutsche-firmen-produzieren-lassen/7162224.html>.
- [9] C. Duhigg and D. Barboza, "In China, human costs are built into an iPad," *The New York Times*, 2013, <http://www.nytimes.com/2012/01/26/business/ieconomy-apples-ipad-and-the-human-costs-for-workers-in-china.html?pagewanted=all>.
- [10] China Daily, "Real cause of tragedy," *China Daily*, 2013, http://europe.chinadaily.com.cn/opinion/2012-11/21/content_15-946621.htm.
- [11] R. Dekker, J. Bloemhof, and I. Mallidis, "Operations research for green logistics—an overview of aspects, issues, contributions and challenges," *European Journal of Operational Research*, vol. 219, no. 3, pp. 671–679, 2012.

- [12] R. Jamshidi, S. Ghomi, and B. Karimi, "Multi-objective green supply chain optimization with a new hybrid memetic algorithm using the Taguchi method," *Scientia Iranica*, vol. 19, no. 6, pp. 1876–1886, 2012.
- [13] A. Nagurney and L. S. Nagurney, "Sustainable supply chain network design: a multicriteria perspective," *International Journal of Sustainable Engineering*, vol. 3, no. 3, pp. 189–197, 2010.
- [14] T. Paksoy, E. Özceylan, and G. Weber, "A multi objective model for optimization of a green supply chain network," *Global Journal of Technology and Optimization*, no. 2, pp. 84–96, 2011.
- [15] S.-C. Tseng and S.-W. Hung, "A strategic decision-making model considering the social costs of carbon dioxide emissions for sustainable supply chain management," *Journal of Environmental Management*, vol. 133, no. 15, pp. 315–322, 2014.
- [16] L. J. Zeballos, C. A. Méndez, A. P. Barbosa-Poavoa, and A. Q. Novais, "Multi-period design and planning of closed loop supply chains with uncertain supply and demand," *Computers and Chemical Engineering*. In press.
- [17] J. M. Cruz, "Dynamics of supply chain networks with corporate social responsibility through integrated environmental decision-making," *European Journal of Operational Research*, vol. 184, no. 3, pp. 1005–1031, 2008.
- [18] Y. Bouchery, A. Ghaffari, Z. Jemai, and Y. Dallery, "Including sustainability criteria into inventory models," *European Journal of Operational Research*, vol. 222, no. 2, pp. 229–240, 2012.
- [19] C. R. Carter and D. S. Rogers, "A framework of sustainable supply chain management: moving toward new theory," *International Journal of Physical Distribution and Logistics Management*, vol. 38, no. 5, pp. 360–387, 2008.
- [20] A. Ait-Alla, M. C. Teucke, M. Lütjen, S. Beheshti-Kashi, and H. R. Karimi, "Robust production planning in fashion apparel industry under demand uncertainty via conditional value-at-risk," *Mathematical Problems in Engineering*, vol. 2014, Article ID 901861, 2014.
- [21] H. R. Karimi, N. A. Duffie, and S. Dashkovskiy, "Local capacity H_{∞} control for production networks of autonomous work systems with time-varying delays," *IEEE Transactions on Automation Science and Engineering*, vol. 7, no. 4, pp. 849–857, 2010.
- [22] A. Mehraei, H. R. Karimi, K. D. Thoben, and B. Scholz-Reiter, "Using metaheuristic and fuzzy system for the optimization of material pull in a push-pull flow logistics network," *Mathematical Problems in Engineering*, vol. 2013, Article ID 359074, 19 pages, 2013.
- [23] A. Mehraei, H.-R. Karimi, and B. Scholz-Reiter, "Toward learning autonomous pallets by using fuzzy rules, applied in a Conwip system," *International Journal of Advanced Manufacturing Technology*, vol. 64, no. 5–8, pp. 1131–1150, 2013.
- [24] W. Bretzke and K. Barkawi, *Nachhaltige Logistik—Antworten Auf Eine Globale Herausforderung*, Springer, Berlin, Germany, 2010.
- [25] World Commission on Environment and Development, *Our Common Future*, Oxford University Press, New York, NY, USA, 1987.
- [26] J. Elkington, *Cannibals with Forks: The Triple Bottom Line of 21st Century Business*, New Society Publishers, 1998.
- [27] T. Dyllick and K. Hockerts, "Beyond the business case for corporate sustainability," *Business Strategy and the Environment*, vol. 11, no. 2, pp. 130–141, 2002.
- [28] C. Tang and S. Zhou, "Research advances in environmentally and socially sustainable operations," *European Journal of Operational Research*, no. 3, pp. 585–594, 2012.
- [29] B. Cetinkaya, R. Cuthbertson, G. Ewer, T. Klaas-Wissing, W. Piotrowicz, and C. Tyssen, *Sustainable Supply Chain Management—Practical Ideas for Moving Towards Best Practice*, Springer, Berlin, Germany, 2011.
- [30] F. Székely and M. Knirsch, "Responsible leadership and corporate social responsibility: metrics for sustainable performance," *European Management Journal*, vol. 23, no. 6, pp. 628–647, 2005.
- [31] S. E. Daniel, D. C. Diakoulaki, and C. P. Pappis, "Operations research and environmental planning," *European Journal of Operational Research*, vol. 102, no. 2, pp. 248–263, 1997.
- [32] Bureau of Labor Statistics, *Incidence Rates of Nonfatal Occupational Injuries and Illnesses By Industry and Case Types*, Bureau of Labor Statistics, Washington, DC, USA, 2011, <http://www.bls.gov/iif/oshsum.htm#11Summary%20Tables>.
- [33] H. Mausser, "Normalization and other topics in multi-objective optimization," in *Proceedings of the Fields-MITACS Industrial Problems Workshop*, O. Grodzewich and O. Romanko, Eds., pp. 89–101, 2006.
- [34] T. L. Saaty, *The Analytic Hierarchy Process*, McGraw-Hill, New York, NY, USA, 1980.
- [35] T. L. Saaty, "A scaling method for priorities in hierarchical structures," *Journal of Mathematical Psychology*, vol. 15, no. 3, pp. 234–281, 1977.
- [36] indexmundi, "Iron Ore Monthly Price," 2013, <http://www.indexmundi.com/commodities/?commodity=iron-ore>.
- [37] A. Beresford, S. Pettit, and Y. Liu, "Multimodal supply chains: iron ore from Australia to China," *Supply Chain Management*, vol. 16, no. 1, pp. 32–42, 2011.
- [38] MetalMiner, "Steel Production Cost Model," 2009, <http://agmetminer.com/steel-production-cost-model>.
- [39] IG Metall, "Tariflöhne: Eisen- und Stahlindustrie," <http://www.igmetall.de/cps/rde/xchg/internet/style.xsl/eisen-und-stahl-loehne-und-gehaelter-913.htm>.
- [40] ThyssenKrupp, "Nachhaltigkeitsbericht 2009, Duisburg: ThyssenKrupp Steel AG," 2009, http://www.econsense.de/sites/all/files/TKS_NH-Bericht_2009_dt_0.pdf.
- [41] Bureau of Labor Statistics, "How to Compute a Firm's Incidence Rate for Safety Management," 2012, <http://www.bls.gov/iif/osheval.htm>.

Research Article

Global Conservative and Multipeakon Conservative Solutions for the Modified Camassa-Holm System with Coupling Effects

Huabin Wen,¹ Yujuan Wang,² Yongduan Song,² and Hamid Reza Karimi³

¹ School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, China

² School of Automation, Chongqing University, Chongqing 400044, China

³ Department of Engineering, Faculty of Engineering and Science, University of Agder, 4898 Grimstad, Norway

Correspondence should be addressed to Yongduan Song; ydsong@cqu.edu.cn

Received 31 January 2014; Accepted 18 March 2014; Published 24 April 2014

Academic Editor: Michael Lütjen

Copyright © 2014 Huabin Wen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper investigates the continuation of solutions to the modified coupled two-component Camassa-Holm system after wave breaking. The underlying problem is rather challenging due to the mutual coupling effect between two components in the system. By introducing a novel transformation that makes use of a skillfully defined characteristic and a set of newly defined variables, the original system is converted into a Lagrangian equivalent system, from which the global conservative solution is obtained, which further allows for the establishment of the multipeakon conservative solution of the system. The results obtained herein are deemed useful for understanding the inevitable phenomenon near wave breaking.

1. Introduction

We consider here the following modified coupled two-component Camassa-Holm system with peakons [1]:

$$\begin{aligned} m_t + 2mu_x + m_x u + (mv)_x + nv_x &= 0, & t > 0, & x \in R, \\ n_t + 2nv_x + n_x v + (nu)_x + mu_x &= 0, & t > 0, & x \in R, \\ m &= u - u_{xx}, & t > 0, & x \in R, \\ n &= v - v_{xx}, & t > 0, & x \in R. \end{aligned} \quad (1)$$

System (1) is a modified version of the new coupled two-component Camassa-Holm system in the following equation; namely,

$$\begin{aligned} m_t &= 2mu_x + m_x u + (mv)_x + nv_x, & t > 0, & x \in R, \\ n_t &= 2nv_x + n_x v + (nu)_x + mu_x, & t > 0, & x \in R, \\ m &= u - u_{xx}, & t > 0, & x \in R, \\ n &= v - v_{xx}, & t > 0, & x \in R, \end{aligned} \quad (2)$$

which, as an extension of the Camassa-Holm (CH) equation, has been established by Fu and Qu to allow for peakon

solitons in the form of a superposition of multipeakons. By parameterizing $\tilde{t} = -t$ for system (2), it then takes the form of (1), which can be rewritten as a Hamiltonian system

$$\frac{\partial}{\partial t} \begin{pmatrix} m \\ n \end{pmatrix} = - \begin{pmatrix} \partial m + m\partial & \partial m + n\partial \\ \partial n + m\partial & \partial n + n\partial \end{pmatrix} \begin{pmatrix} \frac{\delta H}{\delta m} = u \\ \frac{\delta H}{\delta n} = v \end{pmatrix} \quad (3)$$

with the Hamiltonian $H = (1/2) \int (mG * m + nG * n) dx$, where $G * m = u$, $G * n = v$, and $G = (1/2)e^{-|x|}$. Particularly, when $u = 0$ (or $v = 0$), the degenerated (1) has the same peakon solitons as the CH equation. We are interested in such system because it exhibits the following conserved quantities, as can be easily verified:

$$\begin{aligned} E_1(u) &= \int_R u dx, & E_2(v) &= \int_R v dx, \\ E_3(u) &= \int_R m dx, & E_4(u) &= \int_R n dx, \\ E_5(u, v) &= \int_R (u^2 + u_x^2 + v^2 + v_x^2) dx. \end{aligned} \quad (4)$$

Note that, when $u = v$, system (1) is reduced to the scalar Camassa-Holm equation as follows:

$$m_t + 4mu_x + 2m_x u = 0. \quad (5)$$

The CH equation, which models the unidirectional propagation of shallow water waves over a flat bottom, has a bi-Hamiltonian structure [3] and is completely integrable [4–6]. The CH equation has attracted considerable attention because it has peaked solitons [4, 7] and experiences wave breaking [4, 8]. The presence of breaking waves means that the solution remains bounded while its slope becomes unbounded in finite time [8, 9]. After wave breaking, the solutions of the CH equation can be continued uniquely as either global conservative [10–13] or global dissipative solutions [14].

As one of the integrable multicomponent generalizations of the CH equation, system (1) has been shown to be locally well posed with global strong solutions which blow up in finite time [1, 2]. Moreover, the existence issue for a class of local weak solutions for the modified coupled CH2 system was also addressed in [1]. It has been known that the continuation of solutions for the system beyond wave breaking has been a challenging problem. In our recent work [15], the problem of continuation beyond wave breaking for the modified coupled CH2 system was studied by applying an approach that reformulates the system (1) into a semilinear system of O.D.E. taking values in a Banach space. Such treatment makes it possible to investigate the continuity of the solutions beyond collision time, leading to the uniquely global solutions of this system. Also the global dissipative and multipeakon dissipative solutions of this system have been established in [16, 17], while, as far as the authors' concern, there is no effort made in the literature on the study of the global conservative as well as multipeakon conservative solutions of such system, another important feature associated with the system. Motivated by our recent work [15–17], in this paper we develop a new approach to establish a global and stable solution for the modified coupled CH2 system, which is conservative and further allows for the construction of the multipeakon conservative solution. The approach utilized in this paper makes use of a novel system transformation, which is different from [15] and is based on a skillfully defined characteristic and a set of newly introduced variables, where the associated energy is introduced as an additional variable so as to obtain a well-posed initial-value problem, facilitating the study on the behavior of wave breaking. It should be stressed that both global stable solution and multipeakon solution are important aspects related to the solutions near wave breaking, while there is no effort made in the literature on the study of multipeakon property of system (1), which is another motivation of this work. Our inspiration of investigating the underlying issue mainly also stems from the early work [10, 11] in the study of the global conservative solution of the CH equation and [13] where the multipeakon solution is obtained for the CH equation. In this work a coupled system is dealt with where the mutual effect between two components makes the analysis more complicated than a single one as considered in [10, 11, 13]. By utilizing the novel transformation method, the inherent

difficulty is circumvented and then the global conservative solutions of (1) are obtained, which then allows for the establishment of the multipeakon conservative solution of system (1).

The remainder of this paper is organized as follows. Section 2 presents the basic equations. In Section 3, by introducing a set of Lagrangian variables, we transform the original system into an equivalent semilinear system and derive the global solutions of the equivalent system. We obtain a global continuous semigroup of weak conservative solutions for the original system in Section 4 and the multipeakon conservative solution in Section 5.

2. The Original System

We first introduce an operator $\Lambda = (1 - \partial_x^2)^{-1}$, which can be expressed by its associated Green's function $G = (1/2)e^{-|x|}$ such as $\Lambda f(x) = G * f(x) = (1/2) \int_{\mathbb{R}} e^{-|x-x'|} f(x') dx'$, for all $f \in L^2(\mathbb{R})$, where $*$ denotes the spatial convolution. Thus we can rewrite (1) as a form of a quasilinear evolution equation:

$$\begin{aligned} u_t + (u + v)u_x + G * (uv_x) + \partial_x G \\ * \left(u^2 + \frac{1}{2}u_x^2 + u_x v_x + \frac{1}{2}v^2 - \frac{1}{2}v_x^2 \right) &= 0, \quad t > 0, \quad x \in \mathbb{R}, \\ v_t + (u + v)v_x + G * (u_x v) + \partial_x G \\ * \left(v^2 + \frac{1}{2}v_x^2 + u_x v_x + \frac{1}{2}u^2 - \frac{1}{2}u_x^2 \right) &= 0, \quad t > 0, \quad x \in \mathbb{R}. \end{aligned} \quad (6)$$

Let us define P_1, P_2, P_3 , and P_4 as

$$\begin{aligned} P_1(t, x) &= G * (uv_x) = \frac{1}{2} \int_{\mathbb{R}} e^{-|x-x'|} (uv_x)(t, x') dx', \\ P_2(t, x) &= G * \left(u^2 + \frac{1}{2}u_x^2 + u_x v_x + \frac{1}{2}v^2 - \frac{1}{2}v_x^2 \right) \\ &= \frac{1}{2} \int_{\mathbb{R}} e^{-|x-x'|} \left(u^2 + \frac{1}{2}u_x^2 + u_x v_x + \frac{1}{2}v^2 - \frac{1}{2}v_x^2 \right) \\ &\quad \times (t, x') dx', \\ P_3(t, x) &= G * (vu_x) = \frac{1}{2} \int_{\mathbb{R}} e^{-|x-x'|} (vu_x)(t, x') dx', \\ P_4(t, x) &= G * \left(v^2 + \frac{1}{2}v_x^2 + u_x v_x + \frac{1}{2}u^2 - \frac{1}{2}u_x^2 \right) \\ &= \frac{1}{2} \int_{\mathbb{R}} e^{-|x-x'|} \left(v^2 + \frac{1}{2}v_x^2 + u_x v_x + \frac{1}{2}u^2 - \frac{1}{2}u_x^2 \right) \\ &\quad \times (t, x') dx'. \end{aligned} \quad (7)$$

Then (1) can be rewritten as

$$\begin{aligned} u_t + (u + v)u_x + P_1 + P_{2,x} &= 0, \quad t > 0, \quad x \in \mathbb{R}, \\ v_t + (u + v)v_x + P_3 + P_{4,x} &= 0, \quad t > 0, \quad x \in \mathbb{R}. \end{aligned} \quad (8)$$

For regular solutions, we get that the total energy

$$E(t) = \int_R u^2(t, x) + u_x^2(t, x) + v^2(t, x) + v_x^2(t, x) dx \quad (9)$$

is constant in time. Thus (8) possesses the H^1 -norm conservation law defined as

$$\|z\|_{H^1} = \|u\|_{H^1} + \|v\|_{H^1} = \left(\int_R [u^2 + u_x^2 + v^2 + v_x^2] dx \right)^{1/2}, \quad (10)$$

where $z(t, x) = (u, v)(t, x)$ denotes the solution of system (8). Note that $z = (u, v) \in H^1 \times H^1$, and so Young's inequality ensures that $P_1, P_2, P_3, P_4 \in H^1$.

3. Global Solutions of the Lagrangian Equivalent System

We reformulate system (8) as follows. For a given initial data $y(0, \xi)$, we define the corresponding characteristic $y(t, \xi)$ as the solution of

$$y_t(t, \xi) = (u + v)(t, y(t, \xi)), \quad (11)$$

and we define the Lagrangian cumulative energy distribution H as

$$H(t, \xi) = \int_{-\infty}^{y(t, \xi)} (u^2 + u_x^2 + v^2 + v_x^2)(t, x) dx. \quad (12)$$

It is not hard to check that

$$\begin{aligned} & (u^2 + u_x^2 + v^2 + v_x^2)_t + ((u + v)(u^2 + u_x^2 + v^2 + v_x^2))_x \\ &= (u^3 - 2uP_2 + v^3 - 2vP_4)_x. \end{aligned} \quad (13)$$

Then it follows from (11) and (13) that

$$\frac{dH}{dt} = [(u^3 - 2uP_2 + v^3 - 2vP_4)(t, y(t, \xi))]_{-\infty}^{\xi}. \quad (14)$$

Throughout the following, we use the notation

$$\begin{aligned} U(t, \xi) &= u(t, y(t, \xi)), & V(t, \xi) &= v(t, y(t, \xi)), \\ M(t, \xi) &= u_x(t, y(t, \xi)), & N(t, \xi) &= v_x(t, y(t, \xi)). \end{aligned} \quad (15)$$

In the following, we drop the variable t for simplification. Here, we take y as an increasing function for any fixed time t for granted (later on we will prove this). Then after the change

of variables $x = y(t, \xi)$ and $x' = y(t, \xi')$, we obtain the following expressions for P_i and $P_{i,x}$ ($i = 1, 2, 3, 4$); namely,

$$\begin{aligned} P_1(t, \xi) &= P_1(t, y(t, \xi)) \\ &= \frac{1}{2} \int_R e^{-|y(\xi) - y(\xi')|} [(UN) y_\xi] (\xi') d\xi', \end{aligned} \quad (16)$$

$$\begin{aligned} P_{1,x}(t, \xi) &= P_{1,x}(t, y(t, \xi)) \\ &= -\frac{1}{2} \int_R \operatorname{sgn}(\xi - \xi') e^{-|y(\xi) - y(\xi')|} [(UN) y_\xi] \\ &\quad \times (\xi') d\xi', \end{aligned} \quad (17)$$

$$\begin{aligned} P_2(t, \xi) &= P_2(t, y(t, \xi)) \\ &= \frac{1}{2} \int_R e^{-|y(\xi) - y(\xi')|} \\ &\quad \times \left[\left(U^2 + \frac{1}{2} M^2 + MN + \frac{1}{2} V^2 - \frac{1}{2} N^2 \right) y_\xi \right] \\ &\quad \times (\xi') d\xi', \end{aligned}$$

$$\begin{aligned} P_{2,x}(t, \xi) &= P_{2,x}(t, y(t, \xi)) \\ &= -\frac{1}{2} \int_R \operatorname{sgn}(\xi - \xi') e^{-|y(\xi) - y(\xi')|} \\ &\quad \cdot \left[\left(U^2 + \frac{1}{2} M^2 + MN + \frac{1}{2} V^2 - \frac{1}{2} N^2 \right) y_\xi \right] \\ &\quad \times (\xi') d\xi', \end{aligned}$$

$$\begin{aligned} P_3(t, \xi) &= P_3(t, y(t, \xi)) \\ &= \frac{1}{2} \int_R e^{-|y(\xi) - y(\xi')|} [(VM) y_\xi] (\xi') d\xi', \end{aligned}$$

$$\begin{aligned} P_{3,x}(t, \xi) &= P_{3,x}(t, y(t, \xi)) \\ &= -\frac{1}{2} \int_R \operatorname{sgn}(\xi - \xi') e^{-|y(\xi) - y(\xi')|} [(VM) y_\xi] \\ &\quad \times (\xi') d\xi', \end{aligned}$$

$$\begin{aligned} P_4(t, \xi) &= P_4(t, y(t, \xi)) \\ &= \frac{1}{2} \int_R e^{-|y(\xi) - y(\xi')|} \\ &\quad \times \left[\left(V^2 + \frac{1}{2} N^2 + MN + \frac{1}{2} U^2 - \frac{1}{2} M^2 \right) y_\xi \right] \\ &\quad \times (\xi') d\xi', \end{aligned}$$

$$\begin{aligned} P_{4,x}(t, \xi) &= P_{4,x}(t, y(t, \xi)) \\ &= -\frac{1}{2} \int_R \operatorname{sgn}(\xi - \xi') e^{-|y(\xi) - y(\xi')|} \\ &\quad \cdot \left[\left(V^2 + \frac{1}{2} N^2 + MN + \frac{1}{2} U^2 - \frac{1}{2} M^2 \right) y_\xi \right] \\ &\quad \times (\xi') d\xi'. \end{aligned} \quad (18)$$

Since $H_\xi = (u^2 + u_x^2 + v^2 + v_x^2) \circ yy_\xi$, then $P_2, P_{2,x}, P_4$, and $P_{4,x}$ can be rewritten as

$$\begin{aligned}
P_2(t, \xi) &= P_2(t, y(\xi)) \\
&= \frac{1}{4} \int_R e^{-|y(\xi) - y(\xi')|} [H_\xi + (U^2 + 2MN - N^2) y_\xi] \\
&\quad \times (\xi') d\xi', \\
P_{2,x}(t, \xi) &= P_{2,x}(t, y(\xi)) \\
&= -\frac{1}{4} \int_R \operatorname{sgn}(\xi - \xi') e^{-|y(\xi) - y(\xi')|} \\
&\quad \times [H_\xi + (U^2 + 2MN - N^2) y_\xi] \\
&\quad \times (\xi') d\xi', \\
P_4(t, \xi) &= P_4(t, y(\xi)) \\
&= \frac{1}{4} \int_R e^{-|y(\xi) - y(\xi')|} [H_\xi + (V^2 + 2MN - M^2) y_\xi] \\
&\quad \times (\xi') d\xi', \\
P_{4,x}(t, \xi) &= P_{4,x}(t, y(\xi)) \\
&= -\frac{1}{4} \int_R \operatorname{sgn}(\xi - \xi') e^{-|y(\xi) - y(\xi')|} \\
&\quad \times [H_\xi + (V^2 + 2MN - M^2) y_\xi] \times (\xi') d\xi'. \tag{19}
\end{aligned}$$

From the definition of the characteristic, it is not hard to check that

$$\begin{aligned}
U_t(t, \xi) &= u_t(t, y) + u_x(t, y) y_t(t, \xi) \\
&= (-P_1 - P_{2,x}) \circ y(t, \xi), \\
V_t(t, \xi) &= v_t(t, y) + v_x(t, y) y_t(t, \xi) \\
&= (-P_3 - P_{4,x}) \circ y(t, \xi), \\
M_t(t, \xi) &= u_{xt}(t, y) + u_{xx}(t, y) y_t(t, \xi) \\
&= \left(-\frac{1}{2}M^2 - \frac{1}{2}N^2 + U^2 + \frac{1}{2}V^2 - P_{1,x} - P_2 \right) \\
&\quad \circ y(t, \xi), \\
N_t(t, \xi) &= v_{xt}(t, y) + v_{xx}(t, y) y_t(t, \xi) \\
&= \left(-\frac{1}{2}N^2 - \frac{1}{2}M^2 + V^2 + \frac{1}{2}U^2 - P_{3,x} - P_4 \right) \\
&\quad \circ y(t, \xi). \tag{20}
\end{aligned}$$

We introduce another variable $\zeta(t, \xi)$ with $\zeta(t, \xi) = y(t, \xi) - \xi$. It will turn out that $\zeta \in L^\infty(R)$. With these new variables, we now derive an equivalent system of (8) as follows:

$$\begin{aligned}
\zeta_t &= U + V, \\
U_t &= -P_1 - P_{2,x}, \\
V_t &= -P_3 - P_{4,x}, \\
M_t &= \left(-\frac{1}{2}M^2 - \frac{1}{2}N^2 + U^2 + \frac{1}{2}V^2 - P_{1,x} - P_2 \right), \\
N_t &= \left(-\frac{1}{2}N^2 - \frac{1}{2}M^2 + V^2 + \frac{1}{2}U^2 - P_{3,x} - P_4 \right), \\
H_t &= U^3 - 2UP_2 + V^3 - 2VP_4, \tag{21}
\end{aligned}$$

where P_1 and P_3 are given in (18), while $P_2, P_{2,x}, P_4$, and $P_{4,x}$ are given in (19). We regard system (21) as a system of ordinary differential equations in the Banach space

$$E = W \times H^1 \times H^1 \times L^2 \times L^2 \times W \tag{22}$$

endowed with the norm

$$\|X\|_E = \|\zeta\|_W + \|U\|_{H^1} + \|V\|_{H^1} + \|M\|_{L^2} + \|N\|_{L^2} + \|H\|_W, \tag{23}$$

for any $X = (\zeta, U, V, M, N, H) \in E$. Here $W = \{f \in C(R) \cap L^\infty(R) \mid f_\xi \in L^2(R)\}$ is a Banach space with the norm given by $\|f\|_W = \|f\|_{L^\infty(R)} + \|f_\xi\|_{L^2(R)}$. Note that $H^1(R) \subset W$.

