Complexity

# Complexity in Manufacturing Processes and Systems 2019

Lead Guest Editor: Rosario Domingo
Guest Editors: Julio Blanco-Fernández and Jorge Luis García-Alcaraz

# Complexity in Manufacturing Processes and Systems 2019

# Complexity in Manufacturing Processes and Systems 2019

Lead Guest Editor: Rosario Domingo
Guest Editors: Julio Blanco-Fernández and Jorge Luis García-Alcaraz

# Contents

Hindawi
Complexity
Volume 2020, Article ID 7286932, 3 pages
https://doi.org/10.1155/2020/7286932

WILEY | Hindawi

*Editorial*

# Complexity in Manufacturing Processes and Systems 2019

**Rosario Domingo**,[1] **Julio Blanco-Fernández**,[2] **and Jorge Luis García-Alcaraz**[3]

[1]*Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain*
[2]*Universidad de La Rioja, Logroño, Spain*
[3]*Autonomous University of Ciudad Juarez, Ciudad Juárez, Mexico*

Correspondence should be addressed to Rosario Domingo; rdomingo@ind.uned.es

Received 2 December 2019; Accepted 2 December 2019; Published 3 July 2020



This Special Issue is continuation of the work published in 2018 with the same name [1]. As is expected, the control or behavior knowledge of factors that influence on the production performance continues to be important in the academic and industrial field. Thus, new algorithms are developed to reduce the uncertainty and improve the efficiency in machines [2], processes [3, 4], and even the supply chain [5], and new methods are proposed to know the relationship among production factors in some sectors [6, 7] or more specific as in assembly lines [8]. In this context, the communication technologies and the advanced manufacturing technologies could be considered because they affect the performance and facilitate to monitoring the production process [9, 10] or new customized product configurations [11].

Fourteen papers have been accepted for this special issue, after a strict review process. Between them, there are four papers devoted to the manufacturing process, and almost all of them show methods to control some manufacturing parameters.

V. Flores and B. Keith in their paper titled "Gradient Boosted Trees Predictive Models for Surface Roughness in High-Speed Milling in the Steel and Aluminum Metalworking Industry" present a model of the surface roughness average that evaluates the influence of the cutting conditions in the machining process, in particular, in high-speed end-milling. The model is constructed after a predictive analysis by gradient boosted trees derivate of machine learning techniques.

M. Du et al. in their paper titled "Intelligent Turning Tool Monitoring with Neural Network Adaptive Learning" investigate the tool state monitoring in the machining process using artificial intelligent techniques. The paper has an experimental procedure applied to inserts in the turning process, whereby the data are taken using vibration and acoustic emission signals for tool monitoring. The data, taken in real time, are treated by multitheory fusion of wavelet decomposition and the ReliefF algorithm and after neural network adaptive learning. The results indicate a good prediction.

M. Jiménez et al. in their paper titled "Additive Manufacturing Technologies: An Overview about 3D Printing Methods and Future Prospects" realize an overview related to additive manufacturing and considering the structural design of the machines, possible alternatives, more significant materials, technological limitations, and new perspectives in professional and academic fields.

T. R. Gazizov et al. devoted the paper "Solving the Complexity Problem in the Electronics Production Process by Reducing the Sensitivity of Transmission Line Characteristics to Their Parameter Variations" to the electronics production process. With a versatile approach, the study analyzes the reduction in sensitivity of transmission line characteristics to their parameter variations. Different results in microstrip lines are presented.

The other group of papers has a more comprehensive approach, and they consider external factors to manufacturing processes, considering the production system, the supply chain, or the enterprise. They are presented in the following paragraphs.

Q. Chen et al. in their paper titled "Dynamic Matching in Cloud Manufacturing considering Matching Costs" explore

the resources' dynamic matching in a manufacturing supply chain that operates under a cost-sharing contract. The authors model the evolution of resource sharing by means of differential equations, study the optimal matching strategies by a two-stage differential game based on the dynamic control approach, and design a cost-sharing contract to coordinate and improve the supply chain's performance. Moreover, a numerical analysis shows the influence of transaction fee and purchasing costs on different geographical regions and strategies.

Y. Changwei and G. Jinhao focus their paper titled "Research on Warning Threshold Value Optimization and Disposal Strategy of Earthquake Early Warning for High-speed Railway under Earthquake" on the design of a model to study the derailment phenomenon of trains and the seismic performance of different structural types. A dynamic simulation allows identifying the safe running speed threshold of high-speed trains under earthquake.

F. Chen et al. in their paper titled "An Integrated Metaheuristic Routing Method for Multiple-Block Warehouses with Ultranarrow Aisles and Access Restriction" develop a routing algorithm that allows integrating an ant colony optimization (ACO) and integer-coded genetic algorithm (GA). This algorithm is applied to internal logistic with ultranarrow aisles and access restrictions. A simulation considering 12 warehouse layouts shows that the new algorithm allows better solutions according to certain efficiency indexes.

J. Qin and Z. Li in their paper titled "Reliability and Sensitivity Analysis Method for a Multistate System with Common Cause Failure" develop a reliability assessment method for a multistate system (MSS) with series-parallel and common cause failure (CCF). The method is based on the universal generating function (UGF) to which is incorporated the CCF. Different applications show the goodness of this method by means of calculus of reliability indexes.

The paper "Pricing and Collection Rate for Remanufacturing Industry considering Capacity Constraint in Recycling Channels" written by L. Xu et al. explores the coordination in a closed-loop supply chain taking into account capacity constraints in recycling channels. The paper considers two scenarios, centralized and decentralized, to evaluate the equilibriums of decisions and profits, according to the game theory. Findings indicate that, in recycling channels, when the capacity constraints surpass a threshold, the profit is modified and that these constraints affect the selling price more than other variables.

F. Brocal et al. in their paper titled "Emerging Risk Management in Industry 4.0: An Approach to Improve Organizational and Human Performance in the Complex Systems" propose an organizational and human performance approach to improve the emerging risk management in systems such as Human-Machine Interactions (HMI) and Human-Robot Interaction (HRI). The concept of emerging risk management in this framework is introduced and its effects on complex production systems.

The paper "The Incentive Model in Supply Chain with Trade Credit and Default Risk" written by H. Cheng et al. is devoted to trade credit analysis in supply chain management with default risk and moral hazard. The authors develop the principal-agent models by means of incentive theory in a newsvendor approach and considering situations of symmetric and asymmetric information. Findings indicate that with an incentive model, the order quantity is lowered in the incentive contract to mitigate the effect of moral hazard.

X. Wang et al. in their paper titled "A Novel MOEA/D for Multiobjective Scheduling of Flexible Manufacturing Systems" present a deadlock-free multiobjective evolutionary algorithm based on decomposition (DMOEA/D) applied to flexible manufacturing systems. This algorithm decomposes a multiobjective scheduling problem in single-objective subproblems and can resolve all these subproblems in a single run. The computational results show that this method finds better Pareto solutions than other traditional approaches.

P. Urgilés et al. in their paper titled "Analysis of the Earned Value Management and Earned Schedule Techniques in Complex Hydroelectric Power Production Projects: Cost and Time Forecast" focus on the uncertainty of industrial projects. The paper analyzes the efficiency of the Earned Value Management technique, as well as its Earned Schedule extension. Simulation indicates that cost forecasting can be predicted very accurately; however, the duration forecasting is not accurate.

The paper "STEAM-ME: A Novel Model for Successful Kaizen Implementation and Sustainable Performance of SMEs in Vietnam" written by T. L. Nguyen proposes a model that allows identifying critical factors during the implementation and sustainable performance of Kaizen concept in small and medium enterprises (SMEs). This model named "STEAM-ME" is created from a survey carried out in 213 participants from 62 SMEs; this survey allows identifying seven factors after an exploratory statistical analysis and structural equation modeling: supports from senior management, training, environment, assessment, motivation, mindset, and engagement of all members in the organization.

This special issue, as the first issue dedicated to Complexity in Manufacturing Processes and Systems, can be interesting for engineers and managers with responsability in manufacturing plants, where new methods and results show how the complexity can be managed in industrial environments. New advances can be made from the findings of these papers.

## Conflicts of Interest

The editors declare that there are no conflicts of interest regarding the publication of this special issue.

## Acknowledgments

The Guest Editorial Team would like to thank all the authors for their interest in selecting the special issue "Complexity in Manufacturing Processes and Systems 2019" and for their valuable contributions.

*Rosario Domingo*
*Julio Blanco-Fernández*
*Jorge Luis García-Alcaraz*

## References


[1] R. Domingo, J. Blanco-Fernández, J. L. García-Alcaraz, and L. Rivera, "Complexity in manufacturing processes and systems," *Complexity*, vol. 2018, Article ID 8738764, 3 pages, 2018.

[2] R. Calvo, R. D'Amato, E. Gómez, and R. Domingo, "Integration of error compensation of coordinate measuring machines into feature measurement: part I-model development," *Sensors*, vol. 16, no. 10, p. 1610, 2016.

[3] C. Garriz and R. Domingo, "Development of trajectories through the kalman algorithm and application to an industrial robot in the automotive industry," *IEEE Access*, vol. 7, pp. 23570–23578, 2019.

[4] M. Oropesa Vento, J. L. García Alcaraz, A. A. Maldonado Macías, and V. Martínez Loya, "The impact of managerial commitment and Kaizen benefits on companies," *Journal of Manufacturing Technology Management*, vol. 27, no. 5, pp. 692–712, 2016.

[5] S. Kusi-Sarpong, H. Gupta, and J. Sarkis, "A supply chain sustainability innovation framework and evaluation methodology," *International Journal of Production Research*, vol. 57, no. 7, pp. 1990–2008, 2019.

[6] J. R. Díaz-Reza, J. R. Mendoza-Fong, J. Blanco-Fernández, J. A. Marmolejo-Saucedo, and J. L. García-Alcaraz, "The role of advanced manufacturing technologies in production process performance: a causal model," *Applied Sciences*, vol. 9, no. 18, p. 3741, 2019.

[7] R. Villanueva-Ponce, J. L. Garcia-Alcaraz, G. Cortes-Robles, J. Romero-Gonzalez, E. Jiménez-Macías, and J. Blanco-Fernández, "Impact of suppliers' green attributes in corporate image and financial profit: case maquiladora industry," *The International Journal of Advanced Manufacturing Technology*, vol. 80, no. 5–8, pp. 1277–1296, 2015.

[8] V. G. Cannas, M. Pero, R. Pozzi, and T. Rossi, "Complexity reduction and kaizen events to balance manual assembly lines: an application in the field," *International Journal of Production Research*, vol. 56, no. 11, pp. 3914–3931, 2018.

[9] J. L. García-Alcaraz, V. Martínez-Loya, J. R. Díaz-Reza, J. Blanco-Fernández, E. Jiménez-Macías, and A. J. G. López, "Effect of ICT integration on SC flexibility, agility and company' performance: the Mexican maquiladora experience," *Wireless Networks*, 2019.

[10] N. F. Alkayem, M. Cao, and M. Ragulskis, "Damage diagnosis in 3D structures using a novel hybrid multiobjective optimization and FE model updating framework," *Complexity*, vol. 2018, Article ID 3541676, 13 pages, 2018.

[11] S. Bednar and E. Rauch, "Modeling and application of configuration complexity scale: concept for customized production," *The International Journal of Advanced Manufacturing Technology*, vol. 100, no. 1–4, pp. 485–501, 2019.

Hindawi
Complexity
Volume 2020, Article ID 8391252, 13 pages
https://doi.org/10.1155/2020/8391252

*Research Article*

# Pricing and Collection Rate for Remanufacturing Industry considering Capacity Constraint in Recycling Channels

**Lang Xu**,[1,2] **Jia Shi**,[1] **and Jihong Chen**[1]

[1]*College of Transport and Communications, Shanghai Maritime University, Shanghai, China*
[2]*School of Administrative Studies, York University, Toronto, Canada*

Correspondence should be addressed to Lang Xu; jerry_langxu@yeah.net

Received 15 March 2019; Revised 17 June 2019; Accepted 11 July 2019; Published 11 June 2020

Guest Editor: Domingo Rosario



This paper explores the decision-making and coordination mechanism of pricing and collection rate in a closed-loop supply chain with capacity constraint in recycling channels, which consists of one manufacturer and one retailer. On the basis of game theory, the equilibriums of decisions and profits in the centralized and decentralized scenarios are obtained and compared. Through the performance analysis of a different scenario, a higher saving production cost and lower competition intensity trigger the members to engage in remanufacturing. Furthermore, we try to propose a two-part tariff contract through bargaining to coordinate supply chain and achieve a Pareto improvement. The results show that when the capacity constraints in recycling channels exceed a threshold, the decisions and profit will change. Additionally, for closed-loop supply chain, the selling price is more susceptible to the influence of capacity constraint in recycling channel than the members' profit.

## 1. Introduction

In recent years, the Considering Capacity Constraint in Recycling Channels closed-loop supply chain (CLSC) is a hot topic in the field of logistics and supply chain management, which has been attracting extensive attention from the industrial and theoretical circles. According to the viewpoint of the whole life cycle, products have been manufactured, wholesaled, retailed and used. Then, the end-of-life products are eliminated by consumers, which are classified, cleaned, and disassembled by the collectors; meanwhile, the valuable parts are implemented in the process of remanufacturing [1]. Remanufacturing increases the time of material utilization, and makes it more energy-saving and environmentally friendly. Studies have shown that recycling and reuse of core components can save about 40%–60% of the cost [2, 3]. Therefore, many famous enterprises like Xerox, Robert Bosch, and Hewlett-Packard improve the performance through remanufacturing. However, many companies still have doubts about whether the manufacturing can bring profit. Thus, it is of great significance to quantitatively explore the economics of remanufacturing and give some managerial guidance to collection and remanufacturing strategy for promoting the development of the remanufacturing industry and improving the level of closed-loop supply chain management.

However, the closed-loop supply chain should simultaneously consider operations of forward and reverse flow. The recycling channels of closed-loop supply chain have their corresponding infinite recycling capacity; otherwise, it goes against the realistic conditions like warehouse, routing, and staff. Under these circumstances, the relationship between capacity constraint and supply chain performance has been unpredictable. The remanufacturing capacity of Benz-cummins, an automobile-engine remanufacturer, only 3000 per year for Cummins, Doitz, and Mercedes-Benz. Further, Wuxi diesel-engine factory is located in Jiangsu Province with an annual capacity of 5000 modified vehicles. Obviously, the development strategies of both companies are affected by the production capacity, so they cannot meet the market demand. In other words, if the recycling capacity is reduced at a certain threshold, the capacity constraint results in a huge influence of decision and profit [4, 5]. In general, the less recycling capacity tends to decrease the collection ability in the reverse channel. Note that this situation may result in a phenomenon, which undercuts corporate profit aggravate environmental harm [6–9]. Furthermore, based on managerial insight, realistically, the imposing capacity constraints will

change the manufacturing and remanufacturing process. Managers; should also take the recycling capacity into account when they decide the strategies. Thus, the limitation in the recycling capacity impacting the supply chain performance is a major challenge for the operation management era.

Therefore, the behavior of collection and remanufacturing for end-of-life products is crucial to our environment and economic. According to the Stackelberg game, we establish a centralized and decentralized model to obtain the decision of pricing and collection rate in a closed-loop supply chain. Furthermore, we adopt the Nash bargaining theory to propose an improved coordinated contract and discuss the performance of optimal profits and decisions. Different from the above, we mainly fill in the research gap with the influence of the capacity in recycling channel on optimal profits and decisions. In addition, we focus on the design of coordination mechanism in a closed-loop supply chain with a manufacturer recycling channel and a retailer recycling channel to achieve a perfect Pareto improvement. The rest of this paper is organized as follows. Section 2 reviews the relative literature of closed-loop supply chain management and coordinated mechanism. In Section 3, the notation and assumption are provided. We formulate the mathematical model and decision analysis in Section 4. Further, we present the numerical analysis to obtain some managerial insights in Section 5. Section 6 provides the conclusions and suggestions for future research.

## 2. Literature Review

The Several relevant literature should be reviewed here to clarify the need for our paper. In order to demonstrate in detail the contributions, we explore the literature spanning across two subsections. In Section 2.1, we focus on the management of a closed-loop supply chain, which contains network design, collection competition and recycling constraint. Then, we address the coordination mechanism for members to improve the performance and achieve improvement in Section 2.2.

*2.1. Closed-Loop Supply Chain Management.* Many scholars discussed the remanufacturing strategy in a closed-loop supply chain, including recycling mode and channel choice, which is an academic problem and achieve an efficient decision. Savaskan et al. [10] based on the game theory to establish three recycling channels to compare the differences among manufacturer collection retailer collection and third-party collection. Then, Savaskan et al. [11] took one manufacturer and two competitive retailers into consideration in determining the decisions of pricing and collection. On this basis, Wu [12] introduced the price and service decisions into a two-stage Stackelberg game, which consists of one manufacturer, one remanufacturer, and one retailer. In addition, this model was extended to a dual-channel recycling model in a closed-loop supply chain between retailer and third-party recycling competition [13]. Zhang and Ren [14] discussed differences in optimal decision and profit in the centralized and decentralized closed-loop supply chain, as well as analyzed the influences of channel preference and sales service on performance. Morteza et al. [15] proposed a fuzzy two-objective model with the random interruptions, which is a price-dependent demand, to construct a closed-loop network between the manufacturer and the retailer. Taleizadeh et al. [16] combined pricing, quality, and collection in a closed-loop supply chain, besides explore the impact of channel structure on decision and profit. From the above, previous studies seldom considered the impacts of competitive intensity on different recycling models.

But beyond that, the other branch is recycling competition. Feng et al. [17] introduced consumer preference in establishing a dual-channel collection model for a closed-loop supply chain, which discuss the recycling competition and recycling configuration between the manufacturer and the recycler. He et al. [18] discussed competitive collection in a closed-loop supply chain, which investigated remanufacturing efficiency and consumer acceptance with channel inconvenience. Further, Liu et al. [6] proposed a price- and quality-dependent competition model between the formal and informal recyclers to explore the impact of governmental policy on the four competitive scenarios. Liu et al. [19] introduced the recycling competition into the decisions of pricing and reverse channel choice and also compared three different dual-channel recycling models. Wang et al. [20] combined the Stackelberg game to obtain the pricing strategies for the three individual collection models with the competitive collection market and demand market. From the above, these literature on the remanufacturing strategy in a two-period Stackelberg game is derived by backward induction. Yet, the remanufacturing behavior can be affected by some realistic factors, which lead to a capacity constraint in recycling channel. How to analyze the impact of capacity constraint on the performance in a closed-loop supply chain is an urgent problem.

Capacity constraint in decision-making is becoming increasingly crucial in real-life scenarios. Sereno and Efthimiadis [21] investigated the investment of firms whether to add the capacity or not, besides discussing how capacity constraints and incentive schemes influence firms' strategies. This is especially circumstance in closed-loop supply chain. Fischetti et al. [22] adopted Benders' decomposition without separability to design an algorithm and conduct a numerical analysis in deriving the capacitated facility location problems. Mota et al. [23] proposed a multi-objective MIP model, which combines the demand uncertainty and capacity restriction, to guarantee the sustainable development of a closed-loop supply chain through locating the facility. Further, Wang et al. [24] and Dominguez et al. [25] discussed respectively the influences of capacity constraint on low-carbon and closed-loop supply chain. Meanwhile, Zhen et al. [26] and Ljubic and Moreno [27] optimized the network of a closed-loop supply chain from the perspective of capacity constraint to demonstrate several decisions, such as capacity locations, network design and technology allocation. However, they mainly focused on the capacity constraint in distribution centers, which ignores the restriction in recycling channels.

*2.2. Coordinated Mechanism in Closed-Loop Supply Chain.* The topic of coordinated mechanism in the supply chain has received great attention in the existing literature. From the view of cost and revenue sharing, Cachon and Lariviere [28] compared the revenue-sharing contract, buyback contract, and wholesale price contract to conclude that the optimal coordinated mechanism. Mafakheri and Nasiri [29] focused

on the extracting value from end-of-life products to investigate the distribution of revenue sharing in a closed-loop supply chain. Hu and Feng [30] analyzed the revenue-sharing contract in a supplier buyer supply chain under the demand and supply uncertainties. Beyond that, some scholars provided the buyback contract [12, 31] and, discount schemes [32, 33], and lead time incentive mechanisms [34, 35]. Different from the traditional supply chain, Heydari and Ghasemi [36] proposed a coordinated model in a reverse supply chain based on the uncertainty for member's risk, which considered the random remanufacturing capacity, to achieve a win–win situation. Further, Li et al. [37] emphatically analyzed the efficiency of a revenue-sharing contract with the perspective of recycling cost as an extra value for the collector. Zhao and Zhu [38] designed a revenue-sharing mechanism to coordinate the manufacturer and retailer considering the uncertainties of remanufacturing rate and market demand. In addition, Xie et al. [39] compared the performance in the centralized and decentralized supply chain, which indicated the influence of channel conflict in a reverse supply chain scenario, to get a revenue-sharing coordinated mechanism. Therefore, for the members in the closed-loop supply chain, eliminating double-marginalization and ensuring profit-maximization has been a urgently problem.

In the coordinated mechanism, profit distribution between the manufacturer and retailer is more and more becoming a hot topic in the field of operation management [7]. Shi and Wu [40] adopted the fuzzy decision theory to the Shapley value for profit distribution, which considered the risk influence and capital appreciation ratio. Dai and Chen [41] designed the mechanism of profit distribution for supplier, manufacturer, and retailer to improve the traditional Shapley value. Zhang and Geng [42] demonstrated a profit allocation model in the evolutionary game model and discussed the impact of relative parameters on the member's cooperation and performance. Chen et al. [43] compared the different coordinated mechanisms with price- and time-dependent demand, as well as proposed the allocation of individual income for all participants. Wu et al. [44] combined the fairness concern into the coordinated mechanism to incentive the members' participation in energy saving. Wang et al. [45] proposed a mixed integer-programming model, which combined the transportation expansion, vehicle time, and collection routing to improve the Shapley value to balance the profits.

Outside of this, in order to improve the efficiency of supply chain coordination, some scholars investigated the contract with several methods. Zhang and Ren [14] provided a coordinated contract to balance the manufacturer and the retailer, which contained revenue sharing and cost sharing. Xie et al. [46] discussed the coordinated mechanism with service effect and sale effect based on the perspectives of revenue and expenditure. In total, there are limited points of previous studies, which didn't achieve perfectly the coordination in a closed-loop supply chain.

Different from the existing literature, this paper focuses on the decision and coordination of a closed-loop supply chain with a manufacturer recycling channel and a retailer recycling channel to obtain managerial insights from the perspective of capacity constraint in recycling channels. The main contributions of this paper are characterized by three fields. The first is to examine the impacts of competitive intensity for recycling channel on optimal decision and profit. The second is to discuss the effectiveness conditions of capacity constraint in recycling channels for optimal decisions and profits. Finally, we propose a two-part tariff contract through bargaining to coordinate closed-loop supply chain and achieve a Pareto improvement.

Table 1: Notations and explanations.

| *Notation* | *Explanation* |
|---|---|
| $c_m$ | The production cost manufactured the new product via raw materials |
| $c_r$ | The production cost manufactured the manufactured product via used products |
| $s$ | The saving production cost between the new and the remanufactured product |
| $k$ | The scaling parameter for collection cost in a recycling channel |
| $\gamma$ | The competition intensity between two collection channels |
| $b$ | The transfer price from manufacturer to retailer to collect used product |
| $z_m$ | The available capacity constraint in the manufacturer's recycling channel |
| $z_r$ | The available capacity constraint in the retailer's recycling channel |
| $p$ | The retailer's selling price for unit product |
| $w$ | The manufacturer's wholesale price for unit product |
| $\tau_m$ | The manufacturer's collection rate for used production |
| $\tau_r$ | The retailer's collection rate for used production |
| $\pi_{sc}$ | The profit function for supply chain |
| $\pi_m$ | The profit function for manufacturer |
| $\pi_r$ | The profit function for retailer |
| $\theta_m$ | The bargaining power for manufacturer |
| $\theta_r$ | The bargaining power for retailer |

## 3. Problem Description and Assumption

In this paper, we investigate the decision and coordination of a closed-loop supply chain, which consists of one-single manufacturer and one-single retailer under capacity constraints in recycling channel. In the forward channel, the manufacturer sells a certain type of product through the retailer to consumers. Further, the manufacturer and the retailer collect used products in the reverse channel. Since the recycling channels for used products can differ, the demands are divided into two types from both the manufacturer and retailer, which are affected by the competition intensity between the two recycling channels. To discuss the decision behavior and coordination mechanism under capacity constraints in recycling channels, we suppose that the manufacturer follows the "lot-for-lot" policy, which is applied in the existing literature. Notation and explanations are listed in Table 1.

In addition, the following assumptions are considered in our mathematical models under the capacity constraints.

Table 2: Decisions with different strategies in the centralized scenario.

| *Strategy* | *Optimal decisions* $(p, \tau_m,$ *and* $\tau_r)$ |
|---|---|
| N-N-P | $p = \frac{k(a+c) - 2as^2(1-\gamma)}{2\left[k - s^2(1-\gamma)\right]}, \tau_m = \frac{s(1-\gamma)(a-c)}{2\left[k - s^2(1-\gamma)\right]}, \tau_r = \frac{s(1-\gamma)(a-c)}{2\left[k - s^2(1-\gamma)\right]}$ |
| N-N-F | $p = \frac{a+c-s}{2}, \tau_m = \frac{1}{2}, \tau_r = \frac{1}{2}$ |
| Y-N-P | $p = \frac{k(a+c-s\cdot z_r) - as^2(1-\gamma)}{2k - s^2(1-\gamma)}, \tau_m = \frac{s(1-\gamma)(a-c+s\cdot z_r)}{2k - s^2(1-\gamma)}, \tau_r = z_r$ |
| Y-N-F | $p = \frac{a+c-s}{2}, \tau_m = 1 - z_r, \tau_r = z_r$ |
| N-Y-P | $p = \frac{k(a+c-s\cdot z_m) - as^2(1-\gamma)}{2k - s^2(1-\gamma)}, \tau_m = z_m, \tau_r = \frac{s(1-\gamma)(a-c-s\cdot z_m)}{2k - s^2(1-\gamma)}$ |
| N-Y-F | $p = \frac{a+c-s}{2}, \tau_m = z_m, \tau_r = 1 - z_m$ |
| Y-Y-P | $p = \frac{a+c-s(z_m+z_r)}{2}, \tau_m = z_m, \tau_r = z_r$ |
| Y-Y-F | $p = \frac{a+c-s}{2}, \tau_m = z_m, \tau_r = z_r$ |

*Assumption 1.* Following the existing literature and many others [10, 18, 47], we suppose that there is no significant difference between new and remanufactured product and the market demand is a linear function $D = a - p$, which shows a trend of monotonically decreasing and continuous with respect to the selling price.

*Assumption 2.* Comparing the new and remanufactured product, we adopt $c$ and $c - s$ to characterize the production cost for two types of products, where the average production cost $\bar{c} = c(1 - \tau_m - \tau_r) + (c - s)(\tau_m + \tau_r)$ represents the saving production cost between the new and the remanufactured product. Thus, we derive the relationship $s < \bar{c} < c$ and $0 < b < s$, which ensure the collection behavior is profitable.

*Assumption 3.* Considering the recycling competition between manufacturer and retailer, we introduce the competition intensity into our research, which reflects the competition level between two channels in collecting used products. Therefore, the larger the competition intensity is, the more fierce the collection is. Consistent with Zou et al. [48], Xu et al. [49] and Jerbia et al. [50], we get a symmetric influence between two channels. We suppose there exists competition between the two collection channels. This paper refers to the competition intensity between two collection channels and it reflects intense competition level between two collection agents in collecting used products. The collection investments for both the manufacturer and the retailer are a quadratic functions with the collection rate, $k\left(\tau_m^2 + \gamma\tau_r^2\right)/2\left(1 - \gamma^2\right)$ and $k\left(\gamma\tau_m^2 + \tau_r^2\right)/2\left(1 - \gamma^2\right)$, which is widely used in the literature [13, 19, 51].

*Assumption 4.* In addition, the recycling quantities from the manufacturer and the retailer are set at a level much lower than the actual levels in collecting used products [5, 52, 53]. To some extent, this is intuitively consistent with reality since the actual conditions often incur a drop recycling quantity. Hence, we adopt $0 \le \tau_m \le z_m$, $0 \le \tau_r \le z_r$ and $0 \le \tau_m + \tau_r \le 1$ as the relationship to characterize the collection rates of manufacturer and retailer.

*Assumption 5.* According to Zhao and Zhu [38], Wang et al., [54], and Zhao and Zhu [55], we consider a symmetric-information Stackelberg game led by the manufacturer and followed by the retailer in closed-loop supply chain, in which the bargaining power of the retailer is limited compared with that of the manufacturer.

## 4. Model Equilibrium

In this section, we explore the decisions of pricing and collection rate for a closed-loop supply chain in the centralized scenario and decentralized scenario. Considering the capacity constraint, the manufacturer and the retailer have two options "*N*" or "*Y*," which mean that the quantities of used products from different recycling channels exceed the capacity constraint or not. Further, we denote "*P*" or "*F*" to illustrate that a part of used products or full of used products from the manufacturer and the retailer turns remanufacturing. We discuss eight strategies to obtain equilibriums in the centralized scenario and decentralized scenario.

*4.1. Centralized Scenario.* Initially, we establish a centralized model for a closed-loop supply chain and get the decisions of pricing and collection rate with capacity constraint in the recycling channel. Under this structure, the manufacturer and the retailer as a system aim to maximize the total profit. Therefore, the profit function for a supply chain is given as follows

$$\begin{aligned} \pi_{sc} &= \left[p - c + s(\tau_m + \tau_r)\right](a - p) - \frac{k\left(\tau_m^2 + \tau_r^2\right)}{2(1-\gamma)} \\ &\text{s.t.} \quad \tau_m \le z_m, \tau_r \le z_r, \tau_m + \tau_r \le 1. \end{aligned} \tag{1}$$

Table 3: Decisions with different strategies in the decentralized scenario.

| *Strategy* | *Optimal decisions* $(p, \tau_m,\ and\ \tau_r)$ |
|---|---|
| N-N-P | $p = \frac{k(3a+c) - a \cdot s^2(3-\gamma)\left(1-\gamma^2\right)}{4k - 3s^2(3-\gamma)\left(1-\gamma^2\right)}, w = \frac{2k(a+c) - s^2(2a-\gamma a+c)\left(1-\gamma^2\right)}{4k - 3s^2(3-\gamma)\left(1-\gamma^2\right)}, \tau_m = \frac{s(a-c)\left(1-\gamma^2\right)}{4k - 3s^2(3-\gamma)\left(1-\gamma^2\right)}, \tau_r = \frac{s(a-c)\left(1-\gamma^2\right)}{4k - 3s^2(3-\gamma)\left(1-\gamma^2\right)}$ |
| N-N-F | $p = \frac{k(3a+c-2s) + as^2(1-\gamma)(1+\gamma)^2}{4k + s^2(1-\gamma)(1+\gamma)^2}, w = \frac{2k(a+c-2s) - s^2\left(1-\gamma^2\right)(2a+\gamma a-c+2s)}{4k + s^2(1-\gamma)(1+\gamma)^2}, \tau_m = \frac{4k - s\left(1-\gamma^2\right)[a-c+s(1-\gamma)]}{4k + s^2(1-\gamma)(1+\gamma)^2}, \tau_r = \frac{s\left(1-\gamma^2\right)(a-c+2s)}{4k + s^2(1-\gamma)(1+\gamma)^2}$ |
| Y-N-P | $p = \frac{k(3a+c-s\cdot z_r) - as^2\left(1-\gamma^2\right)}{4k - s^2\left(1-\gamma^2\right)}, w = \frac{2k(a+c+s\cdot z_r) - s^2\left(1-\gamma^2\right)(a+s\cdot z_r)}{4k - s^2\left(1-\gamma^2\right)}, \tau_m = \frac{s\left(1-\gamma^2\right)(a-c+s\cdot z_r)}{4k - s^2\left(1-\gamma^2\right)}, \tau_r = z_r$ |
| Y-N-F | $p = \frac{3a+c-s}{4}, w = \frac{a+c-s(1-2z_r)}{2}, \tau_m = 1-z_r, \tau_r = z_r$ |
| N-Y-P | $p = \frac{k(3a+c-s\cdot z_m) - as^2(2-\gamma)\left(1-\gamma^2\right)}{4k - s^2(2-\gamma)\left(1-\gamma^2\right)}, w = \frac{2k(a+c-s\cdot z_m) - s^2\left(1-\gamma^2\right)(a-\gamma a-c+s\cdot z_m)}{4k - s^2(2-\gamma)\left(1-\gamma^2\right)}, \tau_m = z_m, \tau_r = \frac{s\left(1-\gamma^2\right)(a-c+s\cdot z_m)}{4k - s^2(2-\gamma)\left(1-\gamma^2\right)}$ |
| N-Y-F | $p = \frac{as\left(1-\gamma^2\right) - k(1-z_m)}{s\left(1-\gamma^2\right)}, w = \frac{s\left(1-\gamma^2\right)[a+s(1-z_m)] - 2k(1-z_m)}{s\left(1-\gamma^2\right)}, \tau_m = z_m, \tau_r = 1-z_m$ |
| Y-Y-P | $p = \frac{3a+c-s(z_m+z_r)}{4}, w = \frac{a+c-s(z_m-z_r)}{2}, \tau_m = z_m, \tau_r = z_r$ |
| Y-Y-F | $p = \frac{3a+c-s}{4}, w = \frac{a+c-s(1-2z_r)}{2}, \tau_m = z_m, \tau_r = z_r$ |

Table 4: Optimal decisions and profits in different scenarios.

| | $\theta_m$ | $F$ | $p$ | $w$ | $\tau_m$ | $\tau_r$ | $\pi_m$ | $\pi_r$ | $\pi_{sc}$ |
|---|---|---|---|---|---|---|---|---|---|
| *Centralized scenario* | — | — | 5.703 | — | 0.462 | 0.45 | — | — | 16.146 |
| *Decentralized scenario* | — | — | 7.843 | 6.062 | 0.313 | 0.313 | 7.947 | 4.105 | 12.052 |
| *Coordination contract* | 0.5 | 6.725 | — | 5.703 | 0.462 | 0.45 | 10.567 | 5.579 | 16.146 |
| | 0.6 | 6.430 | — | 5.703 | 0.462 | 0.45 | 10.861 | 5.285 | 16.146 |
| | 0.7 | 6.135 | — | 5.703 | 0.462 | 0.45 | 11.156 | 4.990 | 16.146 |
| | 0.8 | 5.840 | — | 5.703 | 0.462 | 0.45 | 11.451 | 4.695 | 16.146 |
| | 0.9 | 5.545 | — | 5.703 | 0.462 | 0.45 | 11.746 | 4.400 | 16.146 |
| | 1.0 | 5.250 | — | 5.703 | 0.462 | 0.45 | 12.041 | 4.105 | 16.146 |

**Proposition 1**. *In the centralized scenario, the equilibrium can be expressed as in Table* 2.

Proof of Proposition 1 is in supplementary materials (available here). Proposition 2 demonstrates the following results: (1) neither the manufacturer's nor retailer's collection rates are not affected by the capacity constraints under the conditions $k > \max(k_1, k_2)$, where $k_1 = s(1-\gamma)(a - c + 2s \cdot z_m)/2z_m$ and $k_2 = s(1-\gamma)(a - c + 2s \cdot z_r)/2z_r$; (2) both the manufacturer's and the retailer's collection rates are closely associated with the capacity constraints under the conditions $k < \min(k_3, k_4)$, where $k_3 = s(1-\gamma)[a - c + s(z_m + z_r)]/2z_m$ and $k_4 = s(1-\gamma)[a - c + s(z_m + z_r)]/2z_r$; (3) only the manufacturer's collection rate is affected by the capacity constraints under the condition $k_1 < k < k_3$; (4) only the retailer's collection rate is affected by the capacity constraints under the conditions $k_2 < k < k_4$.

*4.2. Decentralized Scenario.* Next, we consider the manufacturer and the retailer are independent, with the aim of maximizing one's profit. According to the Stackelberg game, the manufacturer first determines the wholesale price and collection rate. Then the retailer decides the selling price and collection rate. Therefore, the profit functions for the manufacturer and the retailer are respectively given as follows:

$$\pi_m = [w - c + s(\tau_m + \tau_r) - b\tau_r](a - p) - \frac{k(\tau_m^2 + \gamma\tau_r^2)}{2(1-\gamma^2)}. \quad (2)$$

$$\pi_r = (p - w + b\tau_r)(a - p) - \frac{k(\gamma\tau_m^2 + \tau_r^2)}{2(1-\gamma^2)} \quad \text{s.t.} \quad \tau_m \le z_m, \tau_r \le z_r, \tau_m + \tau_r \le 11 \quad (3)$$

**Proposition 2**. *In the decentralized scenario, the equilibrium can be expressed as in Table* 3.

Proof of Proposition 2 is in supplementary materials (available here). Proposition 2 demonstratesthe following results: (1) neither the manufacturer's nor retailer's collection rates are not affected by the capacity constraints under the conditions $k > \max(k_5, k_6)$, where $k_5 = s(1-\gamma^2)[a - c + sz_m(3-\gamma)](4z_m)$ and $k_6 = s(1-\gamma^2)[a - c + sz_r(3-\gamma)]/4z_r$; (2) both the manufacturer's and retailer's collection rates are closely associated with the capacity constraints under the conditions $k < \min(k_7, k_8)$, where $k_7 = s(1-\gamma^2)[1 - c + s(z_m + z_r)]/4z_m$ and $k_8 = s(1-\gamma^2)[a - c + s(z_m + z_r)]/4z_r$; (3) only the manufacturer's collection rate is affected by capacity constraints under the condition $k_5 < k < k_7$; (4) only the retailer's collection rate is affected by capacity constraints under the conditions $k_6 < k < k_8$.

Next, we compare the equilibriums in the decentralized scenario and centralized scenarios to investigate the differences in optimal performance.

**Proposition 3**. *From the above equilibrium, the following orders can be obtained:* $p^C \le p^D, \tau_m^C \ge \tau_m^D, \tau_r^C \ge \tau_r^D$ *and* $\pi_{sc}^C \ge \pi_{sc}^D$.

This relationship can be derived through algebraic comparison. From the above, it indicates that the centralized scenario results in a lower selling price and a higher collection rate, while the minimal total profit occurs under the decentralized scenario without coordination and the maximal total profit occurs under the centralized scenario. Further, a higher saving production cost *s and lower competition intensity k* trigger the manufacturer and the retailer to engage in remanufacturing. In addition, the centralized supply chain is hard to be carried out due to the independence for the manufacturer and the retailer to promote the maximization of supply chain profit. Therefore, the gaps between the optimal performances provide a substantial motivation to coordinate the members. Hence, we will propose a contract to solve the effect of double marginalization in the supply chain, which is caused by profit conflict. The collection strategy adopted by manufacturer and the retailer depends on the values of capacity constraint in the centralized and decentralized scenarios.

*4.3. Coordinated Contract.* In reality, it is difficult to make an agreement possible to be accepted by both the manufacturer and the retailer, which is explained that the upstream and downstream as an independent entity cannot transfer their own decision-making authority and execute centralized scenario. In this section, we provide a side-payment self-enforcing contract to coordinate the members with capacity constraint in recycling channels. The condition for a closed-loop supply chain to achieve the performances of a centralized scenario is negotiated in an agreement, which makes the allocation of profit fair and reasonable [7, 56, 57].

In a coordinated contract, we design a two-part tariff contract through bargaining to achieve a Pareto improvement and eliminate the effect of double marginalization. According to the above, we assume that the retailer sells products with the wholesale price set by the manufacturer and the sub-game

Table 5: Sensitivity analysis for the performances considering capacity constraints.

| *Fluctuation of* $\gamma$ | −50% | −37.5% | −25% | −12.5% | 0% | 12.5% | 25% | 37.5% | 50% |
|---|---|---|---|---|---|---|---|---|---|
| *Centralized scenario* | Case 2 | Case 2 | Case 5 | Case 5 | Case 5 | Case 1 | Case 1 | Case 1 | Case 1 |
| $p$ | 5.65 | 5.65 | 5.662 | 5.683 | 5.703 | 5.737 | 5.778 | 5.819 | 5.859 |
| $\tau_m$ | 0.579 | 0.55 | 0.529 | 0.496 | 0.462 | 0.428 | 0.393 | 0.359 | 0.326 |
| $\tau_r$ | 0.45 | 0.45 | 0.45 | 0.45 | 0.45 | 0.428 | 0.393 | 0.359 | 0.326 |
| $\pi_{sc}$ | 16.704 | 16.579 | 16.442 | 16.298 | 16.146 | 15.988 | 15.831 | 15.678 | 15.527 |
| $\lambda_1$ | 0 | 0 | + | + | + | 0 | 0 | 0 | 0 |
| $\lambda_2$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\lambda_3$ | + | + | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Decentralized scenario* | Case 1 | Case 1 | Case 1 | Case 1 | Case 1 | Case 1 | Case 1 | Case 1 | Case 1 |
| $p$ | 7.809 | 7.815 | 7.823 | 7.832 | 7.843 | 7.856 | 7.868 | 7.883 | 7.898 |
| $w$ | 6.039 | 6.043 | 6.049 | 6.055 | 6.062 | 6.070 | 6.079 | 6.088 | 6.099 |
| $\tau_m$ | 0.352 | 0.344 | 0.335 | 0.325 | 0.313 | 0.300 | 0.285 | 0.269 | 0.252 |
| $\tau_r$ | 0.352 | 0.344 | 0.335 | 0.325 | 0.313 | 0.300 | 0.285 | 0.269 | 0.252 |
| $\pi_m$ | 8.137 | 8.095 | 8.042 | 8.036 | 7.947 | 7.892 | 7.835 | 7.775 | 7.715 |
| $\pi_r$ | 4.259 | 4.224 | 4.286 | 4.146 | 4.105 | 4.063 | 4.020 | 3.976 | 3.933 |
| $\pi_{sc}$ | 12.396 | 12.319 | 12.328 | 12.182 | 12.052 | 11.955 | 11.855 | 11.751 | 11.648 |
| $\lambda_1$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\lambda_2$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\lambda_3$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

equilibrium is equivalent to that in the centralized scenario. Further, the manufacturer pays a fixed fee $F$ to the retailer, which is negotiated by the members' bargaining powers to supply chain coordination. The bargaining powers for manufacturer and retailer are $\theta_m$ and $\theta_r$. Therefore, the model can be characterized as follows

$$\pi_m = [w - c_m + c_s(\tau_m + \tau_r) - b\tau_r](a - w) - \frac{k\left(\tau_m^2 + \gamma\tau_r^2\right)}{2\left(1-\gamma^2\right)} - F. \tag{4}$$

$$\begin{aligned} \pi_r &= F - \frac{k\left(\gamma\tau_m^2+\tau_r^2\right)}{2\left(1-\gamma^2\right)} \\ \text{s.t.} \quad & \tau_m \le z_m, \tau_r \le z_r, \tau_m + \tau_r \le 1. \end{aligned} \tag{5}$$

**Proposition 4**. *In this coordinated contract, the fixed fee should be satisfied as* $F = \theta_m\left[\pi_r^D + C(\tau_r^*)\right] + \theta_r\left(\pi_{sc}^C - \pi_m^D\right)$.

Proof of Proposition 4 is in supplementary materials (available here) discusses the two-part tariff contract coordinates effectively the performances in the decentralized scenario to achieve the best in the centralized scenario. Meanwhile, it also obtains a Pareto improvement, which indicates that the individual members' profits with the improved two-part tariff contract will be improved.

**Proposition 5.** *The two-part tariff contract through bargaining can effectively coordinate the decentralized closed-loop supply chain and achieves the performances similar to that in the centralized supply chain as follows* $w^T = p^C$, $\tau_m^T = \tau_m^C$, $\tau_r^T = \tau_r^C$ *and* $\pi_{sc}^T = \pi_{sc}^C$.

Proposition 5 indicates that the two-part tariff can improve the performances of a closed-loop to that of a centralized scenario. The manufacturer and the retailer negotiate the fixed fee $F$ which should ensure the members' profits are not lower than those in the decentralized scenario. From the retailer's and manufacturer's profit function, the retailer benefits from the manufacturer's payment and the manufacturer covers the input in saving the production cost. Further, the lower the selling price is, the more consumers purchase the more products. In addition, it implies the two-part tariff contract through bargaining is beneficial for economic development and environmental protection in the sustainable operation of a closed-loop supply chain.

## 5. Numerical Analysis

In this section, we provide a numerical example to discuss theoretical results in managerial insights and analyze the influence of relevant parameters on optimal performance. Considering the values of coefficients used in the existing literature [13, 18, 58–60], we suppose that: $a = 10$, $\gamma = 0.35$, $c = 2.5$, $s = 1.2$, $k = 7.25$, $z_m = 0.6$, and $z_r = 0.45$.

*5.1. Case Analysis.* According to the above analysis, the optimal decisions and profits for each effective case in the centralized and decentralized scenarios are shown in Table 4.

From Table 4, we can obtain that:

(1) In the centralized supply chain, the optimal profit is obtained when $\lambda_1 > 0$, $\lambda_2 = 0$, and $\lambda_3 = 0$ since the KKT conditions should be satisfied. The supply chain can achieve maximal profit when the manufacturer's collection rate does not exceed the capacity constraint in the recycling channel and the retailer's collection rate exceeds the capacity constraint in the recycling channel. The optimal collection strategy is to collect the used product through the manufacturer since the capacity constraint from the manufacturer is greater

Table 6

(a) Sensitivity analysis for $z_m$ on the optimal decision and profit.

| *Fluctuation of* $z_m$ | −60% | −45% | −30% | −15% | 0% | 15% | 30% | 45% | 60% |
|---|---|---|---|---|---|---|---|---|---|
| *Centralized scenario* | Case 3 | Case 7 | Case 7 | Case 5 | Case 5 | Case 5 | Case 5 | Case 5 | Case 5 |
| $p$ | 5.837 | 5.782 | 5.728 | 5.703 | 5.703 | 5.703 | 5.703 | 5.703 | 5.703 |
| $\tau_m$ | 0.24 | 0.33 | 0.42 | 0.462 | 0.462 | 0.462 | 0.462 | 0.462 | 0.462 |
| $\tau_r$ | 0.448 | 0.45 | 0.45 | 0.45 | 0.45 | 0.45 | 0.45 | 0.45 | 0.45 |
| $\pi_{sc}$ | 15.888 | 16.055 | 16.137 | 16.146 | 16.146 | 16.146 | 16.146 | 16.146 | 16.146 |
| $\lambda_1$ | 0 | + | + | + | + | + | + | + | + |
| $\lambda_2$ | + | + | + | 0 | 0 | 0 | 0 | 0 | 0 |
| $\lambda_3$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Decentralized scenario* | Case 3 | Case 1 | Case 1 | Case 1 | Case 1 | Case 1 | Case 1 | Case 1 | Case 1 |
| $p$ | 7.867 | 7.843 | 7.843 | 7.843 | 7.843 | 7.843 | 7.843 | 7.843 | 7.843 |
| $w$ | 6.106 | 6.062 | 6.062 | 6.062 | 6.062 | 6.062 | 6.062 | 6.062 | 6.062 |
| $\tau_m$ | 0.24 | 0.313 | 0.313 | 0.313 | 0.313 | 0.313 | 0.313 | 0.313 | 0.313 |
| $\tau_r$ | 0.310 | 0.313 | 0.313 | 0.313 | 0.313 | 0.313 | 0.313 | 0.313 | 0.313 |
| $\pi_m$ | 7.929 | 7.947 | 7.947 | 7.947 | 7.947 | 7.947 | 7.947 | 7.947 | 7.947 |
| $\pi_r$ | 4.069 | 4.105 | 4.105 | 4.105 | 4.105 | 4.105 | 4.105 | 4.105 | 4.105 |
| $\pi_{sc}$ | 12.098 | 12.052 | 12.052 | 12.052 | 12.052 | 12.052 | 12.052 | 12.052 | 12.052 |
| $\lambda_1$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\lambda_2$ | + | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\lambda_3$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

(b) Sensitivity analysis for $z_r$ on the optimal decision and profit.

| *Fluctuation of* $z_r$ | −60% | −45% | −30% | −15% | 0% | 15% | 30% | 45% | 60% |
|---|---|---|---|---|---|---|---|---|---|
| *Centralized scenario* | Case 5 | Case 5 | Case 5 | Case 5 | Case 5 | Case 1 | Case 1 | Case 1 | Case 1 |
| $p$ | 5.876 | 5.832 | 5.789 | 5.746 | 5.703 | 5.694 | 5.694 | 5.694 | 5.694 |
| $\tau_m$ | 0.444 | 0.448 | 0.453 | 0.458 | 0.462 | 0.463 | 0.463 | 0.463 | 0.463 |
| $\tau_r$ | 0.18 | 0.248 | 0.315 | 0.383 | 0.45 | 0.463 | 0.463 | 0.463 | 0.463 |
| $\pi_{sc}$ | 15.731 | 15.906 | 16.033 | 16.113 | 16.146 | 16.147 | 16.147 | 16.147 | 16.147 |
| $\lambda_1$ | + | + | + | + | + | 0 | 0 | 0 | 0 |
| $\lambda_2$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\lambda_3$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Decentralized scenario* | Case 5 | Case 5 | Case 1 | Case 1 | Case 1 | Case 1 | Case 1 | Case 1 | Case 1 |
| $p$ | 7.893 | 7.962 | 7.843 | 7.843 | 7.843 | 7.843 | 7.843 | 7.843 | 7.843 |
| $w$ | 6.182 | 6.221 | 6.062 | 6.062 | 6.062 | 6.062 | 6.062 | 6.062 | 6.062 |
| $\tau_m$ | 0.293 | 0.296 | 0.313 | 0.313 | 0.313 | 0.313 | 0.313 | 0.313 | 0.313 |
| $\tau_r$ | 0.18 | 0.248 | 0.313 | 0.313 | 0.313 | 0.313 | 0.313 | 0.313 | 0.313 |
| $\pi_m$ | 7.734 | 7.857 | 7.947 | 7.947 | 7.947 | 7.947 | 7.947 | 7.947 | 7.947 |
| $\pi_r$ | 3.810 | 3.774 | 4.105 | 4.105 | 4.105 | 4.105 | 4.105 | 4.105 | 4.105 |
| $\pi_{sc}$ | 11.544 | 11.631 | 12.052 | 12.052 | 12.052 | 12.052 | 12.052 | 12.052 | 12.052 |
| $\lambda_1$ | + | + | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\lambda_2$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\lambda_3$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

than that of the retailer. However, in the decentralized scenario, the collection rates from the manufacturer and the retailer are equal because of the value of the capacity constraint in the two channels.

(2) Compared with the decentralized scenario, the centralized supply chain can result in improved performances. From this point, the selling price is lower in the centralized scenario than that in the decentralized scenario, whereas the collection rates and profit are higher. Hence, this means that the centralized system can help to enhance the overall efficiency of the supply chain. In addition, the contract can coordinate the members in the profit's distribution, which means that the profits of the manufacturer and the retailer are higher than those of the decentralized supply chain, and the selling price is lower and the collection rates are higher.

(3) The two-part tariff contract through bargaining has a significant implication for improving the economic performance and environmental benefit. Therefore, the optimal profits are improved to achieve the performance of a centralized scenario and create a win–win situation for the manufacturer and the retailer through the coordinated contract if the incentive compatibility constraint is guaranteed. Moreover, the profit of supply chain in the coordinated contract is constant. When the manufacturer and the retailer integrate as a whole system, the fluctuations of bargaining powers do not influence the total profit. However, under this coordinated contract, the manufacturer's profit and retailer's profit increase as the one's own bargaining power increases. Intuitively, we find that the capacity of recycling channel plays a critical role in the decision and coordination from the above case analysis. From a managerial viewpoint, we suggest that the closed-loop supply chain can integrate the capacity constraints in both recycling channels, to fulfill consumers' demand with a cost-effective method.

*5.2. Sensitivity Analysis.* To illustrate the effects of competitive intensity in a recycling channel on optimal performance, we provide the impact of competitive intensity on performance. Therefore, we assume that the maximal fluctuation of $\gamma$ is ±50% of the baseline values, the results of which are presented in Table 5.

From Table 5, we can conclude that:

(1) In the centralized scenario, changes in the optimal decisions and profit are larger (more than 3.6% and 7.2%) when the value of $\gamma$ fluctuates in the range of [−50%, 50%]. This means that the selling price and collection rate are sensitive to the competitive intensity in the centralized scenario. Moreover, the selling price appears as monotonic increasing trend and the profits show a monotonic decreasing trend with the competitive intensity in recycling channels. However, the difference in the collection rates for members first increases and then decreases with competitive intensity increasing.

(2) In the decentralized scenario, the difference between the collection rates for the manufacturer and the retailer is more insensitive to the fluctuation of $\gamma$, which shows that the manufacturer and the retailer achieve the same collection rates. This is because, compared with the recycling channels in a no competitive situation, the larger the competition intensity is, the more expanse in collection has been invested in to obtain a same collection rate. Therefore, the decreasing collection rate with the competition intensity increasing results in a rise in the production cost, which directly leads to the higher wholesale price and selling price.

Comparing the above situations, we evaluate how the effect of competitive intensity determines the operational performance of the whole supply chain system by measuring the capacity constraint of recycling channels. Interestingly, due to the double marginal effect, the centralized supply chain is more adaptable to the changes in capacity constraints. In other words, centralized supply is more conducive to the improvement of performance between the manufacturer and the retailer to achieve a higher profit. Further, we demonstrate the effect of capacity constraint in recycling channel on the optimal decisions and profits in closed-loop supply chain and give a sensitivity analysis for the capacity constraint. Considering the maximal fluctuations of $z_m$ and $z_r$ are respectively ±60% of the baseline values, the calculation results are as presented in Table 6.

From Table 6, we can obtain that:

(1) In the centralized scenario, when $z_m$ decreases by 30%, the total profit for the supply chain drops by at most 1.6% and the selling price rises by at most 2.3%, the collection rates decrease by at most 49.1% and 0.6% respectively. Meanwhile, the total profit for the supply chain drops by at most 2.6% and the selling price rises by at most 3.1%, and the collection rates decreases by at most 4.3% and 61.3% when $z_r$ decreases by 30%. Obviously, this means that the selling price is more sensitive than the total profit for the supply chain to capacity constraint in recycling channels. Further, the manufacturer's collection rate is more affected and the retailer's collection rate is not much affected by $z_m$, whereas the situation in the change of $z_r$ demonstrates the opposite. Additionally, the optimal decisions and profits will change if the capacity constraint in recycling channels exceeds a certain threshold.

(2) In the decentralized scenario, when $\lambda_1 = 0, \lambda_2 = 0$, and $\lambda_3 = 0$, the optimal decision and profit are not affected by the change of $z_m$ and $z_r$. However, the collection rate for the manufacturer shows a significant declining trend and the collection rate for the retailer is negligible with the value of $z_m$ when $\lambda_1 = 0$, $\lambda_2 > 0$ and $\lambda_3 = 0$; this situation is the contrary with the change of $z_r$ when $\lambda_1 = 0$, $\lambda_2 > 0$ and $\lambda_3 = 0$. It indicates that the members in the decentralized scenario result in a double marginalization effect, that is, the members determines a higher selling price and a lower collection rate. Therefore, only if the capacity constraint in recycling channels has a huge reduction can the decision and profit be effected. This indicates that the government should provide a more attention to ensure the recycling capacity for members in closed-loop supply chain. Specifically, the capacity constraints in recycling channels may cause a smoothing effect in the fabrication of both the manufacturing ability and the market demand.

## 6. Conclusions

This paper studies the decision and coordination in a closed-loop supply chain considering capacity constraints in recycling channels, which the manufacturer and retailer should determine the optimal selling prices and collection rate to balance their profits. Comparing the performances in the centralized scenario and decentralized scenario, we obtain the following results; (1) The centralized scenario results in a lower selling price and a higher collection rate, while the minimal profit occurs under the decentralized scenario without coordination and the maximal profit occurs under the centralized scenario. (2) A higher saving production cost and lower competition intensity trigger the manufacturer and the retailer to engage in remanufacturing. To coordinate closed-loop supply chain, we propose a two-part tariff contract and combine the bargain theory to achieve a Pareto improvement. In addition, the numerical analysis discussed the following managerial insights; (1) The optimal decisions and profits will change when the capacity constraint in recycling channel exceeds a certain threshold. Further, the selling price is more sensitive than the total profit for supply chain to capacity constraint in recycling channels. Hence, the closed-loop supply chain should find ways, through environmental propaganda and remanufacturing technologies to expand the capacity of recycling channels. (2) The optimal collection strategy is to collect the used product through the manufacturer since the capacity constraint from the manufacturer is greater than that of the retailer. Moreover, the difference in the centralized scenario between the collection rates from the manufacturer and the retailer are not equal to that of the centralized scenario because of the value of the capacity constraint in the two channels. Obviously, it is significantly important that the cooperation between the manufacturer and the retailer avoids the double marginal effect. (3) Specifically, the capacity constraints in recycling channels create a smoothing effect in the fabrication of both the manufacturing ability and market demand. Further, the two-part tariff contract through bargaining has a significant implication for improving the economic performance and environmental benefit, which improves the optimal profit to achieve the performance of centralized scenario and make a win–win situation for members via the coordinated contract if the incentive compatibility constraint is guaranteed.

This paper does not consider the uncertainty of market demand and asymmetric information, which are the future research to explore the equilibrium of a closed-loop supply chain with capacity constraints. In addition, introducing the insight of government into the decision and coordination for a closed-loop supply chain is another significant topic.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This research was supported in part by the National Nature Science Foundation of China (Nos. 71373157 and 51879156).

## Supplementary Materials

The proof for propositions. (*Supplementary Materials*)

## References

[1] A. S. Margarete and P. Ken, "Meeting the closed-loop challenge: The case of remanufacturing," *California Management Review*, vol. 46, no. 2, pp. 74–89, 2004.

[2] J.-M. Chen and C.-I. Chang, "The coopetitive strategy of a closed-loop supply chain with remanufacturing," *Transportation Research Part E: Logistics and Transportation Review*, vol. 48, no. 2, pp. 387–400, 2012.

[3] R. Giuntini and K. Gaudette, "Remanufacturing: the next great opportunity for boosting US productivity," *Business Horizons*, vol. 46, no. 6, pp. 41–48, 2003.

[4] J. Heydari, K. Govindan, and R. Sadeghi, "Reverse supply chain coordination under stochastic remanufacturing capacity," *International Journal of Production Economics*, vol. 202, pp. 1–11, 2018.

[5] D. Mohammaditabar, S. H. Ghodsypour, and A. Hafezalkotob, "A game theoretic analysis in capcity-constrained supplier-selection and cooperation by considering the total supply chain inventory costs," *International Journal of Production Economics*, vol. 181, pp. 87–97, 2016.

[6] H. H. Liu, M. Lei, H. H. Deng, G. Keong Leong, and T. Huang, "A dual channel, quality-based price competition model for the WEEE recycling market with government subsidy," *Omega*, vol. 59, pp. 290–302, 2016.

[7] L. Xu and C. X. Wang, "Sustainable manufacturing in a closed-loop supply chain considering emission reduction and remanufacturing," *Resources, Conservation and Recycling*, vol. 131, pp. 297–304, 2018.

[8] L. Xu, F. Xie, Q. Yuan, and J. Chen, "Pricing and carbon footprint in a two-echelon supply chain under cap-and-trade regulation," *International Journal of Low-carbon Technologies*, vol. 14, no. 2, pp. 212–221, 2019.

[9] Q. Zhu, Y. Feng, and S.-B. Choi, "The role of customer relational governance in environmental and economic performance improvement through green supply chain management," *Journal of Cleaner Production*, vol. 155, pp. 46–53, 2018.

[10] R. C. Savaskan, S. Bhattacharya, and L. N. Wassenhove, "Closed-loop supply chain models with product remanufacturing," *Management Science*, vol. 50, pp. 239–252, 2004.

[11] R. C. Savaskan and L. N. Van Wassenhove, "Reverse Channel Design: The Case of Competing Retailers," *Management Science*, vol. 50, no. 2, pp. 1–14, 2006.

[12] D. S. Wu, "Coordination of competing supply chains with news-vendor and buyback contract," *International Journal of Production Economics*, vol. 144, no. 1, pp. 1–13, 2013.

[13] M. Huang, M. Song, L. H. Lee, and W. K. Ching, "Analysis for strategy of closed-loop supply chain with dual recycling channel," *International Journal of Production Economics*, vol. 144, no. 2, pp. 510–520, 2013.

[14] C.-T. Zhang and M.-L. Ren, "Closed-loop supply chain coordination strategy for the manufacture of patented products

under competitive demand," *Applied Mathematical Modelling*, vol. 40, no. 13–14, pp. 6243–6255, 2016.

[15] G. A. Morteza, G. Seyed, T. M. Reza, and A. Jabbarzadeh, "A fuzzy pricing model for a green competitive closed-loop supply chain network design in the presence of disruptions," *Journal of Cleaner Production*, vol. 188, pp. 425–442, 2018.

[16] A. A. Taleizadeh, N. Alizadeh-Basban, and B. R. Sarker, "Coordinated contracts in a two-echelon green supply chain considering pricing strategy," *Computers & Industrial Engineering*, vol. 124, pp. 249–275, 2018.

[17] L. P. Feng, K. Govindan, and C. F. Li, "Strategic planning: design and coordination for dual-recycling channel reverse supply chain considering consumer behavior," *European Journal of Operational Research*, vol. 260, no. 2, pp. 601–612, 2017.

[18] Q. D. He, N. M. Wang, Z. Yang, Z. He, and B. Jiang, "Competitive collection under channel inconvenience in closed-loop supply chain," *European Journal of Operational Research*, vol. 275, no. 1, pp. 155–166, 2019.

[19] L. W. Liu, Z. J. Wang, L. Xu, X. Hong, and K. Govindan, "Collection effort and reverse channel choices in a closed-loop supply chain," *Journal of Cleaner Production*, vol. 144, pp. 492–500, 2017.

[20] J. Wang, Z. Zhou, and M. Yu, "Pricing models in a sustainable supply chain with capacity constraint," *Journal of Cleaner Production*, vol. 222, pp. 57–76, 2019.

[21] L. Sereno and T. Efthimiadis, "Capacity constraints, transmission investments and incentive schemes," *Energy policy*, vol. 119, pp. 8–27, 2018.

[22] M. Fischetti, I. Ljubic, and M. Sinnl, "Benders decomposition without separability: a computational study for capacitated facility location problems," *European Journal of Operational Research*, vol. 253, no. 3, pp. 557–569, 2016.

[23] B. Mota, M. I. Gomes, A. Carvalho, and A. P. Barbosa-Povoa, "Sustainable supply chains: an integrated modelling approach under uncertainty," *Omega*, vol. 77, pp. 32–57, 2018.

[24] N. M. Wang, Q. D. He, and B. Jiang, "Hybrid closed-loop supply chains with competition in recycling and product markets," *International Journal of Production Economics*, 2019.

[25] R. Dominguez, B. Ponte, S. Cannella, and M. J. Framinan, "On the dynamics of closed-loop supply chains with capacity constraints," *Computers & Industrial Engineering*, vol. 128, pp. 91–103, 2019.

[26] L. Zhen, Y. Wu, S. Wang, Y. Hu, and W. Yi, "Capacitated closed-loop supply chain network design under uncertainty," *Advanced Engineering Informatics*, vol. 38, pp. 306–315, 2018.

[27] I. Ljubic and E. Moreno, "Outer approximation and submodular cuts for maximum capture facility location problems with random utilities," *European Journal of Operational Research*, vol. 266, no. 1, pp. 45–56, 2018.

[28] G. P. Cachon and M. A. Lariviere, "Supply chain coordination with revenue-sharing contracts: strengths and limitations," *Management Science*, vol. 51, no. 1, pp. 30–44, 2005.

[29] F. Mafakheri and F. Nasiri, "Revenue sharing coordination in reverse logistics," *Journal of Cleaner Production*, vol. 59, pp. 185–196, 2013.

[30] B. Hu and Y. Feng, "Optimization and coordination of supply chain with revenue sharing contracts and service requirement under supply and demand uncertainty," *International Journal of Production Economics*, vol. 183, pp. 185–193, 2017.

[31] J. Heydari, T. M. Choi, and S. Radkhah, "Pareto improving supply chain coordination under a money-back guarantee service program," *Service Science*, vol. 9, no. 2, pp. 91–105, 2017.

[32] S. K. Chaharsooghi, J. Heydari, and I. N. Kamalabadi, "Simultaneous coordination of order quantity and reorder point in a two-stage supply chain," *Computers & Operations Research*, vol. 38, no. 12, pp. 1667–1677, 2011.

[33] Z. Dimitris, I. George, and B. Apostolos, "Supply chain coordination under discrete information asymmetries and quantity discounts," *Omega*, vol. 53, pp. 21–29, 2015.

[34] J. Heydari, "Coordinating supplier's reorder point: a coordination mechanism for supply chains with long supplier lead time," *Computers & Operations Research*, vol. 48, pp. 89–101, 2014.

[35] J. Heydari, Z. A. Payam, and T. M. Choi, "Coordinating supply chains with stochastic demand by crashing lead times," *Computers & Operations Research*, vol. 100, pp. 394–403, 2018.

[36] J. Heydari and M. Ghasemi, "A revenue sharing contract for reverse supply chain coordination under stochastic quality of returned products and uncertain remanufacturing capacity," *Journal of Cleaner Production*, vol. 197, pp. 607–615, 2018.

[37] S. Li, Z. Zhu, and L. Huang, "Supply chain coordination and decision making under consignment contract with revenue sharing," *International Journal of Production Economics*, vol. 120, no. 1, pp. 88–99, 2009.

[38] S. Zhao and Q. Zhu, "Remanufacturing supply chain coordination under the stochastic remanufacturability rate and the random demand," *Annals of Operations Research*, vol. 257, no. 1–2, pp. 661–695, 2017.

[39] J. Xie, L. Liang, L. Liu, and P. Ieromonachou, "Coordination contracts of dual-channel with cooperation advertising in closed-loop supply chains," *International Journal of Production Economics*, vol. 183, pp. 528–538, 2017.

[40] J. G. Shi and G. D. Wu, "Study on the supply chain alliance profit allocation based on improved Shapely value," *International Conference on Management Science*, vol. 50, pp. 7–512, 2009.

[41] B. Dai and H. X. Chen, "Profit allocation mechanism for carrier collaboration in pickup and delivery service," *Computers & Industrial Engineering*, vol. 62, no. 2, pp. 633–643, 2012.

[42] Y. Zhang and H. Geng, "An analysis on distribution of cooperative profit in supply chain based on multi-agent," *International Conference on Medical Physics and Biomedical Engineering*, vol. 33, pp. 698–704, 2012.

[43] W. Y. Chen, B. Kucukyazici, V. Verter, and M. J. Sáenz, "Supply chain design for unlocking the value of remanufacturing under uncertainty," *European Journal of Operational Research*, vol. 247, no. 3, pp. 804–819, 2015.

[44] Q. Wu, H. Ren, W. Gao, J. Ren, and C. Lao, "Profit allocation analysis among the distributed energy network participants based on game-theory," *Energy*, vol. 118, pp. 783–794, 2017.

[45] C. Wang, W. Wang, and R. Huang, "Supply chain enterprise operations and government carbon tax decisions considering carbon emissions," *Journal of Cleaner Production*, vol. 152, pp. 271–280, 2017.

[46] J. Xie, W. Zhang, L. Liang, Y. Xia, J. Yin, and G. Yang, "The revenue and cost sharing contract of pricing and servicing policies in a dual-channel closed-loop supply chain," *Journal of Cleaner Production*, vol. 191, pp. 361–383, 2018.

[47] X. Hong, Z. Wang, D. Wang, and Z. Zhang, "Decision models of closed-loop supply chain with remanufacturing under hybrid dual-channel collection," *International Journal of Advanced Manufacturing Technology*, vol. 68, no. 5–8, pp. 1851–1865, 2013.

[48] Z.-B. Zou, J.-J. Wang, G.-S. Deng, and H. Chen, "Third-party remanufacturing mode selection: outsourcing or authorization,"

*Transportation Research Part E: Logistics and Transportation Review*, vol. 87, pp. 1–19, 2016.

[49] L. Xu, C. Wang, Z. Miao, and J. Chen, "Governmental subsidy policies and supply chain decisions with carbon emission limit and consumer's environmental awareness," *RAIRO-Operations Research*, vol. 53, pp. 1675–1690, 2019.

[50] R. Jerbia, M. K. Boujelben, M. A. Sehli, and Z. Jemai, "A stochastic closed-loop supply chain network design problem with multiple recovery options," *Computers & Industrial Engineering*, vol. 118, pp. 23–32, 2018.

[51] Y. Li, L. Xu, and D. Li, "Examining relationships between the return policy, product quality, and pricing strategy in online direct selling," *International Journal of Production Economics*, vol. 144, no. 2, pp. 451–460, 2013.

[52] S. Cannella, R. Dominguez, B. Ponte, and J. M. Framinan, "Capacity restrictions and supply chain performance: Modelling and analyzing load-dependent lead times," *International Journal of Production Economics*, vol. 204, pp. 264–277, 2018.

[53] J. Wu, F. Jiang, and Y. He, "Pricing and horizontal information sharing in a supply chain with capacity constraint," *Operations Research Letters*, vol. 46, pp. 402–408, 2018.

[54] Y. Wang, X. Ma, Z. Li, Y. Liu, M. Xu, and Y. Wang, "Profit distribution in collaborative multiple center vehicle routing problem," *Journal of Cleaner Production*, vol. 144, pp. 203–219, 2017.

[55] S. L. Zhao and Q. H. Zhu, "A risk-averse marketing strategy and its effect on coordination activities in a remanufacturing supply chain under market fluctuation," *Journal of Cleaner Production*, vol. 171, pp. 1290–1299, 2018.

[56] J. H. Gao, H. S. Han, L. T. Hou, and H. Wang, "Pricing and effort decisions in a closed-loop supply chain under different channel power structures," *Journal of Cleaner Production*, vol. 112, pp. 2043–2057, 2016.

[57] D. Ghosh and J. Shah, "Supply chain analysis under green sensitive consumer demand and cost sharing contract," *International Journal of Production Economics*, vol. 164, pp. 319–329, 2015.

[58] J. Xu, Y. Chen, and Q. Bai, "A two-echelon sustainable supply chain coordination under cap-and-trade regulation," *Journal of Cleaner Production*, vol. 135, pp. 42–56, 2016.

[59] X. P. Hong, L. Xu, P. Du, and Wenjuan Wang, "Joint advertising, pricing and collection decisions in a closed-loop supply chain," *International Journal of Production Economics*, vol. 167, pp. 12–22, 2015.

[60] L. Xu, C. Wang, and J. Zhao, "Decision and coordination in the dual-channel supply chain considering cap-and-trade regulation," *Journal of Cleaner Production*, vol. 197, pp. 551–561, 2018.

## Author(s) Name(s)

It is very important to confirm the author(s) last and first names in order to be displayed correctly on our website as well as in the indexing databases:

**Author 1**
Given Names: Lang
Last Name: Xu

**Author 2**
Given Names: Jia
Last Name: Shi

**Author 3**
Given Names: Jihong
Last Name: Chen

It is also very important for each author to provide an ORCID (Open Researcher and Contributor ID). ORCID aims to solve the name ambiguity problem in scholarly communications by creating a registry of persistent unique identifiers for individual researchers.

To register an ORCID, please go to the Account Update page (http://mts.hindawi.com/update/) in our Manuscript Tracking System and after you have logged in click on the ORCID link at the top of the page. This link will take you to the ORCID website where you will be able to create an account for yourself. Once you have done so, your new ORCID will be saved in our Manuscript Tracking System automatically.

Hindawi
Complexity
Volume 2019, Article ID 8398356, 16 pages
https://doi.org/10.1155/2019/8398356

*Research Article*

# Dynamic Matching in Cloud Manufacturing considering Matching Costs

**Qi Chen, Qi Xu, and Cui Wu**

*Glorious Sun School of Business and Management, Donghua University, Shanghai 200051, China*

Correspondence should be addressed to Qi Xu; xuqi@dhu.edu.cn

Received 4 April 2019; Revised 8 June 2019; Accepted 17 June 2019; Published 4 July 2019

Guest Editor: Julio Blanco-Fernández



As a service-oriented business platform model, the nature of cloud manufacturing is to realise the manufacturing resources' sharing, which will largely benefit resources supplier, resources demander, and platform operator. However, it also faces some new problems. One of the most critical issues is how to dynamically match resources of supply and demand to maximise profits of all parties while considering matching costs. This paper investigates the resources' dynamic matching in a manufacturing supply chain that operates under a cost-sharing contract and consists of two independent and competing manufacturers and a resource-service platform. We first use differential equation to model the evolution of resource-sharing and capture the effect of matching service efforts on market demand. Next, we study the optimal matching strategies by a two-stage differential game based on the dynamic control approach. Then, we design a cost-sharing contract to coordinate and improve the supply chain's performance. Finally, a numerical example is provided to illustrate the impact of platform transaction fees and matching costs on the feasible region of the corresponding contract.

## 1. Introduction

Information technologies, such as the Internet of Things, cloud computing, and cyber-physical systems, impact daily life through their powerful data-processing capacities. For example, e-commerce has become an indispensable means of shopping over the past two decades, and a small number of giant e-commerce companies, such as Amazon, eBay, and Alibaba, have emerged to dominate the market. However, the Internet also enables other types of transaction, such as sharing. Online networks facilitate the sharing of computing and manufacturing resources in supply chain. This has resulted in collaborative consumption and collaborative production: peer-to-peer exchanges for obtaining, providing, or sharing access to goods and services, facilitated by community-based online platforms. The manufacturing industry is undergoing a major transformation enabled by cloud computing. The main thrust of cloud computing is to provide on-demand computing services with high reliability, scalability, and availability in a distributed environment. Cloud technologies have had profound impacts on production management in manufacturing [1]. O'Rourke [2] stated that new information technologies have helped to drive the development of 'lean manufacturing', with factories using such technologies being better equipped to rapidly deliver the products that customers want. As information technologies become embedded in all aspects of production, 'network-centric' manufacturing advances throughout value chains and each element becomes 'smart', thereby optimising efficiency throughout a product's life-cycle [3].

Learning from cloud computing, researchers have proposed a model of 'cloud manufacturing', in which uniform manufacturing resources are shared through online networking. In this model, manufacturing capabilities and resources are shared via a cloud platform. The status of idle resources is updated and released in real time to facilitate online transactions and identify the most sustainable and robust manufacturing route possible [4]. Figure 1 presents a simplified model of the common features of cloud manufacturing. The cloud manufacturing architecture defines three common roles (although the exact nomenclature for each role varies in the literature): the supplier (which offers services or resources on the platform), the demander (which requests services or resources through the cloud), and the

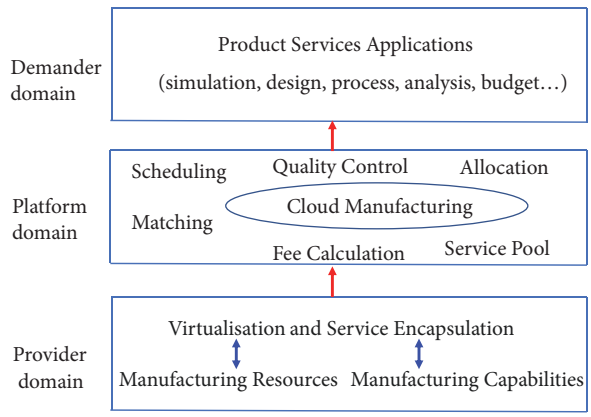



Figure 1: Cloud manufacturing architecture, adapted from [1, 5, 9].

platform manager [5–7]. The demander utilises resources or services for manufacturing purposes and the supplier provides these resources or services by renting, leasing or lending equipment or other resources for short-term periods. The cloud platform manages the use, performance, and delivery of services and negotiates the relationship between supply and demand; it acts as an intermediary, providing connectivity and transport to enable the exchange of services between consumers and providers [8]. In this regard, cloud manufacturing and e-commerce share some similarities, the main difference being that commodities are traded on an e-commerce platform whereas manufacturing services are exchanged on a manufacturing platform. In the reallocation process, idle manufacturing resources and capabilities are connected through cloud computing and other information technologies, eventually forming a supply chain for manufacturing resource sharing.

In China, Shandong province has built an 'industrial cloud platform' that incorporates regional factories and technological resources. Factories and individuals can request various services and resources through the platform at low rental costs. The platform has already had economic and environmental benefits [4]. The transition from traditional manufacturing to a service-oriented model occurred gradually, as lessons from the 'sharing economy' were adapted to the manufacturing sector. Unlike in the sharing economy, in which users share consumer products, the platform facilitates the sharing of idle manufacturing resources. A number of platforms have already implemented business models that closely resemble cloud manufacturing. For example, MFG.com, the world's largest contract manufacturing marketplace, provides a fast and efficient platform for exchanging manufacturing resources. Similarly, 1688.com, China's leading e-commerce platform for domestic small enterprise trading, adopted similar strategies for sharing manufacturing resources. As of 2018, 1688.com's business model covered 16 industries and a wide range of supply services, from raw materials to industrial products, clothing, apparel, and household items. Manufacturing resource sharing has obvious benefits for resources supplier, resources demander and platform operator. However, it also introduces new management challenges. One of the most critical issues is optimising the dynamic matching of supply and demand to maximise cooperation between the various parties while considering matching costs.

The goal of matching is to connect consumer demand to the right products or services. To improve matching, all parties in the supply chain (supplier, demander, and platform manager) must invest in the matching effort. As Figure 1 shows, each party in the cloud manufacturing system incurs a distinct set of matching costs [9, 10]: (i) the supplier (the resource or service provider) incurs service-realisation costs, i.e., the cost of updating the platform to reflect the current status (availability and quality) of the resources, services, and capabilities; (ii) the platform manager incurs aggregation and generation costs, i.e., the costs of computing, storage, and scheduling; and (iii) the platform demander incurs invocation costs related to business operations, i.e., consultation, market analysis and investigation, purchase, insurance, etc. Optimising the allocation of resources and services for the supply chain is complex because it requires ensuring that the supplier, demander, and platform manager each benefit. In the process of reallocating supply chain resources, how to integrate, share, and optimise the allocation of supply chain resources so that the resources provider, cloud platform, and resources consumers can get the greatest benefits is an important issue faced by supply chain enterprises. The aim of this paper is therefore to identify matching strategies that can achieve this optimal solution.

There has been extensive research on performance analysis and supply-demand matching for manufacturing resources and services. In cloud manufacturing, operators use searching and matching algorithms to find suitable services to satisfy users' requests. Several resource-service discovery frameworks are described in the literature. Tao et al. [11] proposed a four-phase method for resource-service matching and searching on service-oriented manufacturing system platforms. A genetic algorithm based model to search for the result that best matches a customer's request is proposed in Zhang et al. [12]. Based on grey correlation theory, a machine tool supply-demand matching method is proposed in Xiao et al. [13]. Wang [14] investigated the cloud manufacturing resource discovery mechanism and proposed a manufacturing resource discovery framework based on the Semantic Web. Capturing user requirements and cloud services matching are important steps for realising on-demand resource-service provision that require the semantic description of manufacturing tasks. Wang et al. [15] investigated the semantic modelling and description of manufacturing tasks in cloud manufacturing system for manufacturing task to be better to match with manufacturing services. Li et al. [16] proposed a multilevel intelligent matching method to realise rapid, efficient and accurate matching. Yin et al. [17] proposed an input, output, precondition, effect matching model based on Web Ontology Language for cloud manufacturing. The model's matching process is divided into three phases: parameter matching, attribute matching, and comprehensive matching. Li et al. [18] proposed an intelligent service searching and matching method of cloud manufacturing according to service type and state information.

The abovementioned studies have mainly examined issues of matching and scheduling with static manufacturing tasks and static candidate resource services in a given period. The dynamic changes typical of the practical process of supply-demand matching and scheduling have not been considered. Cheng et al. [19] proposed a supply-demand matching hypernetwork of manufacturing services, comprising a manufacturing service network, a manufacturing task network, and hyperedges between those two networks. Subsequently, based on the results in [19], Cheng et al. [20] formulated a model for revealing the matchable correlations between each service (supply) and each task (demand), subject to dynamic demand. Cloud manufacturing systems contain many dynamic elements. The number of users and the number of manufacturing tasks change dynamically. Additionally, in an environment of distributed resources, the relative independence of various economic entities also leads to dynamic changes in the sharing relationship. Cheng at al [19, 20] only consider the dynamic complexity caused by changes in the numbers of users and manufacturing tasks. They analysed supply-demand matching in cloud manufacturing from a technical perspective but neglected operations management concerns. From the latter perspective, the goal of matching is to connect consumer demand to the right products or services. This generally involves facilitating information exchange between a supplier and a demander. As matching becomes more successful, sharing increases. To improve matching, all parties in the supply chain (supplier, demander, and platform manager) must invest in the matching effort. However, this investment becomes an issue as the platform's matching abilities improve. Crucially, when the number of sharing transactions on the platform increases, the matching costs also increase. Matching costs have not been considered in previous studies. Thus, our study has an important difference from the abovementioned studies, which is that we investigate the complex relations and conflicts of interest arising from the sharing of resources through cloud manufacturing from the perspective of operations management.

Game theory is a powerful theoretical tool for analysing conflict and cooperation behaviour among rational individuals and, as such, can be useful for optimising manufacturing resource sharing and management, from locating services and supply-demand matching to transactions [21]. Games can be either cooperative or noncooperative depending on whether parties share a formal agreement. Enterprises have variously competitive and cooperative relationships, depending on the functional dependency of their services or products. The service composition in cloud manufacturing should therefore ensure the functional realisation of composite services while guaranteeing that each enterprise profits. Game theory is uniquely suited to this type of problem. The key to apply game theory in service composition in cloud manufacturing is to design proper utility functions for each enterprise by comprehensively considering their service attributes (including economic attributes) and constructing gaming models or mechanisms for the appropriate service interactions [22]. Differential games offer a promising approach. For example, De Giovanni [23] and Amrouche et al. [24] developed differential game models to incorporate channel dynamics. Here, we use differential equations to model the dynamic evolution of manufacturing resource sharing and capture the effect of matching efforts on market demand. By applying optimal control theory, we derive matching strategies for both centralised and decentralised systems. We also design a cost-sharing contract to improve the performance of the decentralised supply chain. Finally, we use a numerical example to examine the feasibility and efficacy of platform transaction fees and other parameters as strategies for optimising the coordination contract. The paper makes three primary contributions, which can be summarised as follows. First, we investigate operational problems for a sharing supply chain from a dynamic matching perspective. Second, we design a coordination contract for the supply chain by accounting for the impact of resource-sharing levels, which can be used to coordinate the decentralised system in dynamic environments. Finally, to the best of our knowledge, our study is the first to explore supply-demand matching issues by applying optimal control theory and game theory to derive optimal solutions.

The study proceeds in six sections. In Section 2, we give descriptions of the notations and assumptions used throughout the paper. Section 3 provides the theoretical results for the optimal strategies under a decentralised decision scenario, Section 4 provides the theoretical results for the optimal strategies under a centralised decision scenario, and Section 5 provides the theoretical results for the optimal strategies under a coordination-contract scenario. The numerical results and sensitivity analyses are represented in Section 6. Finally, Section 7 concludes the study and discusses its implications for management.

## 2. Problem Description and the Basic Model

*2.1. Problem Formulation.* We consider a supply chain formed of two independent and competing manufacturers, labelled $d$ and $s$, and a resource-service platform, labelled $p$, in which manufacturer $s$ (i.e., the supplier) has surplus manufacturing resources, whereas manufacturer $d$ (i.e., the demander) lacks such resources. The platform has a strong reputation and the supplier sells its manufacturing resources to the demander through the platform. Ultimately, the two manufacturers produce homogeneous products and sell them to consumers. Deciding the optimal efforts for matching to enhance sharing is the primary objective of the players, which wish to increase demand and subsequently profits by adopting the optimal operational strategies. A simplified channel structure of the sharing supply chain is presented in Figure 2.

Table 1 provides the notation used throughout the study.

Supply-demand matching within the supply chain is a complex issue. The level at which manufacturing resources are shared within a dynamic framework can be investigated using the following equation:

$$\begin{aligned} R(t)' &= \left\{\alpha A_s(t) + \beta A_p(t) + \gamma A_d(t)\right\} - \varphi R(t), \\ R(0) &= R_0 \geq 0, \end{aligned} \tag{1}$$

Table 1: Notation and descriptions.

| | |
|---|---|
| Variable | |
| $A_s(t)$ | matching effort of supplier |
| $A_d(t)$ | matching effort of demander |
| $A_p(t)$ | matching effort of platform |
| $R(t)$ | manufacturing resources sharing level |
| $\varepsilon$ | the platform's support rate |
| $V$ | the value function |
| $J$ | profit |
| Parameter | |
| $\pi_s$ | the margin profit of supplier |
| $\pi_d$ | the margin profit of demander |
| $c$ | fees from demander to platform |
| $\omega$ | purchasing cost of demander |
| $\rho$ | discount rate |

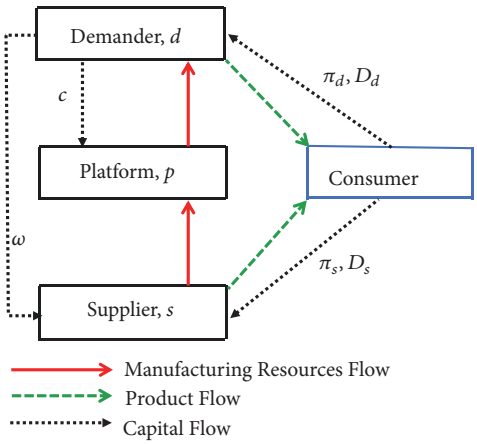



Figure 2: Channel structure of the sharing supply chain.

where $\alpha$, $\beta$, and $\gamma$ represent the marginal contribution of matching efforts to the sharing level, which we call *matching effectiveness*, and $\varphi$ is the sharing level's decay rate. Matching effectiveness captures the relationship between each supply chain member's investment in matching and the sharing level. The sharing level's decay rate might reflect several scenarios; for instance, it could suggest that a manufacturing resource needs to be improved due to an increase in the number of product categories or attributes, which in turn can result in a decrease in the sharing level over time.

Each supply chain member's matching costs are convex and increasing, indicating that the matching efforts' marginal costs increase and are assumed to be quadratic:

$$C\left(A_s(t)\right) = \frac{\mu_s}{2}{A_s}^2(t),$$
$$C\left(A_p(t)\right) = \frac{\mu_p}{2}{A_p}^2(t), \quad (2)$$
$$C\left(A_d(t)\right) = \frac{\mu_d}{2}{A_d}^2(t),$$

where $\mu_s$, $\mu_d$, and $\mu_p$ are the positive cost parameters. This cost function is commonly applied in the literature [23–25].

The level of manufacturing resource sharing has a positive external spill-over effect on the supply chain's supplier and demander. Customer demand depends on both the marginal profit and the level at which manufacturing resources are being shared (i.e., the sharing level). The demand functions can be expressed as follows:

$$D_s(R(t), t) = a - \pi_s + \theta(\pi_d - \pi_s) + \eta_s R(t), \quad (3)$$
$$D_d(R(t), t) = a - \pi_d + \theta(\pi_s - \pi_d) + \eta_d R(t), \quad (4)$$

where $a$ represents the potential market size, $\theta > 0$ denotes cross-price sensitivity between the two manufacturers, and $\eta > 0$ represents the effects of the sharing level on market demand. This is similar to the demand functions used in [24, 26–30], which depict the substitution effect between two independent and competing manufacturers.