Differentiating (21) with respect to the variable ξ yields

$$\begin{aligned}
\zeta_{\xi t} &= U_\xi + V_\xi, \\
U_{\xi t} &= \frac{1}{2}H_\xi + \left(\frac{1}{2}U^2 + MN - N^2 - P_2 - P_{1,x} \right) y_\xi, \\
V_{\xi t} &= \frac{1}{2}H_\xi + \left(\frac{1}{2}V^2 + MN - M^2 - P_4 - P_{3,x} \right) y_\xi, \\
H_{\xi t} &= (3U^2 - 2P_2)U_\xi - 2UP_{2,x}y_\xi \\
&\quad + (3V^2 - 2P_4)V_\xi - 2VP_{4,x}y_\xi, \tag{24}
\end{aligned}$$

which are semilinear with respect to the variables y_ξ, U_ξ, V_ξ , and H_ξ .

To obtain the uniqueness of solutions, one proceeds as follows. By proving that all functions on the right-hand side of (21) are locally Lipschitz continuous, the local existence of solutions will follow from the standard theory of ordinary differential equations in Banach spaces. In a second step, we will then prove that this local solution can be extended globally in time. Note that global solutions of (21) may not exist for all initial data in E . However they exist when the initial data $\bar{X} = (\bar{\zeta}, \bar{U}, \bar{V}, \bar{M}, \bar{N}, \bar{H})$ belongs to the set Γ which is defined as follows.

Definition 1. The set Γ is composed of all $X = (\zeta, U, V, M, N, H) \in E$ such that

$$(i) \quad (\varsigma, U, V, H) \in [W^{1,\infty}(R)]^4, \quad (25)$$

$$(ii) \quad y_\xi \geq 0, H_\xi \geq 0, \quad y_\xi + H_\xi > 0 \quad \text{a.e.},$$

$$\lim_{\xi \rightarrow -\infty} H(\xi) = 0, \quad (26)$$

$$(iii) \quad y_\xi H_\xi = y_\xi^2 U^2 + U_\xi^2 + y_\xi^2 V^2 + V_\xi^2 \quad \text{a.e.}, \quad (27)$$

where $W^{1,\infty}(R) = \{f \in C(R) \cap L^\infty(R) \mid f_\xi \in L^\infty(R)\}$ and $y(\xi) = \varsigma(\xi) + \xi$.

Lemma 2. *Let $\mathfrak{R}_1 : E \rightarrow W$ and let $\mathfrak{R}_2 : E \rightarrow H^1$, or let $\mathfrak{R}_2 : E \rightarrow W$ be two locally Lipschitz maps. Then, the product $X \rightarrow \mathfrak{R}_1(X)\mathfrak{R}_2(X)$ is also a locally Lipschitz map from E to H^1 or from E to W .*

Theorem 3. *Given initial data $\bar{X} = (\bar{\varsigma}, \bar{U}, \bar{V}, \bar{M}, \bar{N}, \bar{H}) \in E$, there exists a time T depending only on $\|\bar{X}\|_E$ such that the system (21) admits a unique solution in $C^1([0, T], E)$.*

Proof. To obtain the local existence of solutions, it suffices to show that $F(X)$, given by

$$F(X) = \left(U + V, -P_1 - P_{2,x} - P_3 - P_{4,x}, -\frac{1}{2}M^2 - \frac{1}{2}N^2 + U^2 \right. \\ \left. + \frac{1}{2}V^2 - P_{1,x} - P_2, -\frac{1}{2}N^2 - \frac{1}{2}M^2 + V^2 + \frac{1}{2}U^2 \right. \\ \left. - P_{3,x} - P_4, U^3 - 2UP_2 + V^3 - 2VP_4 \right) \quad (28)$$

with $X = (\varsigma, U, V, M, N, H)$, is a Lipschitz function on any bounded set of E which is a Banach space.

Our main task is to prove the Lipschitz continuity of P_i and $P_{i,x}$ ($i = 1, 2, 3, 4$) given by (18) and (19) from E to $H^1(R)$. We first prove that $P_{2,x}$ given in (19) is locally Lipschitz continuous from E to $H^1(R)$ and the others follow in the same way. Let us write

$$P_{2,x}(\xi) = -\frac{e^{-\varsigma(\xi)}}{4} \int_R \chi_{\{\xi' < \xi\}} e^{-|\xi - \xi'|} e^{\varsigma(\xi')} \\ \times [H_\xi + (U^2 + 2MN - N^2)(1 + \varsigma_\xi)](\xi') d\xi' \\ + \frac{e^{\varsigma(\xi)}}{4} \int_R \chi_{\{\xi' > \xi\}} e^{-|\xi - \xi'|} e^{-\varsigma(\xi')} \\ \times [H_\xi + (U^2 + 2MN - N^2)(1 + \varsigma_\xi)](\xi') d\xi', \quad (29)$$

where χ_Ω denotes the indicator function of a given set Ω . Let

$$P_{2,x}^1(\xi) = -\frac{e^{-\varsigma(\xi)}}{4} \int_R \chi_{\{\xi' < \xi\}} e^{-|\xi - \xi'|} e^{\varsigma(\xi')} \\ \times [H_\xi + (U^2 + 2MN - N^2)(1 + \varsigma_\xi)](\xi') d\xi', \\ P_{2,x}^2(\xi) = \frac{e^{\varsigma(\xi)}}{4} \int_R \chi_{\{\xi' > \xi\}} e^{-|\xi - \xi'|} e^{-\varsigma(\xi')} \\ \times [H_\xi + (U^2 + 2MN - N^2)(1 + \varsigma_\xi)](\xi') d\xi'. \quad (30)$$

We rewrite $P_{2,x}^1(\xi)$ as

$$P_{2,x}^1(\xi) = -\frac{e^{-\varsigma(\xi)}}{2} \Lambda \circ R(X)(\xi), \quad (31)$$

where R is the operator from E to $L^2(R)$ given as

$$R(X)(\xi) = \chi_{\{\xi' < \xi\}} e^\varsigma [H_\xi + (U^2 + 2MN - N^2)(1 + \varsigma_\xi)]. \quad (32)$$

Since the operator Λ (given in Section 2) is linear and continuous from $H^{-1}(R)$ to $H^1(R)$ and $L^2(R)$ is continuously embedded in $H^{-1}(R)$, we have $\Lambda \circ R(X) \in H^1$. It is not hard to check that R is locally Lipschitz continuous from E to $L^2(R)$ and therefore from E to $H^{-1}(R)$. Thus $\Lambda \circ R$ is locally Lipschitz continuous from E to $H^1(R)$. Since the mapping $X \rightarrow e^{-\varsigma}$ is locally Lipschitz continuous from E to W , by Lemma 2, we deduce that $P_{2,x}^1(\xi)$ is locally Lipschitz continuous from E to $H^1(R)$. Similarly, $P_{2,x}^2(\xi)$ is also locally Lipschitz continuous and therefore $P_{2,x}(\xi)$ is locally Lipschitz continuous. One proceeds in the same way and proves that $P_1, P_{1,x}, P_3$, and $P_{3,x}$ defined by (18) and P_2, P_4 , and $P_{4,x}$ defined by (19) are locally Lipschitz continuous from E to $H^1(R)$. We rewrite the solutions of (21) as

$$X(t) = \bar{X} + \int_0^t F(X(\tau)) d\tau. \quad (33)$$

Thus the theorem follows from the standard contraction argument of ordinary differential equations. \square

It remains to prove the existence of global solutions of (21). Theorem 3 gives us the existence of local solutions of (21) for initial data in E . In the following, we will only consider initial data that belongs to \tilde{E} given by $\tilde{E} = E \cap [(W^{1,\infty}(R))^3 \cap (L^2)^2 \cap W^{1,\infty}(R)]$. To obtain that the solution of (21) belongs to \tilde{E} , we have to specify the initial condition for (24). Let

$$\Omega = \left\{ \xi \in R \mid |\bar{\varsigma}_\xi(\xi)| \leq \|\bar{\varsigma}_\xi\|_{L^\infty}, |\bar{U}_\xi(\xi)| \leq \|\bar{U}_\xi\|_{L^\infty}, \right. \\ \left. |\bar{V}_\xi(\xi)| \leq \|\bar{V}_\xi\|_{L^\infty}, |\bar{H}_\xi(\xi)| \leq \|\bar{H}_\xi\|_{L^\infty} \right\}. \quad (34)$$

We have $\text{meas}(\Omega^c) = 0$. For $\xi \in \Omega^c$, $(c_\xi, U_\xi, V_\xi, H_\xi)(0, \xi)$ is taken as $(0, 0, 0, 0)$, and $(c_\xi, U_\xi, V_\xi, H_\xi)(0, \xi)$ is given as $(\bar{c}_\xi, \bar{U}_\xi, \bar{V}_\xi, \bar{H}_\xi)(\xi)$, for $\xi \in \Omega$.

The global existence of the solution for initial data in Γ relies essentially on the fact that the set Γ is preserved by the flow as the next lemma shows.

Lemma 4. *Given initial data $\bar{X} = (\bar{c}, \bar{U}, \bar{V}, \bar{M}, \bar{N}, \bar{H}) \in \Gamma$, one considers the local solution $X(t) = (c, U, V, M, N, H)(t) \in C([0, T], E)$ of (21) with initial data \bar{X} for some $T > 0$. One then gets that $X(t) \in \Gamma$ for all $t \in [0, T]$. Moreover, for a.e. $t \in [0, T]$, $y_\xi(t, \xi) > 0$, for a.e. $\xi \in R$, and $\lim_{\xi \rightarrow \pm\infty} H(t, \xi)$ exists and is independent of time for all $t \in [0, T]$.*

Proof. We first show that $X(t) \in \Gamma$ for all $t \in [0, T]$. For any given initial data $\bar{X} \in \tilde{E}$, we get that the local solution $(c, U, V, H)(t)$ of (21) belongs to $[W^{1,\infty}(R)]^4$, which satisfies (25) for all $t \in [0, T]$. We now show that (27) holds for any $\xi \in \Omega$ and therefore a.e.. Consider a fixed $\xi \in \Omega$ and drop it in the notation if there is no ambiguity. On the one hand, it follows from (24) that

$$\begin{aligned} & (y_\xi H_\xi)_t \\ &= y_{\xi t} H_\xi + y_\xi H_{\xi t} \\ &= (U_\xi + V_\xi) H_\xi + y_\xi \left[(3U^2 - 2P_2) U_\xi + (3V^2 - 2P_4) V_\xi \right. \\ & \quad \left. - 2UP_{2,x} y_\xi - 2VP_{4,x} y_\xi \right] \\ &= U_\xi H_\xi + V_\xi H_\xi + 3U^2 U_\xi y_\xi - 2P_2 U_\xi y_\xi + 3V^2 V_\xi y_\xi \\ & \quad - 2P_4 V_\xi y_\xi - 2UP_{2,x} y_\xi^2 - 2VP_{4,x} y_\xi^2, \end{aligned} \quad (35)$$

and, on the other hand,

$$\begin{aligned} & (y_\xi^2 U^2 + U_\xi^2 + y_\xi^2 V^2 + V_\xi^2)_t \\ &= 2y_\xi y_{\xi t} U^2 + 2y_\xi^2 U U_t + 2U_\xi U_{\xi t} + 2y_\xi y_{\xi t} V^2 \\ & \quad + 2y_\xi^2 V V_t + 2V_\xi V_{\xi t} \\ &= U_\xi H_\xi + V_\xi H_\xi + 3U^2 U_\xi y_\xi + 3V^2 V_\xi y_\xi - 2P_2 U_\xi y_\xi \\ & \quad - 2P_4 V_\xi y_\xi - 2UP_{2,x} y_\xi^2 - 2VP_{4,x} y_\xi^2. \end{aligned} \quad (36)$$

Hence, $(y_\xi H_\xi)_t = (y_\xi^2 U^2 + U_\xi^2 + y_\xi^2 V^2 + V_\xi^2)_t$. Notice that $y_\xi H_\xi(0) = (y_\xi^2 U^2 + U_\xi^2 + y_\xi^2 V^2 + V_\xi^2)(0)$; then $y_\xi H_\xi(t) = (y_\xi^2 U^2 + U_\xi^2 + y_\xi^2 V^2 + V_\xi^2)(t)$ for all $t \in [0, T]$ and (27) has been proved. We now prove the inequalities in (26). Set $t^* = \sup\{t \in [0, T] \mid y_\xi(t') \geq 0 \text{ for all } t' \in [0, t]\}$. Assume that $t^* < T$. Since $y_\xi(t)$ is continuous with respect to t , we have $y_\xi(t^*) = 0$. It follows from (27) that $U_\xi(t^*) = V_\xi(t^*) = 0$. Furthermore, (24) implies that $y_{\xi t}(t^*) = U_\xi(t^*) + V_\xi(t^*) = 0$ and $y_{\xi t t}(t^*) = (U_{\xi t} + V_{\xi t})(t^*) = H_\xi(t^*)$. If $H_\xi(t^*) = 0$, then $(y_\xi, U_\xi, V_\xi, H_\xi)(t^*) = (0, 0, 0, 0)$ which implies that $(y_\xi, U_\xi, V_\xi, H_\xi)(t) = 0$ for all $t \in [0, T]$ by the uniqueness

of the solution of system (24). This contradicts the fact that $y_\xi(0) + H_\xi(0) > 0$ for all $\xi \in \Omega$. If $H_\xi(t^*) < 0$, then $y_{\xi t t}(t^*) < 0$. Since $y_\xi(t^*) = y_{\xi t}(t^*) = 0$, there exists a neighborhood ω of t^* such that $y_\xi(t) < 0$ for all $t \in \omega \setminus \{t^*\}$. This contradicts the definition of t^* . Hence, $H_\xi(t^*) > 0$. We now have $y_{\xi t t}(t^*) > 0$, which conversely implies that $y_\xi(t) > 0$ for all $t \in \omega \setminus \{t^*\}$, which contradicts the fact that $t^* < T$. Thus we have proved $y_\xi(t) \geq 0$ for all $t \in [0, T]$. We now prove that $H_\xi \geq 0$ for all $t \in [0, T]$. This follows from (27) when $y_\xi(t) > 0$. If $y_\xi(t) = 0$, then $U_\xi(t) = V_\xi(t) = 0$ from (27). As we have seen, $H_\xi < 0$ would imply that $y_\xi(t') < 0$ for some t' in a punctured neighborhood of t , which is impossible. Hence, $H_\xi \geq 0$ for all $t \in [0, T]$. Now we get that $y_\xi(t) + H_\xi(t) \geq 0$ for all $t \in [0, T]$. If $y_\xi(t') + H_\xi(t') = 0$ for some t' , it then follows that $(y_\xi, U_\xi, V_\xi, H_\xi)(t') = 0$ which implies that $(y_\xi, U_\xi, V_\xi, H_\xi)(t) = 0$ for all $t \in [0, T]$, which contradicts the fact that $y_\xi(0) + H_\xi(0) > 0$ for all $\xi \in \Omega$. Hence, $y_\xi(t) + H_\xi(t) > 0$. This completes the proof that $X(t) \in \Gamma$ for all $t \in [0, T]$.

We now prove that $y_\xi(t, \xi) > 0$ for almost all t . Define the set $\Theta = \{(t, \xi) \in [0, T] \times R \mid y_\xi(t, \xi) = 0\}$. It follows from Fubini's theorem that

$$\text{meas}(\Theta) = \int_R \text{meas}(\Theta_\xi) d\xi = \int_{[0, T]} \text{meas}(\Theta_t) dt, \quad (37)$$

where $\Theta_\xi = \{t \in [0, T] \mid y_\xi(t, \xi) = 0\}$ and $\Theta_t = \{\xi \in R \mid y_\xi(t, \xi) = 0\}$. From the above proof, we know that, for all $\xi \in \Omega$, Θ_ξ consists of isolated points that are countable. This means that $\text{meas}(\Theta_\xi) = 0$. Since $\text{meas}(\Omega^c) = 0$, it thus follows from (37) that $\text{meas}(\Theta_t) = 0$ for almost every $t \in [0, T]$. This implies that $y_\xi(t, \xi) > 0$ for almost all t and therefore $y(t, \xi)$ is strictly increasing and invertible with respect to ξ .

For any given $t \in [0, T]$, since $H_\xi(t) \geq 0$ and $H(t, \xi) \in L^\infty(R)$, we know that $H(t, \pm\infty)$ exist. We have the following:

$$H(t, \xi) = H(0, \xi) + \int_0^t (U^3 - 2P_2 U + V^3 - 2P_4 V)(\tau, \xi) d\tau. \quad (38)$$

Let $\xi \rightarrow \pm\infty$. Since U, V, P_2, P_4 are bounded in $L^\infty([0, T] \times R)$ and $\lim_{\xi \rightarrow \pm\infty} U(t, \xi) = \lim_{\xi \rightarrow \pm\infty} V(t, \xi) = 0$ as $U(t, \cdot), V(t, \cdot) \in H^1(R)$ for all $t \in [0, T]$, (38) implies that $H(t, \pm\infty) = H(0, \pm\infty)$ for all $t \in [0, T]$. Since $\bar{X} \in \Gamma$, it follows that $H(0, \pm\infty) = 0$ for all $t \in [0, T]$. \square

Theorem 5. *For any initial data $\bar{X} = (\bar{y}, \bar{U}, \bar{V}, \bar{M}, \bar{N}, \bar{H}) \in \Gamma$, there exists a unique global solution $X(t) = (y, U, V, M, N, H)(t) \in C^1(R^+, E)$ for the system (21). Moreover, for all $t \in R^+$, we have $X(t) \in \Gamma$, which constructs a continuous semigroup.*

Proof. To ensure that the local solution $X = (c, U, V, M, N, H) \in C([0, T], E)$ of system (21) can be extended to a global solution, it suffices to show that

$$\sup_{t \in [0, T]} \|c(t, \cdot), U(t, \cdot), V(t, \cdot), M(t, \cdot), N(t, \cdot), H(t, \cdot)\|_E < \infty. \quad (39)$$

Since $H(t, \xi)$ is an increasing function with respect to ξ for all t and $\lim_{\xi \rightarrow \infty} H(t, \xi) = \lim_{\xi \rightarrow \infty} H(0, \xi)$, we have $\sup_{t \in [0, T]} \|H(t, \cdot)\|_{L^\infty(R)} = \|\bar{H}\|_{L^\infty(R)} < \infty$. We now consider a fixed $t \in [0, T]$ and drop it for simplification. Since $U_\xi(\xi) = V_\xi(\xi) = 0$ when $y_\xi(\xi) = 0$ and $y_\xi(\xi) > 0$, for a. e. ξ , it follows from (27) that

$$\begin{aligned} U^2(\xi) &= 2 \int_{-\infty}^{\xi} U(\xi') U_\xi(\xi') d\xi' \\ &= 2 \int_{\{\xi' < \xi | y_\xi(\xi') > 0\}} U(\xi') U_\xi(\xi') d\xi' \\ &\leq \int_{\{\xi' < \xi | y_\xi(\xi') > 0\}} \left(y_\xi U^2 + \frac{U_\xi^2}{y_\xi} \right) (\xi') d\xi' \\ &\leq \int_R H_\xi(\xi') d\xi' = H(\xi), \end{aligned} \quad (40)$$

which implies that

$$\sup_{t \in [0, T]} \|U^2(t, \cdot)\|_{L^\infty} \leq \sup_{t \in [0, T]} \|H(t, \cdot)\|_{L^\infty(R)} = \|\bar{H}\|_{L^\infty(R)} < \infty, \quad (41)$$

and therefore

$$\sup_{t \in [0, T]} \|U(t, \cdot)\|_{L^\infty} < \infty. \quad (42)$$

Similarly,

$$\sup_{t \in [0, T]} \|V(t, \cdot)\|_{L^\infty} < \infty. \quad (43)$$

We can obtain from the governing equation (21) that

$$|\zeta(t, \xi)| \leq |\zeta(0, \xi)| + \sup_{t \in [0, T]} (\|U(t, \cdot)\|_{L^\infty} + \|V(t, \cdot)\|_{L^\infty}) T, \quad (44)$$

and then $\sup_{t \in [0, T]} \|\zeta(t, \cdot)\|_{L^\infty} < \infty$. We can also get from the governing equation (21) that

$$\sup_{t \in [0, T]} \|M(t, \cdot)\|_{L^\infty} < \infty, \quad \sup_{t \in [0, T]} \|N(t, \cdot)\|_{L^\infty} < \infty. \quad (45)$$

From the identity $H_\xi = (U^2 + M^2 + V^2 + N^2)y_\xi$, we can deduce that

$$\left| (U^2 + 2MN - N^2) y_\xi \right| \leq (U^2 + M^2 + N^2 + N^2) y_\xi \leq 2H_\xi, \quad (46)$$

which implies that

$$\begin{aligned} |P_{2,x}| &\leq \frac{1}{2} \left| \int_R e^{-1y(\xi) - y(\xi')} [H_\xi + (U^2 + 2MN - N^2) y_\xi] (\xi') d\xi' \right| \\ &\leq \frac{1}{2} \left| \int_R e^{-1y(\xi) - y(\xi')} 3H_\xi(\xi') d\xi' \right| \\ &\leq C \sup_{t \in [0, T]} \|H(t, \cdot)\|_{L^\infty(R)} < \infty. \end{aligned} \quad (47)$$

Therefore, $\|P_{2,x}\|_{L^\infty} < \infty$. It is not hard to know that $\|P_{2,x}\|_{L^2} \leq C \|e^{-y(\xi)}\|_{L^2} \cdot \sup_{t \in [0, T]} \|H(t, \cdot)\|_{L^\infty(R)} < \infty$. Similarly, one can obtain that the bounds hold for $P_1, P_{1,x}, P_2, P_3, P_{3,x}, P_4$, and $P_{4,x}$. Let

$$\begin{aligned} Z(t) &= \|U(t, \cdot)\|_{L^2} + \|U_\xi(t, \cdot)\|_{L^2} + \|V(t, \cdot)\|_{L^2} \\ &\quad + \|V_\xi(t, \cdot)\|_{L^2} + \|\zeta_\xi(t, \cdot)\|_{L^2} + \|H_\xi(t, \cdot)\|_{L^2}. \end{aligned} \quad (48)$$

Using the integrated version of (21) and (24), after taking the L^2 -norms on both sides, we obtain

$$Z(t) \leq Z(0) + C \int_0^t Z(\tau) d\tau. \quad (49)$$

It follows from Gronwall's inequality that $\sup_{t \in [0, T]} Z(t) < \infty$. Hence, we infer that the map $S_t : \Gamma \rightarrow \Gamma \times R^+$ defined as

$$S_t(\bar{X}) = X(t) \quad (50)$$

generates a continuous semigroup from the standard theory of ordinary differential equations. \square

4. Global Conservative Solutions of the Original System

We show that the global solution of the equivalent system (21) yields a global conservative solution of the original system (8), which constructs a continuous semigroup in this section.

To obtain the global conservative solution of the original system, we have to establish the correspondence between the Lagrangian equivalent system and the original system.

Let us first introduce the subsets F and F_α of Γ given by

$$F = \{X = (y, U, V, M, N, H) \in \Gamma \mid y + H \in G\}, \quad (51)$$

$$F_\alpha = \{X = (y, U, V, M, N, H) \in \Gamma \mid y + H \in G_\alpha\},$$

where G is defined as

$$\begin{aligned} G &= \{f \text{ is invertible} \mid f - Id, \\ &\quad f^{-1} - Id \text{ both belong to } W^{1,\infty}(R)\}. \end{aligned} \quad (52)$$

And, for any $\alpha > 1$, the subsets G_α of G are given by

$$G_\alpha = \left\{ f \in G \mid \|f - Id\|_{W^{1,\infty}(R)} + \|f^{-1} - Id\|_{W^{1,\infty}(R)} \leq \alpha \right\}, \quad (53)$$

with a useful characterization. If $f \in G_\alpha$ ($\alpha \geq 0$), then $1/(1 + \alpha) \leq f_\xi \leq 1 + \alpha$ a.e. Conversely, if f is absolutely continuous, $f - Id \in L^\infty(R)$ and there exists $c \geq 1$ such that $1/c \leq f_\xi \leq c$ a.e., and then $f \in G_\alpha$ for some α depending only on c and $\|f - Id\|_{L^\infty(R)}$. With this useful characterization of G_α , it is not hard to prove that the space F is preserved by the governing equation (21). Notice that the map $(f, X) \rightarrow X \circ f$ defines a group action of G on F ; we consider the quotient space F/G of F with respect to the group action. The equivalence relation on F is defined as follows: for any $X, X' \in F$, if there exists

$f \in G$ such that $X' = X \circ f$, we claim that X and X' are equivalent.

We denote the projection $\Pi : F \rightarrow F/G$ by $\Pi(X) = [X]$. For any $X = (y, U, V, M, N, H) \in F$, we introduce the mapping $Y : F \rightarrow F_0$ given by $Y(X) = X \circ (y + H)^{-1}$. It is not hard to prove that $Y(X) = X$ when $X \in F_0$ and $Y(X \circ f) = Y(X)$ for any $X \in F$ and $f \in G$. Hence, we can define the map $\tilde{Y} : F/G \rightarrow F_0$ as $\tilde{Y}([X]) = Y(X)$, for any representative $[X] \in F/G$ of $X \in F$. For any $X \in F_0$, we have $\tilde{Y} \circ \Pi(X) = Y(X) = X$. Hence, $\tilde{Y} \circ \Pi|_{F_0} = Id_{F_0}$ and any topology defined on F_0 is naturally transported into F/G by this isomorphism. That is, if we equip F_0 with the metric induced by the E -norm; that is, $d_{F_0}(X, X') = \|X - X'\|_E$, for all $X, X' \in F_0$, which is complete, then the topology on F/G is defined by a complete metric given by $d_{F/G}([X], [X']) = \|Y(X) - Y(X')\|_E$ for any $[X], [X'] \in F/G$. Let us denote by $S : F \times R^+ \rightarrow F$ the continuous semigroup which to any initial data $\bar{X} \in F$ associates the solution $X(t)$ of (21). The system (8) is invariant with respect to relabeling. That is, for any $t > 0$, $S_t(X \circ f) = S_t(X) \circ f$, for an $X \in F$ and $f \in G$. Thus the map $\tilde{S}_t : F/G \rightarrow F/G$ given by $\tilde{S}_t([X]) = [S_t X]$ is well-defined, which generates a continuous semigroup.

To obtain a semigroup of solution for (8), we have to consider the space D , which characterizes the solutions in the original system:

$$D = \{(z, \mu) \mid z \in H^1(R) \times H^1(R), \mu_{ac} = (u^2 + u_x^2 + v^2 + v_x^2) dx\}, \quad (54)$$

where $z = (u, v)$ and μ is a positive finite Radon measure with μ_{ac} as its absolute continuous part.

We now establish a bijection between F/G and D to transport the continuous semigroup obtained in the Lagrangian equivalent system (functions in F/G) into the original system (functions in D).

We first introduce the mapping $L : D \rightarrow F/G$, which transforms the original system into the Lagrangian equivalent system defined as follows.

Definition 6. For any $(z, \mu) \in D$, let

$$y(\xi) = \sup \{y \mid \mu(-\infty, y) + y < \xi\}, \quad (55)$$

$$\begin{aligned} U(\xi) &= u \circ y(\xi), & V(\xi) &= v \circ y(\xi), \\ M(\xi) &= u_x \circ y(\xi), & N(\xi) &= v_x \circ y(\xi), \end{aligned} \quad (56)$$

$$H(\xi) = \xi - y(\xi), \quad (57)$$

with $z = (u, v)$. We define $L(z, \mu) \in F/G$ as the equivalence class of (y, U, V, M, N, H) .

Remark 7. From the definition of y, U, V, M, N, H in (55)–(57), we can check that $X = (y, U, V, M, N, H) \in E$, which also satisfies (25). Moreover, we have $y + H = Id$ from (57), which implies that $X = (y, U, V, M, N, H) \in F_0$.

Furthermore, if μ is absolutely continuous, then $\mu = (u^2 + u_x^2 + v^2 + v_x^2) dx$ and

$$\int_{-\infty}^{y(\xi)} (u^2 + u_x^2 + v^2 + v_x^2) dx + y(\xi) = \xi, \quad (58)$$

for all $\xi \in R$.

We are led to the mapping M , which corresponds to the transformation from the Lagrangian equivalent system into the original system. In the other direction, we obtain the energy density μ in the original system, by pushing forward by y the energy density $H_\xi d\xi$ in the Lagrangian equivalent system, where the push-forward $f_\# \nu$ of a measure ν by a measurable function f is defined as

$$f_\# \nu(B) = \nu(f^{-1}(B)), \quad (59)$$

for all the Borel set B . Give any element $[X] \in F/G$, and let (z, μ) be defined as

$$z(x) = Z(\xi) \quad \text{for any } \xi \text{ such that } x = y(\xi), \quad (60)$$

$$\mu = y_\#(H_\xi d\xi), \quad (61)$$

where $z(x) = (u, v)(x)$ and $Z(\xi) = (U, V)(\xi)$. We get that $(z, \mu) \in D$, which does not depend on the representative $X = (y, U, V, M, N, H) \in F$ of $[X]$ that we choose. We denote by $M : F/G \rightarrow D$ the map to any $[X] \in F/G$ and $(z, \mu) \in D$ given by (60)–(61), which conversely transforms the Lagrangian equivalent system into the original system.

We claim that the transformation from the original system into the Lagrangian equivalent system is a bijection.

Theorem 8. *The maps M and L are well-defined and $L^{-1} = M$. That is,*

$$L \circ M = Id_{F/G}, \quad M \circ L = Id_D. \quad (62)$$

Proof. Let $[X]$ in F/G be given. We consider $X = (y, U, V, M, N, H) = \tilde{Y}([X])$ as a representative of $[X]$ and (z, μ) given by (60)–(61) for this particular X . From the definition of \tilde{Y} , we have $X \in F_0$. Let $\bar{X} = (\bar{y}, \bar{U}, \bar{V}, \bar{M}, \bar{N}, \bar{H})$ be the representative of $L(z, \mu)$ in F_0 given by Definition 6. We have to prove that $\bar{X} = X$ and therefore $L \circ M = Id_{F/G}$. Let

$$g(x) = \sup \{\xi \in R \mid y(\xi) < x\}. \quad (63)$$

Using the fact that y is increasing and continuous, it follows that

$$y(g(x)) = x \quad (64)$$

and $y^{-1}((-\infty, x)) = (-\infty, g(x))$. From (61) and since $H(-\infty) = 0$, for any $x \in R$, we get the following:

$$\mu((-\infty, x)) = \int_{y^{-1}((-\infty, x))} H_\xi d\xi = \int_{-\infty}^{g(x)} H_\xi d\xi = H(g(x)). \quad (65)$$

Since $X \in F_0$ and $y + H = Id$, we have

$$\mu((-\infty, x)) + x = g(x). \quad (66)$$

From the definition of \bar{y} , it follows that

$$\bar{y}(\xi) = \sup \{x \in R \mid g(x) < \xi\}. \quad (67)$$

For any given $\xi \in R$, using the fact that y is increasing and (64), it follows that $\bar{y}(\xi) \leq y(\xi)$. If $\bar{y}(\xi) < y(\xi)$, there then exists x such that $\bar{y}(\xi) < x < y(\xi)$ and (67) implies that $g(x) \geq \xi$. Conversely, since y is increasing, then $x = y(g(x)) < y(\xi)$ implies that $g(x) < \xi$, which gives us a contradiction. Hence, we have $\bar{y} = y$. Since $y + H = Id$, it follows directly from the definitions that $\bar{H} = H, \bar{U} = U, \bar{V} = V, \bar{M} = M$, and $\bar{N} = N$. We thus proved that $L \circ M = Id_{F/G}$.

Given (z, μ) in D , we denote by (y, U, V, M, N, H) the representative of $L(z, \mu)$ in F_0 given by Definition 6. Let $(\bar{z}, \bar{\mu}) = M \circ L(z, \mu)$ and g be defined as before by (63). The same computation that leads to (66) now gives

$$\bar{\mu}((-\infty, x)) + x = g(x). \quad (68)$$

Given $\xi \in R$, we consider an increasing sequence x_i converging to $y(\xi)$ which is guaranteed by (55) such that $\mu((-\infty, x_i)) + x_i < \xi$. Let i tend to infinity. Since $F(x) = \mu((-\infty, x))$ is lower semi-continuous, we have $\mu((-\infty, y(\xi))) + y(\xi) \leq \xi$. Take $\xi = g(x)$ and then we get

$$\mu((-\infty, x)) + x \leq g(x). \quad (69)$$

By the definition of g , there exists an increasing sequence ξ_i converging to $g(x)$ such that $y(\xi_i) < x$. It follows from the definition of y in (55) that $\mu((-\infty, x)) + x \geq \xi_i$. Passing to the limit, we obtain $\mu((-\infty, x)) + x \geq g(x)$ which, together with (69), yields

$$\mu((-\infty, x)) + x = g(x). \quad (70)$$

We obtain that $\bar{\mu} = \mu$ by comparing (70) and (68). It is clear from the definitions that $\bar{z} = z$. Hence, $(\bar{z}, \bar{\mu}) = (z, \mu)$ and $M \circ L = Id_D$. \square

The topology defined in F/G can be transported into D , which is guaranteed by the fact that we have established a bijection between the two equivalent systems. We define the metric d_D on D as

$$d_D((z, \mu), (\bar{z}, \bar{\mu})) = d_{F/G}(L(z, \mu), L(\bar{z}, \bar{\mu})), \quad (71)$$

which makes the bijection L between D and F/G into an isometry. Since F/G equipped with $d_{F/G}$ is a complete metric space, D equipped with the metric d_D is also a complete metric space. For each $t \in R$, we define the mapping $T_t : D \rightarrow D$ as

$$T_t = M\tilde{S}_t L. \quad (72)$$

Then a continuous semigroup of conservative weak solutions for the original system is obtained as the following theorem shows.

Theorem 9. Let $(\bar{z}, \bar{\mu}) \in D$ be given. If one denotes by $t \rightarrow (z(t), \mu(t)) = T_t(\bar{z}, \bar{\mu})$ the corresponding trajectory, then $z = (u, v)$ is a weak solution of the modified coupled two-component Camassa-Holm equation (8), which constructs a continuous semigroup. Moreover, μ is a weak solution of the following transport equation for the energy density:

$$\mu_t + [(u + v)\mu]_x = (u^3 - 2P_2u + v^3 - 2P_4v)_x. \quad (73)$$

Furthermore, for all t , it holds that

$$\mu(t)(R) = \mu(0)(R) \quad (74)$$

and, for almost all t ,

$$\begin{aligned} \mu(t)(R) &= \mu_{ac}(t)(R) = \|z(t)\|_{H^1}^2 \\ &= \|u(t)\|_{H^1}^2 + \|v(t)\|_{H^1}^2 = \mu(0)(R). \end{aligned} \quad (75)$$

Thus the unique solution described here is a conservative weak solution of the system (8).

Proof. To prove that $z = (u, v)$ is a weak solution of the original system (8), it suffices to show that, for all $\phi \in C^\infty(R^+ \times R)$ with compact support,

$$\begin{aligned} &\int_{R^+ \times R} [-u\phi_t + (u + v)u_x\phi](t, x) dx dt \\ &= - \int_{R^+ \times R} [(P_1 + P_{2,x})\phi](t, x) dx dt, \\ &\int_{R^+ \times R} [-v\phi_t + (u + v)v_x\phi](t, x) dx dt \\ &= - \int_{R^+ \times R} [(P_3 + P_{4,x})\phi](t, x) dx dt, \end{aligned} \quad (76)$$

where $P_1, P_{2,x}, P_3$, and $P_{4,x}$ are given by (8). Let the solution $(y, U, V, M, N, H)(t)$ of (21) be a representative of $L(z(t), \mu(t))$. On the one hand, since $y(t, \xi)$ is Lipschitz continuous and invertible with respect to ξ , for almost all t , we then can use the change of variables $x = y(t, \xi)$ and obtain

$$\begin{aligned} &\int_{R^+ \times R} [-u\phi_t + (u + v)u_x\phi](t, x) dx dt \\ &= \int_{R^+ \times R} [- (Uy_\xi)(t, \xi) \phi_t(t, y(t, \xi)) \\ &\quad + ((U + V)U_\xi)(t, \xi) \phi(t, y(t, \xi))] d\xi dt. \end{aligned} \quad (77)$$

By using the identities $y_t = U + V$ and $y_{\xi t} = U_\xi + V_\xi$, it then follows from (21) that

$$\begin{aligned} &\int_{R^+ \times R} [-Uy_\xi\phi_t(t, y) + (U + V)U_\xi\phi(t, y)] d\xi dt \\ &= \frac{1}{2} \int_{R^+ \times R^2} \left\{ -UNy_\xi + \frac{1}{2} \operatorname{sgn}(\xi - \xi') \right. \\ &\quad \left. \times [H_\xi + (U^2 + 2MN - N^2)y_\xi] \right\} (\xi') \\ &\quad \cdot e^{-|y(\xi) - y(\xi')|} \phi(t, y(\xi)) y_\xi(\xi) d\xi' d\xi dt. \end{aligned} \quad (78)$$

On the other hand, using the change of variables $x = y(t, \xi)$ and $x' = y(t, \xi')$ and since y is increasing with respect to ξ , we have the following:

$$\begin{aligned}
 & - \int_{R^+ \times R} [(P_1 + P_{2,x}) \phi] (t, x) dx dt \\
 & = \frac{1}{2} \int_{R^+ \times R^2} \left[-uv_x + \operatorname{sgn}(\xi - \xi') \right. \\
 & \quad \times \left(u^2 + \frac{1}{2}u_x^2 + u_x v_x + \frac{1}{2}v^2 - \frac{1}{2}v_x^2 \right) \quad (79) \\
 & \quad \times (t, y(\xi')) \cdot e^{-|y(\xi) - y(\xi')|} \\
 & \quad \times \phi(t, y(\xi)) y_\xi(\xi') y_\xi(\xi) d\xi' d\xi dt.
 \end{aligned}$$

We obtain from (27) that

$$\begin{aligned}
 & - \int_{R^+ \times R} [(P_1 + P_{2,x}) \phi] (t, x) dx dt \\
 & = \frac{1}{2} \int_{R^+ \times R^2} \left\{ -UNy_\xi + \frac{1}{2} \operatorname{sgn}(\xi - \xi') \right. \\
 & \quad \times [H_\xi + (U^2 + 2MN - N^2) y_\xi] \quad (80) \\
 & \quad \left. \cdot e^{-|y(\xi) - y(\xi')|} \phi(t, y(\xi)) y_\xi(\xi) d\xi' d\xi dt.
 \end{aligned}$$

By comparing (78) and (80), we know that

$$\begin{aligned}
 & \int_{R^+ \times R} [-Uy_\xi \phi_t(t, y) + (U + V)U_\xi \phi(t, y)] d\xi dt \\
 & = - \int_{R^+ \times R} [(P_1 + P_{2,x}) \phi] (t, x) dx dt. \quad (81)
 \end{aligned}$$

Hence, the first identity in (76) holds. The second identity in (76) follows in the same way. One can easily check that $\mu(t)$ is solution of (73). From the definition μ in (61), we get

$$\mu(t)(R) = \int_R H_\xi d\xi = H(t, \infty), \quad (82)$$

which is constant in time from Lemma 2. Thus, we have proved (74).

Since $y_\xi(t, \xi) > 0$ a.e., for almost every $\xi \in R$, it then follows from (27) that

$$\begin{aligned}
 \mu(t)(B) & = \int_{y^{-1}(B)} H_\xi d\xi \\
 & = \int_{y^{-1}(B)} \left(U^2 + \frac{U_\xi^2}{y_\xi^2} + V^2 + \frac{V_\xi^2}{y_\xi^2} \right) y_\xi d\xi, \quad (83)
 \end{aligned}$$

for any Borel set B . Since y is one-to-one and $u_x \circ yy_\xi = U_\xi$, $v_x \circ yy_\xi = V_\xi$ a.e. and then (83) implies that

$$\mu(t)(B) = \int_B (u^2 + u_x^2 + v^2 + v_x^2)(t, x) dx. \quad (84)$$

Hence, (75) is proved (and the solution is conservative), which completes the proof. \square

5. Multipeakon Conservative Solutions of the Original System

In this section, we will derive a new system of ordinary differential equations for the multipeakon solutions which is well posed even when collisions occur, and the variables (y, U, V, M, N, H) will be used to characterize multipeakons in a way that avoids the problems related to blowing up.

Solutions of the modified coupled two-component Camassa-Holm system may experience wave breaking in the sense that the solution develops singularities in finite time, while keeping the H^1 -norm finite. Continuation of the solution beyond wave breaking imposes significant challenge as can be illustrated in the case of multipeakons, which are special solutions of the modified coupled two-component Camassa-Holm system of the following form:

$$\begin{aligned}
 u(t, x) & = \sum_{i=1}^n p_i(t) e^{-|x - q_i(t)|}, \\
 v(t, x) & = \sum_{i=1}^n r_i(t) e^{-|x - q_i(t)|}, \quad (85)
 \end{aligned}$$

where $(p_i(t), r_i(t), q_i(t))$ satisfy the explicit system of ordinary differential equations:

$$\begin{aligned}
 \dot{p}_i & = \sum_{j=1, j \neq i}^n (p_i p_j + r_i r_j) \operatorname{sgn}(q_j - q_i) e^{-|q_i - q_j|}, \\
 \dot{r}_i & = \sum_{j=1, j \neq i}^n (p_i p_j + r_i r_j) \operatorname{sgn}(q_j - q_i) e^{-|q_i - q_j|}, \quad (86) \\
 \dot{q}_i & = - \sum_{j=1}^n (p_j + r_j) e^{-|q_i - q_j|}.
 \end{aligned}$$

Peakons interact in a way similar to that of solitons of the CH equation, and wave breaking may appear when at least two of the q_i coincide. Clearly, if the q_i remain distinct, the system (86) allows for a global smooth solution. It is not hard to see that $z = (u, v)$ is a global weak solution of system (8) by inserting that solution into (85). In the case where $p_i(0)$ and $r_i(0)$ have the same sign for all $i = 1, 2, \dots, n$, (86) admits a unique global solution, where the $q_i(t)$ remain distinct and the peakons are traveling in the same direction. However, when two peakons have opposite signs, collisions may occur, and, if so, the system (86) blows up.

Let us consider initial data $\bar{z} = (\bar{u}, \bar{v})$ given by

$$\begin{aligned}
 \bar{u}(x) & = \sum_{i=1}^n p_i e^{-|x - \xi_i|}, \\
 \bar{v}(x) & = \sum_{i=1}^n r_i e^{-|x - \xi_i|}. \quad (87)
 \end{aligned}$$

Without loss of generality, we assume that the p_i and r_i are all nonzero and that the ξ_i are all distinct. The aim is to characterize the unique and global weak solution from Theorem 9 with initial data (87) explicitly. Since the variables

p_i and r_i blow up at collisions, they are not appropriate to define a multipeakon in the form of (85). We consider the following characterization of multipeakons given as continuous solutions $z = (u, v)$, which are defined on intervals $[y_i, y_{i+1}]$ as the solutions of the Dirichlet problem

$$z - z_{xx} = 0, \quad (88)$$

with boundary conditions $z(t, y_i(t)) = z_i(t)$ and $z(t, y_{i+1}(t)) = z_{i+1}(t)$. The variables y_i denote the position of the peaks, and the variables z_i denote the values of z at the peaks. In the following we will show that this property persists for conservative solutions.

Let us set $A = R \setminus \{\xi_1, \dots, \xi_n\}$. The next lemma gives us the functions \bar{U} , \bar{V} , and \bar{H} which belong to $C^2(A)$ (they even belong to $C^\infty(A)$).

Lemma 10. *Let $\bar{X} = (\bar{y}, \bar{U}, \bar{V}, \bar{M}, \bar{N}, \bar{H}) \in F$ such that $(\bar{y}, \bar{U}, \bar{V}, \bar{H}) \in [C^2(A)]^4$ is given, and then the solution (y, U, V, H) of (21) with initial data \bar{X} belongs to $C^1(R^+, [C^2(A)]^4)$.*

Proof. To prove this Lemma, one proceeds as in Theorem 3 by using the contraction argument and replacing E by

$$\bar{E} = E \cap \left[(C^2(A))^3 \cap (L^2)^2 \cap C^2(A) \right], \quad (89)$$

endowed with the norm

$$\begin{aligned} \|X\|_{\bar{E}} &= \|X\|_E + \|y - Id\|_{W^{2,\infty}(A)} + \|U\|_{W^{2,\infty}(A)} \\ &\quad + \|V\|_{W^{2,\infty}(A)} + \|H\|_{W^{2,\infty}(A)}. \end{aligned} \quad (90)$$

Our main task is to prove the Lipschitz continuity of P_i and $P_{i,x}$ ($i = 1, 2, 3, 4$) from \bar{E} to $H^1(R) \cap C^2(A)$. We first show that $P_{2,x}$ is Lipschitz continuous from \bar{E} to $H^1(R) \cap C^2(A)$ and the others follow in the same way. Given a bounded set $B = \{X \in \bar{E} \mid \|X\|_{\bar{E}} \leq C_B\}$ where C_B is a positive constant, from Theorem 3, we get that

$$\|P_{2,x}(X) - P_{2,x}(\bar{X})\|_{L^\infty(R)} \leq C \|X - \bar{X}\|_E \leq C \|X - \bar{X}\|_{\bar{E}} \quad (91)$$

for a constant C depending only on C_B . We can compute the derivative of $P_{2,x}$ given as

$$P_{2,x\xi} = -\frac{1}{2}H_\xi + \left(P_2 - \frac{1}{2}U^2 - MN + N^2 \right) (1 + \zeta_\xi). \quad (92)$$

From Lemma 2, $P_{2,x\xi}$ is Lipschitz continuous from \bar{E} to $C(A)$ and therefore $P_{2,x}$ is Lipschitz continuous from \bar{E} to $C^1(A)$. Similarly, we obtain the same results for P_1, P_2, P_3 , and P_4 and $P_{1,x}, P_{3,x}, P_{4,x}$. We can also compute the derivative of $P_{2,x\xi}$ on A as follows:

$$\begin{aligned} P_{2,x\xi\xi} &= -\frac{1}{2}H_{\xi\xi} + \left(P_{2,x}y_\xi - UU_\xi - M_\xi N - MN_\xi + 2NN_\xi \right) y_\xi \\ &\quad + \left(P_2 - \frac{1}{2}U^2 - MN + N^2 \right) y_{\xi\xi}. \end{aligned} \quad (93)$$

Since $P_{2,x\xi\xi}$ is locally Lipschitz maps from \bar{E} to $C(A)$, we then get that $P_{2,x}$ is locally Lipschitz continuous from \bar{E} to $C^2(A)$. The same results can be obtained for the other P_i and $P_{i,x}$ ($i = 1, 2, 3, 4$) by the same way. From the standard contraction argument, the local existence of solutions of (21) can be proved in \bar{E} . As far as global existence is concerned, $\|X\|_{W^{1,\infty}(R)}$ does not blow up for initial data in $W^{1,\infty}(R)$. For the second derivative, for any $\xi \in A$, we get that

$$\begin{aligned} y_{\xi\xi t} &= U_{\xi\xi} + V_{\xi\xi}, \\ U_{\xi\xi t} &= \frac{1}{2}H_{\xi\xi} + \left(\frac{1}{2}U^2 + MN - N^2 - P_2 - P_{1,x} \right) y_{\xi\xi} \\ &\quad + \left(UU_\xi + M_\xi N + MN_\xi - 2NN_\xi \right) y_\xi \\ &\quad - \left(P_{2,x} + P_1 - UN \right) y_\xi^2, \\ V_{\xi\xi t} &= \frac{1}{2}H_{\xi\xi} + \left(\frac{1}{2}V^2 + MN - M^2 - P_4 - P_{3,x} \right) y_{\xi\xi} \\ &\quad + \left(VV_\xi + N_\xi M + NM_\xi - 2MN_\xi \right) y_\xi \\ &\quad - \left(P_{4,x} + P_3 - VM \right) y_\xi^2, \\ H_{\xi\xi t} &= - \left(2UP_{2,x} + 2VP_{4,x} \right) y_{\xi\xi} + \left(3U^2 - 2P_2 \right) U_{\xi\xi} \\ &\quad + \left(3V^2 - 2P_4 \right) V_{\xi\xi} + 6UU_\xi^2 - 4P_{2,x}U_\xi y_\xi \\ &\quad + 6VV_\xi^2 - 4P_{4,x}V_\xi y_\xi - 2UP_2 y_\xi^2 + U^3 y_\xi^2 \\ &\quad + 2UMNy_\xi^2 - 2UN^2 y_\xi^2 + UH_\xi y_\xi - 2VP_4 y_\xi^2 \\ &\quad + V^3 y_\xi^2 + 2VMNy_\xi^2 - 2VM^2 y_\xi^2 + VH_\xi y_\xi. \end{aligned} \quad (94)$$

The system (94) is affine with respect to $y_{\xi\xi}, U_{\xi\xi}, V_{\xi\xi}, H_{\xi\xi}$. Thus, we get that

$$\begin{aligned} \|X_{\xi\xi}(t, \cdot)\|_{L^\infty(A)} \\ \leq \|X_{\xi\xi}(0, \cdot)\|_{L^\infty(A)} + C + C \int_0^t \|X_{\xi\xi}(\tau, \cdot)\|_{L^\infty(A)} d\tau, \end{aligned} \quad (95)$$

where C is a constant depending only on $\sup_{t \in [0, T]} \|X\|_{W^{1,\infty}(R)}$ which is bounded on any time interval $[0, T]$. It follows from Gronwall's lemma that $\|X\|_{W^{2,\infty}(A)}$ does not blow up and therefore the solution is globally defined in \bar{E} . \square

We now prove that $\bar{X} = (\bar{y}, \bar{U}, \bar{V}, \bar{M}, \bar{N}, \bar{H})$ is a representative of $z = (u, v)$ in the Lagrangian system; that is, $[\bar{X}] = L(\bar{z}, \bar{\mu})$, where $\bar{X} = (\bar{y}, \bar{U}, \bar{V}, \bar{M}, \bar{N}, \bar{H})$ is given by

$$\bar{y}(\xi) = \xi, \quad (96a)$$

$$\bar{U}(\xi) = \bar{u}(\xi), \quad \bar{V}(\xi) = \bar{v}(\xi), \quad (96b)$$

$$\bar{M}(\xi) = \bar{u}_x(\xi), \quad \bar{N}(\xi) = \bar{v}_x(\xi),$$

$$\bar{H}(\xi) = \int_{-\infty}^{\xi} (\bar{u}^2 + \bar{u}_x^2 + \bar{v}^2 + \bar{v}_x^2) dx. \quad (96c)$$

We first check that $\bar{X} \in F$. Since $\bar{z} = (\bar{u}, \bar{v})$ is a multipeakon, we get that $\bar{z} = (\bar{u}, \bar{v}) \in W^{1,\infty}(R) \cap H^1(R)$ from (87). Hence, \bar{U}, \bar{V} , and \bar{H} all belong to $W^{1,\infty}(R)$ while $\bar{y} - Id$ is identically zero. Due to the exponential decay of (\bar{u}, \bar{v}) and (\bar{u}_x, \bar{v}_x) and $\bar{H}_\xi \in L^\infty(R)$, we get that $\bar{H}_\xi \in L^2(R)$. The properties (25)–(27) are straightforward to check. It is not hard to check that $M([\bar{X}]) = (\bar{z}, (\bar{u}^2 + \bar{u}_x^2 + \bar{v}^2 + \bar{v}_x^2)dx)$ and, therefore, since $L \circ M = Id$, we get that $[\bar{X}] = L(\bar{z}, (\bar{u}^2 + \bar{u}_x^2 + \bar{v}^2 + \bar{v}_x^2)dx)$.