*2.2. The Objective Function.* Assuming an infinite time horizon and a positive discount rate$\rho$, the objective functions are

$$J_s = \max_{A_s} \int_0^\infty e^{-\rho t}\left\{\pi_s D_s(t) - C\left(A_s(t)\right) + \omega R(t)\right\} dt, \quad (5)$$

$$J_p = \max_{A_p} \int_0^\infty e^{-\rho t}\left\{cR(t) - C\left(A_p(t)\right)\right\} dt, \quad (6)$$

$$J_d = \max_{A_d} \int_0^\infty e^{-\rho t}\left\{\pi_d D_d(t) - C\left(A_d(t)\right) - (\omega + c) R(t)\right\} dt. \quad (7)$$

To recapitulate, (1), (5), (6), and (7) define a differential game with three players, three control variables $A_s(t)$, $A_d(t)$, and $A_p(t)$, and one state variable $R(t)$. The controls are constrained by $A_s(t) \geq 0$, $A_d(t) \geq 0$, and $A_p(t) \geq 0$. The state constraint $R(t) \geq 0$ is automatically satisfied. We assume that the game is played à la Stackelberg, with the platform acting as the leader and the two manufacturers as followers (see [24, 25] for examples of the Stackelberg differential game).

## 3. The Optimal Strategies in the Decentralised System

We start by analysing the first scenario, in which the players implement a noncooperative program. Under decentralised decision-making, the supplier, platform, and demander maximise their own profits, respectively. The platform is the channel leader and does not offer subsidies to the demander. We use the superscript '$N$' to signify the decentralised system scenario.

The supply chain game can be conceptualised in two stages. In the first stage, the platform decides the matching efforts $A_p(t)$. In the second, both the supplier and demander make their decisions, respectively. In particular, the supplier determines the matching efforts $A_s(t)$ and the demander determines the matching efforts $A_d(t)$. The sequence of the events is shown in Figure 3.

From this point forward, the time argument is omitted. Let ${V_s}^N$, ${V_p}^N$, and ${V_d}^N$ denote the players' value

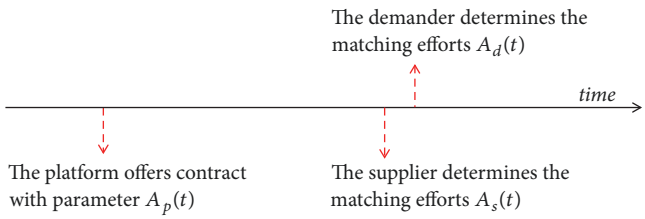



Figure 3: Sequence of events under the decentralised scenario.

functions. To obtain the optimal dynamic matching policy, we follow the literature [24, 25, 31] and use the Hamiltonian–Jacobi–Bellman (HJB) equations:

$$\rho V_s^N(R) = \max_{A_s \geq 0} \left\{ \pi_s D_s - C(A_s) + \omega R + V_s^{N\prime} \left( \alpha A_s + \beta A_p + \gamma A_d - \varphi R \right) \right\}, \tag{8}$$

$$\rho V_p^N(R) = \max_{A_p \geq 0, \varepsilon \geq 0} \left\{ cR - C(A_p) + V_p^{N\prime} \left( \alpha A_s + \beta A_p + \gamma A_d - \varphi R \right) \right\}, \tag{9}$$

$$\rho V_d^N(R) = \max_{A_d \geq 0} \left\{ \pi_d D_d - C(A_d) - (\omega + c) R + V_d^{N\prime} \left( \alpha A_s + \beta A_p + \gamma A_d - \varphi R \right) \right\}. \tag{10}$$

This puts us in a position to propose optimal strategies for the supply chain with decentralised decision-making. Proposition 1 characterises the equilibrium strategies.

**Proposition 1.** *In the decentralised system, the equilibrium results of the differential game between the supplier, platform, and demander are as follows.*

*(i) The equilibrium matching efforts are given by*

$$A_s^{N*} = \frac{\alpha(\pi_s \eta_s + \omega)}{\mu_s(\rho + \varphi)}, \tag{11}$$

$$A_d^{N*} = \frac{\gamma(\pi_d \eta_d - \omega - c)}{\mu_d(\rho + \varphi)}, \tag{12}$$

$$A_p^{N*} = \frac{\beta c}{\mu_p(\rho + \varphi)}. \tag{13}$$

*(ii) The manufacturing resource-sharing level in the supply chain is given by*

$$R^{N*} = K^N + \left( R_0 - K^N \right) e^{-\varphi t}, \tag{14}$$

*where the parameter* $K^N = \{\alpha^2 \mu_p \mu_d (\pi_s \eta_s + \omega) + \beta^2 \mu_s \mu_d c + \gamma^2 \mu_s \mu_p (\pi_d \eta_d - \omega - c)\} / (\mu_s \mu_p \mu_d \varphi(\varphi + \rho))$.

*(iii) The optimal profit functions for the supplier, platform, and demander are given by*

$$J_s^{N*} = e^{-\rho t} V_s^N \left( R^{N*} \right), \tag{15}$$

$$J_p^{N*} = e^{-\rho t} V_p^N \left( R^{N*} \right), \tag{16}$$

$$J_d^{N*} = e^{-\rho t} V_d^N \left( R^{N*} \right), \tag{17}$$

*where the parameters* $a_1^N, a_2^N, a_3^N$ *and* $b_1^N, b_2^N, b_3^N$ *are the coefficients of the linear value functions*

$$\begin{aligned} V_s^N \left( R^{N*} \right) &= a_1^Y R^{N*} + b_1^N \\ V_p^N \left( R^{Y*} \right) &= a_2^Y R^{N*} + b_2^N \\ V_d^N \left( R^{N*} \right) &= a_3^Y R^{N*} + b_3^N, \end{aligned} \tag{18}$$

*which are determined in the proof for Proposition 1 (see the Appendix).*

Proposition 1 shows that the sharing level $R^{N*}$ is positive. This means that all party members are involved in the supply chain. The matching efforts $A_s^{N*}$, $A_d^{N*}$, and $A_p^{N*}$ should be positive and decreasing at decay rate $\varphi$. In contrast, the matching efforts $A_s^{N*}$, $A_d^{N*}$, and $A_p^{N*}$ increase in terms of effectiveness parameters $\alpha$, $\beta$, and $\gamma$, respectively. This indicates that when the investment is efficient, the supplier, platform, and demander are motivated to invest more in supply-demand matching.

## 4. Optimal Strategies in the Centralised System

In this section, we examine the performance of a centralised supply chain. Supply chain members integrate to set the optimal matching efforts in view of maximising the total supply chain profit. In this game, $A_s(t)$, $A_d(t)$, and $A_p(t)$ are decision variables. We use the superscript '$I$' to signify the centralised decision scenario.

Assuming an infinite time horizon and a positive discount rate $\rho$, the objective function of the supply chain in the centralised system is given as

$$\begin{aligned} J_{sc} = \max \quad & \int_0^\infty e^{-\rho t} \left\{ \pi_s D_s(t) + \pi_d D_d(t) - C(A_s(t)) - C\left(A_p(t)\right) - C(A_d(t)) \right\} dt \\ s.t. \quad & R(t)' = \alpha A_s(t) + \beta A_p(t) + \gamma A_d(t) - \varphi R(t), \\ & R(0) = R_0. \end{aligned} \tag{19}$$

From this point forward, the time argument is omitted. Let ${V_{sc}}^I$ denote the supply chain system's value functions; the HJB equation is

$$\begin{aligned}\rho {V_{sc}}^I(R) = \max_{A_s,A_p,A_d} \Big\{ & \pi_s D_s + \pi_d D_d - C(A_s) \\ & - C(A_p) - C(A_d) \\ & + {V_{sc}}^{I'} \left(\alpha A_s + \beta A_p + \gamma A_d - \varphi x\right)\Big\}.\end{aligned} \tag{20}$$

We are now in a position to propose optimal strategies for the supply chain with centralised decision-making. Proposition 2 characterises the equilibrium strategies.

**Proposition 2.** *With centralised decision-making, the equilibrium results of the differential game between the supplier, the platform, and the demander are as follows.*

*(i) The equilibrium matching efforts are given by*

$${A_s}^{I*} = \frac{\alpha(\pi_s\eta_s + \pi_d\eta_d)}{\mu_s(\rho+\varphi)}, \tag{21}$$

$${A_d}^{I*} = \frac{\gamma(\pi_s\eta_s + \pi_d\eta_d)}{\mu_d(\rho+\varphi)}, \tag{22}$$

$${A_p}^{I*} = \frac{\beta(\pi_s\eta_s + \pi_d\eta_d)}{\mu_p(\rho+\varphi)}. \tag{23}$$

*(ii) The sharing level of manufacturing resources in the supply chain is given by*

$$R^{I*} = K^I + \left(R_0 - K^I\right)e^{-\varphi t}, \tag{24}$$

*where the parameter* $K^I = (\alpha^2\mu_p\mu_d + \beta^2\mu_s\mu_d + \gamma^2\mu_s\mu_p)(\pi_s\eta_s + \pi_d\eta_d)/(\mu_s\mu_p\mu_d\varphi(\varphi+\rho))$.

*(iii) The optimal profit function of the supply chain system is given by*

$${J_{sc}}^{I*} = e^{-\rho t}{V_{sc}}^I\left(R^{I*}\right), \tag{25}$$

*where the parameters* $a^I$ *and* $b^I$ *are the coefficients of the linear function* ${V_{sc}}^I(R^{I*}) = a^I R^{N*} + b^I$*, which are determined in the Proof of Proposition 2 (see the Appendix).*

**Proposition 3.** *Compared with optimal strategies and profit functions in the decentralised and centralised systems, one has* ${A_s}^{I*} > {A_s}^{N*}$, ${A_d}^{I*} > {A_d}^{N*}$, ${A_p}^{I*} > {A_p}^{N*}$, *and* ${J_{sc}}^{I*} > {J_{sc}}^{N*}$.

We provide the proof for Proposition 3 in the Appendix. These relationships are derived through algebraic comparison. The matching efforts are higher in the centralised system, which means that the total profit is lower in the decentralised system. Hence, there is a need to design an appropriate contract to improve system efficiency.

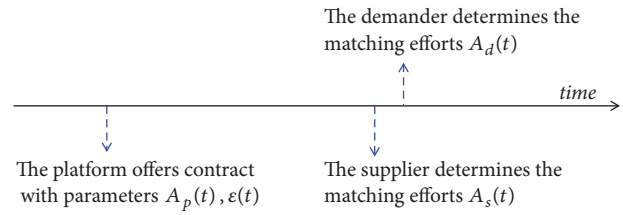



Figure 4: Sequence of the events under the coordination-contract scenario.

## 5. Optimal Strategies under the Coordination Contract

In this scenario, the platform is the channel leader and supports the demander's matching efforts. We use the superscript '$Y$' to signify the coordination-contract scenario. $\varepsilon(t)$ denotes the platform's support rate, which represents the amount that the platform contributes to the demander's matching efforts within the interval $[0,1]$. We are motivated by the coordination method used in [23] to develop a committed dynamic cost-sharing contract capable of coordinating the supply chain and improving the decentralised supply chain's performance. The contract provisions are structured as follows. In the game's first stage, the platform decides the matching efforts and the support rate $\varepsilon(t)$. In the second stage, both the supplier and demander make their decisions, respectively. In particular, the supplier determines the matching efforts $A_s(t)$ and the demander determines the matching efforts $A_d(t)$. The sequence of the events is shown in Figure 4.

Assuming an infinite time horizon and a positive discount rate $\rho$, the objective functionals of supply chain members under the coordination-contract scenario are

$$J_s^Y = \max_{A_s} \int_0^\infty e^{-\rho t}\left\{\pi_s D_s(t) - C(A_s(t)) + \omega R(t)\right\} dt, \tag{26}$$

$$\begin{aligned}J_p^Y = \max_{A_p,\varepsilon} \int_0^\infty e^{-\rho t} \Big\{ & cR(t) - C\left(A_p(t)\right) \\ & - \varepsilon(t) C\left(A_d(t)\right)\Big\} dt,\end{aligned} \tag{27}$$

$$\begin{aligned}J_d^Y = \max_{A_d} \int_0^\infty e^{-\rho t} \{ & \pi_d D_d(t) - (\omega + c) R \\ & - (1-\varepsilon(t)) C(A_d)\} dt.\end{aligned} \tag{28}$$

From this point forward, the time argument is omitted. Let ${V_s}^Y$, ${V_p}^Y$, and ${V_d}^Y$ denote the players' value functions; the HJB equations are

$$\begin{aligned}\rho {V_s}^Y(R) = \max_{A_s \geq 0} \Big\{ & \pi_s D_s - C(A_s) + \omega R \\ & + {V_s}^{Y'}\left(\alpha A_s + \beta A_p + \gamma A_d - \varphi R\right)\Big\},\end{aligned} \tag{29}$$

$$\begin{aligned}\rho {V_p}^Y(R) = \max_{A_p \geq 0, \varepsilon \geq 0} \Big\{ & cR - C\left(A_p\right) - \varepsilon C(A_d) \\ & + {V_p}^{Y'}\left(\alpha A_s + \beta A_p + \gamma A_d - \varphi R\right)\Big\},\end{aligned} \tag{30}$$

$$\begin{aligned}\rho {V_d}^Y(R) = \max_{A_d \geq 0} \Big\{ & \pi_d D_d - (\omega + c) R - (1-\varepsilon) C\left(A_d\right) \\ & + {V_d}^{Y'}\left(\alpha A_s + \beta A_p + \gamma A_d - \varphi R\right)\Big\}.\end{aligned} \tag{31}$$

We are now in a position to propose optimal strategies for the cost-sharing contract system. Proposition 4 characterises the equilibrium strategies.

**Proposition 4.** *Under the coordination-contract scenario, the equilibrium results of the differential game between the supplier, the platform, and the demander are as follows.*

*(i) The equilibrium matching efforts and platform's support rate are given by*

$$A_s^{Y*} = \frac{\alpha(\pi_s\eta_s + \omega)}{\mu_s(\rho + \varphi)}, \tag{32}$$

$$A_d^{Y*} = \frac{\gamma(\pi_d\eta_d - \omega + c)}{2\mu_d(\rho + \varphi)}, \tag{33}$$

$$A_p^{Y*} = \frac{\beta c}{\mu_p(\rho + \varphi)}, \qquad \varepsilon = \frac{-\pi_d\eta_d + \omega + 3c}{\pi_d\eta_d - \omega + c}. \tag{34}$$

*(ii) The sharing level of manufacturing resources in the supply chain is given by*

$$R^{Y*} = K^Y + \left(R_0 - K^Y\right)e^{-\varphi t}, \tag{35}$$

*where the parameter* $K^Y = (2\alpha^2\mu_p\mu_d(\pi_s\eta_s + \omega) + 2\beta^2 c\mu_s\mu_d + \gamma^2\mu_s\mu_p(\pi_d\eta_d - \omega + c))/2\mu_s\mu_p\mu_d\varphi(\varphi + \rho)$.

*(iii) The optimal profit functions of supply chain members are given by*

$$J_s^{Y*} = e^{-\rho t}V_s^Y\left(R^{Y*}\right), \tag{36}$$

$$J_p^{Y*} = e^{-\rho t}V_p^Y\left(R^{Y*}\right), \tag{37}$$

$$J_d^{Y*} = e^{-\rho t}V_d^Y\left(R^{Y*}\right), \tag{38}$$

*where the parameters* $a_1^Y$, $a_2^Y$, $a_3^Y$ *and* $b_1^Y$, $b_2^Y$, $b_3^Y$ *are the coefficients of the linear value functions*

$$V_s^Y\left(R^{Y*}\right) = a_1^Y R^{Y*} + b_1^Y$$
$$V_p^Y\left(R^{Y*}\right) = a_2^Y R^{Y*} + b_2^Y \tag{39}$$
$$V_d^Y\left(R^{Y*}\right) = a_3^Y R^{Y*} + b_3^Y,$$

*which are determined in the proof of Proposition 4 (see the Appendix).*

Next, we compare each supply chain member's profits and the total channel profits with the corresponding values in the above three scenarios. Our objective is to identify the effect of the cost-sharing contract on all channel members' profits to determine whether the cost-sharing contract increases profits and thus improves coordination. For notational convenience, let $K_1 = \max\{0, (\pi_d\eta_d - \omega)/3\}$ and $K_2 = \pi_d\eta_d - \omega$; the interval $(K_1, K_2)$ is the coordination contract's feasible region. We then arrive at the following proposition.

**Proposition 5.** *The strategies and payoffs in the decentralised scenario (N), cost-sharing contract scenario (Y), and centralised decision scenario (I) are related as follows:*

*(i) The supplier equilibrium matching efforts,* $A_s^{N*} = A_s^{Y*} < A_s^{I*}$.

*(ii) The platform equilibrium matching efforts,* $A_p^{N*} = A_p^{Y*} < A_p^{I*}$.

*(iii) The demander equilibrium matching efforts,* $A_d^{N*} < A_d^{Y*} < A_d^{I*}$.

*(iv) The optimal profits,* $J_s^{N*} < J_s^{Y*}$, $J_p^{N*} < J_p^{Y*}$, $J_d^{N*} < J_d^{Y*}$, *and* $J_{sc}^{N*} < J_{sc}^{Y*}$ *for* $K_1 < c < K_2$.

We provide the proof for Proposition 5 in the Appendix. These relationships are derived through algebraic comparison. Proposition 5 shows that all supply chain members incur higher profits in the cost-sharing contract scenario than the decentralised decision-making scenario. Clearly, cost-sharing with the platform provides the greatest benefit to the demander: when the platform manager covers any share of the matching costs, it helps improve the demander's profitability. As the matching costs are lowered, the demander can offer a higher level of matching effort, which subsequently drives up market demand for the resource or service. This increase in market demand more than compensates for the cost shared by the platform.

This result illustrates why matching involves increased collaboration between the demander and the platform manager through cost-sharing contracts and other mechanisms. However, because the comparison of the supplier, platform, demander, and supply chain profits poses some degree of analytical complexity, we now turn to numerical computation to verify our theoretical findings.

## 6. Numerical Example

In this section, we conduct numerical analyses to gain managerial insights. Set $\pi_s = 5$, $\pi_d = 5$, $\mu_s = 10$, $\mu_p = 15$, $\mu_d = 14$, $\eta_s = 0.9$, $\eta_d = 1.7$, $x_0 = 0.25$, $\alpha = 2$, $\beta = 2$, $\gamma = 3$, $\varphi = 0.5$, $\theta = 0.5$, $a = 5$, $c = 4$, $\omega = 0.6$, and $\rho = 0.9$. In Section 6.1, we compare the operational performance of the decentralised ($N$), cost-sharing contract ($Y$), and centralised decision ($I$) scenarios, focusing on the dynamic strategies, the sharing level and profits. In Section 6.2, we examine the impacts of the platform transaction fee and purchasing cost on the feasible region of the corresponding contract and obtain some useful insights.

Before we proceed, recall that the profit functions are linear in the value function $V$ and can be written as an exponential function multiplied by the value function, i.e., $J = e^{-\rho t}V$, which makes it sufficient for comparison. Thus, to compare $J_s^*$, $J_p^*$, $J_d^*$, and $J_{sc}^*$, we compare the values of $V_s^*$, $V_p^*$, $V_d^*$, and $V_{sc}^*$, respectively. Define $\Delta V_s = V_s^{Y*} - V_s^{N*}$, $\Delta V_p = V_p^{Y*} - V_p^{N*}$, $\Delta V_d = V_d^{Y*} - V_d^{N*}$ and $\Delta J_s = J_s^{Y*} - J_s^{N*}$, $\Delta J_p = J_p^{Y*} - J_p^{N*}$, $\Delta J_d = J_d^{Y*} - J_d^{N*}$. Similarly, a comparison between $\Delta V_s$, $\Delta V_p$, and $\Delta V_d$ is equivalent to a comparison between the profit functions $\Delta J_s$, $\Delta J_p$, and $\Delta J_d$, respectively.

Table 2: Optimal strategies in supply chain systems.

| | Decentralised (N) | Cost sharing (Y) | Centralised (I) |
|---|---|---|---|
| $A_s^*$ | 0.73 | 0.73 | 1.86 |
| $A_p^*$ | 0.41 | 0.41 | 1.99 |
| $A_d^*$ | 0.56 | 0.85 | 1.24 |

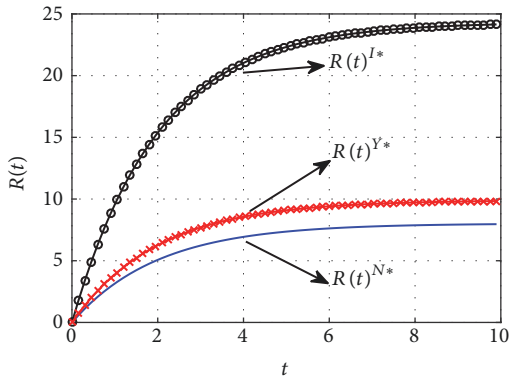



Figure 5: Dynamic change in manufacturing resources sharing level.

### 6.1. Optimal Solutions

*6.1.1. Comparisons of Optimal Solutions.* According to Propositions 1, 2, and 4, we can obtain the optimal matching efforts and sharing level in the decentralised, cost-sharing contract, and centralised decision scenarios. From Table 2, we can see that $A_s^{N*} = A_s^{Y*} < A_s^{I*}$, $A_p^{N*} = A_p^{Y*} < A_p^{I*}$, and $A_d^{N*} < A_d^{Y*} < A_d^{I*}$. The matching efforts in the centralised structure are higher than those in the decentralised scenario, and the optimal matching efforts in the decentralised scenario are equal to or less than those in the cost-sharing contract scenario. This is consistent with the conclusions in Proposition 5. Figure 5 shows changes to the sharing level over time. Here, the corresponding optimal resources sharing level are given as follows:

$$R(t)^* = \begin{cases} 8.02 - 5.52e^{-0.5t}, & \text{Decentralized } (N) \\ 9.9 - 7.4e^{-0.5t}, & \text{Cost sharing } (Y) \\ 24.32 - 21.82e^{-0.5t}, & \text{Centralized } (I). \end{cases} \quad (40)$$

Figure 5 shows that the optimal resource-sharing levels in the centralised decision-making system are higher than those in the decentralised and cost-sharing scenarios, as the matching efforts are higher in the centralised decision system.

*6.1.2. Comparison of Profits.* In this subsection, we compare the profits across the three models; we provide the results in Figure 6. The profit in the centralised decision scenario is the highest, followed by the cost-sharing contract and the decentralised scenario, respectively, which verifies Proposition 5.

Figure 7 presents a comparison between each supply chain member's profits before and after cost-sharing. The equilibrium values in the cost-sharing contract are in the

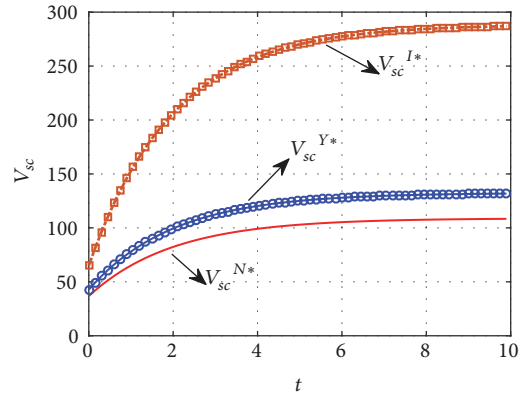



Figure 6: Optimal profit comparison in scenarios $N$, $Y$, and $I$.

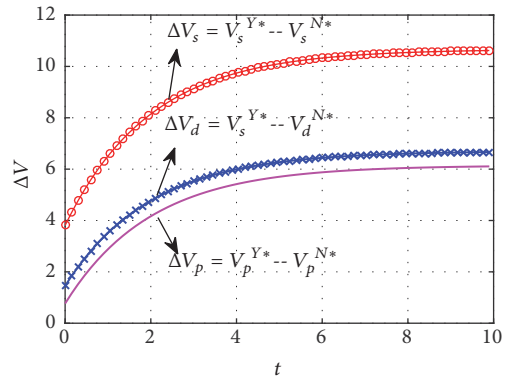



Figure 7: Profit comparison for supply chain parties before and after cost-sharing.

following order in comparison with the decentralised supply chain values: $V_s^{Y*} > V_s^{N*}$, $V_p^{Y*} > V_p^{N*}$, $V_d^{Y*} > V_d^{N*}$ for $K_1 < c < K_2$. This indicates that the supplier, platform manager, and demander all enjoy higher profits in the cost-sharing contract than in the decentralised supply chain case. The cost-sharing contract effectively improves the performance of the decentralised supply chain. The cost-sharing contract achieves Pareto improvement for the supplier, the platform manager, and the demander under certain conditions. Any share of matching costs helps improve the demander's profitability. As such, the demander can provide a higher matching effort; this increases market demand, which more than compensates for the cost shared by the platform.

Moreover, the supplier's profit increases are the highest, followed by the demander and the platform manager, respectively. In the cost-sharing contract scenario, the platform supports the demander's matching efforts and the supplier does not incur any additional matching costs. Cost-sharing lowers the demander's burden in the supply chain structure; the demander thus enjoys greater benefit from the matching decision (see Columns 1 and 2 in Table 2). A comparison of the platform's profit shows that the platform manager incurs higher profits in the cost-sharing contract than the decentralised scenario. The contract thus also benefits the platform. These results illustrate why matching involves

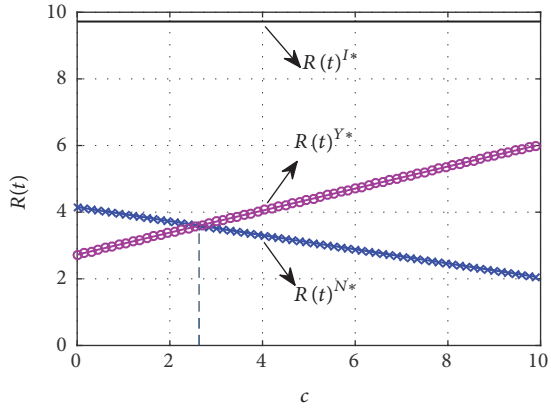



Figure 8: Impact of the transaction fee on the level of manufacturing resource sharing.

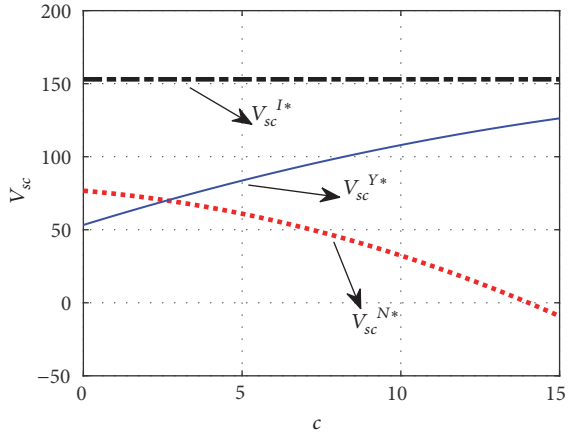



Figure 9: Impact of the transaction fee on optimal profit.

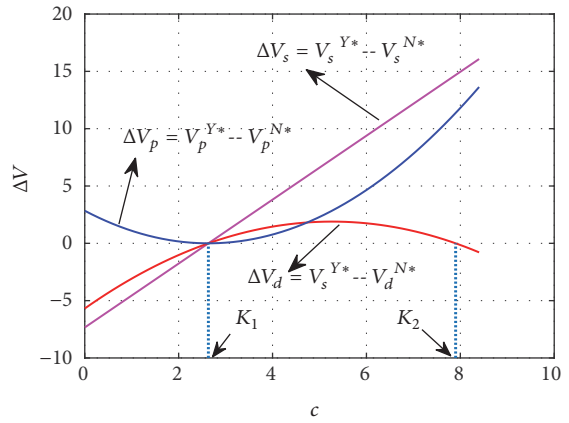



Figure 10: Pareto improvement effect of the transaction fee for the cost-sharing contract.

increased collaboration between the demander and the platform through cost-sharing and other mechanisms.

*6.2. Sensitivity Analysis of the Platform Transaction Fee (c).* We first investigate the effects of platform transaction fees on the sharing level. In Figure 8, the sharing level $R(t)^*$ is plotted as a function of the platform transaction fee $c$. The sharing level decreases as the transaction fee increases in the decentralised decision system due to the fact that the demander's marginal profit decreases with the increase of the transaction fee. As in Section 2.1, the demand function depends on the marginal profit and the sharing level. Thus, the larger the transaction fee, the smaller the market demand. Accordingly, in Figure 9, we see that the profit in the decentralised decision system decreases when the transaction fee $c$ is raised. In contrast, the sharing level increases in tandem with the platform transaction fee in the cost-sharing contract scenario because the platform's support rate $\varepsilon^*(t)$ increases with the transaction fee $c$ (see Proposition 4). The larger the transaction fee, the larger the support rate. As such, the demander has a greater incentive to increase its matching efforts; this drives up the market demand, which more than compensates for the cost shared by the platform. Accordingly, Figure 9 shows that in the cost-sharing contract scenario, profit increases with $c$.

However, we also find that the cost-sharing contract does not always achieve Pareto improvement for all parties (i.e., the value can fall outside the feasible region). Figure 10 shows that only when the value of $c$ is between $K_1$ and $K_2$ can the cost-sharing contract adequately coordinate the supply chain such that all parties benefit. Specifically, when the purchasing cost $\omega$ increases, the win-win region becomes smaller in Figure 11. This implies that, as the value of $\omega$ increases, the degree of flexibility in coordinating the supply chain decreases.

## 7. Conclusions

In this paper, we discussed the challenges of supply-demand matching for manufacturing resource- and service-sharing by considering the sharing level in a complex and dynamic environment. Applying optimal control theory, we identified the optimal matching strategies for decentralised, centralised and cost-sharing contract systems. Our main contribution lies in the following. First, we considered the dynamic evolution feature of the sharing level, which we set as a state variable. Second, we optimised matching-effort strategies through differential game models to coordinate the decentralised supply chain. Finally, we conducted a numerical analysis to illustrate the effect of the platform transaction fee and purchasing costs on equilibria and coordination.

In particular, we obtained the following results. (1) A cost-sharing contract effectively improves the performance of the decentralised supply chain. All channel members (i.e., the manufacturing resource or service supplier, platform manager, and resource or service demander) incur higher profits in the cost-sharing contract system than the decentralised system. (2) The cost-sharing contract does not always achieve Pareto improvement for all parties. (3) Numerical analysis shows that the platform transaction fee and purchasing costs affect the win-win region and optimal strategies. A larger purchasing cost will limit the degree of flexibility with which supply chain members coordinate the supply chain, thus providing manufacturers and the service platform with guidance to improve profitability. Our study contributes to

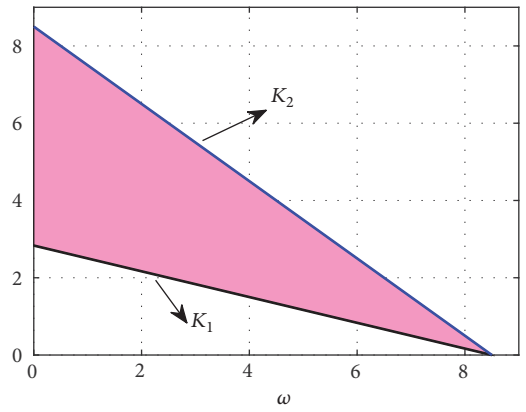



Figure 11: Impact of $\omega$ on the feasible region of the cost-sharing contract.

the burgeoning field of idle manufacturing resource sharing within supply chains and collaboration between channel partners.

## Appendix

### A.

*Proof of Proposition 1.* We need to establish the existence of bounded and continuously differentiable value functions ${V_s}^N$, ${V_p}^N$, and ${V_d}^N$ such that it is a unique solution $R^N$ to differential equation (1) and the HJB equations. We first determine the players' necessary conditions from the HJBs. Because the game is played à la Stackelberg and the platform is the leader, we first derive the decision variables for the supplier and the demander for the second game stage. The optimisation problem of supplier is given as

$$\begin{aligned} J_s^N = \max_{A_s} \quad & \int_0^\infty e^{-\rho t}\left\{\pi_s D_s(t) - C\left(A_s(t)\right) + \omega R(t)\right\} dt \\ s.t. \quad & R(t)' = \alpha A_s(t) + \beta A_p(t) + \gamma A_d(t) - \varphi R(t), \\ & R(0) = R_0. \end{aligned} \tag{A.1}$$

Let the value functions ${V_s}^N = \max_{A_s} \int_t^\infty e^{-\rho(\tau-t)}\{\pi_s D_s(t) - C(A_s(t)) + \omega R(t)\}d\tau$; the optimal profit function of manufacturing resources supplier then is given as

$$J_s^N = e^{-\rho t}{V_s}^N. \tag{A.2}$$

Similarly, the optimal profit function of the platform and demander are given by

$$\begin{aligned} J_p^N &= e^{-\rho t}{V_p}^N, \\ J_d^N &= e^{-\rho t}{V_d}^N. \end{aligned} \tag{A.3}$$

The supplier's HJB is

$$\rho {V_s}^N(R) = \max_{A_s \geq 0}\left\{\pi_s D_s - \frac{\mu_s {A_s}^2}{2} + \omega R + {V_s}^{N\prime} R'\right\}, \tag{A.4}$$

and its maximisation provides the necessary condition for matching efforts:

$${A_s}^N = \frac{\alpha {V_s}^{N\prime}}{\mu_s}. \tag{A.5}$$

Similarly, the demander's HJB is

$$\begin{aligned} &\rho {V_d}^N(R) \\ &= \max_{A_d \geq 0}\left\{\pi_d D_d - \frac{\mu_d {A_d}^2}{2} - (\omega + c) R + {V_d}^{N\prime} R'\right\}, \end{aligned} \tag{A.6}$$

and its maximisation provides the necessary condition for matching efforts:

$$A_d^N = \frac{\gamma {V_d}^{N\prime}}{\mu_d}. \tag{A.7}$$

Substituting (A.5) and (A.7) into the platform's HJB gives

$$\rho {V_p}^N(R) = \max_{F \geq 0}\left\{cR - \frac{\mu_p A_p^2}{2} + {V_p}^{N\prime} R'\right\}, \tag{A.8}$$

and by performing the maximisation of the right-hand side we obtain

$$A_p^N = \frac{\beta {V_p}^{N\prime}}{\mu_p}. \tag{A.9}$$

By inserting (A.5), (A.7), and (A.9) inside the HJBs, we obtain the following three algebraic equations:

$$\begin{aligned} \rho {V_s}^N(R) &= \pi_s\left(a - \pi_s + \theta\pi_d - \theta\pi_s + \eta_s R\right) \\ &\quad + \frac{\left(\alpha {V_s}^{N\prime}\right)^2}{(2\mu_s)} + \omega R \\ &\quad + {V_s}^{N\prime}\left(\frac{\beta^2 {V_p}^{N\prime}}{\mu_p} + \frac{\gamma^2 {V_d}^{N\prime}}{\mu_d} - \varphi R\right), \end{aligned} \tag{A.10}$$

$$\begin{aligned} \rho {V_p}^N(R) &= cR + \frac{\left(\beta {V_p}^{N\prime}\right)^2}{2\mu_p} \\ &\quad + {V_P}^{N\prime}\left(\frac{\alpha^2 {V_s}^{N\prime}}{\mu_s} + \frac{\gamma^2 {V_d}^{N\prime}}{\mu_d} - \varphi R\right), \end{aligned} \tag{A.11}$$

$$\begin{aligned} \rho {V_d}^N(R) &= \pi_d\left(a - \pi_d + \theta\pi_s - \theta\pi_d + \pi_d R\right) \\ &\quad + \frac{\left(\gamma {V_d}^{N\prime}\right)^2}{(2\mu_d)} - (\omega + c) R \\ &\quad + {V_d}^{N\prime}\left(\frac{\alpha^2 {V_s}^{N\prime}}{\mu_s} + \frac{\beta^2 {V_p}^{N\prime}}{\mu_p} - \varphi R\right). \end{aligned} \tag{A.12}$$

Following the literature [24, 25, 31], we obtain the following linear forms for the value functions:

$$V_s^N(R) = a_1^N R + b_1^N,$$
$$V_p^N(R) = a_2^N R + b_2^N, \quad (A.13)$$
$$V_d^N(R) = a_3^N R + b_3^N,$$

in which $a_1^N$, $a_2^N$, and $a_3^N$ and $b_1^N$, $b_2^N$, and $b_3^N$ are constants. From formula (A.13), we have

$$V_s^{N\prime} = a_1^N,$$
$$V_p^{N\prime} = a_2^N, \quad (A.14)$$
$$V_d^{N\prime} = a_3^N.$$

We substitute $V_s^N$, $V_p^N$, and $V_d^N$ from (A.13) and their derivatives from (A.14) into ((A.10)-(A.12)) and collect terms corresponding to $R$. By solving the algebraic equations, we have

$$a_1^N = \frac{(\pi_s\eta_s + \omega)}{(\rho+\varphi)},$$
$$a_2^N = \frac{c}{(\rho+\varphi)},$$
$$a_3^N = \frac{(\pi_d\eta_d - \omega - c)}{(\rho+\varphi)}$$
$$b_1^N = \frac{\pi_s(a - \pi_s + \theta\pi_d - \theta\pi_s)}{\rho} + \frac{\alpha^2}{2\mu_s\rho}\left(\frac{\pi_s\eta_s+\omega}{\rho+\varphi}\right)^2 + \frac{\pi_s\eta_s+\omega}{\rho(\rho+\varphi)^2}\left[\frac{\beta^2 c}{\mu_p} + \frac{\gamma^2(\pi_d\eta_d - \omega - c)}{\mu_d}\right]$$
$$b_2^N = \frac{\beta^2}{2\mu_p\rho}\left(\frac{c}{\rho+\varphi}\right)^2 + \frac{c}{\rho(\rho+\varphi)^2}\left[\frac{\alpha^2(\pi_s\eta_s+\omega)}{\mu_s} + \frac{\gamma^2(\pi_d\eta_d - \omega - c)}{\mu_d}\right]$$
$$b_3^N = \frac{\pi_d(a - \pi_d + \theta\pi_s - \theta\pi_d)}{\rho} + \frac{\gamma^2}{2\mu_d\rho}\left(\frac{\pi_d\eta_d - \omega - c}{\rho+\varphi}\right)^2 + \frac{\pi_d\eta_d - \omega - c}{\rho(\rho+\varphi)^2}\left[\frac{\alpha^2(\pi_s\eta_s+\omega)}{\mu_s} + \frac{\beta^2 c}{\mu_p}\right]. \quad (A.15)$$

Substituting (A.14) into (A.5), (A.7), and (A.9), the equilibrium matching efforts are given by

$$A_s^{N*} = \frac{\alpha(\pi_s\eta_s+\omega)}{\mu_s(\rho+\varphi)},$$
$$A_p^{N*} = \frac{\beta c}{\mu_p(\rho+\varphi)}, \quad (A.16)$$
$$A_d^{N*} = \frac{\gamma(\pi_d\eta_d - \omega - c)}{\mu_d(\rho+\varphi)}.$$

Next, substituting $A_s^{N*}$, $A_p^{N*}$, and $A_d^{N*}$ into differential equation (1) and using the initial conditions of (1), the general solution of the differential equation for the sharing level $R$ is

$$R^{N*} = K^N + \left(R_0 - K^N\right)e^{-\varphi t}, \quad (A.17)$$

where $K^N = \{\alpha^2\mu_p\mu_d(\pi_s\eta_s+\omega) + \beta^2\mu_s\mu_d c + \gamma^2\mu_s\mu_p(\pi_d\eta_d - \omega - c)\}/(\mu_s\mu_p\mu_d\varphi(\varphi+\rho))$.

This completes the proof. □

## B.

*Proof of Proposition 2.* We need to establish the existence of bounded and continuously differentiable value function $V_{sc}^I$ such that there exists a unique solution $R^I$ to differential equation (1) and the HJB equations. We first determine the players' necessary conditions from the HJBs. The optimisation problem of supplier is given as

$$J_{sc}^I = \max_{A_s,A_p,A_d} \int_0^\infty e^{-\rho t}\left\{\pi_s D_s + \pi_d D_d - C(A_s) - C\left(A_p\right) - C(A_d)\right\}dt$$
$$s.t. \quad R(t)' = \alpha A_s(t) + \beta A_p(t) + \gamma A_d(t) - \varphi R(t), \quad (B.1)$$
$$R(0) = R_0.$$

Let the value functions ${V_{sc}}^I = \max_{A_s,A_p,A_d} \int_t^\infty e^{-\rho(\tau-t)}\{\pi_s D_s + \pi_d D_d - C(A_s) - C(A_p) - C(A_d)\}d\tau$; the optimal profit function of supply chain then is given as

$$J_{sc}^I = e^{-\rho t}{V_{sc}}^I. \tag{B.2}$$

The supply chain's HJB is

$$\begin{aligned}\rho {V_{sc}}^I(R) = \max_{A_s,A_p,A_d} \Big\{&\pi_s D_s + \pi_d D_d - C(A_s) \\ &- C(A_p) - C(A_d) + {V_{sc}}^{I\prime} R'\Big\},\end{aligned} \tag{B.3}$$

and its maximisation provides the necessary condition for matching efforts:

$$\begin{aligned} A_s^I &= \frac{\alpha {V_{sc}}^{I\prime}}{\mu_s}, \\ A_p^I &= \frac{\beta {V_{sc}}^{I\prime}}{\mu_p}, \\ A_d^I &= \frac{\gamma {V_{sc}}^{I\prime}}{\mu_d}. \end{aligned} \tag{B.4}$$

By inserting (B.4) inside the HJB we obtain the following algebraic equations:

$$\begin{aligned}\rho {V_{sc}}^I(R) = \max_{A_s,A_p,A_d} \Bigg\{&\pi_s(a - \pi_s + \theta\pi_d - \theta\pi_s + \eta_s R) \\ &+ \pi_d(a - \pi_d + \theta\pi_s - \theta\pi_d + \eta_d R) + \frac{\left(\alpha {V_{sc}}^{I\prime}\right)^2}{(2\mu_s)} \\ &+ \frac{\left(\beta {V_{sc}}^{I\prime}\right)^2}{(2\mu_p)} + \frac{\left(\gamma {V_{sc}}^{I\prime}\right)^2}{(2\mu_d)} + {V_{sc}}^{I\prime}\varphi R\Bigg\}.\end{aligned} \tag{B.5}$$

Thus, we obtain the following linear forms for the value functions:

$${V_{sc}}^I(R) = a^I R + b^I, \tag{B.6}$$

in which $a^I$ and $b^I$ are constants. From formula (B.6), we have

$${V_{sc}}^{I\prime} = a^I. \tag{B.7}$$

We substitute ${V_{sc}}^I$ from (B.6), as well as its derivative from (B.7), into (B.5), and collect terms corresponding to $R$. By solving the algebraic equations, we have

$$\begin{aligned} a^I &= \frac{(\pi_s\eta_s + \pi_d\eta_d)}{(\rho+\varphi)} \\ b^I &= \frac{\pi_s(a - \pi_s + \theta\pi_d - \theta\pi_s)}{\rho} \\ &\quad + \frac{\pi_d(a - \pi_d + \theta\pi_s - \theta\pi_d)}{\rho} \\ &\quad + \frac{1}{2\rho}\left(\frac{\pi_s\eta_s + \pi_d\eta_d}{\rho+\varphi}\right)^2\left(\frac{\alpha^2}{\mu_s} + \frac{\beta^2}{\mu_p} + \frac{\gamma^2}{\mu_d}\right) \end{aligned} \tag{B.8}$$

Substituting ((B.7)-(B.8)) into (B.4) the equilibrium matching efforts are given by

$$\begin{aligned} {A_s}^{I*} &= \frac{\alpha(\pi_s\eta_s + \pi_d\eta_d)}{\mu_s(\rho+\varphi)}, \\ {A_p}^{I*} &= \frac{\beta(\pi_s\eta_s + \pi_d\eta_d)}{\mu_p(\rho+\varphi)}, \\ {A_d}^{I*} &= \frac{\gamma(\pi_s\eta_s + \pi_d\eta_d)}{\mu_d(\rho+\varphi)}. \end{aligned} \tag{B.9}$$

Next, substituting ${A_s}^{I*}$, ${A_p}^{I*}$, ${A_d}^{I*}$ into differential equation (1) and using the initial conditions of (1), the general solution of the differential equation for the sharing level $R$ is

$$R^{I*} = K^I + \left(R_0 - K^I\right)e^{-\varphi t}, \tag{B.10}$$

where $K^I = (\alpha^2\mu_p\mu_d + \beta^2\mu_s\mu_d + \gamma^2\mu_s\mu_p)(\pi_s\eta_s + \pi_d\eta_d)/(\mu_s\mu_p\mu_d\varphi(\varphi+\rho))$. This completes the proof. □

## C.

*Proof of Proposition 3.* To prove the first item, we need to establish that all decision variables of Proposition 1 are positive, i.e., ${A_d}^{N*} > 0$. This implies, in turn, $\pi_d\eta_d - \omega - c > 0$ (from (A.16)). Straightforward comparisons, using the values in (A.16) and (B.9), lead to the results:

$${A_s}^{I*} - {A_s}^{N*} = \frac{\alpha(\pi_d\eta_d - \omega)}{(\mu_s(\rho+\varphi))} > 0, \tag{C.1}$$

$${A_p}^{I*} - {A_p}^{N*} = \frac{\beta(\pi_s\eta_s + \pi_d\eta_d - c)}{\left(\mu_p(\rho+\varphi)\right)} > 0, \tag{C.2}$$

$${A_d}^{I*} - {A_d}^{N*} = \frac{\gamma(\pi_s\eta_s + \omega + c)}{(\mu_d(\rho+\varphi))} > 0, \tag{C.3}$$

$$\begin{aligned} {J_{sc}}^{I*} - {J_{sc}}^{N*} &= e^{-\rho t}\left({V_{sc}}^{I*} - {V_{sc}}^{N*}\right) \\ &= \frac{3\gamma^2(\pi_s\eta_s + \pi_d\eta_d)(\pi_s\eta_s + \omega - c) + \gamma^2(\pi_d\eta_d - \omega + c)^2}{8\mu_d\rho(\rho+\varphi)^2} \\ &> 0. \end{aligned} \tag{C.4}$$

This completes the proof. □

## D.

*Proof of Proposition 4.* We need to establish the existence of bounded and continuously differentiable value functions ${V_s}^Y$, ${V_p}^Y$, and ${V_d}^Y$ such that it is a unique solution $R^Y$ to differential equation (1) and the HJB equations. We first determine the players' necessary conditions from the HJBs. Because the game is played à la Stackelberg and the platform is the leader, we first derive the decision variables for the supplier and the demander for the second game stage. The optimisation problem of supplier is given as

$$\begin{aligned} J_s^Y = \max_{A_s} \quad & \int_0^\infty e^{-\rho t}\left\{\pi_s D_s(t) - C\left(A_s(t)\right) + \omega R(t)\right\} dt \\ s.t. \quad & R(t)' = \alpha A_s(t) + \beta A_p(t) + \gamma A_d(t) - \varphi R(t), \\ & R(0) = R_0 . \end{aligned} \tag{D.1}$$

Let the value functions ${V_s}^Y = \max_{A_s} \int_t^\infty e^{-\rho(\tau-t)}\{\pi_s D_s(t) - C(A_s(t)) + \omega R(t)\}d\tau$; the optimal profit function of manufacturing resources supplier then is given as

$$J_s^Y = e^{-\rho t}{V_s}^Y. \tag{D.2}$$

Similarly, the optimal profit function of the platform and demander are given by

$$\begin{aligned} J_p^Y &= e^{-\rho t}{V_p}^Y, \\ {J_d}^Y &= e^{-\rho t}{V_d}^Y. \end{aligned} \tag{D.3}$$

The supplier's HJB is

$$\rho {V_s}^Y(R) = \max_{A_s \geq 0}\left\{\pi_s D_s - \frac{\mu_s {A_s}^2}{2} + \omega R + {V_s}^{N\prime} R'\right\}, \tag{D.4}$$

and its maximisation provides the necessary condition for matching efforts:

$$A_s^Y = \frac{\alpha {V_s}^{Y\prime}}{\mu_s}. \tag{D.5}$$

Similarly, the demander's HJB is

$$\begin{aligned} \rho {V_d}^Y(R) = \max_{A_d \geq 0}\Bigg\{ & \pi_d D_d - \frac{(1-\varepsilon)\mu_d {A_d}^2}{2} \\ & - (\omega + c) R + {V_d}^{N\prime} R'\Bigg\}, \end{aligned} \tag{D.6}$$

and its maximisation provides the necessary condition for matching efforts:

$$A_d^Y = \frac{\gamma {V_d}^{Y\prime}}{\left(\mu_d(1-\varepsilon)\right)}. \tag{D.7}$$

Substituting (D.5); (D.7) into the platform's HJB gives

$$\begin{aligned} \rho {V_p}^Y(R) = \max_{A_p \geq 0, \varepsilon \geq 0}\Bigg\{ & cR - \frac{\mu_p {A_p}^2}{2} - \frac{\varepsilon\left(\gamma {V_d}^{Y\prime}\right)^2}{2\mu_d(1-\varepsilon)^2} \\ & + {V_p}^{Y\prime}\left(\frac{\alpha^2 {V_s}^{Y\prime}}{\mu_s} + \beta A_p + \frac{\gamma^2 {V_d}^{Y\prime}}{\mu_d(1-\varepsilon)} - \varphi R\right)\Bigg\}, \end{aligned} \tag{D.8}$$

while performing the maximisation of the right-hand side we obtain

$$\begin{aligned} A_p^Y &= \frac{\beta {V_p}^{Y\prime}}{\mu_p}, \\ \varepsilon &= \frac{\left(2{V_p}^{Y\prime} - {V_d}^{Y\prime}\right)}{\left(2{V_p}^{Y\prime} + {V_d}^{Y\prime}\right)}. \end{aligned} \tag{D.9}$$

By inserting (D.5), (D.7), and (D.9) inside the HJBs, we obtain the following three algebraic equations:

$$\begin{aligned} & \rho {V_s}^Y(R) \\ & = \pi_s\left(a - \pi_s + \theta\pi_d - \theta\pi_s + \eta_s R\right) + \frac{\left(\alpha {V_s}^{Y\prime}\right)^2}{(2\mu_s)} \\ & \quad + \omega R \\ & \quad + {V_s}^{Y\prime}\left(\frac{\beta^2 {V_p}^{Y\prime}}{\mu_p} + \frac{\gamma^2\left(2{V_p}^{Y\prime} + {V_d}^{Y\prime}\right)}{(2\mu_d)} - \varphi R\right), \end{aligned} \tag{D.10}$$

$$\begin{aligned} & \rho {V_p}^Y(R) \\ & = cR + \frac{\left(\beta {V_p}^{Y\prime}\right)^2}{\left(2\mu_p\right)} + \frac{\gamma^2\left(2{V_p}^{Y\prime} + {V_d}^{Y\prime}\right)^2}{(8\mu_d)} \\ & \quad + {V_p}^{Y\prime}\left(\frac{\alpha^2 {V_s}^{Y\prime}}{\mu_s} - \varphi R\right), \end{aligned} \tag{D.11}$$

$$\begin{aligned} & \rho {V_d}^N(R) \\ & = p_d\left(a - p_d + \theta p_s - \theta p_d + \eta_d R\right) \\ & \quad + \frac{\gamma^2 {V_d}^{Y\prime}\left(2{V_p}^{Y\prime} + {V_d}^{Y\prime}\right)}{(4\mu_d)} - (\omega + c) R \\ & \quad + {V_d}^{Y\prime}\left(\frac{\alpha^2 {V_s}^{Y\prime}}{\mu_s} + \frac{\beta^2 {V_p}^{Y\prime}}{\mu_p} - \varphi R\right), \end{aligned} \tag{D.12}$$

Thus, we obtain the following linear forms for the value functions:

$$\begin{aligned} V_s^Y(R) &= a_1^Y R + b_1^Y, \\ V_p^Y(R) &= a_2^Y R + b_2^Y, \\ V_d^Y(R) &= a_3^Y R + b_3^Y, \end{aligned} \tag{D.13}$$

in which $a_1^Y$,$a_2^Y$, and $a_3^Y$and $b_1^Y$, $b_2^Y$, and $b_3^Y$are constants. From formula (D.13), we have

$$\begin{aligned} V_s^{Y\prime} &= a_1^Y, \\ V_p^{Y\prime} &= a_2^Y, \\ V_d^{Y\prime} &= a_3^Y. \end{aligned} \tag{D.14}$$

We substitute $V_s^Y$, $V_p^Y$, $V_d^Y$ from (D.13), as well as their derivatives from (D.14), into ((D.10)-(D.12)), and collect terms corresponding to $R$. By solving the algebraic equations, we have

$$\begin{aligned} a_1^Y &= \frac{(\pi_s\eta_s+\omega)}{(\rho+\varphi)}, \\ a_2^Y &= \frac{c}{(\rho+\varphi)}, \\ a_3^Y &= \frac{(\pi_d\eta_d-\omega-c)}{\rho}+\varphi \\ b_1^Y &= \frac{\pi_s(a-\pi_s+\theta\pi_d-\theta\pi_s)}{\rho}+\frac{\alpha^2(\pi_s\eta_s+\omega)^2}{2\mu_s\rho(\rho+\varphi)^2} \\ &\quad+\frac{\pi_s\eta_s+\omega}{\rho(\rho+\varphi)^2}\left[\frac{\beta^2 c}{\mu_p}+\frac{\gamma^2(\pi_d\eta_d-\omega+c)}{2\mu_d}\right] \\ b_2^Y &= \frac{\beta^2}{2\mu_p\rho}\left(\frac{c}{\rho+\varphi}\right)^2+\frac{\gamma^2(\pi_d\eta_d-\omega+c)^2}{8\mu_d\rho(\rho+\varphi)^2} \\ &\quad+\frac{c\alpha^2(\pi_s\eta_s+\omega)}{\mu_s\rho(\rho+\varphi)^2} \\ b_3^Y &= \frac{\pi_d(a-\pi_d+\theta\pi_s-\theta\pi_d)}{\rho} \\ &\quad+\frac{\gamma^2(\pi_d\eta_d-\omega-c)(\pi_d\eta_d-\omega+c)}{4\mu_d\rho(\rho+\varphi)^2} \\ &\quad+\frac{\pi_d\eta_d-\omega-c}{\rho(\rho+\varphi)^2}\left[\frac{\alpha^2(\pi_s\eta_s+\omega)}{\mu_s}+\frac{\beta^2 c}{\mu_p}\right] \end{aligned} \tag{D.15}$$

Substituting (D.14) into (D.5), (D.7), and (D.9), the equilibrium matching efforts are given by

$$\begin{aligned} A_s^{Y*} &= \frac{\alpha(\pi_s\eta_s+\omega)}{\mu_s(\rho+\varphi)}, \\ A_p^{Y*} &= \frac{\beta c}{\mu_p(\rho+\varphi)}, \\ \varepsilon &= \frac{3c-\pi_d\eta_d+\omega}{\pi_d\eta_d-\omega+c}, \\ A_d^{Y*} &= \frac{\gamma(\pi_d\eta_d-\omega+c)}{2\mu_d(\rho+\varphi)}. \end{aligned} \tag{D.16}$$

Next, substituting $A_s^{Y*}$, $A_p^{Y*}$, and $A_d^{Y*}$ into differential equation (1) and using the initial conditions of (1), the general solution of the differential equation for the sharing level $R$ is

$$R^{Y*} = K^Y + \left(R_0 - K^Y\right)e^{-\varphi t}, \tag{D.17}$$

where $K^Y = \{2\alpha^2\mu_p\mu_d(\pi_s\eta_s+\omega)+2\beta^2 c\mu_s\mu_d+\gamma^2\mu_s\mu_p(\pi_d\eta_d-\omega+c)\}/(2\mu_s\mu_p\mu_d\varphi(\varphi+\rho))$.

This completes the proof. □

## E.

*Proof of Proposition 5.* To prove the first item, we need to establish that all decision variables are positive, i.e., $A_d^{N*} > 0$, $\varepsilon > 0$, and $A_d^{Y*}$. This implies, in turn, $\pi_d\eta_d - \omega - c > 0$ (from (A.16)), $\pi_d\eta_d - \omega + c > 0$ (from (D.16)), and $3c - \pi_d\eta_d + \omega > 0$ (from (D.16)). It can be easily shown that

$$\max\left\{0, \frac{(\pi_d\eta_d-\omega)}{3}\right\} < c < \pi_d\eta_d - \omega. \tag{E.1}$$

Straightforward comparisons, using the values in (A.16), (B.9), and (D.16), lead to the results:

$$A_s^{N*} = A_s^{Y*} < A_s^{I*}, \tag{E.2}$$

$$A_p^{N*} = A_p^{Y*} < A_p^{I*}, \tag{E.3}$$

$$A_d^{N*} < A_d^{Y*} < A_d^{I*}, \tag{E.4}$$

$$\begin{aligned} J_s^{Y*} - J_s^{N*} &= e^{-\rho t}\left(V_s^{Y*} - V_s^{N*}\right) \\ &= \frac{e^{-\rho t}\gamma^2(\pi_s\eta_s+\omega)(3c-\pi_d\eta_d+\omega)}{2\mu_d\rho(\rho+\varphi)^2} > 0, \end{aligned} \tag{E.5}$$

$$\begin{aligned} J_p^{Y*} - J_p^{N*} &= e^{-\rho t}\left(V_p^{Y*} - V_p^{N*}\right) \\ &= \frac{e^{-\rho t}\gamma^2(3c-\pi_d\eta_d+\omega)^2}{8\mu_d\rho(\rho+\varphi)^2} > 0, \end{aligned} \tag{E.6}$$

$$\begin{aligned} J_d^{Y*} - J_d^{N*} &= e^{-\rho t}\left(V_d^{Y*} - V_d^{N*}\right) \\ &= \frac{e^{-\rho t}\gamma^2(\pi_d\eta_d-\omega-c)(3c-\pi_d\eta_d+\omega)}{4\mu_d\rho(\rho+\varphi)^2} > 0, \end{aligned} \tag{E.7}$$

From the above inequalities ((E.5)-(E.7)), we get

$$J_{sc}^{Y*} - J_{sc}^{N*} = \left(J_s^{Y*} + J_p^{Y*} + J_d^{Y*}\right) - \left(J_s^{N*} + J_p^{N*} + J_d^{N*}\right) > 0. \tag{E.8}$$

This proves Proposition 5 and completes the proof. □

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

## Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (71572033; 71832001).

## References

[1] X. Xu, "From cloud computing to cloud manufacturing," *Robotics and Computer-Integrated Manufacturing*, vol. 28, no. 1, pp. 75–86, 2012.

[2] D. O'Rourke, "The science of sustainable supply chains," *Science*, vol. 344, no. 6188, pp. 1124–1127, 2014.

[3] W. B. Bonvillian, "Advanced manufacturing policies and paradigms for innovation," *Science*, vol. 342, no. 6163, pp. 1173–1175, 2013.

[4] K. Li, W. Xiao, and S. Yang, "Scheduling uniform manufacturing resources via the Internet: A review," *Journal of Manufacturing Systems*, vol. 50, pp. 247–262, 2019.

[5] F. Tao, L. Zhang, V. C. Venkatesh, Y. Luo, and Y. Cheng, "Cloud manufacturing: a computing and service-oriented manufacturing model," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, vol. 225, no. 10, pp. 1969–1976, 2011.

[6] D. Schaefer, J. L. Thames, R. D. Wellman Jr., D. Wu, and D. W. Rosen, "Distributed collaborative design and manufacture in the cloudmotivation, infrastructure, and education," *The ASEE Computers in Education (CoED) Journal*, vol. 3, no. 4, p. 1, 2012.

[7] L. Zhang, Y. Luo, F. Tao et al., "Cloud manufacturing: a new manufacturing paradigm," *Enterprise Information Systems*, vol. 8, no. 2, pp. 167–187, 2014.

[8] D. Wu, J. L. Thames, D. W. Rosen, and D. Schaefer, "Towards a cloud-based design and manufacturing paradigm: looking backward, looking forward," *Innovation*, vol. 17, no. 18, 2012.

[9] O. Fisher, N. Watson, L. Porcu, D. Bacon, M. Rigley, and R. L. Gomes, "Cloud manufacturing as a sustainable process manufacturing route," *Journal of Manufacturing Systems*, vol. 47, pp. 53–68, 2018.

[10] Y. Cheng, D. Zhao, A. R. Hu, Y. L. Luo, F. Tao, and L. Zhang, "Multi-view Models for Cost Constitution of Cloud Service in Cloud Manufacturing System," in *Advances in Computer Science and Education Applications*, vol. 202, pp. 225–233, Springer, Berlin, Heidelberg, 2011.

[11] F. Tao, Y. Hu, D. Zhao, and Z. Zhou, "Study on resource service match and search in manufacturing grid system," *The International Journal of Advanced Manufacturing Technology*, vol. 43, no. 3-4, pp. 379–399, 2008.

[12] M. Zhang, C. Li, Y. Shang, and C. Li, "Research on resource service matching in cloud manufacturing," *Manufacturing Letters*, vol. 15, Article ID S2213846318300154, pp. 50–54, 2018.

[13] X. Gong, C. Yin, and X. Li, "A grey correlation based supply–demand matching of machine tools with multiple quality factors in cloud manufacturing environment," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 3, pp. 1025–1038, 2019.

[14] W. Wang and L. Fei, "The research of cloud manufacturing resource discovery mechanism," in *Proceedings of the in International Conference on Computer Science Education*, 2012.

[15] T. Wang, S. Guo, and C.-G. Lee, "Manufacturing task semantic modeling and description in cloud manufacturing system," *The International Journal of Advanced Manufacturing Technology*, vol. 71, no. 9-12, pp. 2017–2031, 2014.

[16] H. Li, L. Zhang, and R. Jiang, "Study of manufacturing cloud service matching algorithm based on OWL-S," in *Proceedings of the 26th Chinese Control and Decision Conference, CCDC 2014*, pp. 4155–4160, Yinchuan, China, June 2014.

[17] C. Yin, Q. Xia, and Z.-W. Li, "Semantic matching technique of cloud manufacturing service based on OWL-S," *Computer Integrated Manufacturing Systems*, vol. 18, no. 7, pp. 1494–1502, 2012.

[18] H.-f. Li, X. Dong, and C.-g. Song, "Intelligent searching and matching approach for cloud manufacturing service," *Computer Integrated Manufacturing Systems*, vol. 18, no. 7, pp. 1485–1493, July 2012.

[19] Y. Cheng, F. Tao, L. Zhang, and D. Zhao, "Dynamic supply-demand matching for manufacturing resource services in service-oriented manufacturing systems: a hypernetwork-based solution framework," in *Proceedings of the ASME 2015 International Manufacturing Science and Engineering Conference*, American Society of Mechanical Engineers, Charlotte, NC, USA, 2015.

[20] Y. Cheng, F. Tao, D. Zhao, and L. Zhang, "Modeling of manufacturing service supply–demand matching hypernetwork in service-oriented manufacturing systems," *Robotics and Computer-Integrated Manufacturing*, vol. 45, pp. 59–72, 2017.

[21] F. Tao, L. Zhang, Y. Liu, Y. Cheng, L. Wang, and X. Xu, "Manufacturing service management in cloud manufacturing: overview and future research directions," *Journal of Manufacturing Science and Engineering*, vol. 137, no. 4, Article ID 040912, pp. 1–11, 2015.

[22] G. P. Cachon and S. Netessine, "Game theory in supply chain analysis," in *Handbook of Quantitative Supply Chain Analysis: Modeling in the eBusiness Era*, D. Simchi-Levi, S. Wu, and Z. Shen, Eds., vol. 74, Kluwer, 2004.

[23] P. De Giovanni, "Quality improvement vs. advertising support: which strategy works better for a manufacturer?" *European Journal of Operational Research*, vol. 208, no. 2, pp. 119–130, 2011.

[24] N. Amrouche, G. Martín-Herrán, and G. Zaccour, "Feedback Stackelberg equilibrium strategies when the private label competes with the national brand," *Annals of Operations Research*, vol. 164, no. 1, pp. 79–95, 2008.

[25] S. Jørgensen and G. Zaccour, "A survey of game-theoretic models of cooperative advertising," *European Journal of Operational Research*, vol. 237, no. 1, pp. 1–14, 2014.

[26] S. Karray, "Cooperative promotions in the distribution channel," *Omega*, vol. 51, pp. 49–58, 2015.

[27] J.-C. Lu, Y.-C. Tsao, and C. Charoensiriwath, "Competition under manufacturer service and retail price," *Economic Modelling*, vol. 28, no. 3, pp. 1256–1264, 2011.

[28] Y. Tian, J. Ma, and W. Lou, "Research on supply chain stability driven by consumer's channel preference based on complexity theory," *Complexity*, vol. 2018, Article ID 7812784, 13 pages, 2019.

[29] L. Qiu-xiang, Z. Yu-hao, and H. Yi-min, "The complexity analysis in dual-channel supply chain based on fairness concern and different business objectives," *Complexity*, vol. 2018, Article ID 4752765, 13 pages, 2018.

[30] J. Ma, H. Ren, M. Yu, and M. Zhu, "Research on the complexity and chaos control about a closed-loop supply chain with dual-channel recycling and uncertain consumer perception," *Complexity*, vol. 2018, Article ID 9853635, 13 pages, 2018.

[31] A. Chutani and S. P. Sethi, "Optimal advertising and pricing in a dynamic durable goods supply chain," *Journal of Optimization Theory and Applications*, vol. 154, no. 2, pp. 615–643, 2012.

Hindawi
Complexity
Volume 2019, Article ID 1536716, 15 pages
https://doi.org/10.1155/2019/1536716

*Research Article*

# Gradient Boosted Trees Predictive Models for Surface Roughness in High-Speed Milling in the Steel and Aluminum Metalworking Industry

**Victor Flores and Brian Keith**

*Department of Computing & Systems Engineering, Universidad Católica del Norte, Angamos, Av. 0610, Antofagasta, Chile*

Correspondence should be addressed to Victor Flores; vflores@ucn.cl

Received 1 April 2019; Accepted 4 June 2019; Published 1 July 2019

Guest Editor: Rosario Domingo



High-speed machining is a technique that maintains a high interest in the manufacture of metal parts for the excellent results it provides, both in surface finish and in economic benefits. In the industry, the tendency is to incorporate data management and analysis techniques to generate information that helps improve the surface roughness results in machining. A good alternative to improve the surface quality results in the manufacture of metal parts is using predictive models of the surface roughness. In this document, we present work done with experimental data obtained from two high-speed machining (HSM) machines with different types of tools and cutting conditions, conducted under an experimental design with interest in three of factors commonly studied to generate surface roughness models: tool characteristics, cutting conditions, and characteristics of the machined material. Steel and aluminum alloys were used in the experimentation. The results are contrasted with prior experiences that use the same experimental design but with different soft computing techniques and they are also contrasted with the results of similar previous works. Our results show accuracies ranging from 61.54% to 88.51% on the datasets, which are competitive results when compared with the other approaches. We also find the axial cut-depth is the most influential feature for the slots datasets and the hardness and diameter of the cutting tool are the most influential features for the geometries datasets.

## 1. Introduction

High-speed machining is a technique that maintains popularity in the manufacture of metal parts, with high levels of usability in the metalworking industry to manufacture metal parts with high levels of quality in surface finish [1–3]. The surface quality achieved with material removal techniques is very important in the manufacture of pieces from alloys of metals and plastic and in general materials that can be subjected to roughing [4]. The surface quality in a machined part depends to a large extent on the combination of factors such as the properties of the machined material, characteristics of the machining center, and the tool used [5, 6].

In the industrial field of mechanical cutting, the tendency is to incorporate data managing and analysis techniques to generate information that helps improve machining surface quality [7, 8]. Soft computing techniques help identify factors that affect the machining process and their most convenient values to obtain the best surface quality [6], while minimizing the associated costs, such as calibration costs, experimentation, and qualities measurement, among others. Surface quality is often related to surface roughness, although it can be calculated from several parameters. In practice, the surface roughness can be evaluated using the parameter roughness average (Ra) [2, 4, 9, 10]; this is the most common industrial parameter for this task according to [11] and previous works such as [10]. The Ra parameter can be measured relatively simple using profilometers [4, 9]. The surface roughness has a great influence on other factors of interest in manufacturing such as friction, electrical and thermal resistance, and the appearance of the machined part, among other factors that can affect its functionality [12]. Also, surface roughness can help to establish the relationship between the lubrication and other elements such as friction or wear between parts [13]. Friction causes wear between parts (particularly metal alloys), while ineffective lubrication increases friction; both of these affect surface roughness [14]. Works such as

[15–17] present predictive models of surface roughness with particular conditions of lubricant use.

Surface roughness average is the most used parameter to estimate the surface quality according to [18]; it is also important because it provides ideas on surface finish [19, 20] and also provides information on the behavior of a surface in contact with other surfaces [9, 21]. In the metal fabrication industry, there are research works that evaluate the machinability of steel pieces according to parameters that influence the process, such as the work described in [22]. There are also several works that present surface roughness predictive models based on soft computing techniques, using parameters related to the cutting process as predictors, such as the characteristics of the machine or the characteristics of the machined material, examples of this type of work in [4, 6, 9, 16, 23]. A brief description of these works is given in the following paragraph.

An extended research has been conducted into the applicability of artificial intelligence and soft computing techniques for surface roughness prediction in mechanical cutting over the last years, follow previous works as [15]; the most common configuration (until 2011) was the multilayer perceptron (MLP) with a single hidden layer; though Bayesian networks [4, 9, 21, 23], genetic algorithms [19, 24, 25], and support vector machines [10, 15, 21] have been widely used for surface roughness prediction.

For example, in [10] the decision trees, Naive Bayes, Nearest Neighbors Classifiers, Multilayer Perceptrons, and Logistic Regression were used to generate methods for the early detection of multitooth tool breakages in the milling process. Decision trees have been used in conjunction with sound signal analysis in [26] in order to generate a predictive model of surface roughness. Also, fuzzy logic is used to generate predictive models of surface roughness in [27, 28]. Another example is the work described in [16], where a predictive model of surface roughness in deep drilling operations under high-speed conditions of steel molds was created using Bayesian networks. Furthermore, Bayesian networks were used in [4, 9] and a combination of Bayesian networks and Tree-augmented Network algorithms was used in [23] to generate a preprocessing model of surface roughness on high-speed machining (HSM) over metal probes.

The profitability of metal cutting operations depends to a great extent on factors such as precision in mechanical cutting, excellent surface finish, and minimum wear of the tool [14, 29]. All the relationships between surface roughness and factors such as lubrication, friction, and the mechanical force applied to the cutting process are closely linked and this is currently a point of great interest in the companies and the profitability of mechanical cutting [30, 31].

There are multiple works on the literature that use soft computing to estimate surface roughness or to study the factors that can affect surface roughness. However, in our review of the literature for this work, there were few works that used decision trees and to the best of our knowledge, no previous works with the specific technique of Gradient Boosted Trees to generate a predictive model of surface roughness.

This document presents a surface roughness prediction model that considers a subset of elements involved in the milling process that is related to the machined piece, the tool, and characteristics of the machine tool. To generate the predictive model of surface roughness, metal alloy pieces commonly used in the industry have been employed. The data used to generate the model are the result of experiments on two different machines and, in each one, various combinations of variables that typically influence the surface quality results in the milling process have been used.

The rest of the document is structured as follows. Section 2 details various concepts and related works in the field of predicting surface quality and machining, it also details predictive models, in particular, and it focuses on Gradient Boosted Trees. Section 3 details the materials and methods used in this work, presenting the experimental description, the implementation, and evaluation of the models, as well as their parametrization. Section 4 presents the results obtained with the implemented models alongside a comparison with other methods. Section 5 contains a discussion of the obtained results and the comparison with other state-of-the-art methods. Section 6 closes with the conclusions of this paper and potential future lines of works.

## 2. Concepts and Related Works

*2.1. Surface Quality and Machining.* The surface quality or surface roughness is intimately connected with the appearance of the machined or manufactured surface, which is normally expressed with a Ra value [3, 11]. In many cases, surface quality must comply with established standard values to be functional in certain industries (such as in the molds used to manufacture parts with plastic injection [16]). There are several parameters to establish the surface roughness value. The surface roughness average is the most widely used in the industry thanks to the ease with which it can be assessed (generally after processing) and the closeness with which it represents the surface texture of the mechanized part [32, 33].

The Ra value is usually calculated by integrating the arithmetic mean of the absolute values of ordinates $f(x)$ within a sampling length (L). Each partial value of surface roughness can be measured using profilometers along the length of sampling L. The standard 4288 (1996) is the internationally used way to measure surface roughness in machining processes and this standard is further complemented with the standard 1302 (1992) which establishes 12 levels of surface roughness; these levels range from 0.006 to 50 nanometers (nm) [34].

Although low-speed machining can provide better surface roughness, it reduces the efficiency of the industry, implies more machining time, and consequently increases production costs [5]. High-speed machining is one of the processes with the greatest economic impact in the metal fabrication industry [2], thanks to the high level of surface finish that can be obtained with it [4], which influences the functional behavior of the resulting piece when subjected to friction, abrupt temperature changes, and other factors that may affect its functioning [6].

Currently, there is plenty of face milling research aimed at predicting surface roughness, a parameter that will decrease with respect to changes in other parameters like increased tool wear or flank wear, cutting force, depth-cut, or feed per tooth. There are several works that have been presented in the last 5 years in the topic of predicting surface roughness based on artificial intelligence techniques. For example, on [22], probes of AISI 4340 steel were used and cutting speed; feed and cut-depth were considered as the governing parameters for surface roughness prediction, while workpiece surface temperature, machining forces, and tool flank wear were taken as measures to check the performance of the estimation. In [30] 100CrMoV5 steel molds were machined with minimum quantity lubrication (MQL) and tool lifetime, flank face, and cutting speed were used as predictive variables.

Fuzzy logic and regression analysis were used in [28] for an empirical model for surface roughness and this model was used to calculate the influence of surface roughness predicted over the product profile of surface finish in a face milling process. Random forests, regression trees, and radial-basis functions were used in [35] to estimate the adequate thresholds values required to avoid rapid tool wear and predict the finish quality on a flat surface using face milling. In the same way, the relation between face milling (over steel-45 workpieces) and parameters like tooth flank, cutting speed, and feed was studied in [36]. Genetic algorithms and the Grey Wolf Optimizer algorithm were used in [24] to generate a prediction of surface roughness in ball-end milling over X210CR12 steel. Some of the works described above use postprocessing techniques to estimate the surface roughness (i.e., [22, 30, 37]), while others, such as [24, 28, 35], apply the techniques to estimate surface roughness in-process.

Furthermore, in [38] the authors present a comparative work of milling models on aluminum alloys A7075 with a diamond-like coating (DLC) tool and tool without DLC; the models were made with the aim of predicting the least wear of the tool. The results were experimentally better when using the tool model with DLC. Also, in [37] a model is presented that relates to the influence of factors such as tool wear and flank wear with the surface quality.

In spite of the high number of works, such as those described above, the in-process measurement of surface roughness is difficult and often unfeasible. Therefore, as it has been said in the introduction, having techniques to predict surface quality using postprocess data is a way of working that is gaining interest in the parts manufacturing industry. In this sense, predictive models have much to contribute.

*2.2. Predictive Models.* In artificial intelligence, predictive tasks are one of the central topics of machine learning that involves inducing a model from training data (known as training instances), then this model can be applied to future instances to predict a target variable of interest [39]. There are several prediction algorithms, such as Logistic Regression, neural networks, decision trees, Bayesian networks, among others. These algorithms typically induce a model to learn to predict the best value of a target variable from training instances of a domain, with the aim of finding an optimal value of the target variable in all future instances of the said domain [40].

Many scientific articles in the literature work with predictive algorithms and particular training instances in a domain selected according to research interests. One advantage of working with training instances in a domain is that the predictive algorithm will find a more precise model that can generate good values of the target variable in the presence of new data.

There are several recent works in which artificial intelligence techniques are used to estimate the surface roughness in machining. In [26] a proposal for a semisupervised approach to the development of roughness prediction models, based on machine learning, is described; also, in [31] a modular road roughness classification system operates with the vehicle's transfer functions (according to ISO 8608) is described; for example [16] applies Bayesian networks in the context of turn-milling using steel to generate a prediction of surface roughness. Also, in [4] neural networks are used for the same purpose. In [6] genetic algorithms were used for the prediction of surface roughness in micromachining of Copper C360. These artificial intelligence techniques have been shown to be able to generate predictive models with excellent accuracy even in with a lack of available data in a domain through the identification of patterns in the data [18].

*2.3. Decision Trees.* Learning based on decision trees is a type of predictive model that uses a decision tree to go from observations of an object (represented as the branches of a tree) to a certain conclusion about a target value of the object (represented by the tree leaves). It is used in statistics, data mining, and machine learning and has had several applications, both at the academic level and in the industry [41].

This classifier is one of the easiest modeling techniques to interpret thanks to its graphic representation; they are didactic and easy to understand. They base their predictions on inductive learning; that is, they consider the values that the different attributes or variables take, creating in this way, a series of rules to be able to determine what value the dependent variable will take based on certain situations. It should be noted that the results delivered by the decision tree depend to a large extent on the volume of data contained in each category. The accuracy of the model with respect to reality will be better the greater the amount of data available of that combination of features.

Finally, in the industry of crafting pieces from machine-cutting, it is highly important to also obtain information about the factors that affect surface roughness and to also have the ability to influence such factors. In particular, decision trees are a useful technique to explain the aforementioned information, according to the works of [21, 42]. In this work, we have selected a boosting model based on decision trees, explained in the next section, to develop our models. This approach is appropriate since we are interested in both predicting surface roughness and explaining the relationships and influences among the different factors affecting the cutting process.

Table 1: Variables used in the study including symbols, units, and values/ranges.

| Variable | Symbol | Unit | Values/Ranges |
|---|---|---|---|
| Axial Depth of Cut | ap | mm | Exp-1 = {0.2, 0.4, 0.6, 1}<br>Exp-2 = {5, 10} |
| Advance Speed-Feed Rate | F | mm/min | Exp-1 = {1500, 1875, 2250, 2625}<br>Exp-2 = {500, 675, 750} |
| Diameter of the Tool | diam | mm | Exp-1 = {6, 8, 10, 12}<br>Exp-2= {10, 12, 16, 20} |
| Hardness of the Material | HB | | 85-150 |
| Rotating Speed | n | rpm | Exp-1 = {5000, 6250, 7500, 8750}<br>Exp-2 = {8000, 9600, 712000, 15000} |
| Radial Depth of Cut | ae | mm | Exp-2 = {0.5, 1} |
| Number of Teeth | flutes | | Exp-1 = {2, 6}<br>Exp-2 = {4} |

*2.4. Gradient Boosted Trees.* The term boosting refers to a family of algorithms that convert weak learners into strong learners, understanding that a weak learner is only slightly better than a random choice, while a strong learner has an almost perfect performance [5]. The Gradient Boosting model is a machine learning technique that can be used in both regression and classification problems. This approach produces a predictive model in the form of an ensemble from weak predictive models, which normally correspond to decision trees as in the particular case used in this work (Gradient Boosted Trees (GBT)). This method builds the model in stages like other boosting methods but also generalizes these by optimizing an arbitrary differentiable loss function [43].

There are many works that use artificial intelligence techniques. The Gradient Boosting model is a machine learning technique that can be used in both regression and classification problems. The techniques of gradient boosting use an ensemble of weak models, in the case of this work trees, which together allow forming a stronger model. The ensemble is constructed in a stage-wise process by gradient descent in function space. The final model is a function that takes as input a vector of attributes $x \in R^n$ to get a score $F(x) \in R$ so that $F_i(x) = F_{i-1}(x) + \gamma_i h_i(x)$, where each $h_i$ is a function that models a single tree and $\gamma_i \in R$ is the weight associated with the $i$-th tree, so that these two terms are learned during the training phase [43].

On the other hand, one of the reasons for using GBT in contrast to other predictive models is that ensemble methods, in general, are usually the classifiers/regressors that deliver the best *out-of-the-box* results. In addition, considering that the basic model used to study the problem has been a classic decision tree, it has been considered natural to use its extension by means of an ensemble method.

One of the advantages of ensemble models is that their tuning process is reliable and easy when compared to other approaches such as artificial neural networks [3]. On the other side, a potential disadvantage of ensemble methods is the lack of interpretability [42]; however, this problem is also shared by other models such as neural networks, and in the case of GBT this can be ameliorated because of the ability to easily extract the importance of each feature, as done in this paper.

## 3. Materials and Methods

High-speed machining is a type of milling operation. Milling is defined in previous works [44] as the process in which the cutting tool rotates at a fixed speed while linearly moving in a perpendicular direction to the axis of the tool.

On the field of machining processes, there are many inputs and parameters that have an influence on the resulting surface roughness; some of them can be controlled while others cannot be directly controlled [2]. The datasets from these processes are usually of limited size, mainly due to the high costs of machining tests [44]. This section describes the experimentation parameters and the sizes of the used datasets. Thus, we now present an overview of our methodology for replicability considerations:

(1) Experiment Preparation. Two different kinds of experiments (Exp-1 and Exp-2) were implemented. Accordingly, two types of probes have been used. For Exp-1, we prepared test pieces of steel F-114 (F-114 is the Spanish notation for this steel (http://www.splav-kharkov.com/en/e_mat_start.php?name_id=87) with 175-220 Brinell of hardness) with dimensions of 190x100x20 mm. For Exp-2, we prepared test pieces of aluminum known as Planoxal (hardness 85-150 Brinell), with dimensions of 180x110x20 mm. Since aluminum is more malleable than steel, in Exp-2 it was important to know the hardness of the material, and so the hardness (HB) was measured before the process in each test specimen of Exp-2.

(2) Experimentation. This step consists of performing the milling for each one of the prepared experiments. The different setup of Exp-1 and Exp-2 are made using all the combinations of values for the predictor variables, these are described in Table 1. In all cases, the surface roughness was measured with the Karl Zeiss Surfcom 130 stylus profilometer.

(3) Preprocessing. In this step, the Exp-1 and Exp-2 datasets are prepared. In detail, this step consists of computing the average values of surface roughness (Ra) and associating this result with the experimental conditions (i.e., the values of the predictor variables).

(4) Model Generation. In this step the models for Exp-1 and Exp-2 are generated through machine learning; the details are in Sections 3.2 to 3.4.

(5) Analysis of Results. This analysis is done in two dimensions: (1) checking the quality of the models and (2) validating its practical utility. For (1), classical machine learning metrics (e.g., recall, precision, and accuracy) were used to measure the performance obtained in the classification of surface roughness. For (2), the validation is done by comparing results (the GBT models) with other classifiers that have been used in the literature.

*3.1. Experimental Description.* As mentioned before, two different experiments were designed to obtain the experimental data: one for machining slots on steel F-114 test specimens (Exp-1) and another for machining geometries using aluminum alloy (Exp-2). Each of these two experiments was performed initially in a machining center and later in a second machining center with different characteristics to the first, to validate the experimental design and the predictive model of surface roughness obtained in the first machining center. Also, in both experiments the machining centers were equipped with high-pressure coolant fluid: Houghton HOCUT b-750 cutting oil at 5%, a type of coolant fluid frequently used in the industry for its high quality and anticorrosive properties [45].

Predictive models have been used by authors on this domain in order to analyze the behavior of machines in particular cutting conditions. For example, Pimenov et al. [35] declare that a predictive model of surface roughness has the ability to understand new machining working conditions. In our case, predictive models help us analyze the value of surface roughness in new experimental cases. A descriptive model is not adequate in this case because it does not directly provide the ability to predict new cases.

The experimentation in HSM is generally expensive; thus, for the experimentation described in this paper the experimental model described in [9, 23] has been used and the same machining centers described in [9]. In detail, the first process of obtaining experimental data was performed with a Kondia HS1000 machining center (hereinafter referred to as M1) with CNC Siemens 840D, axis = 3, spindle speed of 24000 rpm, and maximum power of 17.5 KW and the second on a Versa model machine (variant 675004) (hereinafter referred to as M2) with CNC Heidenhain TNCi530, axis = 5, Spindle speed of 15000 rpm, and maximum power of 50.0KW.

In each of the experiments (Exp-1 and Exp-2) conditions were set related to the tool, machine (cutting conditions), and material to be machined or test specimen. In that order, said characteristics are described below for each experiment. Table 1 shows the information of the variables used on Exp-1 and Exp-2; the symbol, units, and admitted values are described for each variable.

Exp-1 was designed to generate linear cuts over steel probes and measure surface roughness on a linear surface; the tool characteristics were its diameter and the number of flutes. Karnasch tools (models 30.6455 and 30.6465) of different diameters were used. Considering standard machining of slots in the manufacture of molds, four different diameters of tools were used: 6,8,10 and 12 mm, and for each type of diameter, variations of 2 and 6 flutes were used per tool (flutes). Two slots lengths were made for each diameter and flutes variation (2x4x2), for a total of 16 experimental-tool combinations.

The characteristics of the cutting conditions are the axial depth of cut (ap), advance speed/feed rate (F), and spindle rotation speed (n). The machining of slots was done with F = 1500 mm/min and n = 5000 rpm initially, and then increments of 25, 50, and 75% of the initial F and n were applied. For each of the experimental-tool combination (described above) variations of ap (0.2, 0.4, 0.6, and 1.0), F (1500, 1875, 2250, and 2625), and n (5000, 6250, 7500, and 8750) were applied (see Table 1). Thus, the complete experimental set included 124 different conditions. All the tests were repeated on each cutting-machine to increase the amount of data to obtain a dataset of 270 samples for M-1 and 150 samples for Machine-2; for each experiment the incomplete records were removed leaving a training dataset of 251 records and 123 records, respectively.

Examples of combinations of the variables described above (Table 1) are shown in Table 2; these values correspond to the parameters used in the tests and to the calculation of the surface roughness (postprocess) according to the way of calculating the Ra value described in the introduction. In synthesis, for Exp-1, five predictive variables and the surface roughness class were used. Both the class and the rest of the continuous variables were discretized as shown in Table 2 and based on the experimental design and the discretization described on previous works such as [4, 9, 16, 23].

In order to calculate the surface roughness average (Ra), partial values of surface roughness in a slot were measured. In order to do this, contact profilometers were used such as described in [4, 23]. For each slot, four partial measurements of surface roughness were made, then a final value of surface roughness per slot was obtained by averaging. The continuous values of surface roughness were grouped in ranges; these ranges were created based on the following criteria: $next\_value = (previous\_value + 60\%\ previous\_value) + dX$, where $dX$ is a variation that considers the error margin from contact measurement (by manual profilometer).