Theorem 11. *Let initial data be given in (87). The solution given by Theorem 9 satisfies $z - z_{xx} = 0$ between the peaks.*

Proof. Let us first prove that $u - u_{xx} = 0$. Assuming that $y_\xi(t, \xi) \neq 0$, we get that

$$u_x \circ y = M, \quad u_{xx} \circ y = \frac{M_\xi}{y_\xi} = \frac{(U_{\xi\xi}y_\xi - y_{\xi\xi}U_\xi)}{y_\xi^3}, \quad (97)$$

and therefore

$$(u - u_{xx}) \circ y = \frac{(Uy_\xi^3 - U_{\xi\xi}y_\xi + y_{\xi\xi}U_\xi)}{y_\xi^3}. \quad (98)$$

We set

$$R = Uy_\xi^3 - U_{\xi\xi}y_\xi + y_{\xi\xi}U_\xi. \quad (99)$$

For a given $\xi \in A$, differentiating (99) with respect to t and after using (21), (24), and (94), we obtain

$$\begin{aligned} \frac{dR}{dt} &= 3Uy_\xi^2y_{\xi t} + U_t y_\xi^3 - U_{\xi\xi t}y_\xi - U_{\xi\xi}y_{\xi t} + y_{\xi\xi t}U_\xi + y_{\xi\xi}U_{\xi t} \\ &= 2UU_\xi y_\xi^2 + 2UV_\xi y_\xi^2 - \frac{H_{\xi\xi}y_\xi}{2} \\ &\quad - (M_\xi N + N_\xi M - 2NN_\xi) y_\xi^2 - U_{\xi\xi}V_\xi \\ &\quad + V_{\xi\xi}U_\xi + \frac{H_\xi y_{\xi\xi}}{2} \\ &= 2U(U_\xi + V_\xi) y_\xi^2 - 2N(M_\xi - N_\xi) y_\xi^2 \\ &\quad - \frac{H_{\xi\xi}y_\xi}{2} + \frac{H_\xi y_{\xi\xi}}{2}. \end{aligned} \quad (100)$$

Differentiating (27) with respect to ξ , we get

$$\begin{aligned} y_{\xi\xi}H_\xi + y_\xi H_{\xi\xi} &= 2y_\xi y_{\xi\xi}U^2 + 2y_\xi^2 UU_\xi + 2U_\xi U_{\xi\xi} \\ &\quad + 2y_\xi y_{\xi\xi}V^2 + 2y_\xi^2 VV_\xi + 2V_\xi V_{\xi\xi}. \end{aligned} \quad (101)$$

After inserting the value of $y_\xi H_{\xi\xi}$ given by (101) into (100) and multiplying the equation by y_ξ , we obtain that

$$\begin{aligned} y_\xi \cdot \frac{dR}{dt} &= UU_\xi y_\xi^3 - U_\xi U_{\xi\xi} y_\xi + (H_\xi y_\xi - y_\xi^2 U^2 - y_\xi^2 V^2 - V_\xi^2) y_{\xi\xi} \\ &\quad + UV_\xi y_\xi^3 - U_{\xi\xi} V_\xi y_\xi + U_\xi V_\xi y_{\xi\xi}. \end{aligned} \quad (102)$$

It follows from (27), and since $y_{\xi t} = (U_\xi + V_\xi)$, that

$$y_\xi \cdot \frac{dR}{dt} = y_{\xi t} \cdot R. \quad (103)$$

We claim that, for any time t such that $y_\xi(t) \neq 0$,

$$\frac{d}{dt} \left(\frac{R}{y_\xi} \right) = \frac{R_t y_\xi - y_{\xi t} R}{y_\xi^2} = 0. \quad (104)$$

We have to prove that R/y_ξ is C^1 in time. Since

$$\begin{aligned} \frac{R}{y_\xi} &= Uy_\xi^2 - U_{\xi\xi} + \frac{y_{\xi\xi}U_\xi}{y_\xi} \\ &= Uy_\xi^2 - U_{\xi\xi} + \frac{y_{\xi\xi}U_\xi}{y_\xi + H_\xi} + \frac{y_{\xi\xi}MH_\xi}{y_\xi + H_\xi} = \frac{J(X, X_\xi, X_{\xi\xi})}{y_\xi + H_\xi}, \end{aligned} \quad (105)$$

for some polynomial J and $X \in C^1(R, \bar{E})$, we get that X, X_ξ , and $X_{\xi\xi}$ are C^1 in time. Since $X(t)$ remains in Γ , for all t , from (26), we have $y_\xi + H_\xi > 0$ and therefore $1/(y_\xi + H_\xi)$ is C^1 in time, which implies that R/y_ξ is C^1 in time. Hence, it holds that

$$R(t, \xi) = K(\xi) y_\xi(t, \xi), \quad (106)$$

for some constant $K(\xi)$ which is independent of time, which leads to

$$y_\xi^2 (u - u_{xx}) \circ y = K(\xi). \quad (107)$$

For the multipeakons at time $t = 0$, we have $y(0, \xi) = \xi$ and $(u - u_{xx})(0, \xi) = 0$ for all $\xi \in A$. Hence,

$$\frac{R}{y_\xi}(t, \xi) = 0, \quad (108)$$

for all time t and all $\xi \in A$. Thus, $(u - u_{xx})(t, \xi) = 0$. Similarly, $(v - v_{xx})(t, \xi) = 0$. \square

For solutions with multipeakon initial data, we have the following result. If $y_\xi(t, \xi)$ vanishes at some point $\bar{\xi}$ in the interval (ξ_i, ξ_{i+1}) , then $y_\xi(t, \xi)$ vanishes everywhere in (ξ_i, ξ_{i+1}) . Moreover, for given initial multipeakon solution $\bar{z}(x) = (\bar{u}, \bar{v})(x) = (\sum_{i=1}^n p_i e^{-|x-\xi_i|}, \sum_{i=1}^n r_i e^{-|x-\xi_i|})$, let (y, U, V, M, N, H) be the solution of system (21) with initial data $(\bar{y}, \bar{U}, \bar{V}, \bar{M}, \bar{N}, \bar{H})$ given by (96a), (96b) and (96c), and then, between adjacent peaks, if $x_i = y(t, \xi_i) \neq x_{i+1} = y(t, \xi_{i+1})$, the solution $z(t, x) = (u, v)(t, x)$ is twice differentiable with respect to the space variable and we have $(z - z_{xx}) = 0$, for $x \in (x_i, x_{i+1})$.

We now start the derivation of a system of ordinary differential equations for multipeakons.

From (21), we get that, for each $i = 1, 2, \dots, n$,

$$\begin{aligned} \frac{dy_i}{dt} &= u_i + v_i, \\ \frac{du_i}{dt} &= -P_{1,i} - P_{2,xi}, \\ \frac{dv_i}{dt} &= -P_{3,i} - P_{4,xi}, \\ \frac{dH_i}{dt} &= u_i^3 - 2u_i P_{2,i} + v_i^3 - 2v_i P_{4,i}, \end{aligned} \tag{109}$$

where $(y_i, u_i, v_i, H_i) = (y, U, V, H)(t, \xi_i)$, $P_{k,i} = P_k(t, \xi_i)$, and $P_{k,xi} = P_{k,x}(t, \xi_i)$, ($k = 1, 2, 3, 4$), respectively. Since the function $y(t, \cdot)$ is invertible, for almost every t , we can use the change of variables $x = y(t, \xi)$ such that $P_{k,i}$ and $P_{k,xi}$ ($k = 1, 2, 3, 4$) can be rewritten as

$$\begin{aligned} P_{1,i} &= \frac{1}{2} \int_R e^{-|y_i-x|} (uv_x) dx, \\ P_{1,xi} &= -\frac{1}{2} \int_R \operatorname{sgn}(y_i-x) e^{-|y_i-x|} (uv_x) dx, \\ P_{2,i} &= \frac{1}{2} \int_R e^{-|y_i-x|} \left(u^2 + \frac{1}{2}u_x^2 + u_x v_x + \frac{1}{2}v^2 - \frac{1}{2}v_x^2 \right) dx, \\ P_{2,xi} &= -\frac{1}{2} \int_R \operatorname{sgn}(y_i-x) e^{-|y_i-x|} \\ &\quad \times \left(u^2 + \frac{1}{2}u_x^2 + u_x v_x + \frac{1}{2}v^2 - \frac{1}{2}v_x^2 \right) dx, \\ P_{3,i} &= \frac{1}{2} \int_R e^{-|y_i-x|} (vu_x) dx, \\ P_{3,xi} &= -\frac{1}{2} \int_R \operatorname{sgn}(y_i-x) e^{-|y_i-x|} (vu_x) dx, \\ P_{4,i} &= \frac{1}{2} \int_R e^{-|y_i-x|} \left(v^2 + \frac{1}{2}v_x^2 + u_x v_x + \frac{1}{2}u^2 - \frac{1}{2}u_x^2 \right) dx, \\ P_{4,xi} &= -\frac{1}{2} \int_R \operatorname{sgn}(y_i-x) e^{-|y_i-x|} \\ &\quad \times \left(v^2 + \frac{1}{2}v_x^2 + u_x v_x + \frac{1}{2}u^2 - \frac{1}{2}u_x^2 \right) dx. \end{aligned} \tag{110}$$

Between two adjacent peaks located at y_i and y_{i+1} , we know that $z = (u, v)$ satisfies $(z - z_{xx}) = 0$ and therefore $z = (u, v)$ can be written as

$$z(x) = \begin{pmatrix} u(x) \\ v(x) \end{pmatrix} = \begin{pmatrix} A_i e^x + B_i e^{-x} \\ C_i e^x + D_i e^{-x} \end{pmatrix}, \tag{111}$$

for $x \in [y_i, y_{i+1}]$, $i = 1, 2, \dots, n-1$, where the constants A_i , B_i , C_i , and D_i depend on $u_i, u_{i+1}, v_i, v_{i+1}, y_i$, and y_{i+1} and read

$$\begin{aligned} A_i &= \frac{e^{-\bar{y}_i}}{2} \left[\frac{\bar{u}_i}{\cosh(\delta y_i)} + \frac{\delta u_i}{\sinh(\delta y_i)} \right], \\ B_i &= \frac{e^{\bar{y}_i}}{2} \left[\frac{\bar{u}_i}{\cosh(\delta y_i)} - \frac{\delta u_i}{\sinh(\delta y_i)} \right], \\ C_i &= \frac{e^{-\bar{y}_i}}{2} \left[\frac{\bar{v}_i}{\cosh(\delta y_i)} + \frac{\delta v_i}{\sinh(\delta y_i)} \right], \\ D_i &= \frac{e^{\bar{y}_i}}{2} \left[\frac{\bar{v}_i}{\cosh(\delta y_i)} - \frac{\delta v_i}{\sinh(\delta y_i)} \right], \end{aligned} \tag{112}$$

where

$$\begin{aligned} \bar{y}_i &= \frac{1}{2} (y_i + y_{i+1}), & \delta y_i &= \frac{1}{2} (y_i - y_{i+1}), \\ \bar{u}_i &= \frac{1}{2} (u_i + u_{i+1}), & \delta u_i &= \frac{1}{2} (u_i - u_{i+1}), \\ \bar{v}_i &= \frac{1}{2} (v_i + v_{i+1}), & \delta v_i &= \frac{1}{2} (v_i - v_{i+1}). \end{aligned} \tag{113}$$

Thus, the constants A_i, B_i, C_i , and D_i uniquely determine $z = (u, v)$ on the interval $[y_i, y_{i+1}]$, and we compute

$$\begin{aligned} \delta H_i &= H_{i+1} - H_i \\ &= \int_{y_i}^{y_{i+1}} (u^2 + u_x^2 + v^2 + v_x^2) dx \\ &= 2\bar{u}_i^2 \tanh(\delta y_i) + 2\delta u_i^2 \coth(\delta y_i) \\ &\quad + 2\bar{v}_i^2 \tanh(\delta y_i) + 2\delta v_i^2 \coth(\delta y_i) \\ &= \delta H_{1i} + \delta H_{2i}, \end{aligned} \tag{114}$$

where $\delta H_{1i} = 2\bar{u}_i^2 \tanh(\delta y_i) + 2\delta u_i^2 \coth(\delta y_i)$ and $\delta H_{2i} = 2\bar{v}_i^2 \tanh(\delta y_i) + 2\delta v_i^2 \coth(\delta y_i)$. At this point, we can get some more understanding of what is happening at a time of collision. Let t^* be a time when the two peaks located at y_i and y_{i+1} collide, that is, such that $\lim_{t \rightarrow t^*} \delta y_i(t) = 0$. Since the solution $z = (u, v)$ remains in H^1 for all time, the function $z = (u, v)$ remains continuous so that we have $\lim_{t \rightarrow t^*} \delta u_i(t) = \lim_{t \rightarrow t^*} \delta v_i(t) = 0$, and when t tends to t^* , A_i, B_i, C_i , and D_i may have a finite limit. However, we know that the first derivative blows up, which implies that $\lim_{t \rightarrow t^*} B_i = -\lim_{t \rightarrow t^*} A_i = \infty$ and $\lim_{t \rightarrow t^*} D_i = -\lim_{t \rightarrow t^*} C_i = \infty$. Thus, δu_i and δv_i tend to zero, respectively, but are slower than δy_i . Indeed, let t tend to t^* in (114), and then, to first order in δy_i , we obtain that

$$\sqrt{\delta u_i^2 + \delta v_i^2} = \sqrt{\frac{\delta H_i}{2}} \cdot \sqrt{\delta y_i + o(\delta y_i)}, \tag{115}$$

which implies that δu_i and δv_i tend to zero at the same rate as $\sqrt{\delta y_i}$. We now turn to the computation of $P_{k,i}$ ($k = 1, 2, 3, 4$) given by (110). Let us write $z = (u, v)$ as

$$\begin{aligned} z(t, x) &= (u(t, x), v(t, x)) \\ &= \left(\sum_{j=0}^n (A_j e^x + B_j e^{-x}) \chi_{(y_j, y_{j+1})}(x), \right. \\ &\quad \left. \sum_{j=0}^n (C_j e^x + D_j e^{-x}) \chi_{(y_j, y_{j+1})}(x) \right). \end{aligned} \quad (116)$$

We have sets $y_0 = -\infty$, $y_{n+1} = \infty$, $u_0 = u_{n+1} = 0$, $v_0 = v_{n+1} = 0$, $A_0 = u_1 e^{-y_1}$, $B_0 = 0$, $A_n = 0$, $B_n = u_n e^{y_n}$, and $C_0 = v_1 e^{-y_1}$, $D_0 = 0$, $C_n = 0$, $D_n = v_n e^{y_n}$. We have

$$\begin{aligned} uv_x &= \sum_{j=0}^n (A_j C_j e^{2x} - A_j D_j + B_j C_j - B_j D_j e^{-2x}) \chi_{(y_j, y_{j+1})}, \\ u^2 + \frac{1}{2} u_x^2 + u_x v_x + \frac{1}{2} v^2 - \frac{1}{2} v_x^2 \\ &= \sum_{j=0}^n \left(\left(\frac{3}{2} A_j^2 + A_j C_j \right) e^{2x} \right. \\ &\quad \left. + (A_j B_j - A_j D_j - B_j C_j + 2C_j D_j) \right. \\ &\quad \left. + \left(\frac{3}{2} B_j^2 + B_j D_j \right) e^{-2x} \right) \chi_{(y_j, y_{j+1})}, \\ v u_x &= \sum_{j=0}^n (A_j C_j e^{2x} - C_j B_j + D_j A_j - B_j D_j e^{-2x}) \chi_{(y_j, y_{j+1})}, \\ v^2 + \frac{1}{2} v_x^2 + u_x v_x + \frac{1}{2} u^2 - \frac{1}{2} u_x^2 \\ &= \sum_{j=0}^n \left(\left(\frac{3}{2} C_j^2 + A_j C_j \right) e^{2x} \right. \\ &\quad \left. + (C_j D_j - C_j B_j - D_j A_j + 2A_j B_j) \right. \\ &\quad \left. + \left(\frac{3}{2} D_j^2 + D_j B_j \right) e^{-2x} \right) \chi_{(y_j, y_{j+1})}. \end{aligned} \quad (117)$$

We set

$$k_{ij} = \begin{cases} -1 & \text{if } i \leq j, \\ 1 & \text{otherwise.} \end{cases} \quad (118)$$

By inserting (117) into (110), we get

$$\begin{aligned} P_{1,i} &= \frac{1}{2} \sum_{j=0}^n \int_{y_j}^{y_{j+1}} e^{-k_{ij}(y_i-x)} \\ &\quad \times (A_j C_j e^{2x} - A_j D_j + B_j C_j - B_j D_j e^{-2x}) dx, \\ P_{2,i} &= \frac{1}{2} \sum_{j=0}^n \int_{y_j}^{y_{j+1}} e^{-k_{ij}(y_i-x)} \left(\left(\frac{3}{2} A_j^2 + A_j C_j \right) e^{2x} \right. \\ &\quad \left. + (A_j B_j - A_j D_j \right. \\ &\quad \left. - B_j C_j + 2C_j D_j) \right. \\ &\quad \left. + \left(\frac{3}{2} B_j^2 + B_j D_j \right) e^{-2x} \right) dx, \\ P_{3,i} &= \frac{1}{2} \sum_{j=0}^n \int_{y_j}^{y_{j+1}} e^{-k_{ij}(y_i-x)} \\ &\quad \times (A_j C_j e^{2x} - C_j B_j + D_j A_j - B_j D_j e^{-2x}) dx, \\ P_{4,i} &= \frac{1}{2} \sum_{j=0}^n \int_{y_j}^{y_{j+1}} e^{-k_{ij}(y_i-x)} \\ &\quad \times \left(\left(\frac{3}{2} C_j^2 + A_j C_j \right) e^{2x} \right. \\ &\quad \left. + (C_j D_j - C_j B_j - D_j A_j + 2A_j B_j) \right. \\ &\quad \left. + \left(\frac{3}{2} D_j^2 + D_j B_j \right) e^{-2x} \right). \end{aligned} \quad (119)$$

It follows from (112) and (114) that

$$\begin{aligned} A_j^2 &= \left[\frac{e^{-\bar{y}_j}}{2} \left(\frac{\bar{u}_j}{\cosh(\delta y_j)} + \frac{\delta u_j}{\sinh(\delta y_j)} \right) \right]^2 \\ &= \frac{e^{-2\bar{y}_j}}{\sinh^2(2\delta y_j)} \left[\bar{u}_j^2 \sinh^2(\delta y_j) \right. \\ &\quad \left. + 2\bar{u}_j \delta u_j \sinh(\delta y_j) \cosh(\delta y_j) \right. \\ &\quad \left. + \delta u_j^2 \cosh^2(\delta y_j) \right] \\ &= \frac{e^{-2\bar{y}_j}}{4 \sinh(2\delta y_j)} \left[\delta H_{1j} + 4\bar{u}_j \delta u_j \right], \end{aligned} \quad (120)$$

$$\begin{aligned} A_j B_j &= \frac{e^{-\bar{y}_j}}{2} \left(\frac{\bar{u}_j}{\cosh(\delta y_j)} + \frac{\delta u_j}{\sinh(\delta y_j)} \right) \\ &\quad \cdot \frac{e^{\bar{y}_j}}{2} \left(\frac{\bar{u}_j}{\cosh(\delta y_j)} - \frac{\delta u_j}{\sinh(\delta y_j)} \right) \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{4 \sinh(2\delta y_j)} [2\bar{u}_j^2 \tanh(\delta y_j) - 2\delta u_j^2 \coth(\delta y_j)] \\
 &= \frac{1}{4 \sinh(2\delta y_j)} [4\bar{u}_j^2 \tanh(\delta y_j) - \delta H_{1j}],
 \end{aligned} \tag{121}$$

$$\begin{aligned}
 A_j C_j &= \frac{e^{-\bar{y}_j}}{2} \left(\frac{\bar{u}_j}{\cosh(\delta y_j)} + \frac{\delta u_j}{\sinh(\delta y_j)} \right) \\
 &\cdot \frac{e^{-\bar{y}_j}}{2} \left(\frac{\bar{v}_j}{\cosh(\delta y_j)} + \frac{\delta v_j}{\sinh(\delta y_j)} \right) \\
 &= \frac{e^{-2\bar{y}_j}}{2 \sinh(2\delta y_j)} [\bar{u}_j \bar{v}_j \tanh(\delta y_j) + \delta u_j \bar{v}_j \\
 &\quad + \delta v_j \bar{u}_j + \delta u_j \delta v_j \coth(\delta y_j)],
 \end{aligned} \tag{122}$$

$$\begin{aligned}
 A_j D_j &= \frac{e^{-\bar{y}_j}}{2} \left(\frac{\bar{u}_j}{\cosh(\delta y_j)} + \frac{\delta u_j}{\sinh(\delta y_j)} \right) \\
 &\cdot \frac{e^{\bar{y}_j}}{2} \left(\frac{\bar{v}_j}{\cosh(\delta y_j)} - \frac{\delta v_j}{\sinh(\delta y_j)} \right) \\
 &= \frac{1}{2 \sinh(2\delta y_j)} [\bar{u}_j \bar{v}_j \tanh(\delta y_j) + \delta u_j \bar{v}_j \\
 &\quad - \delta v_j \bar{u}_j - \delta u_j \delta v_j \coth(\delta y_j)],
 \end{aligned} \tag{123}$$

$$\begin{aligned}
 B_j C_j &= \frac{e^{\bar{y}_j}}{2} \left(\frac{\bar{u}_j}{\cosh(\delta y_j)} - \frac{\delta u_j}{\sinh(\delta y_j)} \right) \\
 &\cdot \frac{e^{-\bar{y}_j}}{2} \left(\frac{\bar{v}_j}{\cosh(\delta y_j)} + \frac{\delta v_j}{\sinh(\delta y_j)} \right) \\
 &= \frac{1}{2 \sinh(2\delta y_j)} [\bar{u}_j \bar{v}_j \tanh(\delta y_j) + \delta v_j \bar{u}_j \\
 &\quad - \delta u_j \bar{v}_j - \delta u_j \delta v_j \coth(\delta y_j)],
 \end{aligned} \tag{124}$$

$$\begin{aligned}
 B_j D_j &= \frac{e^{\bar{y}_j}}{2} \left(\frac{\bar{u}_j}{\cosh(\delta y_j)} - \frac{\delta u_j}{\sinh(\delta y_j)} \right) \\
 &\cdot \frac{e^{\bar{y}_j}}{2} \left(\frac{\bar{v}_j}{\cosh(\delta y_j)} - \frac{\delta v_j}{\sinh(\delta y_j)} \right) \\
 &= \frac{e^{2\bar{y}_j}}{2 \sinh(2\delta y_j)} [\bar{u}_j \bar{v}_j \tanh(\delta y_j) - \delta u_j \bar{v}_j \\
 &\quad - \delta v_j \bar{u}_j + \delta u_j \delta v_j \coth(\delta y_j)],
 \end{aligned} \tag{125}$$

$$\begin{aligned}
 B_j^2 &= \left[\frac{e^{\bar{y}_j}}{2} \left(\frac{\bar{u}_j}{\cosh(\delta y_j)} - \frac{\delta u_j}{\sinh(\delta y_j)} \right) \right]^2 \\
 &= \frac{e^{2\bar{y}_j}}{4 \sinh(2\delta y_j)} [\delta H_{1j} - 4\bar{u}_j \delta u_j].
 \end{aligned} \tag{126}$$

Thus, from (120)–(126), we can obtain that

$$\begin{aligned}
 &\int_{y_j}^{y_{j+1}} e^{-k_{ij}(y_i-x)} A_j^2 e^{2x} dx \\
 &= \frac{e^{-k_{ij}y_i} \cdot e^{k_{ij}\bar{y}_j}}{2(2+k_{ij}) \sinh(2\delta y_j)} \sinh((2+k_{ij})\delta y_j) \\
 &\quad \times [\delta H_{1j} + 4\bar{u}_j \delta u_j],
 \end{aligned} \tag{127}$$

$$\begin{aligned}
 &\int_{y_j}^{y_{j+1}} e^{-k_{ij}(y_i-x)} A_j C_j e^{2x} dx \\
 &= \frac{e^{-k_{ij}y_i} \cdot e^{k_{ij}\bar{y}_j}}{(2+k_{ij}) \sinh(2\delta y_j)} \sinh((2+k_{ij})\delta y_j) \\
 &\quad \cdot [\bar{u}_j \bar{v}_j \tanh(\delta y_j) + \delta u_j \bar{v}_j + \delta v_j \bar{u}_j \\
 &\quad + \delta u_j \delta v_j \coth(\delta y_j)],
 \end{aligned} \tag{128}$$

$$\begin{aligned}
 &\int_{y_j}^{y_{j+1}} e^{-k_{ij}(y_i-x)} A_j B_j dx \\
 &= \frac{e^{-k_{ij}y_i} \cdot e^{k_{ij}\bar{y}_j}}{2 \sinh(2\delta y_j)} \sinh(\delta y_j) \\
 &\quad \times [4\bar{u}_j^2 \tanh(\delta y_j) - \delta H_{1j}],
 \end{aligned} \tag{129}$$

$$\begin{aligned}
 &\int_{y_j}^{y_{j+1}} e^{-k_{ij}(y_i-x)} A_j D_j dx \\
 &= \frac{e^{-k_{ij}y_i} \cdot e^{k_{ij}\bar{y}_j}}{\sinh(2\delta y_j)} \sinh(\delta y_j) \\
 &\quad \cdot [\bar{u}_j \bar{v}_j \tanh(\delta y_j) + \delta u_j \bar{v}_j - \delta v_j \bar{u}_j - \delta u_j \delta v_j \coth(\delta y_j)],
 \end{aligned} \tag{130}$$

$$\begin{aligned}
 &\int_{y_j}^{y_{j+1}} e^{-k_{ij}(y_i-x)} B_j C_j dx \\
 &= \frac{e^{-k_{ij}y_i} \cdot e^{k_{ij}\bar{y}_j}}{\sinh(2\delta y_j)} \sinh(\delta y_j) \\
 &\quad \cdot [\bar{u}_j \bar{v}_j \tanh(\delta y_j) + \delta v_j \bar{u}_j - \delta u_j \bar{v}_j - \delta u_j \delta v_j \coth(\delta y_j)],
 \end{aligned} \tag{131}$$

$$\int_{y_j}^{y_{j+1}} e^{-k_{ij}(y_i-x)} C_j D_j dx$$

$$= \frac{e^{-k_{ij}y_i} \cdot e^{k_{ij}\bar{y}_j}}{2 \sinh(2\delta y_j)} \sinh(\delta y_j) [4\bar{v}_j^2 \tanh(\delta y_j) - \delta H_{2j}],$$
(132)

$$\int_{y_j}^{y_{j+1}} e^{-k_{ij}(y_i-x)} B_j^2 e^{-2x} dx$$

$$= \frac{e^{-k_{ij}y_i} \cdot e^{k_{ij}\bar{y}_j}}{2(k_{ij}-2) \sinh(2\delta y_j)} \sinh((k_{ij}-2)\delta y_j)$$

$$\times [\delta H_{1j} - 4\bar{u}_j \delta u_j],$$
(133)

$$\int_{y_j}^{y_{j+1}} e^{-k_{ij}(y_i-x)} B_j D_j e^{-2x} dx$$

$$= \frac{e^{-k_{ij}y_i} \cdot e^{k_{ij}\bar{y}_j}}{(k_{ij}-2) \sinh(2\delta y_j)} \sinh((k_{ij}-2)\delta y_j)$$

$$\cdot [\bar{u}_j \bar{v}_j \tanh(\delta y_j) - \delta u_j \bar{v}_j - \delta v_j \bar{u}_j + \delta u_j \delta v_j \coth(\delta y_j)].$$
(134)

It thus follows from (127)–(134) that

$$P_{1,i} = \sum_{j=0}^n \frac{e^{-k_{ij}y_i} \cdot e^{k_{ij}\bar{y}_j}}{6 \cosh(\delta y_j)} [2k_{ij} \bar{u}_j \bar{v}_j \sinh^2(\delta y_j) \tanh(\delta y_j)$$

$$+ 2k_{ij} \delta u_j \delta v_j \sinh(\delta y_j) \cosh(\delta y_j)$$

$$+ 2\delta u_j \bar{v}_j \cosh^2(\delta y_j)$$

$$+ 2\delta v_j \bar{u}_j \cosh^2(\delta y_j)$$

$$+ \delta u_j \bar{v}_j + \delta v_j \bar{u}_j - 3\bar{u}_j \bar{v}_j \tanh(\delta y_j)$$

$$+ 3\delta u_j \delta v_j \coth(\delta y_j)],$$

$$P_{2,i} = \sum_{j=0}^n \frac{e^{-k_{ij}y_i} \cdot e^{k_{ij}\bar{y}_j}}{4 \cosh(\delta y_j)} [\delta H_{1j} \cosh^2(\delta y_j)$$

$$+ 4k_{ij} \bar{u}_j \delta u_j \sinh^2(\delta y_j)$$

$$+ 2\bar{u}_j^2 \tanh(\delta y_j) + 4\bar{v}_j^2 \tanh(\delta y_j)$$

$$- \delta H_{2j} - \frac{4}{3} \bar{u}_j \bar{v}_j \tanh(\delta y_j)$$

$$+ \frac{8}{3} \delta u_j \delta v_j \coth(\delta y_j)]$$

$$+ \frac{4}{3} \bar{u}_j \bar{v}_j \cosh(\delta y_j) \sinh(\delta y_j)$$

$$+ \frac{4}{3} \delta u_j \delta v_j \coth(\delta y_j) \cosh^2(\delta y_j)$$

$$+ \frac{4}{3} k_{ij} \delta u_j \bar{v}_j \sinh^2(\delta y_j)$$

$$+ \frac{4}{3} k_{ij} \delta v_j \bar{u}_j \sinh^2(\delta y_j)].$$
(135)

We can also write $P_{1,i}$ and $P_{2,i}$ as

$$P_{1,i} = \sum_{j=0}^n P_{1,ij}, \quad P_{2,i} = \sum_{j=0}^n P_{2,ij},$$
(136)

where

$$P_{1,ij} = \begin{cases} \frac{1}{6} u_1 v_1 e^{y_1 - y_i}, & \text{for } j = 0, \\ \frac{e^{-k_{ij}y_i} \cdot e^{k_{ij}\bar{y}_j}}{6 \cosh(\delta y_j)} \\ \times [2k_{ij} \bar{u}_j \bar{v}_j \sinh^2(\delta y_j) \tanh(\delta y_j) \\ + 2k_{ij} \delta u_j \delta v_j \sinh(\delta y_j) \cosh(\delta y_j) \\ + 2\delta u_j \bar{v}_j \cosh^2(\delta y_j) \\ + 2\delta v_j \bar{u}_j \cosh^2(\delta y_j) + \delta u_j \bar{v}_j \\ + \delta v_j \bar{u}_j - 3\bar{u}_j \bar{v}_j \tanh(\delta y_j) \\ + 3\delta u_j \delta v_j \coth(\delta y_j)], & \text{for } j = 1, \dots, n-1, \\ -\frac{1}{6} u_n v_n e^{-y_n + y_i} & \text{for } j = n, \end{cases}$$

$$P_{2,ij} = \begin{cases} \frac{1}{4} u_1^2 e^{y_1 - y_i} + \frac{1}{6} u_1 v_1 e^{y_1 - y_i}, & \text{for } j = 0, \\ \frac{e^{-k_{ij}y_i} \cdot e^{k_{ij}\bar{y}_j}}{4 \cosh(\delta y_j)} \\ \times [\delta H_{1j} \cosh^2(\delta y_j) \\ + 4k_{ij} \bar{u}_j \delta u_j \sinh^2(\delta y_j) + 2\bar{u}_j^2 \tanh(\delta y_j) \\ + 4\bar{v}_j^2 \tanh(\delta y_j) - \delta H_{2j} - \frac{4}{3} \bar{u}_j \bar{v}_j \tanh(\delta y_j) \\ + \frac{8}{3} \delta u_j \delta v_j \coth(\delta y_j) \\ + \frac{4}{3} \bar{u}_j \bar{v}_j \cosh(\delta y_j) \sinh(\delta y_j) \\ + \frac{4}{3} \delta u_j \delta v_j \cdot \coth(\delta y_j) \cosh^2(\delta y_j) \\ + \frac{4}{3} k_{ij} \delta u_j \bar{v}_j \sinh^2(\delta y_j) \\ + \frac{4}{3} k_{ij} \delta v_j \bar{u}_j \sinh^2(\delta y_j)], & \text{for } j = 1, \dots, n-1, \\ \frac{1}{4} u_n^2 e^{-y_n + y_i} + \frac{1}{6} u_n v_n e^{-y_n + y_i}, & \text{for } j = n. \end{cases}$$
(137)

The terms $P_{3,i}$, $P_{4,i}$, and $P_{k,ix}$ ($k = 1, 2, 3, 4$) can be computed in the same way and we have

$$\begin{aligned}
 P_{1,ix} &= -\sum_{j=0}^n k_{ij} P_{1,ij}, & P_{2,ix} &= -\sum_{j=0}^n k_{ij} P_{2,ij}, \\
 P_{3,ix} &= -\sum_{j=0}^n k_{ij} P_{3,ij}, & P_{4,ix} &= -\sum_{j=0}^n k_{ij} P_{4,ij}.
 \end{aligned}
 \tag{138}$$

The result can be summarized in the following theorem.

Theorem 12. Assume that $\bar{y}_i = \xi_i$, $\bar{z}_i = (\bar{u}_i, \bar{v}_i) = (\bar{u}(\xi_i), \bar{v}(\xi_i))$ and $\bar{H}_i = \int_{-\infty}^{\xi_i} (\bar{u}^2 + \bar{u}_x^2 + \bar{v}^2 + \bar{v}_x^2) dx$ for $i = 1, \dots, n$ with a multipeakon initial data $\bar{z} = (\bar{u}, \bar{v})$ as given by (87). Then, there exists a global solution (y_i, u_i, v_i, H_i) of (109), (136), and (138) with initial data $(\bar{y}_i, \bar{u}_i, \bar{v}_i, \bar{H}_i)$. On each interval $[y_i(t), y_{i+1}(t)]$, one defines $z(t, x) = (u, v)(t, x)$ as the solution of the Dirichlet problem $z - z_{xx} = 0$ with boundary conditions $z(t, y_i(t)) = z_i(t)$ and $z(t, y_{i+1}(t)) = z_{i+1}(t)$ for each time t . Thus $z = (u, v)$ is a conservative solution of the modified coupled two-component Camassa-Holm system, which is the multipeakon conservative solution.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported in part by the Major State Basic Research Development Program 973 (no. 2012CB215202), the National Natural Science Foundation of China (no. 61134001), and Key Laboratory of Dependable Service Computing in Cyber Physical Society (Chongqing University), Ministry of Education.

References

- [1] Y. Fu, Y. Liu, and C. Qu, “Well-posedness and blow-up solution for a modified two-component periodic Camassa-Holm system with peakons,” *Mathematische Annalen*, vol. 348, no. 2, pp. 415–448, 2010.
- [2] Y. Fu and C. Qu, “Well posedness and blow-up solution for a new coupled Camassa-Holm equations with peakons,” *Journal of Mathematical Physics*, vol. 50, no. 1, article 012906, 2009.
- [3] A. Constantin, “The Hamiltonian structure of the Camassa-Holm equation,” *Expositiones Mathematicae*, vol. 15, no. 1, pp. 53–85, 1997.
- [4] R. Camassa and D. D. Holm, “An integrable shallow water equation with peaked solitons,” *Physical Review Letters*, vol. 71, no. 11, pp. 1661–1664, 1993.
- [5] A. Constantin, “On the scattering problem for the Camassa-Holm equation,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 457, no. 2008, pp. 953–970, 2001.
- [6] H. R. Dullin, G. A. Gottwald, and D. D. Holm, “An integrable shallow water equation with linear and nonlinear dispersion,” *Physical Review Letters*, vol. 87, no. 19, Article ID 194501, 2001.

- [7] A. Constantin and D. Lannes, “The hydrodynamical relevance of the Camassa-Holm and Degasperis-Procesi equations,” *Archive for Rational Mechanics and Analysis*, vol. 192, no. 1, pp. 165–186, 2009.
- [8] A. Constantin and J. Escher, “Wave breaking for nonlinear nonlocal shallow water equations,” *Acta Mathematica*, vol. 181, no. 2, pp. 229–243, 1998.
- [9] A. Constantin, “Existence of permanent and breaking waves for a shallow water equation: a geometric approach,” *Annales de l’Institut Fourier*, vol. 50, no. 2, pp. 321–362, 2000.
- [10] A. Bressan and A. Constantin, “Global conservative solutions of the Camassa-Holm equation,” *Archive for Rational Mechanics and Analysis*, vol. 183, no. 2, pp. 215–239, 2007.
- [11] H. Holden and X. Raynaud, “Global conservative solutions of the Camassa-Holm equation—a Lagrangian point of view,” *Communications in Partial Differential Equations*, vol. 32, no. 10–12, pp. 1511–1549, 2007.
- [12] H. Holden and X. Raynaud, “Periodic conservative solutions of the Camassa-Holm equation,” *Annales de l’Institut Fourier*, vol. 58, no. 3, pp. 945–988, 2008.
- [13] H. Holden and X. Raynaud, “Global conservative multipeakon solutions of the Camassa-Holm equation,” *Journal of Hyperbolic Differential Equations*, vol. 4, no. 1, pp. 39–64, 2007.
- [14] A. Bressan and A. Constantin, “Global dissipative solutions of the Camassa-Holm equation,” *Analysis and Applications*, vol. 5, no. 1, pp. 1–27, 2007.
- [15] L. Tian, Y. Wang, and J. Zhou, “Global conservative and dissipative solutions of a coupled Camassa-Holm equations,” *Journal of Mathematical Physics*, vol. 52, Article ID 063702, 2011.
- [16] Y. Wang and Y. Song, “On the global existence of dissipative solutions for the modified coupled Camassa-Holm system,” *Soft Computing*, vol. 17, no. 11, pp. 2007–2019, 2013.
- [17] Z. Shen, Y. Wang, H. R. Karimi, and Y. Song, “On the multi-peakon dissipative behavior of the modified coupled Camassa-Holm model for shallow water system,” *Mathematical Problems in Engineering*, vol. 2013, Article ID 107450, 11 pages, 2013.

Research Article

Objective Attributes Weights Determining Based on Shannon Information Entropy in Hesitant Fuzzy Multiple Attribute Decision Making

Yingjun Zhang,¹ Yizhi Wang,¹ and Jingping Wang²

¹ Beijing Key Lab of Traffic Data Analysis and Mining, School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

² School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China

Correspondence should be addressed to Yingjun Zhang; hitzj@163.com

Received 21 December 2013; Revised 18 March 2014; Accepted 18 March 2014; Published 7 April 2014

Academic Editor: Hamid Reza Karimi

Copyright © 2014 Yingjun Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Hesitant fuzzy set has been an important tool in dealing with multiple attribute decision making (MADM) problems, especially for the decision making situation when only some values of membership are possible for an alternative on attributes. However, determining attributes weights in hesitant fuzzy MADM is still an open problem. In this paper, we propose an objective weighting approach based on Shannon information entropy, which expresses the relative intensities of attribute importance to signify the average intrinsic information transmitted to the decision maker. Furthermore, we construct a hesitant fuzzy MADM approach based on the TOPSIS method and a weighted correlation coefficient proposed in this paper. Finally, we utilize a supplier selection example to validate the objective attributes weights determining method and the proposed hesitant fuzzy MADM approach.

1. Introduction

For a multiple attribute decision making (MADM) problem, an expert (decision maker) must evaluate every alternative to each attribute, assess attributes weights, and select the most desirable one from all the alternatives [1]. The MADM method provides a feasible and effective way to rank all the alternatives among noncommensurable and conflicting attributes. Some representative approaches, such as the technique for order preference by similarity to ideal solution (TOPSIS) [2], the simple additive weighted (SAW) method [3], the ordered weighted averaging (OWA) method [4], and the analytic hierarchy process (AHP) [5], have been successfully utilized in dealing with MADM problems. However, in most MADM situations, the preference of attributes over alternatives provided by decision makers is usually not sufficient for the crisp numerical data, because things are fuzzy, uncertain, and probably influenced by the subjectivity of the decision makers, or the knowledge and data about the problem domain are insufficient during the decision

making process [6]. To solve those MADM problems with uncertainty, research has extended the use of fuzzy set theory introduced by Zadeh to develop various MADM methods [7]. On the basis of Zadeh's fuzzy set, Torra and Narukawa [8, 9] introduced hesitant fuzzy set (HFS), which allows the membership of an element to a set derived from a few different values. HFS provides an effective way in dealing with decision making situation when only some values of membership are possible for an alternative on attributes. After the pioneering work of Torra and Narukawa, a fair proportion of literatures have done the theoretical and application research within the framework of HFS. Torra and Narukawa [8, 9] firstly generalized existing operations on fuzzy sets to HFS and discussed their application in decision making. Xu and Xia introduced distance, similarity, and correlation coefficient on HFS and utilized these operations to solve hesitant fuzzy MADM problems on the basis of TOPSIS method [10–12]. Xia and Xu [13] and Zhang [14] proposed various aggregation operators on HFS, which are effective in dealing with different decision making situations.

Chen et al. [15] presented a clustering method for HFSs based on correlation coefficient. Yu et al. [16] extended Choquet integral on HFS and discussed its application in uncertain decision making situations. Zhu et al. [17] took into account the interrelationships among arguments and proposed hesitant fuzzy geometric Bonferroni means in dealing with hesitant fuzzy MADM problems. Wei [18] presented hesitant fuzzy prioritized operators for solving the decision making problems with different priority levels' attributes. On the basis of HFS and interval-valued fuzzy set, generalized HFS and interval-valued HFS were proposed and utilized in decision making situations [19–21]. Rodriguez et al. [22] utilized hesitant fuzzy linguistic term set as an extension of HFS to solve uncertain qualitative decision making settings.

It is necessary to select appropriate attributes weights in decision making situations since the varied values of attributes weights may result in different ranking order of alternatives. Generally speaking, the attributes weights are divided into objective attributes and subjective attributes according to the ways of information acquisition [23]. The subjective attributes weights are obtained by preference information on the attributes given by the decision maker, who provides subjective intuition or judgments on specific attributes. AHP method [5] and Delphi method [24] are classical approaches for determining subjective attributes weights based on the preference of decision maker. Determining objective attributes weights depends on the decision making matrix. The existing objective attributes weights determining approaches include entropy-based method and optimization methods [25–27]. Most research pertaining to MADM analysis under hesitant fuzzy environment has been utilized depending on existing attributes weights [11–14, 18]. However, there is little research focusing on the problems of assessing objective attributes weights in hesitant fuzzy MADM.

In this paper, we present a new objective attributes weighting method based on Shannon information entropy in hesitant fuzzy MADM. The new objective attributes weighting method expresses the relative intensities of attribute importance to signify the average intrinsic information derived from decision maker, which has an emphasis on the discrimination among data to assess attributes weights. On the basis of known attributes weights, we construct a hesitant fuzzy MADM approach based on TOPSIS method and a weighted correlation coefficient proposed in this paper. The weighted correlation coefficient proposed in this paper takes into account the divergence among different elements on two HFSs. Therefore, the new correlation coefficient is helpful to reflect the attributes' importance in decision making situations.

The rest of this paper is organized as follows. Section 2 presents the concept of HFS and some of its basic operations. Section 3 presents a new weighted correlation coefficient within the framework of HFSs. Section 4 recalls Shannon information entropy and proposes a new objective attributes weighting method and a hesitant fuzzy MADM approach. Section 5 illustrates the proposed objective weighting method and the hesitant fuzzy MADM approach through a supplier selection example. Section 6 draws a conclusion.

2. HFS and Its Operations

In this section, we briefly recall HFS and some of its relevant operations.

2.1. HFS

Definition 1 (see [8, 9]). Let X be a reference set; a HFS on X is defined in terms of a function that when applied to X returns a subset of $[0, 1]$, which can be represented by the following mathematical symbol:

$$E = \{x, h_E(x) \mid x \in X\}, \quad (1)$$

where $h_E(x)$ is a set of values in $[0, 1]$, denoting the possible membership degrees of the element $x \in X$ to the set E .

For convenience, we call $h_E(x)$ a hesitant fuzzy element (HFE).

2.2. Operations on HFS

Definition 2 (see [9]). Given a HFS $h_E(x)$, Torra defined its lower and upper bounds as follows:

$$\begin{aligned} h_E^-(x) &= \min h_E(x), \\ h_E^+(x) &= \max h_E(x), \end{aligned} \quad (2)$$

where $h_E^-(x)$ and $h_E^+(x)$ denote the lower and upper bound of $h_E(x)$, respectively.

Obviously, $(x, h_E^-(x), 1 - h_E^+(x))$ is an intuitionistic fuzzy set.

Definition 3 (see [8]). Given a HFS $h_E(x)$, the full set of HFS is as follows: $h_E(x) = \{1\}$ for all $x \in X$.

The definition of HFS implies that the number of values in different HFEs may be different. Xu and Xia [11] introduced $l(h_E(x))$ to denote the number of values in $h_E(x)$. Assume the elements in $h_E(x)$ are in ascending order, and $h_E^{\sigma(j)}(x)$ is the j th largest value in $h_E(x)$.

Definition 4 (see [13]). For a HFE $h_E(x)$, the score function $s(h_E(x))$ is defined as follows:

$$s(h_E(x)) = \frac{\sum_{j=1}^{l(h_E(x))} h_E^{\sigma(j)}(x)}{l(h_E(x))}, \quad (3)$$

where $s(h_E(x)) \in [0, 1]$.

According to the score function on HFSs, Xia and Xu [13] introduced a method for ranking HFEs. For two HFEs h_1 and h_2 , if $s(h_1) > s(h_2)$, then $h_1 > h_2$; if $s(h_1) = s(h_2)$, then $h_1 = h_2$.

3. Correlation Coefficient on HFS

Correlation coefficient on fuzzy set plays an important role in both theoretical and application fields, such as fuzzy pattern recognition, fuzzy clustering, artificial intelligence, and uncertain decision making. In some situations [11],

the weights of each element $x_i \in X$ should be taken into account; we introduce a weighted correlation coefficient on HFS.

Definition 5. For two HFSs A and B on $X = \{x_1, x_2, \dots, x_n\}$, a weighted correlation coefficient $C(A, B)$ is defined as

$$C(A, B) = \frac{1}{2} (C_1(A, B) + C_2(A, B)),$$

$$C_1(A, B) = 1 - \left[\sum_{i=1}^n \left| \frac{w_i \bar{h}_A(x_i)}{\left(\sum_{i=1}^n (w_i \bar{h}_A(x_i))^\lambda\right)^{1/\lambda}} - \frac{w_i \bar{h}_B(x_i)}{\left(\sum_{i=1}^n (w_i \bar{h}_B(x_i))^\lambda\right)^{1/\lambda}} \right|^\lambda \right]^{1/\lambda},$$

$$C_2(A, B) = 1 - \left[\sum_{i=1}^n \left| \frac{w_i \tilde{h}_A(x_i)}{\left(\sum_{i=1}^n (w_i \tilde{h}_A(x_i))^\lambda\right)^{1/\lambda}} - \frac{w_i \tilde{h}_B(x_i)}{\left(\sum_{i=1}^n (w_i \tilde{h}_B(x_i))^\lambda\right)^{1/\lambda}} \right|^\lambda \right]^{1/\lambda}, \tag{4}$$

where $\bar{h}_A(x_i) = \sum_{j=1}^{l(h_A(x_i))} h_A^{\sigma(j)}(x_i) / l(h_A(x_i))$, $\bar{h}_B(x_i) = \sum_{j=1}^{l(h_B(x_i))} h_B^{\sigma(j)}(x_i) / l(h_B(x_i))$, $\tilde{h}_A(x_i) = \sum_{j=1}^{l(h_A(x_i))} h_A^{\sigma(j)}(x_i) / \sum_{i=1}^n \sum_{j=1}^{l(h_A(x_i))} h_A^{\sigma(j)}(x_i)$, $\tilde{h}_B(x_i) = \sum_{j=1}^{l(h_B(x_i))} h_B^{\sigma(j)}(x_i) / \sum_{i=1}^n \sum_{j=1}^{l(h_B(x_i))} h_B^{\sigma(j)}(x_i)$, $\lambda > 0$, $w_i \in [0, 1]$, and $\sum_{i=1}^n w_i = 1$.

It is clear that $C(A, B)$ satisfies the following three properties of correlation coefficient on HFS:

- (1) $|C(A, B)| \leq 1$,
- (2) if $A = B$, then $C(A, B) = 1$,
- (3) $C(A, B) = C(B, A)$.

Proof. Obviously, $C_1(A, B) \leq 1$ and $C_2(A, B) \leq 1$; $C(A, B) \leq 1$ holds. In the following part, we prove that $C(A, B) \geq -1$. Since

$$\left[\sum_{i=1}^n \left| \frac{w_i \bar{h}_A(x_i)}{\left(\sum_{i=1}^n (w_i \bar{h}_A(x_i))^\lambda\right)^{1/\lambda}} - \frac{w_i \bar{h}_B(x_i)}{\left(\sum_{i=1}^n (w_i \bar{h}_B(x_i))^\lambda\right)^{1/\lambda}} \right|^\lambda \right]^{1/\lambda}$$

$$\leq \sum_{i=1}^n \max \left\{ \frac{(w_i \bar{h}_A(x_i))^\lambda}{\sum_{i=1}^n (w_i \bar{h}_A(x_i))^\lambda}, \frac{(w_i \bar{h}_B(x_i))^\lambda}{\sum_{i=1}^n (w_i \bar{h}_B(x_i))^\lambda} \right\} \leq 2, \tag{5}$$

we get $C_1(A, B) \leq -1$. $|C_1(A, B)| \leq 1$ holds. The proof of $|C_2(A, B)| \leq 1$ is the same as the proof process of $|C_1(A, B)| \leq 1$. It implies that $|C(A, B)| \leq 1$ holds.

Obviously $C(A, B)$ satisfies properties (2). The proof is completed. \square

4. Hesitant Fuzzy MADM

In this section, we propose an approach based on the TOPSIS method and the weighted correlation coefficient proposed in this paper to solve the hesitant fuzzy MADM problem with unknown attributes weights; particular emphasis is put on determining objective attributes weights based on Shannon information entropy.

4.1. Hesitant Fuzzy MADM Problem Description. A MADM problem can be regarded as a decision matrix whose elements denote the evaluation information of all alternatives in relation to an attribute. A hesitant fuzzy MADM problem is defined as below.

Assume that there are m alternative measures, $A = \{A_1, A_2, \dots, A_m\}$, to be performed over n attributes, $X = \{x_1, x_2, \dots, x_n\}$. The hesitant fuzzy decision matrix D is expressed as follows:

$$D = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{m1} & h_{m2} & \cdots & h_{mn} \end{bmatrix}, \tag{6}$$

where $h_{ij} = h_{A_i}(x_j)$ ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$) denotes a HFE.

4.2. Shannon Information Entropy. Shannon information entropy quantifies the expected value of the information contained in a message. Entropy is typically measured in bits, nats, or bans. Shannon information entropy is the average unpredictability in a random variable, which is equivalent to its information content.

Definition 6. Let $X = \{x_1, x_2, \dots, x_n\}$ be a discrete random variable with probability mass function $P(X)$; the entropy H of X is defined as

$$E(X) = D[-\ln(P(X))], \tag{7}$$

where D is the expected value operator. When taken from a finite sample, the entropy can explicitly be written as

$$E(x) = -\sum_i P(x_i) \ln P(x_i). \tag{8}$$

In relation to $P(x_i) = 0$, the value of $0 \ln P(0)$ is taken to be 0.

4.3. Objective Attributes Weighting Method. We first define a score matrix of a hesitant fuzzy decision making matrix.

TABLE 1: Hesitant fuzzy decision matrix D .

	X_1	X_2	X_3	X_4
A_1	{0.5, 0.4, 0.3}	{0.9, 0.8, 0.7, 0.1}	{0.5, 0.4, 0.2}	{0.9, 0.6, 0.5, 0.3}
A_2	{0.5, 0.3}	{0.9, 0.7, 0.6, 0.5, 0.2}	{0.8, 0.6, 0.5, 0.1}	{0.7, 0.4, 0.3}
A_3	{0.7, 0.6}	{0.9, 0.6}	{0.7, 0.5, 0.3}	{0.6, 0.4}
A_4	{0.8, 0.7, 0.4, 0.3}	{0.7, 0.4, 0.2}	{0.8, 0.1}	{0.9, 0.8, 0.6}
A_5	{0.9, 0.7, 0.6, 0.3, 0.1}	{0.8, 0.7, 0.6, 0.4}	{0.9, 0.8, 0.7}	{0.9, 0.7, 0.6, 0.3}

Definition 7. Let D be a hesitant fuzzy decision making matrix as (6). One calls

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m1} & s_{m2} & \cdots & s_{mn} \end{bmatrix} \quad (9)$$

a score matrix of $D_{m \times n}$, where $s_{ij} = s(h_{ij})$ is the score value of h_{ij} .