In accordance with ISO: 4288 (1996) and ISO: 1302 (2002), there are several discrete values for Ra that are related to continuous values (all values in nanometers) [4, 23]: 0.10 = Mirror, 0.20 = Polished, 0.40 = Ground, 0.80 = Smooth, 1.60 = Fine, 3.20 = Semifine, 6.30 = Medium, 12.50 = Semirough, and 25.0 and 50.0 = Clear. Thus, in accordance with previous works such as [4, 9, 16, 23], the labels Smooth, Fine, Semifine, and Medium were created for surface roughness average as shown in Table 2. Ra values greater than 10.1 were discarded

Table 2: Examples of machining cutting conditions and Ra values in Exp-1.

| Machine | diam (mm) | flutes | ap (mm) | F (mm min$^{-1}$) | n (rpm) | Ra (nm) |
|---|---|---|---|---|---|---|
| 1 | 6 | 2 | 0.2 | 1500 | 5000 | Smooth |
| | 6 | 6 | 0.4 | 1500 | 6250 | Smooth |
| | 8 | 2 | 0.6 | 1875 | 5000 | Semi-Fine |
| | 8 | 6 | 1 | 2250 | 6250 | Smooth |
| | 10 | 6 | 0.2 | 2625 | 5000 | Semi-Fine |
| | 10 | 2 | 0.4 | 1875 | 8750 | Fine |
| 2 | 6 | 2 | 0.2 | 2250 | 5000 | Smooth |
| | 6 | 2 | 0.4 | 2250 | 7500 | Semi-Fine |
| | 6 | 6 | 0.6 | 1500 | 8750 | Smooth |
| | 8 | 2 | 1 | 1875 | 6250 | Fine |
| | 8 | 6 | 0.2 | 2250 | 5000 | Smooth |
| | 10 | 6 | 0.4 | 2625 | 7500 | Smooth |

Table 3: Discretization of variables of the Exp-1 model.

| State | F [lower, upper) | n [lower, upper) | ap | Ra [lower, upper) |
|---|---|---|---|---|
| 0 | [1000, 1500] | [4000, 5000] | 0.2 | Smooth = [0.70, 1.59] |
| 1 | (1500, 1875] | (5000, 6250] | 0.4 | Fine = [1.6, 3.20) |
| 2 | [1875, 2250] | (6250, 7500] | 0.6 | Semi-Fine = [3.20, 6.30) |
| 3 | [2250, 2626) | (7500, 8750) | 1 | Medium = [6.30, 10.10] |

because these values are of little use for the metal parts manufacturing industry or the aerospace industry. Examples from the datasets and the discrete values of surface roughness average are shown in Table 2. Table 3 shows the ranges and discrete values for variables and for the surface roughness average in the case of the slots.

Exp-2 was designed to generate nonlinear cuts (geometries) to measure surface roughness on a radial surface. The tool characteristic was tool diameter (diam); tools Karnash of different diameters (10, 12, 16, and 20 mm) were used, but the same number of flutes (4 flutes). The characteristics of the cutting conditions for Exp-2 are ae, ap, f, and n (see Table 1). For Exp-2, circumferences with a radius of 3.5 cm and height of 1,5 cm were initially considered and two different radial cut-depths were made (1 mm and 0.50 mm).

The machining of geometries was performed with n = 8000 rpm initially; after making increments of 20% for each subsequent value the set n = {8000, 9600, 12000, 15000} was generated. Also, the initial value of F was 500 and two representative values based on the analysis of previous experimental results were selected; thus, the set F = {500, 675, 750} was considered.

As it has been said above, the surface roughness labels were assigned in accordance with the average roughness established in ISO: 4288 and ISO: 1302 (see Table 3). For each machined circumference, six partial measurements of surface roughness were made, then, a final value of surface roughness per geometry was obtained by averaging. Thus, the complete experimental set for geometries included 164 different conditions. As discussed in the previous description of Exp-1, all the tests were repeated on each cutting-machine to increase the amount of data. After excluding incomplete data, a total of 431 records were obtained for M-1 and a total of 242 records were obtained for M-2. Examples from the datasets and the discrete values of surface roughness are shown in Table 4.

A summary of both the class labels and the continuous variables discretization (described above) is shown in Table 5. As a summary, Table 6 shows values related to the class and the dataset in each case.

*3.2. Implementation of the Models.* To obtain the models described above, RapidMiner Studio 7.6 ® (free version) will be used. This tool allows for obtaining various machine learning models from the data [46]; in particular, we can use it to build GBT models. The workflow used with this tool is shown in Figure 1. This process is applied to each one of the four distinct datasets evaluated.

The datasets have been acquired following the experimental design described in [9, 23] and those datasets have been prepared as described in Section 3.1 of this document. Therefore, the workflow in this work starts in the Read Data task (see Figure 1).

*3.3. Evaluation of the Methods.* In this research, a series of methods will be evaluated, measuring the performance obtained in the classification of the variables of interest. The different evaluation metrics to be used in this work will be detailed below; in particular, the performance indicators to be

Table 4: Examples of cutting conditions of geometries and Ra values in Exp-2.

| Machine | F (mm min$^{-1}$) | diam (mm) | ae (mm) | ap (mm) | HB | n (rpm) | Ra (nm) |
|---|---|---|---|---|---|---|---|
| M1 | 500 | 10 | 10 | 10 | 111 | 8000 | Smooth |
| | 675 | 10 | 10 | 5 | 110 | 8000 | Smooth |
| | 750 | 10 | 5 | 10 | 111 | 8000 | Smooth |
| | 550 | 12 | 10 | 5 | 112 | 8000 | Smooth |
| | 675 | 12 | 5 | 10 | 111 | 8000 | Fine |
| | 750 | 12 | 5 | 5 | 110 | 8000 | Fine |
| | 500 | 12 | 10 | 10 | 111 | 9600 | Fine |
| | 675 | 12 | 5 | 5 | 111 | 9600 | Fine |
| | 750 | 12 | 10 | 10 | 111 | 9600 | Fine |
| M2 | 500 | 12 | 10 | 10 | 85 | 12000 | Smooth |
| | 675 | 12 | 10 | 5 | 85 | 12000 | Smooth |
| | 750 | 12 | 10 | 10 | 85 | 15000 | Fine |
| | 500 | 16 | 10 | 5 | 85 | 15000 | Smooth |
| | 675 | 16 | 5 | 10 | 85 | 15000 | Smooth |
| | 750 | 16 | 5 | 5 | 85 | 15000 | Smooth |
| | 500 | 20 | 5 | 10 | 85 | 15000 | Smooth |
| | 675 | 20 | 5 | 5 | 87 | 15000 | Smooth |

Table 5: Discretization of the continuous variables of the Exp-2 model.

| State | HB [lower, upper) | ae | ap | Ra [lower, upper) |
|---|---|---|---|---|
| 0 | [85, 100) | 0.5 | 5 | Smooth = [0.7, 1.6) |
| 1 | [100, 115) | 0.5 | 5 | Fine = [1.6, 3.2) |
| 2 | [115, 130) | 1 | 10 | Semi-Fine = [2.1, 4.1) |
| 3 | [130, 145) | 1 | 10 | Ground = [4.1, 6.1) |

Table 6: Additional information for each dataset.

| Data set | Description of the data |
|---|---|
| Slots | Maximal number of slots per test piece: 31 |
| | Minimum number of slots per test piece: 15 |
| | Number of experiments (4 variations of ap for each slot): 462 |
| | Number of partial measurements of surface roughness: 1848 |
| Geometries | Maximal number of geometries per test piece: 8 |
| | Minimum number of geometries per test piece: 6 |
| | Number of experiments (2 variations of ae for each geometry): 768 |
| | Number of partial measurements of surface roughness: 1638 |

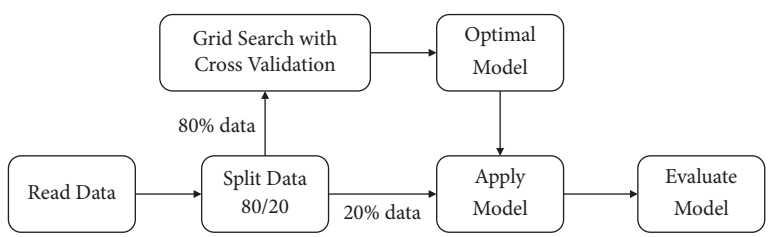



Figure 1: The workflow implemented in RapidMiner.

Table 7: Confusion matrix.

| | | True Class | |
|---|---|---|---|
| | | Positive | Negative |
| Predicted Class | Positive | $a$ | $b$ |
| | Negative | $d$ | $d$ |

used to compare methods are *recall*, *precision*, and *accuracy*. These metrics are standard in the machine learning literature and the presentation made in the following paragraphs is based on the work of Sammut [47].

Based on the data obtained in the experiments, we will obtain a confusion matrix. This matrix will facilitate the analysis needed to determine where classification errors occur. The confusion matrix is a table that shows the distribution of errors in the different categories. The performance indicators necessary to evaluate the performance of the classifier to be implemented, specifically *accuracy*, *recall*, and *precision,* will be calculated using this matrix. An example of the structure of this matrix is shown in Table 7 for the case of two classes (positive and negative in this example).

$a$ is the number of correct predictions for positive instances, $b$ is the number of incorrect predictions for negative instances, $c$ is the number of incorrect predictions for positive instances, and $d$ is the number of correct predictions for negative instances. The simplest indicator to evaluate the performance of a classifier is *accuracy (acc)*, corresponding to the ratio of correctly classified examples on the total of examples in the dataset [48]. This indicator can be calculated based on the data of the confusion matrix according to (1) (it is assumed that the dataset is not empty).

$$acc = \frac{a + d}{a + b + c + d} \tag{1}$$

The other indicators, that is, *precision* and *recall*, are understood as measures of relevance. *Precision* is the proportion of true positives *(a)* among the elements predicted as positive. Conceptually, *precision* refers to the dispersion of the set of values obtained from repeated measurements of a quantity. Specifically, a high *precision* value ($p$) implies a low dispersion in the measurements. *Recall* ($r$) is the proportion of true positives predicted among all elements classified as positive, that is, the fraction of relevant instances that have been classified. *Precision* and *recall* are calculated according to (2) and (3) (assuming $a + b \neq 0$ and $a + c \neq 0$, respectively).

$$p = \frac{a}{a + b} \tag{2}$$

$$r = \frac{a + d}{a + c} \tag{3}$$

Recall and precision and particularly well-suited for unbalanced datasets [10]. Thus, given the unbalanced nature of our data, these metrics are appropriate for the evaluation of the dataset. Meanwhile, accuracy is not necessarily useful in the case of unbalanced data (e.g., it is easy to create an "accurate" classifier by choosing the most recurring class, but this would hardly be useful); however, it allows obtaining a general view of the performance of our models when taken in the context of the other metrics.

*3.4. Parameterization of the Models.* The original data was split into 80% for training and hyperparameter tuning and 20% for testing the final model in order to obtain an unbiased estimate. The optimal model was found using K-fold cross-validation with 3-folds and using a grid search. For GBT the optimized parameters were

(i) number of trees: {10, 20, 30, 40, 50, 60, 70, 80, 90, 100};

(ii) maximal depth: {1, 2, 3, 4, 5, 6, 7, 8};

(iii) minimum rows: {4, 5, 6, 7, 8, 9, 10}.

There are other parameters in this model that remained fixed for simplicity and thus they were not optimized. In particular, they are the number of bins (20), the learning rate (0.1), and the sample rate (1.0). The optimal hyperparameters alongside its accuracy on cross-validation and the test set are shown in Table 8.

## 4. Results

*4.1. Gradient Boosted Trees for Slots.* The results obtained are shown for the slots dataset of both machines with the models given by GBT; both the confusion matrices and the models obtained in each case are presented. Table 9 shows the results obtained with GBT for the slots dataset of M-1. The final accuracy is 78.00% on the test set. The results seem to be mainly balanced; however, the Semifine class presents the lowest precision and recall of each one of the classes.

Table 10 shows the importance of each of the variables with respect to the slots dataset M1. As can be seen here, the most important variable corresponds to axial cut-depth (ap), followed by the rotation speed (n). On the other hand, the diameter of the tool (diam), the feed rate (F), and the number of teeth (flutes) are considered less relevant for a prediction according to the analysis carried out in this dataset, since they take importance values of around 20% or less.

Table 11 shows the results obtained with GBT for the slots dataset of M-2. In particular, it should be noted that the final accuracy is 61.54%, the lowest one in all the performed experiments. The failure of the model seems to occur with the Semifine class, which by analyzing the results of the confusion matrix seems to be hard to distinguish from the Fine class.

Table 12 shows the importance of each of the variables with respect to the M2 slots dataset. As can be seen here, the variable of greater importance corresponds to the axial cut-depth (ap) and then the diameter of the tool (diam) in second place. Noting that all other variables, except those two, provide an importance value lower than 20%, they are considered of less importance for the predictive capacity in this model.

*4.2. Gradient Boosted Trees for Geometries.* The results obtained for the geometries dataset of both machines with the models given by GBT are shown below; both the confusion matrices and the models obtained in each case are presented.

Table 8: List of hyperparameters used for both machines with each model and general results.

| Experiment | Number of Trees (per class) | Maximal Depth | Minimum Rows | Accuracy on Grid Search Cross-Validation | Accuracy on Test Set |
|---|---|---|---|---|---|
| Slots (M1) | 40 | 4 | 5 | 82.51% +- 1.76% (82.50% micro average) | 78.00% |
| Slots (M2) | 60 | 6 | 4 | 89.67% +- 4.64% (89.62% micro average) | 61.54% |
| Geometries (M1) | 100 | 8 | 10 | 87.51% +- 2.93% (87.50% micro average) | 88.51% |
| Geometries (M2) | 40 | 4 | 8 | 89.57% +- 0.74% (89.58% micro average) | 85.71% |

Table 9: Confusion matrix for the slots dataset M1 (GBT).

| | True Smooth | True Fine | True Semi-Fine | True Medium | Class Precision |
|---|---|---|---|---|---|
| Pred. Smooth | 6 | 0 | 0 | 0 | 100.00% |
| Pred. Fine | 3 | 19 | 4 | 0 | 73.08% |
| Pred. Semi-fine | 1 | 1 | 6 | 2 | 60.00% |
| Pred. Medium | 0 | 0 | 0 | 8 | 100.00% |
| Class Recall | 60.00% | 95.00% | 60.00% | 80.00% | *Acc: 78.00%* |

Table 10: Importance of the variables for the slots dataset M1.

| Variable | Importance | Relative Importance | Percentage (%) |
|---|---|---|---|
| Axial Depth of Cut | 219.5075 | 1.0000 | 47.15% |
| Ration Speed | 107.3826 | 0.4892 | 23.06% |
| Diameter of the Tool | 86.7459 | 0.3952 | 18.63% |
| Advance Speed | 49.5846 | 0.2259 | 10.65% |
| Number of Teeth | 2.3563 | 0.0107 | 0.51% |

Table 11: Confusion matrix for the slots dataset M2 (GBT).

| | True Smooth | True Fine | True Semi-Fine | True Medium | Class Precision |
|---|---|---|---|---|---|
| Pred. Smooth | 2 | 1 | 0 | 66.67% | 2 |
| Pred. Fine | 0 | 12 | 4 | 75.00% | 0 |
| Pred. Semi-fine | 0 | 5 | 2 | 28.57% | 0 |
| Class Recall | 100.00% | 66.67% | 33.33% | *Acc: 61.54%* | 100.00% |

Table 12: Importance of the variables for the slots dataset M2.

| Variable | Importance | Relative Importance | Percentage (%) |
|---|---|---|---|
| Axial Depth of Cut | 62.5060 | 1.0000 | 38.09 |
| Diameter of the Tool | 51.2562 | 0.8200 | 31.23 |
| Rotation Speed | 23.4585 | 0.3753 | 14.29 |
| Number of Teeth | 17.0898 | 0.2734 | 10.41 |
| Advance Rate | 9.8038 | 0.1568 | 5.97 |

Table 13 shows the results obtained with GBT for the geometries dataset of M-1. In particular, it should be noted that the final accuracy is 88.51%. The best classification obtained has been for the label "Fine", with a 100% of precision for the 97.22% of the total of the cases for this classification. Note that again the metrics present a good performance for a problem of four classes (compared to a random choice classifier). The results, in this case, seem to be mostly balanced, with no class bringing down the performance in a major way.

Table 13: Confusion matrix for the M1 geometries dataset (GBT).

| | True Smooth | True Fine | True Semi-Fine | True Medium | Class Precision |
|---|---|---|---|---|---|
| Pred. Smooth | 23 | 3 | 0 | 0 | 88.46% |
| Pred. Fine | 0 | 12 | 1 | 3 | 75.00% |
| Pred. Semi-fine | 0 | 0 | 35 | 0 | 100.00% |
| Pred. Medium | 2 | 1 | 0 | 7 | 70.00% |
| Class Recall | 92.00% | 75.00% | 97.22% | 70.00% | *Acc: 88.51%* |

Table 14: Importance of the variables for the M1 geometries dataset.

| Variable | Importance | Relative Importance | Percentage (%) |
|---|---|---|---|
| Hardness of the Material | 413.6821 | 1.0000 | 45.82 |
| Diameter of the Tool | 188.2179 | 0.4550 | 20.85 |
| Advance Speed-Feed Rate | 122.5317 | 0.2962 | 13.57 |
| Rotation Speed | 118.9302 | 0.2875 | 13.17 |
| Radial Depth of Cut | 39.7190 | 0.0960 | 4.40 |
| Axial Depth of Cut | 19.8198 | 0.0479 | 2.20 |

Table 15: Confusion matrix for the M2 geometries dataset (GBT).

| | True Smooth | True Fine | True Semi-Fine | True Medium | Class Precision |
|---|---|---|---|---|---|
| Pred. Smooth | 6 | 0 | 0 | 0 | 100.00% |
| Pred. Fine | 3 | 14 | 0 | 2 | 73.68% |
| Pred. Semi-fine | 1 | 0 | 14 | 0 | 93.33% |
| Pred. Medium | 0 | 1 | 0 | 8 | 88.89% |
| Class Recall | 60.00% | 93.33% | 100.00% | 80.00% | *Acc: 85.71%* |

Table 16: Importance of the variables for the M2 geometries dataset.

| Variable | Importance | Relative Importance | Percentage (%) |
|---|---|---|---|
| Diameter of the Tool | 169.6404 | 1.0000 | 33.13 |
| Rotation Speed | 119.9600 | 0.7071 | 23.42 |
| Hardness of the Material | 97.2817 | 0.5735 | 19.00 |
| Radial Depth of Cut | 92.2594 | 0.5439 | 18.02 |
| Advance Speed-Feed Rate | 17.5635 | 0.1035 | 3.43 |
| Axial Depth of Cut | 15.4129 | 0.0909 | 3.00 |

Table 14 shows the importance of each one of the variables with respect to the M1 geometries dataset. As can be seen here, the most important variable corresponds to the hardness of the material (HB), followed by the diameter of the tool (diam). The rest of the variables seem to be less relevant according to the analysis carried out in this dataset, being below a 20% of importance on this experiment.

Table 15 shows the results obtained with GBT for the geometries dataset of M-2. In particular, it should be noted that the final accuracy is 85.71%. Similar to the results of the classification in M1, the best classification obtained has been for the label "Fine", with a 93.33% of precision for the 100% of the total of cases. In general, the result of M2 is similar to the one from M1. Again, in this case, there does not seem to be any class that is bringing the classification results down in any major way. Although the precision of the Fine class is lower than the other ones, the recall of the Smooth class is also comparably lower than the results of the other ones.

Table 16 shows the importance of each of the variables with respect to the M2 geometries dataset. As can be seen here, the most important variable corresponds to the diameter of the tool (diam) and then rotation speed (n). All the other variables have a lower than 20% importance, with "feed rate" and "ap" being particularly low. Note that "ap" is the variable of least importance in both machines, while the diameter seems to be important in both cases.

Figure 2 summarizes the importance of the variables according to the GBM models used for all experiments and machines. In particular, Figure 2 shows that in Exp-1 the axial depth of cut (ap) is the most important variable, but for Exp-2 the hardness of the material (HB) and diameter of the tool (diam) are the most important variables, although arguably the rotation speed (n) could be considered important too, depending on the machine. These results highlight the importance of the particular characteristics of each one of machining centers, because despite using the same working

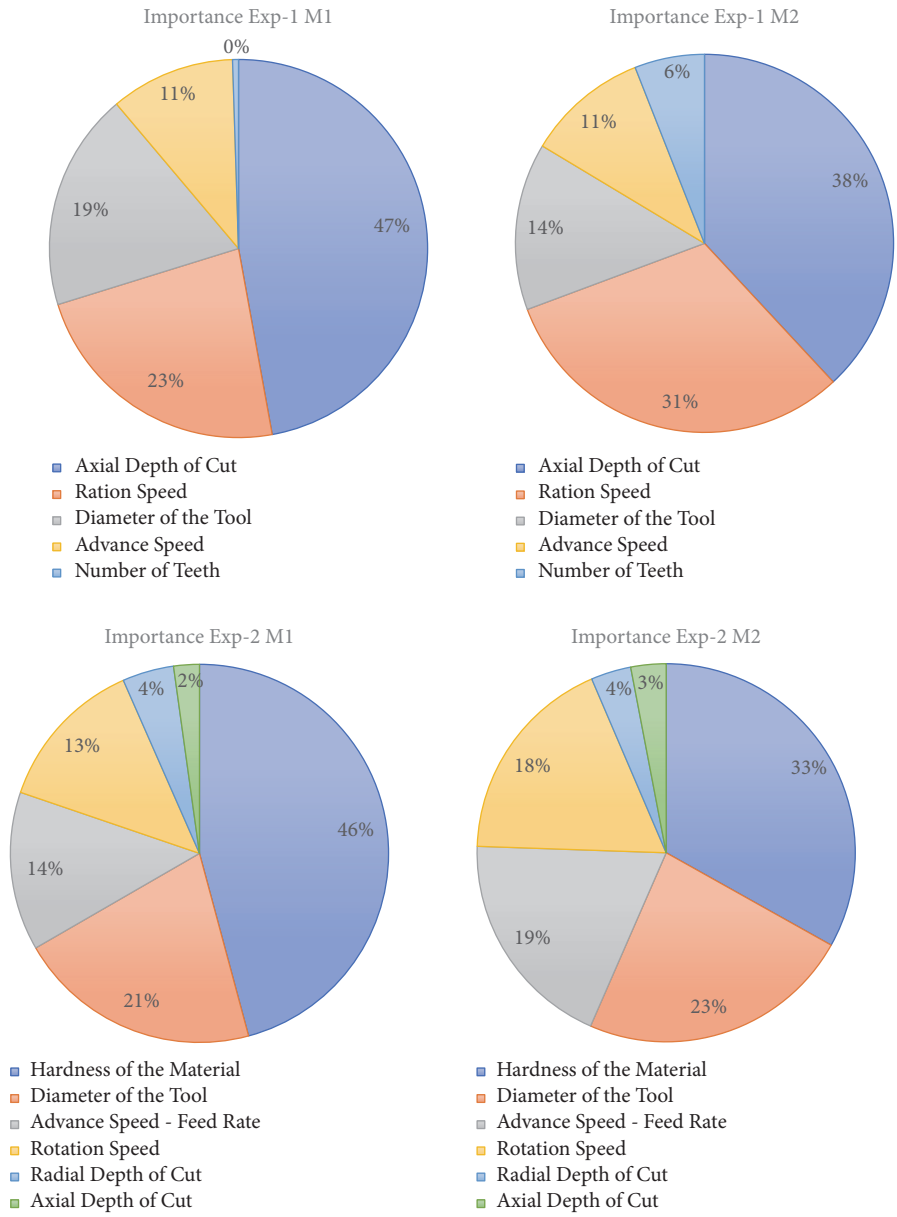



Figure 2: Pie charts for the importance of variables in each experiment and machine.

model in both machining centers (same forces, "ap", etc.) the results are very different in each case.

*4.3. Comparison with Other Methods.* We compare the GBT models with other classifiers that have been used in the literature, such as SVM with RBF kernels or other powerful classifiers like Random Forests. In order to have a fair comparison, we find the optimal parametrization for each one of the compared algorithms using the same strategy as before.

For SVM we used a 1-vs-1 scheme for multiclass classification and we tried two kinds of kernels (RBF and Linear) with the parameters C in the range (0.00001, 0.0001, 0.001, 0.01, 0.1, 1.0, 10, 100, 1000, 10000, 100000) and gamma in the range (0.00001, 0.0001, 0.001, 0.01, 0.1, 1.0, 10, 100, 1000, 10000, 100000). For Random Forests we considered as hyperparameters the Number of Trees (range: 10, 20, 30, 40, 50, 60, 70, 80, 90, 100) and Maximal Depth (range: 1, 2, 3, 4, 5, 6, 7, 8), which are analogous to the parameters of GBT.

The optimal hyperparameters are then used to train 10 classifiers using a hold-out strategy with randomly sampled data from the original dataset. The idea is now to find an estimate of the accuracy of each classifier on the test set and average the results of the 10 classifiers for the different samples. All methods were evaluated using accuracy. We perform a 1-factor (i.e., choice of method) Analysis of Variance (ANOVA) for each dataset and report the corresponding p-values.

Table 17: Comparison between models in each dataset.

| Dataset | GBT | SVM | Random Forests | p-value |
|---|---|---|---|---|
| Slots M1 | Accuracy: 83.20% +- 6.27% (40, 4, 5) | Accuracy: 80.20% +- 5.02% (RBF, 10, 0.10) | Accuracy: 87.60% +- 6.64% (80, 5) | 0.2746 |
| Slots M2 | Accuracy: 72.69% +- 7.19% (60, 6, 4) | Accuracy: 75.77% +- 4.88% (RBF, 10, 1.0) | Accuracy: 77.31% +- 9.80% (30, 8) | 0.4308 |
| Geometries M1 | Accuracy: 85.29% +- 2.56% (100, 8, 10) | Accuracy: 82.76% +- 2.47% (RBF, 1000, 1.0) | Accuracy: 86.44% +- 3.20% (30, 8) | 0.0263 |
| Geometries M2 | Accuracy: 83.68% +- 4.98% (40, 4, 8) | Accuracy: 83.27% +- 4.64% (RBF, 1000, 1.0) | Accuracy: 86.12% +- 4.26% (20, 8) | 0.3817 |

The results are shown in Table 17 alongside the optimal parametrization in each case, for SVM the optimal hyperparameters are shown in the format given by the tuple (Kernel, C, gamma), for GBT we have the format (Number of Trees, Maximal Depth, Minimum Rows), and for Random Forests we have the format (Number of Trees, Maximal Depth).

The ANOVA reveals that there is no statistically significant difference for the slots (M1 and M2) and Geometries M2, while there is a potentially significant difference in the performance of the methods on Geometries M1. Further inspection reveals that there's a difference between RBF SVM and the other two methods, but there's no significant difference between GBT and Random Forests. This suggests that the results from GBT are competitive with other classifiers in the state of the art.

## 5. Discussion

Previous works generate surface roughness predictive models using several soft computing techniques; however, there are no standard models to predict surface roughness, the models being usually generated under specific conditions of machining, coolant, machine tool, and tool. In the literature, it is possible to find works in which artificial neural networks or Bayesian networks are applied to generate predictive models of surface roughness; also classic decision trees have been used as a technique for pattern identification in the behavior of variables that influence surface roughness in the industry [41], but not many works were found where techniques based on decision trees are used to predict surface roughness (such as, for example, Random Forest or Gradient Boosted Trees).

A recently published work is [35] where they use random forest (RF), multilayer perceptron (MLP), regression trees (RT), and radial base functions (RBF); this paper presents a comparative study of the surface roughness prediction quality; RF is the one that provides the best results in terms of *accuracy*, followed by RT and MLP. In our work, we use GBT, obtaining better results in terms of accuracy than similar works performed with the same experimental design and reported, for example, in [4, 23].

In this work, we use real training data and we have also obtained a graphical representation of knowledge using classic decision trees to complement the results obtained by GBT; in this way the joint result provides greater graphic expressivity regarding conditional influences and the values of the predictor variables on the class labels than, for example, Bayesian networks. This is important since it can be easily used to create a domain representation model and can also be interpreted to generate rules that contain dynamic knowledge of the machining process, which facilitates the construction of knowledge and inference bases for an eventual expert system of surface quality prediction in real time under concrete tool conditions, material to be machined, and machine tool.

All the pieces-of-knowledge derived from this research and other obtained from previous related works can be used, for example, to generate inference rules that help to establish the impact of measures of predictor variables on the surface roughness class. Figure 3 shows the syntax diagram for if-statements, based on what was said previously. For example, if_statement_1 shows that the axial depth of cut (ap) has the highest influence on workpiece roughness, but if the machine is M-1, then the least important variable is the number of teeth (flutes), while for M-2 the variable of least importance is the advance speed (F). The If-Then statement described above can be generalized for mechanical cutting with machines that have similar characteristics to M-1 or M-2, and some characteristics of the cutting process using steel F-114 can now be tested.

Finally, the ability of Gradient Boosted Trees to determine the importance of each variable with respect the labels could be considered similar to the ability of Bayesian networks to model influences between these variables and the labels. This is very important in the domain of micro- and nanomachining, where precision in the machining influences heavily the final results. Again, the knowledge gained here can be combined with previous knowledge that has been obtained by elicitation from experts or from state of the art, analysis of results, among others, so that this knowledge can be represented in a knowledge base and used by a Rule-based System or Decision Support System. In order to obtain conclusions about surface roughness as influenced by the predictive variables or to be able to predict surface roughness given a dataset or a particular data record.

## 6. Conclusions

The integration of AI algorithms into computational solutions and techniques for analyzing large amounts of data are gaining interest in the modern industry. This integration is part of what some authors call the sixth technological revolution [27]. In keeping with this idea, this document provides a model of the surface roughness average (Ra) as a

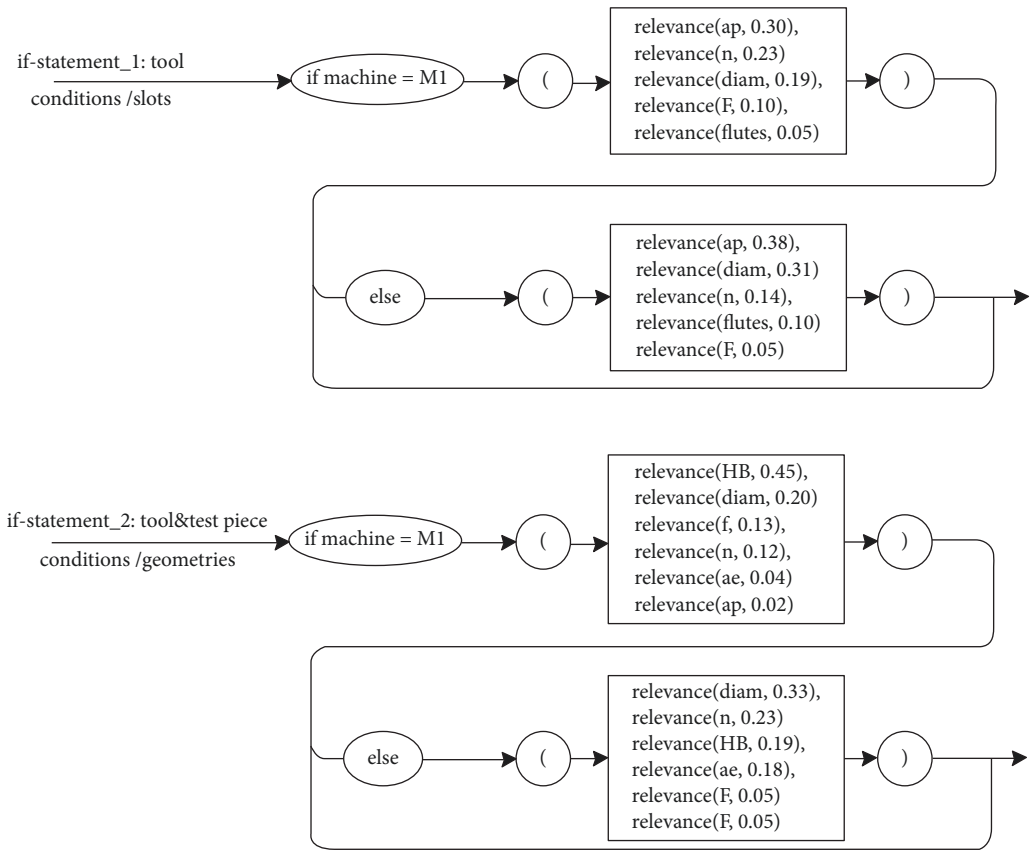



Figure 3: Syntax diagrams for two if-statements related to machining conditions and the relevance of variables for M-1 and M-2.

result of a predictive analysis using Gradient Boosted Trees. The models presented in this work evaluate the influence of cutting conditions, the characteristics of two different end-milling machines, and the material (steel and aluminum test pieces) on the surface quality of high-speed machining.

Previous works, such as [1, 32], provide mathematical models to predict surface roughness considering operational characteristics of HSM; the second group of works provide surface roughness models using various soft computing and AI techniques; specifically, they apply ANNs (for example, [17, 24]), Genetic Algorithms (for example, [1, 25]), or Bayesian networks (for example, [4, 13, 21]). Generally, these models consider aspects related to the tool, the machine, and the material to be machined, and some other works incorporate other aspects such as the lubricant or coolant (for example, the works in [15, 49]).

As a soft computing and AI approach, this work falls in the second group, with the advantage that the resulting model in each experiment has a high accuracy value in comparison with other techniques such as decision trees, thanks to the accuracy of the GBT algorithm. A potential improvement that could be made to our model is incorporating the characteristics of the coolant used and study the behavior of the conditional influences on the surface roughness.

The main contributions of this work are the predictive model itself and the subsequent analysis of variable importance. In particular, our results show accuracies ranging from 61.54% to 88.51% on the datasets, which are competitive results when compared with the other approaches shown in this paper. An important advantage of applying this model is that we have been able to analyze which variables have the most impact on the predictive ability of our model in a natural way. In this context, we find that the axial cut-depth is the most influential feature for the slots datasets. The axial cut deep is the axial contact length between the cutting tool and the workpiece [44]; this means that when developing an experimental model for working with steel F-114, the influence of the cut-depth and the hardness of the material must be carefully considered. On the other hand, the hardness and the diameter of the cutting tool are the most influential features for the geometries datasets. Thus, similar considerations as before could be held for these variables in this case.

As a potential future line of work and for practical applications, the results of this study could form the basis of a decision support system. In particular, such a system could use the knowledge contained in the predictive models as a core knowledge base. In particular, since the predictive model GBT is based on decision trees, it would be relatively easy to extract the information from the tree branches as a knowledge base. These branches could be used as rule sets for technicians and other users to interpret and understand the current behavior of the machining systems. Furthermore, this knowledge base could be enriched with more data as

it becomes available, although this avenue of work would eventually lead to concerns of scalability of such a decision support system.

Finally, as has been said in the discussion, the application of the GBT algorithm (derived from decision trees) is not very common in the industry, even though decision trees are one of the most widely used machine learning techniques because of the ease they provide in generating clear production rules and being easily understood by end users. This brings confidence to the predictive models presented here in the presence of new cases that might be taken as input to predict surface roughness.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The authors want to thank the Centro de Automática y Robótica at CSIC (Spain) for the use of the Kondia machine tool where a part of the experimentation was made. The authors also acknowledge the collaboration of Nicolas Correa S.A. and thank Dr. Andrés Bustillo from Nicolas Correa, who enabled the rest of experimentation in this company, using the Versa machining center.

## References

[1] S. P. Leo Kumar, "Experimental investigations and empirical modeling for optimization of surface roughness and machining time parameters in micro end milling using Genetic Algorithm," *Measurement*, vol. 124, pp. 386–394, 2018.

[2] P. G. Benardos and G.-C. Vosniakos, "Predicting surface roughness in machining: a review," *The International Journal of Machine Tools and Manufacture*, vol. 43, no. 8, pp. 833–844, 2003.

[3] A. Bustillo, J.-F. Díez-Pastor, G. Quintana, and C. García-Osorio, "Avoiding neural network fine tuning by using ensemble learning: Application to ball-end milling operations," *The International Journal of Advanced Manufacturing Technology*, vol. 57, no. 5-8, pp. 521–532, 2011.

[4] V. Flores, M. Correa, and Y. Quin~ones, "Performance of predicting surface quality model using soft computing, a comparative study of results," in *International Work-Conference on the Interplay Between Natural and Artificial Computation*, pp. 233–242, Springer, Corunna, Spain, 2016.

[5] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*, CRC Press, Boca Raton, Fla, USA, 2012.

[6] R. K. Das, A. K. Sahoo, P. C. Mishra, R. Kumar, and A. Panda, "Comparative machinability performance of heat treated 4340 Steel under dry and minimum quantity lubrication surroundings," *Procedia Manufacturing*, vol. 20, pp. 377–385, 2018.

[7] G. Liu, B. Zou, C. Huang, X. Wang, J. Wang, and Z. Liu, "Tool damage and its effect on the machined surface roughness in high-speed face milling the 17-4PH stainless steel," *The International Journal of Advanced Manufacturing Technology*, vol. 83, no. 1-4, pp. 257–264, 2016.

[8] H. Liu, J. Zhang, X. Xu, and W. Zhao, "Experimental study on fracture mechanism transformation in chip segmentation of Ti-6Al-4V alloys during high-speed machining," *Journal of Materials Processing Technology*, vol. 257, pp. 132–140, 2018.

[9] V. M. Flores, M. Correa, and J. R. Alique, "A pre-process model to predict the surface quality in high-speed milling," *Revista Iberoamericana de Automática e Informática Industrial RIAI*, vol. 8, no. 1, pp. 38–43, 2011.

[10] A. Bustillo and J. J. Rodríguez, "Online breakage detection of multitooth tools using classifier ensembles for imbalanced data," *International Journal of Systems Science*, vol. 45, no. 12, pp. 2590–2602, 2014.

[11] ISO-4287., "Geometrical product specifications (GPS)—surface texture: profile method—terms, definitions and surface texture parameters. international organization for standardization".

[12] P. Netake and S. Chinchanikar, "Hard turning under minimum quality lubrication: Modelling of cutting force and surface roughness," *Journal of Advanced Engineering Research*, vol. 2, pp. 17–22, 2015.

[13] A. Kothuru, S. P. Nooka, and R. Liu, "Application of audible sound signals for tool wear monitoring using machine learning techniques in end milling," *International Journal of Advanced Manufacturing Technology*, vol. 95, no. 9-12, pp. 3797–3808, 2018.

[14] G. Liu, C. Huang, B. Zou et al., "Feasibility study of taking high-speed dry milling as the final manufacturing process by the standard of service performance," *The International Journal of Advanced Manufacturing Technology*, vol. 95, no. 5-8, pp. 2897–2906, 2018.

[15] B. Boswell, M. N. Islam, I. J. Davies, Y. R. Ginting, and A. K. Ong, "A review identifying the effectiveness of minimum quantity lubrication (MQL) during conventional machining," *The International Journal of Advanced Manufacturing Technology*, vol. 92, no. 1-4, pp. 321–340, 2017.

[16] A. Bustillo and M. Correa, "Using artificial intelligence to predict surface roughness in deep drilling of steel components," *Journal of Intelligent Manufacturing*, vol. 23, no. 5, pp. 1893–1902, 2012.

[17] M. Mia, M. H. Razi, I. Ahmad et al., "Effect of time-controlled MQL pulsing on surface roughness in hard turning by statistical analysis and artificial neural network," *The International Journal of Advanced Manufacturing Technology*, vol. 91, no. 9-12, pp. 3211–3223, 2017.

[18] M. Hutson, "Artificial Intelligence prevails at predicting Supreme Court decisions," *Science*, vol. 360, no. 6383, 2018.

[19] Y. Altintas and M. Weck, "Chatter stability of metal cutting and grinding," *CIRP Annals - Manufacturing Technology*, vol. 53, no. 2, pp. 619–642, 2004.

[20] M. Chaudhari, B. F. Jogi, and R. Pawade, "Comparative study of part characteristics built using additive manufacturing (fdm)," *Procedia Manufacturing*, vol. 20, pp. 73–78, 2018.

[21] B. Lela, D. Bajić, and S. Jozić, "Regression analysis, support vector machines, and Bayesian neural network approaches to modeling surface roughness in face milling," *The International Journal of Advanced Manufacturing Technology*, vol. 42, no. 11-12, pp. 1082–1088, 2009.

[22] A. Das, A. Mukhopadhyay, S. K. Patel, and B. B. Biswal, "Comparative assessment on machinability aspects of AISI

4340 alloy steel using uncoated carbide and coated cermet inserts during hard turning," *Arabian Journal for Science and Engineering*, vol. 41, no. 11, pp. 4531–4552, 2016.

[23] M. Correa, C. Bielza, M. d. Ramirez, and J. Alique, "A Bayesian network model for surface roughness prediction in the machining process," *International Journal of Systems Science*, vol. 39, no. 12, pp. 1181–1192, 2008.

[24] M. Sekulic, V. Pejic, M. Brezocnik, M. Gostimirović, and M. Hadzistevic, "Prediction of surface roughness in the ball-end milling process using response surface methodology, genetic algorithms, and grey Wolf optimizer algorithm," *Advances in Production Engineering & Management*, vol. 13, no. 1, pp. 18–30, 2018.

[25] Y. Shrivastava and B. Singh, "Stable cutting zone prediction in CNC turning using adaptive signal processing technique merged with artificial neural network and multi-objective genetic algorithm," *European Journal of Mechanics - A-Solids*, vol. 70, pp. 238–248, 2018.

[26] C. K. Madhusudana, H. Kumar, and S. Narendranath, "Fault diagnosis of face milling tool using decision tree and sound signal," in *Proceedings of the 2018 International Conference on Materials Manufacturing and Modelling, ICMMM 2017*, pp. 12035–12044, India, March 2017.

[27] M. Knell, *Nanotechnology and the Challenges of Equity, Equality and Development. Yearbook of Nanotechnology in Society 2*, S. Cozzens and J. Wetmore, Eds., Springer, Dordrecht, Netherlands, 2010.

[28] P. Kovac, D. Rodic, V. Pucovsky, B. Savkovic, and M. Gostimirovic, "Application of fuzzy logic and regression analysis for modeling surface roughness in face milliing," *Journal of Intelligent Manufacturing*, vol. 24, no. 4, pp. 755–762, 2013.

[29] C. Zhang, W. Li, P. Jiang, and P. Gu, "Experimental investigation and multi-objective optimization approach for low-carbon milling operation of aluminum," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 231, no. 15, pp. 2753–2772, 2016.

[30] S. Werda, A. Duchosal, G. Le Quilliec, A. Morandeau, and R. Leroy, "Minimum quantity lubrication advantages when applied to insert flank face in milling," *The International Journal of Advanced Manufacturing Technology*, vol. 92, no. 5-8, pp. 2391–2399, 2017.

[31] C. Gorges, K. Öztürk, and R. Liebich, "Impact detection using a machine learning approach and experimental road roughness classification," *Mechanical Systems and Signal Processing*, vol. 117, pp. 738–756, 2019.

[32] A. E. Charola, C. Grissom, E. Erder, M. Wachowiak, and D. Oursler, "Measuring surface roughness: three techniques," in *Proceedings of the 8th International Congress on Deterioration and Conservation of Stone*, J. Riederer, Ed., pp. 1421–1434, Berlin, Germany, 1996.

[33] H. Wu, C. Li, K. Fang et al., "Effect of machining on the stress corrosion cracking behavior in boiling magnesium chloride solution of austenitic stainless steel," *Materials and Corrosion*, vol. 69, no. 4, pp. 519–526, 2017.

[34] T. J. Ko, J. W. Park, H. S. Kim, and S. H. Kim, "On-machine measurement using a noncontact sensor based on a CAD model," *The International Journal of Advanced Manufacturing Technology*, vol. 32, no. 7-8, pp. 739–746, 2007.

[35] D. Y. Pimenov, A. Bustillo, and T. Mikolajczyk, "Artificial intelligence for automatic prediction of required surface roughness by monitoring wear on face mill teeth," *Journal of Intelligent Manufacturing*, vol. 29, no. 5, pp. 1045–1061, 2018.

[36] D. Y. Pimenov, "Experimental research of face mill wear effect to flat surface roughness," *Journal of Friction and Wear*, vol. 35, no. 3, pp. 250–254, 2014.

[37] Y. Houchuan, C. Zhitong, and Z. ZiTong, "Influence of cutting speed and tool wear on the surface integrity of the titanium alloy Ti-1023 during milling," *The International Journal of Advanced Manufacturing Technology*, vol. 78, no. 5-8, pp. 1113–1126, 2015.

[38] J. Yan, Y. Murakami, and J. P. Davim, "Tool design, tool wear and tool life," in *Machining Dynamics*, K. Cheng, Ed., pp. 117–149, Springer, London, UK, 2009.

[39] L. Munkhdalai, T. Munkhdalai, O. Namsrai, J. Lee, and K. Ryu, "An empirical comparison of machine-learning methods on bank client credit assessments," *Sustainability*, vol. 11, no. 699, pp. 2–23, 2019.

[40] S. Visweswaran and G. F. Cooper, "Learning instance-specific predictive models," *Journal of Machine Learning Research*, vol. 11, pp. 3333–3369, 2010.

[41] R. Lior, "Data mining with decision trees: theory and applications," *World Scientific*, vol. 81, pp. 11–50, 2014.

[42] A. Van Assche and H. Blockeel, "Seeing the forest through the trees: Learning a comprehensible model from an ensemble," in *Proceedings of the In European Conference on Machine Learning*, pp. 418–429, Springer, Berlin, Heidelberg, Germany, 2007.

[43] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

[44] M. Grzenda and A. Bustillo, "Semi-supervised roughness prediction with partly unlabeled vibration data streams," *Journal of Intelligent Manufacturing*, vol. 30, no. 2, pp. 1–13, 2019.

[45] M. Cortina, J. I. Arrizubieta, E. Ukar, and A. Lamikiz, "Analysis of the influence of the use of cutting fluid in hybrid processes of machining and Laser Metal Deposition (LMD)," *Coatings*, vol. 8, no. 2, pp. 61–71, 2018.

[46] M. Hofmann and R. Klinkenberg, Eds., *RapidMiner: Data Mining Use Cases and Business Analytics Applications*, CRC Press, 2013.

[47] C. Sammut and G. I. Webb, *Encyclopedia of machine learning*, Springer Science & Business Media, 2011.

[48] C. K. Chow and C. N. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE Transactions on Information Theory*, vol. 14, no. 3, pp. 462–467, 1968.

[49] H. Sohrabpoor, S. P. Khanghah, and R. Teimouri, "Investigation of lubricant condition and machining parameters while turning of AISI 4340," *The International Journal of Advanced Manufacturing Technology*, vol. 76, no. 9-12, pp. 2099–2116, 2015.

Hindawi
Complexity
Volume 2019, Article ID 2089763, 13 pages
https://doi.org/10.1155/2019/2089763

*Review Article*

# Emerging Risk Management in Industry 4.0: An Approach to Improve Organizational and Human Performance in the Complex Systems

**F. Brocal**,[1,2] **C. González**,[3] **D. Komljenovic**,[4] **P. F. Katina**,[5] **and Miguel A. Sebastián**[3]

[1]*Department of Physics, Systems Engineering and Signal Theory, Escuela Politécnica Superior, Universidad de Alicante, Campus de Sant Vicent del Raspeig s/n, 03690, Sant Vicent del Raspeig, Alicante, Spain*
[2]*University Institute of Physics Applied to Sciences and Technologies, Universidad de Alicante, Alicante, Spain*
[3]*Manufacturing and Construction Engineering Department, ETS de Ingenieros Industriales, Universidad Nacional de Educación a Distancia, Calle Juan del Rosal, 12, 28040 Madrid, Spain*
[4]*Institut de Recherche d'Hydro-Québec (IREQ), 1800, boul. Lionel-Boulet, Varennes, Canada J3X 1S1*
[5]*Department of Informatics and Engineering Systems, University of South Carolina Upstate, 800 University Way, Media 211, Spartanburg, SC 29303, USA*

Correspondence should be addressed to F. Brocal; francisco.brocal@ua.es

Received 5 March 2019; Accepted 20 May 2019; Published 19 June 2019

Guest Editor: Jorge Luis García-Alcaraz



Industry 4.0 in the contemporary operating context carries important sources of complexity. This context generates both traditional risks and emerging risks that must be managed. The management of these risks includes both industrial risks and occupational risks, since they are heavily interlinked. The human factor can be considered the main link between both types of risks. Thus, understanding risks originating from human errors and organizational weaknesses as causes of accidents and other disruptions in complex systems requires elaborating sophisticated modeling approaches. Therefore, the objective of this paper is to propose an organizational and human performance approach to improve the emerging risk management linked to the complex systems, like as Human-Machine Interactions (HMI) and Human-Robot Interaction (HRI). To fulfill this objective, we first introduce the concept of emerging risk linked to human factor. Then, we introduce the concept of emerging risk management in the Industry 4.0 context. Under this complex context, we expose the concept considering the current models of risk management. Finally, we discuss how enhancing human and organizational performance can be achieved through risk management in complex systems linked to Industry 4.0. Therefore, we conclude that while Industry 4.0 brings numerous advantages, it must contend with emerging risks and challenges associated with organizational and human factors. These emerging risks include industrial risks as well as occupational risks. Moreover, the human factor aspect of Industry 4.0 is directly linked to industrial emerging and occupational emerging via context of operations. To cope with these new challenges, it is necessary to develop new approaches. One of such approaches is Complex System Governance. This approach is discussed along with the need for adequate organizational and human performance models dealing with, for example, experience from other domains such as nuclear, space, aviation, and petrochemical.

## 1. Introduction

The concept "Industry 4.0" has its origin in a "strategic initiative" of the German government in 2011 [1]. In its simplest form, this concept is a name given to the current trend of automation and data exchange in manufacturing technologies. Industry 4.0 encompasses elements of cyber-physical systems (CPS), Internet of Things (IoT), cloud computing, and cognitive computing [2]. Industry 4.0 and its synonyms, such as Smart Manufacturing, Smart Production, Internet of Things, or the 4th Industrial Revolution [3], have been identified as major contributors in the context of digital and automated manufacturing environment [4]. Kagermann et al. [1] note that such concepts are linked to smart manufacturing systems configured with digital networking of production. In

addition, the concept "advanced manufacturing" is related to a greater interval of industrial modernization. The scientific interest around these terms is clearly increasing during the last years [5].

In contrast to conventional forecast based production planning, Industry 4.0 enables real-time planning of production plans as well as dynamic self-optimisation [6]. The productive model based on Industry 4.0, in addition to improving processes in terms of efficiency and quality, can help improve safety as well as sustainability and the industry image [7]. However, industrial processes configured for advanced manufacturing processes and technologies can increase levels of complexity of production processes and introduce high dynamism into processes creating the so-called smart working environments [8, 9]. Taking the context of critical infrastructures as a reference, Zio [10] suggests that the industrial processes under the Industry 4.0 paradigm can be considered complex, made up of many components interacting in a network structure. These components can be physical and cyber-physical, functioning heterogeneously, organized in a hierarchy of subsystems, and contributing to system as a whole.

However, there is scarcity of literature discussing issue of complexity in the context of Industry 4.0. Table 1 is developed to support this thesis. The first scientific publications emerged in 2012. Thereafter, there has been increased interest in the concept of Industry 4.0.

In relation to enterprise production and operation, Industry 4.0 has four objectives [11]: environmental sustainability, safety, agility, and high efficiency. It has been suggested that Industry 4.0 environment will lead to a new revolution in the domains of safety management practices with "out of the box" thinking [4]. However, this suggests that the management of those organizations brings new challenges given significant uncertainties associated with increasing complexity [12].

Suffice to say that traditional industry has its own problems. When these problems are coupled with emerging paradigm of Industry 4.0 along with emerging complex, there emerges a need for development of rigorous and sophisticated approaches for risk management (i.e., traditional and emerging risk management). Roig and Brocal [13] suggest management of emerging risks through combination of different approaches to provide more enlightened decision-making.

Interestingly, despite an increase in number of scientific publications on the subject of "Industry 4.0" (Table 1) the number of publications addressing "risk" and "Industry 4.0" is very low. Table 2 highlights this concern with the following search in "Web of Science" (search standards used in Tables 1–3; the results are null for the following queries: TS=("Industry 4.0" and "risk management"); TS=("Industry 4.0" "emerging risk")).

The management of these emerging risks includes both industrial risks and occupational risks, since they are heavily linked [14]. The human factor can be considered the main link between both types of risks. In such cases, concepts of Human-Machine Interactions (HMI), Human-Computer Interaction (HCI), and Human-Robot Interaction (HRI) can be considered among the most important [4, 7, 15]. In an equivalent way to the results shown in Table 2, the number of scientific publications that integrate the concepts Industry 4.0, safety, and occupational safety is still very scarce, as shown in Table 3. In this way, Badri et al. [8] indicate that published research on the integration of occupational health and safety (OH&S) in the Industry 4.0 context is rarely cited.

Table 1: Number of scientific publications on complexity and Industry 4.0 (Results from the Web of Science. Timespan: 2010-2018; All databases; Field tag: Topic).

| Year | Complexity | Industry 4.0 | Industry 4.0 and complexity |
|---|---|---|---|
| 2018 | 70810 | 1555 | 69 |
| 2017 | 73548 | 1136 | 63 |
| 2016 | 82644 | 561 | 48 |
| 2015 | 67325 | 191 | 9 |
| 2014 | 59552 | 69 | 4 |
| 2013 | 55519 | 23 | 3 |
| 2012 | 51013 | 4 | 0 |
| 2011 | 48151 | 0 | 0 |
| 2010 | 46884 | 0 | 0 |

Thus, understanding risks originating from human errors and organizational weaknesses as causes of accidents and other disruptions in complex systems requires elaborating sophisticated modeling approaches. Therefore, the objective of the present research is to define an approach for organizational and human performance that can be used to improve the emerging risk management linked to the complex systems under paradigm of Industry 4.0. To obtain this objective, the rest of this paper is organized as follows:

(i) Section 2 introduces the concept of emerging risk linked to human factor in Industry 4.0 context.

(ii) Section 3 introduces the concept of emerging risk management in the Industry 4.0 context. Under this complex context, authors expose the concept considering the current models of risk management.

(iii) Section 4 is organized to provide ways to improve organizational performance with the goal of improving safety in the complex systems linked to Industry 4.0

## 2. Emerging Risks

The definitions and risk models used in the professional and scientific fields are numerous. In this regard, Aven [16] has made a classification of definitions on risk most commonly used. Said author considers that the definitions collected in Table 4 are the most relevant. Aven [16] argues that the definition or model (3) is the most appropriate.

From a standardized perspective, ISO 31000:2018 standard indicates that a risk is usually expressed in terms of risk sources, potential events, their consequences, and their likelihood [20]. In OH&S field, the ISO 45001:2018 standard defines an OH&S risk (the terms "OH&S risk" and

Table 2: Number of scientific publications on risk and emerging risk in the Industry 4.0 (Results from the Web of Science. Timespan: 2010-2018; All databases; Field tag: Topic).

| Year | Risk | Risk management | Industry 4.0 and risk | Emerging risk | Management and emerging risk |
|---|---|---|---|---|---|
| 2018 | 335.842 | 7155 | 66 | 84 | 19 |
| 2017 | 324.537 | 8226 | 55 | 88 | 21 |
| 2016 | 310.952 | 7762 | 27 | 90 | 17 |
| 2015 | 289.990 | 7210 | 7 | 71 | 19 |
| 2014 | 265.015 | 6461 | 1 | 81 | 20 |
| 2013 | 247.840 | 6173 | 0 | 74 | 17 |
| 2012 | 224.108 | 5884 | 0 | 66 | 16 |
| 2011 | 207.076 | 6050 | 0 | 62 | 14 |
| 2010 | 193.041 | 6092 | 0 | 79 | 11 |

Table 3: Number of scientific publications on safety and occupational safety in the Industry 4.0 (Results from the Web of Science. Timespan: 2010-2018; All databases; Field tag: Topic).

| Year | Safety | Industry and safety | Industry 4.0 and safety | Occupational safety | Industry 4.0 and occupational safety |
|---|---|---|---|---|---|
| 2018 | 291.426 | 13.270 | 81 | 1940 | 7 |
| 2017 | 262.601 | 14.352 | 58 | 2248 | 7 |
| 2016 | 246.359 | 11.658 | 23 | 1858 | 1 |
| 2015 | 228.041 | 11.239 | 7 | 1536 | 1 |
| 2014 | 216.289 | 10.651 | 3 | 1606 | 0 |
| 2013 | 182.296 | 10.390 | 1 | 1596 | 0 |
| 2012 | 154.996 | 9.189 | 0 | 1690 | 0 |
| 2011 | 124.767 | 8.616 | 0 | 1343 | 0 |
| 2010 | 109.836 | 7.259 | 0 | 1229 | 0 |

"occupational risks" are used as equivalent in this paper) as the combination of the likelihood of occurrence of a work-related hazardous event or exposure (s) and the severity of injury and ill health that can be caused by the event or exposure(s) [21]. These definitions are among the models on risk collected in Table 4, highlighting the adaptation of model (2) to the definition of occupational risk.

The application of these definitions and models needs adaptations and new approaches when dealing with emerging risks, which are discussed in the following sections.

*2.1. Emerging Risk Concept.* Generally, when the term "emerging risk" is mentioned, this term refers to any risk that is new and/or increasing. However, other perspectives do exist. For example, the International Risk Governance Council [22, 23] suggests that emerging risk should be viewed from a "systemic" perspective. A systemic perspective of risk includes both (i) emerging and (ii) new conditions. First, the risk is emerging when it is new in a broad sense, as, for example, in the case of new technologies, new materials, etc. Second, the risk is emerging when being familiar or traditional; it is presented under new or unfamiliar conditions (e.g., the reemergence of the poliovirus). In both cases, uncertainty is the main characteristic of emerging risk [24]. From this perspective, the emerging risk adapts especially well to model (4) (R=U). However, the IRGC [24] notes that this uncertainty is related to the probabilities and/or consequences of the emerging risk. In this way, model (2) would also be applicable.

Brocal et al. [17] have proposed a theoretical framework through which the new and emerging risk (NER) concept has been modeled, especially in industrial processes. The risk model used as a reference is the following: a risk (R) is a structure consisting of five components: the source of risk (SR), causes (C), events (E), consequences (CO), and the likelihood (L); this set may be expressed as (8):

$$(8)\ R = (SR, C, E, CO, L)$$

From this framework, Brocal et al. [25] have developed TICHNER (Technique to Identify and Characterize NERs) technique that aims to identify and characterize the NERs generated by a manufacturing system. TICHNER is based in the definition of NER considered in the reports published by EU-OSHA [26–29]. This definition has been codified by Brocal et al. [17] through the so-called conditions that define an NER (C1, C2, C3, C4, C5, and C6), which are collected in Table 5.

Given that a NER is any risk that is new and/or increasing, Brocal et al. [17] consider that a risk is new (NR, new risk) when it can be associated with conditions C1, C2, and C3. Considering model (8), condition C1 is linked to new SR and new C. The novel aspect can be both technological and organizational. C2 and C3 are linked to new SR, C, E, and CO. In this case, the novel aspect of C2 is associated with changes in social perceptions and the novel aspect of condition C3 to new scientific knowledge about risk. These authors consider that a risk is increasing (IR, increasing risk) when it can be associated with conditions C4, C5, and C6. Condition

Table 4: Main models on risk used in the professional and scientific fields (adapted from [16]).

| Model | | Description |
|---|---|---|
| (1) | R=E | Risk=Expected value (loss) |
| (2) | R=P&C | Risk=Probability and scenarios/Consequences/severity of consequences |
| (3) | R=C&U | Risk=Consequences/damage/severity of these + Uncertainty |
| (4) | R=U | Risk=Uncertainty |
| (5) | R=OU | Risk=Objective Uncertainty |
| (6) | R=C | Risk=Event or consequence |
| (7) | R=ISO | Risk=Event or consequence |

Table 5: Possible combinations between the Ci conditions and the risk components (model (8)) that can form a NER: NR and/or IR (adapted from [17]).

| | Conditions | Risk Components (model 8) | | | | |
|---|---|---|---|---|---|---|
| | | Source of Risk (SR) | Causes (C) | Events (E) | Consequences (CO) | Likelihood (L) |
| NEW | C1: New technological or organizational variable | (SR,C1) | (C,C1) | — | — | — |
| | C2: New social perception | (SR,C2) | (C,C2) | (E,C2) | (CO,C2) | — |
| | C3: New scientific knowledge | (SR,C3) | (C,C3) | (E,C3) | (CO,C3) | — |
| INCREASING | C4: Increase in the number of sources of risk | (SR,C4) | — | — | — | — |
| | C5: Increase in the likelihood of exposure | — | — | — | — | (L,C5) |
| | C6: Increase health consequences | — | — | — | (CO,C6) | — |

C4 is linked to the increase of SR, C5 to the increase of L (exposure level and/or the number of people exposed), and C6 to the increase of CO (seriousness of health effects and/or the number of people affected).

Considering the above aspects, Table 5 shows the different possible combinations between the Ci conditions and the risk components that can form an NER (NR and/or IR). Thus, depending on the combination in each case of interest, that is, the characterization of emerging risk according to Brocal et al. [25], the risk could be analyzed with one of the models in Table 4.

The terms "new and emerging risks (NERs)" and "emerging risks" are used as equivalent in this paper. However, some significant differences can be found in the technical and scientific literature. These differences, according to Brocal [30] and Cantonnet et al. [31], point at a clear problem of consensus on terminology and interpretation around emerging risk. The works of Brocal et al. [17, 25] propose approaches and models to reduce the associated uncertainty.

It would be desirable that the terminology regarding these risks (i.e., "new and emerging risks," "emerging risks") is standardized. In this case, the CWA 16649: 2013 document may prove to be the first step. Currently, International Organization for Standardization (ISO) is developing the ISO 31050 standard—Guidance for managing emerging risks to enhance resilience [32].

*2.2. Emerging Risks Linked to Human Factor.* The effects of Industry 4.0 on OH&S generate advantages and drawbacks that could generate emerging risks [8]. These emerging risks include both industrial risks and occupational risks, since they are heavily related [14]. Arguably, human factor is the main link between industrial emerging risks and occupational emerging risks in advanced manufacturing processes. This may as well be the case, given the complexity of operational and the shifting business environment that continues to overwhelm human cognitive capacities [8, 33–37].

Consequently, one often tries to use mental and intellectual "shortcuts" in finding "easy" explanations or solutions taking into account directly visible parts of the whole context only. The complexity of modern systems generates the opacity where some significant risks become systemic and may be well hidden and lurking beneath until conditions reunite for their full display. One of direct consequences of those changes is the nature of the risks which continue to occur. While undesirable events such as industrial accidents, process and supply chain disruptions, or bankruptcies formerly occurred from known causes and factors, contemporary events usually originate from unanticipated interactions between elements with no visible links [12, 38–40]. In tightly connected complex systems and their environment, it creates conditions for cascading events throughout them [10, 38, 41–49] due to increased interdependencies [50, 51]. Several authors argue that analyzing the performance of complex systems should also imply a careful examination of low level events as well as organizational factors such as safety culture and the incentive system, which shape human performance and affect the risk of errors [37, 39, 52–56].

*2.3. Emerging Risks Linked to Complex Systems.* The technological evolution, including the introduction of the concept of Industry 4.0, and the contemporary operating context significantly contribute to increasing complexity [12]. Some authors categorize it as a "structural complexity" issued from the heterogeneity of system components across different

technological domains. It is a result of increased integration of various technological systems. There is also the "dynamic complexity" which is revealed by an emergent (usually unanticipated) system behavior as a reaction to local perturbations and stimuli in its environment [10]. For example, the integration of modern technologies and new control rules (Industry 4.0) creates additional opacity in systems.

In this regard, concepts such as Human-Machine Interactions (HMI) and Human-Robot Interaction (HRI) can be considered among the most important [4, 7, 15]. HMI and HRI can be considered sources of risk (SR). From this perspective, Brocal et al. [25] have applied the TICHNER in order to identify and characterize emerging risks linked to HMI and HRI. These risks are heavily linked to conditions C1, C4, and C5. In this regard, tasks that involve human control of automated equipment are increasing [57], including the expansion of HRI [58]. With the increase and development of the cobots, these emerging risks are evolving towards scenarios with greater uncertainty, due to (i) the greater flexibility and mobility of the cobots as well as their greater interaction with the workers [8, 57] and (ii) the challenge of increasing the sophistication of the tasks carried out by this type of robots [58].

The increase of this intelligent equipment can lead to connecting the causes of human error with the "smart machine error" [8]. Likewise, the increase in the number of degrees of freedom of the robots also increases their complexity, increasing the risk of entrapment and the difficulty for humans to predict their movements [15]. As the manufacturing tasks become more individualized and more flexible, the machines in smart factory are required to do variable tasks collaboratively without reprogramming [59]. The reliability of such devices is more difficult to predict as the complexity of these systems increases [8].

Brocal et al. [5] have classified the above emerging risks into two groups, in light of their consequences (CO): accidents risks and psychosocial and musculoskeletal risks. The accidents risks in automated environments are closely linked by human errors [60]. Among other causes, it highlights human monitoring and control of processes through control systems, instead of direct control [61], which can lead to a reduction of the practical knowledge of the process [62] and therefore result in over-reliance on automated safety systems [60]. In relation to psychosocial risks and musculoskeletal risks, they can be studied independently or jointly, since they are interrelated, especially in automated contexts. In these contexts, considering the work of Flaspöler et al. [60], the main interrelated causes (C) are low physical activity; static postures; high mental workload, for example, during the monitoring and control of the processes indicated by Chidambaram [61]; reduced privacy at work (mainly because new technologies allow closer and more intrusive supervision); and increase of problems in decision-making.

## 3. Emerging Risk Management

The increase in organizational complexity in manufacturing processes is changing from centralized decision-making towards decentralized instances. In decentralized instances, decision-making can be taken by workers or by equipment where artificial intelligence is integrated [57]. This increase in the complexity of Industry 4.0 affects the OH&S especially in terms of work content, management, and other organizational factors [8].

Based on the work of Brocal et al. [5], risk management systems in the context of Industry 4.0 can be classified into four hierarchically interrelated types: risk governance, risk management, OH&S management system, and emerging risk management. The main models, according to this classification, are discussed in the following section below.

*3.1. Risk Management Models.* From a global level, IRGC [63] proposes a framework for risk governance. This framework provides a guide for the design and application of comprehensive assessment and management strategies to deal with these risks. SRA [64] defines risk governance as the application of governance principles to the identification, assessment, management, and communication of risk. Thus, the process of risk management may be considered integrated into the overall process of risk governance.

In relation to risk management, ISO 31000:2018 standard provides guidelines and a common approach to managing any type of risk. This standard defines risk management as coordinated activities to direct and control an organization with regard to risk. SRA [64] defines risk management as those activities to handle risk such as prevention, mitigation, adaptation, or sharing.

In OH&S field, the ISO 45001:2018 standard defines management system as a set or interrelated or interacting elements of an organization to establish policies and objectives and processes to achieve those objectives; and OH&S management system is a management system or part of a management system used to achieve the OH&S policy.

*3.2. Emerging Risk Management Models.* IRGC [65] addresses emerging risks linked to technology and industrial processes. The methodological aspects for early identification and management of emerging risks are described by the IRGC Guidelines for Emerging Risk Governance [24]. These methodological aspects are flexible and adaptable, especially in complex and uncertain contexts [24].

The IRGC [24] has reviewed emerging risk governance frameworks, and it has selected five:

(i) ENISA: European Union Agency for Network and Information Security

(ii) EFSA: European Food Safety Authority

(iii) Swiss Re SONAR system

(iv) Dutch framework (emerging risks related to the use of chemicals)

(v) CEN workshop agreement (CWA) 16649:2013 (emerging risks related to technology)

CWA 16649:2013 proposes the Emerging Risk Management Framework (ERMF). The whole process is based on the

concept that emerging risks go through a maturation process [24]. This ERMF is based on the risk management frameworks defined by IRGC [63] and ISO 31000: 2009. Currently, International Organization for Standardization (ISO) is developing the ISO 31050 standard—Guidance for managing emerging risks to enhance resilience [32]. This new standard has taken CWA 16649: 2013 as one of its references [66].

## 4. Ways to Improve Organizational Performance to Improve Safety in the Complex Systems

As discussed above, the introduction of the concept of Industry 4.0 brings numerous advantages, but also some new issues. It includes, among other things, emerging risks related to rising complexity of technological systems. One has limited knowledge upon them due to lack of long-term observation data. This situation is fairly challenging for management, organizations, and individual workers as a whole.

This section presents a discussion on how to enhance human and organizational performance aiming at improving risk management in complex systems linked to Industry 4.0.

*4.1. Challenges Related to Organizational and Human Performance.* Komljenovic et al. [12] provide a comprehensive review of research works regarding organizational and human performance in analysis of industrial accidents where complexities, systemic risks as well as organizational and human performance, are seriously involved.

A constant deviation toward danger or failure seems to be one of their key characteristics. The latter is practically impossible to grasp in traditional of chain-of-event analyses.

All this requires overcoming the traditional static approach to risk, through the development of dynamic risk management models oriented towards the organizational and human performance, which is strongly linked to the complex systems characteristic of Industry 4.0.

*4.1.1. Dynamic Management.* Industry 4.0 provides digital management of operations in new technological devices, improves working conditions, and generates a safe manufacturing environment for workers [4, 8]. In this way, this technological development allows the integration of advanced safety systems. Among them, the use of Virtual Reality (VR) stands out to create a risk-free virtual learning and training environment, applications for the use with mobile devices or wearable computing, and the use of RFID technology [7]. Among the main applications of RFID technology is the Workforce Management and the Enhancing Safety, which allows, for example, real-time control of access to dangerous areas, as well as emergencies [67].

In the work environment of Industry 4.0, a wide range of examples of smart materials, smart personal protective equipment, and other advances technological applications is improving the OSH [9]. Such context production workers provide immediate feedback of production conditions via real-time data through their own smart phones and tablets [6]. In this context, the safety management is one of the most important issues [7], where Podgorski et al. [9] have proposed a framework based on new dynamic risk management paradigm. In this way, one of the challenges of Industry 4.0 is the difficulty of managing dynamic risk, as well as the availability and presence of experts in OSH [8].

Thus, during the last few years, new approaches and methodologies have been developed for risk assessment and management considering the dynamic evolution of risk [68]. The main objective of the dynamic risk assessment is to provide an estimate of the risk over time that reflects the current condition of the system, considering for it the integration of the effective aggregation of heterogeneous information [69].

*4.1.2. Organizational Performance.* It seems that the understanding of events is changing given that one of the main sources of risks (SR) nowadays is the organization [12]. Majority of undesirable events (E) has essentially organizational components. One considers "organizational" factors that understand a collective behavior (e.g., centralization and decentralization and organizational clarity). These characteristics come from the evolution of two factors: (i) the category of fences (physical or nonphysical) insuring a safe operating environment and (ii) the new couplings between factors that were formerly nearly independent. The latter is even pushed further with Industry 4.0.

Barriers that allow normal operation progressed with both the complication of work and the more involved persons. Therefore, this new context complicates the detection of flaws in these barriers, leading to undesirable events and failures. Such situations bring degradation of operational and safety margins. They may be locally and individually acceptable, but the sum of effects may have important unanticipated consequences that are not captured by a local analysis. The complexity of the operating environment involves a solution at the organizational level in order to cope with new challenges [12, 34, 39, 45, 70–74].

There are also studies investigating the hypothesis that modern enterprises depend on the deployment of cognitive capacities [70]. The authors argue that there are severe limitations on these capacities in a phenomenon they identify as "functional stupidity." "Functional stupidity" represents a lack of critical thinking, a deliberate condition of ignorance evading defiance to the status quo. This type of pathology is usually widespread in settings ruled by economy in persuading [75, 76]. Such a situation may lead to types of administration curbing or marginalizing suspicion and suppress effective communications. On the other hand, this context brings a structure of interior discussions in such ways favoring positive storylines and marginalizing undesirable ones. Thus, those situations may lead to reducing capacities of the innovation and creativity creating more fragile organizational structures that are vulnerable to both internal and external threats and perturbations, particularly in complex modern enterprises.

*4.1.3. Human Performance.* The behavior of people is basically shaped by their milieu. Marais et al. [77] analyze situations where one does not always anticipate inadequate

behavior. It seems that autonomous decisions across organization may lead to unexpected combination with harmful, often surprising, impacts on both the performance and safety. Additionally, Katina [75, 78] using a synthesis of literature suggests that decision-making and actions are influences by norms and personal attitudes, organizational structures, knowledge base, and social context, degree of connectivity, race and ethnicity, mass media, and national ideology. This suggests a need for consideration of cognition, system, and environment as well as their interplay. Furthermore, a tangible reduction of undesirable events, such as accidents, major process interruptions, or bankruptcies, involves a deep understanding of human and organizational performance issues.

Some research works propose the framework of complexity leadership theory which may help getting a better human and organizational performance [79, 80].

As far as the rationality of decision-making is concerned, several research works indicate that one cannot assume that it is always rational [12, 35, 72]. Cognitive and motivational biases are usually part of the decision-making process. Cognitive bias is defined as "a systematic discrepancy between the "correct" answer in a judgmental task, initiated through a formal normative rule, and the decision maker's actual response to such a task." Motivational biases describe a distortion in decision-making regarding outcomes, results, or choices [72]. Those biases usually have a negative impact on the wanted outcomes and performance of organizations.

*4.1.4. Complex Systems.* The complexity and the opacity of modern systems bring difficulties to the staff to predict its overall behavior as a function of its individual components. The complexity is a system property and results from interactions between its components/subsystems, humans, HMI, HRI, etc. It generates unanticipated and emergent comportment of the system, often intensified by ill adapted operator's actions to those situations.

HRI can be a paradigmatic industrial and occupational example of this complex and challenging context, where Vasic and Billard [15] propose the design of new sensing technology and of fast sensor fusion algorithms (i) to track multiple moving targets in real time, (ii) to achieve robust detection of human motion in order to build good predictive systems, (iii) to ensure robust detection of contact between robots and surrounding living agents in multiple points, and (iv) to develop fast responsive controllers that can replant trajectories in complex, cluttered environment in real time.

Several research works highlight the importance of detecting and cautiously analyzing warning flags, precursors, near-misses, and "low-level" events in order to avoid system level break-downs, process interruptions, and/or major accidents. Therefore, organization should have enough organizational, economic, and technological resilience and flexibility applicable in a large number of different and (un)anticipated situations [12, 36, 37, 39, 55, 81].

As far as human performance is concerned, it is important to understand the error itself. Some research works have shown that both success and failure pathways apply the same intellectual processes, and only the consequence changes. So, the undesired outcome qualifies an action as an error, and it is essential to find its cause. Analyses shall find out why an event occurred ("direct cause" related to preventive and mitigating barriers as well as error precursors) and why it was not stopped ("fundamental cause"). It also has to investigate the organization and its performance (expanded fundamental causes). Considering that it is almost impossible to determine a true causality in complex systems, those analyses become a difficult undertaking in a modern industrial setting. Stock and Seliger [57] propose approaches to address the human factor and social change in Industry 4.0. For this, these authors propose through new technologies increasing the efficiency of the training of the workers, as well as increasing their intrinsic and extrinsic motivation.

*4.2. Potential Approaches for Improving Human and Organizational Performance.* As discussed above, the introduction of the concept of Industry 4.0 brings numerous challenges. To cope with those new challenges, it is necessary to implement both a systematic return of experience (internal and external) and a continuous improvement process and to increase organizational resilience and robustness to unexpected events. However, increasing resilience shall be thought about wisely in order to preserve competiveness, further development, sustainability, and economic viability of an organization.

The organizational resilience is a developing concept. It will not be discussed in detail here, but there are some suggested references upon the topic [10, 34, 45, 70, 71, 82–84]. This section presents suggestions regarding key elements which may reinforce the organizational resilience in the context of Industry 4.0.

*4.2.1. Modeling and Measuring Human and Organizational Performance.* Understanding impacts and risks of humans and organizations as contributors to mishaps, disruptions, and accidents in complex systems requires an adequate model. The model has to go further than the "simple" approach of linearly analyzing preventive and mitigating barriers, which provides quite narrow insights of the events.

Although some models exist [12, 39, 52, 53], these models are not necessarily adaptive to the context of Industry 4.0 risks. Therefore, suggesting the development of methods which will enable modeling and measuring human and organizational performance in the context of Industry 4.0 is a novel approach.

In this regard, it is necessary to adequately take account of the complexity of today's organizations as well as their operating context. This complexity necessitates a new way of reasoning and managing contemporary organizations. The traditional approaches in modeling, analyzing, and management are not entirely adequate to do it, and new methods are necessary [12, 36].

Actually, there are numerous research works suggesting that contemporary organizations should be considered, analyzed, modeled, and managed as Complex Adaptive Systems (CAS) or Complex Adaptive Systems of Systems (CASoS) [12, 36, 73, 85]. This claim is particularly relevant with the introduction of the concept of Industry 4.0. The theory of the complexity and characteristics of CASoS are not discussed

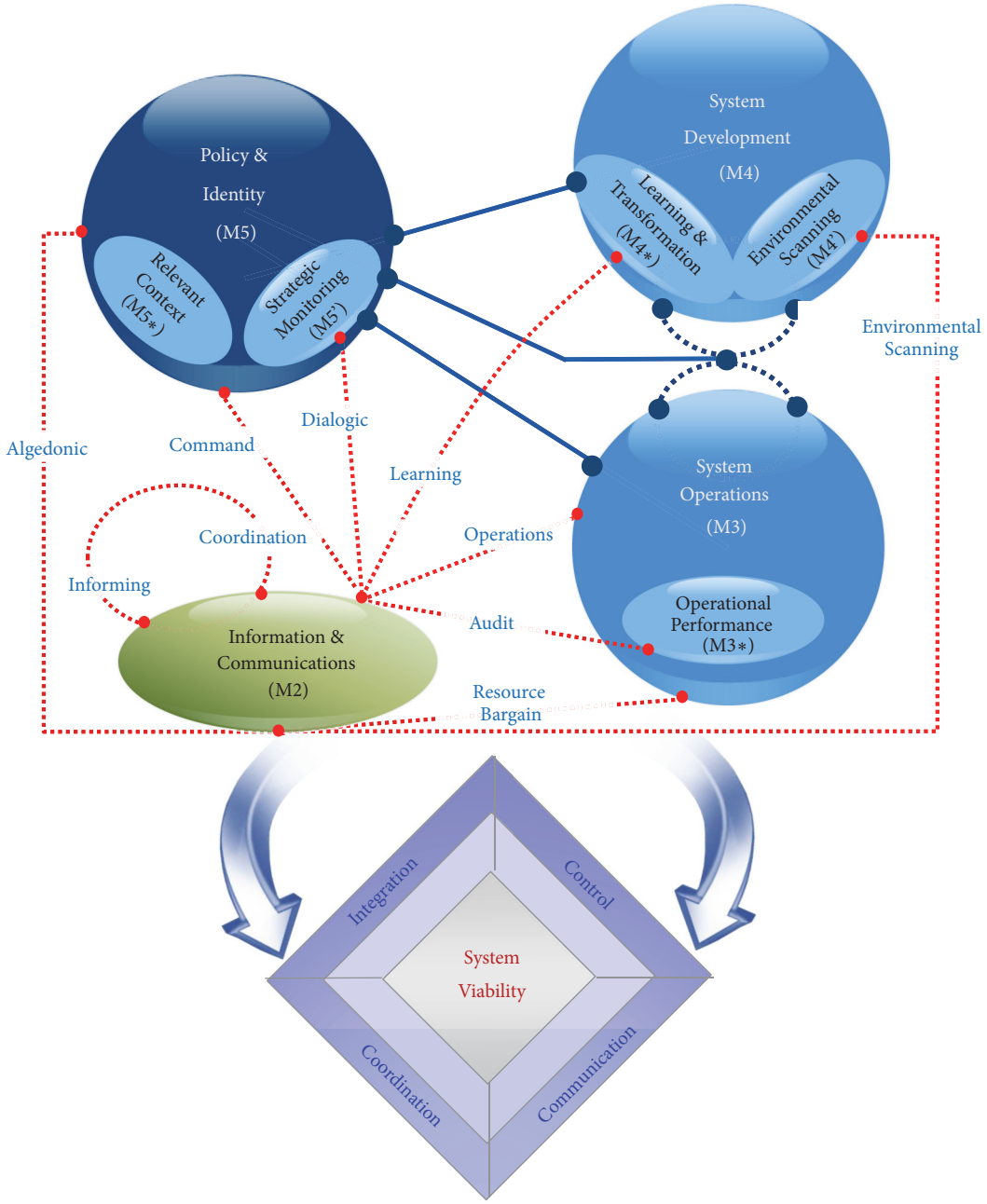



Figure 1: CSG metasystem functions (modified from [19]).

here in detail. It is suggested to consult relevant references on the topic [10, 12, 36, 40, 43, 45, 49, 71, 74, 86, 87].

*4.2.2. Adaptation and Introduction of the Concept of Complex System Governance (CSG) for Industry 4.0.* A possible approach for dealing with complexity in organizational setting is the application of emerging research of Complex System Governance. Complex System Governance is an emerging field, representing an approach to improve system performance through purposeful execution of design for and evolution of important metasystem roles [19, 88–90]. In this emerging field of CSG, the main roles include evolving and improving system performance essential functions enabling control, communication, coordination, and integration [19]. Figure 1 is based on the research of Keating et al. [19] and provides the overall description of the SG approach (Table 6 provides elaboration of the different functions in the CSG model as suggested by Keating and Bradley's [18] research). In CSG, the "metasystem" is deliberately used to imply functions and roles beyond the purview of individual systems in the "system of systems" cluster [91–93]. In Table 6, individual systems can be represented as M1. However, M1 is not listed as the concern for CSG is as the control and communication beyond individual systems (i.e., the metasystem level).

Keating and Bradley [18] provide the systemic representation of CSG indicated in Table 6 in which the purpose

Table 6: CSG's overall description based on Keating and Bradley's [18].

| Areas of concern | Metasystem function | Primary role |
|---|---|---|
| System identity | M5: Policy and identity | Focusing on overall steering and trajectory for the Industry 4.0 systems in the fulfillment of their missions. Maintaining identity and balance between current and future focus. |
| | M5∗: System context | Focusing on the specific context within which the metasystem of an industry is embedded. Context is the set of circumstances, factors, conditions, or patterns that enable or constrain execution of systems in Industry 4.0 setting. |
| | M5': Strategic system monitoring | Focusing on oversight of the Industry 4.0 performance indicators at a strategic level. This includes identification of performances that exceed as well as those the fail to meet the established expectations. |
| System development | M4: System development | Developing and maintaining current and future models of systems in question within the Industry 4.0 schema as well as concentrating on the long-range development for the industry to enable the realization of future feasibilities. |
| | M4∗: Learning and transformation | Focusing on facilitation of learning based on correction of design errors (first order learning) in the metasystem roles of the industry as well as planning for revolution of the industry (second order learning). |
| | M4': Environmental scanning | Focusing on designs, and monitoring systems that can be used to sense operating environment of the industry to detect environmental trends, patterns, or events that can have implications on the current and future state of the industry. |
| System operations | M3: System operations (M3) | Focusing on the day to day execution of industry functions (operations) |
| | M3∗: Operational performance | Monitoring system performance to identify and assess aberrant conditions, exceeded thresholds, and/or irregularities. |
| System information | M2: Information and communications | Creating designs and mechanisms that enable the information flow and consistent interpretation of exchanges information and data necessary to execute industry functions. |

is to provide an organizing construct for the interrelated functions necessary to perform CSG. In CSG approach, the center of any systemic intervention is people with their strengths and weaknesses. In essence, this is recognition that design, execution, and evolution of a systemic intervention are accomplished by people. As such, people become the central driving force behind systemic intervention including challenges related to machine-human interfaces. The effectiveness in intervention must be a function of those who design, conduct, and have participatory roles in the intervention effort [94]. The CSG approach includes a system mapping followed by an investigation to uncover causes of weaknesses (i.e., pathologies) in the governance structure. Pathology acts inhibiting system performance. While CSG is focused on functions enabling viability of the system, it also calls for systemic identification and assessment of system pathologies affecting governance [75]. Interestingly, present research suggests that solutions in the CSG paradigm might involve going beyond technology "fixes" to include "socio" policy, context, and environment.

Pyne et al. [85] further explore concepts, methods, and tools that may help managers to cope with constantly increasing complexity issues. Furthermore, it is suggested that effective problem solving in complex domains (i.e., Industry 4.0) may need different level of "more systemic" approaches capable of matching the uncertain, complex, and dynamic behavior characterize today's industries. Proponents of CSG approach suggest exploring the challenges in moving the CSG from the theoretical/conceptual formulation to practice.

A potential area of practice is Industry 4.0. Again, if one takes Industry 4.0 to mean the current trend of automation and data exchange in manufacturing technologies and encompassing CPS, IoT, cloud computing, and cognitive computing, there remains a case to be made for the utility of CSG in the various areas of Industry 4.0. At the onset, it appears CSG might be used to address issues related to themes of viability, governance, control, communication, coordination, and integration, as well as malfunctions (pathology) that may emerge in the different facets of Industry 4.0.

*4.2.3. Benefiting from Experience of Other Industries at Risk.* Main industries at risk such as nuclear, space, aviation, and petrochemical have developed different ways of coping with human and organizational performance issues. Their return of experience may be beneficial for analyzing ways to improve it within the context of Industry 4.0. Some experience and practices from the nuclear power industry are depicted below.

The Institute of Nuclear Power Operations (INPO) identifies key aspects to accomplish the quality in the integrated risk management [81]. Some attributes are enumerated below:

(i) Behaviors: the expected actions for the stages of risk management are suggested for all organizational levels from individuals to corporate executives.

(ii) Organizational characteristics for effective integrated risk management: a set of principles, policies, practices, oversight, and training are recommended for achieving an all-inclusive risk management process which is elaborated and implemented.

(iii) Integrated risk management warning flags: the warning flags aim at helping the staff and managers to detect undesirable conditions affecting an integrated risk management. The former are categorized by defenses aiming at minimizing risky events. The staff and managers should analyze them for stimulating discussions and draw lessons.

Such an approach may help reinforcing the overall resilience and robustness. This activity also involves a cautious analysis of organizational factors such as incentive systems that influence human performance and impact the risk of errors [37]. This feature is also termed "Antifragility" by some researchers [42, 95]. The antifragility puts forward an idea of adaptive organizations with regard to complex and continuously altering internal and external operating context. In this case, one can suggest that internal and external complexification is also partly due to the introduction of the concept of Industry 4.0 which sometimes may overwhelm human capacity to grasp relevant factors in acting and decision-making. This implies that the paradigm of "only" technological solutions to emerging risks may be misleading especially without consideration of role of the human and organizational factor. For example, one becomes inclined to more risk taking by putting undue confidence into the technology without knowing its limits and vulnerabilities.

Therefore, improving and managing human performance risks (including those coming from machine-human interfaces) based on the experience from the nuclear power industry may also include other elements such as the following [48]:

(i) Frequently discuss risks, complexity, and their interdependence

(ii) Perform gap analysis between outlooks and observations of behavior

(iii) Enthusiastically lobby different opinion to avoid deliberate carelessness (reduce cognitive and motivational biases)

(iv) Analyze and discuss past behaviors, including informal messages at the corporate level

(v) Doubt and uncertainty should not go unchallenged

(vi) Ask to demonstrate that a system is appropriately safe to function or not sufficiently unsafe to be shut down

(vii) Be cautious regarding findings from a root cause investigation that one points out to the negligence; in this case, only part of the story has been exposed

(viii) Enlarge the scope of defense-in-depth concept to embrace the concept of complex systems and their intrinsic nonlinearity

## 5. Conclusions

Undoubtedly, Industry 4.0 brings numerous advantages. However, this paradigm also carries emerging risks and

challenges related to organizational and human performance. These emerging risks include both industrial risks and occupational risks. Arguably, human factor is the main link between industrial emerging risks and occupational emerging risks in the Industry 4.0 context.