In the following part, we present a new objective attributes weighting method as follows.

Step 1. Calculate the score matrix S of D .

Step 2. Normalize the score matrix S as follows:

$$\bar{S} = \begin{bmatrix} \bar{s}_{11} & \bar{s}_{12} & \cdots & \bar{s}_{1n} \\ \bar{s}_{21} & \bar{s}_{22} & \cdots & \bar{s}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{s}_{m1} & \bar{s}_{m2} & \cdots & \bar{s}_{mn} \end{bmatrix}, \quad (10)$$

where $\bar{s}_{ij} = s_{ij} / \sum_{i=1}^m s_{ij}$ ($i = 1, \dots, m; j = 1, \dots, n$).

Step 3. Determine the attributes weights.

Let

$$E_j = -\frac{1}{\ln m} \sum_{i=1}^m \bar{s}_{ij} \ln \bar{s}_{ij}, \quad j = 1, \dots, n. \quad (11)$$

The attribute weight w_j ($j = 1, \dots, n$) is defined by

$$w_j = \frac{1 - E_j}{\sum_{j=1}^n (1 - E_j)}. \quad (12)$$

The above objective attributes weighting method utilizes Shannon information entropy to express the relative intensities of attribute importance and the divergence among attributes. And, then, the attributes weights are determined through (12).

4.4. Hesitant Fuzzy MADM Approach. With respect to the hesitant fuzzy MADM problem in Section 4.1, we present a hesitant fuzzy MADM approach based on TOPSIS method and the weighted correlation coefficient on HFS defined in Section 3. The schematic structure of the proposed hesitant

fuzzy MADM approach is shown in Figure 1, and the detailed decision steps of this approach are listed as below.

Step 1. On the basis of Definitions 2 and 3, we define the hesitant fuzzy positive solution $A^+ = \{\sigma_1^+, \sigma_2^+, \dots, \sigma_n^+\}$, where $\sigma_j^+ = \{1\}$ ($j = 1, 2, \dots, n$).

Step 2. Determine the objective attributes weights $w = [w_1, w_2, \dots, w_n]$ based on the newly objective weighting method in Section 4.3.

Step 3. Calculate the weighted correlation coefficient $C(A_i, A^+)$ ($i = 1, 2, \dots, m$) using (4).

Step 4. Rank all the alternatives A_i based on $C(A_i, A^+)$ ($i = 1, 2, \dots, m$) and select the most desirable one.

5. Illustrative Example and Discussion

In this section, we utilize a supplier selection example to illustrate the proposed method for determining objective attributes weights in a MADM problem under hesitant fuzzy environment.

5.1. Illustrative Example. Suppose the supplier selection problem refers to 5 possible alternatives $A = \{A_1, A_2, A_3, A_4, A_5\}$ on 4 attributes $X = \{x_1, x_2, x_3, x_4\}$. The attributes weights for these problems are completely unknown. The hesitant fuzzy decision making matrix D of A on X is shown in Table 1.

Step 1. Using Definitions 2 and 3, the hesitant fuzzy positive solution is defined as

$$A^+ = \{\{1\}, \{1\}, \{1\}, \{1\}\}. \quad (13)$$

Step 2. Based on the new objective attributes weighting method, the process of determining attributes weights is as follows.

Firstly, calculate the score matrix of D based on Definition 7:

$$S = \begin{bmatrix} 0.4000 & 0.6250 & 0.3667 & 0.5750 \\ 0.4000 & 0.5800 & 0.5000 & 0.4667 \\ 0.6500 & 0.7500 & 0.5000 & 0.5000 \\ 0.5500 & 0.4333 & 0.4500 & 0.7667 \\ 0.5200 & 0.6250 & 0.8000 & 0.6250 \end{bmatrix}. \quad (14)$$

TABLE 2: The correlation coefficients and ranking results of all the alternatives.

	$C(A_1, A)$	$C(A_2, A)$	$C(A_3, A)$	$C(A_4, A)$	$C(A_5, A)$	Ranking results
$C _{\lambda=1}$	0.7050	0.7981	0.8599	0.7100	0.9093	$A_5 > A_3 > A_2 > A_4 > A_1$
$C _{\lambda=2}$	0.7095	0.8052	0.8633	0.6885	0.9095	$A_5 > A_3 > A_2 > A_1 > A_4$

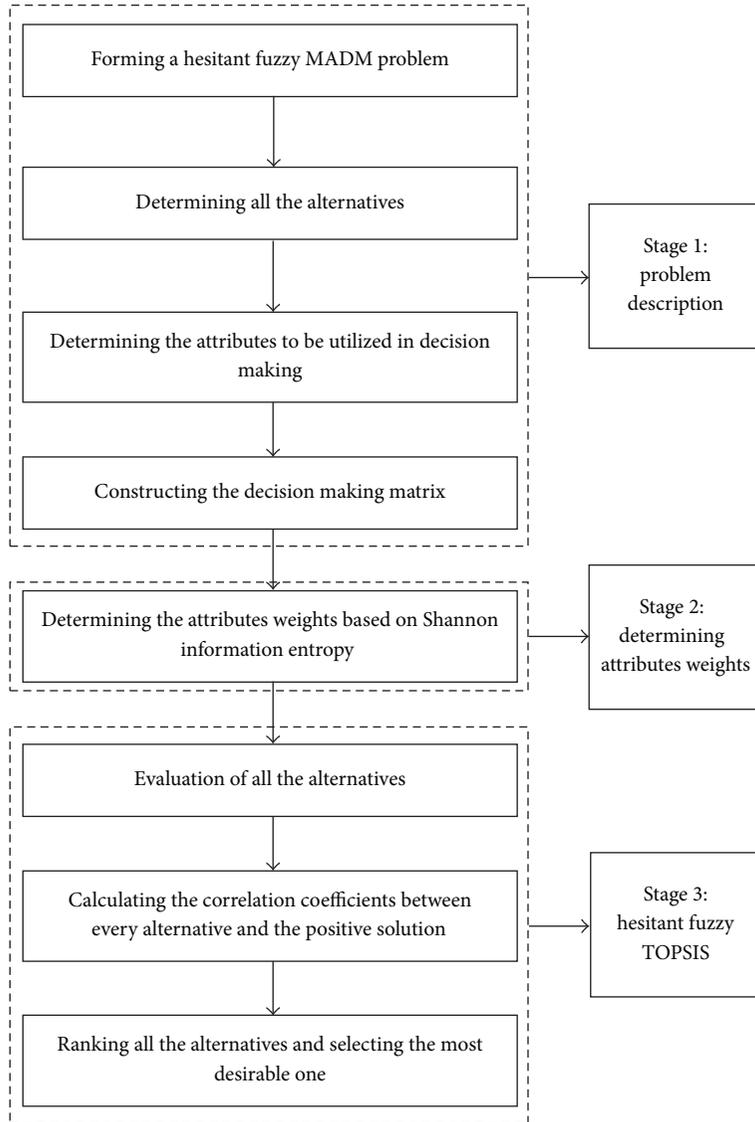


FIGURE 1: The schematic structure of the proposed hesitant fuzzy MADM approach.

Secondly, we get the the normalized score matrix \bar{S} based on (10):

$$\bar{S} = \begin{bmatrix} 0.1587 & 0.2074 & 0.1401 & 0.1960 \\ 0.1587 & 0.1925 & 0.1911 & 0.1591 \\ 0.2579 & 0.2489 & 0.1911 & 0.1705 \\ 0.2183 & 0.1438 & 0.1720 & 0.2614 \\ 0.2063 & 0.2074 & 0.3057 & 0.2131 \end{bmatrix}. \quad (15)$$

Thirdly, the attribute weight vector is determined by (11) and (12):

$$w = [0.2094 \ 0.1735 \ 0.4304 \ 0.1867]^T. \quad (16)$$

Step 3. Calculate the correlation coefficient $C(A_i, A^+)$ ($i = 1, 2, 3, 4$) based on $C(\lambda = 1, 2)$ (4), and the results are shown as in Table 2.

TABLE 3: The cross-entropy and ranking order of all the alternatives.

	A_1	A_2	A_3	A_4	A_5	Ranking results
C_r^+	0.3080	0.2988	0.1990	0.2949	0.1583	$A_5 > A_3 > A_4 > A_2 > A_1$
C_r^-	0.3233	0.3083	0.3936	0.3704	0.4826	$A_5 > A_3 > A_4 > A_1 > A_2$
C_r	0.4879	0.4921	0.3358	0.4433	0.2470	$A_5 > A_3 > A_4 > A_1 > A_2$

Step 4. Rank all the alternatives based on the obtained correlation coefficient values, and the ranking results are shown as in Table 2.

5.2. Comparison with Existing Methods. In [12], Xu and Xia proposed a method based on hesitant fuzzy entropy and cross-entropy (*Method Xu*) to deal with the hesitant fuzzy MADM problem with completely unknown attribute weight information. In the following part, we utilize *Method Xu* in an example in Section 5.1.

Step 1. Determine the attributes weights based on the weighting method of *Method Xu*.

Firstly, calculate the hesitant fuzzy entropy matrix according to (44) with $q = 2$ in [12]; we get the entropy matrix E of D :

$$E = \begin{bmatrix} 0.9615 & 0.8790 & 0.9283 & 0.9742 \\ 0.9613 & 0.9725 & 0.9898 & 0.9872 \\ 0.9135 & 0.7527 & 1.0000 & 1.0000 \\ 0.9900 & 0.9804 & 0.9891 & 0.7187 \\ 0.9923 & 0.9368 & 0.6481 & 0.9357 \end{bmatrix}. \quad (17)$$

Secondly, calculate the attributes weights based on (53) in [12]:

$$w = [0.1218 \ 0.3214 \ 0.2987 \ 0.2580]^T. \quad (18)$$

Step 2. Based on the hesitant fuzzy cross-entropy (see (43) with $q = 2$ in [12]), calculate the positive cross-entropy $C_r^+(A_i)$ ($i = 1, \dots, 5$), the negative cross-entropy $C_r^-(A_i)$ ($i = 1, \dots, 5$), and the closeness degree $C_r(A_i)$ ($i = 1, \dots, 5$) of the alternative A_i ($i = 1, \dots, 5$) to the ideal solution by (54)–(56) in [12] as in Table 3.

Step 3. Rank all the alternatives based on $C_r(A_i)$ ($i = 1, \dots, 5$) as in Table 3.

5.3. Discussion. Both *Method Xu* and the newly hesitant fuzzy MADM approach proposed in this paper choose A_5 as the most desirable one for the decision making example in Section 5.1. The results show that proposed MADM approach is effective for addressing hesitant fuzzy MADM problem with completely unknown attribute weight information. Instead of the decision making method in [12], which utilizes hesitant fuzzy entropy to determine attributes weights depending on the credibility of the input data, we employ Shannon information entropy to express the relative intensities of attribute importance and determine the objective attributes weights.

6. Conclusion

By applying hesitant fuzzy sets to uncertain MADM problems, we can get more accurate choice from the incomplete and complex information derived from decision makers in real life. In this paper, we propose a new objective attributes weighting method based on Shannon information entropy to express the relative intensities of attribute importance and determine the objective attributes weights. Furthermore, we propose a MADM approach based on the new objective weighting method and a weighted correlation coefficient introduced in this paper. The results of the example in Section 5.1 indicate that the newly MADM approach is feasible and effective in dealing with hesitant fuzzy MADM problems with completely unknown attributes weights. The newly hesitant fuzzy MADM approach offers a useful way to solve uncertain decision making problems derived from supplier selection, public risk, medical diagnosis, and other aspects.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors are very grateful to Professor Hamid Reza Karimi and two anonymous referees for their constructive comments and suggestions that have led to an improved version of this paper. This research was supported by Project funded by China Postdoctoral Science Foundation under Grant no. 2013M540848 and the Fundamental Research Funds for the Central Universities under Grant no. 2013JBM023.

References

- [1] W. Ho, X. Xu, and P. K. Dey, "Multi-criteria decision making approaches for supplier evaluation and selection: a literature review," *European Journal of Operational Research*, vol. 202, no. 1, pp. 16–24, 2010.
- [2] M. Behzadian, S. K. Otaghsara, M. Yazdani, and J. Ignatius, "A state-of-the-art survey of TOPSIS applications," *Expert Systems with Applications*, vol. 39, no. 17, pp. 13051–13069, 2012.
- [3] Y. S. Huang, W. C. Chang, W. H. Li, and Z. L. Lin, "Aggregation of utility-based individual preferences for group decision-making," *European Journal of Operational Research*, vol. 229, no. 2, pp. 462–469, 2013.
- [4] R. R. Yager, "OWA aggregation over a continuous interval argument with applications to decision making," *IEEE Transactions*

- on Systems, Man, and Cybernetics B: Cybernetics*, vol. 34, no. 5, pp. 1952–1963, 2004.
- [5] R. J. Ormerod and W. Ulrich, “Operational research and ethics: a literature review,” *European Journal of Operational Research*, vol. 228, no. 2, pp. 291–307, 2013.
- [6] I. N. Durbach and T. J. Stewart, “Modeling uncertainty in multi-criteria decision analysis,” *European Journal of Operational Research*, vol. 223, no. 1, pp. 1–14, 2012.
- [7] R. A. Ribeiro, “Fuzzy multiple attribute decision making: a review and new preference elicitation techniques,” *Fuzzy Sets and Systems*, vol. 78, no. 2, pp. 155–181, 1996.
- [8] V. Torra and Y. Narukawa, “On hesitant fuzzy sets and decision,” in *Proceedings of the IEEE International Conference on Fuzzy Systems*, pp. 1378–1382, Jeju Island, Republic of Korea, August 2009.
- [9] V. Torra, “Hesitant fuzzy sets,” *International Journal of Intelligent Systems*, vol. 25, no. 6, pp. 529–539, 2010.
- [10] Z. Xu and M. Xia, “On distance and correlation measures of hesitant fuzzy information,” *International Journal of Intelligent Systems*, vol. 26, no. 5, pp. 410–425, 2011.
- [11] Z. Xu and M. Xia, “Distance and similarity measures for hesitant fuzzy sets,” *Information Sciences*, vol. 181, no. 11, pp. 2128–2138, 2011.
- [12] Z. Xu and M. Xia, “Hesitant fuzzy entropy and cross-entropy and their use in multiattribute decision-making,” *International Journal of Intelligent Systems*, vol. 27, pp. 799–822, 2012.
- [13] M. Xia and Z. Xu, “Hesitant fuzzy information aggregation in decision making,” *International Journal of Approximate Reasoning*, vol. 52, no. 3, pp. 395–407, 2011.
- [14] Z. Zhang, “Hesitant fuzzy power aggregation operators and their application to multiple attribute group decision making,” *Information Sciences*, vol. 234, pp. 150–181, 2013.
- [15] N. Chen, Z. Xu, and M. Xia, “Correlation coefficients of hesitant fuzzy sets and their applications to clustering analysis,” *Applied Mathematical Modelling*, vol. 37, no. 4, pp. 2197–2211, 2013.
- [16] D. Yu, W. Zhang, and Y. Xu, “Group decision making under hesitant fuzzy environment with application to personnel evaluation,” *Knowledge-Based Systems*, vol. 52, pp. 1–10, 2013.
- [17] B. Zhu, Z. Xu, and M. Xia, “Hesitant fuzzy geometric Bonferroni means,” *Information Sciences*, vol. 205, pp. 72–85, 2012.
- [18] G. Wei, “Hesitant fuzzy prioritized operators and their application to multiple attribute decision making,” *Knowledge-Based Systems*, vol. 31, pp. 176–182, 2012.
- [19] B. Farhadinia, “Information measures for hesitant fuzzy sets and interval-valued hesitant fuzzy sets,” *Information Sciences*, vol. 240, pp. 129–144, 2013.
- [20] G. Qian, H. Wang, and X. Feng, “Generalized hesitant fuzzy sets and their application in decision support system,” *Knowledge-Based Systems*, vol. 37, pp. 357–365, 2013.
- [21] G. Wei, X. Zhao, and R. Lin, “Some hesitant interval-valued fuzzy aggregation operators and their applications to multiple attribute decision making,” *Knowledge-Based Systems*, vol. 46, pp. 43–53, 2013.
- [22] R. M. Rodriguez, L. Martinez, and F. Herrera, “Hesitant fuzzy linguistic term sets for decision making,” *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 1, pp. 109–119, 2012.
- [23] E. K. Zavadskas, Z. Turskis, L. Ustinovichius, and G. Shevchenko, “Attributes weights determining peculiarities in multiple attribute decision making methods,” *Engineering Economics*, no. 1, pp. 32–43, 2010.
- [24] C. L. Hwang and M. J. Lin, *Group Decision Making Under Multiple Criteria: Methods and Applications*, Springer, Berlin, Germany, 1987.
- [25] T. Y. Chen and C. H. Li, “Determining objective weights with intuitionistic fuzzy entropy measures: a comparative analysis,” *Information Sciences*, vol. 180, no. 21, pp. 4207–4222, 2010.
- [26] D. F. Li, “TOPSIS-based nonlinear-programming methodology for multiattribute decision making with interval-valued intuitionistic fuzzy sets,” *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 2, pp. 299–311, 2010.
- [27] Y. Zhang, P. Li, Y. Wang, P. Ma, and X. Su, “Multiattribute decision making based on entropy under interval-valued intuitionistic fuzzy environment,” *Mathematical Problems in Engineering*, vol. 2013, Article ID 526871, 8 pages, 2013.

Research Article

Fault Diagnosis for Compensating Capacitors of Jointless Track Circuit Based on Dynamic Time Warping

Wei Dong^{1,2}

¹ Tsinghua National Laboratory for Information Science and Technology, Beijing 100084, China

² Department of Automation, Tsinghua University, Beijing 100084, China

Correspondence should be addressed to Wei Dong; weidong@tsinghua.edu.cn

Received 31 January 2014; Accepted 17 February 2014; Published 25 March 2014

Academic Editor: Hamid R. Karimi

Copyright © 2014 Wei Dong. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aiming at the problem of online fault diagnosis for compensating capacitors of jointless track circuit, a dynamic time warping (DTW) based diagnosis method is proposed in this paper. Different from the existing related works, this method only uses the ground indoor monitoring signals of track circuit to locate the faulty compensating capacitor, not depending on the shunt current of inspection train, which is an indispensable condition for existing methods. So, it can be used for online diagnosis of compensating capacitor, which has not yet been realized by existing methods. To overcome the key problem that track circuit cannot obtain the precise position of the train, the DTW method is used for the first time in this situation to recover the function relationship between receiver's peak voltage and shunt position. The necessity, thinking, and procedure of the method are described in detail. Besides the classical DTW based method, two improved methods for improving classification quality and reducing computation complexity are proposed. Finally, the diagnosis experiments based on the simulation model of track circuit show the effectiveness of the proposed methods.

1. Introduction

Track circuit is one of the most important basic equipment in train control system. It is used for detection of train's occupancy on the track, train-ground communication, and detection of broken rails, and its functions are crucial in ensuring safe operation of trains. In track circuit, the rails are taken as parts of the working circuit and the current flows from transmitter to receiver through the rails (Figure 1). If some train is located between the transmitter and the receiver (this interval is called "block section"), then the circuit will be shorted by the wheels and axles of the train (Figure 2). So a remarkable drop in the receiver's voltage will indicate a train's occupancy. In this way, the train control system can obtain the information which block sections are occupied and which block sections are empty. Based on this information, the train control system can release movement authorities to each train in order to avoid collisions between trains. If the track circuits have some faults, the most serious consequence is to cause collisions of trains, resulting in catastrophes. Even if the fail-safety mechanism plays its role and no catastrophic

accident happens, the transportation efficiency of railway will be affected seriously, causing considerable economic loss. So the timely fault diagnosis for track circuit is significant to both safety and efficiency of the railway transportation system.

In the practical application, as track circuits must be laid along the rails and the length of one track circuit is about 1~2 kilometers, the application scope of track circuits is quite large. And due to the special structure and complex working environment, track circuits are easily affected by temperature, humidity, ballast resistance, electromagnetic interference, and mechanical vibration, which results in a high fault rate of track circuit [1-4]. As a result of the above two aspects, the loss caused by faults of track circuits is very huge. For example, according to the report from the official website of the Ministry of Railways of China, up to the end of 2012, the number of track circuits laid along Chinese railways reaches about 600 thousand, and the equipment assets of track circuits reach about 60 billion RMB. And according to the statistical data released by Chinese railway department, there were about 8 thousand faults happening in the signal and communication systems of Chinese railways per year.

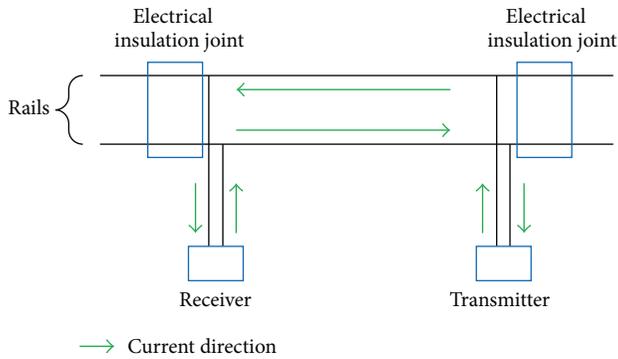


FIGURE 1: Basic structure of track circuit (empty).

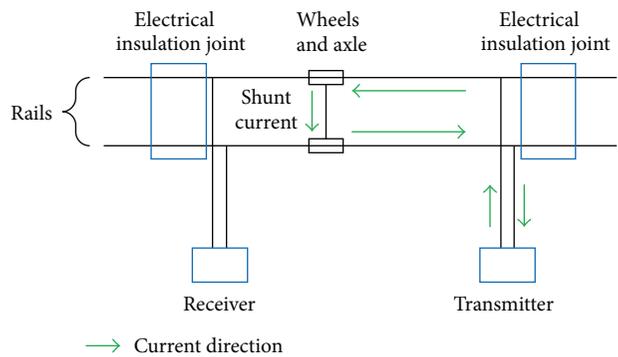


FIGURE 2: Basic structure of track circuit (occupied).

The majority of these faults were caused directly or indirectly by track circuits, resulting in a total traffic delay time of about 4 hundred thousand minutes. In other words, each fault caused an average traffic delay of about 50 minutes [5].

Among different fault types of track circuit, compensating capacitor's fault is one of the most important fault types. Compensating capacitors are used to compensate the inductive impedance of the rails, so that the current signal can be transmitted along the rails far enough. To this end, compensating capacitors are installed at equal intervals along the rails, and the installation interval is about 80~100 meters (Figure 3). Due to influence of bad weather and mechanical vibration and so on, compensating capacitors are prone to be faulty. Once some compensating capacitors are faulty, the signal transmitted along the rails will be attenuated rapidly, which may result in red light faults. When a track circuit has this kind of faults, the track circuit will declare that it is occupied (i.e., in red light state) when there is no train on the track in fact. In this case, to find the faulty compensating capacitors, maintenance personnel have to check the compensating capacitors one by one along the very long rails, which not only consumes lots of human resources but also causes a long delay in train's operation. For example, in the spring of 2009, the large-range faults of the compensating capacitors in Chinese Beijing-Tianjin intercity high-speed railway were caused by snowstorm. In order to repair the faults as soon as possible, on the condition of lacking related diagnosis technologies, the railway department was forced

to replace all the compensating capacitors within 2 days, resulting in a huge consumption of human and material resources.

Despite the importance, the problem of fault diagnosis for track circuit has not been solved very well and there are still many difficulties. In industrial community, up to now, the most commonly adopted methods for fault diagnosis of track circuit are periodic maintenance and field inspection [4], resulting in low working efficiency, high working strength, and long repair time. In academic community, a lot of work has been done. Chen et al. [3, 4] and Huang et al. [6] gave fault diagnosis methods for track circuit based on neurofuzzy systems, which can detect and diagnose the most common faults of track circuit. However, due to lack of dynamic diagnostic information, this method cannot be used to diagnose compensating capacitor's faults. Aiming at the features of track circuit's multiple fault modes and uncertain and nonaccurate fault symptom information, Oukhellou et al. [7] gave an information fusion method for fault diagnosis of compensating capacitor based on Dempster-Shafer classifier. However, this method can only be used by inspection train, resulting in that it cannot be used for online diagnosis.

Similar to [7], most existing methods for fault diagnosis of compensating capacitors in track circuit are based on the shunt current ("shunt current" is the current flowing through the wheels and axle of a train when the track circuit is occupied (see Figure 2)) of inspection train. For example, Côme et al. [8, 9] proposed noiseless independent factor analysis methods for fault diagnosis of compensating capacitors in track circuit. Zhao et al. [10] designed a kind of onboard autotest system to detect the faults of track circuit's compensating capacitors in real time. Zhao et al. [11] used the induced voltage recorded by cab signal as fault feature and gave a fault diagnosis method for track circuit based on discrete binary wavelet transform (DBWT) and wavelet ridge (WR), finally detecting the changes of instantaneous frequency to diagnose the disconnection fault of compensating capacitor. Zhao et al. [12] used the transmission line theory to construct a track circuit model with cascade form of multiple four terminal networks. Based on the constructed model, the typical fault modes of compensating capacitor were analyzed through the fault features of track voltage, shunt current, and input impedance of main track circuit. Linhai et al. [13] proposed a wavelet analysis based method to detect the integrity of the compensating capacitors of UM71 track circuit, by making use of the recorded data in the cab signal recorder. Zhao and Mu [14] analyzed induced voltage envelope of cab signal (IVECS) of track circuit by simulation model under different faulty conditions and proposed a compensating capacitor diagnosis method based on adaptive optimal kernel time-frequency representation (AOK-TFR). Zhao et al. [15] adopted B-spline DBWT and improved Hilbert-Huang transformation for noise reduction and fault feature retrieve and proposed a diagnosis method for compensating capacitor based on cab signal recorder information. Zhao et al. [16] proposed a diagnosis method for compensating capacitor based on the regression model of the shunt current. This method adopted Levenberg-Marquardt

(L-M) algorithm to verify the model and generalized S-transform (GST) to compute the instantaneous frequency so as to locate the faulty capacitors. Zhao et al. [17] analyzed the influence law of compensating capacitor's fault on the induced voltage envelope of the cab signal and proposed a comprehensive fault diagnosis method based on genetic algorithm to locate faulty compensating capacitors. Lin-Hai et al. [18] presented a collaborative fault diagnosis system for compensating capacitors in track circuit using AOK-TFR and adaptive genetic algorithm (AGA) based on the cab signal. Sun et al. [19] presented a diagnosis approach for compensating capacitors based on empirical mode decomposition (EMD) and Teager energy operator (TEO) theory, which can detect multiple capacitor faults.