Addressing these issues calls for effectiveness in dealing with traditional static approach to risk, for example, through the development of dynamic risk management models oriented towards the organizational and human performance. However, there is also a need to develop robust approaches capable of dealing with emerging risks associated with Industry 4.0. In this research, Complex System Governance is suggested.

However, there remains a need for case applications, clearly articulating the potential of such approaches. Moreover, there is no need to be limited to such approaches as Complex System Governance. This research can be extended to include technological approaches such as Blockchain Technology and the discovery of deep systemic pathological issues affecting Industry 4.0.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was funded by the Spanish Ministry of Economy and Competitiveness, with the title "Analysis and Assessment of technological requirements for the design of a New and Emerging Risks standardized management SYStem (A2NERSYS)" with reference DPI2016-79824-R.

## References

[1] H. Kagermann, W. Wahlster, and J. Helbig, *Recommendations for Implementing the Strategic Iniciative Industrie 4.0*, Acatech e National Academy of Science and Engineering, 2013.

[2] A. J. C. Trappey, C. V. Trappey, U. Hareesh Govindarajan, A. C. Chuang, and J. J. Sun, "A review of essential standards and patent landscapes for the Internet of Things: A key enabler for Industry 4.0," *Advanced Engineering Informatics*, vol. 33, pp. 208–229, 2017.

[3] B. Marr, "Why everyone must get ready for the 4th industrial revolution," 2016, https://www.forbes.com/sites/bernardmarr/2016/04/05/why-everyone-must-get-ready-for-4th-industrial-revolution/.

[4] S. S. Kamble, A. Gunasekaran, and S. A. Gawankar, "Sustainable Industry 4.0 framework: a systematic literature review identifying the current trends and future perspectives," *Process Safety and Environmental Protection*, vol. 117, pp. 408–425, 2018.

[5] F. Brocal, M. A. Sebastián, and C. González, "Advanced manufacturing processes and technologies," in *Management of Emerging Public Health Issues and Risks*, B. Roig, K. Weiss, and V. Thireau, Eds., Chapter 2, pp. 31–64, Academic Press, 2019.

[6] A. Sanders, C. Elangeswaran, and J. Wulfsberg, "Industry 4.0 implies lean manufacturing: research activities in industry 4.0 function as enablers for lean manufacturing," *Journal of Industrial Engineering and Management*, vol. 9, no. 3, pp. 811–833, 2016.

[7] T. D. Oesterreich and F. Teuteberg, "Understanding the implications of digitisation and automation in the context of Industry 4.0: a triangulation approach and elements of a research agenda for the construction industry," *Computers in Industry*, vol. 83, pp. 121–139, 2016.

[8] A. Badri, B. Boudreau-Trudel, and A. S. Souissi, "Occupational health and safety in the industry 4.0 era: A cause for major concern?" *Safety Science*, vol. 109, pp. 403–411, 2018.

[9] D. Podgórski, K. Majchrzycka, A. Dąbrowska, G. Gralewicz, and M. Okrasa, "Towards a conceptual framework of OSH risk management in smart working environments based on smart PPE, ambient intelligence and the Internet of Things technologies," *International Journal of Occupational Safety and Ergonomics*, vol. 23, no. 1, pp. 1–20, 2017.

[10] E. Zio, "Challenges in the vulnerability and risk analysis of critical infrastructures," *Reliability Engineering & System Safety*, vol. 152, pp. 137–150, 2016.

[11] F. Qian, W. Zhong, and W. Du, "Smart process manufacturing—perspective fundamental theories and key technologies for smart and optimal manufacturing in the process industry," *Engineering Journal*, vol. 3, no. 2, pp. 154–160, 2017.

[12] D. Komljenovic, G. Loiselle, and M. Kumral, "Organization: a new focus on mine safety improvement in a complex operational and business environment," *International Journal of Mining Science and Technology*, vol. 27, no. 4, pp. 617–625, 2017.

[13] B. Roig and F. Brocal, "Introduction: needs on emerging risk and management," in *Management of Emerging Public Health Issues and Risks*, B. Roig, K. Weiss, and V. Thireau, Eds., pp. 17–21, Academic Press, 2019.

[14] F. Brocal, C. González, G. Reniers, V. Cozzani, and M. A. Sebastián, "Risk management of hazardous materials in manufacturing processes: links and transitional spaces between occupational accidents and major accidents," *Materials*, vol. 11, no. 1915, pp. 1–23, 2018.

[15] M. Vasic and A. Billard, "Safety issues in human–robot interaction," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 197–204, IEEE, Piscataway, New Jersey, NJ, USA, 2013.

[16] T. Aven, "The risk concept-historical and recent development trends," *Reliability Engineering & System Safety*, vol. 99, pp. 33–44, 2012.

[17] F. Brocal, M. A. Sebastián, and C. González, "Theoretical framework for the new and emerging occupational risk modeling and its monitoring through technology lifecycle of industrial processes," *Safety Science*, vol. 99, pp. 178–186, 2017.

[18] C. B. Keating and J. M. Bradley, "Complex system governance reference model," *International Journal of System of Systems Engineering*, vol. 6, no. 1-2, pp. 33–52, 2015.

[19] C. B. Keating, P. F. Katina, and J. M. Bradley, "Complex system governance: Concept, challenges, and emerging research," *International Journal of System of Systems Engineering*, vol. 5, no. 3, pp. 263–288, 2014.

[20] International Organization for Standardization (ISO), "Risk management –guidelines," ISO 31000:2018, Geneva, 2018.

[21] International Organization for Standardization (ISO), "Occupational health and safety management systems-Requirements with guidance for use," ISO 45001:2018, Geneva, 2018.

[22] International Risk Governance Council (IRGC), *The Emergence of Risks*, Contributing Factors, IRGC, Geneva, Switzerland, 2010.

[23] International Risk Governance Council (IRGC), *Emerging Risks Sources, Drivers and Governance Issues*, IRGC, Geneva, Switzerland, 2010.

[24] International Risk Governance Council (IRGC), *Guidelines for Emerging Risk Governance*, International Risk Governance Council (IRGC), Lausann, Switzerland, 2015.

[25] F. Brocal, C. González, and M. A. Sebastián, "Technique to identify and characterize new and emerging risks: a new tool for application in manufacturing processes," *Safety Science*, vol. 109, pp. 144–156, 2018.

[26] E. Flaspöler, D. Reinert, and E. Brun, *Expert Forecast on Emerging Physical Risks Related to Occupational Safety and Health*, EU-OSHA (European Agency for Safety and Health at Work), Luxembourg., 2005.

[27] E. Brun, R. O. Beeck, S. Van Herpe et al., *Expert Forecast on Emerging Biological Risks Related to Occupational Safety and Health*, EU-OSHA (European Agency for Safety and Health at Work), Luxembourg, 2007.

[28] E. Brun, M. Milczarek, N. Roskams et al., *Expert Forecast on Emerging Psychosocial Risks Related to Occupational Safety and Health*, Office for Official Publications of the European Communities, Luxembourg, 2007.

[29] E. Brun, R. Op de Beeck, S. Van Herpe et al., *Expert Forecast on Emerging Chemical Risks Related to Occupational Safety and Health*, EU-OSHA (European Agency for Safety and Health at Work), Luxembourg, 2009.

[30] F. Brocal, "Uncertainties and challenges when facing new and emerging occupational risks," *Archivos de Prevención de Riesgos Laborales*, vol. 19, no. 1, pp. 6–9, 2016.

[31] M. L. Cantonnet, J. C. Aldasoro, and J. Iradi, "New and emerging risks management in small and medium-sized Spanish enterprises," *Safety Science*, vol. 113, pp. 257–263, 2019.

[32] International Organization for Standardization (ISO), "Guidance for managing emerging risks to enhance resilience," ISO/NP 31050, Geneva, 2019, ISO/NP 31050, https://www.iso.org/standard/54224.html.

[33] S. Arbesman, *Overcomplicated: Technology at the Limits of Comprehension*, Current, New York, NY, USA, 2016.

[34] J. Bourgon, *A New Synthesis of Public Administration, Servicing in the 21st Century*, The School of Public Studies, McGill-Queen's University Press, 2011.

[35] D. Kahneman, *Thinking, Fast and Slow, Farrar, Straus and Giroux*, New York, NY, USA, 2012.

[36] D. Komljenovic, M. Gaha, G. Abdul-Nour, C. Langheit, and M. Bourgeois, "Risks of extreme and rare events in Asset Management," *Safety Science*, vol. 88, pp. 129–145, 2016.

[37] M. E. Paté-Cornell, "On 'black swans' and 'perfect storms', risk analysis and management when statistics are not enough," *Risk Analysis*, vol. 32, no. 11, pp. 1823–1833, 2012.

[38] S. Dekker, P. Cilliers, and J.-H. Hofmeyr, "The complexity of failure: Implications of complexity theory for safety investigations," *Safety Science*, vol. 49, no. 6, pp. 939–945, 2011.

[39] K. Grigoriou, A. Labib, and S. Hadleigh-Dunn, "Learning from rare events: An analysis of ValuJet Flight 592 and Swissair Flight SR 111 crashes," *Engineering Failure Analysis*, vol. 96, pp. 311–319, 2019.

[40] G. Rzevski and P. Skobelev, *Managing Complexity*, WIT Press, Massachusetts, Mass, USA, 2014.

[41] T. Aven and M. Ylönen, "A risk interpretation of sociotechnical safety perspectives," *Reliability Engineering & System Safety*, vol. 175, pp. 13–18, 2018.

[42] T. Aven, "Implications of black swans to the foundations and practice of risk assessment and management," *Reliability Engineering & System Safety*, vol. 134, pp. 83–91, 2015.

[43] L. Bukowski, "System of systems dependability – Theoretical models and applications examples," *Reliability Engineering & System Safety*, vol. 151, pp. 76–92, 2016.

[44] L. A. Cox, "Confronting deep uncertainties in risk analysis," *Risk Analysis*, vol. 32, no. 10, pp. 1607–1629, 2012.

[45] D. Helbing, "Globally networked risks and how to respond," *Nature*, vol. 497, no. 7447, pp. 51–59, 2013.

[46] N. G. Leveson, *Engineering a Safer World, Systems Thinking Applied to Safety*, The MIT Press, Cambridge, Massachusetts, Matt, USA, 2011, http://mitpress.mit.edu/books/engineering-safer-world.

[47] N. G. Leveson, "Applying system thinking to analyze and learn from events," *Safety Science*, vol. 49, pp. 55–64, 2011.

[48] D. Mosey, "Looking beyond operator – putting people in the mix, nei magazine," collaboration with Ken Ellis, Managing Direction of World Association of Nuclear Power Operators (WANO), 2014, http://www.neimagazine.com/features/feature-looking-beyond-the-operator-4447549/, http://www.neimagazine.com/features/featureputting-people-in-the-mix-4321534/.

[49] O. Renn, K. Lucas, A. Haas, and C. Jaeger, "Things are different today: the challenge of global systemic risks," *Journal of Risk Research*, pp. 1–16, 2017.

[50] P. F. Katina, C. Ariel Pinto, J. M. Bradley, and P. T. Hester, "Interdependency-induced risk with applications to healthcare," *International Journal of Critical Infrastructure Protection*, vol. 7, no. 1, pp. 12–26, 2014.

[51] S. M. Rinaldi, J. P. Peerenboom, and T. K. Kelly, "Identifying, understanding, and analyzing critical infrastructure interdependencies," *IEEE Control Systems Magazine*, vol. 21, no. 6, pp. 11–25, 2001.

[52] Department of Energy, *Human Performance Improvement Handbook*, vol. 1, DOE Standard, Washington, Wash, DC, USA, 2009, https://www.standards.doe.gov/standards-documents/1000/1028-BHdbk-2009-v1.

[53] IAEA, *Managing Human Performance to Improve Nuclear Facility Operation*, Vienna, Austria, 2013.

[54] R. Moura, M. Beer, E. Patelli, J. Lewis, and F. Knoll, "Learning from accidents: Interactions between human factors, technology and organisations as a central element to validate risk studies," *Safety Science*, vol. 99, pp. 196–214, 2017.

[55] US NRC, *Safety Culture Communicator, Case Study 4: April 2010 Upper Big Branch Mine Explosion – 29 Lives Lost*, Washington, Wash, D.C., USA, 2012, http://pbadupws.nrc.gov/docs/ML1206/ML12069A003.pdf.

[56] B. Wahlström, "Organisational learning - reflections from the nuclear industry," *Safety Science*, vol. 49, no. 1, pp. 65–74, 2011.

[57] T. Stock and G. Seliger, "Opportunities of sustainable manufacturing in industry 4.0," *Procedia CIRP*, vol. 40, pp. 536–541, 2016.

[58] T. B. Sheridan, "Human–robot interaction: status and challenges," *Human Factors*, vol. 58, no. 4, pp. 525–532, 2016.

[59] W. Wang, X. Zhu, L. Wang, Q. Qiu, and Q. Cao, "Ubiquitous robotic technology for smart manufacturing system," *Computational Intelligence and Neuroscience*, vol. 2016, Article ID 6018686, 14 pages, 2016.

[60] E. Flaspöler, A. Hauke, P. Pappachan et al., *The Human Machine Interface as an Emerging Risk*, Office for Official Publications of the European Communities, Luxembourg, 2010.

[61] P. Chidambaram, "Perspectives on human factors in a shifting operational environment," *Journal of Loss Prevention in the Process Industries*, vol. 44, pp. 112–118, 2016.

[62] N. Stacey, P. Ellwood, S. Bradbrook, J. Reynolds, and H. Williams, "Key trends and drivers of change in information and communication technologies and work location," Foresight on new and emerging risks in OSH. Working report, European Agency for Safety and Health at Work (EU-OSHA), 2017.

[63] International Risk Governance Council (IRGC), *White Paper on Risk Governance. Towards an Integrative Approach*, IRGC, Geneva, 2005.

[64] Society for Risk Analysis (SRA), SRA glossary, SRA, 2015.

[65] International Risk Governance Council (IRGC), *Improving the Management of Emerging Risks*, IRGC, Geneva, Switzerland, 2011.

[66] International Organization for Standardization (ISO), "ISO 31050 guidance for managing emerging risks to enhance resilience: thriving in a world growing in uncertainty," ISO 31050-Leaflet_v09aj14082018, 2018, https://www.eu-vri.eu/file-handler.ashx?file=16526.

[67] M. Roberti, "How is RFID being used in the construction industry? ask experts forum 6," 2013.

[68] N. Paltrinieri and G. Scarponi, "Addressing dynamic risk in the petroleum industry by means of innovative analysis solutions," *Chemical Engineering Transactions*, vol. 36, pp. 451–456, 2014.

[69] X. Yang, S. Haugen, and N. Paltrinieri, "Clarifying the concept of operational risk assessment in the oil and gas industry," *Safety Science*, vol. 108, pp. 259–268, 2018.

[70] M. Alvesson and A. Spicer, "Stupidity based theory of organizations," *Journal of Management Studies*, vol. 49, no. 7, pp. 1194–1220, 2012.

[71] T. Homer-Dixon, "Complexity science, shifting the trajectory of civilization," *Oxford Leadership Journal*, vol. 2, no. 1, pp. 2–15, 2011.

[72] G. Montibeller and D. Winterfeldt, "Cognitive and motivational biases in decision and risk analysis," *Risk Analysis*, vol. 35, no. 7, pp. 1230–1251, 2015.

[73] R. D. Stacey and C. Mowles, "Strategic management and organisational dynamic," in *The Challenge of Complexity to Ways of Thinking about Organizations*, Pearson, London, UK, 7th edition, 2016.

[74] M.-C. Therrien, A. Valiquette-L'Heureux, and J.-M. Normandin, "Tightly coupled governance for loosely coupled wicked problems: the train explosion in Lac-Mégantic case," *International Journal of Risk Assessment and Management*, vol. 19, no. 4, pp. 260–277, 2016.

[75] P. F. Katina, "Metasystem pathologies (M-Path) method: phases and procedures," *Journal of Management Development*, vol. 35, no. 10, pp. 1287–1301, 2016.

[76] C. I. Barnard, "Functions and pathology of status systems in formal organizations," in *Industry and Society*, W. F. Whyte, Ed., pp. 46–83, McGraw-Hill, New York, NY, USA, 1946.

[77] K. Marais, J. H. Saleh, and N. G. Leveson, "Archetypes for organizational safety," *Safety Science*, vol. 44, no. 7, pp. 565–582, 2006.

[78] P. F. Katina, "Individual and societal risk (RiskIS): beyond probability and consequence during hurricane katrina," in *Disaster Forensics: Understanding Root Cause and Complex Causality*, A. J. Masys, Ed., pp. 1–23, Springer International Publishing, Geneva, Switzerland, 2016.

[79] M. Schneider and M. Somers, "Organizations as complex adaptive systems: implications of Complexity Theory for leadership research," *The Leadership Quarterly*, vol. 17, no. 4, pp. 351–365, 2006.

[80] M. Uhl-Bien, R. Marion, and B. McKelvey, "Complexity leadership theory: shifting leadership from the industrial age to the knowledge era," *The Leadership Quarterly*, vol. 18, no. 4, pp. 298–318, 2007.

[81] Institute of Nuclear Power Operations (INPO), "Excellence in integrated risk management; the elements, attributes, and behaviors that exemplify excellence in integrated risk management INPO 12-008," 2013, http://nuclearsafety.info/wp-content/uploads/2017/03/INPO-12-008-Excellence-in-Integrated-Risk-Management.pdf.

[82] F. Bouaziz and Z. S. Hachicha, "Strategic human resource management practices and organizational resilience," *Journal of Management Development*, vol. 37, no. 7, pp. 537–551, 2018.

[83] A. De Galizia, C. Simon, P. Weber, B. Iung, C. Duval, and E. Serdet, "Markers and petterns of organizational resilience for risk analysis," *IFAC-PapersOnLine*, vol. 49, no. 19, pp. 432–437, 2016.

[84] D. Mendonça and W. A. Wallace, "Factors underlying organizational resilience: The case of electric power restoration in New York City after 11 September 2001," *Reliability Engineering & System Safety*, vol. 141, pp. 83–91, 2015.

[85] J. C. Pyne, C. B. Keating, and P. F. Katina, "Enhancing utility manager's capability for dealing with complex issues," in *Proceedings of the Water Environment Federation Technical Exhibition and Conference (WEFTEC '16)*, vol. 2016, New Orleans.

[86] B. Johnson and A. Hernandez, "Exploring engineered complex adaptive systems of systems," *Procedia Computer Science*, vol. 95, pp. 58–65, 2016.

[87] A. Stirling, "Keep it complex," *Nature*, vol. 468, no. 7327, pp. 1029–1031, 2010.

[88] D. Baugh, "Environmental scanning implications in the governance of complex systems," *International Journal of System of Systems Engineering*, vol. 6, no. 1-2, pp. 127–143, 2015.

[89] C. B. Keating and V. Ireland, "Editorial: complex systems governance - issues and applications," *International Journal of System of Systems Engineering*, vol. 7, no. 1-3, pp. 1–21, 2016.

[90] C. B. Keating and P. F. Katina, "Complex system governance development: a first generation methodology," *International Journal of System of Systems Engineering*, vol. 7, no. 1-3, pp. 43–74, 2016.

[91] B. Carter, "A metasystem perspective and implications for governance," *International Journal of System of Systems Engineering*, vol. 6, no. 1-2, pp. 90–100, 2015.

[92] K. D. Palmer, "Meta-systems engineering," in *Proceedings of the 10th Annual International Symposium of the International Council on Systems Engineering*, 2000, http://www.archonic.net/MSE04.PDF.

[93] G. R. Djavanshir, R. Khorramshahgol, and J. Novitzki, "Critical characteristics of metasystems: Toward defining metasystems' governance mechanism," *IT Professional*, vol. 11, no. 3, pp. 46–49, 2009.

[94] J. C. Pyne, C. B. Keating, P. F. Katina, and J. M. Bradley, "Systemic intervention methods supporting complex system governance initiatives," *International Journal of System of Systems Engineering*, vol. 8, no. 3, pp. 285–309, 2018.

[95] N. N. Taleb, *Antifragile, Things that Gain from Disorder*, Random House, New York, NY, USA, 2012.

Hindawi
Complexity
Volume 2019, Article ID 1280285, 14 pages
https://doi.org/10.1155/2019/1280285

*Research Article*

# An Integrated Metaheuristic Routing Method for Multiple-Block Warehouses with Ultranarrow Aisles and Access Restriction

**Fangyu Chen**,[1] **Gangyan Xu**,[2] **and Yongchang Wei**[1]

[1] *Institute of Operations Management and System Engineering, School of Business Administration, Zhongnan University of Economics and Law, Wuhan 430073, China*
[2] *School of Architecture, Harbin Institute of Technology, Shenzhen, China*

Correspondence should be addressed to Yongchang Wei; ivanwilts306@163.com

Received 4 April 2019; Accepted 26 May 2019; Published 18 June 2019

Guest Editor: Jorge Luis García-Alcaraz



A problem-specific routing algorithm integrating ant colony optimization (*ACO*) and integer-coded genetic algorithm (*GA*) is developed to address the newly observed limitations imposed by ultranarrow aisles and access restriction, which exist in the largest e-commerce enterprise with self-run logistics in China. Those limitations prohibit pickers from walking through the whole aisle, and the access restriction even allows them to access the pick aisles only from specific entrances. The ant colony optimization is mainly responsible for generating the initial chromosomes for the genetic algorithm, which then searches the near-optimal solutions of picker-routing with our novel chromosome design by recording the detailed information of access modes and subaisles. To demonstrate the merits of the proposed algorithm, a comprehensive simulation for comparison is conducted with 12 warehouse layouts with real order information. The simulation results show that the proposed hybrid algorithm is superior to dedicated heuristics in terms of solution quality. The impacts of the parameters with respect to warehouse layout on the picking efficiency are analyzed as well. Setting more connect aisles and cross aisles is suggested to effectively optimize the picking-service efficiency in the presence of access limitations.

## 1. Introduction

For most e-commerce enterprises or online distributors, a well-designed order-picking strategy (from warehouse location [1] to routing method [2]) can promote the efficiency of the entire supply chain [3]. Given the advent of e-commerce, the significance of order-picking has grown [4, 5]. Therefore, we concentrate on a picker-routing problem based on the observation in a Chinese e-commerce enterprise.

The order service time is one of the main performance measurements of warehouse systems. In a picker-to-parts warehouse, Petersen [6] pointed out that service time can be represented as a positive correlation function of one picker's travel distance. Accordingly, an appropriate routing method that identifies the shortest picking route can minimize the order service time and therefore improve the picking efficiency. One alternative option that is capable of reducing warehouse management costs is to increase the space utilization, which is typically made by warehouse owners to minimize the pick aisles' width, culminating in the widespread adoption of the narrow-parallel aisles in picker-to-parts warehouses. In a narrow-aisle warehouse with multiple order pickers, congestions may happen [7].

Recently, we observed an interesting warehouse system with ultranarrow aisles in a realistic Chinese online retailer with self-run logistics, *D* Company, meaning that the pick aisles' widths are narrowed further, relative to the aforementioned narrow aisles. The ultranarrow aisles impede order pickers to enter while pushing their picking carts. They should deposit their carts at the entrance. When all the picks are finished, they should walk out the aisles to get their carts. The connect aisles and cross aisles, which are located in the warehouse, allow the pickers to change working aisle, because the width of them is sufficient for walking with carts, as shown

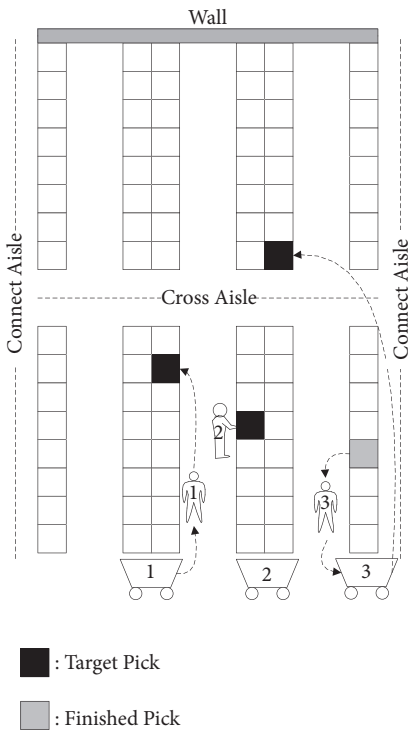



Figure 1: Order picking in the ultranarrow aisles.

in Figure 1. For such a warehouse, there is little utility for existing routing methods that permit the order pickers to walk through the whole pick aisles. Additionally, constraints in physical space lead to access restrictions. For example, the racks that are located in the front of a warehouse system may be placed against the front wall; order pickers can only access these aisles through the back entrances. To promote the e-commerce boom, this paper aims to propose efficient routing methods for the warehouses taking both ultranarrow aisles and access restriction into consideration.

We propose a method by integrating ant colony optimization (*ACO*) and genetic algorithm (*GA*), which is named as *AGNA* (hybrid *ACO* and *GA* for ultranarrow aisles and access restriction). The novelties of *AGNA* are as follows. Firstly, to avoid traversing aisles, all the aisles with picks are restricted to be visited by specified access modes, and the access mode will map the picks at the certain entrances. Secondly, to deal with the unique characteristics induced by access restriction, the mechanism of determining the feasible travel routes between designated entrances is embedded in *AGNA*. Thirdly, to improve the convergence speed and performance of the routing method, *ACO* is responsible for generating the initial chromosome for *GA*, and the integer-coded *GA* is designed to search the near-optimal solutions mainly through our novel chromosome design that represents all the access modes for all the aisles with picks. To validate the effectiveness of the improved routing method, a comprehensive simulation study was performed with the data collected from the aforementioned *D* company. Meanwhile, the impact of the warehouse layout is elucidated. We found that the *AGNA* method can be readily deployed in different warehouse-layout scenarios due to its flexibility.

The main contributions of this research lie in the following aspects. Firstly, we developed a hybrid *AGNA*-based routing method to address the methodological challenges for the warehouses with ultranarrow aisles and access restriction. To cope with the unique characteristics of routing problem in this kind of warehouses, many innovative mechanisms with respect to feasibility judgment and chromosome coding are significantly incorporated in our hybrid *AGNA*-based algorithm. Secondly, with the real data of customer order information and warehouse configuration collected from the *D* company, a simulation study is employed to show the effectiveness and efficacies of the proposed method. The negative impacts brought by new features are analyzed as well.

The remainder of this paper is arranged as follows. Section 2 presents a review of the related literature and introduces some existing dedicated heuristics for single-block warehouses. Some of these heuristics can be applied in the warehouse concerned in this paper after improvement, and the modified heuristics are compared with our designed *AGNA* method. In Section 3, the optimization model tailored to the concerned routing problem in the warehouse is defined. The *AGNA* is presented in Section 4. A comprehensive simulation is implemented to validate the efficiency of the proposed method in Section 5. Finally, the paper is concluded in Section 6 and our ongoing research topics are introduced as well.

## 2. Literature Review

The literature closely related to our research can be grouped into two main streams: one is the impact of aisles width, which constitutes a key feature in our research, while the other is the picker-routing problem. In the following, we will review these two streams of literature separately.

*2.1. The Impact of Aisle Width.* Reducing the pick aisles' width is an alternative to enlarge the space utilization [7]. Nevertheless, narrow aisles lead to more frequent congestion. In this regard, by changing the ratio of the pick : walk-time, analytical models to assess the order picker congestion (blocking) in both narrow-aisle and wide-aisle warehouses are developed by Parikh and Meller [8, 9]. Hong et al. [10] proposed a batching method in warehouses with narrow aisle. Similarly, Chen et al. [11, 12] introduced some routing methods for narrow-aisle warehouses with the consideration of congestion.

Most of the previous literature primarily considers the traditional wide-aisle or narrow-aisle warehouse systems. Although Chen et al. [13] proposed a set of heuristic routing methods for the ultranarrow aisle warehouse system with consideration of access restriction, these heuristics are developed on the existing methods for basic warehouse layout. These methods have the advantage in the ease of implementation. Nevertheless, they can deviate from the optimal solutions seriously, which further increase operational costs. To this end, to get a better optimization effect on picking efficiency, we concentrate on designing a judicious routing method based on metaheuristics.

*2.2. Routing Method.* We can divide all the existing routing methods in the related literature into two groups: dedicated heuristics and metaheuristics. The dedicated heuristics have found prevalent application for order picking, given the ease of their deployment [14–16]. In the existing dedicated heuristics, only the *Largest Gap* and *Return* methods forbid the order pickers from traversing the subaisles for most situations, as shown in Figure 2. However, these two methods cannot be readily deployed in the warehouse system concerned in this paper. For example, the *Largest Gap* method requests the order pickers to walk through the first aisle with picks to reach the back of the block.

For more complicated environments, metaheuristics are more suitable [17–19]. Except for generating picking routes in complicated environments such as multiple-order pickers [12], much of the relevant literature has focused on designing integrated solutions for both order-batching and picker-routing [20–22]. Although many metaheuristics have been presented, such as Particle Swarm Optimization [23] and Simulated Annealing [24], *ACO* is the most popular for the picker-routing problem [25–27]. To design the picking route for the warehouse of interest in our research, we design a routing method by integrating *ACO* and *GA*. This integrated method can dynamically control the invocation time of both *ACO* and *GA* and improve the convergence speed and performance of the algorithm.

## 3. Problem Definition

*3.1. Warehouse Configuration and Assumption.* The related literature [4, 28] and practice focus commonly on a multiple-block picker-to-parts warehouse, which is concerned in our paper as well, as shown in Figure 3. The main difference of our research is that the widths of all the parallel pick aisles are assumed to be ultranarrow. Additionally, some aisles may have access restriction on designated entrances. The *CA*s (cross aisles) are located in the front-back direction. Likewise, there are *CNA* (connect aisles) in the left-right direction. The widths of *CA* and *CNA* are sufficient to permit order pickers walking with their picking carts, which allows order pickers to change working aisles. The *cross point* is the intersection between a *CA* and a *CNA*. The warehouse is divided into multiple blocks (*B*) by these *CA*s and *CNA*s in front-back and left-right directions. Each single pick aisle is also divided into a number of subaisles (*S*).

A *Row* contains the blocks arranged in the left-right direction, and correspondingly a *Column* refers to those in the front-back direction. For an *S*, $E_f$ and $E_b$ indicate the front entrance and back entrance, respectively. For a single block, there are two adjacent *CA*s, one is $CA_f$ (*front cross aisle*), and the other is $CA_b$ (*back cross aisle*). For an order picker, the working cross aisle is denoted as $CA_c$ (*current cross aisle*). To describe a route from one subaisle to another, $S_b$ (*Begin Subaisle*) denotes the subaisle where the order picker begins, and $S_t$ (*Target Subaisle*) denotes the subaisle which the order picker is going to visit. Moreover, the front entrances of $S_b$ and $S_t$ are $E_{bf}$ and $E_{tf}$, respectively, and their back entrances are $E_{bb}$ and $E_{tb}$, respectively. Given this layout, some assumptions are noteworthy. Firstly, in such a warehouse, the two-sided picking policy is chosen by the order pickers [29]. Meanwhile, we ignore the time cost on reaching the different layers and sides of shelves.

In the warehouse of e-commerce enterprise, the dynamic arrival of customers' orders means that we can composite all the orders into batches depending only on their arrival time. We assume that the warehouse adopts random-storage policy [30]. We impose restriction on the capacity of picking carts to avoid overload [31], and the picker congestion is ignored. For simplicity, we ignore the difference in the order pickers' walking speed and the time of finishing a picking task is assumed to be constant [32, 33]. The order pickers have no constraint and can change walking direction in any aisles [34]. The picking operation begins and finishes at the depot, where the operating time is ignored in this paper.

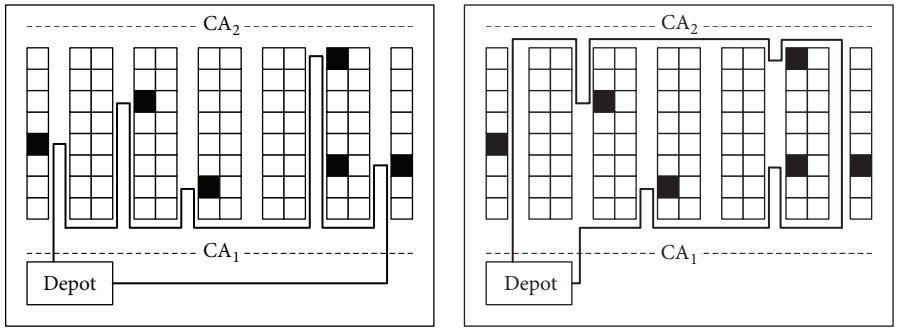
*3.2. Routing Model.* According to the *Largest Gap* and *Return* methods, to avoid traversing and deal with the access restriction, the most effective picking solution for each subaisle with picks can be illustrated by three access modes: *Front-return*, *Back-return*, and *Gap-return* (as shown in Figure 4). The *Front-return* mode inquires the order pickers to walk into and depart from a subaisle through the front entrance after visiting all the picks. Correspondingly, in the *Back-return* mode, the order pickers only use the back entrance for both entry and departure. In the *Gap-return* mode, the order pickers enter a pick aisle but make a U-turn at the largest gap. The order pickers may have to visit the subaisle twice: once by $E_f$ and another by $E_b$. Because the largest gap may cut the subaisle into two parts, this implies that the *Gap-return* mode corresponds to two access modes: accessing the front part and the back part. The *Gap-return* mode does not necessitate the order pickers to visit the remaining part immediately after they finish one part of a subaisle. Once the access mode is determined for a subaisle, the picking route in this subaisle is determined as well, and the picks therein will be mapped to the entrance which should be visited by the order pickers. The *Front-return* mode will map all the picks to $E_f$, the *Back-return* mode will map all the picks to $E_b$, and the *Gap-return* mode will map the picks to $E_f$ and $E_b$, respectively. Then, routing methods only need to find a satisfactory moving route passing all the entrances with mapped picks. Finally, an integrated route is generated by combining this moving route and the picking routes in subaisles with picks. Consequently, the routing problem can be equivalently expressed by determining the best access modes and finding the best visiting sequence for all the subaisles with picks.

Based on the assumptions and the access modes aforementioned, the objective is to minimize the total travel distance of serving a batch. Then, we can formulate a 0-1 integer linear optimization model.

*Indices and Sets*

$\mathbf{\Omega}$: Set of subaisles with picks and the depot

$i, j \in \mathbf{\Omega}$: Indexes of subaisles with picks and depot, in which $i = 0$ means the depot




(a) Return from single block (b) Largest Gap from single block

Figure 2: A sketch of *Return* and *Largest Gap*.

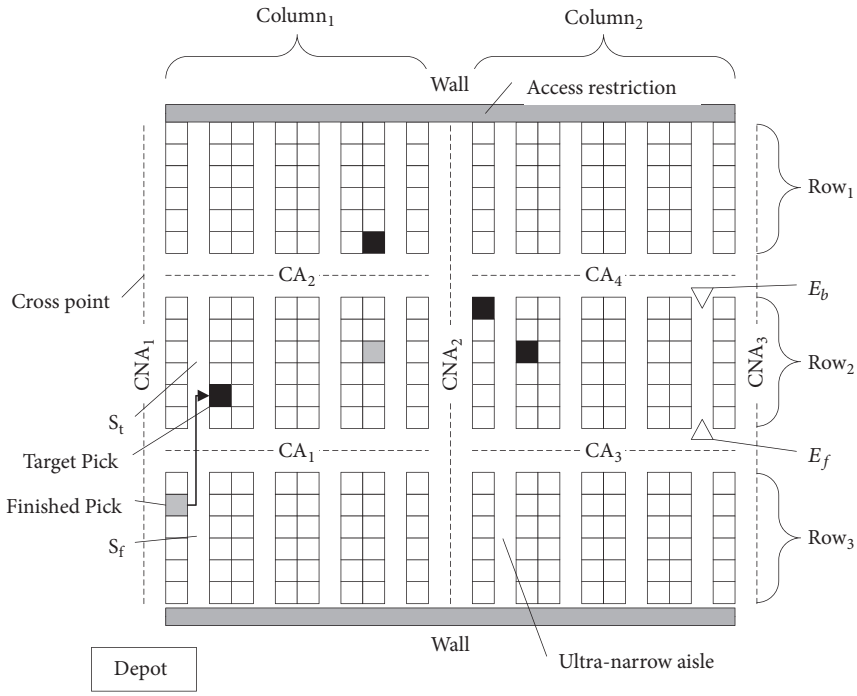



Figure 3: Multiple-block warehouse system with ultranarrow aisles and access restriction.

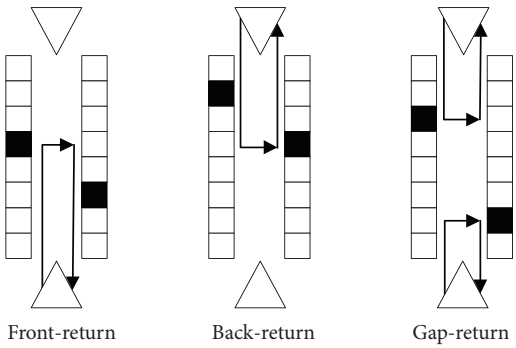



Figure 4: Access modes for subaisles with picks.

$\mathbf{R}_{i,j}$: Set of the feasible moving routes from subaisles $i$ to $j$

$r \in \mathbf{R}_{i,j}$: Index of the feasible moving routes from subaisles $i$ to $j$

$\mathbf{M}_j$: Set of the feasible access modes of subaisle $j$

$m \in \mathbf{M}_j$: Index of the feasible access modes of subaisle $j$

*Parameters and Sets*

$\mathbf{D}_{ij}$: Set of the moving distances of the feasible moving routes from subaisles $i$ to $j$

$d_{ijr} \in \mathbf{D}_{ij}$: Moving distance of the $r^{th}$ feasible moving route from subaisles $i$ to $j$

$\mathbf{D}'_j$: Set of the picking distances of the feasible access modes of subaisle $j$

$d'_{jm} \in \mathbf{D}'_j$: Picking distance of the $m^{th}$ feasible access mode of subaisle $j$

*Decision Variables*

$x_{ij}$: Binary decision that equals 1 ($x_{ij} = 1$) means that subaisle $i$ is visited directly after subaisle $j$

$y_{ijr}$: Binary decision that equals 1 ($y_{ijr} = 1$) means that the $r^{th}$ feasible moving route is selected to move from subaisles $i$ to $j$

$z_{jm}$: Binary decision that equals 1 ($z_{jm} = 1$) means that the $m^{th}$ feasible access mode is selected to visit subaisle $i$ to $j$

*Model Formulation*

$$\min \left\{ \sum_{i \in \mathbf{\Omega}} \sum_{j \in \mathbf{\Omega}} x_{ij} \left( \sum_{k \in \mathbf{T}_{ij}} y_{ijr} d_{ijr} + \sum_{m \in \mathbf{M}_j} z_{jm} d'_{jm} \right) \right\} \quad (1)$$

subject to

$$\sum_{r \in \mathbf{R}_{ij}} y_{ijr} = 1, \quad \forall i, j \in \mathbf{\Omega} \quad (2)$$

$$\sum_{m \in \mathbf{M}_j} z_{jm} = 1, \quad \forall j \in \mathbf{\Omega} \quad (3)$$

$$x_{ij} \in \{0, 1\}, \quad \forall i, j \in \mathbf{\Omega} \quad (4)$$

$$y_{ijr} \in \{0, 1\}, \quad \forall i, j \in \mathbf{\Omega}, \ \forall r \in \mathbf{R}_{ij} \quad (5)$$

$$z_{jm} \in \{0, 1\}, \quad \forall j \neq 0 \in \mathbf{\Omega}, \ \forall m \in \mathbf{M}_j \quad (6)$$

Function (1) is the objective function that minimizes the total travel distance of serving a batch. Constraint (2) ensures that only one moving route is selected when the order pickers move from subaisle $i$ to $j$. Constraint (3) ensures that each subaisle selects only one access mode. Constraints (4) to (6) define the variable domains.

The routing model differs from the basic picker-routing problem, which can be defined as the Steiner-TSP (Steiner Travelling Salesman Problem) [14]. All the cities (subaisles) in the TSP need to be visited only once. However, if a subaisle (city) chooses the *Gap-return* mode, the order pickers may visit this subaisle twice: once by $E_f$ and another time by $E_b$. For our model, this removes the necessity for the common constraint in TSP that each subaisle (city) has exact only one predecessor and successor. Besides, when the *Gap-return* mode is determined for a subaisle and the order pickers are required to visit $E_f$ immediately after $E_b$, the common constraint in TSP that there is only a single tour is likewise rendered unnecessary for our model.

Furthermore, the impact of access restriction cannot be incorporated into the model. When the selected blocks have access restriction, some access modes will become infeasible for the subaisles. For example, if $CA_f$ of a block has access restriction, only the *Back-return* mode may be chosen for the subaisles in this block. Moreover, the choice of the access mode will affect the moving route between subaisles. For example, if the *Front-return* mode is selected for both $S_b$ and $S_t$, the order pickers will have to enter and leave these subaisles by $E_{bf}$ and $E_{tf}$, signifying that the moving route between $S_b$ and $S_t$ must pass $E_{bf}$ and $E_{tf}$.

Generally, there is an interaction between the values of $Z_{jm}$ and $y_{ijr}$. Before the routing-construction begins, for each subaisle, we can just get a set of the "selectable" access modes and the corresponding detailed picking routes for these access modes are also contingent upon the access restriction constraint. The set of the "selectable" moving routes between any two subaisles can be confirmed in this stage as well.

When the order picker chooses the access mode for $S_t$, by extracting from the set of the "selectable" moving routes, we can determine the set of feasible moving routes between $S_b$ and $S_t$. The access mode of the current $S_t$ will also affect the choice of the moving route from the current $S_t$ to the next $S_t$, once the access mode of the next $S_t$ is determined. Due to the aforementioned characteristics, the accurate set of $\mathbf{D}_{ij}$ can not be determined prior to the selection of optimal routing solutions. Thus, our problem cannot be solved simply by optimization solvers, such as Lingo and CPLEX, which motivates us to design metaheuristic-based routing methods.

## 4. Solution Methods

In this section, a hybrid algorithm integrating *ACO* (ant colony optimization) and integer-coded *GA* (genetic algorithm) is developed to deal with the specified routing problem. On one hand, the *ACO* generates the initial chromosome for the integer-coded *GA* during the initialization. This structure can not only dynamically control the invocation time of both the *GA* and *ACO* but also improve the convergence speed and performance of the algorithm. On the other hand, the integer-coded *GA* searches a near-optimal solution for picker-routing. The total travel distance of serving a batch could be defined as the fitness values of the integer-coded chromosomes obtained from the *GA*.

*4.1. Distance Initialization.* For different access modes of one subaisle, the corresponding picking distances and picking routes differ from each other. Similarly, the moving route and moving distance between $S_b$ and $S_t$ are also affected by their access modes. The access mode for a subaisle can be determined only when this subaisle is chosen as $S_t$. After that, its picking route and moving route can be correspondingly confirmed. To represent the routing problem as a combination optimization problem manageable by the *AGNA*, all the picking routes in subaisles and the moving routes between any two subaisles for different access modes should be recorded during the initialization.

For a subaisle, each access mode will determine the corresponding picking route and picking distance. This applies especially to the *Gap-return* mode, in which the front and the back parts lead to different picking routes and picking distances. This means that, for each subaisle without access restriction, four kinds of picking routes are usually available: for the *Front-return* mode, for the *Back-return* mode, for the

front part of the *Gap-return* mode, and for the back part of the *Gap-return* mode. In the initialization, all the corresponding picking routes and picking distances will be calculated and recorded as the input for the routing method.

Moreover, there are four cases of the moving route and moving distance between $S_b$ and $S_t$, which are affected by the access modes of $S_b$ and $S_t$.

*Case 1.* The *Front-return* mode is chosen for both $S_b$ and $S_t$, or the *Gap-return* mode is chosen for both of them, in which only the front part of $S_b$ is completed and the front part of $S_t$ needs to be visited. Then, the moving route will start from $E_{bf}$ and end at $E_{tf}$.

*Case 2.* The *Front-return* mode is chosen for $S_b$, or the *Gap-return* mode is chosen, in which only the front part is completed. Conversely, the *Back-return* mode is chosen for $S_t$, or the *Gap-return* mode is chosen, in which only the back part needs to be visited. Then, the moving route will start from $E_{bf}$ and end at $E_{tb}$.

*Case 3.* The *Back-return* mode is chosen for $S_b$, or the *Gap-return* mode is chosen, in which only the back part is completed. Conversely, the *Front-return* mode is chosen for $S_t$, or the *Gap-return* mode is chosen, in which only the front part needs to be visited. Then, the moving route will start from $E_{bb}$ and end at $E_{tf}$.

*Case 4.* The *Back-return* mode is chosen for both $S_b$ and $S_t$, or the *Gap-return* mode is chosen for both of them, in which only the back part of $S_b$ is completed and the back part of $S_t$ needs to be visited. Then, the moving route will start from $E_{bb}$ and end at $E_{tb}$.

It deserves attention that the moving route and its moving distance between $S_b$ and $S_t$ still need to be computed for each case. The moving distance is usually calculated by the path between any two designated entrances with the Manhattan distance (Theys et al., 2010). However, the Manhattan distance cannot fully satisfy the layout with the ultranarrow aisles. According to the diversified spatial relationship between the designated entrances of $S_b$ and $S_t$, as shown in Figure 5, the detailed definitions of the moving route and moving distance are described as follows [13].

*(i) Entrances in the Same CA*. The order pickers can walk directly from the designated entrance of $S_b$ to the designated entrance of $S_t$; see Route ⓐ shown in Figure 5.

*(ii) Entrances in Different CAs but in the Same Column*. Usually, two optional moving routes between the designated entrance of $S_b$ and the designated entrance of $S_t$ are available to the order pickers. For simplicity, the shorter one is chosen; refer to Route ⓑ shown in Figure 5.

*(iii) Entrances in Different Columns*. The order pickers walk along $CA_c$ by leaving $S_b$ from the chosen entrance, until getting to the *CNA* adjacent to the *CA* which the chosen entrance of $S_t$ belongs to. Then, after setting this *CA* as $CA_c$, the order pickers walk along $CA_c$ and enter $S_t$; see Route ⓒ shown in Figure 5.

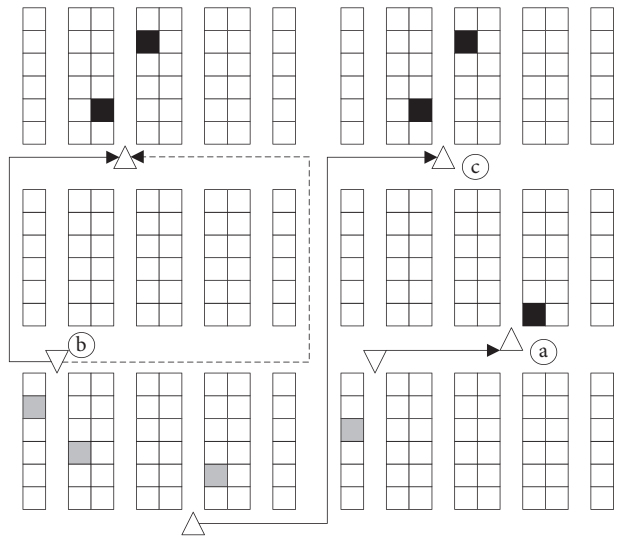



Figure 5: The definition rules of the moving routes.

All the moving routes and picking routes with respect to different access modes and their distances are recorded and will be used for route construction and fitness-value evaluation. If some access modes are infeasible due to access restriction, the related picking distance and moving distance will be recorded as infinity.

*4.2. ACO-Based Chromosome Initialization.* Based on the initialized distances, the *ACO* is employed to initialize the population of the routing chromosomes. We assume that the ant colony contains $N(N \geq 2)$ ants, which denote the order pickers. The ant finds $S_t$ from the unfinished subaisles and determines the access mode of $S_t$. Next, it records the moving distance from $S_b$ to $S_t$ and the picking distance in $S_t$. When all the subaisles are finished, the ant will walk back to the depot. Then, $L$, the total travel distance, is composed by all the selected picking distances and moving distances, which will be seemed as the evaluation criterion. Before the *ACO* starts, the initial value of the pheromone is calculated by

$$\tau_{ijr}(0) = \frac{1}{\sum_{i\in\Omega}\sum_{j\in\Omega}\left|\mathbf{D}_{ij}\right|} \tag{7}$$

where $\tau_{ijr}(0)$ is the pheromone concentration in the $r^{th}$ feasible moving route from subaisle $i$ to $j$ at time 0 and $\sum_{i\in\Omega}\sum_{j\in\Omega}|\mathbf{D}_{ij}|$ is the sum of the number of feasible moving routes among all the subaisles to be visited by the ants or order pickers.

Each ant chooses $S_t$ and determines the moving route randomly depending on the transition probability $p_{ijr}(t)$ that

subaisle $j$ will be visited immediately after finishing subaisle $i$ at time $t$ by the $r^{th}$ moving route, which is represented as

$$p_{ijr}(t) = \begin{cases} \dfrac{\left[\tau_{ijr}(t)\right]^{\alpha}\left[\eta_{ijr}\right]^{\beta}}{\sum_{j\in\mathbf{U}}\left[\tau_{ijr}(t)\right]^{\alpha}\left[\eta_{ijr}\right]^{\beta}}, & \text{if } j \in \mathbf{U}, \\ 0, & \text{otherwise}, \end{cases} \tag{8}$$

where $\tau_{ijr}(t)$ is the pheromone concentration on the $r^{th}$ moving route from subaisle $i$ to $j$ at time $t$, $\eta_{ijr}$ measures the visibility on this route, and $\mathbf{U}$ is the set of unfinished subaisles. The value of $\eta_{ijr}$ is usually defined as the reciprocal of the sum of the moving distance of the $r^{th}$ moving route from subaisle $i$ to $j$ and the corresponding picking distances in subaisles $i$ and $j$. The definition of the transition probability leads to high superiority to the shortest travel and picking route, and $\tau_{ijr}(t)$ impels the familiar moving route and picking route is chosen by these ants. It is well recognized that the parameters $\alpha$ and $\beta$ coordinate the relative importance between the pheromone concentration and the route visibility ($\alpha \geq 1, \beta \geq 1$). Usually, $\alpha$ and $\beta$ are valued according to the former observation.

When all the ants complete their picking tasks, the pheromone concentration on all the routes is updated by

$$\tau_{ijr}(t+1) = (1-\rho)\,\tau_{ijr}(t) + \Delta\tau_{ijr}(t, t+1) \tag{9}$$

where $\rho$ is the recession parameter of the pheromone concentration ($0 < \rho < 1$). $\Delta\tau_{ijr}(t, t+1)$ is the alteration of the pheromone concentration laid on moving route $r$ from subaisle $i$ to subaisle $j$ between times $t$ and $t+1$, as determined by

$$\Delta\tau_{ijr}(t, t+1) = \sum_{n=1}^{N} \Delta\tau_{ijr}^{n}(t, t+1) \tag{10}$$

where $\Delta\tau_{ijr}^{n}(t, t+1)$ is the pheromone concentration left on the route $r$ from subaisles $i$ to $j$ by the $n^{th}$ ant from times $t$ to $t+1$, which is defined by

$$\Delta\tau_{ijr}^{n}(t, t+1) = \begin{cases} \dfrac{1}{L_n}, & \text{if the } n^{th} \text{ ant chooses the } r^{th} \text{ route to move from sub-aisle } i \text{ to sub-aisle } j \text{ between times } t \text{ and } t+1, \\ 0, & \text{otherwise}, \end{cases} \tag{11}$$

where $L_n$ means the total travel distance of the picking route completed by the $n^{th}$ ant.

When all the ants complete the trips, some solutions will be saved in a queue ($\mathbf{Q}$). If a new solution outside the queue but better than the worst solution in the queue is found, it will replace that worst solution in the queue. The *ACO* will stop after the iterations equal the predetermined number. When it terminates, the solutions saved in the queue will be used in the initialization of chromosomes in the integer-coded *GA*.

### 4.3. Integer-Coded Genetic Algorithm (GA)

*4.3.1. Chromosomal Representation and Initial-Population Generation.* During the route construction process by *ACO*, the order pickers choose $S_t$ first and then decide the access mode. The unselected access modes of $S_t$ may be ignored easily in the later iterations. To avoid converging to local optimum, the integer-coded *GA* generates the near-optimal solutions by searching the access modes and the visiting sequence of all the subaisles with picks. The value of each gene in a chromosome not only records the visiting sequence of a certain subaisle but also determines the access mode of this subaisle. In our algorithm, if a gene is associated with the $i^{th}$ subaisle and the $m^{th}$ mode, the value of this gene is defined as $2^{i-1} * 3^{m-1}$. As is corresponding to all the access modes, we set $m = 1$ as the *Front-return* mode, $m = 2$ as the *Back-return* mode, $m = 3$ as accessing the front part by the *Gap-return* mode, and $m = 4$ as accessing the front part by the *Gap-return* mode. Consequently, this value formula helps us identify the unique subaisle and access mode for each gene.

It is noteworthy that although the *ACO* has already generated a visiting sequence passing all the subaisles with feasible access modes, this solution may be insufficiently effective and ignores other unselected access modes. The chromosomal representation should include all the subaisles and all the access modes for them, and the length of chromosome will be $\sum_{i\in\Omega} |\mathbf{M}_i|$. Therefore, the solution of *ACO* can be used as the encoding basis for the first few genes, and the remaining genes of the chromosome will be supplemented by random encoding. Figure 6 illustrates an example of the integer-coded chromosomal representation including four subaisles without access restriction, which means that the length of the chromosome is 16. The solutions generated by *ACO* are sequentially represented as $\{i = 1, m = 4\}$, $\{i = 4, m = 1\}$, $\{i = 2, m = 3\}$, $\{i = 3, m = 2\}$, $\{i = 2, m = 4\}$, and $\{i = 1, m = 3\}$. Such sequential representation implies visits to all the subaisles with picks by the following sequence: (1) accessing the back part of the $1^{st}$ subaisle by the *Gap-return* mode; (2) accessing the $4^{th}$ subaisle by the *Front-return* mode; (3) accessing the front part of the $2^{nd}$ subaisle by the *Gap-return* mode; (4) accessing the $3^{rd}$ subaisle by the *Back-return* mode; (5) accessing the back part of the $2^{nd}$ subaisle by the *Gap-return* mode; and (6) accessing the front part of the $1^{st}$ subaisle by the *Gap-return* mode. Hence, in the chromosomal initialization, the first six genes will be valued as 27, 8, 18, 12, 54, and 9, whereas the remaining 10 genes will

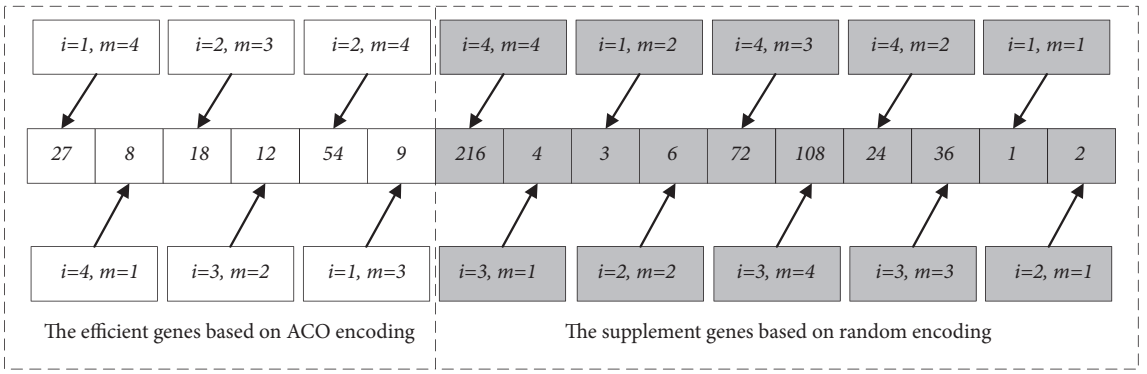



Figure 6: The illustration of integer-coding.

be randomly coded to cover the unselected access modes for further optimization.

The population of chromosome is $P$. When the encoding is finished for all the chromosomes, the crossover and mutation operation will begin.

*4.3.2. Fitness Value Evaluation.* The total travel distances of all the integer-coded chromosomes are chosen as their fitness values. Because each chromosome contains all the access modes for each subaisle with picks, not all the genes need to be calculated in the evaluation of the fitness values. For example, if the $i^{th}$ subaisle has already been indicated by a former gene to choose the *Front-return* mode, the remaining genes that indicate the $i^{th}$ subaisle to be accessed by other modes will be excluded. During the chromosomal decoding, the value of each gene should be validated to avoid any repeated or missed visits to the subaisles. If the subaisle is unfinished and the access mode is feasible, the gene will pass the validation and be calculated in the fitness values. The procedure of validation for each gene can be described as in Figure 7.

The calculation of the total travel distance of each chromosome is described as follows.

*Step 1.* Set $L = 0$ and $S_b = S_0$ to signify that the order pickers start from the depot.

*Step 2.* Get the next gene.

*Step 3.* Decode the gene to get the subaisle and the access mode.

*Step 4.* If the subaisle $i$ and access mode $m$ pass the validation, set $S_t = S_i$; then proceed to Step 5; otherwise, go back to Step 2.

*Step 5.* Based on the access modes of $S_b$ and $S_t$, determine the moving route $r$ from $S_b$ to $S_t$. Get the moving distance $d_{btr}$ and the picking distance $d'_{tr}$; then set $L = L + d_{btr} + d'_{tr}$ and $S_b = S_t$.

*Step 6.* If all the subaisles are finished, go on with Step 7; otherwise, revert to Step 2.

*Step 7.* Return to the depot and determine the moving route $r$; set $S_t = S_0$ and $L = L + d_{btr}$.

*Step 8.* The reciprocal value of $L$ is the fitness value of this solution.

*Step 9.* If $L$ is associated with a better fitness value than that of the near-optimal solution so far in the last population generation, set $L_{near-optimal} = L$.

*4.3.3. Crossover and Mutation.* The designating parents for the undergoing crossover and mutation are adopted by the approach of proportional selection. In this approach, the parents are selected randomly. However, the parents with higher fitness values tend to be selected with higher probability. Because the genes in the front of a chromosome are more likely to be calculated in the fitness value, the effective genes should be included in the crossover and mutation. To improve the evolutionary speed, the Four-Point Crossover method is adopted to create new offspring. Four segmentations with equal lengths are determined; the second and fourth gene segmentations are selected in crossover operation. Then, we use the Order Crossover in the Legality Repair. Figure 8(a) illustrates the crossover operation.

The mutation operation guarantees the diversity and avoids premature convergence. *SBM* (Swap-based mutation) – which selects two random elements (one from the first third part of the chromosome and the other from the remaining part) and exchanges their positions – is applied to generate the offspring. Figure 8(b) illustrates these mutation operators.

*4.4. Stopping Criteria.* The *GA* algorithm terminates in two situations. The first one is as follows: once the iterations reaches a predetermined number of generations $G$, the *GA* will terminate. The second criteria is that the *GA* will terminate if the best solution remains the same for a fixed number of generations $ep$. The same criteria can be seen in the *ACO* part, and the corresponding parameters are $I$ and $C$, respectively. Figure 9 describes the overall procedure of *AGNA*.

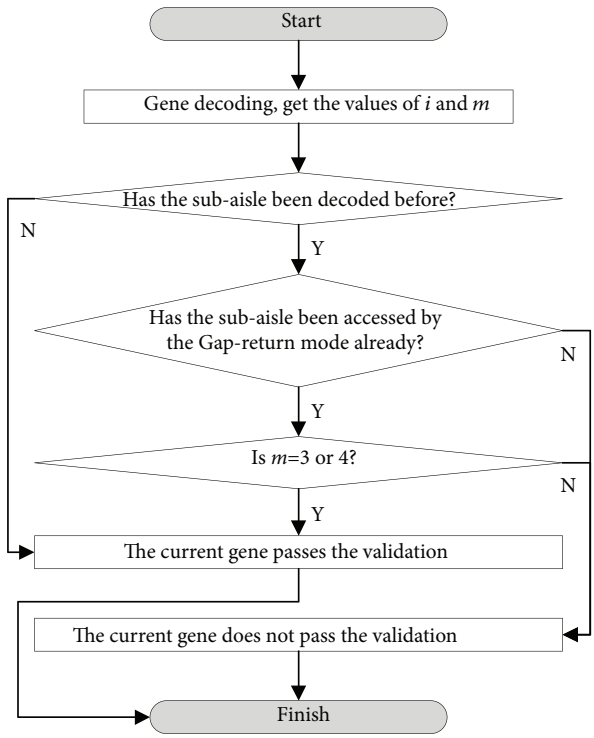



Figure 7: The procedure of validation for each single gene.

## 5. Experimental Study

*5.1. Experimental Design.* We conduct a comprehensive experiment to evaluate the effectiveness of the proposed method, which contains 12 kinds of warehouse layout scenarios extracted from a realistic warehouse system in *D* company, as shown in Figure 10. The detailed configuration parameters are shown in Table 1, which is referred to Chen et al. [13].

There are usually at least 11 aisles in every single block. For each single subaisle, the minimum number of storage locations is 384. The average length and width of each single subaisle are 12.048 m and 0.8 m, respectively. For each single shelf, the average width is 0.7 m. Furthermore, the average widths of each single *CNA* and *CA* are 3.64 m and 3.3 m, respectively.

*5.2. Experimental Results.* The simulation is input with the realistic order information over one month in this aforementioned warehouse. For all the scenarios, at least 547,754 orders arrived in this month. The order information only contains the designated storage locations, without other customer-related information. According to the practical situation, we restrict each picking cart to carry 60 items at most. Besides, there are usually only 1 to 2 items in an online customer order. To maximize the utilization of picking cart, orders are combined into batches. In this online warehouse, we use the Variable *Time-Window Batching* rule, also known as *First-Come-First-Served*, to composite all the orders into batches according to their arrival time. After the batching operation, 13,060 batches are generated. There are 1,165 picking tasks at least assigned to the scenario with single block. To compare our hybrid *AGNA* method with the dedicated heuristics, we also improved the basic *Return* and *Largest Gap* methods to adapt to the warehouse concerned in this paper, which, respectively, are named as the *RNA* (Return for ultranarrow aisles and access restriction) and *LNA* (Largest Gap for ultranarrow aisles and access restriction). For the *RNA* and *LNA*, each subaisle with picks also needs to choose access mode. The *RNA* has the *Front-return* and *Back-return* modes, and *LNA* has the *Front-return*, *Back-return*, and *Gap-return* modes. The procedures of *RNA* and *LNA* are described as follows [13].

*Step 1.* Sort and index all the *CA*s progressively in the *Column*s and then in the *Row*s.

*Step 2.* Choose feasible access modes with the shortest picking distance for all the subaisles with picks. Then, the picks will be mapped to the designated entrances.

*Step 3.* If the picking tasks are mapped to an entrance, the *CA* with this entrance can be labelled as a *CA* with picks.

*Step 4.* By traversing the indexes of the *CA*s incrementally, set the first found *CA* with picks as $CA_c$; then the order pickers will walk to the left *cross point* of $CA_c$.

*Step 5.* Enter $CA_c$, and visit all the designated entrances with mapped picking tasks.

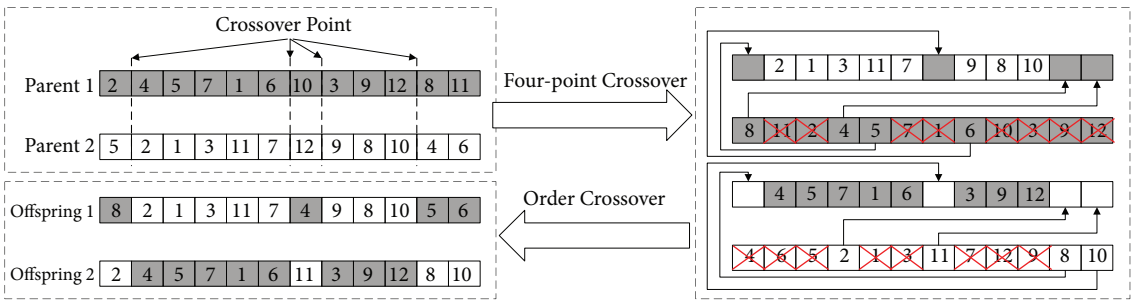
*Step 6.* If all the *CA*s with picks are visited, go on with Step 9; otherwise, go on with Step 7.

*Step 7.* If the next *CA* with picks is in the same *Column* as $CA_c$, leave $CA_c$ from the nearer *cross point*; if the next *CA* with picks is in different *Column*s, leave $CA_c$ from the right *cross point* of $CA_c$.

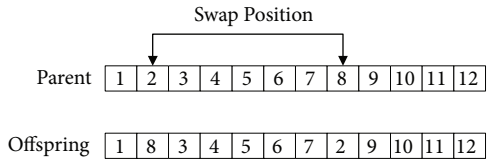
*Step 8.* Reach the next *CA* with picks, and set this *CA* as $CA_c$. Then, revert to Step 5.

*Step 9.* Leave $CA_c$ and return to the depot.

The simulation platform is developed by C#.net on a 4.00GHz PC. In all the *AGNA* solutions, the values of the following search parameters, as shown in Table 2, are validated to be robust in the previous research [11] and *ANOVA* test. For *RNA* and *LNA*, the average computation time for a batch was less than 1 s. For *AGNA*, this time was less than 5 s. The results show that, in eight scenarios, *AGNA* gets the best performance. In the remaining four scenarios, the three routing methods get the same results, as shown in Table 3. Data in italic font means the best result in the corresponding scenario. Owing to the flexible definition of the visiting sequence for a subaisle, *AGNA* can optimize the travel route effectively. Additionally, the access modes are derived from existing basic heuristics, with which the order pickers are familiar. Accordingly, the order pickers can easily accept the newly designed routing method. Therefore, *AGNA* can be readily applied to the setting of multiple-block warehouse.




(a) Cross operation




(b) Mutation operation

Figure 8: The illustration of crossover and mutation.

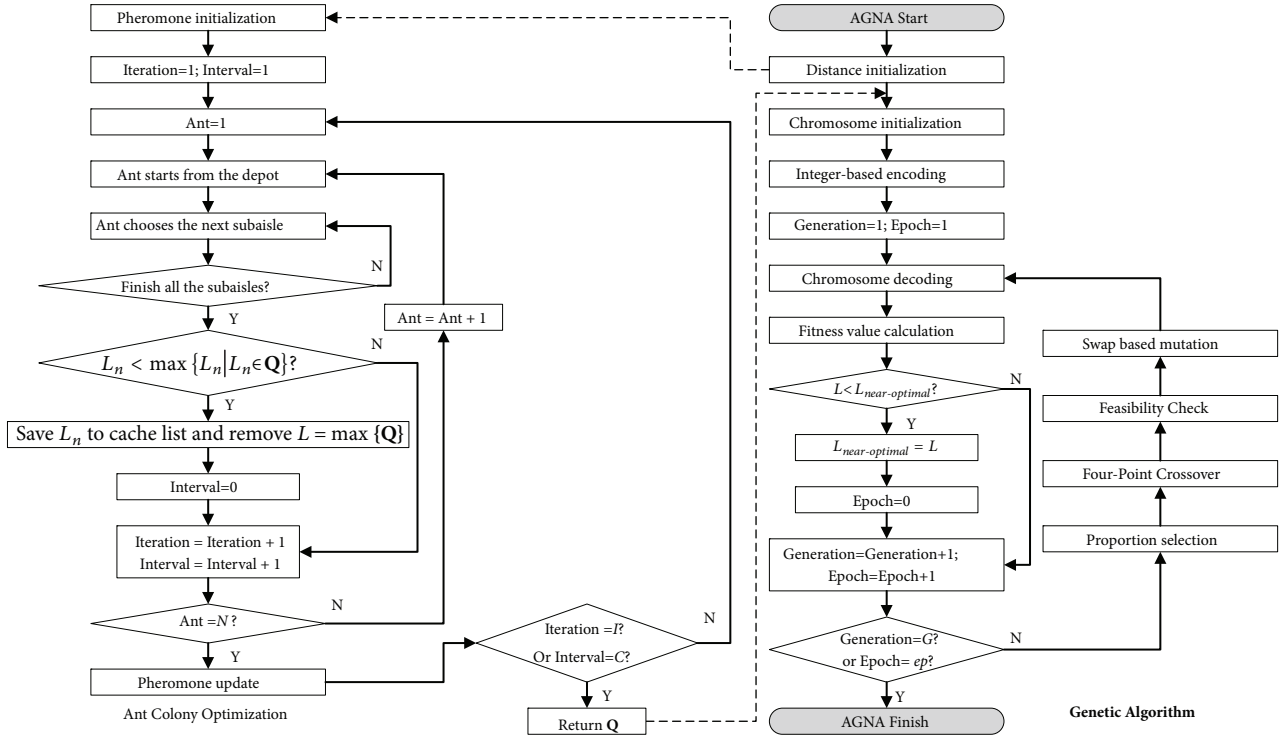



Figure 9: The flowchart of *AGNA*.

*5.3. Discussions.* Based on the results, we can analyze the effects of the following parameters.

*(1) Access Restriction*. In four scenarios, all the methods produce the same results. All of them are the layouts having only one *Row*, and access restriction is imposed on one side for all the blocks implying that all the subaisles are allowed to be visited from only one entrance. This means that only one identical access mode with the same travel distance will be chosen by all the routing methods. For the same reason, *AGNA* outperforms *RNA* and *LNA* in the other eight scenarios with fewer access restrictions (see Table 4). We calculate the percentage differences by $A/R = 100 \times (1 - r(A)/r(R))$, $A/L = 100 \times (1 - r(A)/r(L))$, and $L/R = 100 \times (1 - r(L)/r(R))$, where $r(R)$, $r(L)$, and $r(A)$ refer to the average travel distances of *RNA*, *LNA*, and *AGNA*, respectively. The imposition of fewer access restrictions will increase the optimization space among these methods.

*(2) Number of CNAs*. Although the *CNA* is only responsible for the order pickers changing working aisles, *AGNA* produces better performance in a scenario with more *CNA*s. For

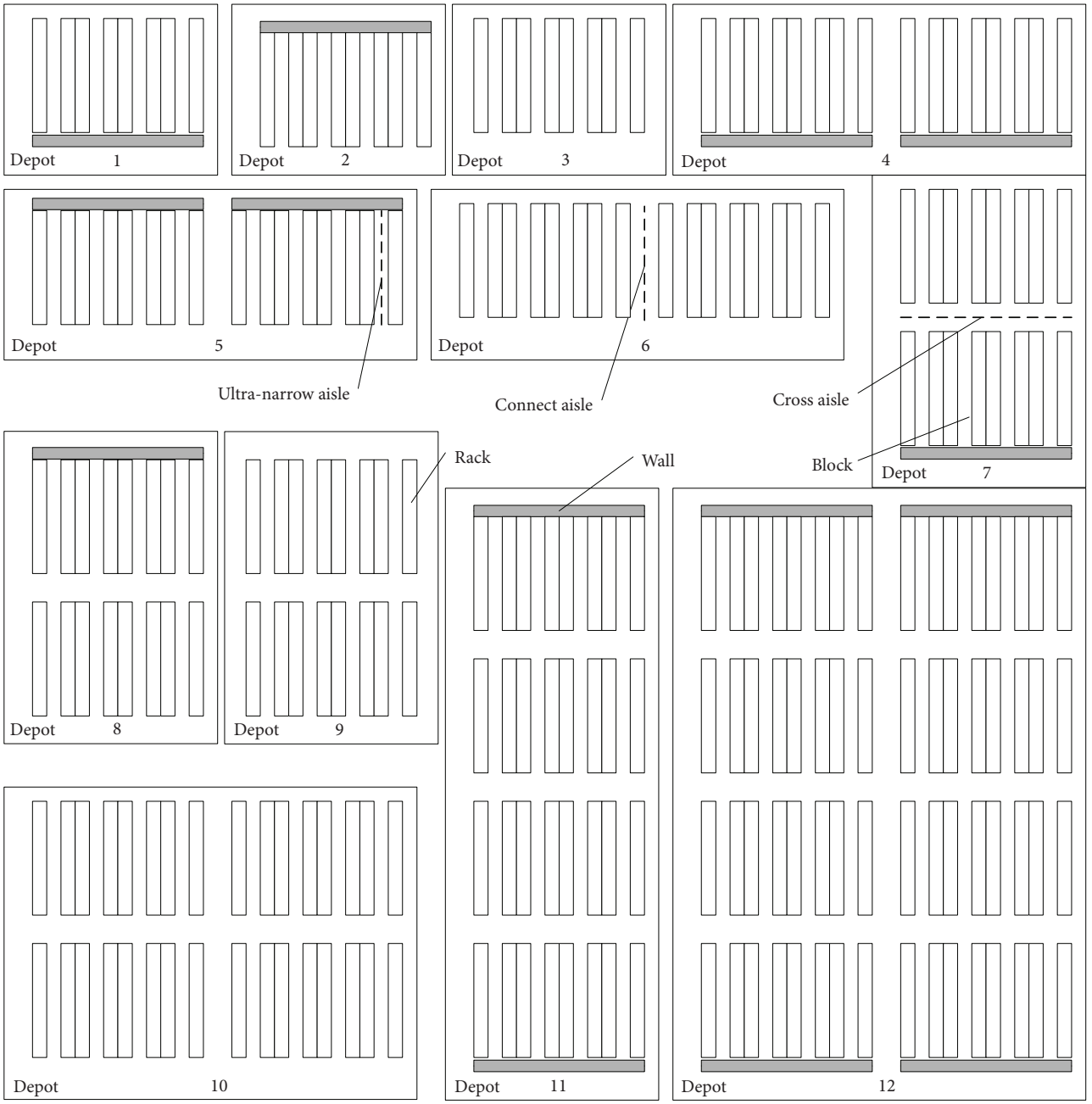



Figure 10: Warehouse scenarios concerned in the experiment.

instance, the mild increase in the number of *CNA*s from two (in the $3^{rd}$ scenario) to three (in the $6^{th}$) translates into a significant synchronous increase in the percentage difference between *AGNA* and *RNA*. Similar comparisons between *AGNA* and *RNA* are shown in Figure 11. The implication of this observation is that the negative impacts of access restriction can be by mitigated by setting the optimal number of *CNA*s.

*(3) Number of CAs*. The presence of more *CA*s contributes to a reduced relative proportion of access restriction, by which the order pickers can enjoy greater flexibility. As shown in Table 4, *AGNA* outperforms other methods in the scenarios with more *CA*s. If the access restriction is inevitable, we suggest that warehouse designers can set more *CA*s.

## 6. Conclusions

This paper studied a new kind of multiple-block warehouse with ultranarrow aisles as observed from one large-scale online retailer in China. In such a warehouse system, an order picker is prohibited from entering the ultranarrow aisles carrying a picking cart. In addition, we also take access restriction into account, which has been imposed

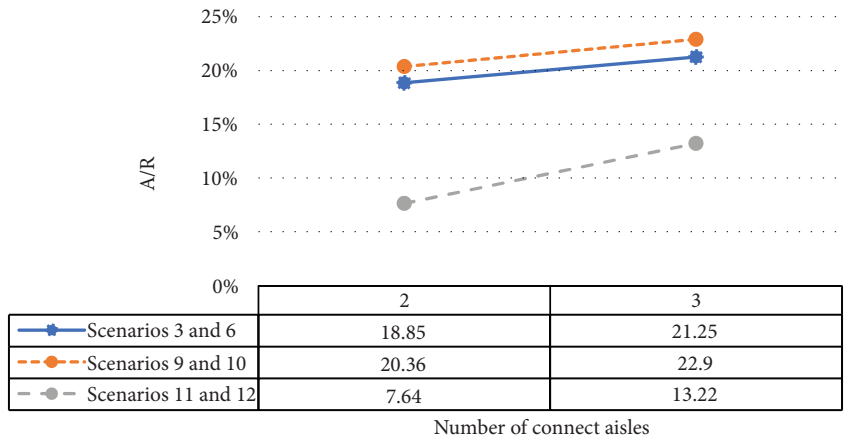



Figure 11: The percentage difference between *AGNA* and *RNA* in scenarios with two or three *CNA*s.

Table 1: Detailed configurations of warehouse scenarios.

| Layout scenario | Number of blocks | Number of cross aisles | Number of connect aisles | Front restriction | Back restriction |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 2 | Yes | No |
| 2 | 1 | 1 | 2 | No | Yes |
| 3 | 1 | 2 | 2 | No | No |
| 4 | 2 | 1 | 3 | Yes | No |
| 5 | 2 | 1 | 3 | No | Yes |
| 6 | 2 | 2 | 3 | No | No |
| 7 | 2 | 2 | 2 | Yes | No |
| 8 | 2 | 2 | 2 | No | Yes |
| 9 | 2 | 3 | 2 | No | No |
| 10 | 4 | 3 | 3 | No | No |
| 11 | 4 | 3 | 2 | Yes | Yes |
| 12 | 8 | 3 | 3 | Yes | Yes |

Table 2: The parameter setting for the *AGNA* algorithm.

| Parameter | *ACO* | *GA* |
|---|---|---|
| Population | $\|\mathbf{\Omega}\|$ | $\|\mathbf{\Omega}\|$ |
| Iterations ($I$) | $2000 * \|\mathbf{\Omega}\|$ | |
| Interval ($C$) | $50 * \|\mathbf{\Omega}\|$ | |
| $\alpha$ | 1 | |
| $\beta$ | 5 | |
| $\rho$ | 0.15 | |
| Crossover rate | | 0.75 |
| Mutation rate | | 0.1 |
| $G$ | | $100 * \|\mathbf{\Omega}\|^2$ |
| $ep$ | | $100 * \|\mathbf{\Omega}\|$ |

on some special pick aisles to limit the order pickers to visit by a designated entrance. In such a warehouse, the existing routing methods cannot be immediately applied without modification. This study designed a routing method by integrating *ACO* and *GA* to address those practical features with the aim to improve operational efficiency. Specifically, the core procedure of the *AGNA* lies in the identification of not only the best access mode but also the near-optimal visiting sequence and moving route that visits all the subaisles with picks.

A comprehensive simulation is deliberately designed to verify the efficacies of the proposed method. The input data for the simulation, including the warehouse configurations and order information, are collected from *D* company. Moreover, two improved dedicated heuristics were also examined for comparison. By simulation, we found that *AGNA* can significantly improve the picking efficiency in most scenarios. Additionally, the mechanism of *AGNA* which deals with the access restriction can be deployed in the warehouses with traditional aisles as well. Moreover, it appears that minimizing access restrictions by setting more *CA*s and *CNA*s can effectively optimize the picking-service efficiency.

In this paper, the proposed method is developed for already-designed warehouse layouts. If the ultranarrow aisles and access restriction are inevitable, designing the optimal layout is a problem that is worthy of further research. Moreover, the designation of storage assignment and order-batching methods in the setting of such a warehouse system also deserves ongoing further research.

## Data Availability

The data will be accessible upon request.

Table 3: The average length of travel distances for all the methods (unit: meter).

| Layout scenario | *RNA* | *LNA* | *AGNA* |
|---|---|---|---|
| 1 | *247.65* | *247.65* | *247.65* |
| 2 | *231.14* | *231.14* | *231.14* |
| 3 | 208.90 | 208.90 | *171.55* |
| 4 | *381.48* | *381.48* | *381.48* |
| 5 | *370.02* | *370.02* | *370.02* |
| 6 | 411.86 | 411.86 | *346.48* |
| 7 | 296.15 | 296.15 | *263.05* |
| 8 | 261.81 | 262.78 | *261.64* |
| 9 | 313.63 | 312.04 | *263.48* |
| 10 | 495.13 | 492.15 | *409.51* |
| 11 | 341.70 | 340.67 | *329.56* |
| 12 | 534.14 | 532.94 | *498.92* |

Table 4: The percentage differences between different methods (unit: %).

| Layout scenario | $A/R$ | $A/L$ | $L/R$ |
|---|---|---|---|
| 1 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 |
| 3 | 17.88 | 17.88 | 0 |
| 4 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 |
| 6 | 15.87 | 15.87 | 0 |
| 7 | 11.18 | 11.18 | 0 |
| 8 | 0.06 | 0.43 | -0.37 |
| 9 | 15.99 | 15.56 | 0.51 |
| 10 | 17.29 | 16.79 | 0.6 |
| 11 | 3.55 | 3.26 | 0.3 |
| 12 | 6.59 | 6.38 | 0.22 |

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China [Grants nos. 71701213 and 71401181] and the MOE (Ministry of Education in China) Project of Humanities and Social Sciences [Grants nos. 15YJC630008 and 14YJC630136].

## References

[1] J. L. García, A. Alvarado, J. Blanco, E. Jiménez, A. A. Maldonado, and G. Cortés, "Multi-attribute evaluation and selection of sites for agricultural product warehouses based on an analytic hierarchy process," *Computers and Electronics in Agriculture*, vol. 100, pp. 60–69, 2014.

[2] R. De Santis, R. Montanari, G. Vignali, and E. Bottani, "An adapted ant colony optimization algorithm for the minimization of the travel distance of pickers in manual warehouses," *European Journal of Operational Research*, vol. 267, no. 1, pp. 120–137, 2018.

[3] J. A. Tompkins, J. A. White, Y. A. Bozer, and J. M. A. Tanchoco, *Facilities Planning*, John Wiley & Sons, 2010.

[4] F. Chen, Y. Wei, and H. Wang, "A heuristic based batching and assigning method for online customer orders," *Flexible Services and Manufacturing Journal*, vol. 30, no. 4, pp. 640–685, 2018.

[5] J. Zhang, X. Wang, and K. Huang, "On-line scheduling of order picking and delivery with multiple zones and limited vehicle capacity," *OMEGA - The International Journal of Management Science*, vol. 79, pp. 104–115, 2018.

[6] C. G. Petersen II, "An evaluation of order picking routeing policies," *International Journal of Operations and Production Management*, vol. 17, no. 11, pp. 1098–1111, 1997.

[7] K. Gue, R. Meller, and J. Skufca, "The effects of pick density on order picking areas with narrow aisles," *Institute of Industrial Engineers (IIE). IIE Transactions*, vol. 38, no. 10, pp. 859–868, 2006.

[8] P. J. Parikh and R. D. Meller, "Estimating picker blocking in wide-aisle order picking systems," *Institute of Industrial Engineers (IIE). IIE Transactions*, vol. 41, no. 3, pp. 232–246, 2009.

[9] P. J. Parikh and R. D. Meller, "A travel-time model for a person-onboard order picking system," *European Journal of Operational Research*, vol. 200, no. 2, pp. 385–394, 2010.

[10] S. Hong, A. L. Johnson, and B. A. Peters, "Batch picking in narrow-aisle order picking systems with consideration for

picker blocking," *European Journal of Operational Research*, vol. 221, no. 3, pp. 557–570, 2012.

[11] F. Chen, H. Wang, C. Qi, and Y. Xie, "An ant colony optimization routing algorithm for two order pickers with congestion consideration," *Computers & Industrial Engineering*, vol. 66, no. 1, pp. 77–85, 2013.

[12] F. Chen, H. Wang, Y. Xie, and C. Qi, "An ACO-based online routing method for multiple order pickers with congestion consideration in warehouse," *Journal of Intelligent Manufacturing*, vol. 27, no. 2, pp. 389–408, 2016.

[13] F. Chen, G. Xu, and Y. Wei, "Heuristic routing methods in multiple-block warehouses with ultra-narrow aisles and access restriction," *International Journal of Production Research*, vol. 57, no. 1, pp. 228–249, 2018.

[14] R. de Koster, T. Le-Duc, and K. J. Roodbergen, "Design and control of warehouse order picking: a literature review," *European Journal of Operational Research*, vol. 182, no. 2, pp. 481–501, 2007.

[15] R. W. Hall, "Distance approximations for routing manual pickers in a warehouse," *Institute of Industrial Engineers (IIE). IIE Transactions*, vol. 25, no. 4, pp. 76–87, 1993.

[16] N. Gademann and S. van de Velde, "Order batching to minimize total travel time in a parallel-aisle warehouse," *Institute of Industrial Engineers (IIE). IIE Transactions*, vol. 37, no. 1, pp. 63–75, 2005.

[17] A. Alvarado-Iniesta, J. L. Garcia-Alcaraz, M. I. Rodriguez-Borbon, and A. Maldonado, "Optimization of the material flow in a manufacturing plant by use of artificial bee colony algorithm," *Expert Systems with Applications*, vol. 40, no. 12, pp. 4785–4790, 2013.

[18] H. Guo, C. Li, Y. Zhang, C. Zhang, and Y. Wang, "A nonlinear integer programming model for integrated location, inventory, and routing decisions in a closed-loop supply chain," *Complexity*, vol. 2018, Article ID 2726070, 17 pages, 2018.

[19] C. Huang, Y. Lan, Y. Liu et al., "A new dynamic path planning approach for unmanned aerial vehicles," *Complexity*, vol. 2018, Article ID 8420294, 17 pages, 2018.

[20] O. Kulak, Y. Sahin, and M. E. Taner, "Joint order batching and picker routing in single and multiple-cross-aisle warehouses using cluster-based tabu search algorithms," *Flexible Services and Manufacturing Journal*, vol. 24, no. 1, pp. 52–80, 2012.

[21] M. Matusiak, R. de Koster, L. Kroon, and J. Saarinen, "A fast simulated annealing method for batching precedence-constrained customer orders in a warehouse," *European Journal of Operational Research*, vol. 236, no. 3, pp. 968–977, 2014.

[22] C. A. Valle, J. E. Beasley, and A. S. da Cunha, "Optimally solving the joint order batching and picker routing problem," *European Journal of Operational Research*, vol. 262, no. 3, pp. 817–834, 2017.

[23] C.-C. Lin, J.-R. Kang, C.-C. Hou, and C.-Y. Cheng, "Joint order batching and picker Manhattan routing problem," *Computers & Industrial Engineering*, vol. 95, pp. 164–174, 2016.

[24] E. H. Grosse, C. H. Glock, and R. Ballester-Ripoll, "A simulated annealing approach for the joint order batching and order picker routing problem with weight restrictions," *International Journal of Operations and Quantitative Management*, vol. 20, no. 2, pp. 65–83, 2014.

[25] T.-L. Chen, C.-Y. Cheng, Y.-Y. Chen, and L.-K. Chan, "An efficient hybrid algorithm for integrated order batching, sequencing and routing problem," *International Journal of Production Economics*, vol. 159, pp. 158–167, 2015.

[26] C.-Y. Cheng, Y.-Y. Chen, T.-L. Chen, and J. Jung-Woon Yoo, "Using a hybrid approach based on the particle swarm optimization and ant colony optimization to solve a joint order batching and picker routing problem," *International Journal of Production Economics*, vol. 170, pp. 805–814, 2015.

[27] J. Li, R. Huang, and J. B. Dai, "Joint optimisation of order batching and picker routing in the online retailer's warehouse in China," *International Journal of Production Research*, vol. 55, no. 2, pp. 447–461, 2017.

[28] A. Scholz, S. Henn, M. Stuhlmann, and G. Wäscher, "A new mathematical programming formulation for the single-picker routing problem," *European Journal of Operational Research*, vol. 253, no. 1, pp. 68–84, 2016.