The common feature of the above research on fault diagnosis of compensating capacitor is that they must depend on the shunt current of inspection train. The problems of these methods can be summarized as follows.

- (1) For a specific railway line, the inspection train has a specific running period and can only be used when the railway line is idle, so these methods cannot be used for online diagnosis. Online diagnosis of compensating capacitor is important, because it can help to find the faulty compensating capacitors in time so as to avoid traffic delay. Usually, one faulty compensating capacitor may not result in failure of track circuit, but more faulty compensating capacitors may result in failure of track circuit. So finding faulty compensating capacitors in time is important for avoiding failure of track circuit.
- (2) If there is no inspection train, to use these methods, some shunt current detecting and fault diagnosis devices must be installed on common trains, which will result in huge cost input and huge workload for modifications of equipment. Besides, considering safety, modifications of on-board equipment are rigorously restricted by the management department, so it is hard to install these devices.
- (3) Since the shunt current must be detected through electromagnetic induction, the detected results are easily interfered by noise.
- (4) These methods take the function relationship between shunt current and shunt position as the diagnostic basis, so the position of the train's first wheel-set must be obtained accurately. But in engineering practice, the precise position of the train is not easy to obtain by track circuit, resulting in that the diagnosis results are easily affected by position error.

To overcome the above problems, a dynamic time warping (DTW) based fault diagnosis method for compensating capacitors in track circuit is proposed in this paper. In order to realize online diagnosis, this method does not depend on the shunt current of inspection train and only collects the voltage signals in receiver of track circuit. However, static voltage information is insufficient to locate the faulty capacitor. In order to promote the diagnosis resolution, this method uses

the dynamic voltage information which is varying when a train is passing the track circuit (in this situation, the structure of track circuit is changing (see Figure 2), so more information can be obtained). Different from the existing methods, this method only needs the function relationship between receiver's peak voltage and shunt time (instead of shunt position) as the diagnostic basis, and the DTW method is used in this situation to recover the original function relationship between receiver's peak voltage and shunt position. Finally, by DTW, the characteristic curve is compared to the sample curves of different faults to locate the faulty capacitor. The contributions and advantages of the method proposed in this paper can be listed as follows.

- (1) This method only uses the ground monitoring signals of track circuit to locate the faulty compensating capacitor, not depending on inspection train. So, this method can be used for online diagnosis of compensating capacitor, which is not yet realized by existing methods.
- (2) This method does not need the outdoor monitoring signals (see Figure 3) or the on-board monitoring signals, so the existing system does not need to be modified greatly, which means a good engineering feasibility.
- (3) Since the precise train position is very hard to obtain based on the existing monitoring system of track circuit (the position information of a train is easily obtained by on-board devices instead of ground devices. However, on-board devices are not suitable for online diagnosis of track circuit), this method does not need the position of the train and adopts DTW to recover the function relationship between receiver's peak voltage and shunt position. So, this method overcomes the problem that the existing methods must depend on the precise position of the train.

This paper is organized as follows. Section 1 gives an overall background and introduces the problem. Section 2 gives a description of track circuit structure and track circuit model. Section 3 gives the fault diagnosis method for track circuit based on dynamic time warping. Section 4 gives the results and discussion. Finally, Section 5 gives the conclusion.

2. Track Circuit and Track Circuit Model

2.1. Track Circuit Structure. The structure of a track circuit is shown in Figure 3. A track circuit consists of four parts: transmitter part, receiver part, main track circuit, and small track circuit. Transmitter part, which is used for signal transmitting, consists of transmitter, lightning protector, cable simulator, SPT ("SPT" is just a code for digital signal cable in Chinese railway. "SP" means "digital signal cable" and "T" means "railway") cable, and matching transformer. Transmitter is used to generate a frequency-shift modulated signal with high stability and precision. Lightning protector is used to protect against lightning impulses which are introduced indoors by the cable. SPT cable is used to transmit signals from indoor transmitter to outdoor steel rails, and

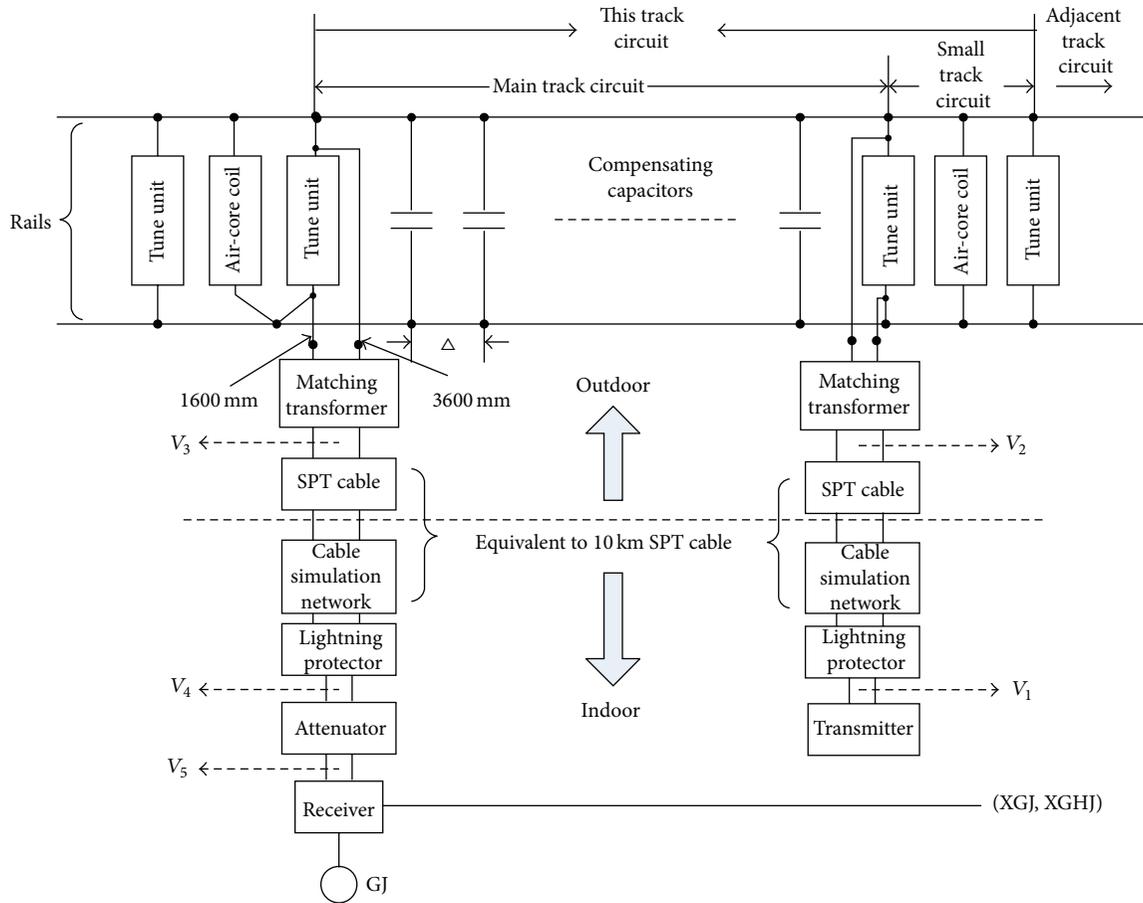


FIGURE 3: Structure of track circuit.

cable simulator is used to compensate the actual SPT cable so that the total equivalent length of SPT cable is equal to 10 km. Matching transformer is used to couple the frequency-shift modulated signal into steel rails or collect the frequency-shift modulated signal from steel rails. Receiver part, which is used for signal receiving, consists of receiver, attenuator, lightning protector, cable simulator, SPT cable, and matching transformer. Except for attenuator, which is used to attenuate the signal for receiving, the others are the same as the transmitter part. Main track circuit, which is the main body of the current loop, consists of two steel rails and some compensating capacitors connected between the two rails. Small track circuit (also called “electrical insulation joint”) consists of tune units, air-core coil, and 29 meters long steel rails. These circuit components constitute different kinds of resonant circuits and are used to separate two signals of different frequencies in different track sections.

Different track circuits have different working frequencies. There are totally four working frequencies for track circuits: 1700 Hz, 2000 Hz, 2300 Hz, and 2600 Hz. The distribution of these four working frequencies is shown in Figure 4. By this distribution, the interference between two adjoining track circuits can be reduced effectively.

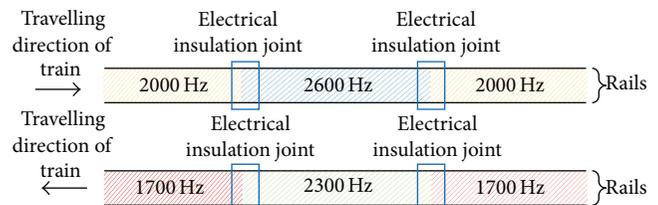


FIGURE 4: Frequency distribution of track circuits.

2.2. *Track Circuit Model.* The track circuit model is used to generate sufficient data for research, so as to make up for the deficiency of data from actual devices. The “SimPowerSystems” tool box in Matlab is used to construct the track circuit model, and receiver’s dynamic peak voltages are calculated by adjust the structure and parameters of the model automatically. The main parameters of the track circuit model are shown in Table 1.

For reasonable simplification of the model, the following assumptions are made.

- (1) Due to the frequency-shift modulation mechanism, the signal on track circuit is not a strict sine wave. However, since the frequency offset (11 Hz) is very

TABLE 1: Main parameters of track circuit model.

Parameter name	Parameter value
Working frequency	2000 Hz
Peak voltage in transmitter	142 V
Capacitance of compensating capacitor	5×10^{-5} F
Shunt resistance	0.15 Ω
Length	1301 m
The number of compensating capacitors	16
Contact resistance of compensating capacitor	1×10^{-4} Ω

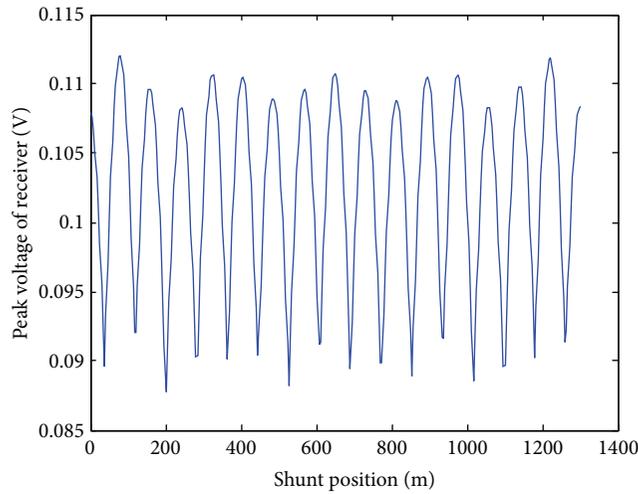


FIGURE 5: Peak voltages of receiver under different shunt positions (when track circuit has no faults, without noise).

small relative to the working frequency (2000 Hz), the track circuit can be considered as a sine steady circuit within an acceptable error range.

- (2) Because the transient signal component decays very quickly when the train is shunting on the rails [20], this paper only considers steady states of track circuit. And the peak voltages of receiver (V_5 , shown in Figure 3) under different shunt positions are picked as fault features (Figure 5).
- (3) Because the shunt resistance is very small (0.04–0.15 Ω), the electrical effect of multiple wheel-sets shunting on the rails is almost the same as the situation of one wheel set [21]. So, this paper only considers one wheel set's shunt on the rails and the shunt resistance is supposed to be 0.15 Ω (for smaller shunt resistance, transmitter's current instead of receiver's voltage can be selected as fault feature, and the diagnosis process is all the same).

2.3. Verification and Validation of Track Circuit Model. In order to verify and validate the track circuit model, a physical model of track circuit has been set up (shown in Figure 6). This physical model consists of completely real

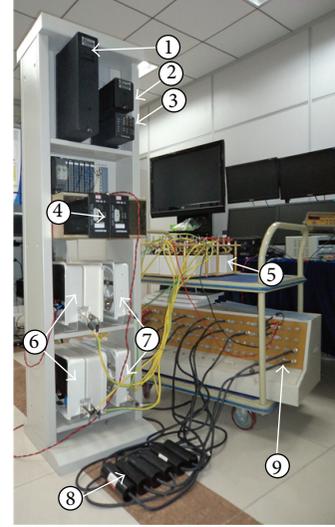


FIGURE 6: The physical model of ZPW-2000A track circuit. ① Transmitter; ② receiver; ③ attenuator; ④ lightning protector and cable simulator; ⑤ small track simulator; ⑥ tune unit; ⑦ air core coil; ⑧ compensating capacitor; ⑨ main track simulator.

TABLE 2: The verification and validation results.

	V_1	V_2	V_3	V_4
Simulation model	142.31 (V)	41.58 (V)	13.46 (V)	1.09 (V)
Physical model	142.56 (V)	41.44 (V)	13.39 (V)	1.06 (V)
Absolute error	-0.25 (V)	0.14 (V)	0.07 (V)	0.03 (V)
Relative error	-0.18%	0.34%	0.52%	2.83%

track circuit devices except for the steel rails, which cannot be installed indoors due to its huge size. The steel rails in main track circuit and small track circuit are replaced by the corresponding track simulators, which have been carefully calibrated according to real steel rails.

The verification and validation scheme is as follows. Four typical peak voltages (V_1, V_2, V_3, V_4 , shown in Figure 3) of track circuit are selected as features for comparison. One set of voltages are obtained from the simulation model, and the other set of voltages are obtained by averaging 100 sets of measured values from the physical model. The comparative results are shown in Table 2, which shows that the deviation between the simulation model and the physical model is very small.

3. Fault Diagnosis Method for Track Circuit Based on Dynamic Time Warping

3.1. Problem Formulation

3.1.1. Targets and Constraints. According to the introduction section, the following targets or constraints should be achieved or met.

- (1) Online fault diagnosis for compensating capacitors of track circuit should be realized. Here, only single disconnection fault of compensating capacitor is

TABLE 3: Fault state list of track circuit.

State symbol	Meaning
F_0	No fault
F_i ($i = 1, 2, 3, \dots, 16$)	Number i (numbered from transmitter to receiver) compensating capacitor disconnects from the rails

TABLE 4: Peak voltage of receiver under different fault states.

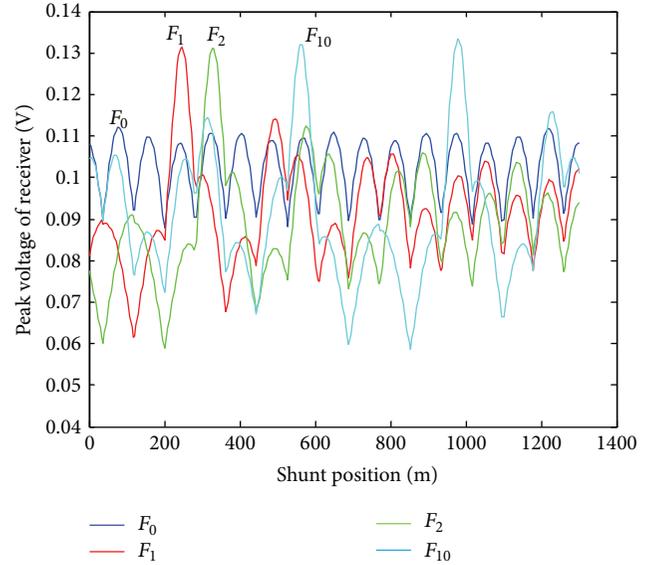
Fault state	Peak voltage of receiver (V)
F_0	0.5917
F_1	0.5343
F_2	0.5257
F_3	0.5246
F_4	0.5313
F_5	0.5276
F_6	0.5258
F_7	0.5299
F_8	0.5275
F_9	0.5272
F_{10}	0.5292
F_{11}	0.5263
F_{12}	0.5275
F_{13}	0.5309
F_{14}	0.5251
F_{15}	0.5258
F_{16}	0.5341

considered, because disconnection fault is the most possible fault type of compensating capacitor and two compensating capacitors are almost impossible to be faulty synchronously.

- (2) Only the indoor voltages or currents of track circuit (see Figure 3) can be obtained for fault diagnosis. Because the outdoor voltages or currents are hard to be measured under existing conditions, and the shunt current cannot be used for online diagnosis, as discussed in Section 1.
- (3) Under existing conditions, a train's exact position information cannot be obtained by track circuit or its monitoring system. So the shunt position information cannot be used for fault diagnosis.

Based on the above targets and constraints, the problem to be solved is to distinguish 17 kinds of states (listed in Table 3) of track circuit by specific fault features. Without loss of generality, the peak voltage of receiver is picked as fault feature.

3.1.2. Insufficiency of Static Voltage Information. (Here, "Static" means there is no train passing the track circuit, so the peak voltage is invariable.) Under different fault states, the peak voltages of receiver (V_5 in Figure 3) are shown in Table 4.

FIGURE 7: Peak voltages of receiver under different shunt positions (for fault states F_0 , F_1 , F_2 , and F_{10} , without noise).

It can be seen from Table 4 that if only static peak voltage is obtained, F_8 and F_{12} , as well as F_6 and F_{15} , cannot be distinguished at all. Besides, if a measurement noise of ± 0.005 V is considered, there are 120 pairs of indistinguishable faults (i.e., $V_{F_i} - V_{F_j} \leq 0.01(V)$), occupying 88% of all fault pairs. In this case, besides F_0 , all other fault states cannot be distinguished, which means the fault can be detected but cannot be located. Even if a measurement noise of ± 0.002 V is considered, there are still 76 pairs of indistinguishable faults.

Besides the peak voltage of receiver, other static voltages or currents can also be selected as fault features, but the results are similar. Moreover, in theory, the phase information of voltages or currents can also be used as fault features. But in engineering practice, the accurate phase information is hard to obtain when the working frequency reaches 2000 Hz, so the phase information is not considered in this paper.

From the above analysis, it can be seen that, if only static voltages or currents are used, it is very hard or even impossible to locate the compensating capacitor's fault.

3.1.3. Advantage of Adopting Dynamic Voltage Information. (Here, "Dynamic" means there is a train passing the track circuit, so the peak voltage will vary according to the train's position.) Since static information is not enough for locating the compensating capacitor's fault, some other methods should be considered according to track circuit's characteristics. It should be noticed that there are trains periodically passing through the track circuit. When a train is passing through the track circuit, the circuit structure of the track circuit will change, so more information about faults can be obtained. Based on this thinking, the peak voltages of receiver under different shunt positions can be taken as fault features (Figure 7).

From Figure 7, it can be seen that different curves under different fault states have remarkable differences. In fact, the

TABLE 5: Description of the diagnosis (classification) experiment based on SVM.

Item	Description
Vector dimension (the number of sampling points in a curve)	156
The number of classes	17
Training data set	1 vector per class (without noise)
Testing data set	50 vectors per class (with ± 0.005 V uniformly distributed measurement noise and smoothing filtering with 3 points)
SVM tool	LIBSVM [29]
Scaling scheme [30]	Vectors in training data set are linearly consistently scaled to the range [0, 1], and testing data set has the same scaling proportion as the training data set
SVM type	C-SVM [29]
Kernel type for SVM	Linear kernel
Penalty factor of the error term in C-SVM	$C = 1$
Testing result	Classification accuracy = 100%

minimal Euclidean distance ($d = \sqrt{\sum_{x=1}^N (V_i(x) - V_j(x))^2 / N}$). Here, $\sqrt{1/N}$ is taken as a constant factor for comparability with the static situation) between these curves is 0.0113 V, which is remarkably larger than the static situation.

To verify the diagnosis effect based on this kind of fault features, a diagnosis (classification) experiment based on support vector machine (SVM) has been done. This experiment is described by Table 5.

From the testing result in Table 5, it can be seen that, based on this kind of fault features, the diagnosis (classification) effect is very good.

3.1.4. Necessity of Adopting Dynamic Time Warping (DTW).

Though the dynamic voltage information is promising to give a good diagnosis effect, there is still a big problem: the train's exact position information cannot be obtained by track circuit or its monitoring system. So, in Figure 7, the horizontal axis cannot be "shunt position," but "shunt time." In this case, if the train travels in a constant speed, the new (actual) curve will have the same shape with the original curve, so there is no problem. But if the train does not travel in a constant speed, except for start point and end point (at the start point, there will be a fall edge, because the train enters the track circuit. At the end point, there will be a rising edge, because the train leaves the track circuit), other points in the new curve cannot be aligned with the corresponding points in the original curve, resulting in a nonlinear stretching deformation of the curve in horizontal direction (shown in Figure 8).

In this case, the diagnosis (classification) effect based on the actual deformed curves will become bad. To verify this point, the SVM classification experiment has been performed

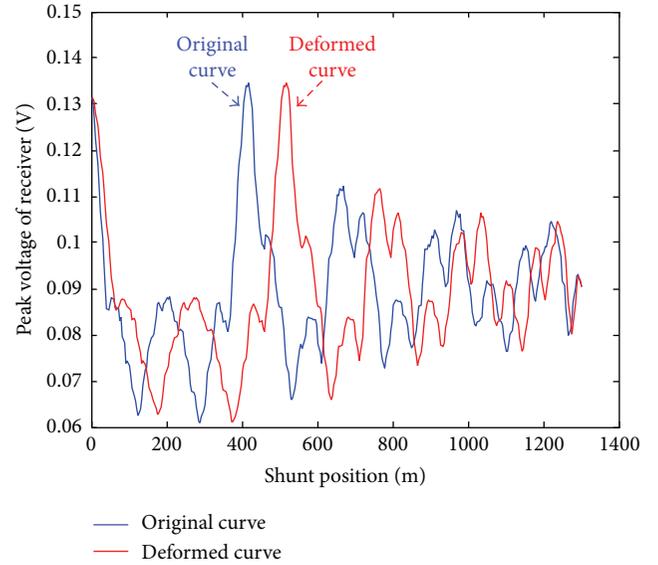


FIGURE 8: Original curve and deformed curve of fault state F3 (with noise and smoothing filtering).

once again with the deformed testing data set (see the appendix for details). This time, the classification accuracy drops to 81.29%, which means there are a lot of incorrect diagnosis results. So, it is necessary to adopt dynamic time warping (DTW) method to overcome the deformation effect.

3.2. DTW Method for Fault Diagnosis of Compensating Capacitors

3.2.1. Introduction of DTW.

Dynamic time warping (DTW) is a well-known dynamic-programming-based method to find an optimal alignment between two given sequences under certain constraints. Intuitively, the sequences are warped in a nonlinear fashion to match each other [22]. DTW was originally used in speech recognition for comparing different speech patterns [23]. DTW was also successfully applied in the fields of data mining and information retrieval, to cope with time deformations of time-dependent data [24, 25].

Among main DTW methods, the classical DTW [22] aims at computing the DTW distance and optimal warping path using the dynamic programming method, which can be realized by calculating cost matrix and accumulated cost matrix. Besides the classical DTW, some variations of DTW [22] aim at avoiding abnormal alignments and promoting computational efficiency by adding step size condition, local weights, and global constraints. The approximate DTW and multiscale DTW [26, 27] were proposed to speed up DTW computations based on the idea to perform the alignment on coarsened versions of the sequences X and Y .

The related concepts and algorithms of the classical DTW are simply introduced as follows [22].

Objective of DTW. The objective of DTW is to find an optimal alignment between two given sequences:

$X := (x_1, x_2, \dots, x_N)$ of length N and $Y := (y_1, y_2, \dots, y_M)$ of length M .

Local Cost Measure and Cost Matrix. To measure the alignment level of two different features $x_i, y_j \in F$, a local cost measure $c : F \times F \rightarrow R_{\geq 0}$ is needed. Based on the local cost measure, the cost matrix $C \in R^{N \times M}$ can be defined by $C(i, j) := c(x_i, y_j)$.

Warping Path. An (N, M) -warping path (“warping path” for short) is a sequence $p = (p_1, p_2, \dots, p_L)$ with $p_l = (i_l, j_l) \in [1 : N] \times [1 : M]$, $(l \in [1 : L])$ satisfying the following conditions:

- (1) boundary condition: $p_1 = (1, 1)$ and $p_L = (N, M)$;
- (2) monotonicity condition: $n_1 \leq n_2 \leq \dots \leq n_L$ and $m_1 \leq m_2 \leq \dots \leq m_L$;
- (3) step size condition: $p_{l+1} - p_l \in \{(1, 0), (0, 1), (1, 1)\}$, $(l \in [1 : L - 1])$.

A warping path defines an alignment between sequence X and sequence Y .

Total Cost. The total cost of a warping path is defined as $c_p(X, Y) := \sum_{l=1}^L c(x_{i_l}, y_{j_l}), (i_l, j_l) \in p$. The total cost represents the alignment level of an alignment defined by the warping path p .

DTW Distance. The minimum total cost is defined as DTW distance:

$$d_{\text{DTW}}(X, Y) := \min \{c_p(X, Y) \mid p \text{ is an } (N, M)\text{-warping path}\}. \quad (1)$$

Optimal Warping Path. The warping paths corresponding to the DTW distance are defined as optimal warping paths, which represent the optimal alignments. An optimal warping path defines a path through the cost matrix, which has the minimum total cost.

The objective of DTW can be summarized as finding the DTW distance and the optimal warping paths.

Accumulated Cost Matrix. The accumulated cost matrix is used to calculate the DTW distance and the optimal warping paths. It is defined as follows:

$$\begin{aligned} D(n, m) &:= d_{\text{DTW}}(X(1:n), Y(1:m)) \\ X(1:n) &:= (x_1, x_2, \dots, x_n), \quad n \in [1 : N] \\ Y(1:m) &:= (y_1, y_2, \dots, y_m), \quad m \in [1 : M]. \end{aligned} \quad (2)$$

Algorithm for DTW Distance. Based on the accumulated cost matrix, the algorithm for calculating DTW distance is as follows:

$$\begin{aligned} D(n, 1) &= \sum_{k=1}^n c(x_k, y_1), \quad n \in [1 : N] \\ D(1, m) &= \sum_{k=1}^m c(x_1, y_k), \quad m \in [1 : M] \\ D(n, m) &= \min \{D(n-1, m-1), D(n-1, m), \\ &\quad D(n, m-1)\} + c(x_n, y_m), \\ &\quad \text{for } n \in [2 : N], \quad m \in [2 : M] \\ d_{\text{DTW}}(X, Y) &= D(N, M). \end{aligned} \quad (3)$$

The computation complexity of this algorithm is $O(NM)$.

Algorithm for Optimal Warping Path. Based on the accumulated cost matrix, the algorithm for calculating optimal warping paths is as follows:

$$p_L = (N, M)$$

$$\text{Suppose } p_l = (n, m)$$

if $(n, m) = (1, 1)$ then l must be 1 and algorithm finishes else

$$p_{l-1} = \begin{cases} (1, m-1), & \text{if } n = 1 \\ (n-1, 1), & \text{if } m = 1 \\ \arg \min \{D(n-1, m-1), \\ D(n-1, m), D(n, m-1)\}, & \text{otherwise.} \end{cases} \quad (4)$$

3.2.2. Basic DTW Method for Fault Diagnosis of Compensating Capacitors. Since the problem is to classify the deformed curves into 17 classes (17 fault states) according to the 17 training samples (see Table 5), instead of directly using SVM, the problem can be solved by using DTW between the deformed curve and the 17 training sample curves. That is, the DTW distances between the deformed curve and the 17 training sample curves can be calculated, and the training sample curve corresponding to the minimum DTW distance will be the target class.

This algorithm can be formally described as follows:

$X_i := (x_1^i, x_2^i, \dots, x_N^i)$ of length N is the training sample for class i and $Y := (y_1, y_2, \dots, y_M)$ of length M is the deformed curve to be classified. Based on the above classical DTW, the classification (diagnosis) result can be calculated as follows:

$$t = \arg \min \{d_{\text{DTW}}(X_i, Y) \mid i \in [1 : 17]\}. \quad (5)$$

According to this algorithm, the deformed curve in Figure 8 can be restored to approach the original curve (shown in Figure 9).

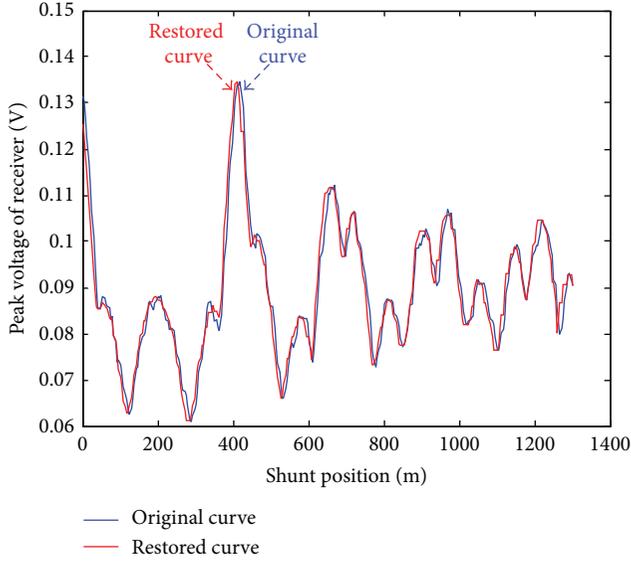


FIGURE 9: Original curve and restored curve of fault state F3 (with noise and smoothing filtering).

TABLE 6: Description of the diagnosis (classification) experiment based on DTW.

Item	Description
Vector dimension (the number of sampling points in a curve)	156
The number of classes	17
Local cost measure	$c(x, y) = x - y $
Training data set	1 vector per class (without noise)
Testing data set	50 vectors per class with: (i) ± 0.005 V uniformly distributed measurement noise (ii) Smoothing filtering with 3 points (iii) Deformation (see Appendix for details)
Testing result	Classification accuracy = 99.88%
Computation complexity	$O(156 \times 156 \times 17 \times (50 \times 17))$
Classification stability (minimum)	-0.0429 (0.1795 for correct classifications)
Classification stability (average)	0.9260

To verify the diagnosis effect of the DTW based method, the deformed testing data set in Section 3.1.4 is again used for the verification experiment. The experiment is described in Table 6.

From the testing result in Table 6, it can be seen that, in spite of the deformation of testing data set, by the DTW based method, the classification accuracy can still reach 99.88%. In Section 3.1.4, the classification accuracy can only reach 81.29% by SVM method.

Besides classification accuracy, in Table 6, the “classification stability” is also used as an index for judging

TABLE 7: Description of the diagnosis (classification) experiment based on DTW (with improved local cost measure).

Item	Description
Local cost measure	$c(x, y) = x - y ^2$, $f''(d) = 2$
Testing result	Classification accuracy = 100%
Classification stability (minimum)	0.4165
Classification stability (average)	5.0881

the classification quality. The stability ψ for one classification is defined as follows:

Suppose Y belongs to class t , $t \in [1 : 17]$

$$d_1 = d_{\text{DTW}}(X_t, Y)$$

$$d_2 = \min \{d_{\text{DTW}}(X_i, Y) \mid i \in [1 : 17], i \neq t\} \quad (6)$$

$$\psi = \frac{(d_2 - d_1)}{\min(d_1, d_2)}.$$

If the classification result is incorrect, then $\psi < 0$; otherwise, $\psi > 0$. Regardless of $\psi < 0$ or $\psi > 0$, bigger ψ implies better classification quality. $\psi = 0$ is a special situation, which means the minimum DTW distance is not unique.

3.2.3. Improvement on DTW Method for Classification Quality. The classification quality of the above DTW method can still be improved by simple modification of the local cost measure. Since local cost measure is a cost measure for one step in a warping path, it can be considered to modify the cost structure to encourage the correct alignment. The modification scheme is as follows:

$$\text{suppose } d = |x - y|,$$

$$c(x, y) = f(d), \quad \text{s.t. } f''(d) > 0. \quad (7)$$

The basic idea here is that $f'(d)$ should increase with d 's increment, which means bigger d should have a higher cost rate. This is because

- (1) small d is tolerable in correct alignment, because it may be caused by noise;
- (2) big d is intolerable in correct alignment, because it is unlikely to be caused by noise.

A simple local cost measure satisfying the condition $f''(d) > 0$ is $c(x, y) = |x - y|^2$. The diagnosis (classification) experiment based on this local cost measure is described in Table 7 (other items are the same as Table 6).

From the testing result and classification stability in Table 7, it can be seen that, with the improved local cost measure, there is a remarkable improvement on the classification quality.

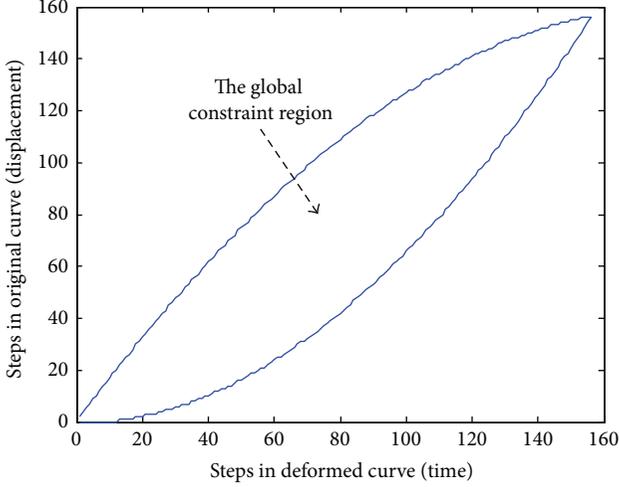


FIGURE 10: The global constraint region based on the dynamic characteristics of CRH3.

TABLE 8: Related symbols for analysis of the boundary conditions of the global constraint region.

Symbol	Description
x	Abscissa in Figure 10
y	Ordinate in Figure 10
s	Displacement (m)
v	Velocity (m/s)
t	Time (s)
a	Acceleration (m/s ²)
v_1	Start velocity on the track circuit (m/s)
v_2	End velocity on the track circuit (m/s)
T	Total time on the track circuit (s)
L	Length of the track circuit (m)
k	Slope of curve in Figure 10
k_1	Start slope of curve in Figure 10
k_2	End slope of curve in Figure 10

3.2.4. Improvement on DTW Method for Computation Complexity. In order to reduce the computation complexity, the global constraints in variations of DTW [22] can be used. However, neither the Sakoe-Chiba band region [22] nor the Itakura parallelogram region [22] is adopted for the global constraint. According to the characteristic of the train's dynamics (see the appendix for details), the following global constraint region (Figure 10), which is generated by boundary gears (maximum gear or minimum gear) and boundary velocities ($v_1 = 0$ or $v_2 = 0$), is adopted for the DTW method applied in fault diagnosis of track circuit.

The boundary conditions of the global constraint region in Figure 10 can be analyzed as follows (the related symbols are defined in Table 8).

First, the lower boundary curve is considered. According to the physical relationships between the above variables, the following formulas can be given:

$$\begin{aligned}
 t &= \frac{xT}{160}, \\
 s &= \frac{yL}{160}, \\
 v &= \frac{ds}{dt} = \frac{dy}{dx} \cdot \frac{L}{T}, \\
 \therefore k &= \frac{dy}{dx} = v \cdot \frac{T}{L}.
 \end{aligned} \tag{8}$$

For reasonable simplification, the acceleration is assumed to be constant; then the following formulas can be given:

$$\begin{aligned}
 v_2^2 - v_1^2 &= 2aL \\
 T &= \frac{2L}{v_2 + v_1}
 \end{aligned} \tag{9}$$

$$v_2 = \sqrt{2aL + v_1^2}$$

$$\begin{aligned}
 \therefore \Delta k &= k_2 - k_1 \\
 &= (v_2 - v_1) \cdot \frac{T}{L} = \frac{2aL}{v_2 + v_1} \cdot \frac{2}{v_2 + v_1} \\
 &= \frac{4aL}{(v_2 + v_1)^2} = \frac{4aL}{2aL + 2v_1^2 + 2\sqrt{2aL + v_1^2} \cdot v_1} \\
 &= \frac{4L}{2L + 2v_1^2/a + 2\sqrt{2L/a + v_1^2/a^2} \cdot v_1}.
 \end{aligned} \tag{10}$$

So, Δk will decrease when v_1 is increasing and will increase when a is increasing. Thus, when v_1 is minimum and a is maximum, the maximum Δk can be gotten, which corresponds to the maximum deformation of the lower boundary curve.

So, the boundary conditions of the lower boundary curve are $v_1 = 0$ and $a = \max$. The upper boundary curve is symmetrical with the lower boundary curve. So the upper boundary conditions are $v_2 = 0$ and $a = -\max$.

In the case of the global constraints, the elements beyond the global constraint region in accumulated cost matrix does not need to be calculated and can be directly set as infinity, so the computation complexity can be reduced to $S_{\text{region}}/S_{\text{square}}$ (area ratio) of the original computation complexity. The diagnosis (classification) experiment based on this global constraint region is described in Table 9 (other items are the same as Table 6).

From the results in Table 9, it can be seen that, with the global constraint, there is a remarkable reduction in the computation complexity, at the cost of one incorrect classification.

TABLE 9: Description of the diagnosis (classification) experiment based on DTW (with the global constraint).

Item	Description
Local cost measure	$c(x, y) = x - y ^2, f''(d) = 2$
Testing result	Classification accuracy = 99.88%
Computation complexity	$O(29.26\% \times (156 \times 156) \times 17 \times (50 \times 17))$
Classification stability (minimum)	-0.0468 (0.1299 for correct classifications)
Classification stability (average)	6.5735

4. Results and Discussion

The results of the above experiments can be summarized in Table 10.

From Table 10, the results of all the experiments can be analyzed as follows.

- (1) The method of comparison of static voltages has very poor classification quality and can hardly be used for fault location of compensating capacitors.
- (2) The method of SVM without DTW has a generally acceptable classification quality, but there are still many incorrect classifications, which will cause trouble in practical applications.
- (3) The classical DTW has good classification accuracy, but the classification stability is poor, which means incorrect classifications are prone to appear under severe conditions such as big noise. Besides, the computation complexity of this method is great, which means it is not suitable for online diagnosis.
- (4) The DTW method with improved local cost measure has very good classification accuracy and good classification stability, but the computation complexity has not been improved.
- (5) The DTW method with global constraint has good classification quality, great classification stability, and reduced computation complexity. The very few incorrect classifications, which are not general situations, may be caused by the nonconservative global constraint region. As a whole, this method is the best one among these methods for online fault diagnosis of compensating capacitors in track circuit.

5. Conclusions

Aiming at the problem of online fault diagnosis for compensating capacitors of jointless track circuit, a dynamic time warping (DTW) based diagnosis method is proposed in this paper. Different from the existing related works, this method only uses the ground monitoring signals of track circuit to locate the faulty compensating capacitor, not depending on the shunt current of inspection train, which is an indispensable condition for existing methods. So, it can be

used for online diagnosis of compensating capacitor, which has not yet been realized by existing methods. Besides, this method does not need the outdoor monitoring signals or the on-board monitoring signals, so the existing system does not need to be modified greatly, which means a good engineering feasibility. To overcome the key problem that track circuit cannot obtain the precise position of the train, the DTW method is used for the first time in this situation to recover the function relationship between receiver's peak voltage and shunt position. Besides the classical DTW based method, two improved methods for improving classification quality and reducing computation complexity are proposed. Finally, the diagnosis experiments based on the simulation model of track circuit show the effectiveness of the proposed methods. Fault diagnosis for compensating capacitors under uncertain ballast resistance and shunt resistance will be the future work.

Appendix

Generation of Deformed Testing Data Set

In order to generate the almost real deformed testing data set, an effective dynamic model of train must be first established. Referring to [28], the dynamic model (state equation) of CRH3 high-speed train is established as follows:

$$Mv \frac{dv}{dx} = F_{\sigma}(v) - r(v) \quad (\text{A.1})$$

$$\frac{dt}{dx} = \frac{1}{v}.$$

Here, x instead of t is taken as independent variable for convenience of integral calculation at given displacement interval $[0, L]$ (L is the length of the track circuit).

The related symbols are described in Table 11.

The traction characteristic of CRH3 high-speed train is described as follows [28]:

$$F_{\sigma}(v) = \begin{cases} c_{\sigma} (300 - 0.284v) \text{ kN}, & \text{for } 0 \leq v \leq 119.7 \text{ km/h}, \sigma > 0 \\ c_{\sigma} (266 \times 119.7/v) \text{ kN}, & \text{for } 119.7 < v \leq 300 \text{ km/h}, \sigma > 0 \\ 0 \text{ kN}, & \text{for } \sigma = 0 \\ c_{\sigma} (-300 + 0.281v) \text{ kN}, & \text{for } 0 \leq v \leq 106.7 \text{ km/h}, \sigma < 0 \\ c_{\sigma} (-270 \times 106.7/v) \text{ kN}, & \text{for } 106.7 < v \leq 300 \text{ km/h}, \sigma < 0, \end{cases} \quad (\text{A.2})$$

where c_{σ} is the gear coefficient for CRH3, which is defined in Table 12.

The resistance characteristic of CRH3 high-speed train is described as follows [28]:

$$r(v) = 6.4M + 130q + 0.14Mv + [0.046 + 0.0065(p - 1)] Av^2. \quad (\text{A.3})$$

TABLE 10: Summary of results of all the experiments.

Method	Classification accuracy	Classification stability (average)	Computation complexity (for one classification)
Comparison of static voltages	Very low (not able to locate the faults)	NA*	NA
SVM (without DTW)	81.29%	NA	NA
Classical DTW	99.88%	0.9260	O (156 × 156 × 17)
DTW with improved local cost measure	100%	5.0881	O (156 × 156 × 17)
DTW with global constraint	99.88%	6.5735	O (29.26% × 156 × 156 × 17)

*“NA” means not applicable.

TABLE 11: Related symbols in the dynamic model of train.

Symbol	Description
x	Displacement (m)
v	Velocity (m/s)
t	Time (s)
M	Mass (kg, 408t for CRH3)
p	The number of cars (8 for CRH3)
F	Tractive force or braking force (N)
σ	Gear
r	Resistance (N)
A	Frontal area (m ² , 9 m ² for CRH3)
q	The number of axles (32 for CRH3)

TABLE 12: Gear coefficients for CRH3.

Gear	Tractive gear				Neutral gear		Braking gear			
	4	3	2	1	0	-1	-2	-3	-4	
c_{σ}	1	0.7	0.5	0.3	0	0.3	0.5	0.7	1	

The differential equation solver “ode45” in Matlab is used for solving the state equation with the initial condition “(v_0 , $t_0 = 0$)” and the integral interval $[0, L = 1301]$. For embodying randomness of samples, the gear σ is randomly selected in $\{-4, -3, -2, -1, 0, 1, 2, 3, 4\}$, and v_0 is randomly selected in $[v_{\min}, v_{\max}]$. v_{\min} and v_{\max} are selected to ensure $v \in (0, 300 \text{ km/h}]$ when the train is running on the track circuit.

The typical velocity-time curve and displacement-time curve generated from the dynamic model are shown in Figures 11 and 12, which are corresponding to Figure 8.

Based on the displacement-time curves generated from the dynamic model, the deformed testing data set can be generated from the original testing data set.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

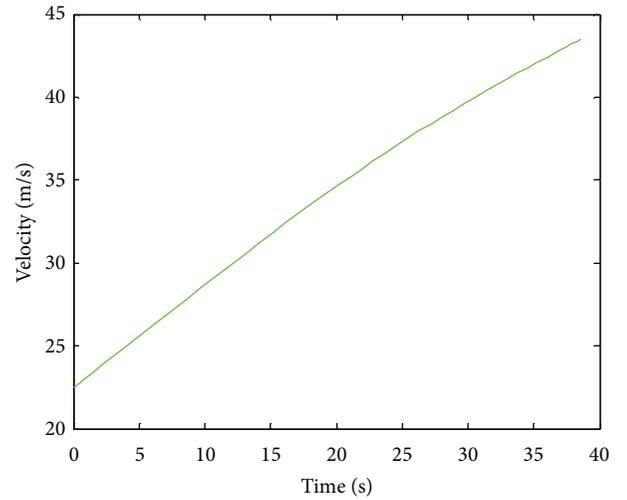


FIGURE 11: Typical velocity-time curve.

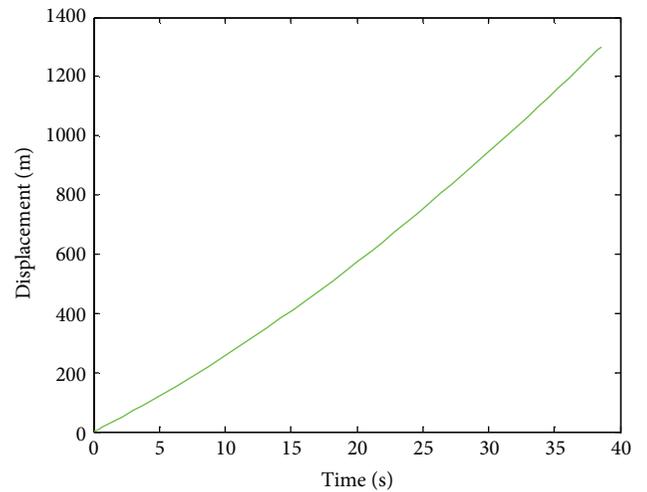


FIGURE 12: Typical displacement-time curve.

Acknowledgments

This work was supported by the NSFC under Grants 61104019 and 61374123 and the Tsinghua University Initiative Scientific Research Program.

References

- [1] C. J. Baker, L. Chapman, A. Quinn, and K. Dobney, "Climate change and the railway industry: a review," *Proceedings of the Institution of Mechanical Engineers C: Journal of Mechanical Engineering Science*, vol. 224, no. 3, pp. 519–528, 2010.
- [2] F. P. García Márquez, R. W. Lewis, A. M. Tobias, and C. Roberts, "Life cycle costs for railway condition monitoring," *Transportation Research E: Logistics and Transportation Review*, vol. 44, no. 6, pp. 1175–1187, 2008.
- [3] J. Chen, C. Roberts, and P. Weston, "Neuro-fuzzy fault detection and diagnosis for railway track circuits," in *Proceedings of the 6th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes (SAFEPROCESS '06)*, no. 1, pp. 1366–1371, September 2006.
- [4] J. Chen, C. Roberts, and P. Weston, "Fault detection and diagnosis for railway track circuits using neuro-fuzzy systems," *Control Engineering Practice*, vol. 16, no. 5, pp. 585–596, 2008.
- [5] Y. Cheng, *Reliability and Safety of Railway Signal System*, China Railway, 2010.
- [6] Z.-W. Huang, X.-Y. Wei, and Z. Liu, "Fault diagnosis of railway track circuits using fuzzy neural network," *Journal of the China Railway Society*, vol. 34, no. 11, pp. 54–59, 2012.
- [7] L. Oukhellou, A. Debiolles, T. Dencœux, and P. Aknin, "Fault diagnosis in railway track circuits using Dempster-Shafer classifier fusion," *Engineering Applications of Artificial Intelligence*, vol. 23, no. 1, pp. 117–128, 2010.
- [8] E. Côme, L. Oukhellou, T. Dencœux, and P. Aknin, "Noiseless independent factor analysis with mixing constraints in a semi-supervised framework. application to railway device fault diagnosis," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5769, no. 2, pp. 416–425, 2009.
- [9] Z. L. Cherfi, L. Oukhellou, E. Côme, T. Dencœux, and P. Aknin, "Partially supervised Independent Factor Analysis using soft labels elicited from multiple experts: application to railway track circuit diagnosis," *Soft Computing*, vol. 16, no. 5, pp. 741–754, 2012.
- [10] H. Zhao, Y. Liu, and B. Liu, "Research on the onboard auto-test system for track circuit compensating capacitors," in *Proceedings of the 10th International Conference on Computer System Design and Operation in the Railway and Other Transit Systems, COMPRAIL (CR '06)*, pp. 965–971, July 2006.
- [11] L. Zhao, Z. Li, and W. Liu, "The compensation capacitors fault detection method of jointless track circuit based on DBWT and WR," in *Proceedings of the IEEE International Conference on Intelligent Computing and Intelligent Systems (ICIS '09)*, pp. 875–879, Shanghai, China, November 2009.
- [12] L. Zhao, H. Li, J. Guo, and W. Liu, "The simulation analysis of influence on jointless track circuit signal transmission from compensation capacitor based on transmission-line theory," in *Proceedings of the 3rd IEEE International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications (MAPE '09)*, pp. 1113–1118, Beijing, China, October 2009.
- [13] Z. Linhai, L. Baosheng, and X. Xun, "Using the wavelet analysis based noise canceling technique to detect the integrity of the compensating capacitors of UM71 track circuit," in *Proceedings of the IEEE International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications (MAPE '05)*, pp. 1268–1274, Beijing, China, August 2005.
- [14] L. Zhao and J. Mu, "Fault diagnosis method for jointless track circuit based on AOK-TFR," *Journal of Southwest Jiaotong University*, vol. 46, no. 1, pp. 84–91, 2011.
- [15] L.-H. Zhao, B.-G. Cai, K.-M. Qiu, and Z.-Q. Lu, "The method of diagnosis of compensation capacitor failures with jointless track circuits based on HHT and DBWT," *Journal of the China Railway Society*, vol. 33, no. 3, pp. 49–54, 2011.
- [16] L.-H. Zhao, J.-J. Xu, W.-N. Liu, and B.-G. Cai, "Compensation capacitor fault detection method in jointless track circuit based on Levenberg-Marquardt algorithm and generalized S-transform," *Control Theory and Applications*, vol. 27, no. 12, pp. 1612–1622, 2010.
- [17] L. Zhao, Y. Ran, and J. Mu, "A comprehensive fault diagnosis method for jointless track circuit based on genetic algorithm," *China Railway Science*, vol. 31, no. 3, pp. 107–114, 2010.
- [18] Z. Lin-Hai, W. Jian-Ping, and R. Yi-Kui, "Fault diagnosis for track circuit using AOK-TFRs and AGA," *Control Engineering Practice*, vol. 20, no. 12, pp. 1270–1280, 2012.
- [19] S. P. Sun, H. B. Zhao, and G. Zhou, "A fault diagnosis approach for railway track circuits trimming capacitors using EMD and teager energy operator," *WIT Transactions on the Built Environment*, vol. 127, pp. 167–176, 2012.
- [20] M. Guang-zhi and X. Xue-shu, "The modeling and simulation of track circuit," *Electric Drive For Locomotives*, no. 1, pp. 41–43, 2004.
- [21] Z. Yong-xian, X. Xue-song, and Y. Jiang-song, "Modeling and simulation of ZPW-2000A jointless track circuit," *Journal of East China Jiaotong University*, no. 3, pp. 64–68, 2009.
- [22] M. Müller, *Information Retrieval For Music and Motion*, Springer Press, 2007.
- [23] L. R. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall Signal Processing Series, 1993.
- [24] E. Keogh, "Exact indexing of dynamic time warping," in *Proceedings of the 28th International Conference on Very Large Data Bases*, pp. 406–417, Hong Kong, China, August 2002.
- [25] S. Dick, A. Meeks, M. Last, H. Bunke, and A. Kandel, "Data mining in software metrics databases," *Fuzzy Sets and Systems*, vol. 145, no. 1, pp. 81–110, 2004.
- [26] S. Chu, E. Keogh, D. Hart, and M. Pazzani, "Iterative deepening dynamic time warping for time series," in *Proceedings of the 2nd SIAM International Conference on Data Mining*, pp. 195–212, Chicago, Ill, USA, April 2002.
- [27] S. . Salvador and P. Chan, "FastDTW: toward accurate dynamic time warping in linear time and space," in *Proceedings of KDD Workshop on Mining Temporal and Sequential Data*, Seattle, Wash, USA, August 2004.
- [28] L. Liang, D. Wei, J. Yindong, Z. Zengke, and T. Lang, "Minimal-energy driving strategy for high-speed electric train with hybrid system model," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 4, pp. 1642–1653, 2013.
- [29] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, article 27, 2011.
- [30] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," <http://www.csie.ntu.edu.tw>.