[29] T.-H. Chang, H.-P. Fu, and K.-Y. Hu, "A two-sided picking model of M-AS/RS with an aisle-assignment algorithm," *International Journal of Production Research*, vol. 45, no. 17, pp. 3971–3990, 2007.

[30] X. Xu, T. Liu, K. Li, and W. Dong, "Evaluating order throughput time with variable time window batching," *International Journal of Production Research*, vol. 52, no. 8, pp. 2232–2242, 2014.

[31] S. Henn and V. Schmid, "Metaheuristics for order batching and sequencing in manual order picking systems," *Computers & Industrial Engineering*, vol. 66, no. 2, pp. 338–351, 2013.

[32] K. J. Roodbergen and R. De Koster, "Routing methods for warehouses with multiple cross aisles," *International Journal of Production Research*, vol. 39, no. 9, pp. 1865–1883, 2001.

[33] J. C.-H. Pan, P.-H. Shih, and M.-H. Wu, "Order batching in a pick-and-pass warehousing system with group genetic algorithm," *OMEGA - The International Journal of Management Science*, vol. 57, pp. 238–248, 2015.

[34] K. J. Roodbergen and R. de Koster, "Routing order pickers in a warehouse with a middle aisle," *European Journal of Operational Research*, vol. 133, no. 1, pp. 32–43, 2001.

Hindawi
Complexity
Volume 2019, Article ID 8431784, 21 pages
https://doi.org/10.1155/2019/8431784

*Research Article*

# Intelligent Turning Tool Monitoring with Neural Network Adaptive Learning

**Maohua Du**,[1] **Peixin Wang**,[1] **Junhua Wang**,[2] **Zheng Cheng**,[1] **and Shensong Wang**[1]

[1]*Department of Mechanical Engineering, Kunming University of Science and Technology, Kunming 650500, China*
[2]*Department of Mechanical Engineering, Tsinghua University, Beijing 100084, China*

Correspondence should be addressed to Maohua Du; 1337289843@qq.com

Received 16 January 2019; Accepted 28 March 2019; Published 10 June 2019

Guest Editor: Julio Blanco-Fernández



Tool state monitoring is a key technology in intelligent manufacturing. But it is still in a research stage and lacks general adaptability for different machining conditions. To overcome this limitation, this work systematically investigates an intelligent, real-time, and visible tool state monitoring through adopting integrated theories and technologies, i.e., (a) through distinctively designed experimental technique with comprehensive consideration of cutting parameters and tool wear values as variables, (b) through bisensor fusion for simultaneous measurements of low and high frequency signals, (c) through multitheory fusion of wavelet decomposition and the Relief-F algorithm for performing dual feature extraction and feature dimension reduction to achieve more accurate state identification using neural network, and (d) through an innovative programming technique of MATLAB-nested labVIEW. This tool monitoring system has neural network adaptive learning ability with the change of the experimental sample signals which are collected simultaneously by selected vibration and acoustic emission (AE) sensors in *all factors* turning experiments. Based on LabVIEW and MATLAB hybrid programming, the waveforms of signals are observed and the significant characteristics of signals are extracted through the time-frequency domain analysis and then the calculation of the energy proportion of each frequency band obtained using 4 levels of wavelet packet decompositions of the vibration signal as well as 8 levels of wavelet multiresolution decompositions of the AE signal; the ensuing Relief-F algorithm is used to further reextract the features that are most relevant to the tool state as input of neural network pattern recognition. Through the BP neural network adaptive learning, tool state recognition model is finally established. The results show that the correct recognition rate of BP network model after samples training is 92.59%, which can more accurately and intelligently monitor the tool wear state.

## 1. Introduction

Tool condition monitoring is an indispensable technology in manufacturing automation and intelligence. The real-time monitoring of cutting tools can not only improve the quality of products and reduce production costs, but also reduce the downtime of machining centers and increase productivity.

Over the years, researchers have done a lot of work on tool condition monitoring and put forward many monitoring methods, which can be classified into two major categories: direct and indirect methods.

Direct methods [1, 2] are measurements of the cutting tool wear using optical, radioactive, electrical resistance sensors, etc. This kind of method directly measures the actual geometric parameters of the cutting tool [3], and the results are intuitive and reliable, and the measuring accuracy is high. However, the cutting zone is generally inaccessible during the cutting processes due to the continuous contact between the cutting tool and the workpiece, so it is impossible to realize online monitoring; and the sudden breakage of the cutting edge is unable to be monitored; and it is not feasible in the case of coolant. Hence it has limitations in practical applications.

Indirect method is a new developed technology on the basis of modern sensors, signal analysis, and pattern recognition. It monitors tool wear through analyzing and processing measured physical signals in cutting processes. It is characterized by real-time performance, versatility, convenient installation, and low cost, so it is much easier to be employed in practice. In general, indirect tool condition monitoring system consists of sensors, the signal conditioner or amplifier,

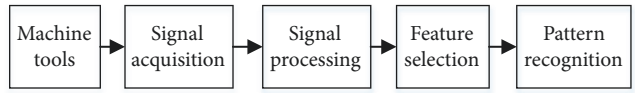



Figure 1: The framework of the tool monitoring system.

and the monitor. Sensors are very important elements which must monitor the target location (near the tip of the tool) as close as possible. Then the received signals by sensors are processed to get useful information. A monitor is a display unit for analyzing signals. The indirect state monitoring system usually consists of four parts: signal acquisition, signal processing, and feature extraction and pattern recognition. Figure 1 shows the framework of tool monitoring system.

Indirect methods have been investigated by many researchers. The commonly monitored parameters with the change of tool include the cutting force, feed and spindle motor currents, vibration signal, acoustic emission (AE) signal, temperature, and surface roughness.

Dimla and Lister [4] developed an online tool wear monitoring system. They used tool dynamometer to measure three vertical components of the cutting force and the tool wear state monitoring was studied through the analyses of time series and fast Fourier transform (FFT) of the signals. The results showed that the cutting force was mainly determined by the cutting conditions, especially the depth of cut and the feed rate; and the vertical component of cutting force and vibration characteristics (z direction) were most sensitive to tool wear. Ghasempoor et al. [5] experimentally verified the feasibility of the online tool state monitoring system by using different cutting force components. They correlated the cutting force components with the tool wear and observed that these parameters were very sensitive to the change of the cutting conditions. However, the prediction of the wear of the crescent tooth was not very accurate, and the possible reason was that the wear of the tool rake face had the counteraction on the component cutting forces. Sikdar and Chen [6] studied the relationship between the wear area of the tool flank and the cutting force in the cutting process. The experimental results showed that the cutting force gradually increased with the increase of the wear area of the tool flank. Nouri et al. [7] described a new method to monitor end milling tool wear in real-time by tracking force model coefficients during the cutting process. The behaviors of these coefficients were shown to be independent from the cutting conditions and correlated with the wear state of the cutting tool. Wang et al. [8] took milling force as the monitoring signal, and wavelet packet theory was used to analyze and extract the energy feature of the signal as a basis for diagnosis and then the Continuous Hidden Markov Model (CHMM) was used to diagnose tool wear state. The results indicated that CHMM tool condition monitoring required less training samples and produced results quickly. However, the diagnostic accuracy seemed obscure due to the lack of contrast between experimental and diagnostic results.

Li et al. [9] presented that the feed cutting force estimation using the current from servo motor was applied to monitor the tool wear in turning. However, whether the feed force component could represent the influence of overall cutting physical parameters on tool wear still needs to be proved.

Many researchers studied the tool state monitoring by analyzing the generated vibration in machining operations. Teti et al. [10] classified the vibration during metal cutting process as free vibration and forced vibration. The two groups are not mutually exclusive. Free vibration independent of metal cutting contains forced vibration caused by other machines or machine components, for example, vibration transmitted through foundations, unbalance of rotating parts, inertia forces of reciprocating parts, and kinematic inaccuracies of drives.

Abouelatta and Madl [11] established the mapping relation between the surface roughness of the workpiece and the vibration by using the vibration signal. The surface roughness was measured using the Surtronic 3+ measuring instrument and the tool vibrations in radial and feed directions were measured by the FFT analyzer. It was found that the tool wear and the surface roughness could be predicted more accurately by the vibration model. Alonso and Salgado [12] developed a tool state monitoring system based on the tool vibration signal by using the singular spectrum analysis (SSA) and cluster analysis. SSA is a nonparametric technique of time series analysis. The sampled vibration signals can be decomposed into a group of time series, and the clustering analysis is used for grouping the time series of SSA decomposition to obtain several independent components in the frequency domain. The results showed that using SSA and clustering analysis to process signals could achieve more reliable results in the development of tool condition monitoring system.

During the metal cutting process, the plastic deformation occurs when the tool cuts into the workpiece. Due to the plastic deformation, transient elastic stress waves are generated inside the material, which quickly releases strain energy from one or more local sources. The released energy is usually called acoustic emission (AE). Li [13] described that AE was one of the most effective methods to monitor tool wear. The main advantage of using AE to monitor tool state is that the frequency range of AE signal is much higher than that of the machine vibrations and environmental noises, and it does not interfere with the cutting process. Kakade et al. [14] predicted the tool wear and chip shape in the milling process through the analysis of AE signal. The tool flank wear and AE signal parameters were measured at the fixed distance interval of the tool pass and the chips were collected at the same time. The results indicated that the AE signal clearly demarcated the cutting action between sharp and worn-out tool and that signal features including ring down counts, rising time, and event duration were capable of reflecting the tool wears.

It is normal and inevitable to generate heat in metal cutting process. The high temperature around the tip of the cutting tool will speed up the damage of the tool. To some extent, the tool temperature determines the tool wear rate. Korkut et al. [15] established the regression analysis (RA) and artificial neural network (ANN) model to predict the tool state by monitoring the chip interface temperature, and the results were verified by experiments. It was found that the proposed model had good accuracy and was suitable for predicting tool state.

Surface roughness is widely used in quality evaluation of machined products, which indirectly indicates tool condition. Cutting conditions have considerable influences on tool wear and surface roughness. Ozel and Karpat [16] used neural network modeling to predict surface roughness and tool flank wear in finish hard turning. The regression model was developed to obtain the specific parameters in the cutting process. The results showed that decrease of the feed rate resulted in better surface roughness but slightly faster tool wear, and the increase of cutting speed resulted in a significant increase of tool wear but better surface roughness; the increase of workpiece hardness resulted in better surface roughness but higher tool wear.

Artificial neural network (ANN) is a mathematical model for information processing by simulating the human brain. It consists of several processing units similar to neurons. It memorizes knowledge and stores it in the network by training and learning samples. Because of high fault tolerance and adaptability, noise suppression, and data driving, neural network classifier has been adopted by some researchers. Balazinski et al. [17] established three models: feedforward backpropagation neural network, fuzzy clustering analysis, and fuzzy reasoning based on neural network by collecting the cutting force signals during the cutting process. The results showed that the three models estimated tool wear with almost the same precision. But the fuzzy clustering analysis method was somewhat difficult to be applied in practice because it had to analyze the dependence of tool wear on the cutting force and the training time of the neural network method was long. Silva et al. [18] utilized two neural networks and an expert system using Taylor's tool life equation to classify the tool wear state; the classification accuracy results, however, were puzzling because of the missing Figures 16–18 and the simple linear regression analysis. The research of monitoring tool wear using classifier fusion by Kannatey-Asibu et al. [19] illustrated that the classifier fusion performance reached 99.7% when a penalty vote was applied on the weighting factor. However, whether the demonstrated technique using AE monitoring is suitable for low frequency signal monitoring or not, further research is still needed.

By and large, the tool state monitoring technology is not mature enough due to the complexity of the relationship between the measured physical quantities by sensors and the tool state, and thus the establishment of the accurate relationship between them is still a hard problem to be solved when we intend to develop a practical tool state monitoring system to meet various machining processes and environments. Specifically, there exist the following main problems. (1) First is the adaptability of monitoring models. Most of the existing monitoring models are suitable for narrow cutting conditions, and the reliability and effectiveness under variable cutting conditions cannot be guaranteed. The metal cutting process is very complex and has a lot of randomness, and different machine tools, cutting tools, workpiece materials and cutting parameters will make the monitored signal characteristics different. Thus it is of great significance to establish a tool state monitoring system that can adapt to the changeable cutting conditions. (2) Second is sample acquisition problem. Sample acquisition in cutting is difficult and costly in some cases. A large number of cutting experiments will cause high costs. Therefore, it is essential to establish a model with self-learning ability and little dependence on training samples. (3) Third is the difficulty of developing the monitoring system based on new sensor and signal analysis technology. Tool state monitoring involves a variety of sensors and signal analysis techniques, and the selection of the suitable sensors and signal analysis methods have not been systematically studied.

Hence, the aim of this study is to investigate a method of achieving a visual, intelligent, and real-time tool state monitoring that is universal and can be used in practice. The implementation of tool wear monitoring is divided into two major parts: experimental details, and design and implementation of tool monitoring functions. The contents include (1) using bisensor fusion to measure the signals (low frequency vibration signal and higher frequency AE signal) in the turning process; (2) being based on hybrid programming technique of LabVIEW and MATLAB, analyzing and processing the signals, and extracting the most relevant features that reflect the tool wear state by the use of dual methods, i.e., the combination of the wavelet theory with the Relief-F algorithm; (3) using the extracted features as training samples of the neural network and then establishing the recognition model of tool state, such that intelligent recognition of the tool state is accomplished.

The designed structure of lathe tool wear monitoring system is shown in Figure 2.

## 2. Experimental Details

Table 1 lists the instrumentation and specifications of the experiments in this work.

*2.1. Selection of Experimental Parameters.* In the turning process, there are many factors affecting the signals of the sensors, such as cutting parameters, tool parameters, the machine tools, degrees of the tool wear, and the machining environment. And so many factors will inevitably bring difficulties to the experimental design, so we need to analyze the specific machining environment and remove some minor factors.

On the premise of some fixed factors, such as the machine tools, the workpiece material, and the cutting tool, the machining parameters are changeable according to the machining requirements. The value of the tool wear is also a variable in the cutting process. Because the experimental objective is to obtain the mapping relations between the tool

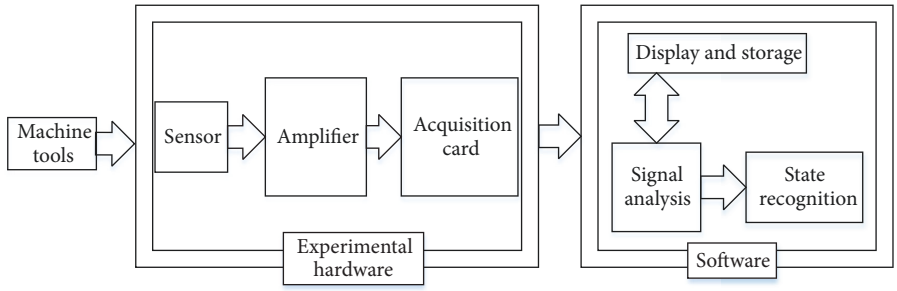



Figure 2: Structural frame of tool condition monitoring.

Table 1: Experimental instrumentation and specifications.

| Name | Specifications |
|---|---|
| Lathe | CA6140 |
| Workpiece material | Aeronautical aluminum alloy 7050, Φ98×500mm bar |
| Cutting tool | Carbide tool YT15 |
| Vibration sensor | Accelerometers; DH112 (Denmark, B & K), frequency range 0.5-7KHz |
| Vibration signal amplifier | 2635 (Denmark, B & K) |
| Data acquisition card | NI9215(US, NI company), 16 bit resolution, sampled at a rate of up to 7 kHz for 100 MS/s |
| Sensor 2 | AE sensor PAC WD US |
| Pre-amplifier | PAC2/4/6 (US) |
| Data acquisition card | PCI-9846H (Taiwan), 16 bit resolution, 40MS/s |

Table 2: Experimental parameters and levels.

| Level | Cutting speed $v_c$ (m/min) | Feed rate $f$ (mm/rev) | Depth of cut $a_p$ (mm) | Tool wear value $VB$(mm) |
|---|---|---|---|---|
| 1 | 64.6 | 0.11 | 0.8 | 0 |
| 2 | 80 | 0.13 | 1.0 | 0.2 |
| 3 | 98.5 | 0.15 | 1.2 | 0.4 |

Table 3: The stages of tool wear.

| Tool wear | sharp tool stage (initial wear) | normal working stage (normal wear) | worn out stage (severe wear) |
|---|---|---|---|
| $VB$(mm) | 0 | 0.2 | 0.4 |

wear and the vibration plus AE signals characteristics, so the experimental factors include the cutting speed, the feed rate, the depth of cut, and tool wear.

According to the specifications of the machine tools, the experimental parameters are designed, as presented in Table 2, which also takes into account three VB values of the tool flank wear corresponding to the indices of tool wear criterions: sharp tool stage (initial wear), normal working stage (normal wear), and worn out stage (severe wear), as listed in Table 3. For each of the experimental cutting conditions, the observations of the vibration and AE signals are made and recorded corresponding to the three wear states of the tool flank.

*2.2. All Factors Experimental Technique.* The orthogonal test is a frequently used method in experimental design. It can greatly reduce the number of trials, the test time, and costs. But the orthogonal test is inappropriate for the modeling and recognition of the neural network since the amount of data is relatively small. Therefore, in this work, the experiments of *all factors* combinations are conducted for obtaining the data of tool state monitoring. According to the test factors and levels designed in Table 2, there are four factors and each factor has three levels. Therefore, the *four factors-three levels* comprise $3^4$=81 groups of experiments.

*2.3. Bisensors Installation and Tool Wear Measurement.* The most common waveform data in condition monitoring are vibration signals and acoustic emissions [20]. In this work, two sensors, i.e. vibration acceleration sensor and AE sensor, are simultaneously selected for measuring the cutting tool signals. The advantages of selecting them lie in that both

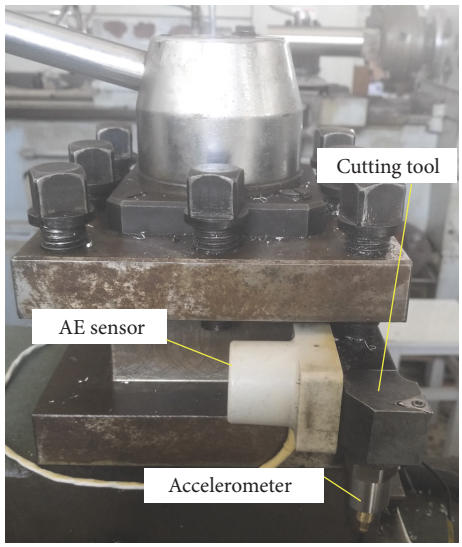



Figure 3: Sensors installation.

vibration and AE signals can sensitively reflect the tool state in real time and that they are suitable for monitoring different machining processes. Figure 3 illustrates the experimental installation of sensors in the turning process. Both the vibration and AE sensors have magnetic bases, which can be installed onto the tool shank magnetically. Two sensors are installed close to the cutting tool. The vibration and the AE sensors are placed in the vertical direction (Z axis) and the horizontal direction (Y axis) of the shank and used to measure vibration and AE signals, respectively.

Three prepared cutting tools for each wear state in the experiments, as shown in Figure 4, are used to cut the workpiece so that the signal data of each tool state can be correspondingly collected by sensors, respectively. Hence each tool wear state corresponds to its own signal data and three tool wear modes (initial, normal, and severe wear) will not mix together. A tool microscope (with five million pixels) that can magnify 100 times is used to measure the tool wear, as shown in Figure 5. It can be connected with the computer, and the tool wear can be observed and measured on the computer screen, and the measurement photos can also be taken.

*2.4. Signal Acquisition Procedure.* The data acquisition is a process of collecting and storing the experimentally measured sensor signals in the turning process, such that it provides reliable sample data for later state recognition.

The flow chart of the signal acquisition procedure is illustrated in Figure 6. In the machining process, the vibration sensor and AE sensor collect the original tool state signals, which are weak and need to be amplified, buffered, and filtered via the amplifiers and data acquisition cards, and then the signals are entered into the computer for storage.

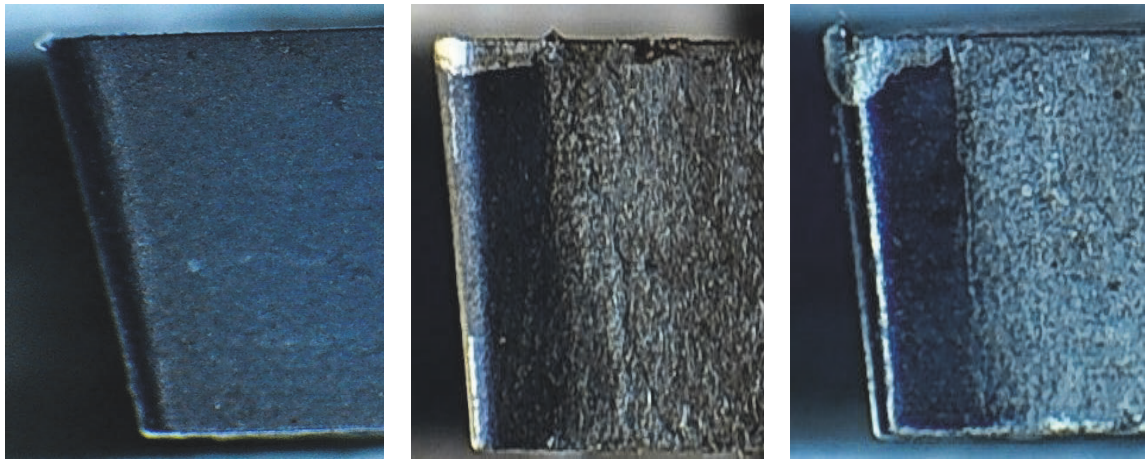
## 3. Tool Monitoring Functions and Implementation Methods

This tool monitoring design includes three functions: signal acquisition or reading, signal processing and feature extraction, and state recognition.

*3.1. Signal Acquisition Function.* The signal acquisition module in this work has the functions of signal display and storage, which is developed using the LabVIEW software. LabVIEW is a graphical programming language. The LabVIEW program is called virtual instrument (VI), which provides lots of graphical controls and functions used in the data analysis and processing in engineering fields. We devise desirable functions and drag graphical controls into the program diagram and then connect the wires according to the needs of programming, and finally the instrument functions and the parameters settings can be realized. LabVIEW also provides the interface with MATLAB, i.e., MATLAB Script. Using MATLAB programming for numerical calculation and analysis can greatly shorten the time of the system development and finish a completely automated test system. LabVIEW software has two human-machine interaction interfaces: the front panel and the rear panel. The front panel provides many graphical controls that are similar to the appearance of actual instruments. Graphical controls on the front panel are manipulated by using icons and wires on the rear panel.

*3.2. Signal Analysis and Method Fusion of Dual Feature Extraction.* In the metal cutting process, the original signals collected by the sensors contain a lot of noises and useless information. They cannot be used directly to identify the tool state and only the extracted characteristics are useful for the state recognition. The collected vibration and AE signals are dynamic signals. The signal components are very complex and the background noise comes from many sources, such as cutting environment, machining conditions, and material properties. There exist both stationary and nonstationary signals. The time domain analysis of signal can only represent time history of the signal amplitude characteristic, but cannot clearly reveal the frequency compositions of the signal. And the frequency domain analysis can represent the frequency structure of the signal and the amplitude of the frequency components, but cannot provide time information and judge the change of local signal with the time. The time-frequency analysis method can reflect the signal characteristics both in time domain and in frequency domain; that is, it can satisfy the requirements of nonstationary signal analysis and processing. Thus the time-frequency domain analysis is used based on the initial time domain analysis and frequency domain analysis.

In this work, firstly, the time domain analysis mainly focuses on the time domain waveform display for observing the waveform change rules, and, secondly, some statistical eigenvalues (mean value $\overline{x}$, variance $\sigma^2$, and root mean square value $X_{rms}$) are calculated according to the sampling number of each signal through LabVIEW programming. The



(a) (b) (c)

Figure 4: Cutting tools of three different wear stages: (a) sharp tool; (b) normal working tool; (c) worn-out tool.

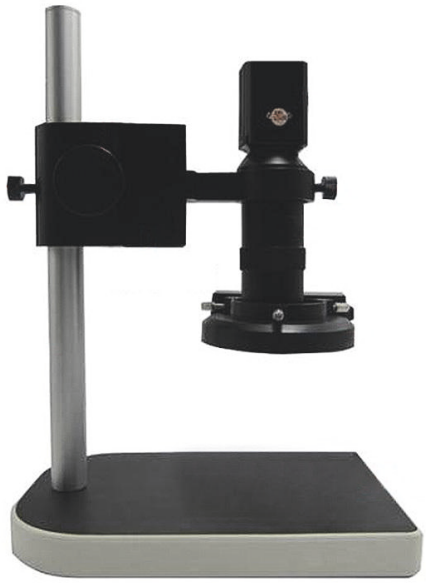


Figure 5: Universal tool microscope.

expressions of three statistical characteristics are presented in the following, respectively.

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1}$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i - \overline{x})^2 \tag{2}$$

$$X_{rms} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \tag{3}$$

where $x_i$ is the collected signal data by the sensor; mean value $\overline{x}$ indicates the constant component of a signal; variance $\sigma^2$ describes the deviation from the mean value of a random signal; root mean square value $X_{rms}$ describes the intensity of a random signal, i=1, 2,...N; and N is sample number.

Figures 7 and 8, respectively, show the typical time domain waveforms of the experimentally collected vibration and AE signals in different cutting tool wear stages (cutting speed 64.6m/min; feed rate 0.11mm/r; depth of cut 1.2mm). The abscissa is sampling time and the ordinate is the amplitude of the signal.

It can be seen from Figure 9 that the frequency components of the vibration signal are not very rich, mainly ranging from 2 to 4 kHz. The signal energy in this frequency band varies noticeably in different wear stages and in other frequency bands it has no observable change. As can be seen from Figure 10, the AE signal contains rich frequency components, which mainly exist between the frequency band 0-50kHz and 100-150 kHz. The signal energies in these two frequency bands change obviously in different tool wear stages, and the energies keep almost unchanged in other frequency bands.

We can see that, according to the time domain analysis and frequency domain analysis on Figures 7–10, it is difficult to describe the characteristics of the tool state and to do the quantitative analysis. Thus further analysis is necessary.

Due to the nature of manufacturing processes, the signals are usually nonstationary and the signal processing approaches that deal with nonstationary signals are more appropriate for process monitoring [3]. We have known that time-frequency analysis is usually used to analyze nonstationary signals which often have both high and low frequency components. And wavelet transform is one of the commonly used time-frequency analysis methods, because it has the zoom function and can adaptively adjust the time-frequency resolution of the signal according to the difference of decomposition scale. Wavelet analysis has good localization property in both time domain and frequency domain and permits adaptive time-frequency representation. Dividing the frequency band into multiple levels can improve

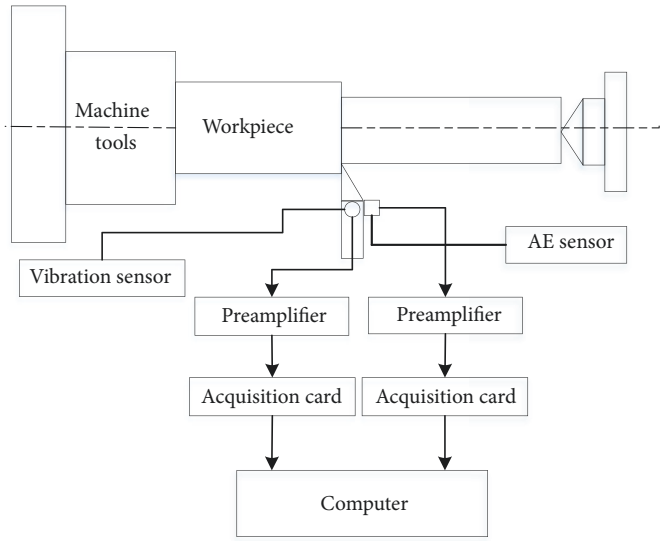



Figure 6: The flow chart of the signal acquisition.

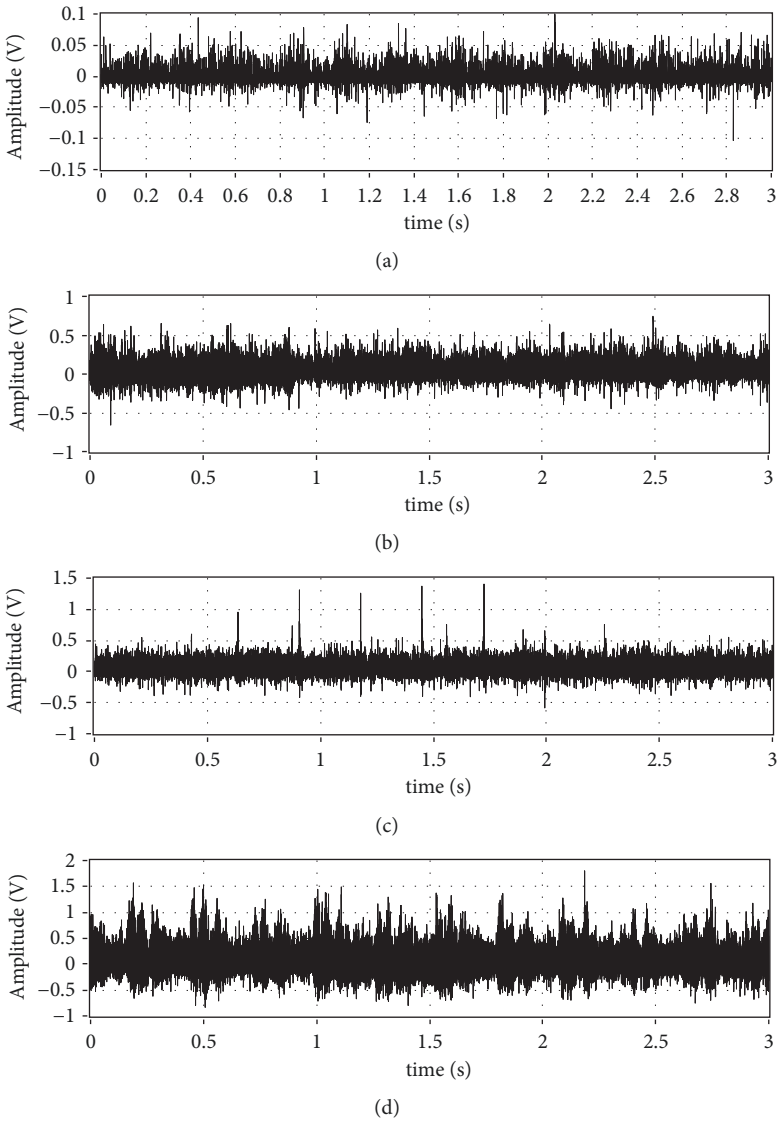



Figure 7: Time domain waveforms of vibration signal: (a) machine idling; (b) initial tool wear; (c) normal tool wear; (d) severe tool wear.

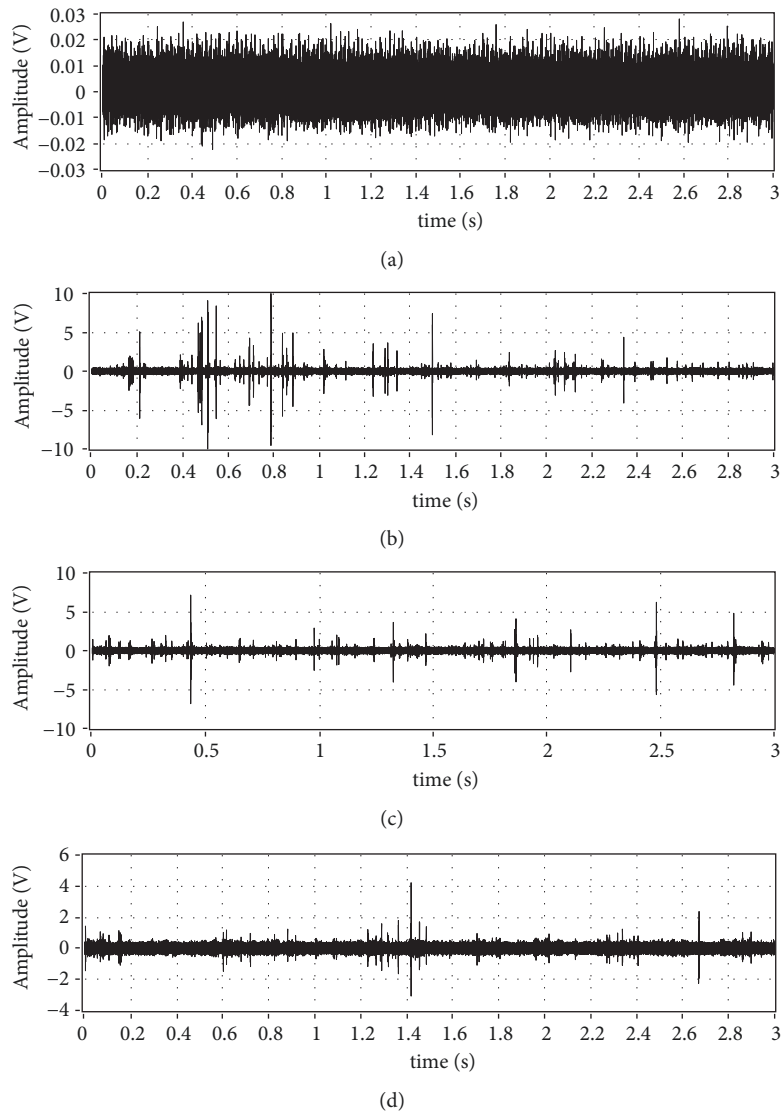



Figure 8: Time domain waveforms of AE signal: (a) machine idling; (b) initial tool wear; (c) normal tool wear; (d) severe tool wear.

frequency resolution of both high and low frequencies. Zhu et al. [3] reviewed that, due to the advantages of wavelet decomposition, wavelet methods have been studied in many aspects, such as signal denoising, feature extraction, and compression, or used as classifier in tool condition monitoring. Gong et al. [21] showed that the wavelet analysis was more sensitive and reliable than the Fourier analysis for recognizing the turning tool wear states. Yoon and Chin [22] verified the reliability of the wavelet transform compared with the spectra method of Fast Fourier transform (FFT).

Continuous wavelet transforms (CWT) are recognized as effective tools for both stationary and nonstationary signal processing. However, it involves much redundant information and takes long computational time. Mallat [23] developed discrete wavelet transform (DWT) which is an algorithm based on the conjugate quadratic filter (CQF) and preferable in the time-frequency analysis because of no redundancy and fast computation. Wavelet and scaling functions of different scales are generated from a scaling function $\phi$(t) with two-scale difference equations [24]:

$$\phi(t) = \sqrt{2}\sum_{k} h(k)\,\phi(2t-k) \tag{4}$$

$$\psi(t) = \sqrt{2}\sum_{k} g(k)\,\phi(2t-k), \tag{5}$$

$$g(k) = (-1)^{k} h(1-k) \tag{6}$$

where $h(k)$ and $g(k)$ are filter coefficients of low-pass and high-pass filters; $\phi$(t) and $\psi$(t) are scaling and wavelet functions at scale space $j$=1, respectively.

Assuming $c_{j,k}$, $d_{j,k}$ are scaling coefficient and wavelet coefficient, respectively, they are derived from the projection

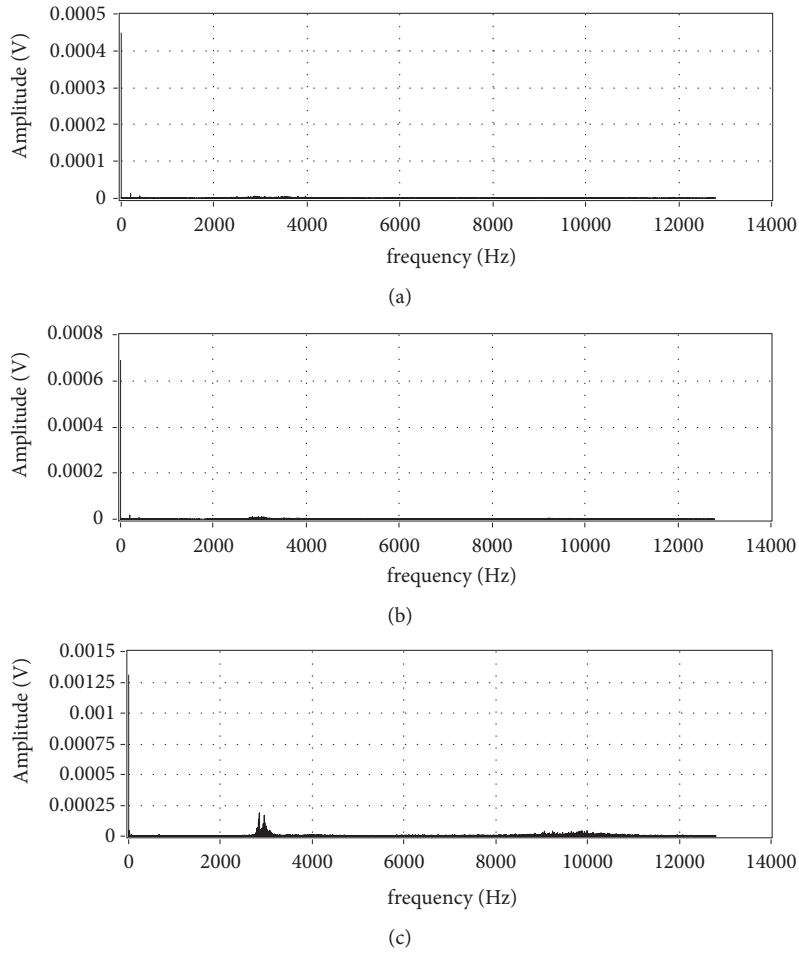



Figure 9: Frequency domain waveforms of the vibration signal (amplified 100 times): (a) initial tool wear; (b) normal tool wear; (c) severe tool wear (cutting speed 64.6m/min; feed speed 0.11mm/r; depth of cut 1.2mm).

of the signal onto the spaces of scaling function $\phi_{j,k}$ (t) and wavelet function $\psi_{j,k}$ (t), *and m* is the filter length, then

$$c_{j,k} = \sum_m h_{m-2k} c_{j+1,m} \tag{7}$$

$$d_{j,k} = \sum_m g_{m-2k} c_{j+1,m} \tag{8}$$

Figure 11 shows a schematic diagram [25] of feature extractions for the signal representation with 3 levels of wavelet packet decompositions. The decomposition process is as follows: assuming that the frequency range of the signal S is $[0, f]$, after the first level decomposition, S is decomposed into two parts, i.e., high frequency part D1 and low frequency part A1. Among them, the frequency range of high frequency signals is $[f/2, f]$, and the low frequency range is $[0, f/2]$. Then, in the second level decomposition, A1 obtained from the first level is decomposed to get the low frequency part AA2 and the high frequency part DA2, and D1 is decomposed to obtain low frequency part AD2 and the high frequency part DD2. Their corresponding frequency ranges are $[0, f/4]$, $[f/4, f/2]$, $[f/2, 3f/4]$, and $[3f/4, f]$, respectively. And likewise, the original signal is decomposed level by level. The decomposition relation of the signal S in Figure 11 can be expressed as the following:

$$\begin{aligned} S = {} & AAA_3 + DAA_3 + ADA_3 + DDA_3 + AAD_3 \\ & + DAD_3 + ADD_3 + DDD_3 \end{aligned} \tag{9}$$

S is initial signal; A1, A2, and A3 are the first, second, and third level of low frequency after decomposition, respectively; D1, D2, and D3 are the first, second, and third level of high frequency after decomposition, respectively.

Figure 12 shows 3 levels of tree structure of wavelet multiresolution decomposition. It can be clearly seen from Figure 12 that multiresolution analysis is actually equivalent to a bandpass filter, and only the low frequency component is decomposed at each level without considering the high frequency. This decomposition process is that the frequency range of the signal S is $[0, f]$, and S is divided into two parts after the first level decomposition, i.e., the high frequency part of D1 and the low frequency part A1. The frequency range of high frequency component is $[f/2, f]$, and the low frequency part is $[0, f/2]$. When the second level is decomposed, the high frequency part D1 obtained from the first level decomposition is no longer decomposed; only

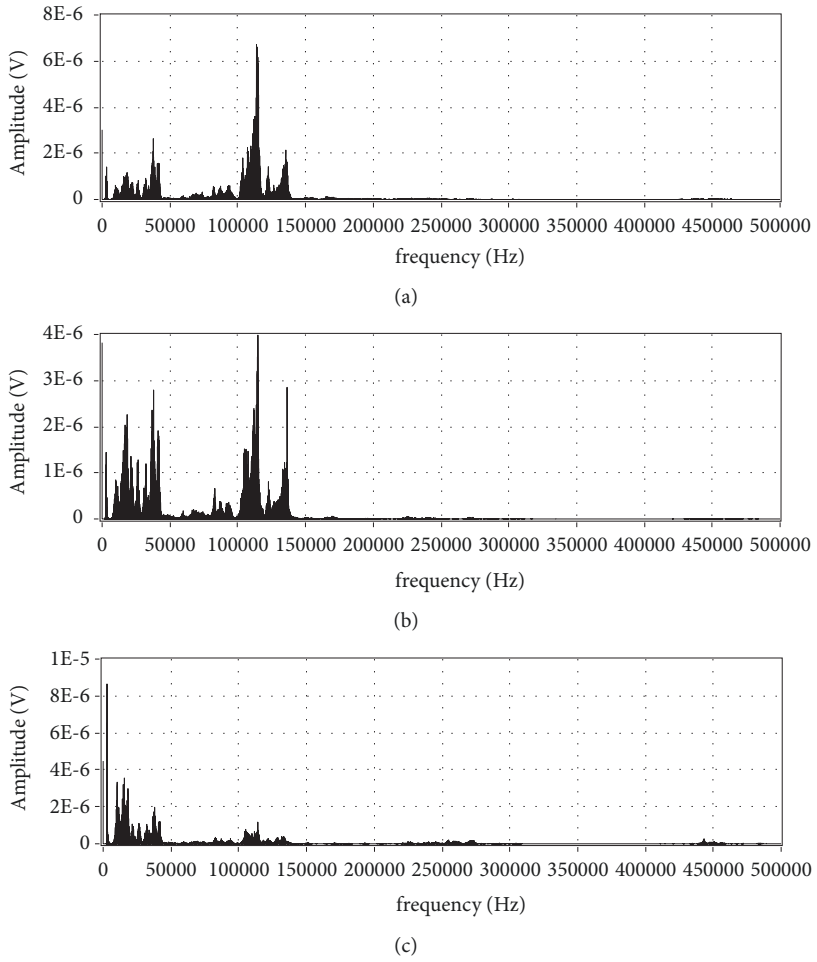



Figure 10: Frequency domain waveforms of AE signal: (a) initial tool wear; (b) normal tool wear; (c) severe tool wear (cutting speed 64.6m/min; feed speed 0.11mm/r; depth of cut 1.2mm).

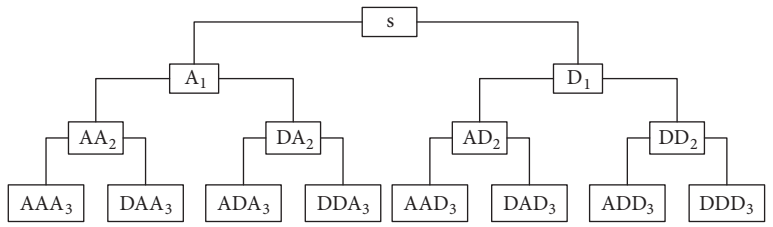



Figure 11: Wavelet packet decomposition structure.

the low frequency part A1 is decomposed, and it is also decomposed into two parts: high frequency part D2 and low frequency part A2. The frequency range of the high frequency part D2 is $[f/4, f/2]$, and the low frequency part A2 has a frequency range of $[0, f/4]$. The third level of decomposition is the same as the second level and only the low frequency part A2 is decomposed into two parts, i.e., the high-frequency part D3 and the low frequency part A3. And likewise, the signal is decomposed level by level. After the signal S is decomposed by the three levels of multiresolution, the signals A3, D3, D2, and D1 are obtained, and the corresponding frequency ranges are $[0, f/8]$, $[f/8, f/4]$, $[f/4, f/2]$, and $[f/2, f]$. Therefore, we have the following formula:

$$S = A_3 + D_3 + D_2 + D_1 \tag{10}$$

When the signal is decomposed by wavelet packet, if the number of decomposed levels is too many, the calculation time of the computer will increase; if the number is too little, the resolution will reduce and many useful information may not be resolved. The number of decomposition levels depends on the signal itself and the requirements of the characteristic parameters. In this work, considering that the frequency of the vibration signal is relatively low and the frequency of AE

Table 4: Frequency bands of 4 levels of wavelet packet decompositions for the vibration signal.

| Sequence No. | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 |
|---|---|---|---|---|---|---|---|---|
| Frequency band (KHz) | 0-0.8 | 0.8-1.6 | 1.6-2.4 | 2.4-3.2 | 3.2-4 | 4-4.8 | 4.8-5.6 | 5.6-6.4 |
| Sequence No. | A9 | A10 | A11 | A12 | A13 | A14 | A15 | A16 |
| Frequency band (KHz) | 6.4-7.2 | 7.2-8 | 8-8.8 | 8.8-9.6 | 9.6-10.4 | 10.4-11 | 11.2-12 | 12-12.8 |

Table 5: Energy proportion ($\times 10^2$) of the vibration signal in each frequency band after wavelet packet decompositions.

| No. | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 |
|---|---|---|---|---|---|---|---|---|
| Initial wear | 9.955 | 2.999 | 17.77 | 5.762 | 3.607 | 3.146 | 17.85 | 6.105 |
| Normal wear | 12.42 | 2.252 | 16.44 | 2.877 | 3.708 | 2.87 | 15.68 | 5.308 |
| Severe wear | 4.575 | 0.965 | 19.20 | 1.283 | 3.218 | 2.298 | 11.19 | 3.675 |
| No. | A9 | A10 | A11 | A12 | A13 | A14 | A15 | A16 |
| Initial wear | 2.115 | 2.325 | 6.591 | 3.453 | 4.055 | 3.371 | 6.933 | 3.953 |
| Normal wear | 2.27 | 2.556 | 8.706 | 4.231 | 3.92 | 3.493 | 8.855 | 4.408 |
| Severe wear | 1.976 | 2.854 | 15.43 | 6.708 | 3.346 | 3.649 | 13.72 | 5.903 |

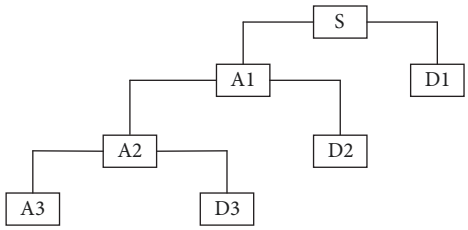



Figure 12: Tree structure for three level of wavelet multiresolution analysis.

signal is high, in order to achieve good frequency resolution effects, the vibration signal is decomposed by four levels of wavelet packet analysis, and the AE signal is decomposed by eight levels of multiresolution analysis. After this, the energy proportion of each frequency band is extracted. This process is realized by mixed programming of LabVIEW and MATLAB. Figure 13(a) shows the block diagram of the 4 levels of wavelet packet decompositions for vibration signal. Figure 13(b) is the front panel of the time-frequency analysis, which can observe the waveform of the frequency bands and the energy proportion of each frequency band after the signal is decomposed by wavelet analysis.

*3.2.1. Wavelet Packet Analysis of Vibration Signal and Characteristic Extraction.* The experimental sampling frequency of the vibration signal is 25.6 KHz. According to the Nyquist sampling theorem (the sampling frequency is two times more than the actual maximum frequency), the available frequency is 12.8 KHz. Therefore, the collected vibration signal is decomposed into four levels by the wavelet packet analysis, and the frequency resolution is 0.8 KHz. Finally, the vibration signal is decomposed into 16 frequency bands (A1-A16), and the frequency distribution of each frequency band is presented in Table 4.

Figure 14 shows the amplitudes and waveforms of the vibration signal with the change of the tool wear state, so the internal relationship of the amplitude with the tool state can be studied by further analysis of the change of each frequency band.

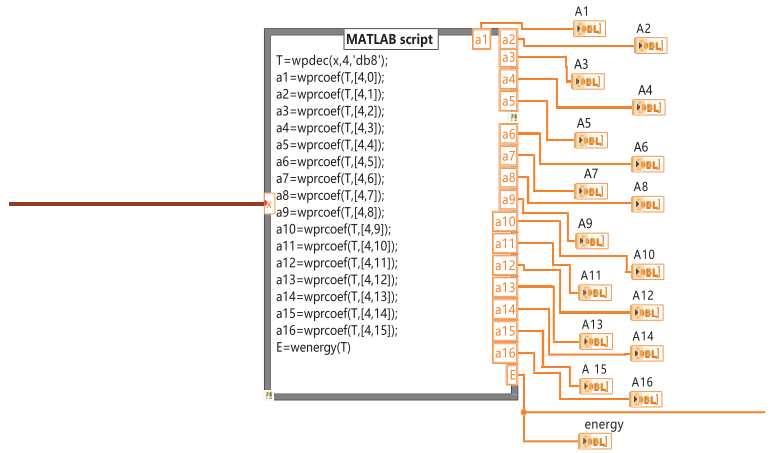
In order to find out the change rules of each frequency band, we quantify the amplitude of Figure 14 by summing up the squared wavelet coefficients of the decomposed 16 frequency bands and then by normalizing to obtain energy proportion of each frequency band. Table 5 lists the energy proportion of the signal for three different tool states under a given cutting condition (cutting speed 64.6m/min; feed rate 0.11mm/r; cutting depth 1.2mm).

The histogram for Table 5 is drawn to give a more direct view of the energy proportion change in each frequency band and the comparison of the energy proportion of three tool wear states in each frequency band, as shown in Figure 15. It can be seen from Figure 15 that the energy of the vibration signal is mainly distributed in several frequency bands (A1, A3, A7, A11, and A15). The energy proportion of frequency bands A2, A4, A7, and A8 decreases with the increase of tool wear, and the energy proportion of frequency bands A11, A12, and A15 increases with the tool wear. The change in other frequency bands is irregular or no obvious change.

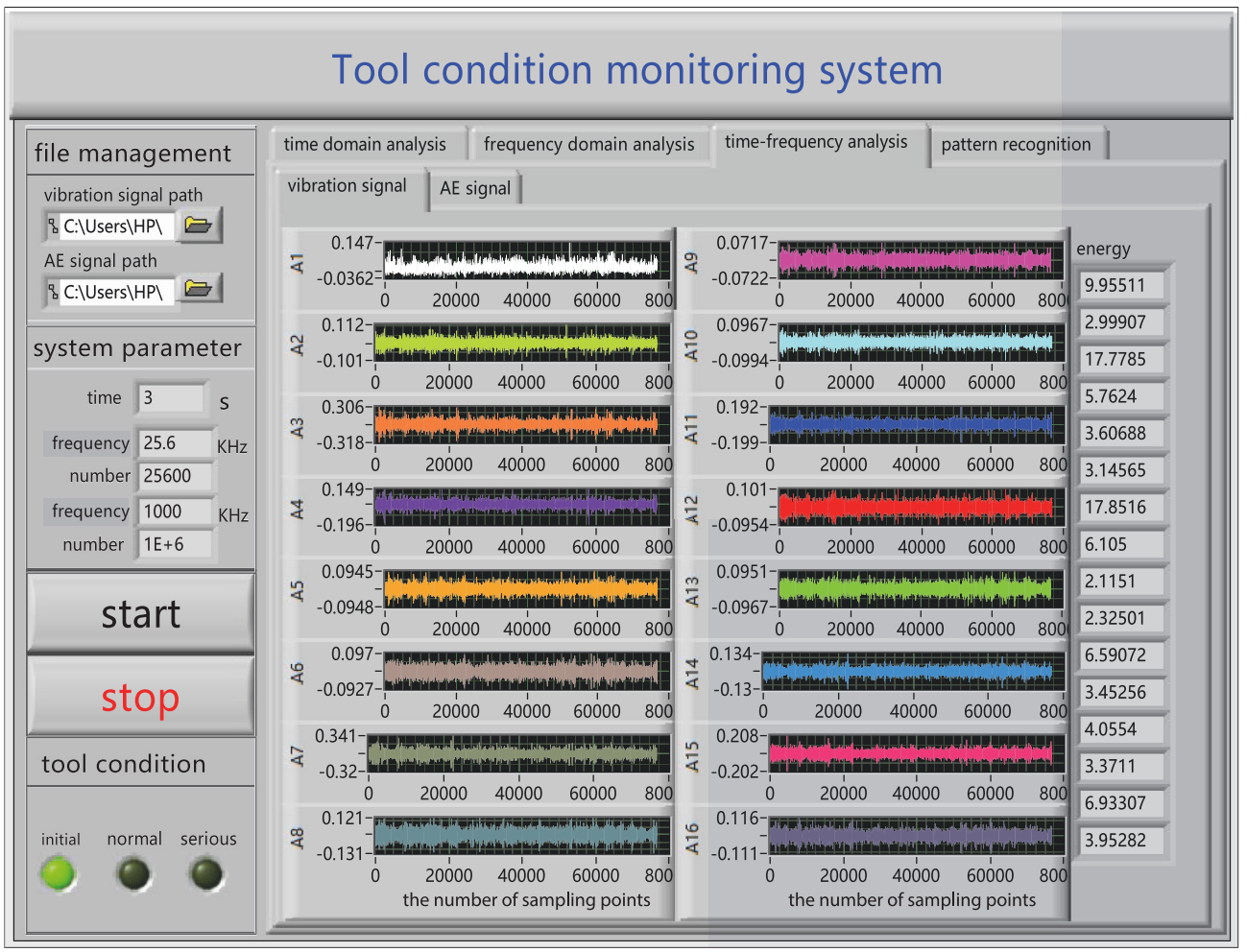
*3.2.2. Wavelet Multiresolution Analysis of AE Signals and Characteristic Extraction.* The experimental sampling frequency of AE signal is 1MHz. According to Nyquist sampling theorem, we know that the available frequency of AE signal is 500 KHz. After 8 levels of wavelet multiresolution analysis, the frequency resolution of AE signal is 1.95 KHz. Finally, AE signal is decomposed into 9 bands (A1, D1-D8), and the distribution of each frequency band is listed in Table 6.

The multiresolution waveform diagram of the AE signal is shown in Figure 16. It can be seen that the amplitudes in some frequency bands vary with the change of the tool state. So we can study the internal relationship of the amplitude with the tool state by further analysis of the change of each frequency band.

The amplitude of each frequency band is quantized by the sum of the squared wavelet coefficients of the AE signal in nine bands obtained after the decompositions and




(a)

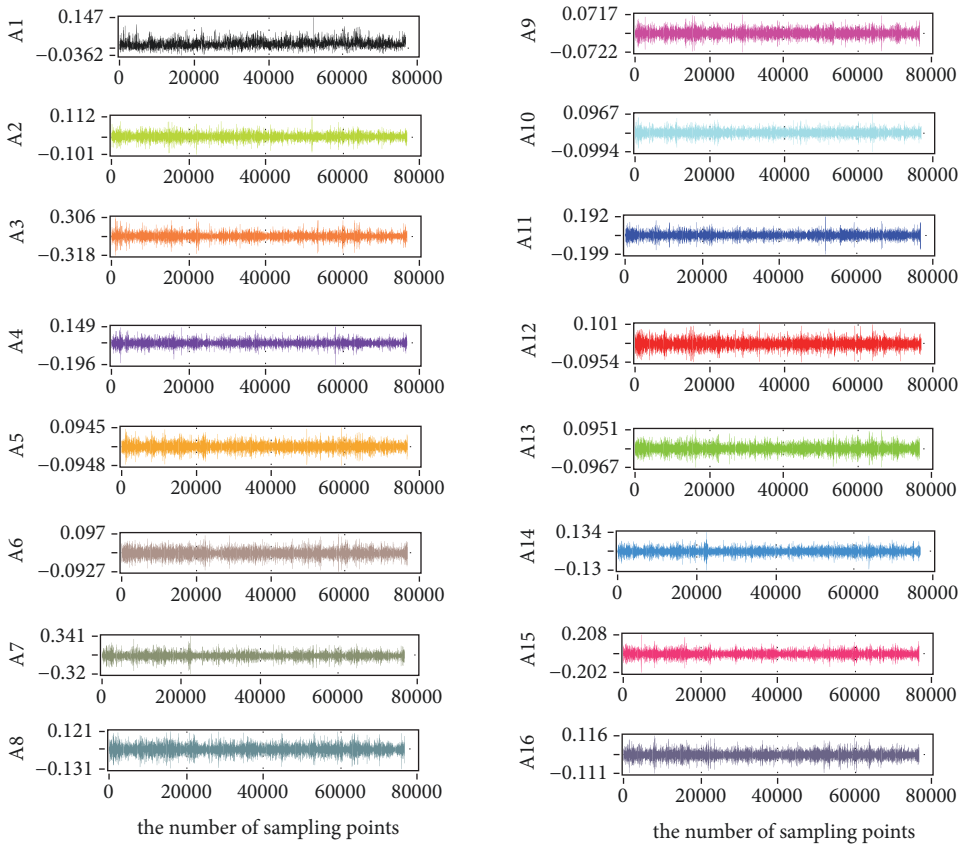



(b)

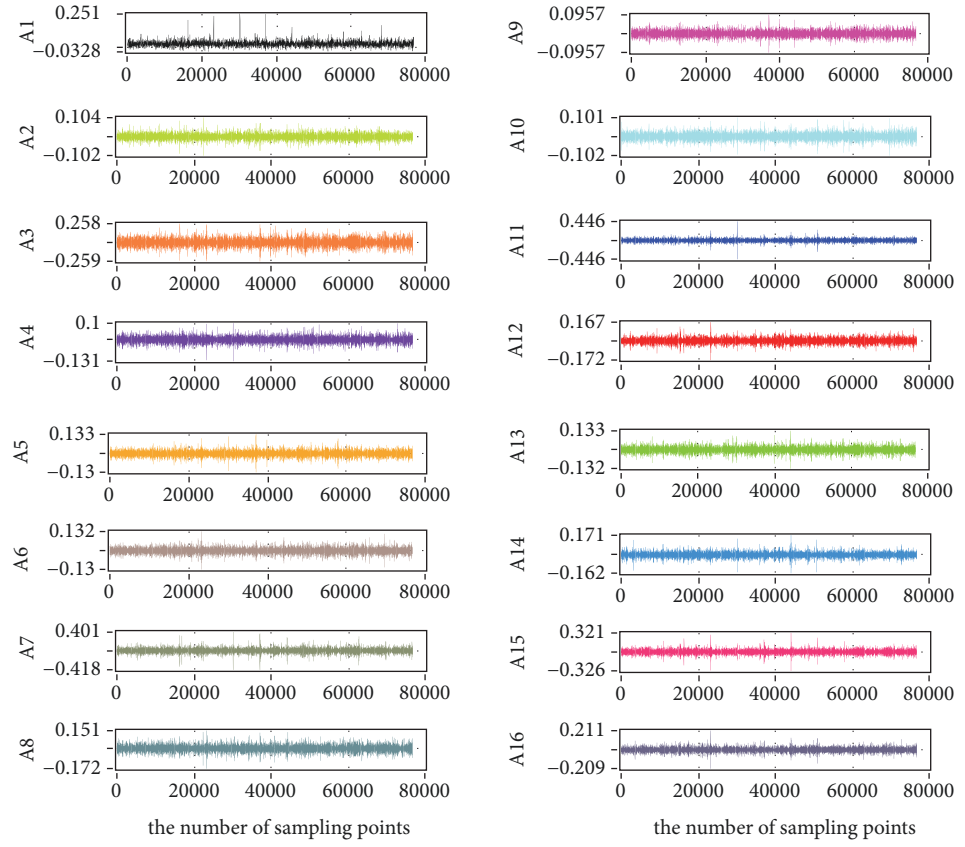
Figure 13: (a) (upper) Block diagram of the 4 levels of wavelet packet decompositions for vibration signal; (b) (lower) front panel of the time-frequency analysis.

Table 6: Frequency bands of AE signal after eight levels of multiresolution decompositions.

| No. | A8 | D8 | D7 | D6 | D5 | D4 | D3 | D2 | D1 |
|---|---|---|---|---|---|---|---|---|---|
| Frequency band (KHz) | 0-1.95 | 1.95-3.9 | 3.9-7.8 | 7.8-15.6 | 15.6-31.25 | 31.25-62.5 | 62.5-125 | 125-250 | 250-500 |

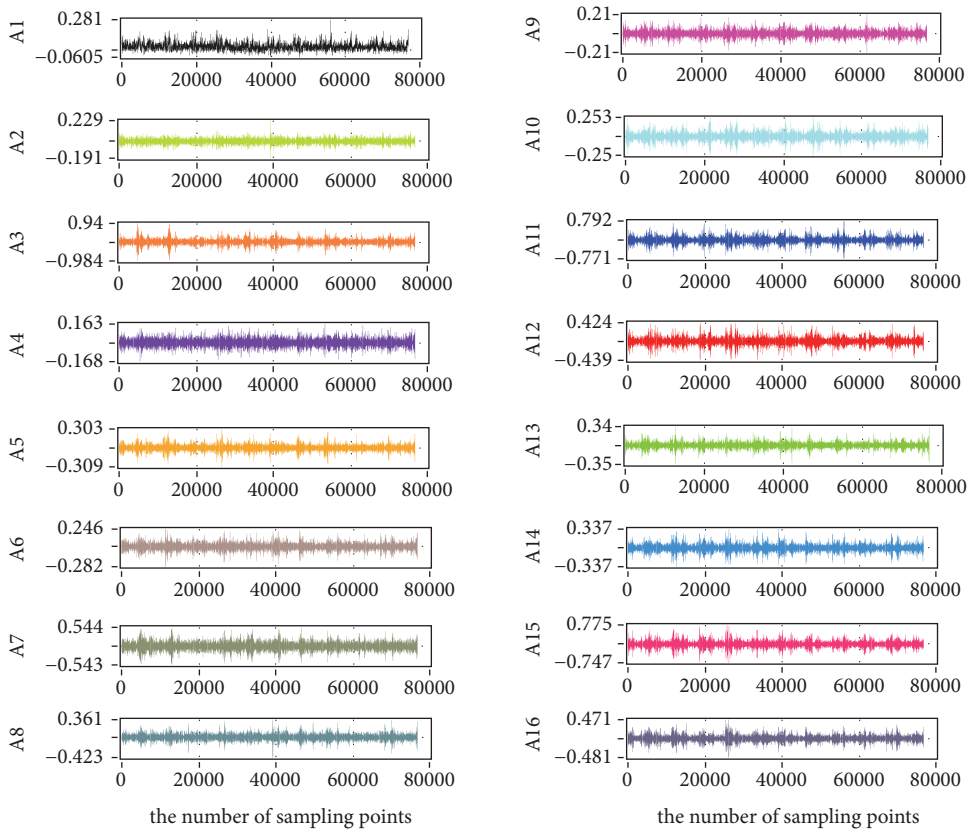



(a) Initial tool wear




(b) Normal tool wear

Figure 14: Continued.




(c) Severe tool wear

Figure 14: Waveforms of the vibration signal after Wavelet packet decompositions.

Table 7: Energy proportion ($\times 10^2$) of AE signal after wavelet decompositions in each frequency band.

| | A8 | D8 | D7 | D6 | D5 | D4 | D3 | D2 | D1 |
|---|---|---|---|---|---|---|---|---|---|
| Initial wear | 0.019 | 0.77 | 0.268 | 3.092 | 8.868 | 11.683 | 47.688 | 26.72 | 0.892 |
| Normal wear | 0.041 | 0.993 | 0.373 | 6.331 | 16.38 | 16.969 | 38.883 | 19.24 | 0.79 |
| Severe wear | 0.178 | 4.411 | 1.116 | 21.599 | 23.418 | 16.601 | 17.31 | 9.754 | 5.613 |

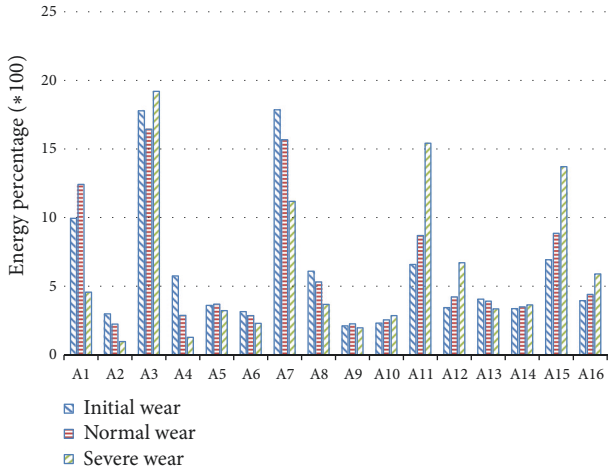
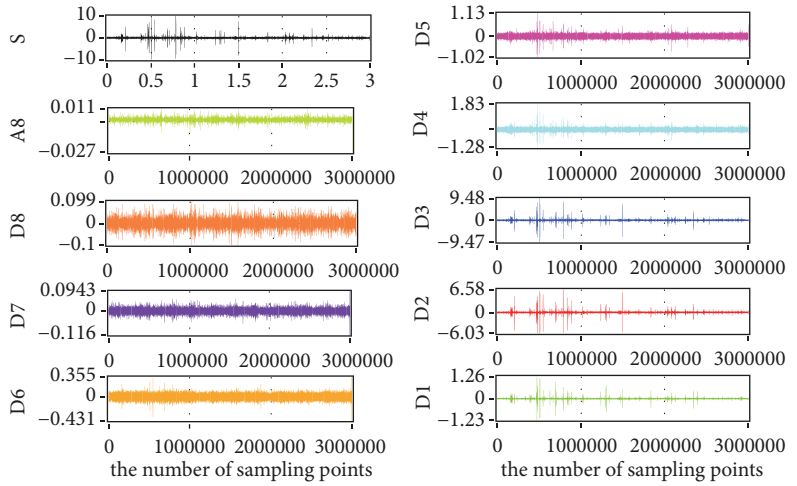



Figure 15: Energy proportion of the vibration signal in each frequency band after wavelet packet decompositions.

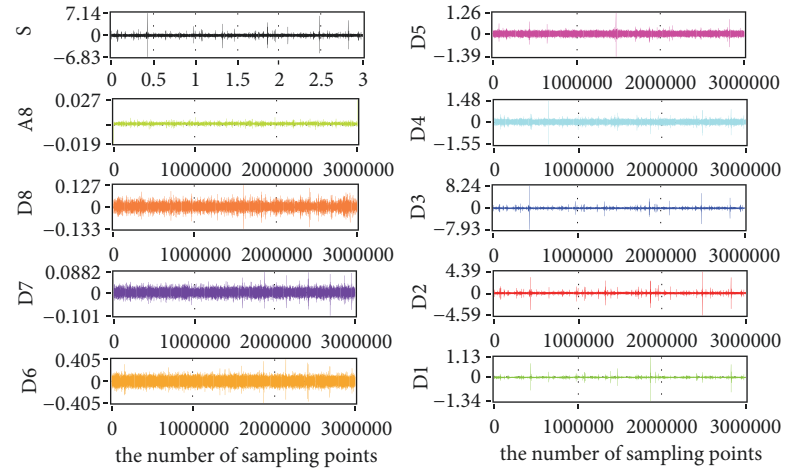
by normalization to obtain the energy proportion of each frequency band. Table 7 lists the energy proportion of AE signal for three different tool states under a given cutting condition (cutting speed 64.6m/min; feed rate 0.11mm/r; depth of cut 1.2mm).

The histogram for Table 7 is shown in Figure 17. It can be seen from Figure 17 that the energy of the AE signal is mainly distributed in several frequency bands of D2, D3, D4, D5, and D6. The energy proportions in D2 and D3 bands decrease with the increase of tool wear, and the energy proportions in the D5 and D6 frequency bands increase with the tool wear.

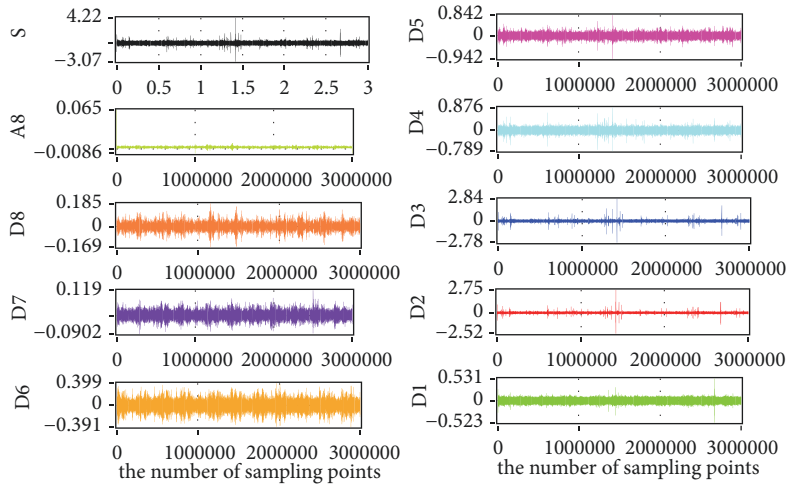
By now, after the signal analyses of 81 groups of experimental sample data in three tool states (initial wear, normal wear, and severe wear), the mean value, variance, and root mean square value for each group of the vibration signal data and the energy proportions of 16 frequency bands after 4 levels of wavelet packet decompositions are extracted. Meanwhile, the mean value, variance and root mean square




(a) Initial tool wear




(b) Normal tool wear




(c) Severe tool wear

Figure 16: Waveforms of AE signal by wavelet multiresolution decompositions (cutting speed 64.6m/min; feed speed 0.11mm/r; depth of cut 1.2mm).

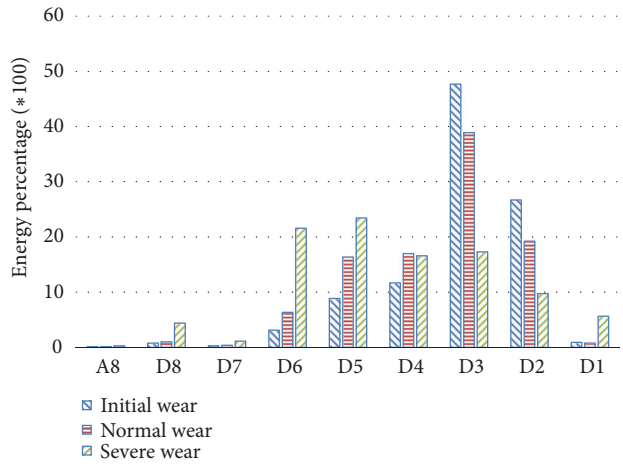



Figure 17: Energy proportion of AE signal after wavelet decompositions in each frequency band.

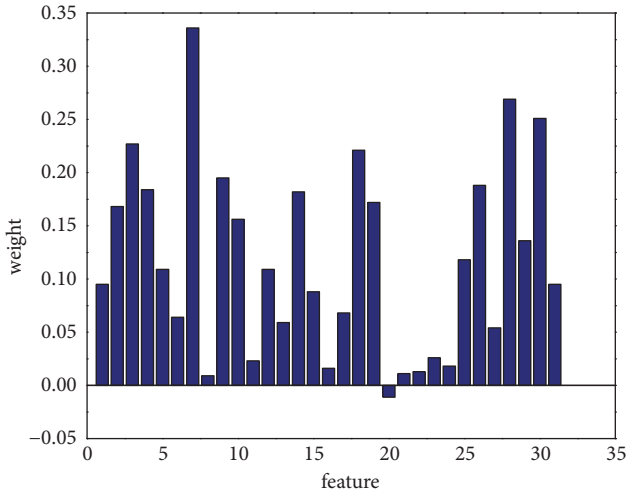



Figure 18: The average weight of the classification ability of each feature.

Table 8: A total of thirty-one extracted characteristics.

| *Extracted characteristics of the vibration signal* | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Feature | Mean value | Variance | Root mean square | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 |
| No. | 13 | 14 | 15 | 16 | 17 | 18 | 19 | | | | | |
| Feature | A10 | A11 | A12 | A13 | A14 | A15 | A16 | | | | | |
| *Extracted characteristics of AE signal* | | | | | | | | | | | | |
| No. | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 |
| Feature | Mean value | Variance | Root mean square | A8 | D8 | D7 | D6 | D5 | D4 | D3 | D2 | D1 |

value for each group of the AE signal data and the energy proportions of 9 frequency bands by the 8 levels of wavelet multiresolution decompositions are also extracted. Therefore, these characteristics for each group of vibration and AE signal data can compose a 31-dimensional fusion characteristic vector, as listed in Table 8.

In Table 8, some characteristics are useless, which not only increase the computation, but also affect the classification effect. Therefore, further feature selection is necessary.

*3.2.3. Feature Re-Extraction with Relief-F Algorithm.* Feature reextraction is to select a few features that are the most relevant with the tool state to compose the feature vectors by calculating each feature in a group of multidimensional features so that feature extraction and dimension reduction are achieved. In this work, a well-known filtering feature selection method called Relief-F algorithm is adopted. The principle of Relief-F algorithm is that the weight of each feature is calculated and then these weights are contrasted,

Table 9: Feature ranking of the classification ability.

| Sequence | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Feature No. | 7 | 28 | 30 | 3 | 18 | 9 | 26 | 14 | 19 | 2 | 10 |
| Weight | 0.33 | 0.26 | 0.24 | 0.22 | 0.21 | 0.19 | 0.18 | 0.17 | 0.16 | 0.16 | 0.15 |
| Sequence | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
| Feature No. | 29 | 25 | 5 | 4 | 12 | 1 | 31 | 15 | 17 | 6 | 13 |
| Weight | 0.13 | 0.11 | 0.10 | 0.10 | 0.10 | 0.09 | 0.09 | 0.08 | 0.06 | 0.06 | 0.05 |
| Sequence | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | | |
| Feature No. | 27 | 11 | 23 | 24 | 16 | 8 | 22 | 21 | 20 | | |
| Weight | 0.05 | 0.02 | 0.02 | 0.01 | 0.01 | 0.00 | 0.01 | 0.01 | -0.1 | | |

the greater the weight value is, the stronger the classification ability will be.

When dealing with multiclassification problems, Relief-F algorithm takes a sample R randomly from the training sample set each time and searches the nearest neighbor samples (near-hit) $H_j$ $(j = 1, 2, \ldots k)$ from the sample set of the same class as R and then finds out the nearest neighbor samples (near-miss) $M_j(C)$ from the sample set of different class from R and then updates the weight $W(A)$ of each feature A, as shown in

$$W(A) = W(A) - \sum_{j=1}^{k} \frac{diff(A, R, Hj)}{(mk)} + \sum_{C \in Class(R)} \frac{\left[\left(p(C) / \left(1 - p(Class(R))\right)\right) \sum_{j=1}^{k} diff\left(A, R, Mj(C)\right)\right]}{(mk)} \tag{11}$$

where diff (A, $R_1$, $R_2$) is the difference of sample $R_1$ and $R_2$ on feature A, $p$ (C) is the proportion of this class; $p$ (Class ($R$)) is the proportion of the class of a randomly selected sample; M$_j$($C$) is the $jth$ nearest neighbor sample in class C $\notin$class($R$); $m$ is the sampling number; and $k$ is the number of nearest samples.

$$diff(A, R_1, R_2) = \begin{cases} \dfrac{\left|R_1[A] - R_2[A]\right|}{[\max(A) - \min(A)]}, & if\ A\ is\ continuous \\ 0, & if\ A\ is\ discrete\ and\ R_1[A] = R_2[A] \\ 1, & if\ A\ is\ discrete\ and\ R_1[A] \neq R_2[A] \end{cases} \tag{12}$$

The weight of each feature is calculated with MATLAB based on Relief-F algorithm. In Table 8, the characteristics are numbered for the convenience of feature selection. We specify that the intentionally selected number of features is 8, and then the corresponding sequence number of the feature level after 8 will be removed. So the most optimal feature vectors for the classification of cutting tool states can be selected.

Figure 18 is the histogram for the averaged 30 calculation results of each feature weight. The distribution of the classification ability of each feature can be seen from the diagram, and the weight of each feature is arranged in order of value and Table 9 is obtained, which clearly reflects the correlation between each feature and the tool wear. The bigger the weight value is, the greater the correlation between them is.

From Figure 18 and Table 9, we can see that the numbers of the feature vectors are 7, 28, 30, 3, 18, 9, 26, and 14 after the Relief-F algorithm calculations. Therefore, checking Table 8, we know that the eight features including frequency bands A4, A6, A11, A15 and the root mean square of the vibration signal, and frequency bands D2, D4, and D6 of AE signal can be selected as the input of the subsequent state recognition.

*3.3. State Recognition Using BP Neural Network.* State recognition plays an important role in tool state monitoring. It is to classify the cutting tool state in the cutting process by inputting the extracted tool state features into the recognition model, enabling the mapping from the feature space to the state space. At present, the most commonly used state recognition methods in tool state monitoring are artificial neural network, fuzzy reasoning, etc.

In this work, the state recognition model is established by using BP neural network based on the analysis results of the signal processing and extraction with the hybrid programming of LabVIEW and MATLAB in Section 3.2 .

The structure of BP artificial neural network is divided into input layer, hidden layer, and output layer, as shown in Figure 19. In the structure, the network is trained by the training sample set of the input. During the training process, the error is constantly propagated and corrected, and the

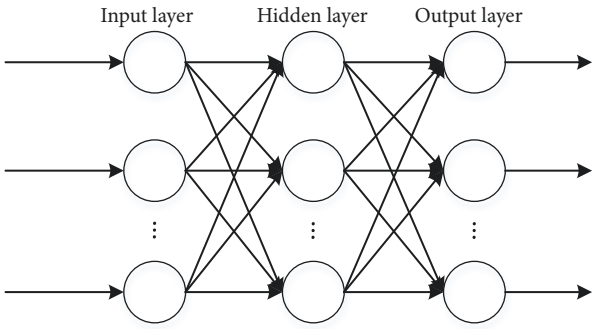



Figure 19: BP neural network structure.

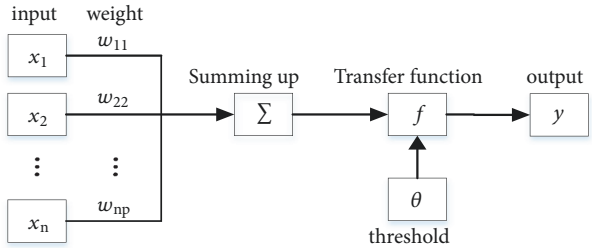



Figure 20: Artificial neuron model.

network parameters, such as the weight and the threshold value (see Figure 20), are also adjusted to enable the network to approach the desired mapping relationship between the input and output.

*3.3.1. The Establishment of Adaptive BP Neural Network Model.* The establishment of adaptive BP neural network model mainly includes the determination of the number of nodes in the input, hidden and output layer, selection of the transfer function, the training algorithm, and the setting of the target error.

*(1) The Number of Input Layer Nodes*. The number of input layer nodes of the BP neural network is equal to the dimension of the input vector. Because 8 features are retained after the feature selection, so the number of the input layer nodes is 8.

*(2) The Number of Output Layer Nodes*. Since there are three kinds of tool wear: initial wear, normal wear, and severe wear, which are represented by binary numbers [1 0 0], [0 1 0], and [0 0 1], respectively, so the number of the output layer nodes is 3.

*(3) The Number of Hidden Layer Nodes*. The number of hidden layer nodes determines the performance of the network model. If the number is too small, then the division of the model space is rough, and the network fault tolerance performance and recognition ability are low; if the number is too large, then the model space division is too fine, and the network convergence speed is slow or not convergent, and the training time is long. Based on the empirical formula (13), preliminary calculations are carried out, and the number of hidden layer nodes is changed and tested repeatedly. Finally, the number of hidden layer nodes is determined to be 11.

$$p = \sqrt{n+m} + a \tag{13}$$

where $p$ is the number of hidden layer nodes; n is the number of input nodes; m is the number of output nodes; and a∈(1, 10).

Thus an adaptive BP neural network model of 8-11-3 structure is obtained, and Figure 20 shows the artificial neuron model. The S type tangent function (Tansig) is selected for the transfer function of the hidden layer neurons, and the S type logarithmic function (logsig) is selected for the output layer neuron transfer function. Levenberg-Marquart's improved algorithm is used as the learning algorithm. The maximum number of training steps is 1000, the target error is 1e-6, and the performance parameter is 0.001.

*3.3.2. Determination of Samples.* After the signal analysis and features selection in Section 3.2, *8 dimensional-81 groups* of *feature-sample* set are formed. This work selects 54 groups of samples as a training set and 27 groups of samples as a test set.

The range of numerical values of different types of features in the samples is inconsistent, so we need to normalize the sample set with MATLAB and to limit the numerical values of different types of features in the range of (-1, 1) according to

$$y = \frac{(y_{\max} - y_{\min}) \times (x - x_{\min})}{x_{\max} - x_{\min}} + y_{\min} \tag{14}$$

where $y_{max}$=1, $y_{min}$= -1; $x$ and $y$ are feature values to be normalized and the corresponding value after normalization, respectively; $x_{max}$ and $x_{min}$ are the maximum and minimum values of the features to be normalized, respectively.

*3.3.3. Training Results of Adaptive BP Neural Networks.* Fifty-four groups of sample data in training set are input into adaptive BP neural network model for training and learning. Figure 21 shows the training state diagram of the BP neural network. It can be seen from Figure 21 that the model reaches the convergence precision at step 49.

## 4. Results Analysis

Twenty-seven groups of test sample data are input into the trained adaptive network model for testing, and the obtained test results are presented in Table 10. From Table 10, we can see that the recognition results of the network model contain two incorrect outputs; hence the correct recognition rate is u=25/27=92.59%. This indicates that the state recognition model established by adaptive BP algorithm can better accomplish the tool state recognition. If the first number of the actual output value of adaptive BP neural network is maximum, then the output shows [1 0 0], indicating that the tool is in the initial wear state; if the second number is

Table 10: Recognition results of cutting tool state.

| No. | Tool wear state | | Output | | Recognition results | Actual results | Result judgment |
|---|---|---|---|---|---|---|---|
| 1 | Initial wear | 0.9999 | 0.5000 | 0.5000 | 1 0 0 | 1 0 0 | Correct |
| 2 | Normal wear | 0.5000 | 0.9999 | 0.5000 | 0 1 0 | 0 1 0 | Correct |
| 3 | Severe wear | 0.5000 | 0.5000 | 0.9998 | 0 0 1 | 0 0 1 | Correct |
| 4 | Initial wear | 0.9999 | 0.5000 | 0.5000 | 1 0 0 | 1 0 0 | Correct |
| 5 | Normal wear | 0.5000 | 0.9999 | 0.5000 | 0 1 0 | 0 1 0 | Correct |
| 6 | Severe wear | 0.5000 | 0.5000 | 0.9999 | 0 0 1 | 0 0 1 | Correct |
| 7 | Initial wear | 0.9999 | 0.5000 | 0.5000 | 1 0 0 | 1 0 0 | Correct |
| 8 | Normal wear | 0.5000 | 0.9991 | 0.5000 | 0 1 0 | 0 1 0 | Correct |
| 9 | Severe wear | 0.5000 | 0.5000 | 0.9999 | 0 0 1 | 0 0 1 | Correct |
| 10 | Initial wear | 0.9998 | 0.5000 | 0.5000 | 1 0 0 | 1 0 0 | Correct |
| 11 | Normal wear | 0.5000 | 0.9999 | 0.5000 | 0 1 0 | 0 1 0 | Correct |
| 12 | Severe wear | 0.5000 | 0.5000 | 0.9994 | 0 0 1 | 0 0 1 | Correct |
| 13 | Initial wear | 0.8765 | 0.5000 | 0.5000 | 1 0 0 | 1 0 0 | Correct |
| 14 | Normal wear | 0.5000 | 0.9999 | 0.5000 | 0 1 0 | 0 1 0 | Correct |
| 15 | Severe wear | 0.5000 | 0.5000 | 0.9999 | 0 0 1 | 0 0 1 | Correct |
| 16 | Initial wear | 0.5003 | 0.7265 | 0.5000 | 0 1 0 | 1 0 0 | Incorrect |
| 17 | Normal wear | 0.5000 | 0.9615 | 0.5000 | 0 1 0 | 0 1 0 | Correct |
| 18 | Severe wear | 0.5000 | 0.5000 | 0.9998 | 0 0 1 | 0 0 1 | Correct |
| 19 | Initial wear | 0.9999 | 0.5000 | 0.5000 | 1 0 0 | 1 0 0 | Correct |
| 20 | Normal wear | 0.5000 | 0.9999 | 0.5000 | 0 1 0 | 0 1 0 | Correct |
| 21 | Severe wear | 0.5000 | 0.5000 | 0.9999 | 0 0 1 | 0 0 1 | Correct |
| 22 | Initial wear | 0.5003 | 0.6465 | 0.5000 | 0 1 0 | 1 0 0 | Incorrect |
| 23 | Normal wear | 0.5000 | 0.9999 | 0.5000 | 0 1 0 | 0 1 0 | Correct |
| 24 | Severe wear | 0.5000 | 0.5000 | 0.9999 | 0 0 1 | 0 0 1 | Correct |
| 25 | Initial wear | 0.9999 | 0.5000 | 0.5000 | 1 0 0 | 1 0 0 | Correct |
| 26 | Normal wear | 0.5000 | 0.9615 | 0.5000 | 0 1 0 | 0 1 0 | Correct |
| 27 | Severe wear | 0.5000 | 0.5000 | 0.9998 | 0 0 1 | 0 0 1 | Correct |

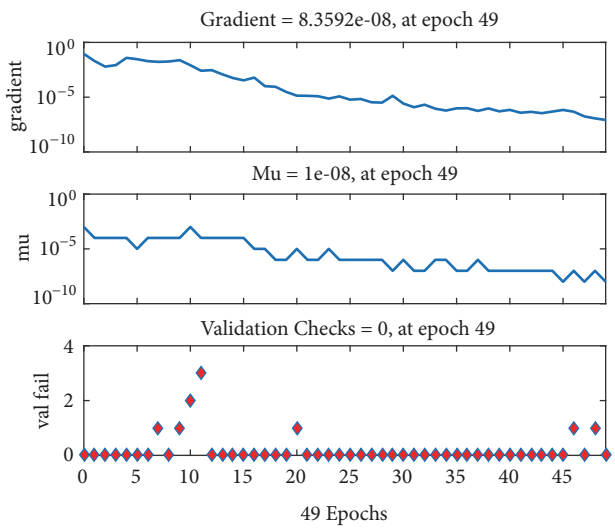



Figure 21: Training state diagram of BP neural network.

maximum, the output is [0 1 0] and the tool is in the normal wear state; and if the third number is maximum, the output is [0 0 1] and the tool is in severe wear state.

## 5. Discussion

This work systematically investigates the approaches of realizing a visual, intelligent, and real-time tool state monitoring based on hybrid programming of LabVIEW and MATLAB. Both the multitheories used in the study and the experimental method can adapt to various machining processes and environments. The reasons are discussed as follows.

*(1) Advantages of Experimental Scheme.* The experimental objective is to obtain the mapping relations between the tool wear and the measured physical signals that are able to reflect the tool state in cutting processes. Therefore, the determination of experimental factors is one of the key issues. For this, the experimental scheme is designed according to the specifications of the machine tools used in this work. Except some fixed factors affecting the signals such as the machining environment and cutting tool parameters, other variable factors such as cutting parameters including the cutting speed, the feed rate, the depth of cut, and tool wear are determined, among which three VB values of the tool flank wear corresponding to the indices of general tool wear criterions, sharp tool stage (initial wear), normal working stage (normal wear), and worn out stage (severe wear), are taken into account, as presented in Tables 2 and 3. Based on

this scheme, *all-factors* experiments are conducted, which is a very important aspect for obtaining the sufficient data of the later task on modeling and recognition of the neural network. Hence, there are four factors and each factor has three levels thereby $3^4$=81 groups of experiments.

Secondly, tool state monitoring involves a variety of sensors and signal analysis techniques, and the selection of the suitable sensors is the second key issue. Due to the nature of manufacturing processes, the signals are usually nonstationary, which often contain both high and low frequency components. Therefore, bisensors, i.e., vibration acceleration sensor and AE sensor, are simultaneously selected for measuring the low frequency vibration signal and higher frequency AE signal in this work. Therefore, the advantages of bisensors lie in that they can sensitively and entirely reflect the tool state in real time and are suitable for monitoring different machining processes.

*(2) Significance of Dual Feature Extraction of Signals*. This work extracts the signal features that are most relevant with the tool state by the use of the method fusion of wavelet theory and Relief-F algorithm. The reason is that dual feature extraction of signals can achieve the dual effects of finding out the most relevant features and feature dimension reduction. Wavelet analysis has good localization property in both time domain and frequency domain and permits adaptive time-frequency representation. Dividing the frequency band into multiple levels can improve frequency resolution of both high and low frequencies. Through the signal analyses of 81 groups of experimental sample data in three tool states (initial wear, normal wear, and severe wear), 4 levels of wavelet packet decompositions of the vibration signal and the 8 levels of wavelet multiresolution decompositions of the AE signals can satisfy the requirement of the frequency resolution and 31 characteristics for each group of signal data are extracted, as listed in Table 8. However, some characteristics are useless in Table 8, which not only increase the computation, but also affect the state recognition effect. Therefore, feature reextraction using Relief-F algorithm is carried out to select a few features that are the most correlated with the tool state. Figure 18 shows the histogram for the averaged 30 calculation results of each feature weight. From Figure 18 and Table 9, we can see that the sequence numbers of the feature vectors are 7, 28, 30, 3, 18, 9, 26, and 14 after the Relief-F algorithm calculations. Therefore, by the inspection of Table 8, they are 8 corresponding features including frequency bands A4, A6, A11, A15, the root mean square of the vibration signal, and frequency bands D2, D4, and D6 of AE signal. That is, the number of features is reduced from 31 to 8.

*(3) Success Rate of State Recognition Using BP Neural Network*. In this investigation, it is found that the number of hidden layer nodes is one of BP network training conditions determining the success of network recognition. The number of hidden nodes is changed and tested repeatedly after the initial calculation using the empirical formula (13). The network is trained by the training sample set of the input. During the training process, the network parameters, such as the weight and the threshold value, also need to be adjusted to enable the network to approach the desired mapping relationship between the input and output. This work selects 54 groups of samples as a training set and 27 groups of samples as a test set. It can be seen from Figure 21 that the adaptive BP neural network model for training and learning reaches the convergence precision at step 49. Test results show that the correct recognition rate can reach 92.59%.

## 6. Conclusions

A tool wear monitoring through hybrid programming technique of LabVIEW and MATLAB is accomplished, which can more accurately complete the intelligent and real-time recognition of the tool state with the correct recognition rate of 92.59%.

Vibration and AE signals are ideal signals for tool monitoring. Bisensor fusion can collect both the relatively low frequency of the vibration signal and the high frequency of AE signal and thus can sensitively and entirely reflect the tool state in real time.

The method fusion of wavelet decompositions and relief-F algorithm can achieve dual extraction of signal features and feature dimension reduction, such that computation is fast and real-time performance is good. The results show that the number of features that are relevant with the tool wear state is reduced from 31 of each group of signal data to 8, which are of the frequency bands most relevant to tool wear states including A4 (2.4-3.2 KHz), A6 (4-4.8 KHz), A11 (8-8.8 KHz), A15 (11.2-12 KHz) of vibration signal and D2 (125-250 KHz), D4 (31.25-62.5 KHz), and D6 (7.8-15.6 KHz) of AE signal. Obviously vibration signal and AE signal can complement each other in tool wear monitoring.

The tool wear monitoring method shows the generality. This is because the experimental scheme, hybrid programming technique, the dual feature extraction, and BP network recognition model in this work have generality, which can be applied to any other different machining operations such as milling, drilling, boring, forming, and shaping, combined with different cutting tools including single point tool (shaping) and multipoint tool (milling, drilling) in machining any kinds of workpiece materials. This can be accomplished only through changing the installation locations of the vibration and AE sensors. If necessary, the sampling parameters can easily be set in the system and can reextract the different features by implementing the Relief-F algorithm and can also reset the parameters of the BP neural network.

## Data Availability

The data used to support the findings of this study are included in the article.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 61562055).

## References

[1] L. Dan and J. Mathew, "Tool wear and failure monitoring techniques for turning-A review," *The International Journal of Machine Tools and Manufacture*, vol. 30, no. 4, pp. 579–598, 1990.

[2] P. Waydande, N. Ambhore, and S. Chinchanikar, "A review on tool wear monitoring system," *Journal of Mechanical Engineering and Automation*, vol. 6, no. 5A, pp. 49–53, 2016.

[3] K. P. Zhu, Y. S. Wong, and G. S. Hong, "Wavelet analysis of sensor signals for tool condition monitoring: a review and some new results," *The International Journal of Machine Tools and Manufacture*, vol. 49, no. 7-8, pp. 537–553, 2009.

[4] D. E. Dimla Sr. and P. M. Lister, "On-line metal cutting tool condition monitoring. I: force and vibration analyses," *The International Journal of Machine Tools and Manufacture*, vol. 40, no. 5, pp. 739–768, 2000.

[5] A. Ghasempoor, J. Jeswiet, and T. N. Moore, "Real time implementation of on-line tool condition monitoring in turning," *The International Journal of Machine Tools and Manufacture*, vol. 39, no. 12, pp. 1883–1902, 1999.

[6] S. K. Sikdar and M. Chen, "Relationship between tool flank wear area and component forces in single point turning," *Journal of Materials Processing Technology*, vol. 128, no. 1-3, pp. 210–215, 2002.

[7] M. Nouri, B. K. Fussell, B. L. Ziniti, and E. Linder, "Real-time tool wear monitoring in milling using a cutting condition independent method," *The International Journal of Machine Tools and Manufacture*, vol. 89, pp. 1–13, 2015.

[8] M. Wang and J. Wang, "CHMM for tool condition monitoring and remaining useful life prediction," *The International Journal of Advanced Manufacturing Technology*, vol. 59, no. 5-8, pp. 463–471, 2012.

[9] X. Li, H.-X. Li, X.-P. Guan, and R. Du, "Fuzzy estimation of feed-cutting force from current measurement-a case study on intelligent tool wear condition monitoring," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 34, no. 4, pp. 506–512, 2004.

[10] R. Teti, K. Jemielniak, G. O'Donnell, and D. Dornfeld, "Advanced monitoring of machining operations," *CIRP Annals - Manufacturing Technology*, vol. 59, no. 2, pp. 717–739, 2010.

[11] O. B. Abouelatta and J. Mádl, "Surface roughness prediction based on cutting parameters and tool vibrations in turning operations," *Journal of Materials Processing Technology*, vol. 118, no. 1-3, pp. 269–277, 2001.

[12] F. J. Alonso and D. R. Salgado, "Analysis of the structure of vibration signals for tool wear detection," *Mechanical Systems and Signal Processing*, vol. 22, no. 3, pp. 735–748, 2008.

[13] X. Li, "A brief review: acoustic emission method for tool wear monitoring during turning," *The International Journal of Machine Tools and Manufacture*, vol. 42, no. 2, pp. 157–165, 2002.

[14] S. Kakade, L. Vijayaraghavan, and R. Krishnamurthy, "In-process tool wear and chip-form monitoring in face milling operation using acoustic emission," *Journal of Materials Processing Technology*, vol. 44, no. 3-4, pp. 207–214, 1994.

[15] I. Korkut, A. Acir, and M. Boy, "Application of regression and artificial neural network analysis in modelling of tool-chip interface temperature in machining," *Expert Systems with Applications*, vol. 38, no. 9, pp. 11651–11656, 2011.

[16] T. Özel and Y. Karpat, "Predictive modeling of surface roughness and tool wear in hard turning using regression and neural networks," *The International Journal of Machine Tools and Manufacture*, vol. 45, no. 4-5, pp. 467–479, 2005.

[17] M. Balazinski, E. Czogala, K. Jemielniak, and J. Leski, "Tool condition monitoring using artificial intelligence methods," *Engineering Applications of Artificial Intelligence*, vol. 15, no. 1, pp. 73–80, 2002.

[18] R. G. Silva, R. L. Reuben, K. J. Baker, and S. J. Wilcox, "Tool wear monitoring of turning operations by neural network and expert system classification of a feature set generated from multiple sensors," *Mechanical Systems and Signal Processing*, vol. 12, no. 2, pp. 319–332, 1998.

[19] E. Kannatey-Asibu, J. Yum, and T. H. Kim, "Monitoring tool wear using classifier fusion," *Mechanical Systems and Signal Processing*, vol. 85, pp. 651–661, 2017.

[20] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp. 1483–1510, 2006.

[21] W. Gong, T. Obikawa, and T. Shirakashi, "Monitoring of tool wear states in turning based on wavelet analysis," *JSME International Journal Series C Mechanical Systems, Machine Elements and Manufacturing*, vol. 40, no. 3, pp. 447–453, 1997.

[22] M. C. Yoon and D. H. Chin, "Cutting force monitoring in the endmilling operation for chatter detection," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, vol. 219, no. 6, pp. 455–465, 2005.

[23] S. G. Mallat, "Theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[24] S. G. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, Newyork, NY, USA, 2nd edition, 1999.

[25] L. L. Zhang and J. Xiao, *Case Book of Mechanical Fault Diagnosis Technology Based on MATLAB*, Higher Education Press, Beijing, China, 1st edition, 2016.

Hindawi
Complexity
Volume 2019, Article ID 6301326, 11 pages
https://doi.org/10.1155/2019/6301326

*Research Article*

# Solving the Complexity Problem in the Electronics Production Process by Reducing the Sensitivity of Transmission Line Characteristics to Their Parameter Variations

**Talgat R. Gazizov, Indira Ye. Sagiyeva, and Sergey P. Kuksenko**

*Department of Television and Control, Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russia*

Correspondence should be addressed to Talgat R. Gazizov; talgat@tu.tusur.ru

Received 1 March 2019; Revised 24 April 2019; Accepted 5 May 2019; Published 3 June 2019

Guest Editor: Julio Blanco-Fernández



In this paper we consider the complexity problem in electronics production process. Particularly, we investigate the ways to reduce sensitivity of transmission line characteristics to their parameter variations. The reduction is shown for the per-unit-length delay and characteristic impedance of several modifications of microstrip transmission lines. It can be obtained by means of making an optimal choice of parameter values, enabling proper electric field redistribution in the air and the substrate. To achieve this aim we used an effective simulation technique and software tools. Taken together, for the first time, they have allowed formulating general approach which is relevant to solve a wide range of similar tasks.

## 1. Introduction

Electronics is more and more widely used in human life. Complexity of electronics is continuously increasing. It results in increasing the number of electronics parameters. Unfortunately, the parameters undergo undesirable variations caused by the manufacturing process. As undesirable results, we obtain the reduced yield ratio in mass production and the reduced quality of a product or the increased production cycle due to necessity of a product to be redesigned or reproduced.

To treat the problem we can consider two main parts of the problem: concerning electronic components and supporting structures for their packaging. As for components, their nomenclature is very wide. Nevertheless, a great attention is paid to ensure their tolerance. As for supporting structures, their number is very small (a printed circuit board (PCB), an integrated circuit (IC), and a chip). The most popular structure is a PCB, while an IC and a chip can be considered as dedicated to electronic components, especially due to the recent trends of system-in-package (SiP) and system-on-chip (SoC) designs. Thus, a PCB is becoming the main concern of a designer, while a microstrip-like-lines are the main paths for signal propagation in PCBs and even ICs and chips as well. Therefore, to assure stable characteristics of the lines are an important task, the importance is illustrated by a representative example in Figure 1 [1]. One can see that a PCB line impedance value, for example, calculated as 55 Ω can really be 45 Ω or 65 Ω. Thus, it is important to seek new ways to minimize sensitivity of critical characteristics of PCB transmission lines. Besides, proper tools for multiple, but quick and accurate, calculations of the characteristics are necessary.

With increasing requirements to electronics characteristics it is necessary to have transmission lines with more stable per-unit-length delay ($\tau$) and characteristic impedance ($Z$). In turn, it leads to the necessity of detailed simulation and investigation of these characteristics.

One of the most widely used high-speed signal transmission lines is a microstrip line (MSL) [2]. The sensitivity of MSL characteristics has been a subject of extensive considerations since the appearance of MSLs and is still a matter of great importance. Particularly, variations of strip thickness and width as the most appropriate parameters to be controlled by a designer in practice have been considered in [3] and are being considered in [4]. (As opposed to sensitivity

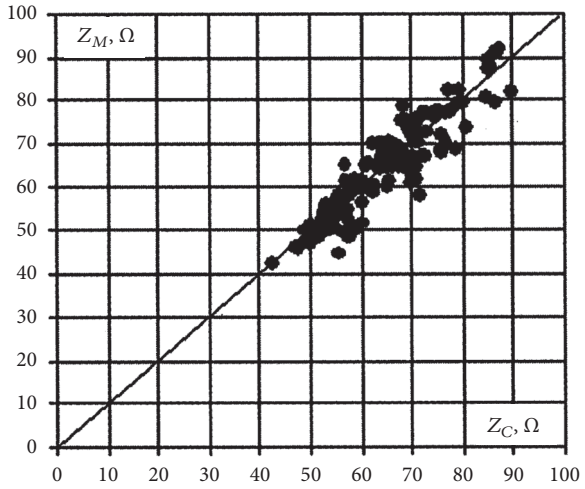



Figure 1: Comparison of measured ($Z_M$) and calculated ($Z_C$) impedance values for 186 test line samples of various cross-sections [1].

minimization, considerable modifications have also been proposed in order to increase the defined characteristics. For example, a slot in the ground plane increases the MSL characteristic impedance and can be properly used in a filter [5]. To accelerate the process of simulating such MSL characteristics, the analytic expressions have been proposed [6].)

However the possibilities to minimize the sensitivity of MSL characteristics are limited by the simplicity of the classical MSL construction. Therefore, various modifications of an MSL, such as suspended and inverted microstrip lines, allowing zero sensitivity of $\tau$ and $Z$, are considered. For example, a detailed study of sensitivity for a particular case of an air gap between a substrate and a ground plane has been considered [7]. A more general case of a nonair dielectric layer between a substrate and a ground plane has also shown possibilities to minimize the sensitivity to changes in the thickness of dielectric layers [8]. In the multilayer PCBs, the varieties of MSLs, for example, MSLs with polygons on different layers which allow obtaining a stable value of the $\tau$, are used [9]. Similar possibilities are revealed in MSL with side grounded conductors located over the substrate [10] and buried in the substrate [11]. The possibility of minimizing the sensitivity arises from electric field redistribution in the layers of the air and the substrate. Similar phenomena occur in the case of placing an additional grounded conductor over a usual MSL and in the case of a totally shielded MSL [12].

To obtain necessary characteristics of a structure, it is required to carry out a multivariate analysis in the range of parameter variations. However, for a structure of an arbitrary cross-section, the quick and accurate expressions are not available. Therefore, it is necessary to use numerical methods, wherein the main costs fall on a multiple linear algebraic systems solution. In this case, depending on the type of the problem, only the matrix or the matrix and the right-hand side can be varied. In the matrix case we can write

$$\mathbf{A}_k \mathbf{x}_k = \mathbf{b}_k, \quad k = 1, 2, \ldots, m, \tag{1}$$

where $\mathbf{A}_k$ are nonsingular matrices, $\mathbf{b}_k$ are corresponding right-hand sides, $k$ is a sequence number of linear system, and $m$ is the number of linear systems to be solved. To solve sequence (1), iterative methods with preconditioning are often used. However, calculating the preconditioner and using it in solving sequence (1) are significantly different, compared to solving one linear system.

The first approach to solving sequence (1) is recomputing a preconditioner for each matrix of this sequence from scratch [13]. However, computational costs increase significantly, which does not make it effective. The second approach is based on the use of a frozen preconditioner, calculated from the first matrix of sequence (1) and used to solve subsequent linear systems. This approach has found wide application in solving nonlinear equation problems [14–16]. Obviously, this approach has less computational complexity, but in practice there are often situations when the solution of the current linear system cannot be obtained due to the stagnation of the iterative process. First of all, this is due to the fact that matrix-to-matrix changes of sequence (1) are significant and the application of a frozen preconditioner becomes ineffective. Therefore, the process of updating the preconditioner is widespread. The remaining approaches contain features of the two described. The third approach lies in updating the preconditioner obtained from the matrix of one of the systems (seed preconditioner), and repeated when necessary. This approach has proven itself in solving the sequences of shifted systems which emerge in problems of numerical optimization and solutions of nonlinear equations performed by the methods of Newtonian type [17–19]. Papers [20–22] were devoted to developing the approaches for updating incomplete $\mathrm{LDL}^T$-decomposition of matrix $\mathbf{A}$, which is used to form a preconditioner. Then, these approaches were generalized on the case of nonsymmetric matrices with updating the preconditioner periodically or before solving the current system. The last approach is based on the adaptive use of information about the Krylov subspaces obtained on the previous steps (recycling of Krylov subspaces) [23–25]. Therefore, there exist appropriate algorithms which can be used (*per se* or after some additional modifications) to perform accelerated simulations of transmission line structures.

Thus, it is useful to analyze the results obtained as well as the techniques and the tools used for these studies. The aim of this paper is to propose a general approach to solving the complexity problem in electronics production process based on the summary of the recent studies devoted to minimizing the sensitivity of microstrip-like lines characteristics to variations of their parameters.

## 2. Structures and Approach for Investigation

Various modifications of the MSL under study are presented in Figures 2–4.

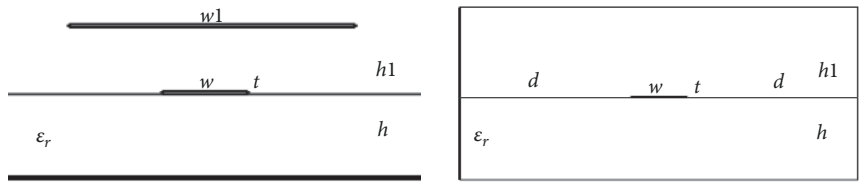



Figure 2: Cross-sections of covered (a) and shielded (b) MSLs.

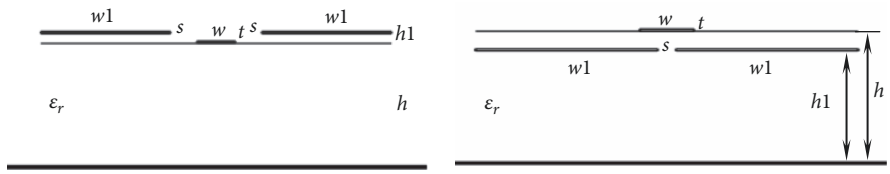



Figure 3: Cross-sections of MSLs with side grounded conductors in air (a) and dipped in substrate (b).

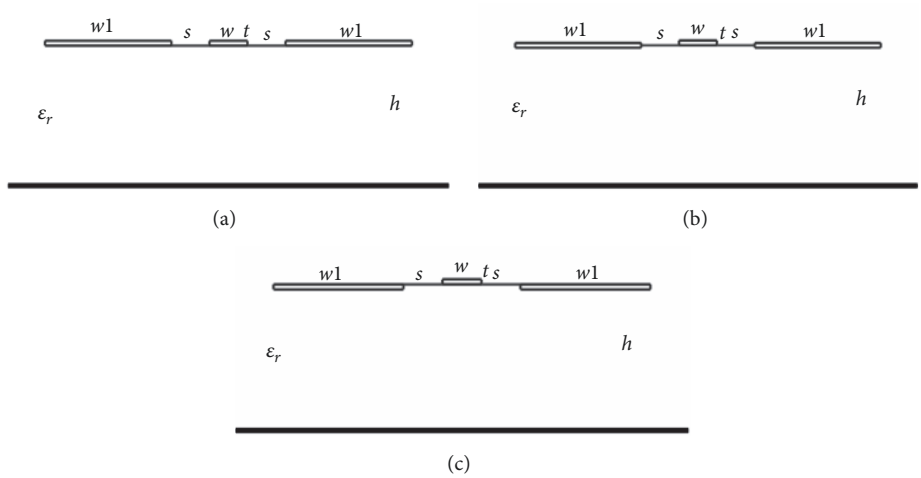



Figure 4: Cross-sections of MSLs with side grounded conductors above (a), among (b), and under (c) air–substrate boundary.

The values of the characteristics ($\tau$ and $Z$) of the investigated lines from Figures 2–4 were calculated by the known formulas:

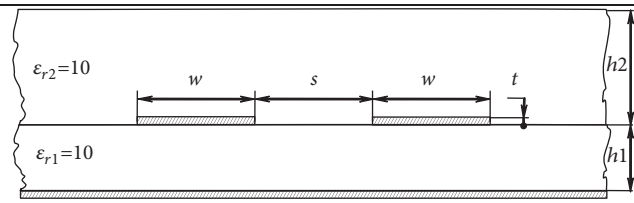
$$\tau = \frac{\left(C/C_0\right)^{0.5}}{v_0}, \qquad Z = \frac{1}{\left(v_0\left(CC_0\right)^{0.5}\right)} \tag{2}$$

where $v_0$ is the speed of light in a vacuum; $C$ and $C_0$ are signal strip diagonal entries of per-unit-length coefficient matrix of electrostatic induction in the real dielectric filling and in the air.

The investigation of a microstrip structure of various characteristics, especially in the first stage should be performed through simulation, as it is less costly and may be more accurate than measurements. In this regard, in this work, the construction of cross-sections and calculations are performed in the TALGAT system available to the authors [28]. Strict full-wave analysis of fields in the investigated lines is rather complicated. This is due to the inhomogeneity of the dielectric medium over the section of the line. As a result, part of the field is concentrated in the dielectric substrate and the rest in the air. Therefore, not pure TEM-mode but quasi-TEM-mode propagates in the lines.

Before using any software, one must validate it properly by using simulation and measurements. However, in case of coupled transmission lines, the off-diagonal coefficients of a per-unit-length matrix have large (up to 25%) error [29]. Therefore, one can use computations by several numerical methods. Indicative examples of such approach are shown in Table 1 [29]. It demonstrates the mentioned large measurement error and the close results of numerical methods. The method of moments (MoM) is the most widely used and tested method among them. Thus, we used the MoM implemented in the TALGAT system. In order to test it, we

Table 1: Results of simulation by Green's Function Method (GFM), Method of Moments (MoM) and Variational Method (VM) and measurement (pF/cm).




| Size (mills) for $h1$=10, $h2$=20, $t$=0.5 | Results | $C_{11}$ | $-C_{12}$ |
|---|---|---|---|
| $w$=40, $s$=10 | GFM | 5.61 | 0.77 |
| | MoM | 5.62 | 0.76 |
| | VM | 5.64 | 0.68 |
| | Measured | 5.59±0.06 | 0.62±0.15 |
| $w$=20, $s$=10 | GFM | 3.78 | 0.70 |
| | MoM | 3.78 | 0.70 |
| | VM | 3.78 | 0.63 |
| | Measured | 3.69±0.04 | 0.38±0.10 |
| $w$=10, $s$=20 | GFM | 2.66 | 0.29 |
| | MoM | 2.65 | 0.30 |
| | VM | 2.67 | 0.24 |
| | Measured | 2.64±0.03 | 0.20±0.05 |
| $w$=10, $s$=10 | GFM | 2.77 | 0.59 |
| | MoM | 2.76 | 0.60 |
| | VM | 2.77 | 0.53 |
| | Measured | 2.75±0.30 | 0.48±0.12 |
| $W$=10, $s$=5 | GFM | 3.00 | 0.97 |
| | MoM | 2.99 | 0.97 |
| | VM | 2.96 | 0.94 |
| | Measured | 2.95±0.03 | 0.92±0.23 |

computed the per-unit-length matrices of various structures, for which there exist published original and obtained results.

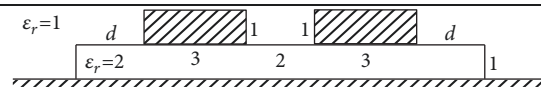
Firstly, we compared results for a simple case of 2 coupled strips on a substrate [26] shown in Table 2. However, we consider the cases of side dielectric walls becoming closer to side strip edges and the similar cases without side dielectric walls (Table 2). Maximum errors are 1.4% for diagonal and 8.1% for off-diagonal values. The coincidence is satisfactory.

Then, we considered the same 3 strips of various positions in a two-layer dielectric medium [27] (Table 3). Comparison of our results and the results obtained with the use of integral equation method showed maximum error 8.8% for **C** and 0.8% for **L** entries. The coincidence is also satisfactory for this more complex structure.

Thus, the performed comparisons showed satisfactory coincidence of the results and the relevance of the TALGAT system for computing per-unit-length matrices for structures of various complexities. Meanwhile, for final testing, one can compare a time response of a structure. There exist indicative and commonly available examples comparing the TALGAT system results with the measurement [30, 31] and electromagnetic analysis [32] results being omitted here.

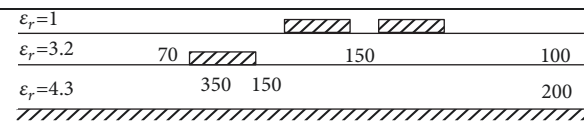
We are using several approaches to reduce the computational complexity of the analysis. The main feature of these approaches is the use of iterative methods. So, when solving the first linear system, we calculated a proper preconditioner. This preconditioner (frozen) is used to solve subsequent linear systems. At the same time, the use of the previous system solution as the initial guess is more preferable than the fixed initial guess [33]. However, as the differences between the matrices increase, the preconditioner efficiency decreases. Therefore, updating or recomputing the preconditioner is needed. We use the second approach. For this purpose we developed methods of adaptive recomputing preconditioner based on the threshold of the number of iterations [33], the average arithmetic complexity [34], and the average solution time [35]. Recomputation of the preconditioner is not necessary if the seed matrix for calculating the preconditioner has been selected properly. In this case, the choice of a seed matrix from the middle of the sequence linear systems allows us to accelerate the overall solution without recomputing the preconditioner. Another very simple way to accelerate the solution is to choose the solution order. So, it was shown that the reverse (from the highest value of the parameter being varied to the smallest) order comparing to the direct (from the lowest value of the parameter being varied to the largest) order is more preferable [36].

Table 2: Comparison of computed entries of matrix **C**.




| $d$ | Results | $C_{11}$ | | $-C_{12}$ | |
|---|---|---|---|---|---|
| | | Without walls | With walls | Without walls | With walls |
| 6.0 | Result [26], pF/m | 92.36 | 92.05 | 8.494 | 8.473 |
| | Our result, pF/m | 91.11 | 91.11 | 9.162 | 9.162 |
| | Error, % | –1.3 | –1.0 | 7.8 | 8.1 |
| 4.0 | Result [26], pF/m | 92.44 | 92.14 | 8.506 | 8.485 |
| | Our result, pF/m | 91.15 | 91.14 | 9.167 | 9.168 |
| | Error, % | –1.4 | –1.1 | 7.7 | 8.0 |
| 2.0 | Result [26], pF/m | 92.40 | 92.10 | 8.539 | 8.517 |
| | Our result, pF/m | 91.15 | 91.08 | 9.179 | 9.182 |
| | Error, % | –1.3 | –1.1 | 7.5 | 7.8 |
| 0.5 | Result [26], pF/m | 91.44 | 90.50 | 8.595 | 8.565 |
| | Our result, pF/m | 90.73 | 89.88 | 9.185 | 9.184 |
| | Error, % | –0.7 | –0.7 | 6.8 | 7.2 |
| 0 | Result [26], pF/m | 89.68 | 87.97 | 8.603 | 8.569 |
| | Our result, pF/m | 89.34 | 87.68 | 9.146 | 9.146 |
| | Error, % | –0.3 | 0.3 | 6.2 | 6.7 |

Table 3: Comparison of computed **C** (pF/m) and **L** (nH/m) entries for 3 strips of various positions in a two-layer dielectric medium.




| Results | $C_{11}$ | $-C_{21}$ | $-C_{31}$ | $C_{22}$ | $-C_{32}$ | $C_{33}$ |
|---|---|---|---|---|---|---|
| Result [27] | 142.1 | 21.7 | 0.9 | 93.5 | 18.1 | 88.0 |
| Our result | 143.6 | 19.8 | 0.9 | 88.6 | 17.7 | 83.1 |
| Error, % | 1.1 | –8.8 | 0 | –5.2 | –2.2 | –5.6 |
| Results | $L_{11}$ | $L_{21}$ | $L_{31}$ | $L_{22}$ | $L_{32}$ | $L_{33}$ |
| Result [27] | 277.7 | 87.8 | 36.8 | 328.6 | 115.8 | 338.0 |
| Our result | 279.4 | 87.6 | 36.5 | 330.7 | 115.5 | 339.0 |
| Error, % | 0.6 | –0.2 | –0.8 | 0.6 | –0.3 | 0.3 |

## 3. Results of Calculations

*3.1. An MSL Covered with a Grounded Conductor and a Shielded MSL.* Constant parameters for the line in Figures 2(a) and 2(b) are strip thickness $t$=18 $\mu$m, substrate thickness $h$=1 mm, and relative dielectric constant of the substrate $\varepsilon_r$=4.5 (fiberglass).

For Figure 2(a) dependencies of $\tau$ on the height grounded conductor ($h1$) above the substrate at different values of the strip width ($w$) are shown in Figure 5(a). A characteristic feature of the dependencies is their intercrossing. Thus, at the beginning of the range of $h1$, the increase of $w$ decreases $\tau$ and at the end increases. In the middle of the range (at $h1$=0.5–0.8 mm) there will be a minimal (up to zero) sensitivity of $\tau$ to the variation of $w$. It is also remarkable that the sensitivity of $\tau$ to the variation of $h1$ decreases with the decrease of $w$. Similar dependencies for $Z$ are shown in Figure 5(b). They increase monotonically and do not intercross.

For Figure 2(b), the characteristics were preliminarily simulated at a distance from the side walls to the strip ($d$) equal to $w$ for $w$=0.1 mm for segment lengths of 10, 5, 2.5, 1.25 $\mu$m (with the uniform segmentation of the boundaries of conductors and dielectrics). The results analysis showed a consequent increase in the $\tau$ value with deviations of 0.4, 0.3, and 0.1% and a decrease in the $Z$ value by 0.7, 0.4, and 0.1%, respectively. Subsequent calculations were performed with a segment length of 10 $\mu$m, providing an acceptable error of less than 0.7%. The dependencies of $\tau$ and $Z$ on the height of the cover above the substrate ($h1$) at different values of $w$ for the distance from the side walls to the strip $d$=$w$, 3$w$ are shown in Figures 6 and 7. The analysis of the dependencies in Figures 6(a) and 7(a) shows that upon varying $h1$ in the

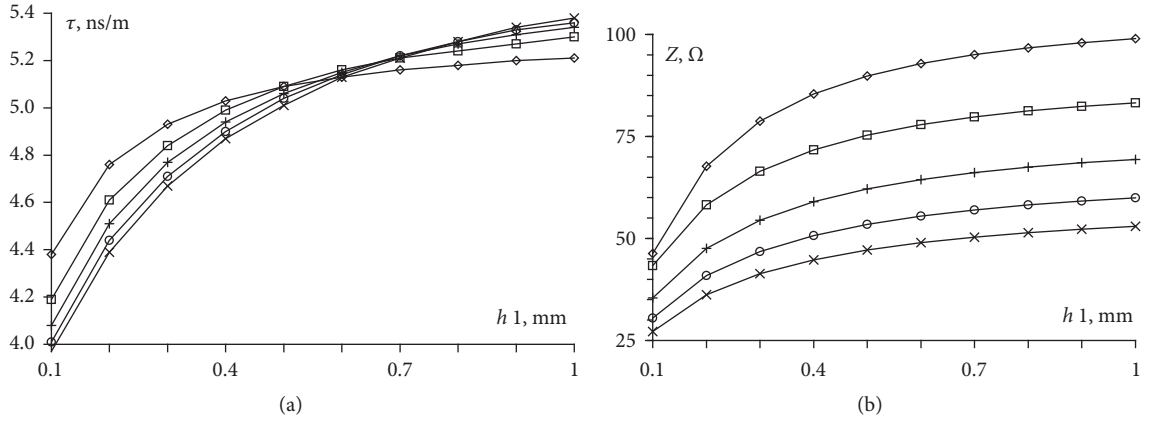



FIGURE 5: Dependencies of $\tau$ (a) and $Z$ (b) on $h1$ with $w$=0.3 (◊); 0.6 (□); 0.9 (+); 1.2 (○); 1.5 (×) mm for Figure 2(a).

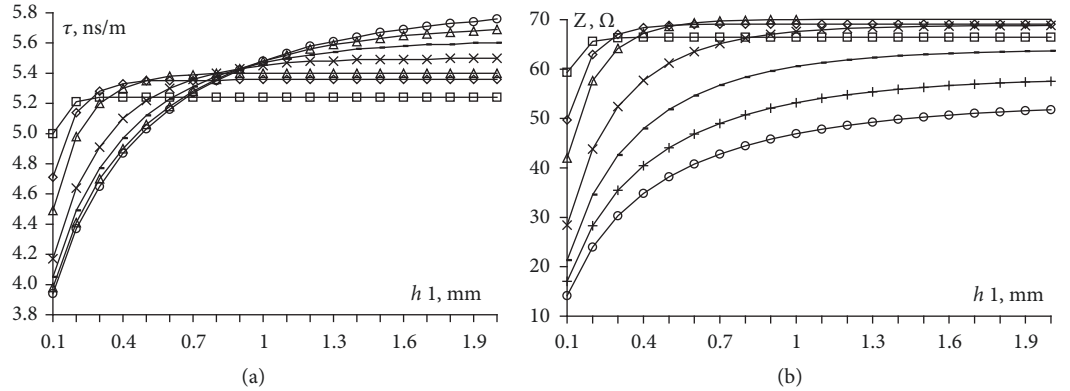



FIGURE 6: Dependencies of $\tau$ (a) and $Z$ (b) on $h1$ with $w$=0.1 (□); 0.2 (◊); 0.3 (Δ); 0.6 (×); 0.9 (-); 1.2 (+); 1.5 (○) mm and $d$=$w$ for Figure 2(b).

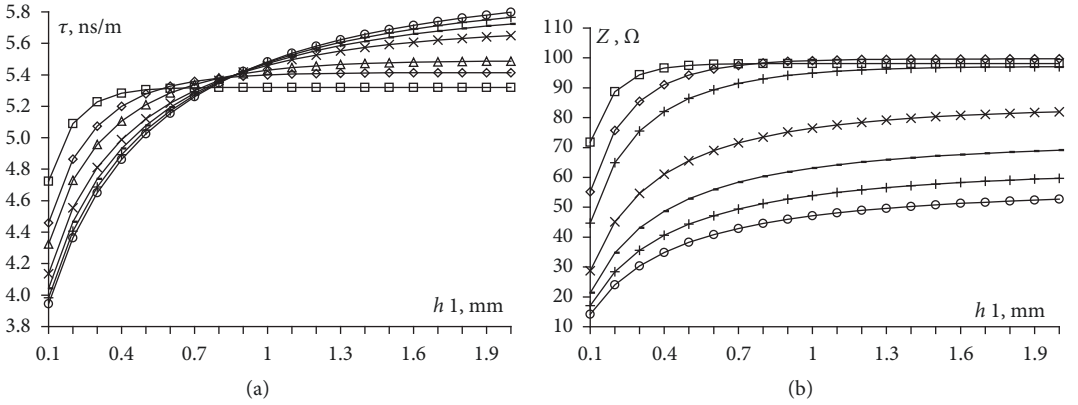



FIGURE 7: Dependencies of $\tau$ (a) and Z (b) on $h1$ with $w$=0.1 (□); 0.2 (◊);0.3 (Δ); 0.6 (×); 0.9 (-); 1.2 (+); 1.5 (○) mm and $d$=3$w$ for Figure 2(b).

entire range, when the strip is widest ($w$=0.6, 0.9, 1.2, 1.5 mm), the value of $\tau$ increases monotonically, and when the width of the strip is small ($w$=0.1, 0.2, 0.3 mm), the zero sensitivity of $\tau$ to the variation of $h1$ is observed, in almost entire range. Dependencies over $w$=0.6; 0.9; 1.2; 1.5 mm intercross at one point ($h1$=0.9 mm); i.e., at this point there will be zero sensitivity of $\tau$ to the change of $w$. As $w$ decreases to 0.1 mm, the intercrossing point of the dependencies shifts to $h1$=0.2 mm. In Figures 6(b) and 7(b) the corresponding dependences for $Z$ are shown. They behave similarly to the dependencies for $\tau$, also showing the possibility of obtaining zero sensitivity to changes in $h1$ and $w$. This fact has special practical importance because the stability of $Z$ is critical for many applications.

Consider the influence of the side walls on the calculated characteristics. Quantitative estimates can be done from the comparison of the relevant dependencies from Figures 5, 6, and 7. Meanwhile, the comparison with the dependencies for

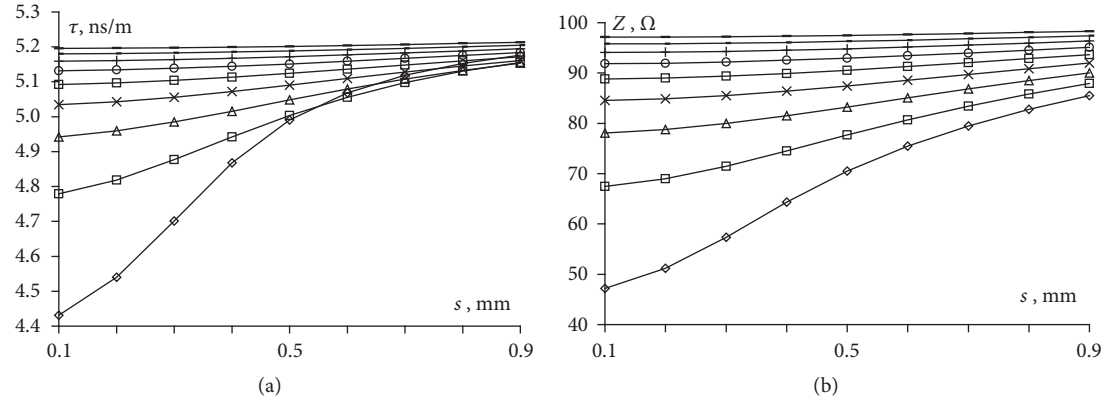



Figure 8: Dependencies of $\tau$ (a) and $Z$ (b) on $s$ with $h1$=0.1 (◊); 0.2 (□); 0.3 (Δ); 0.4 (×); 0.5 (□); 0.6 (○); 0.7 (+); 0.8 (-); 0.9(–) mm for Figure 3(a).

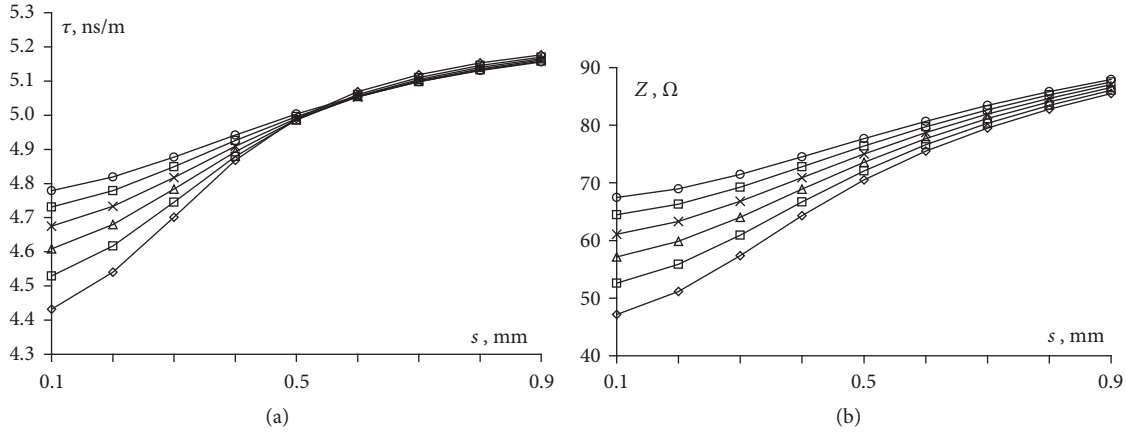



Figure 9: Dependencies of $\tau$ (a) and $Z$ (b) on $s$ with $h1$=0.1 (◊); 0.12 (□); 0.14 (Δ); 0.16 (×); 0.18 (□); 0.2 (○) mm for Figure 3(a)

the covered MSL (without side walls) of the same parameters allows us to assume that it is the presence of side walls, by increasing the edge capacitances, that gives the possibility of obtaining zero sensitivity of $\tau$ and $Z$ over a wide range of values of $h1$.

*3.2. MSLs with Side Grounded Conductors.* Consider the results for various modifications of MSLs with side grounded conductors (Figures 3 and 4). To build these structures we chose the following cross-sectional parameters (they are close to typical): the width of the signal conductor is $w$=0.3 mm, the thickness of the signal and side grounded conductors is $t$=18 $\mu$m, the width of the side conductors is $w1$=1 mm, the thickness of the dielectric substrate is $h$=1 mm, and the relative permittivity of the substrate is $\varepsilon_r$=4.5.

In the TALGAT software we built the geometric models of the line cross-section and calculated (using the method of moments) the matrices (3×3) of per-unit-length coefficients of electrostatic induction, taking into account the dielectric as well as ignoring it. Calculations were performed, changing the distance ($2s+w$) between the side conductors located in the air, for $h1$=0.1–0.9 mm (with a segment length of 5 $\mu$m for Figure 3(a)). It can be seen from Figure 8 that with increasing $s$ the values of $\tau$ and $Z$ smoothly increase. At low values of $h1$ and $s$, changes in $\tau$ and $Z$ are more pronounced, and an increase in $h1$ leads to an increase in the values of $\tau$ and $Z$. The approaching of the side conductors to the air–substrate boundary has a special effect on the characteristics of $\tau$: for small values of $h1$, the characteristics intersect. Therefore, we performed similar calculations for $h1$=0.1–0.2 mm with a step of 0.02 mm (Figure 9). One can see a similar behavior of dependencies for small $s$. However, at $s$=0.5–0.9 mm, the minimum (close to zero) sensitivity of $\tau$ to the change in $h1$ is revealed, which can be used to obtain a stable delay.

For Figure 3(b) we performed calculations for the change of distance between the side conductors ($s$), dipped in the substrate, for the height of the side conductors $h1$=0.1–0.9 mm (Figure 10). It is seen that with the increase of $s$, the value of $\tau$ gradually decreases, while $Z$ increases. At low values of $h1$, the changes of $\tau$ and $Z$ are small, but the growth of $h1$ leads to an increase in the value of $\tau$ and a decrease in the value of $Z$, while at small values of $s$ the changes of $\tau$ and $Z$ are more significant. Additionally, we performed simulation with a smaller step near the air–substrate interface: for $h1$=0.8; 0.82; 0.84; 0.86; 0.88; 0.9 mm (Figure 11). The analysis of Figure 11 shows a similar behavior of dependencies, but it reveals its specific character as well. It is expressed in

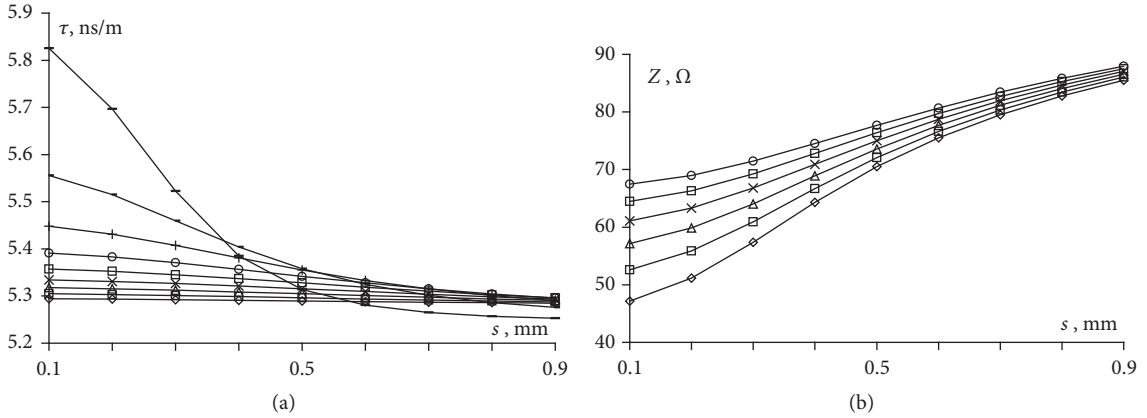



Figure 10: Dependencies of $\tau$ (a) and $Z$ (b) on $h1$ with $h1$=0.1 (◊); 0.2 (□); 0.3 (Δ); 0.4 (×); 0.5 (□); 0.6 (○); 0.7 (+); 0.8 (-); 0.9(–) mm for Figure 3(b).

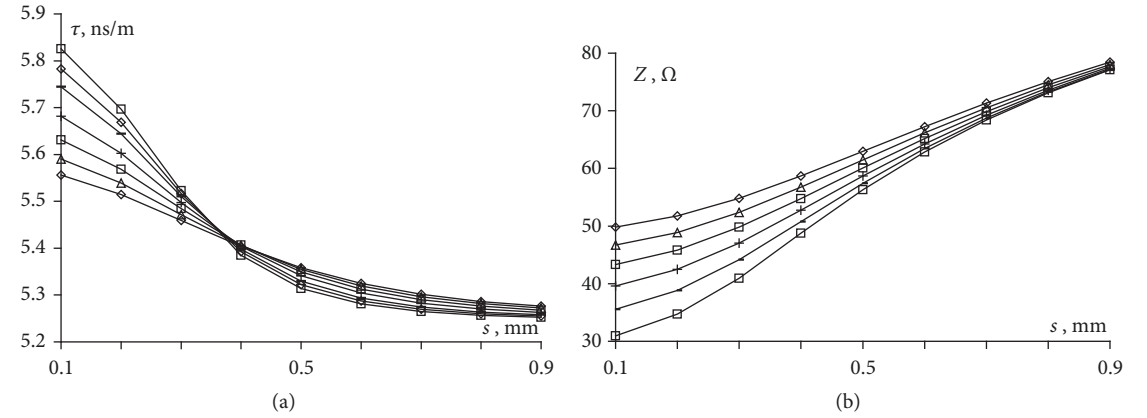



Figure 11: Dependencies of $\tau$ (a) and $Z$ (b) on with $h1$=0.8 (◊); 0.82 (Δ); 0.84 (□); 0.86 (+); 0.88 (–); 0.9 (□) mm for Figure 3(b).

the increased influence of the side conductors when they approach the air–substrate interface for low values of $s$. When $s$=0.1 mm, the value of $\tau$ increases from 5.56 to 5.82 ns/m. We noticed that, for large values of $s$, the approaching of the side conductors to the air–substrate boundary does not increase but, instead, decreases the values of $\tau$. When $s$=0.6 mm, this decrease is maximal and is from 5.33 ns/m to 5.29 ns/m. When $s$=0.38 mm, the change of $h1$ value in the whole range hardly ever changes the values of $\tau$ and, therefore, zero sensitivity of $\tau$ to changes of $h1$ is possible. Thus, we can obtain the required $Z$ value in the range from 48 to 59 Ω by changing the value of $h1$.

As $s$ increases, the values of $\tau$ (Figure 12(a)) and $Z$ (Figure 12(b)) gradually increase, but the change of $\tau$ is much smaller. The deepening of grounded conductors reduces the sensitivity of $\tau$ to changes in $s$. The change of $\tau$ over the entire range of $s$ is less than 2%. It can be assumed that with certain parameters of MSLs, their sensitivity can be reduced to zero. In this regard, in addition to $t$=18 $\mu$m, the values of $\tau$ and $Z$ are calculated for typical values of the thickness of the conductors ($t$=35, 70, 105 $\mu$m) as $s$ varies (Figures 13–15). Consider first the graphs for $\tau$ (Figures 13(a), 14(a), and 15(a)). With the increase of $s$, the value of $\tau$ smoothly increases, but not in all cases. Thus, the deepening of grounded conductors reduces the sensitivity of $\tau$ to changes in $s$, and even more, with increasing the thickness of the conductors, down to zero sensitivity of $\tau$. It can be assumed that with certain parameters of MSLs, the sensitivity can be reduced to almost zero in a wide range of $s$ values. For example, the value of $\tau$ for Figure 3(b) at $t$=35 $\mu$m (Figure 13(a)) is changed only by 0.8%. The graphs for $\tau$ at $t$=105 $\mu$m (Figure 15(a)) are also indicative, since the graph for $\tau$, with the deepening of the lateral conductors, turns from monotonously increasing into monotonously decreasing. It is obvious that there is such a value of the deepening of the conductors, at which the graph for $\tau$ will look very much like a horizontal straight line in the maximum range of $s$ values. The analysis of the graphs for $Z$ (Figures 12(b), 13(b), 14(b), and 15(b)) shows a slight influence of the position of the lateral conductors. Thus, it becomes possible to choose the line parameters to get the required $Z$ value with the minimum sensitivity of $\tau$ to the variations of $s$.

## 4. Conclusion

We have presented the systematic results of our study into the values of $\tau$ and $Z$ of modified MSLs. Comparison of the

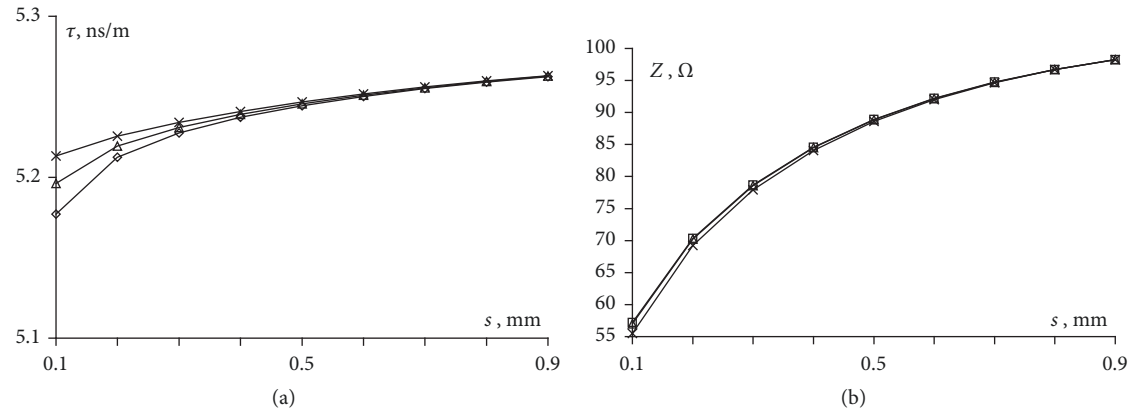



Figure 12: Dependencies of $\tau$ (a) and $Z$ (b) on $s$ for Figure 4 $a$ (□), $b$ (Δ), $c$ (×) with $t$=18 $\mu$m.

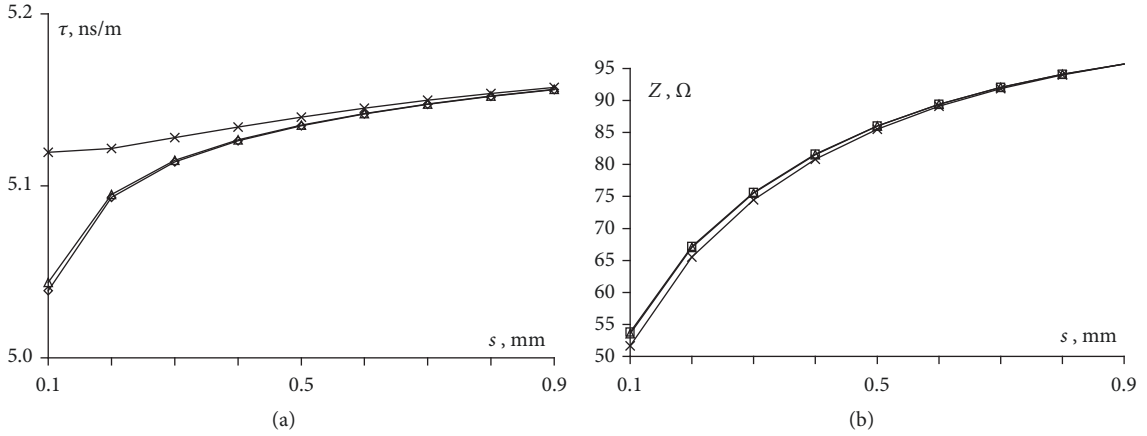



Figure 13: Dependencies of $\tau$ (a) and $Z$ (b) on $s$ for Figure 4 a (□), b (Δ), c (×) at $t$=35 $\mu$m.

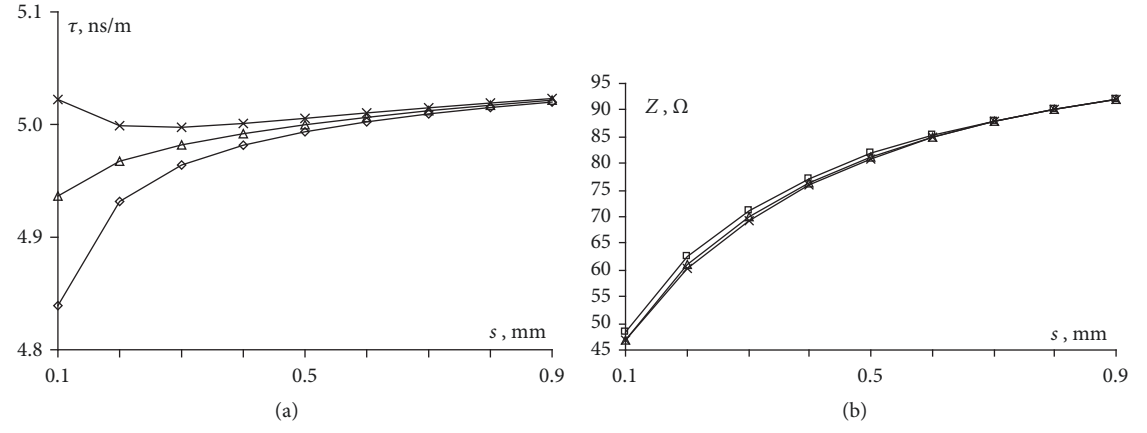



Figure 14: Dependencies of $\tau$ (a) and $Z$ (b) on $s$ for Figure 4 a (□), b (Δ), c (×) with $t$=70 $\mu$m.

MSL covered with a grounded conductor and the shielded MSL showed that the presence of side walls, by increasing the edge capacitances, allows minimal sensitivity in a wide range of upper conductor height values. For the MSL with side grounded conductors, their proximity to the air–substrate interface has a special influence on the characteristics under study. In particular, it becomes possible to select the line parameters to get the required $Z$ value with the minimum sensitivity of $\tau$ to the change in $s$. In addition, we revealed the possibility of zero sensitivity of $\tau$ to the change in the distance between the grounded conductors and the air–substrate interface when a given value of $Z$ is obtained.

The presented results have been obtained for particular values of line parameters. However, it is easy to obtain similar

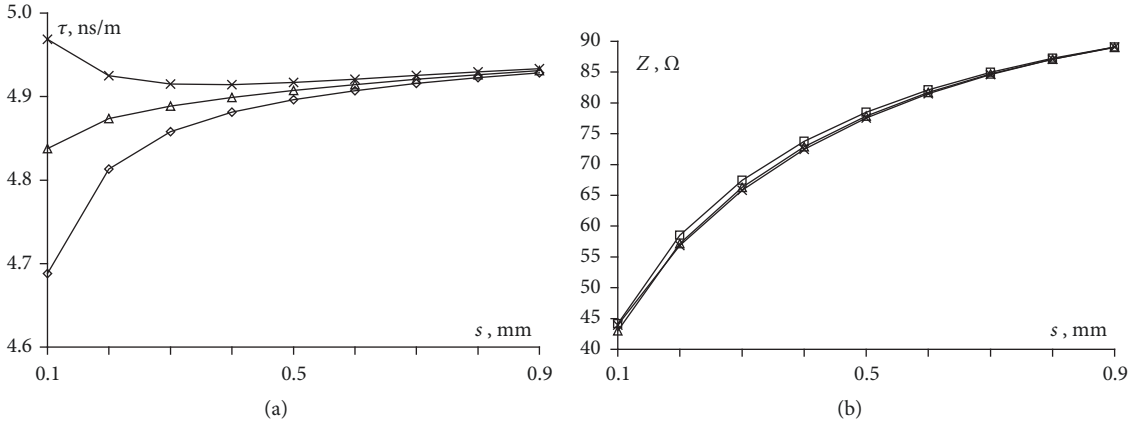



Figure 15: Dependencies of $\tau$ (a) and $Z$ (b) on $s$ for Figure 4 a (□), b (Δ), c (×) at $t$=105 $\mu$m.

dependencies for other values of the parameters and even other parameters. Besides, transmission lines with arbitrary number and shapes of conductors and dielectrics can be studied easily [28] to manufacture transmission lines with stable characteristics.

For multiple calculations, a lot of described accelerations of linear system solutions can be effectively used. In this paper, for quick estimations, we have used calculations of transmission line parameters only. However, for more comprehensive analysis, you can calculate the frequency or time response and use optimization by genetic algorithms similarly to [37].

Thus, taken together, for the first time, the above-mentioned techniques form a general approach to solving complexity problem in the electronics production process through the reduction of sensitivity of transmission line characteristics to their parameter variations. Versatility of the approach allows solving a wide range of tasks similar to considered examples.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This research was supported by The Ministry of Education and Science of the Russian Federation (project 8.9562.2017/8.9).

## References

[1] E. Bogatin and S. Zimmer, "Achieving impedance control targets," *Printed Circuit Design and Manufacture*, vol. 21, no. 4, pp. 28–31, 2004.

[2] L. G. Maloratsky, "Using modified microstrip lines to improve circuit performance," *High Frequency Electronics*, vol. 10, no. 5, pp. 38–52, 2011.

[3] R. P. Owens and M. H. N. Potok, "Analytical methods for calculating the characteristic impedance of finite-thickness microstrip lines," *International Journal of Electronics*, vol. 41, no. 4, pp. 399–403, 1976.

[4] M. K. Hamood, "Line thickness for various characteristic impedance of microstrip line," *Tikrit Journal of Pure Science*, vol. 18, no. 3, pp. 140–144, 2013.

[5] A. Arbabi, A. Boutejdar, M. Mahmoudi, and A. Omar, "Increase of characteristic impedance of microstrip line using a simple slot in metallic ground plane," in *Proceedings of the First IEEE International Conference on Communications and Electronics*, pp. 478–481, Hanoi, Vietnam, October 2006.

[6] R. Sharma, T. Chakravarty, S. Bhooshan, and A. B. Bhattacharyya, "Characteristic impedance of a microstrip-like interconnect line in presence of ground plane aperture," *International Journal of Microwave Science and Technology*, vol. 2007, Article ID 41951, 5 pages, 2007.

[7] R. S. Tomar and P. Bhartia, "Effects of manufacturing tolerances on the electrical performance of suspended and inverted microstrip lines," *International Journal of Infrared and Millimeter Waves*, vol. 6, no. 9, pp. 807–829, 1985.

[8] T. Gazizov and N. Leontiev, "Reduction of high-speed signal distortions in double-layered dielectric PCB interconnects," in *Proceedings of the Digest of 6-th Topical Meeting on Electrical Performance of Electronic Packaging*, pp. 67–69, San Jose, CA, USA, 1997.

[9] T. R. Gazizov, V. K. Salov, and S. P. Kuksenko, "Stable delay of microstrip line with side grounded conductors," *Wireless Communications and Mobile Computing*, vol. 2017, Article ID 1965739, 5 pages, 2017.

[10] I. Ye. Sagiyeva and T. R. Gazizov, "Modeling of microstrip line characteristics with side grounded conductors near airsubstrate boundary," *Journal of Physics: Conference Series [Electronic Resources]*, vol. 1118, pp. 1–6, 2018.

[11] I. Sagiyeva and T. Gazizov, "Side grounded conductors dipped in a substrate of a microstrip line, as a tool of line characteristics control," *Scientific Journal of Science and Technology*, vol. 19, no. 2, pp. 303–307, 2018.

[12] I. Y. Sagiyeva and T. R. Gazizov, "Decrease of microstrip line characteristics sensitivity at the expense of a shielding,"

in *Actual Problems of Radiophysics: Proceedings of the VII International Conference "APR–2017"*, pp. 67–70, 2018.

[13] J. D. Tebbens and M. Tůma, "Preconditioner updates for solving sequences of linear systems in matrix-free environment," *Numerical Linear Algebra with Applications*, vol. 17, no. 6, pp. 997–1019, 2010.

[14] P. N. Brown and Y. Saad, "Hybrid Krylov methods for nonlinear systems of equations," *SIAM Journal on Scientific Computing*, vol. 11, no. 3, pp. 450–481, 1990.

[15] A. Meister and C. Vömel, "Efficient preconditioning of linear systems arising from the discretization of hyperbolic conservation laws," *Advances in Computational Mathematics*, vol. 14, no. 1, pp. 49–73, 2001.

[16] D. Knoll and D. Keyes, "Jacobian-free Newton–Krylov methods: a survey of approaches and applications," *Journal of Computational Physics*, vol. 193, no. 2, pp. 357–397, 2004.

[17] M. Benzi and D. Bertaccini, "Approximate inverse preconditioning for shifted linear systems," *BIT Numerical Mathematics*, vol. 43, no. 2, pp. 231–244, 2003.

[18] F. Durastante, "Interpolant update of preconditioners for sequences of large linear systems," in *Proceedings of the Mathematical Methods, Computational Techniques and Intelligent Systems (MAMECTIS '15)*, vol. 41, pp. 40–47, 2015.

[19] D. Bertaccini, "Efficient preconditioning for sequences of parametric complex symmetric linear systems," *Electronic Transactions on Numerical Analysis*, vol. 18, pp. 49–64, 2004.

[20] G. Meurant, "On the incomplete cholesky decomposition of a class of perturbed matrices," *SIAM Journal on Scientific Computing*, vol. 23, no. 2, pp. 419–429, 2001.

[21] S. Bellavia, V. De Simone, D. di Serafino, and B. Morini, "Efficient preconditioner updates for shifted linear systems," *SIAM Journal on Scientific Computing*, vol. 33, no. 4, pp. 1785–1809, 2011.

[22] S. Bellavia, V. D. Simone, D. di Serafino, and B. Morini, "A Preconditioning framework for sequences of diagonally modified linear systems arising in optimization," *SIAM Journal on Numerical Analysis*, vol. 50, no. 6, pp. 3280–3302, 2012.

[23] J. Baglama, D. Calvetti, G. H. Golub, and L. Reichel, "Adaptively preconditioned GMRES algorithms," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 243–269, 1998.

[24] M. L. Parks, E. de Sturler, G. Mackey, D. D. Johnson, and S. Maiti, "Recycling krylov subspaces for sequences of linear systems," *SIAM Journal on Scientific Computing*, vol. 28, no. 5, pp. 1651–1674, 2006.

[25] D. Loghin, D. Ruiz, and A. Touhami, "Adaptive preconditioners for nonlinear systems of equations," *Journal of Computational and Applied Mathematics*, vol. 189, no. 1-2, pp. 362–374, 2006.

[26] J. Venkataraman, S. M. Rao, A. R. Djordjevic, T. K. Sarkar, and Y. Naiheng, "Analysis of arbitrarily oriented microstrip transmission lines in arbitrarily shaped dielectric media over a finite ground plane," *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT–33, no. 10, pp. 952–960, 1985.

[27] W. Delbare and D. De Zutter, "Space-domain green's function approach to the capacitance calculation of multiconductor lines in multilayered dielectrics with improved surface charge modeling," *IEEE Transactions on Microwave Theory and Techniques*, vol. 37, no. 10, pp. 1562–1568, 1989.

[28] S. Kuksenko, T. Gazizov, A. Zabolotsky et al., "New developments for improved simulation of interconnects based on method of moments," in *Proceedings of the 2015 International Conference on Modeling, Simulation and Applied Mathematics (MSAM'15)*, Advances in Intelligent Systems Research, pp. 293–301, Phuket, Thailand, August 2015.

[29] M. R. Scheinfein and O. A. Palusinski, "Methods of calculation of electrical parameters for electronic packaging applications," *Transactions of the Society for Computer Simulation*, vol. 4, no. 3, pp. 187–254, 1987.

[30] R. S. Surovtsev, A. V. Nosov, A. M. Zabolotsky, and T. R. Gazizov, "Possibility of protection against uwb pulses based on a turn of a meander microstrip line," *IEEE Transactions on Electromagnetic Compatibility*, vol. 59, no. 6, pp. 1864–1871, 2017.

[31] A. T. Gazizov, A. M. Zabolotskii, and T. R. Gazizov, "Measurement and simulation of time response of printed modal filters with broad-side coupling," *Journal of Communications Technology and Electronics*, vol. 63, no. 3, pp. 270–276, 2018.

[32] P. E. Orlov, T. R. Gazizov, and A. M. Zabolotsky, "Short pulse propagation along microstrip meander delay lines with design constraints: Comparative analysis of the quasi-static and electromagnetic approaches," *Applied Computational Electromagnetics Society Journal*, vol. 31, no. 3, pp. 238–243, 2016.

[33] R. R. Akhunov, S. P. Kuksenko, and T. R. Gazizov, "Acceleration of multiple iterative solution of linear algebraic systems in computing the capacitance of a microstrip line in wide ranges of its sizes," *Journal of Mathematical Sciences*, vol. 207, no. 5, pp. 686–692, 2015.

[34] T. R. Gazizov, S. P. Kuksenko, and R. R. Akhunov, "Acceleration of multiple solution of linear systems for analyses of microstrip structures," *Mathematical Models and Methods in Applied Sciences*, vol. 9, pp. 721–726, 2015.

[35] R. R. Akhunov, T. R. Gazizov, and S. P. Kuksenko, "Multiple solution of systems of linear algebraic equations by an iterative method with the adaptive recalculation of the preconditioner," *Computational Mathematics and Mathematical Physics*, vol. 56, no. 8, pp. 1382–1387, 2016.

[36] S. P. Kuksenko, R. R. Akhunov, and T. R. Gazizov, "Choosing order of operations to accelerate strip structure analysis in parameter range," *Journal of Physics: Conference Series [Electronic Resources]*, vol. 1015, no. 3, pp. 1–6, 2018.

[37] A. O. Belousov and T. R. Gazizov, "Systematic approach to optimization for protection against intentional ultrashort pulses based on multiconductor modal filters," *Complexity*, vol. 2018, Article ID 5676504, 15 pages, 2018.

Hindawi
Complexity
Volume 2019, Article ID 5734149, 14 pages
https://doi.org/10.1155/2019/5734149

*Research Article*

# A Novel MOEA/D for Multiobjective Scheduling of Flexible Manufacturing Systems

**Xinnian Wang,[1] Keyi Xing,[1] Chao-Bo Yan,[1] and Mengchu Zhou[2]**

[1]*State Key Laboratory for Manufacturing Systems Engineering, and Systems Engineering Institute, Xi'an Jiaotong University, Xi'an, China*
[2]*Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, USA*

Correspondence should be addressed to Keyi Xing; kyxing@mail.xjtu.edu.cn

Received 22 January 2019; Accepted 7 May 2019; Published 2 June 2019

Guest Editor: Rosario Domingo



This paper considers the multiobjective scheduling of flexible manufacturing systems (FMSs). Due to high degrees of route flexibility and resource sharing, deadlocks often exhibit in FMSs. Manufacturing tasks cannot be finished if any deadlock appears. For solving such problem, this work develops a deadlock-free multiobjective evolutionary algorithm based on decomposition (DMOEA/D). It intends to minimize three objective functions, i.e., makespan, mean flow time, and mean tardiness time. The proposed algorithm can decompose a multiobjective scheduling problem into a certain number of scalar subproblems and solves all the subproblems in a single run. A type of a discrete differential evolution (DDE) algorithm is also developed for solving each subproblem. The mutation operator of the proposed DDE is based on the hamming distance of two randomly selected solutions, while the crossover operator is based on Generalization of Order Crossover. Experimental results demonstrate that the proposed DMOEA/D can significantly outperform a Pareto domination-based algorithm DNSGA-II for both 2-objective and 3-objective problems on the studied FMSs.

## 1. Introduction

The *multiobjective optimization problem* (MOP, the abbreviations and their meanings are given in Table 1) is an optimization problem that may have a number of conflicting objectives to be considered, and decision-makers need to determine an optimal trade-off among the objectives. This problem presents in many real-life applications [1–4]. A *Pareto optimal* solution to an MOP is a candidate for the optimal trade-off [5]. Most MOPs have a lot of or even an infinite number of Pareto optimal solutions, and the set of all the Pareto optimal solutions in the objective space is called as *Pareto front* (PF). The decision-makers desire a fair approximation to the PF to make final decisions. *Multiobjective evolutionary algorithms* (MOEAs) work with populations of candidate solutions to MOPs and therefore are able to find good approximations to the PF in a single run.

To produce Pareto optimal vectors that are distributed evenly along the PF and therefore can approximate the PF, many MOEAs evaluate the solutions based on Pareto domination like SPEA2 [6] and NSGA-II [7]. Multiobjective evolutionary algorithm based on decomposition (MOEA/D) [8], which decomposes an MOP into a number of single-objective subproblems, is another effective MOEA framework. In an MOEA/D, the objective to each subproblem is a weighted aggregation of some individual objectives. Based on the distances between the weighted aggregation vectors, the neighborhood relationships among the subproblems can be found out. Each subproblem is then solved according to the information mainly from the neighboring subproblems. There are a few significant studies that apply it in dealing with multiobjective problems in different areas. Chen et al. [9] proposed a multiobjective discrete method called MODTLBO/D for solving the community detection problem of complex networks. They adopted a multiobjective decomposition method and introduced a neighbor-based mutation. For multiobjective job shop scheduling problem, Zhao et al. [10] developed an improved multi-objective evolutionary algorithm based on decomposition (IMOEA/D). Experimental results demonstrate that their IMOEA/D can converge better than Pareto dominance-based MOEAs.

Table 1: Descriptions of abbreviations.

| Abbreviation | Description | Abbreviation | Description |
|---|---|---|---|
| MOP | Multi-objective optimization problem | DE | Differential evolution |
| PF | Pareto front | DDE | Discrete differential evolution |
| MOEA | Multi-objective evolutionary algorithm | GOX | Generalization of order crossover |
| SPEA2 | Improving the strength Pareto evolutionary algorithm | PNS | Petri net for scheduling |
| NSGA-II | Nondominated sorting genetic algorithm II | DAP | Deadlock avoidance policy |
| MOEA/D | MOEA based on decomposition | EP | External population |
| MODTLBO/D | Multi-objective discrete teaching-learning-based optimization with decomposition | $S^3PRs$ | Systems of simple sequential processes with resources |
| IMOEA/D | Improved MOEA/D | DNSGA-II | Deadlock-free NSGA-II |
| FMS | Flexible manufacturing system | *NPS* | Number of Pareto solutions |
| AGV | Automated guided vehicle | *MID* | Mean ideal distance |
| PN | Petri net | *RAS* | Rate of achievement to objectives simultaneously |

A flexible manufacturing system (FMS) is a computer-controlled manufacturing system that is comprised of finite resources and can process multitypes of jobs. Different from the traditional production environments such as job shops and flow shops, FMSs often encounter deadlock problems due to high degrees of route flexibility and resource sharing. Once a deadlock appears, the whole or partial system will be indefinitely blocked and cannot finish manufacturing tasks. Developing effective control and scheduling approaches to avoid deadlocks while optimizing the performance of the system is of paramount importance in practice.

Scheduling of FMSs contains both deadlock control and optimization of objectives and therefore is more difficult. A few works have been published on this area [4, 11–23]. Most of them involve deadlock problems and some also concerns different optimization objectives such as the back-tracking and distance travel of AGVs [4], AGV's fleet size [11], and total energy consumption [21, 22]. However, none of them takes multiobjective optimization into consideration.

This work addresses the multiobjective scheduling problem of deadlock-prone FMSs for the first time and proposes a deadlock-free multiobjective evolutionary algorithm based on decomposition (DMOEA/D) and Petri net (PN) models. Our DMOEA/D can decompose a multiobjective FMS scheduling problem into several single-objective subproblems and optimizes all the subproblems in a single run by evolving a population of solutions. In each generation, a solution for solving a subproblem is reproduced by a new proposed discrete differential evolution (DDE) algorithm. In DDE, the mutation operator is based on the hamming distance between two randomly selected solutions. The Generalization of Order Crossover (GOX) [24] is used as the crossover operator. Two illustrated examples are used to demonstrate the efficiency of the proposed algorithm. Since there is no work reported for the multiobjective scheduling of FMSs considering deadlock situations, we just compare our proposed DMOEA/D with a NSGA-II based MOEA on these examples.

This work presents an effective approach for organizations and production managers to enhance their competitiveness in manufacturing versatile, route-flexible, and time-critical productions concerning two or more different scheduling objectives.

The rest of the paper is organized as follows. Section 2 introduces the FMSs studied in the paper and their Petri net models and defines the considered multiobjective scheduling problem. Section 3 introduces some basics of MOEA/D and DDE and develops a multiobjective scheduling algorithm based on them. A performance comparison between two developed scheduling algorithms is made, and the experimental results are show in Section 4. Section 5 concludes the paper.

## 2. PN Models of FMSs

PN is a widely used mathematical tool for modeling the dynamic behaviors of FMSs and many other manufacturing systems [15–17, 25–27]. This section briefly introduces some basics of PNs first and then the PN models of FMSs for multiobjective scheduling. For more details of PNs, readers may refer to [28].

*2.1. Definitions of Petri Nets.* Let $\mathbb{Z}_0 = \{0, 1, 2, \ldots\}$ and $\mathbb{Z}_k = \{1, 2, \ldots, k\}$. A marked PN is a 4-tuple $N = (P, T, F, M_0)$, where $P$ is a finite set of places, $T$ is a finite set of transitions, $F \subseteq (P \times T) \cup (T \times P)$ is the set of directed arcs, and $M_0 : P \longrightarrow \mathbb{Z}_0$ is the initial marking of $N$. For a given node $x \in P \cup T$, its preset is defined as ${}^\bullet x = \{y \in P \cup T \mid (y, x) \in F\}$ and postset $x^\bullet = \{y \in P \cup T \mid (x, y) \in F\}$. Given a marking $M$ and a place $p \in P$, denote $M(p)$ as the number of tokens in $p$ at $M$. A string $\alpha = x_1 x_2 \cdots x_k$ is called a *path* in the PN, where $x_i \in P \cup T$ and $(x_i, x_{i+1}) \in F$, $i \in \mathbb{Z}_{k-1}$.

A transition $t \in T$ is enabled at $M$ if $\forall p \in {}^\bullet t$, $M(p) > 0$, denoted as $M[t >$. An enabled transition $t$ can fire at $M$, yielding $M'$, denoted as $M[t > M'$, where $M'(p) = M(p) - 1$, $\forall p \in {}^\bullet t \backslash t^\bullet$, $M'(p) = M(p) + 1$, $\forall p \in t^\bullet \backslash {}^\bullet t$, and otherwise $M'(p) = M(p)$. A sequence of transitions $\alpha = t_1 t_2 \cdots t_k$ is feasible from $M$ if $M_i[t_i > M_{i+1}$ holds for $i \in \mathbb{Z}_k$, where $M_1 = M$.

*2.2. Placed Timed PN Models of FMSs.* An FMS studied in our work is comprised of $m$ types of resources, denoted as $R = \{r_i, i \in \mathbb{Z}_m\}$, and is capable to process $n$ types of jobs, denoted as $Q = \{q_j, j \in \mathbb{Z}_n\}$. More definitions and constraints are described as follows:

(1) The number of type-$q_j$ jobs to be processed is $\varphi(q_j)$, and the total number of jobs is $\Phi = \sum_{j \in \mathbb{Z}_n} \varphi(q_j)$.

(2) The capacity of type-$r_i$ resources is denoted as $C(r_i)$, which is a positive integer and marks the maximum jobs that type $r_i$ resources can handle simultaneously.

(3) A processing route of a type-$q_j$ job, $w_k = o_{k1} o_{k2} \cdots o_{kl(w_k)}$, is a predefined sequence of operations, where $l(w_k)$ is the total number of operations in route $w_k$ and $o_{kl}$ is the $l$th operation in $w_k$. A job may be processed on more than one route and can choose its routes while processing. Let $\Omega = \{w_k \mid k \in \mathbb{Z}_{|\Omega|}\}$ be the set of all processing routes and $\Omega_j \subseteq \Omega$ be the set of routes for type-$q_j$ jobs, respectively.

(4) Each operation needs one unit resource, and any two consecutive operations of a job need different resource types. For operation $o_{kl}$, let $R(o_{kl})$ denote the resource needed for processing $o_{kl}$.

(5) The processing time $d(o_{kl})$ for operation $o_{kl}$ is predefined. Let $c(o_{kl})$ denote the completion time of $o_{kl}$.

(6) No pre-emption is allowed.

Now we can establish a PN model for FMS considered in this paper.

For type-$q_j$ jobs, let $o_{js}$ and $o_{je}$ be two virtual operations representing the storages of raw and processed type-$q_j$ jobs, respectively. Operations $o_{js}$ and $o_{je}$ do not need any resource. Then, route $w_k$ with these two fictitious operations for type-$q_j$ jobs can be defined as $w_k = o_{js} o_{k1} o_{k2} \cdots o_{kl(w_k)} o_{je}$. Identical operations shared by different routes are merged as one operation.

A processing route $w_k = o_{js} o_{k1} o_{k2} \cdots o_{kl(w_k)} o_{je}$ in the PN model is modelled as a path of places and transitions $a_k = p_{js} t_{k1} p_{k1} t_{k2} p_{k2} \cdots t_{kl} p_{kl} \cdots t_{kl(w_k)} p_{kl(w_k)} t_{k(l(w_k)+1)} p_{je}$, where $p_{js}$ and $p_{je}$ represent operations $o_{js}$ and $o_{je}$, respectively, and $p_{kl}$ is an operation place that represents operation $o_{kl}$; $t_{kl}$ is a transition that indicates the start of $o_{kl}$ and the completion of $o_{k(l-1)}$. A token in operation place $p_{kl}$ means that operation $o_{kl}$ of a job is being processed. In these ways, the marked PN model of processing routes for type-$q_j$ jobs is defined as

$$N_j = \left(P_j \cup \left\{p_{js}, p_{je}\right\}, T_j, F_j, M_{j0}\right) \tag{1}$$

where $P_j = \cup_{1 \le k \le |\Omega_j|} \{p_{k1}, p_{k2}, \ldots, p_{kl(w_k)}\}$, $T_j = \cup_{1 \le k \le |\Omega_j|} \{t_{k1}, t_{k2}, \ldots, t_{k(l(w_k)+1)}\}$, and $F_j = \cup_{1 \le k \le |\Omega_j|} \{(p_{js}, t_{k1}), (t_{k1}, p_{k1}), (p_{k1}, t_{k2}), \ldots, (p_{kl(w_k)}, t_{k(l(w_k)+1)}), (t_{k(l(w_k)+1)}, p_{je})\}$. $M_{j0}$ is the initial marking, where $M_{j0}(p_{js}) = \varphi(q_j)$ and $M_{j0}(p) = 0$, $\forall p \in P_j \cup \{p_{je}\}$. In $N_j$, $\forall t \in T_j$, $|t^\bullet| = |{}^\bullet t| = 1$. An operation place $p$ is a split place if $p \in P_j$, $|p^\bullet| > 1$. At a split place, a job is able to choose the processing routes.

For type-$r_i$ resources, assign a resource place denoted also by $r_i$. Tokens in $r_i$ represent the number of available type-$r_i$ resources. Let $C(r_i)$ denote the initial marking of $r_i$.

Denote $R(p)$ and $P_R$ as the resource needed by operation place $p$ and the set of all resource places, respectively. Then, in our PN model, add arcs from $R(p)$ to each transition in ${}^\bullet p$ denoting the occupation of $R(p)$, and add arcs from each transition in $p^\bullet$ to $R(p)$ denoting the releasing of $R(p)$. The set of all arcs related to resource places is denoted as $F_R$. Then, the marked PN model of our studied FMS can be defined as

$$N = \left(P \cup P_s \cup P_f \cup P_R, T, F, M_0\right) \tag{2}$$

where $P = \cup_{j \in \mathbb{Z}_n} P_j$, $P_s = \{p_{js} \mid j \in \mathbb{Z}_n\}$, $P_f = \{p_{je} \mid j \in \mathbb{Z}_n\}$, $T = \cup_{j \in \mathbb{Z}_n} T_j$, $F = F_Q \cup F_R$, and $F_Q = \cup_{j \in \mathbb{Z}_n} F_j$. The initial marking $M_0$ is defined as $M_0(p_{js}) = \varphi(q_j)$, $\forall p_{js} \in P_s$; $M_0(p) = 0$, $\forall p \in P \cup P_f$; and $M_0(r_i) = C(r_i)$, $\forall r_i \in P_R$.

In this paper, processing times needed by operations are described by a place-timed PN. Each operation place $p$ is assigned with a time delay $d(p)$, denoting its processing time, $d(p) = 0$, $\forall p \in P_s \cup P_f \cup P_R$. Such PN model for an FMS is called PNS [17].

*Example 1.* Consider an FMS shown in Figure 1. It contains four machines $r_1$, $r_2$, $r_3$, and $r_4$. Machines $r_1$, $r_2$, and $r_4$ can hold one job at the same time while machine $r_3$ can hold two. Then the resource set is $R = \{r_1, r_2, r_3, r_4\}$, with initial markings $C(r_1) = C(r_2) = C(r_4) = 1$, and $C(r_3) = 2$. This FMS is able to process two types of jobs, $q_1$ and $q_2$. Type-$q_1$ jobs can be processed sequentially on $r_1$, $r_2$, and $r_4$, or on $r_1$, $r_3$, and $r_4$; while type-$q_2$ jobs are processed on $r_4$, $r_3$, and $r_1$. Thus, there are two processing routes for type-$q_1$ jobs, $w_1 = o_{1s} o_{11} o_{12} o_{13} p_{1e}$ and $w_2 = o_{1s} o_{21} o_{22} o_{23} o_{1e}$, while type-$q_2$ jobs are processed on one route, $w_3 = o_{2s} o_{31} o_{32} o_{33} o_{2e}$. Note that $o_{11} = o_{21}$ and $o_{13} = o_{23}$ are processed on machines $r_1$ and $r_4$, respectively. Then routes $w_1$ and $w_2$ can be modelled as $w_1 = p_{1s} t_{11} p_{11} t_{12} p_{12} t_{13} p_{13} t_{14} p_{1e}$ and $w_2 = p_{1s} t_{11} p_{11} t_{22} p_{22} t_{23} p_{13} t_{14} p_{1e}$, respectively, while route $w_3$ is modelled as $w_3 = p_{2s} t_{31} p_{31} t_{32} p_{32} t_{33} p_{33} t_{34} p_{2e}$. Figure 2(a) shows the PN model of this system, in which the jobs of types $q_1$ and $q_2$ are to be processed are 2 and 1, respectively.

When all operations of all jobs are completed, the system reaches its final marking, denoted as $M_f$, where $M_f(p_{ie}) = M_0(p_{js})$ and $\forall p_{je} \in P_f$; $M_f(p) = 0$, $\forall p \in P \cup P_s$; and $M_f(r_i) = C(r_i)$, $\forall r_i \in P_R$. A sequence of transitions $\alpha$ is complete and feasible if $M_0[a > M_f$. Such a sequence in the PNS is a *feasible schedule* to the considered FMS.

Note that an improper schedule of a PNS may lead to deadlock situations, and the system can thus never reach its final marking. Consider a reachable marking $M_1 = 2p_{22} + p_{31} + r_1 + r_2$ in Figure 2(b). It is clear that a circular wait for resources $r_3$ and $r_4$ arises and no transition can be fired.

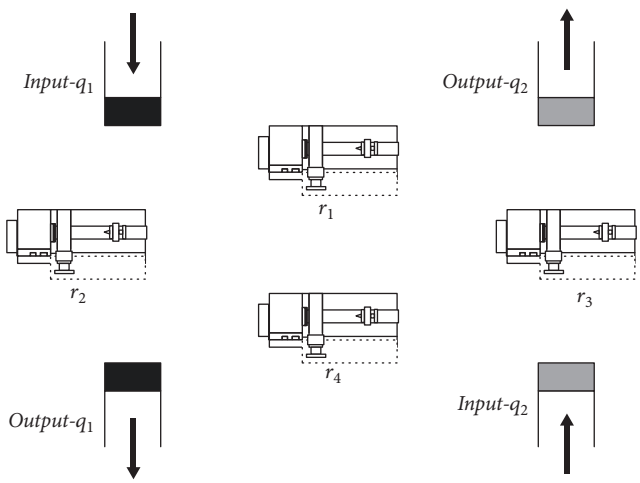



Figure 1: The FMS in Example 1.

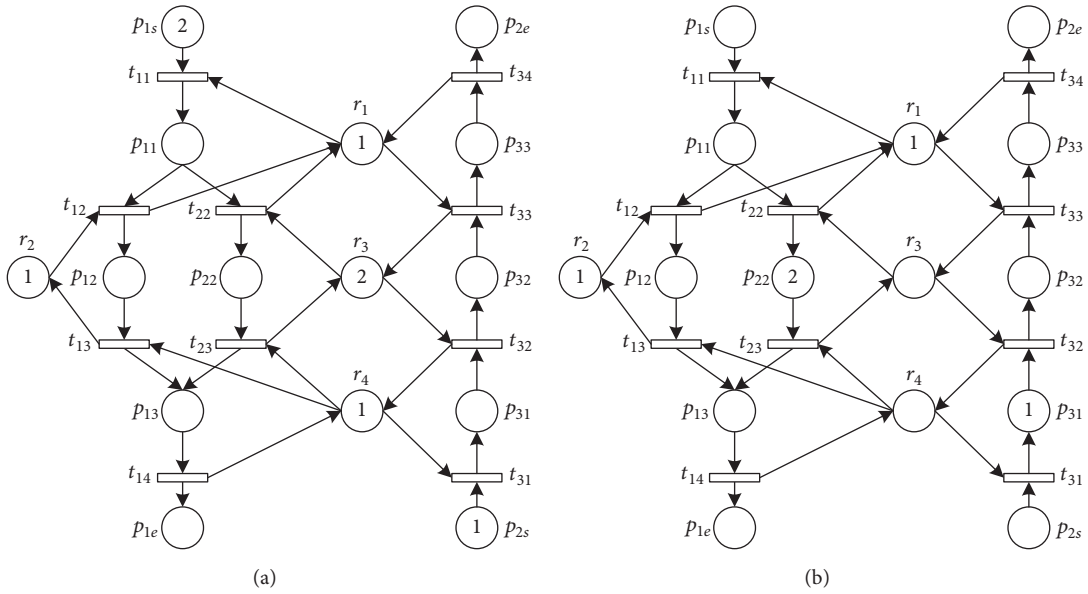



Figure 2: (a) The PNS of an FMS and (b) its deadlock situation.

## 3. DMOEA/D for the Scheduling of FMSs

A multiobjective optimization problem (MOP) is formally stated as follows:

$$\begin{aligned} &\text{Minimize} \quad F(x) = \left(f_1(x), f_2(x), \ldots, f_L(x)\right)^T \\ &\text{Subject to} \quad x \in \Pi \end{aligned} \tag{3}$$

where $\Pi$ is the decision (variable) space and $f_i(x)$ is the objective functions, $i \in \mathbb{Z}_L$. A feasible *solution* to MOP can be interpreted into a feasible schedule. Then, the multiobjective scheduling problem studied in this paper is to find a feasible solution $x^*$ so that $F(x^*)$ is as small as possible.

MOEA/D solves an MOP problem by decomposing it into several single-objective optimization subproblems and then solves all the subproblems in a single run by evolving a population of solutions [8]. The population in each generation is composed of best found solutions for these subproblems. The neighbourhood relations among the subproblems are calculated according to the distances between their coefficient vectors. In MOEA/D, each subproblem is then solved based on the information of its neighbouring subproblems.

There are a few studies using MOEA/D in different areas recently [5, 8–10, 29–31]. In this work, we focus on the multiobjective scheduling of FMSs and propose a novel DMOEA/D by combining MOEA/D and DDE. PNSs are used to model the systems and deadlock controllers are embedded to avoid deadlocks. The varieties of DMOEA/D are described as follows.

*3.1. Representation, Interpretation, and Reparation.* In this paper, a solution $x$ in DMOEA/D consists of two sections $x = (S_r; S_o)$, where $S_r$ stores the route information and $S_o$ is a permutation with repetition of all jobs. Let $l(w_s)$ denote

the length of route $w_s$. Then, each job $J_i$ appears $l(J_i)$ times in $S_o$, where $l(J_i) = \max\{l(w_s) \mid w_s \text{ is a route of job } J_i\}$. The $j$th operation of $J_i$ is represented as the $j$th $J_i$ in $S_o$, and $S_o$ is uniquely interpreted as a sequence of operations $\sigma(S_o)$. Then $x$ is rewritten as $x = (S_r; \sigma(S_o))$. By associating each operation to its corresponding transition, a sequence of transitions $\alpha(x)$, which is taken as a schedule in PNS, can be interpreted from solution $x$.

*Example 2.* Reconsider the PNS in Figure 1(a). There are three jobs to be processed: two type-$q_1$ jobs, $J_1$ and $J_2$, and one type-$q_2$ job, $J_3$. The type-$q_1$ jobs have two routes, $w_1$ and $w_2$, while type-$q_2$ job has only one, $w_3$. Suppose that $J_1$ and $J_2$ are processed by routes $w_1$ and $w_2$, respectively. Then, $S_{r1} = (w_1, w_2, w_3)$ can be taken as the first section of a solution $x_1$. Note that the route lengths $l(J_i)$ are 5, 5, and 4, respectively. Thus, we can represent section $S_o$ as a permutation with repetition that consists of five $J_1$'s, five $J_2$'s, and four $J_3$'s, for example, $S_{o1} = (J_1, J_1, J_3, J_2, J_3, J_3, J_1, J_2, J_1, J_2, J_2, J_2, J_1, J_3)$. Then $x_1$ is represented as $x_1 = (S_{r1}; S_{o1}) = (w_1, w_2, w_3; J_1, J_1, J_3, J_2, J_3, J_3, J_1, J_2, J_1, J_2, J_2, J_2, J_1, J_3)$, $S_{o1}$ is interpreted as a sequence of operations $\sigma(S_{o1}) = (O_{11}, O_{12}, O_{31}, O_{21}, O_{32}, O_{33}, O_{13}, O_{22}, O_{14}, O_{23}, O_{24}, O_{25}, O_{15}, O_{34})$, and $x_1$ is interpreted as a sequence of transitions $\alpha(x_1) = (t_{11}t_{12}, t_{31}, t_{11}, t_{32}, t_{33}, t_{13}, t_{22}, t_{14}, t_{23}, t_{24}, t_{15}, t_{15}, t_{34})$, or for more details, $\alpha(x_1) = (t_{11}[O_{11}], t_{12}[O_{12}], t_{31}[O_{31}], t_{11}[O_{21}], t_{32}[O_{32}], t_{33}[O_{33}], t_{13}[O_{13}], t_{22}[O_{22}], t_{14}[O_{14}], t_{23}[O_{23}], t_{24}[O_{24}], t_{15}[O_{25}], t_{15}[O_{15}], t_{34}[O_{34}])$.

The sequence of transitions interpreted from a solution may be infeasible and contain deadlock situations. The feasibility of each solution must be examined and the infeasible ones are repaired. There are some works addressing the deadlock problem of FMSs [32–36], and deadlock-free operations of FMSs can thus be guaranteed. In this work, an Amending Algorithm proposed by [17] is used to obtain a feasible sequence of transitions from $M_0$ to $M_f$. This Amending Algorithm is based on the deadlock avoidance policy (DAP) proposed by [34]. An optimal polynomial complexity DAP for $S^3$PRs without $\xi$-resources is developed by a one-step look-ahead approach; for $S^3$PRs with $\xi$-resources, a suboptimal DAP is also derived by reducing the net. A deadlock search algorithm is then proposed for prohibiting deadlock situations.

*3.2. Objective Functions.* Three objective functions are used in the work, makespan, mean completion time, and mean tardiness time.

For a given solution $x$, let $\alpha(x) = t_1 t_2 \cdots t_n$ be its corresponding sequence of transitions. Let $f(t_k[O_{ij}])$ be the firing time of transition $t_k$ and $C_i$ be the completion time of the last operation of job $J_i$. Note that $f(t_k[O_{ij}])$ is also the start time of operation $O_{ij}$. According to the predefined operation sequence and the order in a schedule (sequence of transitions), transition $t_k[O_{ij}]$ ($k \in \mathbb{Z}_n$) can be fired only after transition $t_{k-1}$ is fired and operation $O_{i(j-1)}$ is completed. Assume that $t_{k-1}$ corresponds to operation $O_{uv}$ and $t_u$ corresponds to operation $O_{i(j-1)}$. We have $f(t_k[O_{ij}]) = \max\{f(t_{k-1}[O_{uv}]), f(t_u[O_{i(j-1)}]) + d(O_{i(j-1)})\}$, and the completion time of job $J_i$ can be calculated as $C_i = f(t_v[O_{il(J_i)}]) + d(O_{il(J_i)})$, where $O_{il(J_i)}$ is the last processed operation of job $J_i$ and $t_v$ corresponds to $O_{il(J_i)}$.

The makespan $C_{max}$ is the completion time of the last job that leaves the system. A minimum $C_{max}$ implies a good utilization of machines. It can be defined as

$$f_1(x) \equiv C_{max} = \max_{1 \le i \le \Phi} \{C_i\} \tag{4}$$

The mean completion time is the average completion time of all the jobs, and its definition is

$$f_2(x) \equiv \overline{C} = \sum_{1 \le i \le \Phi} \frac{C_i}{\Phi} \tag{5}$$

Tardiness time of a job is a time delay after the predefined due time. Jobs completed after the due time may cause compensation for customers and loss of goodwill. Let $D_i$ and $T_i$ be the due time and the tardiness time of job $J_i$, respectively. In this paper, $D_i$ is defined as $D_i = 1.5 \times \sum_{1 \le j \le l(J_i)} d(O_{ij})$ and $T_i$ is defined as $T_i = \max\{C_i - D_i, 0\}$. The mean tardiness time is the average tardiness time of all the jobs, and it can be expressed as

$$f_3(x) \equiv \overline{T} = \sum_{1 \le i \le \Phi} \frac{T_i}{\Phi} \tag{6}$$

*3.3. Decomposition of Multiobjective Optimization.* A general MOEA/D must decompose an MOP into $K$ single-objective optimization problems, where $K$ is the population size. Several approaches have been proposed for this task and this paper uses the Tchebycheff approach.

Given a weight vector $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_L)^T$, $\lambda_i \ge 0$, $\forall i \in \mathbb{Z}_L$, and $\sum_{i \in \mathbb{Z}_L} \lambda_i = 1$. Then, in the Tchebycheff approach, the optimal solution to the single-objective optimization problem below

$$\begin{aligned} &\text{Minimize} \quad g^{te}(x \mid \lambda) = \max_{i \in \mathbb{Z}_L} \{\lambda_i (f_i(x) - z_i^*)\}^T \\ &\text{Subject to} \quad x \in \Pi \end{aligned} \tag{7}$$

is a Pareto optimal solution to (3), where $z^* = (z_1^*, z_2^*, \ldots, z_L^*)^T$ is the reference point, i.e., $z_i^* = \min_{x \in \Pi} f_i(x), \forall i \in \mathbb{Z}_L$. For each optimal solution of (7), there exists a corresponding Pareto optimal solution of (3). Hence, we can obtain different Pareto optimal solutions by setting different weight vectors.

*3.4. Discrete Differential Evolution Algorithm.* Differential evolution (DE) [37] generally works on a population of candidate solutions, which are represented by floating point numbers. DE generates a mutated solution by adding a weighted difference between two randomly selected solutions to a third one. A trial solution is then generated by crossing the mutated and target solutions. The selection operator in DE determines whether the target solution can be retained in the next generation or not.

From the description above, it can be known that the traditional DE cannot be directly used in the multiobjective

scheduling of FMSs, since it is originally designed for solving continuous optimization problems and can only generate solutions of floating point numbers [38]. In this subsection, we propose a novel discrete DE (DDE) for generating the solutions to the single-objective optimization problems in our DMOEA/D.

The mutated solution at the $k$th generation $x_k^v$ is constructed based on 3 randomly selected solutions, $x_k^a$, $x_k^b$, and $x_k^c$. Let $D$ denote the hamming distance between solutions $x_k^b$ and $x_k^c$. In our DDE, the mutation operator is defined as follows:

$$x_k^v = \begin{cases} Mut_1\left(x_k^a, p_m\right), & \text{if } D = 0 \\ Mut_D\left(x_k^a, p_m\right), & \text{if } D > 0 \end{cases} \tag{8}$$

where $Mut_i$ is the mutation with $i$ iterations and $p_m$ is the mutation probability. At each iteration of $Mut_i(x_k^a, p_m)$, the algorithm generates a random number $r \in [0, 1]$. If $r < p_m$, select one job in $x_k^a$ randomly and insert it to another position in $x_k^a$; otherwise, $x_k^a$ remains unchanged. Note that the mutation operator is iterated at least once, even if there is no difference between $x_k^b$ and $x_k^c$.

The trial solution $x_k^u$ in DDE algorithm is generated by crossing the target solution $x_k^i$ and the mutated solution $x_k^v$. GOX is used for achieving this. In our DDE, GOX is only applied in the second section of the solution, i.e., $S_o$. The length of the crossover string is chosen randomly between $N/4$ and $3N/4$.

*Example 3.* Assume that $x_1 = (S_r^1; S_o^1) = (w_1, w_2, w_3; J_1, J_1, J_3, J_2, J_3, J_3, J_1, J_2, J_1, J_2, J_2, J_2, J_1, J_3)$ is a target solution and $x_2 = (S_r^2; S_o^2) = (w_1, w_1, w_3; J_1, J_2, J_2, J_1, J_3, J_1, J_2, J_3, J_2, J_1, J_3, J_1, J_2, J_3)$ is a mutated solution. Then, according to Section 3.1, their corresponding sequences of operations are $\sigma(S_o^1) = (O_{11}, O_{12}, O_{31}, O_{21}, O_{32}, O_{33}, O_{13}, O_{22}, O_{14}, O_{23}, O_{24}, O_{25}, O_{15}, O_{34})$ and $\sigma(S_o^2) = (O_{11}, O_{21}, O_{22}, O_{12}, O_{31}, O_{13}, O_{23}, O_{32}, O_{24}, O_{14}, O_{33}, O_{15}, O_{25}, O_{34})$, respectively. Let the length of a crossover string be 7 and the crossover starts at operation $O_{31}$. So the crossover string in $S_o^2$ is $\sigma' = (J_3, J_1, J_2, J_3, J_2, J_1, J_3)$ and the corresponding operations are $O_{31}$, $O_{13}$, $O_{23}$, $O_{32}$, $O_{24}$, $O_{14}$, and $O_{33}$. Then, delete the jobs that correspond to these operations from $S_o^1$ and insert $\sigma'$ into $S_o^1$. The so obtained $S_o^3$ is $(J_1, J_1, J_2, \boldsymbol{J_3}, \boldsymbol{J_1}, \boldsymbol{J_2}, \boldsymbol{J_3}, \boldsymbol{J_2}, \boldsymbol{J_1}, \boldsymbol{J_3}, J_2, J_2, J_1, J_3)$. The first section of the trial solution is inherited directly from $x_1$. Then, we have the trial solution $x_k^u = (w_1, w_2, w_3; J_1, J_1, J_2, J_3, J_1, J_2, J_3, J_2, J_1, J_3, J_2, J_2, J_1, J_3)$. Note that the obtained trial solution $x_k^u$ is infeasible and hence should be repaired by Amending Algorithm.

The selection operator in DDE decides whether a trial solution $x_k^u$ should be a member of the population in the next generation. It is stated as follows:

$$x_k^i = \begin{cases} x_k^u, & \text{if } g^{te}\left(x_k^u \mid \lambda, z\right) \le g^{te}\left(x_{k-1}^i \mid \lambda, z\right) \\ x_{k-1}^i, & \text{otherwise} \end{cases} \tag{9}$$

where $z = (z_1, z_2, \ldots, z_L)^T$ and $z^i$ is the best objective $f_i$ found so far.

*3.5. Framework of DMOEA/D.* Let $\lambda^1, \lambda^2, \ldots, \lambda^K$ be a set of uniform spread weight vectors. DMOEA/D decomposes MOP into $K$ single-objective subproblems with the Tchebycheff approach, and the $j$th subproblem is

$$\begin{aligned} \text{Minimize} \quad & g^{te}\left(x \mid \lambda^j, z^*\right) \\ & = \max_{i \in \mathbb{Z}_L} \left\{\lambda_i^j \left(f_i(x) - z_i^*\right)\right\}^T \\ \text{Subject to} \quad & x \in \Pi \end{aligned} \tag{10}$$

where $\lambda^j = (\lambda_1^j, \lambda_1^j, \ldots, \lambda_i^j)^T$. Given a weight vector $\lambda^j$, define the neighbourhood of $\lambda^j$ as a set of its closest weight vectors in $\{\lambda^1, \lambda^2, \ldots, \lambda^K\}$. Then, all the subproblems that correspond to these weight vectors in the neighbourhood of $\lambda^j$ constitute the neighbourhood of the $j$th subproblem. The population contains the best found solutions for all the subproblems. While DMOEA/D optimizes a subproblem, it only exploits the current solutions to the neighbourhood of the subproblem.

In each generation, our proposed DMOEA/D using a Tchebycheff approach contains a population of $K$ solutions $x^1, \ldots, x^K \in \Pi$, where $x^j$ is the current solution to the $j$th subproblem and an external population (EP) for storing nondominated solutions found in DMOEA/D. Then, our proposed DMOEA/D can be stated as in Algorithm 1.

From the algorithm, it can be known that $E(j)$ contains the indexes of $H$ closest vectors of $\lambda^j$. This paper uses the Euclidean distance to evaluate how close any two weight vectors are. Hence, $\lambda^j$ itself is its closest vector, $j \in E(j)$. The $k$th subproblem is in the neighbourhood of the $j$th subproblem if $k \in E(j)$. While considering the $j$th subproblem in DMOEA/D, since $x^a$, $x^b$, and $x^c$ are the best found solutions to the respective neighbours of the $j$th subproblem, the reproduced solution $x^u$ based on them should be a promising one to the $j$th subproblem. Then, amend $x^u$ by using Amending Algorithm [17]. The so-obtained solution $x^u$ is therefore feasible and probably better for the neighbours of the $j$th subproblem. Then, for each neighbour $j^h$ of the $j$th subproblem, replace $x^{j^h}$ with $x^u$ if $x^u$ is better for the $j$th subproblem. The external population EP is also updated based on the new solutions.

Since obtaining the accurate reference point $z^*$ is usually very time-consuming, we use $z$ as a substitute of $z^*$ in $g^{te}$. $z$ is initialized by a problem-specific method and updated according to the quality of generated trial solutions.

## 4. Illustrative Examples

To the authors' knowledge, there is still no work reported for the multiobjective scheduling of deadlock-prone FMSs. Thus, to demonstrate the effectiveness of DMOEA/D, we compare it with a NSGA-II (one of the best known MOEAs) based MOEA. Since the DAP used in Section III-A is also embedded, we rename the compared algorithm as DNSGA-II.

**Input**
$K$: the number of sub-problems used in DMOEA/D;
$\lambda^1, \lambda^2, \ldots, \lambda^K$: the set of uniform spread weight vectors;
$H$: the size of neighborhood of each weight vector;
**Initialize**
Set EP $= \varnothing$, initialize population $x^1, \ldots, x^K \in \Pi$;
Set $F^j = F(x^j)$, $j \in \mathbb{Z}_K$, initialize $z = (z_1, z_2, \ldots, z_L)^T$, where $z_i = \min_{x \in \Pi} f_i(x)$;
Compute the Euclidean distances between any two weight vectors and figure out the $H$ closest ones of each weight vector;
Set $E(j) = \{j^1, j^2, \ldots, j^H\}$, where $\lambda^{j^1}, \lambda^{j^2}, \ldots, \lambda^{j^H}$ are the $H$ closest weight vectors to $\lambda^j$.
**While**(the stopping criterion is not met) {
  **For** ($j = 1, 2, \ldots, K$) {
    Randomly select three different neighbors $x^a$, $x^b$, and $x^c$ from $E(j)$;
    Generate the mutated solution $x^v$ from $x^a$, $x^b$, and $x^c$;
    Generate the trial solution $x^u$ from $x^j$ and $x^v$; Amend $x^u$;
    **For** ($i = 1, 2, \ldots, L$) {
      **If** ($z_i < f_i(x^u)$) {
        Set $z_i = f_i(x^u)$; }
    } // End **For**
    **For** (each index $j^h \in E(j)$) {
      **If**($g^{te}(x^u \mid \lambda^{j^h}, z) \leq g^{te}(x^{j^h} \mid \lambda^{j^h}, z)$) {
        Set $x^{j^h} = x^u$; $F^{j^h} = F(x^u)$; }
    } // End **For**
    Remove all the vectors dominated by $F(x^u)$ from EP;
    **If**(there is no vector dominates $F(x^u)$ and $F(x^u)$ do not exist in EP) {
      Add $F(x^u)$ into EP; }
  } // End **For**
} // End **While**
**Output** EP

Algorithm 1: Algorithm DMOEA/D.

The stop criteria of two tested algorithms are all set as 1000 generations. The population size is set as $K = 100$. For DMOEA/D, the weight vectors are determined by a parameter $I$. Each weight vector chooses a value from $\{0/I, 1/I, \ldots, I/I\}$ and the number of the weight vectors is $K = C_{I+L-1}^{L-1}$. For problems with 2-objective, $I$ is set to 99 since $C_{99+2-1}^{2-1} = K = 100$; for problems with 3-objective, $I$ is set to 13 since $C_{13+3-1}^{3-1} = 105 > K = 100$ and DMOEA/D randomly chooses 100 values out of 105. The number of neighbourhoods is $H = 20$. The cross rate $p_c$ and the mutation rate $p_m$ are set as 0.8 and 0.2, respectively. All the algorithms are coded by C++ and simulated on a desktop PC with 3.2 GHz processor and 8 GB memory.

*4.1. Performance Metrics.* The number of Pareto solutions (*NPS*) obtained in a run is an important metric for MOEA. More Pareto solutions mean more candidate schedules for decision makers and therefore have better performance.

Mean ideal distance (*MID*) evaluates how close the solutions on a PF to the ideal point (often referred to point (0, 0)). It can be defined as

$$MID = \sum_{1 \leq j \leq NPS} \frac{\Delta_j}{NPS} \tag{11}$$

where $\Delta_j = \sqrt{\sum_{1 \leq i \leq L} f_{ji}^2}$ and $f_{ji}$ is the $i$th objective of the $j$th Pareto solution. Less value of *MID* indicates better performance.

The rate of achievement to objectives simultaneously (*RAS*) evaluates the closeness of objectives. It can be represented as

$$RAS = \sum_{1 \leq j \leq NPS} \frac{\left(\sum_{1 \leq i \leq L} \left(f_{ji} / \min_{1 \leq i \leq L} f_{ji} - 1\right)\right)}{NPS} \tag{12}$$

The lower *RAS* value there is, the better solution quality we have.

*4.2. Case 1.* In this subsection, the FMS example in *Example* 1 is used to test the performance of the algorithms. The processing time of operations is listed in Table 2. 20 instances (FMS01- FMS20) designed by [39] are tested.

Firstly, we study the 2-objective problem with respect to makespan and mean completion time. The scheduling results of FMS01- FMS20 are shown in Table 3. The algorithms make 10 independent runs for each instance, and the metrics are averaged.

From Table 3, it can be seen that DMOEA/D outperforms DNSGA-II on all 3 metrics for the 2-objective problem. DMOEA/D achieves more *NPS* in 16 instances out of 20.

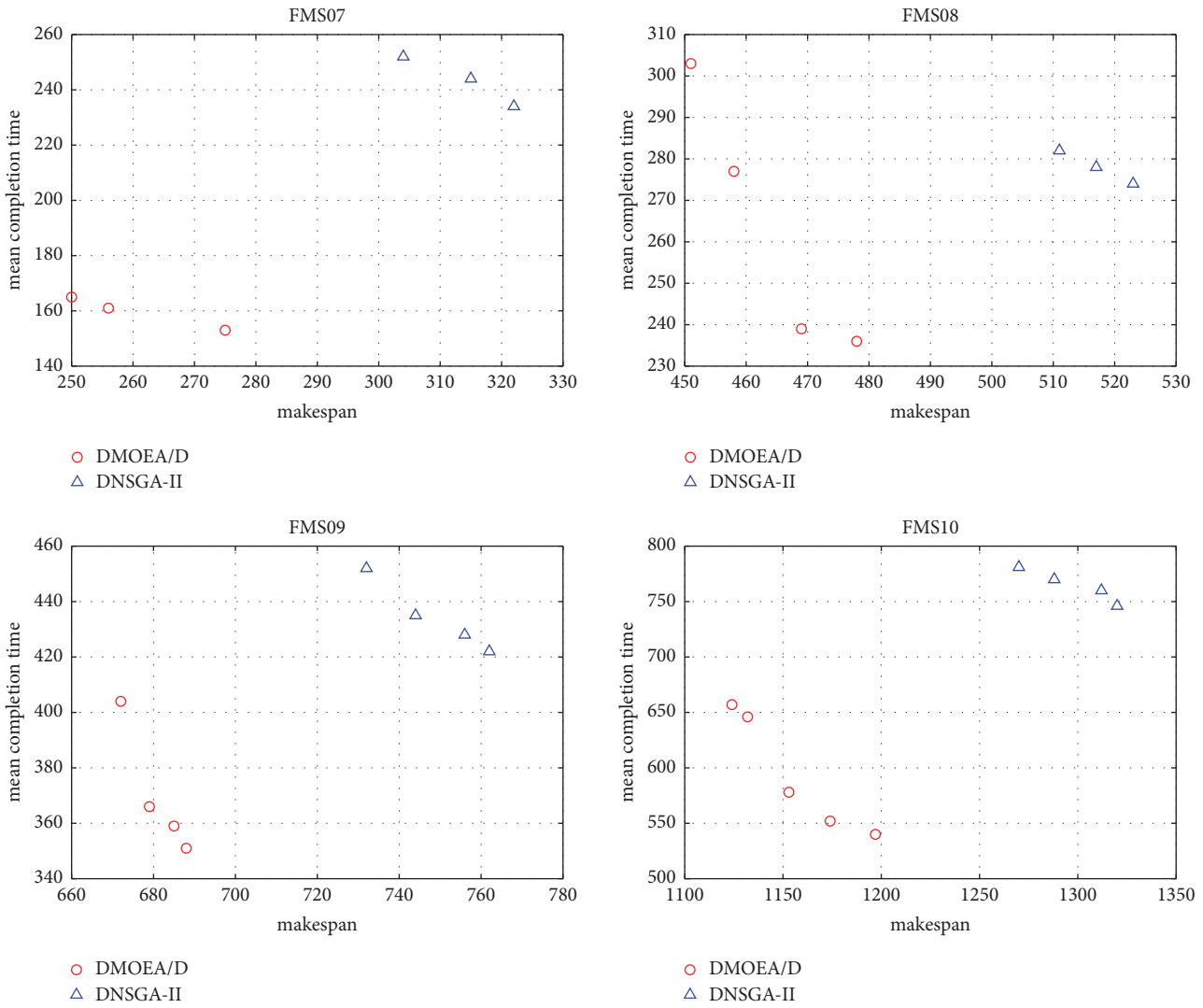



Figure 3: Nondominated solutions of 2-objective problem on instances FMS07-FMS10.

Table 2: Processing time of operations for the FMS in *Example* 1.

| $q_1$ | | $q_2$ |
|---|---|---|
| $w_1$ | $w_2$ | $w_3$ |
| $d(O_{11})$: 25 | $d(O_{21})$: 25 | $d(O_{31})$: 26 |
| $d(O_{12})$: 23 | $d(O_{22})$: 20 | $d(O_{32})$: 21 |
| $d(O_{13})$: 27 | $d(O_{23})$: 27 | $d(O_{33})$: 24 |

This means that DMOEA/D can obtain more Pareto solutions than DNSGA-II for the studied problem. For the metric *MID*, DMOEA/D significantly outperforms DNSGA-II since it obtains less *MID* values for all 20 instances. This indicates that DMOEA/D gets a better convergence than DNSGA-II. For the metric *RAS*, it seems that the solutions of DMOEA/D have better closeness of objectives as it obtains lower *RAS* values than DNSGA-II in 17 instances out of 20.

Figure 3 shows a part of nondominated solutions of 2-objective on FMS02-FMS05. We can see that most solutions found by DMOEA/D dominate the ones found by DNSGA-II. DMOEA/D achieves a better approximation to PF than DNSGA-II does.

Then, the 3-objective problem is studied. The scheduling results are listed in Table 4. Each instance makes 10 independent runs and the metrics are averaged.

As seen in Table 4, the results for the 3-objective problem seem in accordance with the conclusions for 2-objective problem. DMOEA/D generally obtains more Pareto solutions (higher *NPS* values in 18 of 20 instances) than DNSGA-II for the studied problem. DMOEA/D also has a better convergence than DNSGA-II, since it obtains lower *MID* values in all 20 instances. The solutions obtained by DMOEA/D have better closeness of objectives as it obtains lower *RAS* values in 19 instances out of 20.

Note that the *RAS* values of instances with very few jobs (FMS01, FMS06, FMS11, and FMS16) are much higher than the ones of other instances. This is due to the fact that a schedule with very few jobs usually has very little tardiness

Table 3: Scheduling results of 2-objective problem (FMS01-FMS20).

| Instances | *NPS* | | *MID* | | *RAS* | |
|---|---|---|---|---|---|---|
| | DNSGA-II | DMOEA/D | DNSGA-II | DMOEA/D | DNSGA-II | DMOEA/D |
| FMS01 | **2.1** | 2.0 | 218.33 | **199.24** | 0.80 | **0.76** |
| FMS02 | 2.0 | **2.7** | 412.45 | **350.20** | 0.84 | **0.79** |
| FMS03 | 2.3 | **2.4** | 657.22 | **584.42** | 0.91 | **0.87** |
| FMS04 | **2.6** | **2.6** | 990.17 | **875.01** | 1.04 | **0.98** |
| FMS05 | 2.8 | **3.1** | 1644.12 | **1436.08** | **1.08** | 1.09 |
| FMS06 | 2.0 | **2.3** | 190.50 | **165.82** | 0.57 | **0.55** |
| FMS07 | 2.8 | **3.9** | 361.71 | **302.64** | 0.61 | **0.57** |
| FMS08 | 3.3 | **3.8** | 602.29 | **537.69** | 0.72 | **0.65** |
| FMS09 | 3.8 | **4.2** | 917.27 | **827.27** | 1.05 | **0.97** |
| FMS10 | 3.9 | **4.8** | 1525.18 | **1354.03** | 1.10 | **1.04** |
| FMS11 | **2.2** | 1.9 | 158.93 | **132.09** | 0.55 | **0.41** |
| FMS12 | 2.4 | **2.5** | 323.24 | **257.43** | **0.56** | 0.58 |
| FMS13 | 3.1 | **4.0** | 568.74 | **516.87** | 0.72 | **0.70** |
| FMS14 | 4.2 | **4.6** | 858.91 | **779.25** | 0.88 | **0.83** |
| FMS15 | 4.5 | **4.7** | 1437.07 | **1320.08** | 0.98 | **1.04** |
| FMS16 | **1.8** | **1.8** | 129.86 | **118.01** | 0.52 | **0.43** |
| FMS17 | 2.4 | **3.0** | 242.64 | **199.14** | 0.55 | **0.54** |
| FMS18 | 3.0 | **3.2** | 547.16 | **508.87** | **0.72** | 0.80 |
| FMS19 | 3.4 | **4.2** | 827.83 | **753.60** | 0.90 | **0.92** |
| FMS20 | 3.6 | **4.3** | 1358.92 | **1182.49** | 0.97 | **0.99** |

time, and the objective function mean tardiness time is thus much smaller than the other two.

The computational times of DNSGA-II and DMOEA/D on FMS01-FMS20 are almost at the same level. For the 2-objectives problem, the computational times for instances with 10, 20, 40, 60, and 100 jobs are around 5s, 10s, 25s, 40s, and 60s, respectively. For the 3-objectives problem, the computational times for instances with 10, 20, 40, 60, and 100 jobs are around 8s, 15s, 30s, 40s, and 65s, respectively.

*4.3. Case 2.* In this subsection, we use a widely researched FMS example to test the performance of our algorithms. This FMS consists of 3 robots $r_1$-$r_3$, 4 machines $m_1$-$m_4$, and can process 3 types of jobs $q_1$-$q_3$. Its PNS with the initial marking of In01 is shown in Figure 4 and the processing time of operations is listed in Table 5. 16 instances (In01-In16) designed by [18] are tested.

The 2-objective problem is studied first. The scheduling results of 16 instances are shown in Table 6. The algorithms make 10 independent runs for each instance and the metrics are averaged.

From Table 6, we can see that DMOEA/D outperforms DNSGA-II on all 3 metrics. It can obtain more Pareto solutions than DNSGA-II in 15 instances out of 16. DMOEA/D also gets a better convergence than DNSGA-II since it obtains less *MID* values in 15 instances out of 16. Moreover, the solutions of DMOEA/D have better closeness of objectives than DNSGA-II, since DMOEA/D obtains lower *RAS* values in 14 instances out of 16.

Figure 5 shows a part of nondominated solutions of 2-objective on instances In01-In04. From Figure 5, we can see that the solutions found by DMOEA/D can dominate the ones obtained by DNSGA-II.

For the 3-objective problem, the scheduling results of In01-In16 are listed in Table 7. Each instance makes 10 independent runs and the metrics are averaged.

As seen in Table 7, DMOEA/D obtains more Pareto solutions than DNSGA-II in 11 of 16 instances. DMOEA/D also has a better convergence than DNSGA-II, since it obtains lower *MID* values in all 16 instances. The solutions obtained by DMOEA/D have better closeness of objectives since it obtains lower *RAS* values in all 16 instances.

The computational times of DNSGA-II and DMOEA/D on In01-In16 are almost at the same level. For the 2-objectives problem, the computational times for instances with 28, 40, 50, and 60 jobs are around 20s, 55s, 90s, and 120s, respectively. For the 3-objective problem, the computational times for instances with 28, 40, 50, and 60 jobs are around 25s, 65s, 95s, and 130s, respectively.

## 5. Conclusions

This paper studies the multiobjective scheduling problem of deadlock-prone FMSs. Combining decomposition approaches and DDE, a novel scheduling algorithm called DMOEA/D is proposed based on the PN model of the studied system. DMOEA/D can decompose a multiobjective scheduling problem into a certain number of single-objective subproblems and solves all the subproblems in a single run. The solutions of the subproblems are reproduced by a new proposed DDE algorithm. The mutation operator of DDE is based on the hamming distance between two selected

Table 4: Scheduling results of 3-objective problem (FMS01-FMS20).

| Instances | *NPS* | | *MID* | | *RAS* | |
|---|---|---|---|---|---|---|
| | DNSGA-II | DMOEA/D | DNSGA-II | DMOEA/D | DNSGA-II | DMOEA/D |
| FMS01 | 3.2 | **4.4** | 246.15 | **218.61** | 10.47 | **8.54** |
| FMS02 | 3.9 | **5.5** | 486.50 | **423.76** | 4.42 | **2.18** |
| FMS03 | **4.3** | 4.1 | 812.69 | **703.04** | 1.89 | **1.41** |
| FMS04 | 4.5 | **4.7** | 1184.47 | **1035.12** | 2.24 | **1.79** |
| FMS05 | 4.4 | **5.2** | 1982.92 | **1802.27** | 2.37 | **2.02** |
| FMS06 | 3.5 | **4.8** | 207.96 | **174.29** | 21.26 | **19.82** |
| FMS07 | 4.5 | **5.0** | 398.74 | **341.40** | 3.10 | **2.87** |
| FMS08 | 4.4 | **6.3** | 739.06 | **621.45** | 1.92 | **1.27** |
| FMS09 | 5.8 | **5.9** | 1115.63 | **965.83** | 2.01 | **1.61** |
| FMS10 | 5.5 | **6.0** | 1870.84 | **1685.44** | 2.24 | **1.96** |
| FMS11 | **3.3** | 2.5 | 170.02 | **137.77** | 30.36 | **28.23** |
| FMS12 | 3.5 | **3.6** | 320.57 | **263.89** | 3.72 | **3.47** |
| FMS13 | 3.9 | **4.9** | 671.11 | **589.06** | 1.82 | **1.08** |
| FMS14 | 5.8 | **6.0** | 1054.72 | **946.01** | 1.87 | **1.62** |
| FMS15 | 5.7 | **6.2** | 1759.81 | **1594.49** | 2.19 | **1.94** |
| FMS16 | 2.2 | **2.6** | 144.56 | **112.42** | **40.66** | 41.47 |
| FMS17 | 3.2 | **2.5** | 247.30 | **191.23** | 12.02 | **10.76** |
| FMS18 | 4.0 | **4.8** | 617.54 | **542.48** | 1.14 | **0.93** |
| FMS19 | 5.1 | **5.2** | 977.95 | **856.20** | 2.06 | **1.47** |
| FMS20 | 5.0 | **5.9** | 1668.03 | **1576.65** | 2.38 | **2.04** |

Table 5: Processing times of operations for the FMS in Figure 4.

| $q_1$ | $q_2$ | | $q_3$ |
|---|---|---|---|
| $w_1$ | $w_2$ | $w_3$ | $w_4$ |
| $d(O_{11})$: 8 | $d(O_{21})$: 4 | $d(O_{21})$: 4 | $d(O_{41})$: 5 |
| $d(O_{12})$: 34 | $d(O_{22})$: 32 | $d(O_{32})$: 23 | $d(O_{42})$: 22 |
| $d(O_{13})$: 5 | $d(O_{23})$: 8 | $d(O_{33})$: 6 | $d(O_{43})$: 4 |
| | $d(O_{24})$: 38 | $d(O_{34})$: 20 | $d(O_{44})$: 17 |
| | $d(O_{25})$: 5 | $d(O_{25})$: 5 | $d(O_{45})$: 6 |

Table 6: Scheduling results of 2-objective problem (In01-In16).

| Instances | *NPS* | | *MID* | | *RAS* | |
|---|---|---|---|---|---|---|
| | DNSGA-II | DMOEA/D | DNSGA-II | DMOEA/D | DNSGA-II | DMOEA/D |
| In01 | 1.4 | **1.6** | 465.13 | **384.32** | **0.69** | 0.73 |
| In02 | 1.3 | **2.0** | 646.31 | **563.24** | 0.81 | **0.78** |
| In03 | 1.5 | **1.7** | 886.41 | **752.05** | 0.99 | **0.81** |
| In04 | 1.7 | **2.1** | 1089.87 | **920.63** | 1.08 | **0.83** |
| In05 | 1.4 | **1.7** | 396.37 | **366.64** | **0.68** | **0.68** |
| In06 | 1.6 | **1.6** | 526.92 | **523.22** | 0.86 | **0.72** |
| In07 | 1.5 | **1.5** | **691.50** | 699.44 | 0.95 | **0.83** |
| In08 | 1.5 | **1.9** | 882.66 | **839.75** | 0.99 | **0.82** |
| In09 | 1.6 | **1.8** | 339.47 | **298.18** | 0.65 | **0.58** |
| In10 | 1.5 | **1.6** | 476.35 | **451.60** | 0.77 | **0.62** |
| In11 | 1.7 | **2.0** | 585.71 | **559.66** | 0.84 | **0.68** |
| In12 | 1.6 | **1.8** | 719.93 | **655.98** | 0.83 | **0.69** |
| In13 | 1.4 | **1.5** | 328.30 | **263.18** | 0.61 | **0.54** |
| In14 | 1.6 | **1.7** | 465.32 | **416.40** | 0.71 | **0.61** |
| In15 | 1.6 | **1.5** | 550.88 | **514.23** | 0.79 | **0.66** |
| In16 | **1.7** | **1.7** | 687.50 | **654.51** | 0.85 | **0.72** |

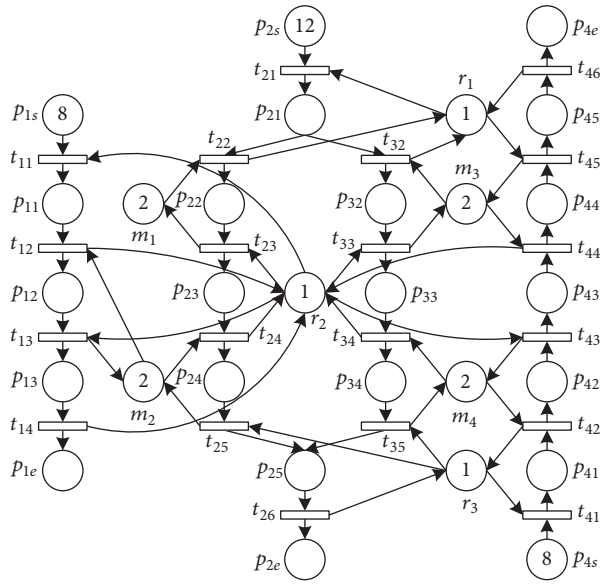



FIGURE 4: PNS model of the studied FMS (In01).

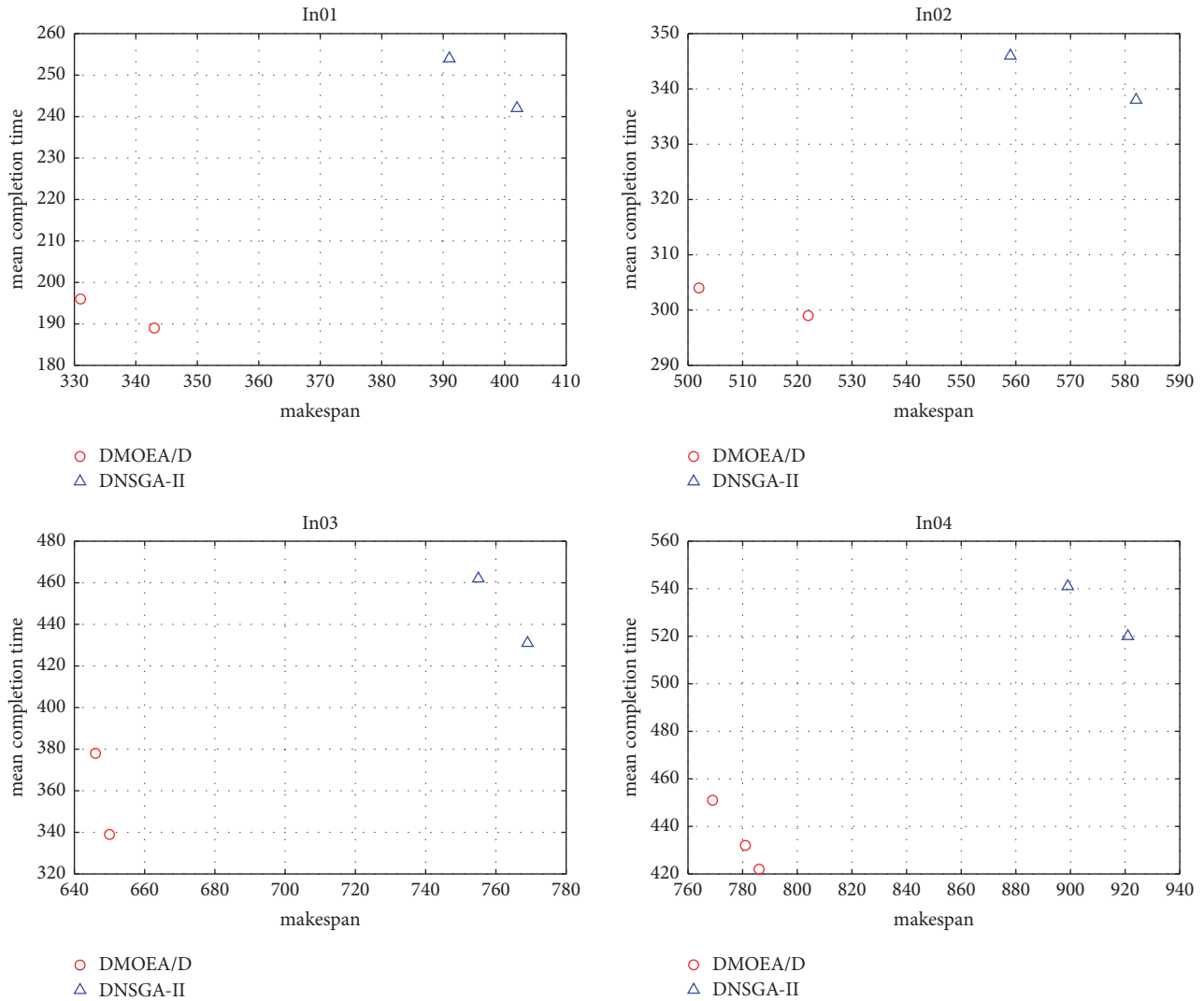



FIGURE 5: Non-dominated solutions of 2-objective problem on instances In01-In04.

Table 7: Scheduling results of 3-objective problem (In01-In16).

| Instances | *NPS* | | *MID* | | *RAS* | |
|---|---|---|---|---|---|---|
| | DNSGA-II | DMOEA/D | DNSGA-II | DMOEA/D | DNSGA-II | DMOEA/D |
| In01 | 1.8 | **2.2** | 467.16 | **412.22** | 4.32 | **3.53** |
| In02 | 1.6 | **1.8** | 793.22 | **662.40** | 3.80 | **1.98** |
| In03 | **2.1** | 2.0 | 1061.19 | **847.26** | 3.52 | **1.79** |
| In04 | **2.1** | **2.1** | 1299.54 | **1056.08** | 3.06 | **1.43** |
| In05 | 1.6 | **1.8** | 423.02 | **372.25** | 4.86 | **3.08** |
| In06 | **2.0** | **2.0** | 674.95 | **577.32** | 4.42 | **2.06** |
| In07 | 2.2 | **2.4** | 956.23 | **725.51** | 3.93 | **1.90** |
| In08 | 2.3 | **2.5** | 1129.71 | **909.89** | 3.69 | **1.58** |
| In09 | **1.9** | **1.9** | 354.21 | **319.51** | 5.23 | **3.22** |
| In10 | 2.5 | **2.6** | 562.76 | **460.27** | 4.97 | **2.38** |
| In11 | 1.9 | **2.1** | 785.41 | **636.38** | 4.48 | **1.76** |
| In12 | 2.5 | **2.8** | 977.70 | **761.75** | 4.01 | **1.61** |
| In13 | **2.1** | 1.8 | 309.61 | **263.67** | 5.82 | **4.58** |
| In14 | 2.3 | **2.5** | 500.26 | **432.11** | 4.78 | **2.49** |
| In15 | 2.5 | **2.7** | 729.43 | **572.32** | 4.73 | **2.01** |
| In16 | 2.1 | **2.4** | 891.41 | **681.56** | 4.24 | **1.79** |

solutions. GOX is used as the crossover operator. Two benchmark examples are used to test the DMOEA/D. Computational results demonstrate that our proposed DMOEA/D can outperform DNSGA-II for both 2-objective and 3-objective problems on the studied FMSs.

The advantage of DMOEA/D over DNSGA-II may attribute to the following: (1) DMOEA/D optimizes several single-objective optimization problems rather than directly solving the MOP, and hence the diversity of the population is easier to maintain, and more evenly distributed Pareto solutions can thus be obtained; and (2) coevolution mechanism between subproblems is used in DMOEA/D, and, hence, DMOEA/D can converge faster and obtain better Pareto solutions than DNSGA-II.

For the future work, one research direction is to introduce other efficient local optimization strategies or heuristics to improve the search capability of the proposed algorithm. Another direction is to extend the proposed method to solve the scheduling problem of FMSs with other objectives or constraints, such as energy consumptions, unreliable resources, and maintenance activities, which may be proved both interesting and useful.

## Data Availability

(1). All data generated by the simulation in this study are included within the article in Tables 3, 4, 6, and 7. (2). Previously reported data in [17, 39] are used to support our study, and they are included in Tables 2 and 5. (3). No data were used to support this study.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported by the National Science Foundation of PR China [grant number 61573278].

## References

[1] A. Alvarado-Iniesta, J. L. García-Alcaraz, M. Piña-Monarrez, and L. Pérez-Domínguez, "Multiobjective optimization of torch brazing process by a hybrid of fuzzy logic and multiobjective artificial bee colony algorithm," *Journal of Intelligent Manufacturing*, vol. 27, no. 3, pp. 631–638, 2016.

[2] J. C. Leyva López, J. J. Solano Noriega, J. L. García Alcaraz, and D. A. Gastélum Chavira, "Exploitation of a medium-sized fuzzy outranking relation based on multi-objective evolutionary algorithms to derive a ranking," *International Journal of Computational Intelligence Systems*, vol. 9, no. 4, pp. 745–764, 2016.

[3] R. Domingo, B. de Agustina, and M. M. Marín, "A multiresponse optimization of thrust forces, torques, and the power of tapping operations by cooling air in reinforced and unreinforced polyamide PA66," *Sustainability*, vol. 10, no. 3, p. 889, 2018.

[4] V. K. Chawla, A. K. Chanda, and S. Angra, "Sustainable multiobjective scheduling for automatic guided vehicle and flexible manufacturing system by a grey wolf optimization algorithm," *International Journal of Data and Network Science*, vol. 2, no. 1, pp. 27–40, 2018.

[5] W. K. Mashwani and A. Salhi, "Multiobjective memetic algorithm based on decomposition," *Applied Soft Computing*, vol. 21, pp. 221–243, 2014.

[6] 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.

[7] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE*

*Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[8] Q. Zhang and H. Li, "MOEA/D: a multiobjective evolutionary algorithm based on decomposition," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 6, pp. 712–731, 2007.

[9] D. Chen, F. Zou, R. Lu, L. Yu, Z. Li, and J. Wang, "Multi-objective optimization of community detection using discrete teaching-learning-based optimization with decomposition," *Information Sciences*, vol. 369, pp. 402–418, 2016.

[10] F. Zhao, Z. Chen, J. Wang, and C. Zhang, "An improved MOEA/D for multi-objective job shop scheduling problem," *International Journal of Computer Integrated Manufacturing*, vol. 30, no. 6, pp. 616–640, 2017.

[11] V. K. Chawla, A. K. Chanda, and S. Angra, "Automatic guided vehicles fleet size optimization for flexible manufacturing system by grey wolf optimization algorithm," *Management Science Letters*, vol. 8, no. 2, pp. 79–90, 2018.

[12] S. Angra, A. K. Chanda, and V. K. Chawla, "Comparison and evaluation of job selection dispatching rules for integrated scheduling of multi-load automatic guided vehicles serving in variable sized flexible manufacturing system layouts: a simulation study," *Management Science Letters*, vol. 8, no. 4, pp. 187–200, 2018.

[13] G. Demesure, M. Defoort, A. Bekrar, D. Trentesaux, and M. Djemai, "Decentralized motion planning and scheduling of AGVs in an FMS," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 4, pp. 1744–1752, 2018.

[14] C. Su, X. Guan, Y. Du, X. Huang, and M. Zhang, "Toward capturing heterogeneity for inferring diffusion networks: a mixed diffusion pattern model," *Knowledge-Based Systems*, vol. 147, pp. 81–93, 2018.

[15] H. H. Xiong and M. Zhou, "Scheduling of semiconductor test facility via petri nets and hybrid heuristic search," *IEEE Transactions on Semiconductor Manufacturing*, vol. 11, no. 3, pp. 384–393, 1998.

[16] I. B. Abdallah, H. A. ElMaraghy, and T. ElMekkawy, "Deadlock-free scheduling in flexible manufacturing systems using Petri nets," *International Journal of Production Research*, vol. 40, no. 12, pp. 2733–2756, 2002.

[17] K. Xing, L. Han, M. Zhou, and F. Wang, "Deadlock-free genetic scheduling algorithm for automated manufacturing systems based on deadlock control policy," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, no. 3, pp. 603–615, 2012.

[18] L. Han, K. Xing, X. Chen, H. Lei, and F. Wang, "Deadlock-free genetic scheduling for flexible manufacturing systems using Petri nets and deadlock controllers," *International Journal of Production Research*, vol. 52, no. 5, pp. 1557–1572, 2014.

[19] O. T. Baruwa, M. A. Piera, and A. Guasch, "Deadlock-free scheduling method for flexible manufacturing systems based on timed colored petri nets and anytime heuristic search," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 5, pp. 831–846, 2015.

[20] J. Luo, K. Xing, M. Zhou, X. Li, and X. Wang, "Deadlock-free scheduling of automated manufacturing systems using petri nets and hybrid heuristic search," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 3, pp. 530–541, 2015.

[21] X. Li, K. Xing, Y. Wu, X. Wang, and J. Luo, "Total energy consumption optimization via genetic algorithm in flexible manufacturing systems," *Computers & Industrial Engineering*, vol. 104, pp. 188–200, 2017.

[22] X. Li, K. Xing, M. Zhou, X. Wang, and Y. Wu, "Modified dynamic programming algorithm for optimization of total energy consumption in flexible manufacturing systems," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 2, pp. 691–705, 2019.

[23] X. Wang, K. Xing, X. Li, and J. Luo, "An estimation of distribution algorithm for scheduling problem of flexible manufacturing systems using Petri nets," *Applied Mathematical Modelling: Simulation and Computation for Engineering and Environmental Systems*, vol. 55, pp. 776–788, 2018.

[24] C. Bierwirth, "A generalized permutation approach to job shop scheduling with genetic algorithms," *OR Spectrum*, vol. 17, no. 2-3, pp. 87–92, 1995.

[25] J.-I. Latorre-Biel, E. Jiménez-Macías, J. Blanco-Fernández, and J. C. Sáenz-Díez, "Optimal design of an olive oil mill by means of the simulation of a petri net model," *International Journal of Food Engineering*, vol. 10, no. 4, pp. 573–582, 2014.

[26] J. I. Latorre-Biel, E. Jiménez-Macías, J. Blanco-Fernández, E. Martínez-Cámara, J. C. Sáenz-Díez, and M. Pérez-Parte, "Decision support system, based on the paradigm of the petri nets, for the design and operation of a dairy plant," *International Journal of Food Engineering*, vol. 11, no. 6, pp. 767–776, 2015.

[27] J.-I. Latorre-Biel, E. Jiménez-Macías, M. Pérez-De-La-Parte, J. C. Sáenz-Díez, E. Martínez-Cámara, and J. Blanco-Fernández, "Compound petri nets and alternatives aggregation petri nets: two formalisms for decision-making support," *Advances in Mechanical Engineering*, vol. 8, no. 11, pp. 1–12, 2016.

[28] T. Murata, "Petri nets: properties, analysis and applications," *Proceedings of the IEEE*, vol. 77, no. 4, pp. 541–580, 1989.

[29] S.-Z. Zhao, P. N. Suganthan, and Q. Zhang, "Decomposition-based multiobjective evolutionary algorithm with an ensemble of neighborhood sizes," *IEEE Transactions on Evolutionary Computation*, vol. 16, no. 3, pp. 442–446, 2012.

[30] L. Ke, Q. Zhang, and R. Battiti, "MOEA/D-ACO: a multiobjective evolutionary algorithm using decomposition and ant colony," *IEEE Transactions on Cybernetics*, vol. 43, no. 6, pp. 1845–1859, 2013.

[31] C. Wang, Z. Ji, and Y. Wang, "A novel memetic algorithm based on decomposition for multiobjective flexible job shop scheduling problem," *Mathematical Problems in Engineering*, vol. 2017, Article ID 2857564, 20 pages, 2017.

[32] Y. Huang, M. Jeng, X. Xie, and S. Chung, "Deadlock prevention policy based on Petri nets and siphons," *International Journal of Production Research*, vol. 39, no. 2, pp. 283–305, 2001.

[33] L. Piroddi, R. Cordone, and I. Fumagalli, "Selective siphon control for deadlock prevention in Petri nets," *IEEE Transactions on Systems, Man and Cybernetics, Part A*, vol. 38, no. 6, pp. 1337–1348, 2008.

[34] K. Y. Xing, M. C. Zhou, H. X. Liu, and F. Tian, "Optimal Petri-net-based polynomial-complexity deadlock-avoidance policies for automated manufacturing systems," *IEEE Transactions on Systems, Man, and Cybernetics: Part A: Systems and Humans*, vol. 39, no. 1, pp. 188–199, 2009.

[35] Y. Feng, K. Xing, M. Zhou, X. Wang, and H. Liu, "Robust deadlock prevention for automated manufacturing systems with unreliable resources by using general Petri nets," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–13, 2018.

[36] Y. X. Feng, K. Y. Xing, M. C. Zhou, F. Tian, and H. X. Liu, "Structural liveness analysis of automated manufacturing systems modeled by $S^4$PRs," *IEEE Transactions on Automation Science and Engineering*, 2019.

[37] R. Storn and K. Price, "Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, no. 4, pp. 341–359, 1997.

[38] Q.-K. Pan, M. F. Tasgetiren, and Y.-C. Liang, "A discrete differential evolution algorithm for the permutation flowshop scheduling problem," *Computers & Industrial Engineering*, vol. 55, no. 4, pp. 795–816, 2008.

[39] L. Han, K. Xing, X. Chen, and F. Xiong, "A Petri net-based particle swarm optimization approach for scheduling deadlock-prone flexible manufacturing systems," *Journal of Intelligent Manufacturing*, vol. 29, no. 5, pp. 1083–1096, 2018.

Hindawi
Complexity
Volume 2019, Article ID 6535726, 8 pages
https://doi.org/10.1155/2019/6535726

*Research Article*

# Reliability and Sensitivity Analysis Method for a Multistate System with Common Cause Failure

**Jinlei Qin and Zheng Li**

*Department of Computer, North China Electric Power University, Baoding, China*

Correspondence should be addressed to Jinlei Qin; jlqin717@163.com

Received 18 March 2019; Accepted 16 May 2019; Published 28 May 2019

Guest Editor: Rosario Domingo



With the increasing complexity of industrial products and systems, some intermediate states, other than the traditional two states, are often encountered during reliability assessments. A system with more than two states is called a multistate system (MSS) which has already become a general phenomenon in the components and/or systems. Moreover, common cause failure (CCF) often plays a very important role in the assessment of system reliability. A method is proposed to assess the reliability and sensitivity of an MSS with CCF. Some components are not only in a failure state that can cause failure itself, but also in a state that can cause the failure of other components with a certain probability. The components that are affected by one type of CCF make up some sets which can overlap on some components. Using the technology of a universal generating function (UGF), the CCF of a component can be incorporated in the expression of its UGF. Consequently, indices of reliability can be calculated based on the UGF expression of an MSS. Sensitivity analysis can help engineers to judge which type of CCF should be eliminated first under various resource limitations. Examples illustrate and validate this method.

## 1. Introduction

Common cause failures (CCFs) are the dependent failures of multiple components originating from a common cause or single occurrence or condition. CCFs can be important contributors to system unavailability or accident risk. For example, lightning events can cause outages of unprotected electronic equipment and voltage surges caused by inappropriate switching can lead to multiple failures of components in a power system. Recognition of the fact that a CCF will increase the joint-failure probability and then reduce system reliability has inspired many researchers to model and estimate the reliability or availability of a system with CCF. Some basic concepts and theories have been investigated and developed [1].

Two fundamental kinds of methods, termed implicit and explicit approaches, are used to incorporate CCFs in system reliability analysis [2]. The implicit approach proceeds as follows. The CCF is first ignored and the system logic is modeled with basic single-failure (component-level) events. The system reliability is given by algebraic probability expressions of the basic components, which are quantified to include the contribution of the CCF such that the system reliability or availability with the CCF can be expressed. This method has been studied in both binary- and multistate system (MSS) reliability [3]. In the explicit approach, the CCF is modeled as a basic cause-event in the system reliability block diagram or system fault tree, appearing as the repeated input to all elements or gates affected by the CCF. As the CCF can occur at random times in a redundant standby safety system, the expressions for the basic event probabilities of the explicit fault tree model have been developed [4].

In the context of imperfect fault coverage (IPC), an uncovered failure may often lead to complete system failure [5]. Aiming at this situation, other methods such as ordered binary decision diagrams and generalized reliability block diagrams have been suggested to assess system reliability [6]. In the phased-mission system with IPC and CCF, a new binary decision diagram approach to reliability analysis was suggested [7]. Myers demonstrated how the coverage effect can be computed using combinational and recursive technologies for four coverage models [8]. In the reliability

redundancy allocation problem three nonlinear optimization models mixing components with the inclusion of CCF events were addressed [9]. The reliability of an IPC system with functional or performance dependence was also investigated [10]. An extended object-oriented Petri net model was proposed for mission reliability simulation of a repairable phased-mission system with both external and internal CCFs [11]. Using the reliability block diagram method, the quantitative study of probability of failing safely (PFS) for a safety instrumented system showed that the CCF increases the PFS [12]. Under the assumption that a single-failure event may lead to simultaneous failure of multiple components, a nonparametric predictive inference for system reliability following the CCF of components was presented [13].

In many cases, the system and/or its components can function in some different states characterized by several performance levels. Such systems are often referred to as MSSs. The theory of the MSS was investigated by Murchland in 1975 [14]. Many researchers have analyzed the reliability which is often seen as an important measure of the MSS to provide the desired performance level [15]. MSS can also be subjected to the CCF which can lead to the failure of an entire system or subsystem [16].

Many studies have investigated the reliability and modeling of the MSS with CCF. For instance, an algorithm was proposed to evaluate the reliability of a complex nonrepairable series-parallel MSS with CCF under the assumption that the failure propagation time is a random value with a given distribution [17]. An MSS reliability analysis method incorporated a CCF into a Bayesian network [18] which is based on a well-defined theory of probabilistic reasoning and the ability to express the complex dependence between random variables. To reduce the computational complexity caused by the MSS, where the number of states will increase rapidly with the number of elements, a universal generating function (UGF) was adopted by Ushakov [19]. Further developments and applications have been studied in detail [20]. Recently, the reliability analysis method based on the UGF has been employed in many environments. A new reliability-evaluation methodology based on the UGF and a recursive algorithm was applied to a multistate weighted k-out-of-n system [21]. To analyze the reliability of the generalized linear multistate consecutively connected system, a UGF-based method was suggested [22]. Using UGF, the evaluation method of reliability characteristics for a nonrepair complex system has been presented [23].

Although much interest has been focused on methods of reliability analysis for systems with CCF, less attention has been paid to the MSS with CCF, which only occurs in several independent elements. The reliability of a transmission MSS considering the CCF effects has been analyzed based on Bayesian network model [18]. This model can clearly express the influence of CCF on system reliability without the computation of minimum cut sets or the determination of algebraic expression of unreliability. However, these parameter values used in $\beta$-factor model reduce its accuracy because these values are based on engineering experiences or published statistics of CCF. An implicit two-stage procedure is proposed to evaluate reliability of MSS with CCF [3]. This procedure can obtain the system reliability function in both analytic and numerical solution. But the procedure can be applied only to series-parallel MSS in which each component can belong only to one common cause group.

Aiming at the research limitations mentioned above, such as the inaccuracies of parameter values and nonoverlap of common cause group, an easily programmed recursive method is suggested for this situation, with the benefit of reduced computational complexity for reliability and performance distribution assessment. Using the sensitivity analysis method, it is easy to locate the optimal solution to eliminate the CCF and will help engineers to find the bottlenecks of system reliability design. This method is based on the application of UGF.

This paper is organized as follows. Section 2 formulates the MSS with CCF. In Section 3, based on the expression of the CCF, the algorithm to calculate the system reliability is presented. Several illustrative examples are shown in Section 4. Section 5 discusses the conclusions.

## 2. System Description

*2.1. Formulations of the MSS.* Under the framework of multistate modeling, an MSS consists of $n$ components connected in series-parallel. A characteristic of the MSS is that any component $i$ can be in one of $m_i + 1$ performance levels corresponding to states $s \in \{0, 1, 2, \ldots, m_i\}$, $i = 1, 2, \ldots, n$. State 0 is for the perfect function, state $m_i$ for complete failure and other intermediate states are for partial failure or degradation. Without loss of generality, the performance level at each state of component $i$ will take values from the set

$$\mathbf{h_i} = \left\{h_{i0}, h_{i1}, \ldots, h_{im_i}\right\}. \tag{1}$$

Then the space of all combinations of performance levels for all components can be given by

$$H^n = \mathbf{h_1} \times \mathbf{h}_2 \times \cdots \times \mathbf{h}_n, \tag{2}$$

where the operator $\times$ denotes the Cartesian product. The event of every performance level for component $i$ at any time is also assumed to a discrete random variable $H_i \in \mathbf{h_i}$. The probability set of $H_i$ can be expressed by

$$\mathbf{p}_i = \left\{p_{i0}, p_{i1}, \ldots, p_{im_i}\right\}. \tag{3}$$

Obviously $\sum_0^{m_i} p_{is} = 1$, because that the $m_i + 1$ states constitute a group of mutually exclusive events, i.e., the component $i$ can be in one and only one state from $s$ at any time. The relationship between the variable $H_i$ and performance $\mathbf{g_i}$ can be given by

$$p_{is} = \Pr\left\{H_i = h_{is}\right\}. \tag{4}$$

Another characteristic of the MSS is that it usually will consist of more than two states. Here, some external factors causing the failures or degradations of a system are not considered. Under this assumption, the performance levels of the MSS will be unambiguously ascertained by performance

levels of its components, whose states will consequently determine the states of the MSS.

Suppose that the MSS has $M + 1$ states and the variable $w_j$ is the performance level corresponding to system state $j \in \{0, 1, 2, \ldots, M\}$. The MSS performance level can be seen as a discrete random variable $W$ taking values from the set $L = \{w_0, w_1, \ldots, w_M\}$. Using (2), the system function can be written as

$$W = \phi\left(H_1, H_2, \ldots, H_n\right) : H^n \longrightarrow L. \tag{5}$$

The above formula is the map of the performance level space from all components to the system. Similarly, the probability of system performance can be expressed as

$$q_j = \Pr\left\{W = w_j\right\}, \quad j = 0, 1, \ldots, M. \tag{6}$$

From the analysis above, the keys of the MSS model are given as follows:

(i) The MSS consists of $n$ components connected in series-parallel.

(ii) Any component $i$ has $m_i + 1$ states. Each state's performance level and probability of a corresponding performance level can be expressed as (1) and (3), respectively.

(iii) The MSS will have $M + 1$ states deduced from the states of the $n$ components as defined in (5) and its performance level will take a value from $L = \{w_0, w_1, \ldots, w_M\}$.

(iv) The probability of the MSS in a corresponding state can be expressed by (6).

*2.2. Indices of Reliability.* According to the origin of the CCF, there are two types of cause: external and internal cause. The CCF caused by the latter is often called a propagated failure because it will affect other system components. In an MSS where the CCF has an internal cause, the failed components will propagate to other components which may be mutually independent or even overlap with one or more components. Then the system performance level will be undermined with an induced decrease in reliability. When the system performance level is reduced to a limit value $\theta$, which is often called system demand, the system will be seen as unacceptable. The reliability of the MSS with CCF can be defined as the probability that the system satisfies the value of $\theta$. From (6), one can obtain

$$R(\theta) = \sum_{j=0}^{M} q_j \Delta\left(w_j, \theta\right), \tag{7}$$

where

$$\begin{aligned} &\Delta\left(w_j, \theta\right) = 1, \quad w_j \geq \theta \\ &\Delta\left(w_j, \theta\right) = 0, \quad otherwise. \end{aligned} \tag{8}$$

Another important measure is the conditional expected performance $\Omega(\theta)$. It expresses the system's expected performance under the condition that the MSS is in an acceptable state. Having the system reliability $R(\theta)$, this measure can be calculated by

$$\Omega(\theta) = \frac{\left(\sum_{j=0}^{M} q_j w_j \Delta\left(w_j, \theta\right)\right)}{R(\theta)}. \tag{9}$$

To calculate these measures, the performance level distribution of a system should be first obtained according to (6). UGF has been proved an effective method for the reliability assessment of different types of MSS. In particular, the series-parallel system is likely to adopt it by the recursive method.

## 3. Analysis Method

The CCF caused by one component corresponds to the state $0'$ with the performance level $h_{i0'} = c$. The performance value can coincide with component performance in the state of local failure and can usually be seen as zero. When the component $i$ is in state $0'$, all the system components affected by this CCF will also be in the failure state with performance $c$. If one component cannot arouse a CCF, the probability mass function (PMF) of the corresponding state should be zeroed: $p_{i0'} = 0$. The UGF of component $i$ incorporating the CCF can be rewritten as

$$u_i(z) = \sum_{k_i=0}^{m_i} p_{ik_i} * z^{h_{ik_i}} + p_{i0'} * z^c. \tag{10}$$

The conditional PMF of any component $i$ that cannot fail due to the CCF can be represented by

$$u_i'(z) = \sum_{k_i=0}^{m_i} \frac{p_{ik_i}}{1 - p_{i0'}} * z^{h_{ik_i}}, \tag{11}$$

where $p_{i0'}$ is the probability of causing the CCF.

Suppose there are $b \leq n$ components that can independently and simultaneously cause the CCF. For component vector $\boldsymbol{\beta} = \{\beta(1), \beta(2), \ldots, \beta(r), \ldots, \beta(b)\}$, $r \in [1, b]$, the probability that the CCF originated from one component is expressed as $p_{\beta(r)0'}$. Since these CCFs can be combined independently, the number of combination is $2^b$.

For any combination $\delta$, $(0 \leq \delta \leq 2^b - 1)$, the CCF originates from component $\beta(r)$ if

$$\varepsilon(r) = \mathrm{mod}_2\left\lfloor \frac{\delta}{2^{r-1}} \right\rfloor = 1, \quad 1 \leq r \leq b. \tag{12}$$

After evaluating the above formula for $\delta = 0$ to $\delta = 2^b - 1$, the set of components corresponding to the combination $\delta$ can be obtained. For every combination $\delta$, the probability of the corresponding CCF is

$$\phi_\delta = \prod_{r=1}^{b} p_{\beta(r)0'}^{\varepsilon(r)} \left(1 - p_{\beta(r)0'}\right)^{1-\varepsilon(r)}, \quad \left(0 \leq \delta \leq 2^b - 1\right). \tag{13}$$

According to the properties of the CCF, other components failed because component $\beta(r)$ can be denoted by $S_{\beta(r)}$.

For every combination $\delta$, the set of components affected by the CCF and aroused the CCF can be obtained by

$$\mu_\delta = \bigcup_{r=1}^{b} S'_{\beta(r)}, \tag{14}$$

where

$$S'_{\beta(r)} = \begin{cases} S_{\beta(r)} \cup \{\beta(r)\}, & \varepsilon(r) = 1 \\ \varnothing, & \varepsilon(r) = 0 \end{cases}. \tag{15}$$

Given that one component fails due to the CCF, the conditional PMF of its performance level can be represented by $\psi(z) = z^c$, which indicates the component can be only in the failure state with performance level $c$.

When the CCF corresponding to combination $\delta$ occurs, all components in the set $\mu_\delta$ will come in the CCF mode, and their UGFs will be replaced by $\psi(z)$. However, the UGFs of components not belonging to the set $\mu_\delta$ must be represented by (11). Then the conditional PMF of the overall system performance will be obtained as $u_\delta(z)$ in the form of a UGF. From (13), the system UGF can be calculated as

$$U_{sys}(z) = \sum_{\delta=0}^{2^b-1} u_\delta(z) * \phi_\delta. \tag{16}$$

Based on the UGF of the whole system, the relevant indices of reliability can be assessed by combining (7) and (9).

Combined with the above analysis, the following steps should be adopted to realize the assessment:

*S1.* Incorporate the CCF into each component's UGF according to (10).

*S2.* Obtain the conditional UGF $u'_i(z)$ of component $i$ by (11).

*S3.* Determine all possible combinations of components aroused by the CCF.

*S4.* For every combination $\delta$, $(0 \le \delta \le 2^b - 1)$, fix its corresponding components which can inspire the CCF according to (12).

*S5.* Calculate the probability of every combination $\delta$ using (13).

*S6.* Determine the set of components affected by the CCF given one combination $\delta$ according to (14).

*S7.* Replace the UGF of all components belonging to set $\mu_\delta$ with $\psi(z)$, and use the UGF $u'_i(z)$ of the components not belonging to it.

*S8.* Based on the physical structure of the series-parallel MSS, represent the conditional UGF $u_\delta(z)$.

*S9.* Obtain the UGF of the entire system $U_{sys}(z)$, according to (16).

After $U_{sys}(z)$ has been obtained, the reliability indices can be further calculated according to (7) and (9). The following section will illustrate the above method.

Table 1: Performance distribution of components.

| Component | Perfect state | | Local failure | Failure to cause CCF |
|---|---|---|---|---|
| | p | h | p | p |
| C11 | 0.88 | 3 | 0.07 | 0.05 |
| C12 | 0.89 | 5 | 0.11 | - |
| C13 | 0.79 | 3 | 0.15 | 0.06 |
| C21 | 0.86 | 6 | 0.14 | - |
| C22 | 0.85 | 2 | 0.15 | - |

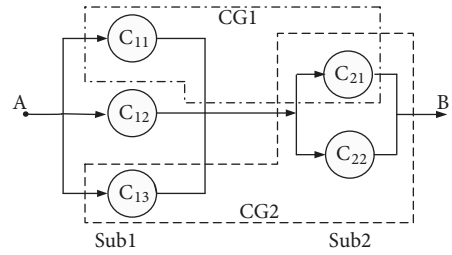



Figure 1: Series-parallel structure MSS.

## 4. Application Examples

*4.1. Basic Application.* A numerical example illustrates the above method for the reliability assessment of the MSS with CCF. A type of MSS with a series-parallel structure, as shown in Figure 1, is a typical configuration. In a flow transition system, matter such as oil, water, or steam is often transmitted from terminal A to terminal B by interconnected components.

For example, in the feeding water system of a power plant, components (pumps) are grouped into two subsystems; sub1 consists of C11, C12, and C13 connected in parallel, and sub2 consists of C21 and C22. Their parameters of performance distribution are listed in Table 1. As shown in this table, all components can fail with a given probability. But only components C11 and C13 can arouse the CCF. C11 can destroy component C21 owing to its nearby location and then constitute the CCF group CG1 with component C21. Component C13 will stimulate the failure of components C21 and C22 at the same time, to form the second CCF group CG2. The reasons for the CCF can be events such as fire or lightning.

According to the above steps, the UGF of every component can be obtained as follows:

$$\begin{aligned} u_{C_{11}}(z) &= 0.88 * z^3 + 0.07 * z^0 + 0.05 * z^c \\ u_{C_{12}}(z) &= 0.89 * z^3 + 0.11 * z^0 + 0.0 * z^c \\ u_{C_{13}}(z) &= 0.79 * z^5 + 0.15 * z^0 + 0.06 * z^c \\ u_{C_{21}}(z) &= 0.86 * z^6 + 0.14 * z^0 + 0.0 * z^c \\ u_{C_{22}}(z) &= 0.85 * z^2 + 0.15 * z^0 + 0.0 * z^c. \end{aligned} \tag{17}$$

Table 2: Value of $\varepsilon(r)$, $\phi_\delta$, and $\mu_\delta$ for any combination $\delta$.

| $\delta$ | $\varepsilon(r)$ | | $\phi_\delta$ | $\mu_\delta$ |
|---|---|---|---|---|
| | $r = 1$ | $r = 2$ | | |
| 0 | 0 | 0 | $= (1 - 0.05) * (1 - 0.06) = 0.893$ | $\varnothing$ |
| 1 | 1 | 0 | $= 0.05 * (1 - 0.06) = 0.047$ | $\{C_{11}, C_{21}\}$ |
| 2 | 0 | 1 | $= (1 - 0.05) * 0.06 = 0.057$ | $\{C_{13}, C_{21}, C_{22}\}$ |
| 3 | 1 | 1 | $= 0.05 * 0.06 = 0.003$ | $\{C_{11}, C_{21}\} \cup \{C_{13}, C_{21}, C_{22}\}$ |

From (11), the conditional UGFs of components take the following forms:

$$\begin{aligned} u'_{C_{11}}(z) &= 0.9263 * z^3 + 0.0737 * z^0 \\ u'_{C_{12}}(z) &= 0.89 * z^3 + 0.11 * z^0 \\ u'_{C_{13}}(z) &= 0.8404 * z^5 + 0.1596 * z^0 \\ u'_{C_{21}}(z) &= 0.86 * z^6 + 0.14 * z^0 \\ u'_{C_{22}}(z) &= 0.85 * z^2 + 0.15 * z^0. \end{aligned} \tag{18}$$

The conditional UGFs of component $C_{12}$, $C_{21}$, and $C_{22}$ can be obtained by removing $0.0 * z^c$ directly. Two components, $\beta(1) = C_{11}, \beta(2) = C_{13}$, cause the CCF, and then $S_{\beta(1)} = \{C_{21}\}$, $S_{\beta(2)} = \{C_{21}, C_{22}\}$. The other variables are listed in Table 2.

According to the physical nature of the components' interconnection, the function within subsystems sub1 and sub2 should take the form of the sum, and the function between two subsystems should take the form of the minimum. Then the conditional UGF of the entire system can be denoted as

$$\begin{aligned} u_\delta(z) &= u_{sub1}(z) \otimes_{\min} u_{sub2}(z) \\ &= \left(u'_{C_{11}}(z) \otimes_{sum} u'_{C_{12}}(z) \otimes_{sum} u'_{C_{13}}(z)\right) \\ &\quad \otimes_{\min} \left(u'_{C_{21}}(z) \otimes_{sum} u'_{C_{22}}(z)\right). \end{aligned} \tag{19}$$

For $\delta = 0$ and $\mu_\delta = \varnothing$, no UGF should be replaced with $\psi(z)$. Equation (19) can be expressed as

$$\begin{aligned} u_0(z) &= \left[\left(0.9263 * z^3 + 0.0737 * z^0\right)\right. \\ &\quad \otimes_{sum} \left(0.89 * z^3 + 0.11 * z^0\right) \\ &\quad \left. \otimes_{sum} \left(0.8404 * z^5 + 0.1596 * z^0\right)\right] \\ &\quad \otimes_{\min} \left[\left(0.86 * z^6 + 0.14 * z^0\right)\right. \\ &\quad \left. \otimes_{sum} \left(0.85 * z^2 + 0.15 * z^0\right)\right] \\ &= 0.6094 * z^8 + 0.2207 * z^6 + 0.0059 * z^5 + 0.0230 \\ &\quad * z^3 + 0.1188 * z^2 + 0.0222 * z^0. \end{aligned} \tag{20}$$

For $\delta = 1$ and $\mu_\delta = \{C_{11}, C_{21}\}$, only $u'_{C_{11}}(z)$ and $u'_{C_{21}}(z)$ will be substituted by $\psi(z) = z^c$. In the case of failure, the performance level is zero. Equation (19) will be rewritten as

$$\begin{aligned} u_1(z) &= \left[\left(z^0\right) \otimes_{sum} \left(0.89 * z^3 + 0.11 * z^0\right)\right. \\ &\quad \left. \otimes_{sum} \left(0.8404 * z^5 + 0.1596 * z^0\right)\right] \\ &\quad \otimes_{\min} \left[\left(z^0\right) \otimes_{sum} \left(0.85 * z^2 + 0.15 * z^0\right)\right] = 0.835 * z^2 \\ &\quad + 0.165 * z^0. \end{aligned} \tag{21}$$

For $\delta = 2$ and $\mu_\delta = \{C_{13}, C_{21}, C_{22}\}$, the UGF of components $C_{13}, C_{21}, C_{22}$ must be replaced by $\psi(z)$ in (19). So one can obtain the following formula:

$$\begin{aligned} u_2(z) &= \left[\left(0.9263 * z^3 + 0.0737 * z^0\right)\right. \\ &\quad \left. \otimes_{sum} \left(0.89 * z^3 + 0.11 * z^0\right) \otimes_{sum} \left(z^0\right)\right] \\ &\quad \otimes_{\min} \left[\left(z^0\right) \otimes_{sum} \left(z^0\right)\right] = z^0. \end{aligned} \tag{22}$$

When $\delta = 3$, the UGF of components $C_{11}, C_{13}, C_{21}, C_{22}$ should be substituted with $\psi(z)$. Then (19) will be written as

$$\begin{aligned} u_3(z) &= \left[\left(z^0\right) \otimes_{sum} \left(0.89 * z^3 + 0.11 * z^0\right) \otimes_{sum} \left(z^0\right)\right] \\ &\quad \otimes_{\min} \left[\left(z^0\right) \otimes_{sum} \left(z^0\right)\right] = z^0. \end{aligned} \tag{23}$$

The whole system's UGF can be obtained by

$$\begin{aligned} U_{sys}(z) &= \sum_{\delta=0}^{3} u_\delta(z) * \phi_\delta \\ &= 0.5442 * z^8 + 0.1971 * z^6 + 0.0053 * z^5 \\ &\quad + 0.0205 * z^3 + 0.1453 * z^2 + 0.0876 \\ &\quad * z^0. \end{aligned} \tag{24}$$

Now the reliability of the entire system with CCF can be depicted as in Figure 2.

According to (7), the reliability of the system with performance $\theta = 4$ can be calculated as follows. Point A in Figure 2 is the reflection of the reliability value:

$$\begin{aligned} R(4) &= \sum_{j=0}^{8} q_j \Delta\left(w_j, 4\right) = 0.5442 + 0.1971 + 0.0053 \\ &= 0.7466. \end{aligned} \tag{25}$$

Table 3: Performance distribution of components.

| Component | Perfection | | Degradation | | Local failure | Failure of causing CCF |
|---|---|---|---|---|---|---|
| | p | h | p | h | p | p |
| C11 | 0.8 | 9 | 0.10 | 7 | 0.05 | 0.05 |
| C12 | 0.75 | 12 | 0.09 | 8 | 0.05 | 0.11 |
| C21 | 0.85 | 8 | 0.12 | 6 | 0.03 | - |
| C22 | 0.90 | 11 | - | - | 0.10 | - |
| C23 | 0.82 | 7 | 0.13 | 5 | 0.05 | - |
| C31 | 0.93 | 13 | - | - | 0.07 | - |
| C32 | 0.82 | 11 | 0.05 | 7 | 0.11 | 0.02 |
| C33 | 0.89 | 9 | - | - | 0.11 | - |
| C34 | 0.82 | 11 | 0.10 | 7 | 0.05 | 0.03 |

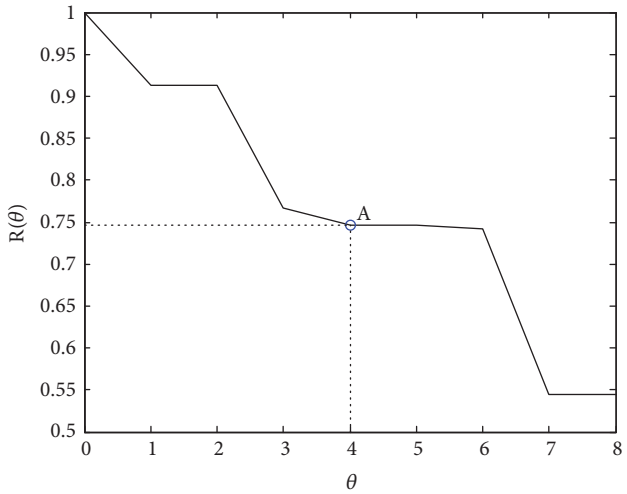



Figure 2: Reliability of the system.

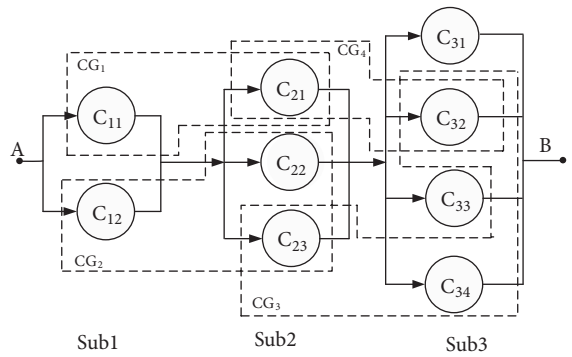



Figure 3: An MSS production system.

Using (9), the conditional expected system performance level at reliability $R(4)$ is

$$\Omega(4) = \frac{(8 * 0.5442 + 6 * 0.1971 + 5 * 0.0053)}{0.7466} = 7.4507. \quad (26)$$

*4.2. Advanced Application.* A more practical example will be depicted and the sensitivity of the CCF can be further analyzed. Suppose an MSS production system consists of nine components connected in series-parallel, as shown in Figure 3.

The system can be divided into the subsystems. Components $C_{22}$, $C_{31}$, and $C_{33}$ have only binary states. All the others can be seen as multistate components with two operating states: perfect and degraded performance.

The CCF can be inspired by four components $\boldsymbol{\beta} = \{C_{11}, C_{12}, C_{32}, C_{34}\}$. The set of components affected by the CCF are $S_{\beta(1)} = \{C_{21}\}$, $S_{\beta(2)} = \{C_{22}, C_{23}\}$, $S_{\beta(3)} = \{C_{21}\}$, and $S_{\beta(4)} = \{C_{23}, C_{32}\}$. These components constitute four CCF groups, $CG_1$, $CG_2$, $CG_3$, and $CG_4$. The performance distributions of system components for operating states and failing states are listed in Table 3.

According to the analysis method suggested above, the reliability curve at different performance levels can be represented as in Figure 4.

There are 21 performance levels corresponding to this MSS, as shown in the results. For example, the data points A, B, C, and D are the reliabilities at performance: $R(5) = 0.9782$, $R(10) = 0.8154$, $R(15) = 0.7228$, and $R(20) = 0.4711$, respectively.

Under the limitation of cost, it is necessary to find the most valuable component than can improve the reliability of the system to the greatest extent when the CCF originating from it is eliminated. In other words, the sensitivity of components can be further analyzed by the suggested method. When the CCF of component $i$ is localized, its UGF will be rewritten as follows, according to (10)

$$u_i(z) = \sum_{k_i=1}^{m_i} p_{ik_i} * z^{h_{ik_i}} + \left(p_{i0} + p_{i0'}\right) * z^0. \quad (27)$$

The CCF can be incorporated in the failure state without arousing other components' failures, and the performance level is also zeroed. Then the system reliability can be reassessed when the CCF in one single component is eliminated. The improvements shown in Table 4 have been compared with the case where no CCF is removed.

In Table 4, the performance level $\theta$ is chosen from 8 to 15 and the reliability $R(\theta)$ corresponds to cases where no CCF

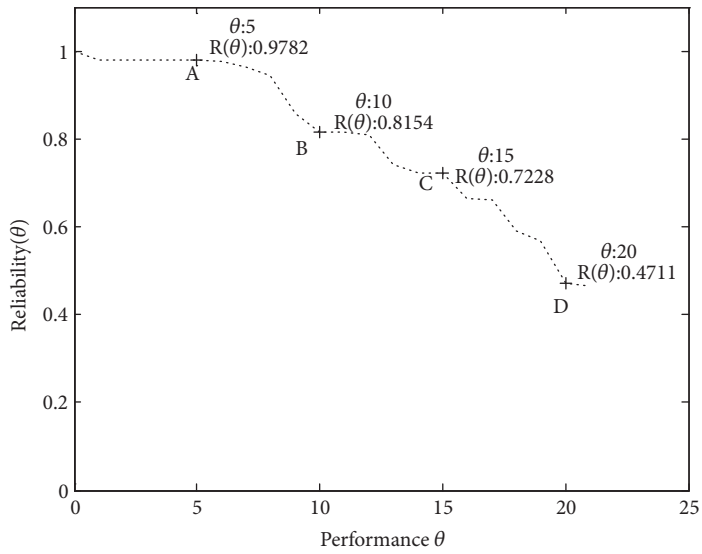



Figure 4: Curve of reliability at different performance levels.

Table 4: Comparison of reliability improvement after the CCF is removed from one single component.

| $\theta$ | R($\theta$) | Improvement after specified component's CCF is eliminated | | | |
|---|---|---|---|---|---|
| | | $\beta(1) = C_{11}$ | $\beta(2) = C_{12}$ | $\beta(3) = C_{32}$ | $\beta(4) = C_{34}$ |
| 8 | 0.943819 | 0.42% | 1.50% | 0.33% | 0.05% |
| 9 | 0.856588 | 0.38% | 10.14% | 0.17% | 0.28% |
| 10 | 0.815374 | 0.40% | 0.00% | 0.17% | 0.43% |
| 11 | 0.815374 | 0.40% | 0.00% | 0.17% | 0.43% |
| 12 | 0.808013 | 0.70% | 0.00% | 0.30% | 0.63% |
| 13 | 0.740249 | 0.00% | 0.00% | 0.31% | 0.54% |
| 14 | 0.722797 | 0.00% | 0.00% | 0.27% | 0.72% |
| 15 | 0.722768 | 0.00% | 0.00% | 0.27% | 0.72% |

is excluded. For the four components that can cause the CCF, the improved reliability is expressed in percentage form after the corresponding CCF has been eliminated.

Moreover, for a specified performance level $\theta$, the best component can be chosen to localize its CCF. For $\theta = 9$, the component $\beta(2) = C_{12}$ should be the best one to be improved because its improving proportion can reach 10.14%. At the same time, when one component has been chosen, the best performance level can be determined according to the extent of improvement. For example, because of some limitations only $\beta(1) = C_{11}$ can be maintained to remove its CCF, and the best performance level corresponding to the biggest reliability improvement, 0.7%, is $\theta = 12$. The analysis of sensitivity to CCF localization will help engineers to find bottlenecks in the system reliability design and to properly direct investment in reliability enhancement.

## 5. Concluding Remarks

In this paper a reliability assessment method is suggested for the series-parallel MSS with CCF. The components that can arouse the CCF may cause failures of different subsets of system components. These subsets can even overlap on some components. This method is based on the UGF application, and the factor of the CCF is incorporated in the expression of the UGF. Because of its repetition for every combination, the method is conveniently programmed by iteration. The indices of reliability and expected conditional performance level can be obtained. The sensitivity of components can be further analyzed. This method will help for reliability engineers to decide which component is the most valuable to be invested in, so as to eliminate its CCF. According to the assumptions in this method, the structure of the MSS is applied only to the series-parallel system. The more complex MSS topologies such as bridge and G:(k/n) structures need to be considered in the future.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported financially in part by a grant from Fundamental Research Funds for the Central Universities (no. 2015MS128; no. 2018MS076).

## References

[1] J. K. Vaurio, "The theory and quantification of common cause shock events for redundant standby systems," *Reliability Engineering & System Safety*, vol. 43, no. 3, pp. 289–305, 1994.

[2] K. N. Fleming and A. Mosleh, "Common cause data analysis and implications in system modeling," in *Proceedings of the International Topical Meeting on Probabilistic Safety Methods and Applications*, pp. 1–12, 1985.

[3] G. Levitin, "Incorporating common-cause failures into nonrepairable multistate series-parallel system analysis," *IEEE Transactions on Reliability*, vol. 50, no. 4, pp. 380–388, 2001.

[4] J. K. Vaurio, "Common cause failure probabilities in standby safety system fault tree analysis with testing—scheme and timing dependencies," *Reliability Engineering & System Safety*, vol. 79, no. 1, pp. 43–57, 2003.

[5] S. V. Amari, J. B. Dugan, and R. B. Misra, "A separable method for incorporating imperfect fault-coverage into combinatorial models," *IEEE Transactions on Reliability*, vol. 48, no. 3, pp. 267–274, 1999.

[6] G. Levitin, "Block diagram method for analyzing multi-state systems with uncovered failures," *Reliability Engineering & System Safety*, vol. 92, no. 6, pp. 727–734, 2007.

[7] L. Xing, "Reliability evaluation of phased-mission systems with imperfect fault coverage and common-cause failures," *IEEE Transactions on Reliability*, vol. 56, no. 1, pp. 58–68, 2007.

[8] A. F. Myers, "k-out-of-n:G system reliability with imperfect fault coverage," *IEEE Transactions on Reliability*, vol. 56, no. 3, pp. 464–473, 2007.

[9] J. E. Ramirez-Marquez and D. W. Coit, "Optimization of system reliability in the presence of common cause failures," *Reliability Engineering & System Safety*, vol. 92, no. 10, pp. 1421–1434, 2007.

[10] L. Xing, B. A. Morrissette, and J. B. Dugan, "Efficient analysis of imperfect coverage systems with functional dependence," in *Proceedings of the Reliability and Maintainability Symposium (RAMS)*, pp. 1–6, 2010.

[11] X.-Y. Wu and X.-Y. Wu, "Extended object-oriented Petri net model for mission reliability simulation of repairable PMS with common cause failures," *Reliability Engineering & System Safety*, vol. 136, pp. 109–119, 2015.

[12] J. Jin, L. Pang, S. Zhao, and B. Hu, "Quantitative assessment of probability of failing safely for the safety instrumented system using reliability block diagram method," *Annals of Nuclear Energy*, vol. 77, pp. 30–34, 2015.

[13] F. P. A. Coolen and T. Coolen-Maturi, "Predictive inference for system reliability after common-cause component failures," *Reliability Engineering & System Safety*, vol. 135, pp. 27–33, 2015.

[14] J. Murchland, "Fundamental concepts and relations for reliability analysis of multi-state systems," in *Reliability and Fault Tree Analysis*, 1975.

[15] X. Janan, "On multistate system analysis," *IEEE Transactions on Reliability*, vol. R-34, no. 4, pp. 329–337, 1985.

[16] A. Shrestha, L. Xing, and S. V. Amari, "Reliability and sensitivity analysis of imperfect coverage multi-state systems," in *Proceedings of the Annual Reliability and Maintainability Symposium: The International Symposium on Product Quality and Integrity, RAMS 2010*, pp. 1–6, USA, January 2010.

[17] G. Levitin, L. Xing, H. Ben-Haim, and Y. Dai, "Reliability of series-parallel systems with random failure propagation time," *IEEE Transactions on Reliability*, vol. 62, no. 3, pp. 637–647, 2013.

[18] J. Mi, Y. Li, H.-Z. Huang, Y. Liu, and X. Zhang, "Reliability analysis of multi-state systems with common cause failure based on Bayesian Networks," *Eksploatacja I Niezawodnosc-Maintenance and Reliability*, vol. 15, no. 2, pp. 169–175, 2013.

[19] I. A. Ushakov, "Optimal standby problems and a universal generating function," *Soviet Journal of Computer and Systems Sciences*, vol. 25, no. 4, pp. 79–82, 1987.

[20] G. Levitin, *Universal Generating Function in Reliability Analysis and Optimization*, Springer, London, UK, 2005.

[21] H. A. Khorshidi, I. Gunawan, and M. Y. Ibrahim, "On reliability evaluation of multistate weighted k-out-of-n system using present value," *The Engineering Economist*, vol. 60, no. 1, pp. 22–39, 2015.

[22] G. Levitin, L. Xing, and Y. Dai, "Linear multistate consecutively-connected systems subject to a constrained number of gaps," *Reliability Engineering & System Safety*, vol. 133, pp. 246–252, 2015.

[23] S. Negi and S. B. Singh, "Reliability analysis of non-repairable complex system with weighted subsystems connected in series," *Applied Mathematics and Computation*, vol. 262, pp. 29–89, 2015.

Hindawi
Complexity
Volume 2019, Article ID 5909785, 11 pages
https://doi.org/10.1155/2019/5909785

*Research Article*

# The Incentive Model in Supply Chain with Trade Credit and Default Risk

**Hong Cheng**[1], **Yingsheng Su**[2], **Jinjiang Yan**[3], **Xianyu Wang**[3] and **Mingyang Li**[1]

[1]*College of Management Science, Chengdu University of Technology, Chengdu, China*
[2]*School of Statistics, Southwestern University of Finance and Economics, Chengdu, China*
[3]*Business School, Sichuan University, Chengdu, China*

Correspondence should be addressed to Yingsheng Su; suys@swufe.edu.cn

Received 29 January 2019; Revised 3 April 2019; Accepted 11 April 2019; Published 2 May 2019

Guest Editor: Rosario Domingo



Trade credit is widely used for its advantages. However, trade credit also brings default risk to the manufacturer due to the uncertain demand. And moral hazard may aggravate the default risk. The purpose of this paper is to investigate the role of moral hazard in trade credit and explore incentive contract under uncertain demand and asymmetric information. We consider a two-echelon supply chain consisting of a risk-neutral retailer ordering a single product from a risk-neutral manufacturer. Market demand is stochastic and is influenced by retailer's sales effort which is his private information. Incentive theory is used to develop the principal-agent model and get the incentive contract from the manufacturer's perspective. Results show that the retailer will reduce his effort level to get more profit and the manufacturer's profit will be reduced, in the case of asymmetric information. Facing this result, the manufacturer will reduce the order quantity in incentive contract to lessen his losses. Numerical examples are provided to illustrate all these theoretical results and to draw managerial insights.

## 1. Introduction

Emerging trends and practices have joint considered capital flow, material flow, and information flow in supply chain [1]. The typical financial resources for firms in supply chain include trade credit, bank credit, and capital budget. This paper focuses on the trade credit. Trade credit is widely used, with a survey indicating that companies were granting trade credits for up to 80% of their business transactions in some European countries [2].

Trade credit cannot be replaced by bank loans for its advantages. Comparatively speaking, trade credit is more accessible and less expensive than bank loans when a buyer faces constrained finance. And trade credit has been proven to be one of the more effective competitive tools for product market lenders [3, 4]. It was found in a study on Spanish firms that SMEs with credit constrained depended on trade credit instead of bank loans, particularly during the financial crisis [5]. It was also concluded that trade credit had advantages for both the supplier and the retailer in the supply chain [6].

Trade credit is widespread and has been shown to be good for the supply chain. But it can also result in default risks and even bankrupt risks to the manufacturer if the retailer is unable to settle the account on time. For example, when demand decreases suddenly at a certain time because of the uncertain market demand, the retailer may not have sufficient funds to pay the manufacturer because of bad sales. Therefore, the unsold losses are transferred from the retailer to the manufacturer because of the trade credit, which results in a loss to the manufacturer, and even results in a domino effect and a supply chain crisis. For example, the 2008 US subprime crisis resulted in an economic downturn in many countries, and some small open economies that had little direct exposure to the US subprime mortgage market suffered from a fall in export demand by up to 30% in the second half of 2008 [7]. This means that providing trade credit could be risky for the manufacturer when facing uncertain demand.

Such risks may be worse when the retailer has private information about his sales effort level which can influence the uncertain demand. Cachon (2003) summarized that the moral hazard caused by retailer's private information of sales

effort level would reduce the manufacturer's profit and the efficiency of the supply chain [8]. Because the effort level is the retailer's private information and the retailer will pay effort to maximize his profit instead of paying as the request. The quantity discount contract can fix that when there is no trade credit. Uncertain demand and asymmetric information are common in commercial activities. Therefore, including the moral hazard and the uncertain demand, which can result in default risk, is critical when analyzing trade credit in supply chain management. When studying trade credit with default risk few papers study moral hazard. Therefore, this paper seeks to fill this gap by studying the moral hazard and default risk in trade credit.

In this paper, our motivation for developing principal-agent models under the situation of symmetric information and the situation of asymmetric information separately is to investigate the impact of moral hazard on trade credit and explore the incentive contract. What is the effect of trade credit with uncertain demand influenced by retailer's sales effort? How does the moral hazard influence the trade credit? What contract can weaken the moral hazard? Our research will attempt to answer these questions.

The remainder of the paper is structured as follows. Section 2 provides a review of existing article about trade credit in supply chain. Section 3 presents assumptions and notations and describes the basic model. The model under symmetric information is developed and analyzed in Section 4. The influence of moral hazard under asymmetric information is analyzed and incentive contract is derived in Section 5, followed by an application example in Section 6. Finally, Section 7 draws conclusions.

## 2. Literature Review

Since Goyal (1985) first introduced the economic order quantity (EOQ) model to study trade credit in supply chains [9], the model has been extended to deteriorating items and two or three level trade credits from the retailer's view to determine specific replenishment policies. For example, Aggarwal and Jaggi (1995) built the model for deteriorating items under trade credit [10], and Teng (2009) extended the model to two-level trade credit [11], in which the supplier offers trade credit to the retailers and the retailer offers trade credit to their customers. Then, Tiwari (2018) considered this study under two-level partial trade credits in supply chain [12]. Besides deteriorating items, Lin et al. (2019) considered the variable capacity utilization under trade credit [13]. Feng and Chan (2019) concerned manufacturer's production cost which follows a learning curve effect and studied the optimal lot-sizing and pricing strategies with both upstream and downstream trade credit [14]. Chung et al. (2014) [15] also developed an economic production quantity inventory model for deteriorating items under a two-level trade credit to determine an optimal replenishment policy. Along with deteriorating items and two-level trade credit, Liao et al. (2017) also considered partial trade credit and two storage facilities in situations when the supplier's own warehouse was not sufficient for the order quantity and a rented warehouse was required for the excess units [16]. Some more literatures focused on the trade credit influence on decision-making in supply chain. For example, Zhong et al. (2018) introduced the trade credit to supply chain network design [1], and Hosseini-Motlagh et al. (2018) introduced trade credit to the collaborative model for coordination in a supply chain consisting of one monopolistic manufacturer and two competing retailers [17].

The literatures above have assumed that market demand is constant or depending on the retail price. And under this assumption, some literatures have sought to coordinate the supply chain inventories under a trade credit environment to enhance supply chain efficiency. Luo (2007), for example, analyzed supply chain coordination using credit periods [18], and Yang et al. (2014) examined the efficiency of quantity discounts on supply chain coordination [19]. These assumptions about the demand mean that there is no default risk for trade credit, which means the borrower can repay the credit loan certainly. In reality, however, the borrower cannot always repay the credit loan because of the bad sales. And this brings the default risk to lenders.

Considering the default risk coming from the probabilities of the supplier and the retailer's bankruptcy, Lee and Rhee (2011) applied a Newsvendor framework to study supply chain coordination using trade credit as a tool from the supplier's perspective [20]. For Lou and Wang (2013) and Wu et al. (2017), the optimal order quantity and credit period were discussed by assuming that the longer the credit period offered to the buyer, the higher the default risk to the seller [21, 22]. Tsao (2018) extended this study to the uncertain default risk [23].

The uncertain demand is an important factor to the default risk and is common in business. Therefore, including uncertain demand is critical when analyzing trade credit in supply chain. Under an uncertain demand scenario, Chaharsooghi and Heydari (2010) and Arkan and Hejazi (2012) proposed incentive schemes to coordinate the supply chain [24, 25]; Heydari (2015) then extended to truckload limitations [26]. Heydari et al. (2017) investigated how a two-level trade credit contributed to coordination under an assumed stochastic and credit-dependent demand [2]. Although these literatures above assumed an uncertain demand, the default risk coming from this uncertain demand under trade credit was not considered.

Considering the default risk coming from the uncertain demand, Kouvelis and Zhao (2012) analyzed a retailer's optimal decisions when seeking a loan from the supplier or from a bank and a supplier's optimal contract parameters [27]. Chen (2015) concluded that the trade credit made both channel members better off, and there was a unique financing equilibrium in wholesale price contracts when retailers were confronted with capital constraints [28]. No more literatures, which consider the default risk coming from the uncertain demand, are found. But the assumption about uncertain demand in supply chain is popular and is the base of newsvendor model. The default risk in credit trade always happens because of bad sales which is not as expected. So, considering the default risk with uncertain demand is necessary in trade credit study.

To reduce the default risk, Yang et al. (2016) proposed a two-period model with a flexible trade credit contract to incentivize the retailer to settle the account early [29]. Big-Data analytics has also been proven to be a useful tool in reducing default risk; when the retailer (supplier) has the power to determine the credit period, the retailer (supplier) can implement Big-Data analytics [30]. In the presence of both retailer and supplier capital constraints, Kouvelis and Zhao (2012) [27] applied a supplier early payment discount scheme and concluded that both supplier and retailer preferred trade credit to bank financing. They also surmised that the reason not all suppliers wanted to offer cheap trade credits was because of symmetric information, such as an asymmetric ability to pay, asymmetric demand information, and asymmetric action about effort.

Asymmetric information is common in commercial activities. Ex ante asymmetric information is called adverse selection and ex post asymmetric information is known as moral hazard. Both adverse selection and moral hazard can be harmful. Therefore, the principal seeks to mitigate them. Using micro-data on Italian SMEs from 1998-2006, it was found that banks could use trade credit as the signal to offer additional loans to reduce asymmetric information [31]. However, there is also asymmetric information in trade credit. The adverse selection has been considered in previous studies, such as Wang (2018), in which the asymmetric credit default risk was discussed [32]. But moral hazard has not been widely considered. Therefore, this paper seeks to fill this gap by accounting for moral hazard.

With the mitigation of moral hazard in mind, some studies have designed incentive contracts under certain demand. Through the development of an agency model, Pindado (2007) found that in certain circumstances, trade credit was offered primarily to mitigate adverse selection rather than to mitigate moral hazard [33]. The question remains therefore as to whether the moral hazard under trade credit can be mitigated. Costello (2013) claimed that long-term rather than short-term contracts could better reduce moral hazard [34]. Csóka et al. (2015) proved that in the asymmetric case the borrowing capacity and the welfare of the society are weakly smaller than in the symmetric case by developing a game model [35]. In which discrete revenue was considered and trade credit was seen as similar to equity financing. Wang and Liu discussed the supply chain coordination under trade credit and moral hazard caused by retailer's private information about effort level [36]. They considered uncertain demand influenced by retailer's effort. But they did not consider the default risk from the uncertain demand.

Overall, from the perspective of supply chain management research, most literatures studying for trade credit adopt EOQ model and assumed constant market demand rate. And they also ignored the default risk for trade credit. A few existing papers focused on enhancing supply chain's efficiency and considered the default risk coming from the uncertain demand, which is a popular assumption in supply chain management. According to the authors' best knowledge, there have been few papers examining the moral hazard and default risk simultaneously in trade credit, under uncertain demand.

Therefore, in this paper, we construct principal-agent models based on newsvendor model to shed some light on the role of default risk and moral hazard in trade credit. This work is an extension of Chen (2015), which studied the trade credit contract considering default risk coming from the uncertain demand [28]. And moral hazard, because of retailer's private information of his sales effort, is introduced in this paper. Our work is distinguished from earlier studies as follows: (1) studying the impact of moral hazard on trade credit and (2) exploring the incentive contract to mitigate the impact of moral hazard. In addition, the decisions of retailers with different working capital are analyzed. At the same time, the condition when the manufacturer adopts the incentive contract is also studied.

## 3. Model Description

*3.1. Assumptions and Notations.* Most of the assumptions and notations are based on Wang and Liu (2018) [36], except the default risk coming from the trade credit. A supply chain with a risk-neutral manufacturer and a risk-neutral retailer is considered. The retailer purchases a single product from the manufacturer. The manufacturer dominates and first sets trade credit contract menu. Then, the retailer chooses the contract and executes his effort according to contract terms and market situation. The retailer is capital constraint. He pays part of the transfer payment with his working capital and arranges to settle the remainder of the account at the end of the sales period, when ordering.

The retailer faces the uncertain market demand. We assume the demand without sales effort effect is $y$ with a cumulative distribution function $F(y)$ and a probability density function $f(y)$. The market demand is affected by the retailer's effort, such as the promotion, maintaining and developing customers, etc. To model retailer's effort, supposing a single effort level, $e$, summarizes the retailer's activities and let $\psi(e)$ be the retailer's cost of exerting effort level $e$, where $\psi(e) = 0$ $\psi'(e) > 0$ and $\psi''(e) > 0$; let $F(y \mid e)$ be the distribution of demand given the retailer's sales effort level $e$, where demand is stochastically increasing in effort, i.e., $F_e(y \mid e) < 0$, $F_{ee}(y \mid e) > 0$ [8].

Because of the uncertain demand, the retailer's sales revenue is uncertain. If the retailer's sales revenue is sufficient to settle the account at the end of the sales period the manufacturer recovers the balance due; otherwise, the manufacturer receives all the retailer's sales revenue but is unable to recover the remaining debt because of the retailer's limited liability.

Finally, to focus on trade credit and incentive strategies, assume the product is seasonal product and neglect the salvage value of unsold items. The time value of the capital is also not considered.

The following notations are used throughout this study.

$p$ is the retailer's unit product retail price;

$c$ is the manufacturer's unit product production cost;

$\eta$ is working capital of the retailer;

$t$ is total transfer payments from retailer to manufacturer;

$w$ is the unit product wholesale price provided by manufacturer;

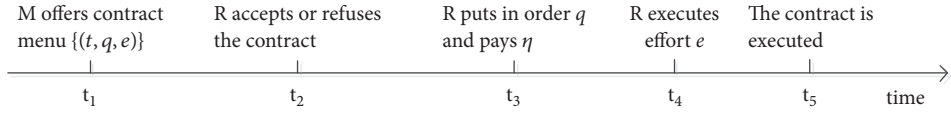



Figure 1: The sequence of events.

$q$ is ordering quantity of the retailer;
$e$ is sales effort level of the retailer;
$\pi_s$ is manufacturer's expected profit;
$\pi_r$ is retailer's expected profit;
$\pi$ supply chain's expected profit.

$e$ is retailer's decision variable. $t$ and $q$ are manufacturer's decision variables and constitute the contract provided by manufacturer.

*3.2. The Basic Model.* The manufacturer grants trade credit to the retailer and finances the retailer's purchase. The sequence of events is illustrated in Figure 1, in which M represents the manufacturer and R represents the retailer. Before making a transaction, the manufacturer provides a contract menu $\{(t, q, e)\}$, which the retailer can accept or reject. Then, the retailer chooses to accept or reject the contract menu. If the retailer refuses it, the transaction ends. If the retailer accepts it, he will put in order $q$ and pay $\eta$, leaving an account $t - \eta$ paid at the end of the sales period. The retailer executes effort $e$ during the sales period according to the contractual requirements. At the end of the sales period, both the manufacturer and retailer execute the contract. If there is adequate sales revenue to pay the account, the manufacturer can recycle $t-\eta$ owed by the retailer. Otherwise, the manufacturer receives all sales revenue $py$ from the retailer and the retailer goes bankrupt.

When the sales revenue of the retailer is enough to cover the credit account, there is $py \geq t - \eta$ which can be written as $y \geq (t - \eta)/p$. Therefore, when $y \geq (t - \eta)/p$, the retailer repays the full credit $t - \eta$ to the manufacturer, and when $y < (t - \eta)/p$ the retailer pays all sales revenue $py$ to the manufacturer and declares bankrupt. Comprehending the above, the manufacturer takes default risk, and his expected profit function is

$$\pi_s = t(q, e) - cq - \int_0^z pF(y \mid e)\, dy \tag{1}$$

where $z = (t - \eta)/p$.

The retailer's expected profit function is

$$\pi_r = pq - \int_z^q pF(y \mid e)\, dy - t(q, e) - \psi(e) \tag{2}$$

The expected profit function of the integrated supply chain is

$$\pi = pq - \int_0^q pF(y \mid e)\, dy - cq - \psi(e) \tag{3}$$

Subsequently, the second partial derivatives of function (3) are $-pf(q \mid e)$ and $-\psi''(e)$, and both of them are not more than zero. With the assumption $pf(q \mid e) \int_0^q pF_{ee}(y \mid e)dy + pf(q \mid e)\psi''(e) - p^2F_e^2(q \mid e) > 0$, which is implied in Cachon with the same profit function and model of the supply chain [37], the Hessian matrix of the binary function $\pi(q, e)$ is positive. Thus, from the supply chain perspective, the optimal quantity and effort level $(q^o, e^o)$ can be solved from the following first order condition (4) and (5).

$$-\int_0^{q^o} pF_{e^o}(y \mid e^o)\, dy - \psi'(e^o) = 0 \tag{4}$$

$$p\overline{F}(q^o \mid e^o) - c = 0 \tag{5}$$

And the expected profit of the supply chain is

$$\pi(q^o, e^o) = pq^o - \int_0^{q^o} pF(y \mid e^o)\, dy - cq^o - \psi(e^o) \tag{6}$$

## 4. The Model under Symmetric Information

In this part, we assume the retailer's effort level can be observed and verified by the manufacturer. The observable and verifiable effort is contractual and enforceable. Therefore, the effort can be signed in the contract designed by the manufacturer to maximize his own profit.

From the sequence as Figure 1, the model of the manufacturer can be written as P1 based on the incentive theory.

$$\begin{aligned} \text{P1:} \quad & \max_{(t,q,e)} \pi_s = t(q, e) - cq - \int_0^z pF(y \mid e)\, dy \\ & \text{s.t.} \quad \pi_r \geq \pi_{r0} \end{aligned} \tag{7}$$

where $\pi_{r0}$ is the retailer's reserved profit, which indicates that the retailer can get profit $\pi_{r0}$ when participating other activities instead of this transaction. The constraint means that the retailer cannot receive more profit in some other activities than in this transaction.

Solving model P1 and Proposition 1 can be got as follows.

**Proposition 1.** *Under symmetry information, the contract is* $(t_1(q^o, e^o), q^o, e^o)$, *where* $t_1(q^o, e^o) = pq^o - \int_{z^o}^{q^o} pF(y \mid e^o)dy - \psi(e^o) - \pi_{r0}$. *With this contract, the retailer retains the reserved profit* $\pi_{r0}$, *and the manufacturer receives all other residual profit of the supply chain* $\pi_{s1} = \pi(q^o, e^o) - \pi_{r0}$.

*Proof.* As the manufacturer does not allow the retailer to earn more than reserved profit, the manufacturer makes the transfer payment $t(q, e) = pq - \int_0^q pF(y \mid e)dy - \psi(e) + \int_0^z pF(y \mid e)dy - \pi_{r0}$ to maximize his profit. With this transfer payment, the retailer's expected profit is

$$\pi_{r1} = \pi_{r0} \tag{8}$$

The manufacture's expected profit is $\pi = pq - \int_0^q pF(y \mid e)dy - cq - \psi(e)$. This is the same as the expected profit function of the integrated supply chain. So, the manufacturer makes decision consistent with the integrated supply chain. And the same as the integrated supply chain, the optimal product transaction volume and the optimal effort are $q^o$ and $e^o$, respectively.

Therefore, the transfer payment function and the profits of both sides are as follows:

$$t_1(q^o, e^o) = pq^o - \int_{z^o}^{q^o} pF(y \mid e^o)\,dy - \psi(e^o) - \pi_{r0} \quad (9)$$

where $z^o = (t_1(q^o, e^o) - \eta)/p$.

$$\pi_{r1} = \pi_{r0} \quad (10)$$

$$\pi_{s1} = \pi(q^o, e^o) - \pi_{r0} \quad (11)$$

□

The retailer only retains the reserved profits $\pi_{r0}$, and the manufacturer receives the full residual supply chain profit $\pi_{s1} = \pi(q^o, e^o) - \pi_{r0}$.

From Proposition 1, we can find that the order quantity, the effort level, and expected profits of both partners have no relationship with the retailer's working capital. Only the total transfer payment is influenced by the retailer's working capital.

*Property 2.* The transfer payment $t_1(q^o, e^o)$ decreases with an increase in $\eta$.

*Proof.* From (9), we can get $\partial t_1(q^o, e^o)/\partial\eta = -F(z^o \mid e^o)/(1 - F(z^o \mid e^o))$. $F(z^o \mid e^o)$ is a cumulative distribution function and is no bigger than 1. So, $\partial t_1(q^o, e^o)/\partial\eta < 0$, which means that the transfer payment $t_1(q^o, e^o)$ decreases with an increase in $\eta$. □

Property 2 means that the retailer with higher working capital will pay less total transfer payment. This is because the retailer with higher working capital brings less default risk to the manufacturer.

## 5. The Model under Asymmetric Information

*5.1. The Moral Hazard.* Under asymmetric information, we assume that the retailer's effort level is his private information. The true effort level of the retailer cannot be observed and verified by the manufacturer or any other third party. And a contract binding the retailer to choose a particular effort level cannot be signed.

In this case, if the manufacturer provides the same contract $(t_1(q^o, e^o), q^o, e^o)$, the retailer will execute the required order quantity $q^o$ because it can be observed. But the retailer will execute his optimal effort $e_2^*$ to maximize his profit, instead of the required effort $e^o$ because it cannot be observed. Therefore, a moral hazard arises.

With asymmetric information and the contract $(t_1(q^o, e^o), q^o, e^o)$, the retailer's expected profit function is

$$\pi_{r2} = pq^o - \int_{z^o}^{q^o} pF(y \mid e)\,dy - t_1(q^o, e^o) - \psi(e) \quad (12)$$

We take the partial derivatives of (12) and obtain $\partial^2\pi_{r2}/\partial e^2 = -\int_{z^o}^{q^o} pF_{ee}(y \mid e)dy - \psi''(e)$. In addition, with the assumption $-\int_{z^o}^{q^o} pF_{ee}(y \mid e)dy - \psi''(e) \le 0$, according to the first-order conditions, the retailer's optimal effort ${e_2}^*$ is the solution of the following equation:

$$-\int_{z^o}^{q^o} pF_{e_2^*}(y \mid e_2^*)\,dy - \psi'(e_2^*) = 0 \quad (13)$$

The profits obtained by the retailer and the manufacturer, respectively, are

$$\pi_{r2} = p\int_{z^o}^{q^o} \left(F(y \mid e^o) - F(y \mid e_2^*)\right)dy + \psi(e^o) - \psi(e_2^*) + \pi_{r0} \quad (14)$$

$$\pi_{s2} = \pi(q^o, e^o) + p\int_0^{z^o} \left(F(y \mid e^o) - F(y \mid e_2^*)\right)dy - \pi_{r0} \quad (15)$$

and the total profit of the supply chain is

$$\pi_2 = \pi(q^o, e^o) + p\int_0^{q^o} \left(F(y \mid e^o) - F(y \mid e_2^*)\right)dy + \psi(e^o) - \psi(e_2^*) \quad (16)$$

We can get Proposition 3 as follows.

**Proposition 3.** *Under asymmetric information and the contract $(t_1(q^o, e^o), q^o, e^o)$, the retailer's effort level $e_2^*$ is the solution to (13), and the expected profits of the retailer, manufacturer, and supply chain are (14), (15), and (16), respectively.*

*Property 4.* There is $e_2^* \le e^o$, and $e_2^*$ increases with $\eta$ increasing.

*Proof.* With the assumption $-\int_{z^o}^{q^o} pF_{ee}(y \mid e)dy - \psi''(e) \le 0$, $-\int_{z^o}^{q^o} pF_e(y \mid e)dy - \psi'(e)$ decreases with $e$ increasing. Equation (4) can be written as $-\int_0^{z^o} pF_{e^o}(y \mid e^o)dy - \int_{z^o}^{q^o} pF_{e^o}(y \mid e^o)dy - \psi'(e^o) = 0$. If $e_2^* > e^o$, then $-\int_{z^o}^{q^o} pF_{e_2^*}(y \mid e_2^*)dy - \psi'(e_2^*) < -\int_{z^o}^{q^o} pF_{e^o}(y \mid e^o)dy - \psi'(e^o)$. Combining with (13), there is $-\int_{z^o}^{q^o} pF_{e^o}(y \mid e^o)dy - \psi'(e^o) > 0$. This means that (4) cannot be satisfied. So, there is $e_2^* \le e^o$. The retailer exerts less effort under asymmetric information than that under symmetric information, which is in line with the moral hazard expectation.

From (13), we have $(-\int_{z^o}^{q^o} pF_{e_2^*e_2^*}(y \mid e_2^*)dy - \psi''(e_2^*))(\partial e_2^*/\partial\eta) + F_{e_2^*}(z^o \mid e_2^*)(\partial t_1/\partial\eta - 1) = 0$. Combining Property 2, we

can get $\partial e_2^*/\partial\eta \geq 0$; $e_2^*$ increases with $\eta$. The retailer exerts more effort when he puts in more working capital. □

*Property 5.* Under this situation, the retailer's profit $\pi_{r2}$ decreases with an increase in $\eta$.

*Proof.* From (14) and (13), we can have $\partial\pi_{r2}/\partial\eta = (F(z^o \mid e_2^*) - F(z^o \mid e^o))/(F(z^o \mid e^o) - 1)$. Because of $F_e(y \mid e) < 0$ and $e_2^* \leq e^o$, there is $\partial\pi_{r2}/\partial\eta \leq 0$. The retailer gets less profit when he has more working capital. □

So, from Property 5, the retailer can get more profit by reducing his effort level, when his working capital is small. But when the retailer has more working capital, he can get less profit by reducing his effort level. This means that moral hazard is more likely to happen, when $\eta$ is small.

*Property 6.* Under moral hazard, there is $\pi_{s2} \leq \pi(q^o, e^o)$, and $\pi_{s2}$ increases with an increase in $\eta$.

*Proof.* With Property 4, the retailer usually executes a lower effort level than the optimal effort to get more profit. Comparing (15) and (11), the reduce of effort causes a change $p\int_0^{z^o}(F(y \mid e_2{}^*) - F(y \mid e^o))dy$ to the manufacturer's profit. Because of $F_e(y \mid e) \leq 0$ and Property 4, $p\int_0^{z^o}(F(y \mid e_2^*) - F(y \mid e^o))dy \leq 0$. So, the manufacturer encounters profit loss $p\int_0^{z^o}(F(y \mid e_2^*) - F(y \mid e^o))dy$ under asymmetric information and the contract $(t(q^o, e^o), q^o, e^o)$.

Because of $F_e(y \mid e) \leq 0$, it is easy to find that $p\int_0^{z^o}(F_e(y \mid e))dy \leq 0$. So, the loss decreases with an increase in $e$. Combining this and Property 2, we come to a conclusion that the loss decreases with an increase in $\eta$. Therefore, the manufacturer's expected profit under moral hazard increases with an increase in $\eta$. □

Combining Properties 5 and 6, the retailer with low working capital could get more profit by reducing effort level but brings more profit loss to the manufacturer. It is better for the manufacturer refusing the retailer with low working capital to reduce moral hazard and profit loss, under asymmetric information. However, trade credit is inevitable in modern commercial activities. When refusing the retailer with little capital is not feasible, the manufacturer needs to take other measures such as incentive contracts to reduce his losses from moral hazard.

*5.2. Incentive Contract.* To reduce the profit loss coming from the moral hazard, the manufacturer needs to encourage the retailer to improve effort level. The manufacturer's incentive model can be developed as P2.

$$\begin{aligned} \text{P2:} \max_{\{(t(q),q)\}} \quad & \pi_s = t(q) - cq - \int_0^z pF(y \mid e)\,dy \\ \text{s.t.} \quad & e = \arg\max_e \pi_r \\ & \pi_r \geq \pi_{r0} \end{aligned} \tag{17}$$

In constraints, arg max means arguments of the maxima [38]. Solved as P1, the transaction quantity $q_3^*$ and retailer's effort $e_3^*$ expected by the manufacturer in the contract are the solutions to the following equations:

$$\begin{aligned} &-\int_0^{q_3^*} pF_{e_3^*}(y \mid e_3^*)\,dy - \psi'(e_3^*) \\ &+\int_0^{z_3^*} pF_{e_3^*}(y \mid e_3^*)\,dy = 0 \end{aligned} \tag{18}$$

$$p - pF(q_3^* \mid e_3^*) - c = 0 \tag{19}$$

where $z_3^* = (t(q_3^*, e_3^*) - \eta)/p$, and

$$\begin{aligned} t(q_3^*, e_3^*) = {} & pq_3^* - \int_0^{q_3^*} pF(y \mid e_3^*)\,dy - \psi(e_3^*) \\ & + \int_0^{z_3^*} pF(y \mid e_3^*)\,dy - \pi_{r0} \end{aligned} \tag{20}$$

With contract $(t(q_3^*, e_3^*), q_3^*, e_3^*)$, the respective profits for the supply chain, the manufacturer, and the retailer are as in (21), (22), and (23).

$$\begin{aligned} \pi_3(q_3^*, e_3^*) = {} & pq_3^* - \int_0^{q_3^*} pF(y \mid e_3^*)\,dy - \psi(e_3^*) \\ & - cq_3^* \end{aligned} \tag{21}$$

$$\pi_{s3} = \pi(q_3^*, e_3^*) - \pi_{r0} \tag{22}$$

$$\pi_{r3} = \pi_{r0} \tag{23}$$

**Proposition 7.** *Under moral hazard, the incentive contract provided by the manufacturer is $(t(q_3^*, e_3^*), q_3^*, e_3^*)$, in which $q_3^*$ and $e_3^*$ are the solutions to (18) and (19), and the transfer payment $t(q_3^*, e_3^*)$ is determined by (20). Therefore, the respective profits of the supply chain, the manufacturer, and the retailer are (21), (22), and (23).*

From Propositions 3 and 7, we can find that the retailer's profit is reduced to his reserved profit. And when $\pi(q^o, e^o) - \pi(q_3^*, e_3^*) \leq \int_0^{z^o} pF(y \mid e_2^*)dy - \int_0^{z^o} pF(y \mid e^o)dy$ is satisfied the manufacturer obtains a higher profit with the incentive contract and the incentive contract is adopted to prevent the moral hazard. Otherwise, the manufacturer obtains a higher profit without an incentive contract and he will accept the profit loss coming from moral hazard instead of adopting incentive measures. Therefore, the inequality can be used to assess whether the manufacturer chooses to take the incentive measure. We can get Property 8 as follows.

*Property 8.* If $\pi(q^o, e^o) - \pi(q_3^*, e_3^*) \leq \int_0^{z^o} pF(y \mid e_2^*)dy - \int_0^{z^o} pF(y \mid e^o)dy$ is satisfied, the manufacturer adopts the incentive contract to motivate the retailer; otherwise the manufacturer does not adopt the incentive contract.

The left part of $\pi(q^o, e^o) - \pi(q_3^*, e_3^*) \leq \int_0^{z^o} pF(y \mid e_2^*)dy - \int_0^{z^o} pF(y \mid e^o)dy$ is the profit loss of supply chain

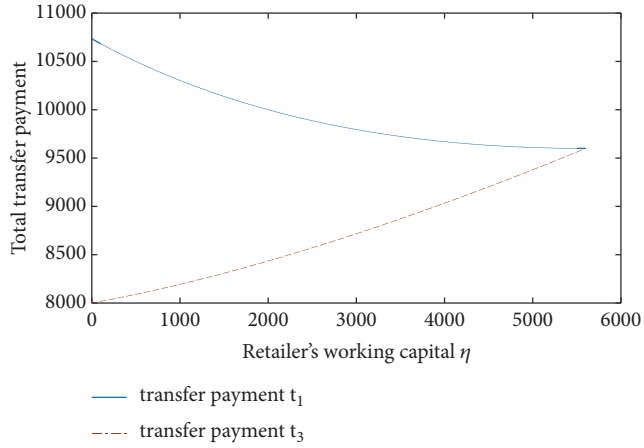



Figure 2: Change trend of total transfer payment with retailer's working capital $\eta$.

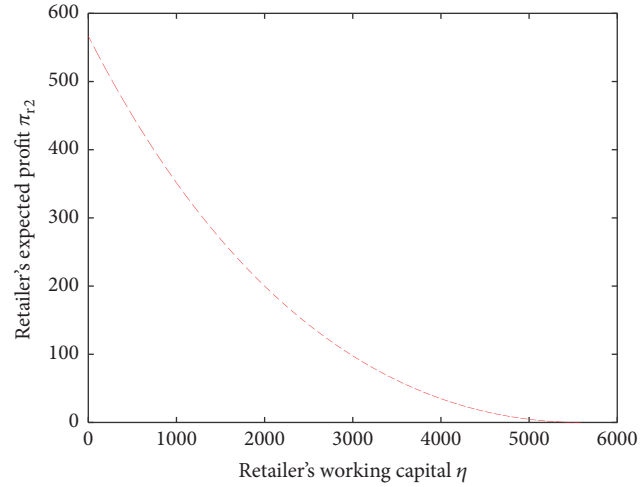



Figure 3: Change trend of retailer's expected profit $\pi_{r2}$ with retailer's working capital $\eta$.

comparing with profit under symmetric information. And the right part is the losses of unmarketable products increased by reducing retailer's effort. Therefore, under asymmetric information, if the profit loss of supply chain under incentive is lower than the unmarketable losses increased with moral hazard, the manufacturer is willing to take the incentive measures. Otherwise, the manufacturer tolerates the moral hazard rather than taking any incentive measures.

## 6. An Illustrating Example

We now employ numerical examples to further analyze the model under symmetric information and asymmetric information. In the supply chain configuration, we suppose that $p = 80$, $c = 40$; assume the retailer's effort cost follows $\psi(e) = e^2/2$ [37]; let the market demand with retailer's effort effects be $y$ and $y$ is subject to uniform distribution, that is, $y \sim U(e, e + 200)$; let the retailer's reserved profit $\pi_{r0} = 0$.

At first, we analyze the optimal order quantity and effort level of the integrated supply chain. The Hessian matrix of the binary function $\pi(q, e)$ is $|H| = 1/2 > 0$. And there are $\partial^2\pi/\partial e^2 = -3/2 < 0$ and $\partial^2\pi/\partial e^2 = -1/2 < 0$. Therefore, according to optimization theory, $\pi(q, e)$ has the only optimal solution $((q^o, e^o) = (120, 40)$ as the order quantity and effort level, and the profit of the supply chain is 2400.

Then, we consider the symmetric information. Under this situation, the retailer gets zero profit and the manufacturer gets all supply chain's profit 2400. The contract is $(t_1(q^o, e^o), q^o, e^o)$ and the total transfer payment is shown in Figure 2, when $\eta \leq 5600$. The total transfer payment increases with retailer's working capital decreasing, which is described in Property 2. When $\eta \geq 5600$, there is $(t - \eta)/p \leq e^o$, which means there is no default risk to the manufacturer and the retailer can pay all the account $t - \eta$ under the demand $e^o$. And when $\eta \geq 5600$, the total transfer payment is 9600.

At last, we consider the situation with asymmetric information. When the manufacturer asks for quantity and effort level $(q^o, e^o) = (120, 40)$ in the contract, the retailer will order quantity $q^o$ and exert effort level $e_2^*$ instead of $e^o$ to get profit $\pi_{r2}$ which is more than zero. As seen in Figure 3 and described in Property 5, the retailer's expected profit decreases with his working capital increasing, when he exerts effort level $e_2^*$.

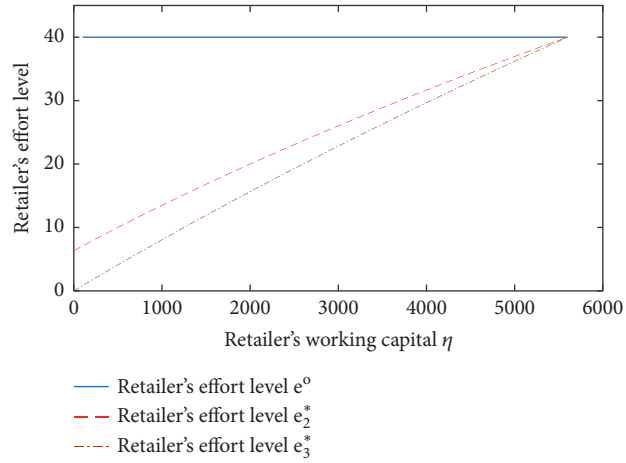



FIGURE 4: Change trend of retailer's effort level with retailer's working capital $\eta$.

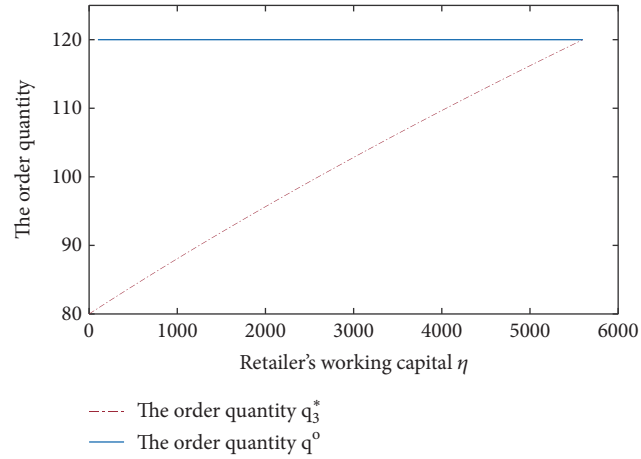



FIGURE 5: Change trend of order quantity with retailer's working capital $\eta$.

When $\eta \geq 5600$, the manufacturer has no default risk, so the retailer's expected profit is zero.

We can draw retailer's effort level $e_2^*$ and $e^o$ in Figure 4. As concluded in Property 4, Figure 4 shows that $e_2^*$ is lower than $e^o$ and increases with $\eta$ increasing when $\eta < 5600$. This is because the default risk is shared by the manufacturer. And when all default risk is taken by the retailer, which means $\eta \geq 5600$, he will exert effort level $e^o$.

Under the situation with asymmetric information, if the manufacturer takes incentive contract, the retailer's effort level $e_3^*$ is shown in Figure 4 and the order quantity $q_3^*$ is shown in Figure 5. The retailer's effort level $e_3^*$ increases with his working capital increasing. But it is lower than the optimal effort level $e^o$ and the effort level $e_2^*$. The order quantity $q_3^*$ increases with retailer's working capital increasing and is below the optimal order quantity. When the manufacturer takes no default risk, effort level is $e^o$ and the order quantity is $q^o$.

As seen in Figure 6, the manufacturer's expected profit increases with an increase in retailer's working capital $\eta$, under the situation with asymmetric information. This means that the manufacturer's loss caused by moral hazard decreases with an increase in $\eta$. These are described in Properties 5 and 6.

The manufacturer can get more profit when he takes incentives contract. And this increased profit comes from the order quantity reduced as seen in Figure 5. When the manufacturer takes no default risk, he gets all the profit of the supply chain under both the moral hazard and the incentive contract.

## 7. Conclusions

Trade credit is widespread for financing in commercial activities. And it is also an important issue in supply chain management. Like a double-edged sword, trade credit has been shown to be good for the supply chain but also results in default risks to the manufacturer and even supply chain crisis. Much literature in supply chain management has studied the retailer's stocking policies and the manufacturer's trade credit contracts. Only a few papers studied trade credit considering

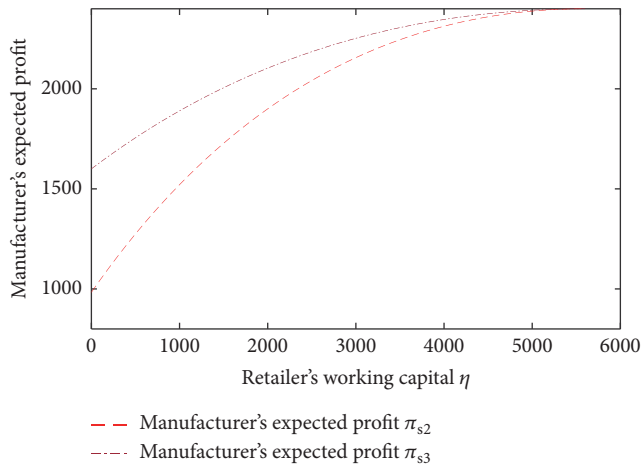



Figure 6: Change trend of manufacturer's expected profit with retailer's working capital $\eta$.

the default risk, and few papers combined the default risk and moral hazard.

This research sheds light on trade credit study in supply chain management with default risk and moral hazard. The default risk is coming from the uncertain market influenced by retailer's sale effort. Retailer's effort level is his private information. The principal-agent models are developed using incentive theory in a newsvendor framework, under the situation of symmetric information and the situation of asymmetric information separately. By model analysis and numerical examples, some conclusions and managerial implications are obtained as follows.

Under the situation of symmetric information, the order quantity, the sales effort, and the profit of the whole supply chain are the same as those in the integrated supply chain. The retailer only retains the reserved profit, and the manufacturer gets all other residual profit of the supply chain by an incentive transfer payment which decreases with an increase in retailer's working capital. The manufacturer should ask more transfer payment when the retailer has less working capital because less working capital means more default risk.

Under the situation of asymmetric information, the retailer pays less sales effort than what he pays under the symmetric information. Without the incentive contract, we can find that the retailer's sales effort increases with an increase in his working capital, but the retailer's profit decreases with an increase in his working capital. At the same time, the manufacturer's profit increases with an increase in the retailer's working capital. This means that when the effort level cannot be observed by the manufacturer, by reducing his effort level, the retailer can get more profit than the residual profit. And the retailer with less working capital can get more profit by executing less effort. This action brings profit loss to the manufacturer. It is better for the manufacturer refusing the retailer with low working capital, to avoid serious moral hazard and profit loss.

By an incentive model, it is found that the order quantity is lowered in the incentive contract to mitigate the impact of moral hazard. And under certain condition the manufacturer can take incentive contract and reduce the order quantity to improve his profit. Specifically, if the profit loss of supply chain by reducing the order quantity is lower than the unmarketable losses by keeping the optimal quantity in contract, the manufacturer should take the incentive measures. Otherwise, the manufacturer should tolerate the moral hazard instead of taking any incentive measures. Furthermore, if the retailer has sufficient working capital, the manufacturer will have no default risk and is advised to take no incentive.

We have considered a supply chain with a risk-neutral manufacturer and a risk-neutral retailer. Nevertheless, the supply chain may consist of multiple retailers and multiple suppliers. What influence does the competition among multiple retailers have on the trade credit? Will the impact of the moral hazard be reduced by the competition? Is there a need for the incentive contract for the manufacturer? the assumption of risk-neutral should also be relaxed, and channel partner's different attitudes toward risk will be considered. For instance, the model can be extended to study the willingness of the risk-aversion manufacturer.

Only retailer's sales effort, which influences the uncertain demand and is his private information, is considered in this research. The manufacturer can also influence the market demand by his promotional activities. And the true effort and cost of these activities are the manufacturer's private information. This means that the moral hazard may be caused by the manufacturer's private information. This case calls for a new incentive scheme to make the retailer accept the contract menu. And this study would be extended to the bilateral moral hazard, in which situation both the retailer and the manufacturer have private information influencing the uncertain demand. The bilateral incentive model may be needed to reduce both the retailer's unsold risk and the manufacturer's default risk.

Another extension would be to introduce the Pareto improving to the incentive contract. In this paper, we design the incentive mechanism and seek the incentive contract to mitigate the impact of moral hazard. It is necessary and has

bothered us for a long time how should Pareto improving be introduced to the incentive mechanism.

In addition, we may extend this framework to the multi-stage trade credit. If the retailer cannot pay the account at the first stage because of the bad sales, should the manufacturer provide trade at the second stage to recover the remaining debt of the first stage? This is interesting and is helpful to manufacturers who have bad account because of trade credit.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

All the authors declare that they have no conflict of any competing financial, professional, or personal interests with other parties.

## Acknowledgments

This work was supported by the MOE (Ministry of Education in China) Project of Humanities and Social Sciences [Grant no. 14YJC630020]; Sichuan Science and Technology Program [Grant.no. 2019JDR0210]; the National Natural Science Foundation of China [Grant nos. 71502037, 71501019]; Chengdu University of Technology [Grant no. KYTD201406]; and Sichuan University [Grant no skqy201768].

## References

[1] Y. Zhong, J. Shu, W. Xie, and Y.-W. Zhou, "Optimal trade credit and replenishment policies for supply chain network design," *OMEGA - The International Journal of Management Science*, vol. 81, pp. 26–37, 2018.

[2] J. Heydari, M. Rastegar, and C. H. Glock, "A two-level delay in payments contract for supply chain coordination: The case of credit-dependent demand," *International Journal of Production Economics*, vol. 191, pp. 26–36, 2017.

[3] P. Pramanik, M. K. Maiti, and M. Maiti, "A supply chain with variable demand under three level trade credit policy," *Computers & Industrial Engineering*, vol. 106, pp. 205–221, 2017.

[4] D. Fabbri and L. F. Klapper, "Bargaining power and trade credit," *Journal of Corporate Finance*, vol. 41, pp. 66–80, 2016.

[5] S. Carbó-Valverde, F. Rodríguez-Fernández, and G. F. Udell, "Trade credit, the financial crisis, and sme access to finance," *Journal of Money, Credit and Banking*, vol. 48, no. 1, pp. 113–143, 2016.

[6] L.-H. Chen and F.-S. Kang, "Coordination between vendor and buyer considering trade credit and items of imperfect quality," *International Journal of Production Economics*, vol. 123, no. 1, pp. 52–61, 2010.

[7] D. Chor and K. Manova, "Off the cliff and back? Credit conditions and international trade during the global financial crisis," *Journal of International Economics*, vol. 87, no. 1, pp. 117–133, 2012.

[8] G. P. Cachon, "Supply chain coordination with contracts," *Handbooks in Operations Research and Management Science*, vol. 11, pp. 227–340, 2003.

[9] S. K. Goyal, "Economic order quantity under conditions of permissible delay in payments," *Journal of the Operational Research Society*, vol. 36, no. 4, pp. 335–338, 1985.

[10] S. P. Aggarwal and C. K. Jaggi, "Ordering policies of deteriorating items under permissible delay in payments," *Journal of the Operational Research Society*, vol. 46, no. 5, pp. 658–662, 1995.

[11] J.-T. Teng, "Optimal ordering policies for a retailer who offers distinct trade credits to its good and bad credit customers," *International Journal of Production Economics*, vol. 119, no. 2, pp. 415–423, 2009.

[12] S. Tiwari, L. E. Cárdenas-Barrón, M. Goh, and A. A. Shaikh, "Joint pricing and inventory model for deteriorating items with expiration dates and partial backlogging under two-level partial trade credits in supply chain," *International Journal of Production Economics*, vol. 200, pp. 16–36, 2018.

[13] F. Lin, T. Jia, F. Wu, and Z. Yang, "Impacts of two-stage deterioration on an integrated inventory model under trade credit and variable capacity utilization," *European Journal of Operational Research*, vol. 272, no. 1, pp. 219–234, 2019.

[14] L. Feng and Y.-L. Chan, "Joint pricing and production decisions for new products with learning curve effects under upstream and downstream trade credits," *European Journal of Operational Research*, vol. 272, no. 3, pp. 905–913, 2019.

[15] K.-J. Chung, L. Eduardo Cárdenas-Barrón, and P.-S. Ting, "An inventory model with non-instantaneous receipt and exponentially deteriorating items for an integrated three layer supply chain system under two levels of trade credit," *International Journal of Production Economics*, vol. 155, pp. 310–317, 2014.

[16] J.-J. Liao, K.-N. Huang, K.-J. Chung, P.-S. Ting, S.-D. Lin, and H. M. Srivastava, "Lot-sizing policies for deterioration items under two-level trade credit with partial trade credit to credit-risk retailer and limited storage capacity," *Mathematical Methods in the Applied Sciences*, vol. 40, no. 6, pp. 2122–2139, 2017.

[17] S.-M. Hosseini-Motlagh, M. Nematollahi, M. Johari, and B. R. Sarker, "A collaborative model for coordination of monopolistic manufacturer's promotional efforts and competing duopolistic retailers' trade credits," *International Journal of Production Economics*, vol. 204, pp. 108–122, 2018.

[18] J. Luo, "Buyer-vendor inventory coordination with credit period incentives," *International Journal of Production Economics*, vol. 108, no. 1-2, pp. 143–152, 2007.

[19] S. Yang, K.-S. Hong, and C. Lee, "Supply chain coordination with stock-dependent demand rate and credit incentives," *International Journal of Production Economics*, vol. 157, no. 1, pp. 105–111, 2014.

[20] C. H. Lee and B.-D. Rhee, "Trade credit for supply chain coordination," *European Journal of Operational Research*, vol. 214, no. 1, pp. 136–146, 2011.

[21] K.-R. Lou and W.-C. Wang, "Optimal trade credit and order quantity when trade credit impacts on both demand rate and default risk," *Journal of the Operational Research Society*, vol. 64, no. 10, pp. 1551–1556, 2013.

[22] C. Wu, Q. Zhao, and M. Xi, "A retailer-supplier supply chain model with trade credit default risk in a supplier-Stackelberg game," *Computers & Industrial Engineering*, vol. 112, pp. 568–575, 2017.

[23] Y.-C. Tsao, "Trade credit and replenishment decisions considering default risk," *Computers & Industrial Engineering*, vol. 117, pp. 41–46, 2018.

[24] S. K. Chaharsooghi and J. Heydari, "Supply chain coordination for the joint determination of order quantity and reorder point

using credit option," *European Journal of Operational Research*, vol. 204, no. 1, pp. 86–95, 2010.

[25] A. Arkan and S. R. Hejazi, "Coordinating orders in a two echelon supply chain with controllable lead time and ordering cost using the credit period," *Computers & Industrial Engineering*, vol. 62, no. 1, pp. 56–69, 2012.

[26] J. Heydari, "Coordinating replenishment decisions in a two-stage supply chain by considering truckload limitation based on delay in payments," *International Journal of Systems Science*, vol. 46, no. 10, pp. 1897–1908, 2015.

[27] P. Kouvelis and W. Zhao, "Financing the newsvendor: supplier vs. bank, and the structure of optimal trade credit contracts," *Operations Research*, vol. 60, no. 3, pp. 566–580, 2012.

[28] X. Chen, "A model of trade credit in a capital-constrained distribution channel," *International Journal of Production Economics*, vol. 159, pp. 347–357, 2015.

[29] H. Yang, W. Zhuo, Y. Zha, and H. Wan, "Two-period supply chain with flexible trade credit contract," *Expert Systems with Applications*, vol. 66, pp. 95–105, 2016.

[30] Y.-C. Tsao, "Managing default risk under trade credit: Who should implement Big-Data analytics in supply chains?" *Transportation Research Part E: Logistics and Transportation Review*, vol. 106, pp. 276–293, 2017.

[31] M. Agostino and F. Trivieri, "Does trade credit play a signalling role? Some evidence from SMEs microdata," *Small Business Economics*, vol. 42, no. 1, pp. 131–151, 2014.

[32] K. Wang, R. Zhao, and J. Peng, "Trade credit contracting under asymmetric credit default risk: screening, checking or insurance," *European Journal of Operational Research*, vol. 266, no. 2, pp. 554–568, 2018.

[33] R. Bastos and J. Pindado, "An agency model to explain trade credit policy and empirical evidence," *Applied Economics*, vol. 39, no. 20, pp. 2631–2642, 2007.

[34] A. M. Costello, "Mitigating incentive conflicts in inter-firm relationships: Evidence from long-term supply contracts," *Journal of Accounting and Economics*, vol. 56, no. 1, pp. 19–39, 2013.

[35] P. Csóka, D. Havran, and N. Csóka, "Corporate financing under moral hazard and the default risk of buyers," *Central European Journal of Operations Research*, vol. 23, no. 4, pp. 763–778, 2015.

[36] Z. Wang and S. Liu, "Supply chain coordination under trade credit and quantity discount with sales effort effects," *Mathematical Problems in Engineering*, vol. 2018, Article ID 2190236, 15 pages, 2018.

[37] J. Heydari and J. Asl-Najafi, "A revised sales rebate contract with effort-dependent demand: a channel coordination approach," *International Transactions in Operational Research*, pp. 1–32, 2018.

[38] Z. Lin, C. Cai, and B. Xu, "Supply chain coordination with insurance contract," *European Journal of Operational Research*, vol. 205, no. 2, pp. 339–345, 2010.

Hindawi
Complexity
Volume 2019, Article ID 3190830, 11 pages
https://doi.org/10.1155/2019/3190830

*Research Article*

# Analysis of the Earned Value Management and Earned Schedule Techniques in Complex Hydroelectric Power Production Projects: Cost and Time Forecast

**P. Urgilés, J. Claver, and M. A. Sebastián**

*Department of Construction and Manufacturing Engineering, National University of Distance Education (UNED), C/Juan del Rosal 12, 28040 Madrid, Spain*

Correspondence should be addressed to P. Urgilés; purgiles1@alumno.uned.es and J. Claver; jclaver@ind.uned.es

Received 21 January 2019; Revised 7 March 2019; Accepted 13 March 2019; Published 1 April 2019

Guest Editor: Julio Blanco-Fernández



All projects take place within a context of uncertainty. That is especially noticeable in complex hydroelectric power generation projects, which are affected by factors such as the large number of multidisciplinary tasks to be performed in parallel, long execution times, or the risks inherent in various fields like geology, hydrology, and structural, electrical, and mechanical engineering, among others. Such factors often lead to cost overruns and delays in projects of this type. This paper analyzes the efficiency of the Earned Value Management technique and its Earned Schedule extension, as means of forecasting costs and deadlines when applied to complex hydroelectric power production projects. It is worth noting that this analysis was based on simulation models applied to real-life projects. The results showed that cost forecasting becomes very accurate over time, whereas duration forecasting is not reliably accurate.

## 1. Introduction

The production of hydroelectric power is based on the installation and operation of large generation plants, which are considered to be highly complex projects. Examples from the international arena demonstrate complexity with statistics revealing significant deviations in terms of costs and execution periods compared to scheduled costs and deadlines. Various studies, such as those by the University of Oxford [1], the Institute for Energy & the Environment [2], and Awojobi & Jenkins [3], analyzed samples of 58 to 235 hydroelectric projects worldwide. The results are surprising, with average values ranging from increases of 27% to 99% over planned costs and increases up to 44% in terms of execution periods.

A hydroelectric power production project is made up of many interrelated and interdependent subsystems. These subsystems entail several engineering disciplines, like electrical, electronic, civil, mechanical, industrial, and environmental engineering, among others. Given such a context, a number of different organizations will naturally be involved in these types of projects. Complex projects are considered to be dynamic systems that develop within an environment of great uncertainty and unpredictability [4]. They call for an exceptional level of management and in fact applying systems developed for ordinary projects is normally insufficient for the level of complexity involved in the execution of such complex projects. The effect of complexity on the project is that it hinders the identification of goals and objectives and it bears upon the duration, costs, and quality of the project, which in turn affects the project's development and management [5–7].

Properly identifying and handling the complexity of a project in its early stages is a critical factor for success. Implementing follow-up and control tasks during the development of a project aims at monitoring performance as well as obtaining projections and measurements that may anticipate possible deviations to the schedule. Time and cost management is closely linked to the monitoring and control of any project and a fundamental support of the processes required to manage project completion within the deadlines. In the same way, they enable the project to be completed

within the approved budget [8–10]. The main factors that affect the complexity of a project are its size, interdependence and interrelations, goals and objectives, and multidiscipline engineering and technology teams, among others [11].

Earned Value Management (EVM) is one of the most widely used techniques worldwide with which to assess a project's performance during its execution. Its application in project management has extended to important institutions like the National Aeronautics and Space Administration (NASA). The basis of this technique was presented by the US Department of Defense (DoD) in the 1960s and was further developed and improved during the 1970s and early 1980s. In 1998, the American National Standards Institute (ANSI) and the Electronic Industries Alliance (EIA) published guidelines for EVM. In the year 2000, the Project Management Institute (PMI) added the terminology and basic formulas of EVM [12]. An important extension to EVM is the Earned Schedule (ES) technique, which analyzes timing performance and provides key indicators based on duration [13].

This paper analyzes the efficiency of EVM and its extension as tools for forecasting costs and duration in complex hydroelectric power production projects. The analysis was based on applying these techniques to simulation models of the schedules of actual power production projects.

## 2. Models and Techniques

To undertake the proposed study, four complex hydroelectric power production projects were considered. Stochastic models were built for all of them, in which the random variables were defined as cost and duration of the jobs that make up the respective execution schedule. The projects used are real-life cases and their characterization is addressed in later sections.

By using numerical methods, simulations were set up that enabled the probability of occurrence by project completion for cost and duration to be calculated. In each simulation, EVM techniques and its ES extension were applied and an efficiency analysis of these techniques as means of forecasting cost and duration was performed. The main stages in the methodology applied are described below.

*2.1. Stochastic Model of the Schedule.* The uncertainty surrounding any complex hydroelectric power production project renders its planning necessarily dynamic and variable over time. In such a context, a simulation model is a powerful tool to enable a virtual study of the project's actual behavior. A good simulation model allows for properties of the project planning and schedules to be depicted, ascertained, and predicted [14, 15]. Furthermore, it is important to bear in mind that the way the schedule is developed has a significant influence on the overall management of any project [16].

The models were based on schedules made up of a distributed series of interrelated tasks, and for each of the tasks likely cost and duration ranges were assigned. Figure 1 shows a graphic illustration of this idea. The dependency between tasks and the allocation of cost and duration to them were defined using conventional techniques: the Program Evaluation and Review Technique (PERT) and the Critical Path Method (CPM). The CPM allowed for total duration of the project to be worked out once the individual duration of each task is known, whereas the PERT provided the means with which to build uncertainty into those durations [17, 18]. The Monte Carlo method was used to simulate the model, which entails a random numerical method that uses statistical sampling techniques to find approximate solutions to quantitative problems [19]. On each run of the models, ten thousand scenarios were simulated for each schedule.

To characterize cost and duration variability in each task, probability distributions were used, specifically the PERT distribution function. In this way, the minimum, most likely, and maximum values were estimated according to American Association of Cost Engineering (AACE) recommendations [20]. In this regard, the research took advantage of the lead author of this paper's 17 years' prior experience in hydroelectric power production projects in the Republic of Ecuador. Cost and duration randomness was estimated on the basis of possible variations in the quantities and technical details of each individual task. Exceptional events that might occur in projects of this type, like errors in prior studies and project designs, lack of financial resources during project execution, or natural phenomena like earthquakes, among others, were not taken into account.

The models were developed on Microsoft Excel using Palisade @Risk software [21]. This software enables the use of Monte Carlo simulation techniques and can generate distributions of possible outcomes of any cell or range of cells on the model spreadsheet [22]. The decision to use @Risk for this research was based on the expertise of the firm that develops the software, which has maintained it on the market since 1987, currently with over 150,000 users in more than 100 countries across the world and with translations into seven languages. Other software options with significant experience are Crystal Ball by Oracle and Risk Simulator by Real Options Valuation [23, 24].

*2.2. EVM Technique and Earned Schedule.* The Earned Value Management technique evaluates the performance of a project during its execution by monitoring the integrated management of its scope, schedule, and costs. Specifically, this technique compares baseline performance with actual performance in terms of duration and costs [25].

To do so, the technique takes a series of fundamental measurements as the basis. The Plan Value (PV) refers to the sum of the planned costs for each task in each time period, from day 1 or the commencement of project implementation up to completion. The Earned Value technique considers this cost planning as the baseline that will serve for future performance comparisons. At another level, Actual Cost (AC) refers to the sum of the costs actually incurred in each task and for each time period. Earned Value (EV) refers to the work actually performed, expressed as a cost. This measurement is calculated by multiplying the percentage of the actual physical progress of each task by its budgeted cost. Finally, Budget at Completion (BAC) is the total budget as estimated in the project plan.

In addition, the Earned Value technique calculates indicators that numerically represent the performance of the

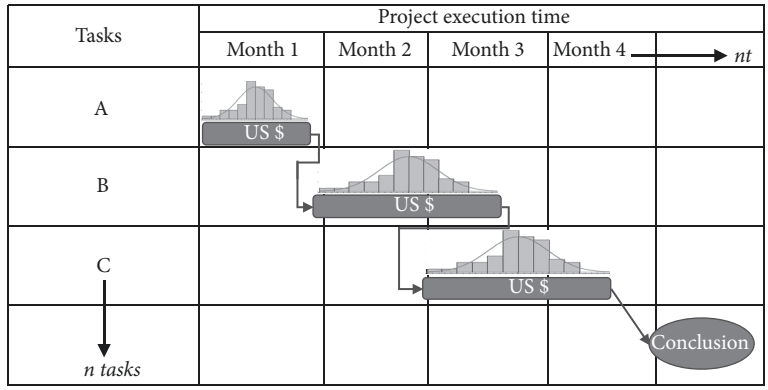



Figure 1: Diagram of a schedule with variable cost and duration.

project in terms of cost and schedule over a given period of time. The main measurements are Cost Variance (CV), Cost Performance Index (CPI), Schedule Variance (SV), and Schedule Performance Index (SPI).

The Earned Value technique analyzes the trends in project performance in terms of cost and duration, i.e., based on the historical performance readings calculated for each period (CV, CPI, SV, and SPI); it makes forward-looking statements and compares them to the initial budget and schedule. The main measurements for forecasting are Estimate at Completion (EAC) and Estimate to Complete (ETC).

Figure 2 is a Cost/Time graph showing the main indicators and predictors the EVM technique calculates.

An important extension to EVM is the technique called Earned Schedule. This technique consists of calculating new values for SV and SP but based on the variable Time and not on Cost, which the Earned Value technique normally calculates. This is achieved by calculating projections on the Earned Value curve for the Plan Value compared to the abscissa of the Time variable [26, 27].

This technique is based on the variables Schedule at Completion (SAC), Earned Schedule (ES), and Actual Time (AT). In addition, the main indicators are Time Estimate at Completion (TEAC) and the Schedule Performance Index (SPI(t)). Figure 2 presents a Cost/Time graph showing the main indicators and predictors that the Earned Schedule technique calculates.

EVM finds application in very varied contexts. It is a technique traditionally applied in the construction industry [28–30], where it has been regularly and successfully applied. But it has also been applied in different highly specialized contexts with a large number of special features and needs. Examples of this can be the projects developed by NASA within the framework of aerospace engineering [12] or the construction of nuclear plants [31]. These examples show the ability of the technique to be adapted to different scenarios. In this sense, there are interesting studies in which EVM is applied to the analysis and improvement of productive processes [32–35].

To verify the efficiency of EVM and its ES extension as techniques for forecasting cost and duration, the indicators contained in these techniques were calculated and compared with results obtained from the simulations, as explained below.

For each of the ten thousand iterations of the model, the cost prediction indicator proposed by EVM and corresponding to EAC was calculated. Furthermore, the simulated final cost (FCS) of the project was calculated and compared with EAC for each iteration. The efficiency of the prediction indicator is confirmed when the difference between EAC and FCS is similar to zero. Accuracy tolerance range was 2% of the total budgeted cost of the project. These calculations and comparisons were made in periods of 5 months up until the scheduled period for project completion.

In addition to that, for each of the ten thousand iterations in the model, the prediction indicator for duration proposed by the ES technique and giving the TEAC was calculated. Furthermore, simulated final project time (FTS) was calculated and compared with TEAC for each iteration. The efficiency of the prediction indicator is confirmed when the difference between the TEAC and FTS is similar to zero. Accuracy tolerance range was 5% of overall scheduled project time. These calculations and comparisons were made in periods of 5 months up until the period scheduled for project completion.

*2.3. Case Studies.* Four actual hydroelectric power production projects were taken as case studies. The basic technical details are given in Table 1. Currently, the projects named Santiago and Cardenillo are at the final design stage and implementation is planned for the coming years. For their part, Mazar Dudas and Sopladora are fully built hydroelectric plants now in operation [36].

The four projects are located on the River Amazon in the southeast of the Republic of Ecuador. The Santiago and Cardenillo projects are located entirely within the province of Morona Santiago at an approximate altitude of 280 and 550 meters above sea level, respectively. The Sopladora project is located in the provinces of Azuay and Morona Santiago at an altitude of 940 meters, while the Mazar Dudas project is located in the province of Cañar at an approximate altitude

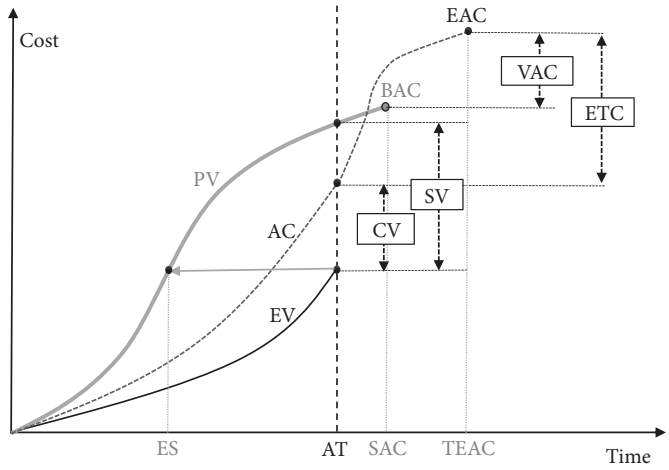



Figure 2: Cost/Time graph showing the main variables in the EVM and Earned Schedule techniques.

Table 1: Basic technical details of the hydroelectric power production projects analyzed.

| Technical information | Hydroelectric project | | | |
|---|---|---|---|---|
| | Santiago | Cardenillo | Mazar Dudas | Sopladora |
| Installed capacity [MW] | 3630 | 595.65 | 20.8 | 487 |
| Plant factor [%] | 47 | 65 | 65 | 60 |
| River | Santiago | Paute | Mazar | Paute |
| Powerhouse type | underground | underground | superficial | underground |
| Turbine Type | Francis | Pelton | Pelton | Francis |
| Units | 6 | 6 | 3 | 3 |
| Hydraulic head [m] | 134 | 372 | 300 | 360 |

of 2,300 meters above sea level. Figure 3 shows their exact locations.

The schedules of the projects under analysis served as the input data for the simulation models. Project schedules list the series of jobs or tasks, how they relate to each other, their costs, and their distribution over time. Table 2 shows the main features of project schedules, where the number of tasks, costs, and total duration are presented in detail. These schedules are the ones prepared and used by the construction companies building the Mazar Dudas and Sopladora plants. In the case of Santiago and Cardenillo, the schedules were created by engineering firms. Depending on each company's criteria, schedules can be more or less detailed in terms of the total number of tasks or steps, which typically range between 120 and 1306. Based on their expertise, the designers and builders of these projects defined duration and total cost within a range of 27 to 75 months and US$56 to US$2,684 million. The forecast duration and costs illustrate that these are large scale projects.

The schedules enabled Accumulated Cost-Duration Curves (S-Curves) to be plotted that represent the baselines for comparing project progress in accordance with the EVM technique. The baselines for the projects in question are presented in Figure 4.

## 3. Results and Discussion

Probability ranges were obtained by simulating schedule implementation at the hydroelectric power production projects in our case study.

Over the full duration of the projects, the models calculated possible ranges of occurrence. Table 3 shows the probabilities calculated for durations equal to or less than the scheduled duration of the projects in this case study. In the same way, estimated duration for each project is shown, with 95% likelihood of occurrence.

The values presented in Table 3 correspond to the results from the relative frequency histograms of duration returned by the stochastic models. Figures 5(a) and 5(b) show the histograms from the Sopladora project model. The abscissa axis represents duration in months, while the top bar shows the calculated probability ranges.

As far as total cost of the projects is concerned, the models estimated possible ranges of occurrence. Table 4 shows the probabilities calculated for costs equal to or less than the planned costs for the projects used in our case study. Estimated costs for each project with a 95% probability of occurrence are also presented.

The values presented in Table 4 correspond to the results of the relative frequency histograms of cost from the stochastic models. Figures 6(a) and 6(b) present the histograms

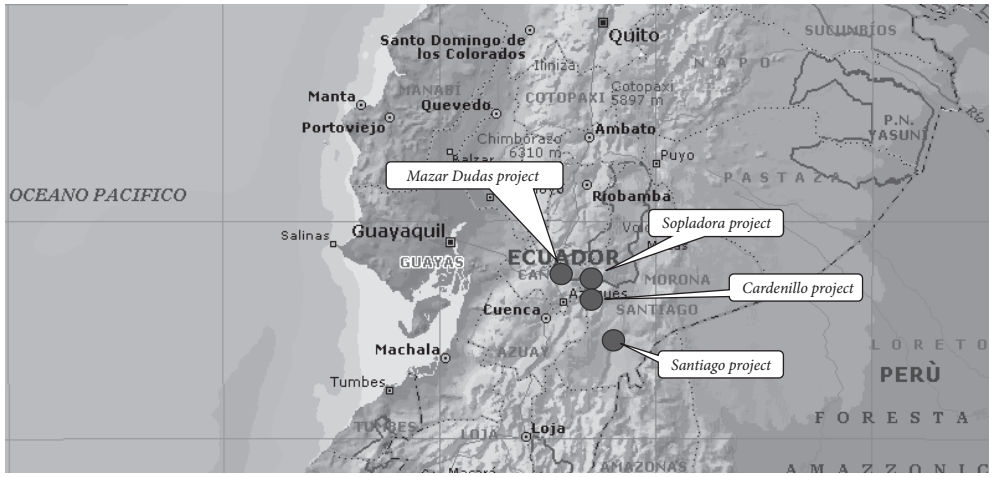



Figure 3: Geographical location of hydroelectric projects in the Republic of Ecuador, prepared from [36].

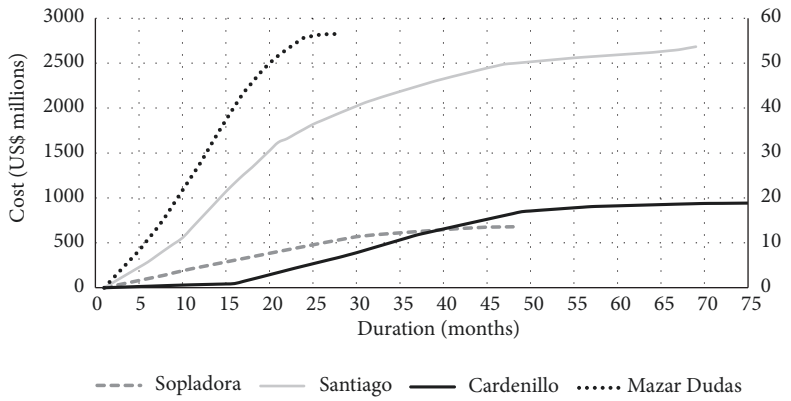



Figure 4: S-Curves that represent the baselines of the projects under analysis. The secondary scale represents the costs of the Mazar Dudas project only.

Table 2: Characteristics of the schedules for the projects under analysis.

| Hydroelectric project | Total tasks | Summary tasks | Scheduled duration [months] | Planned cost [million US$] |
|---|---|---|---|---|
| Santiago | 797 | 29 | 68 | 2684 |
| Cardenillo | 120 | 37 | 75 | 944 |
| Mazar Dudas | 1306 | 29 | 27 | 56 |
| Sopladora | 562 | 29 | 47 | 678 |

Table 3: Outcome of simulation models for total duration.

| Hydroelectric project | Schedule at Completion (SAC) [months] | Duration probability equal to or less than SAC [%] | Estimated duration for 95% probability [months] |
|---|---|---|---|
| Santiago | 68 | 6.8 | 78 |
| Cardenillo | 75 | 4.2 | 83 |
| Mazar Dudas | 27 | 7.9 | 32 |
| Sopladora | 47 | 14.1 | 55 |

Table 4: Results of simulation models in regard to total costs.

| Hydroelectric project | Budget at Completion (BAC) [million US$] | Probability of cost equal to or less than BAC [%] | Estimated cost for 95% probability [million US$] |
|---|---|---|---|
| Santiago | 2684 | 26.1 | 2839.2 |
| Cardenillo | 944 | 26.6 | 999.1 |
| Mazar Dudas | 56 | 21.6 | 59.29 |
| Sopladora | 678 | 22.7 | 714.4 |

Table 5: Accuracy probability rates in cost forecasts (EVM) and duration forecasts (ES) for the Santiago project.

| Analysis period [months] | Earned Value (EV) [millions] [Prob. 95%] | Cost Accuracy Probability (%) [range 2% ± 26.8 million] | Duration Accuracy Probability (%) [range 5% ± 1.7 months] |
|---|---|---|---|
| 5 | 299.00 | 23.00 | 23.20 |
| 10 | 685.70 | 28.30 | 31.80 |
| 15 | 1204.90 | 38.90 | 32.10 |
| 20 | 1613.90 | 43.20 | 36.10 |
| 25 | 1859.90 | 55.60 | 37.70 |
| 30 | 2055.70 | 72.00 | 40.10 |
| 35 | 2215.60 | 86.20 | 42.40 |
| 40 | 2356.20 | 96.20 | 45.30 |
| 45 | 2476.60 | 99.50 | 46.30 |
| 50 | 2528.90 | 100.00 | 47.30 |
| 55 | 2570.30 | 100.00 | 54.60 |
| 60 | 2607.10 | 100.00 | 61.80 |
| 65 | 2653.00 | 100.00 | 52.00 |
| 68 | 2678.00 | 100.00 | 52.10 |
| Completion | 2684.32 | 100.00 | 100.00 |

resulting from the Cardenillo project model. The abscissa axis represents cost in millions of US dollars, while the upper bar shows the estimated probability ranges.

By applying the EVM technique and its ES extension to the models, the probable accuracy of the techniques was determined in regard to the results obtained from the simulations. The closer this probability is to 100%, the greater the efficiency and accuracy of the prediction are.

In regard to the cost and duration forecast, the models produced relative frequency histograms in which the probability ranges were calibrated at 2% of the total project cost and 5% of the total scheduled duration of the project. Figure 7(a) presents an example using the cost histogram for the Santiago project for period 40 and with the probability range calibrated to US $-26.8 and $+26.8 million, representing the 2% range of the total budgeted cost. Meanwhile, Figure 7(b) shows another example, this time the duration histogram for the Mazar Dudas project for period 10 and with the probability range calibrated to +0.7 and -0.7 months, representing 5% of the total scheduled duration.

Details of the cost and duration results obtained for each of the projects in our case study are given below.

*3.1. Santiago Project.* As far as forecast cost upon completion of the Santiago project is concerned, probability readings were obtained that reflect increased efficiency or accuracy of EVM in the various periods of time under analysis. This situation is set out in Table 5. In month 5 of execution, the probability of the predicted cost meeting the cost of the simulation was 23.00%, while in month 45 of execution, that probability reached 99.50% and in the following months it remained at 100% until completion. These probability calculations were based on a range of 2% of total budgeted cost (± US$ 26.8 million).

As far as the forecast for project duration is concerned, as Table 5 shows, the probability rates obtained for the different periods of time in the assessment improved their accuracy over time, reaching 52.10% in month 68, which was the last month in the project schedule. Probability calculations were based on a range of 5% of scheduled duration (±1.7 months).

*3.2. Cardenillo Project.* Turning to the cost forecast upon completion of the Cardenillo project, probability rates were obtained that demonstrated an increase in EVM accuracy or efficiency over the different periods of time under analysis, as seen in Table 6. In month 5 of execution, the probability of the predicted cost matching the simulated cost was 17.80 %, whereas in month 50 of execution, probability had risen to 99.60% and in the following months remained at 100% until completion. Probability calculations were based on a range of 2% of total budgeted cost (±US$ 9.4 million).

Table 6 shows forecast duration, in which the probability rates reveal improved accuracy over time, reaching 59.00% in period 70, which is very close to the scheduled

Table 6: Accuracy probability rates in cost forecasts (EVM) and duration forecasts (ES) for the Cardenillo project.

| Analysis period [months] | Earned Value (EV) [millions] [Prob. 95%] | Cost Accuracy Probability (%) [range 2% ± 9.4 million] | Duration Accuracy Probability (%) [range 5% ± 1.9 months] |
|---|---|---|---|
| 5 | 17.05 | 17.80 | 21.90 |
| 10 | 32.76 | 21.30 | 34.80 |
| 15 | 46.94 | 22.00 | 35.40 |
| 20 | 169.70 | 55.60 | 37.80 |
| 25 | 291.60 | 72.80 | 44.80 |
| 30 | 417.60 | 76.80 | 52.20 |
| 35 | 558.50 | 87.70 | 57.90 |
| 40 | 677.00 | 94.90 | 59.60 |
| 45 | 784.50 | 98.70 | 60.30 |
| 50 | 860.70 | 99.60 | 64.70 |
| 55 | 896.50 | 100.00 | 58.40 |
| 60 | 914.40 | 100.00 | 62.00 |
| 65 | 928.06 | 100.00 | 60.00 |
| 70 | 939.17 | 100.00 | 59.00 |
| 75 | 943.00 | 100.00 | 95.60 |
| Completion | 944.00 | 100.00 | 100.00 |

Table 7: Accuracy probability rates in cost forecasts (EVM) and duration forecasts (ES) for the Mazar Dudas project.

| Analysis period [months] | Earned Value (EV) [millions] [Prob. 95%] | Cost Accuracy Probability (%) [range 2% ± 0.6 million] | Duration Accuracy Probability (%) [range 5% ± 0.7 months] |
|---|---|---|---|
| 5 | 10.85 | 71.00 | 26.70 |
| 10 | 24.75 | 94.20 | 28.10 |
| 15 | 40.51 | 99.80 | 28.80 |
| 20 | 52.19 | 100.00 | 30.40 |
| 25 | 56.41 | 100.00 | 42.00 |
| 27 | 56.50 | 100.00 | 44.10 |
| Completion | 56.55 | 100.00 | 100.00 |

rate upon completion. Only in period 75, which represents scheduled completion, do we see a leap in probable accuracy up to 95.60%. The probability calculations were based on a range of 5% of scheduled duration (±1.9 months).

*3.3. Mazar Dudas Project.* Forecast cost upon completion of the Mazar Dudas project returns probability rates that indicate greater accuracy, as shown in Table 7. In month 5 of execution, the probability of the predicted cost matching the simulated cost was 71.00%, while in month 15 of execution, the probability was 99.80% and in the following months it stood at 100% until completion. The probability calculations were based on a range of 2% of total planned cost (±US$ 0.6 million).

As Table 7 demonstrates in regard to forecast duration of the project, the probabilities obtained improved their accuracy over time, to reach a value of 44.10% in period 27, which is the last period in the project schedule. The probability calculations were based on a range of 5% of scheduled duration (± 0.7 months).

*3.4. Sopladora Project.* From the results presented in Table 8, it appears that the forecast cost upon completion of the Sopladora project shows probability rates with increased accuracy over time. In month 5 of execution, a probability of 64.30% was obtained, whereas in month 30 of execution, probability had risen to 97.80%. In the following months, a probability rate of 100% was maintained until completion. The probability calculations were based on a range of 2% of total budgeted cost (±US$ 6.8 million).

As far as the forecast of duration for the Sopladora project goes, as Table 8 shows, the probabilities obtained improved their accuracy over time up to a value of 74.20% in period 47. The probability calculations were based on a range of 5% of scheduled duration (± 1.2 months).

Having seen the results obtained for the various projects, let us now discuss them as a whole. As far as cost forecasts using EVM are concerned, in Figure 8, one can see how the probability of accuracy in all four projects under study performs. From the execution start-up periods, the tendency is for accuracy to increase in all the projects; i.e., the efficiency of this cost forecasting is seen to improve over time. It is clear

Table 8: Accuracy probability rates in cost forecasts (EVM) and duration forecasts (ES) for the Sopladora project.

| Analysis period [months] | Earned Value (EV) [millions] [Prob. 95%] | Cost Accuracy Probability (%) [range 2% ± 6.8 million] | Duration Accuracy Probability (%) [range 5% ± 1.2 months] |
|---|---|---|---|
| 5 | 102.78 | 64.30 | 40.60 |
| 10 | 212.74 | 76.00 | 40.80 |
| 15 | 310.30 | 87.80 | 43.60 |
| 20 | 407.50 | 93.20 | 44.10 |
| 25 | 502.90 | 95.90 | 44.50 |
| 30 | 577.70 | 97.80 | 47.20 |
| 35 | 621.20 | 100.00 | 42.80 |
| 40 | 656.30 | 100.00 | 43.30 |
| 45 | 676.00 | 100.00 | 59.90 |
| 47 | 678.00 | 100.00 | 74.20 |
| Completion | 678.04 | 100.00 | 100.00 |

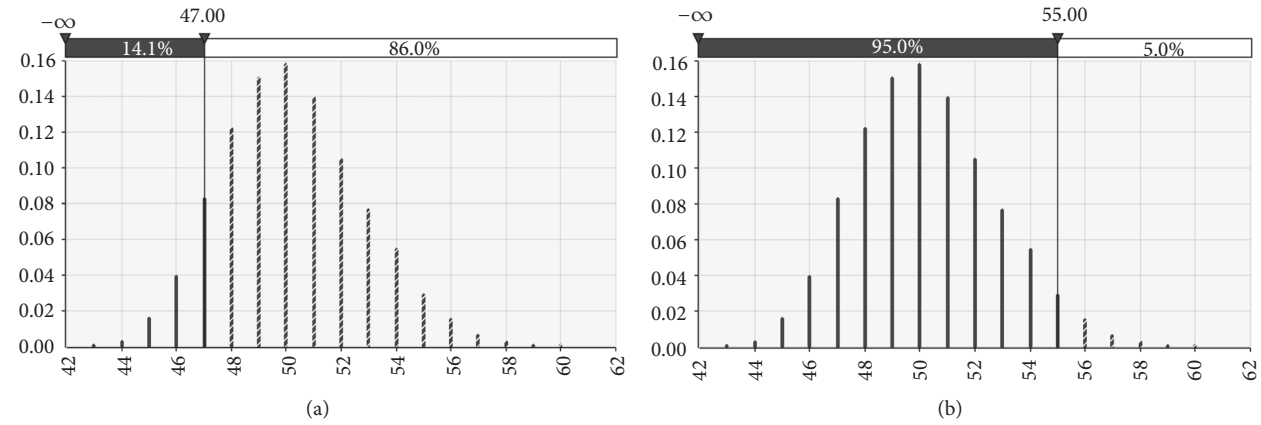



Figure 5: Relative frequency histograms of the duration of the Sopladora project. (a) Duration probability equal to or less than 47 months (SAC). (b) Duration calculated with 95% probability equal to 55 months.

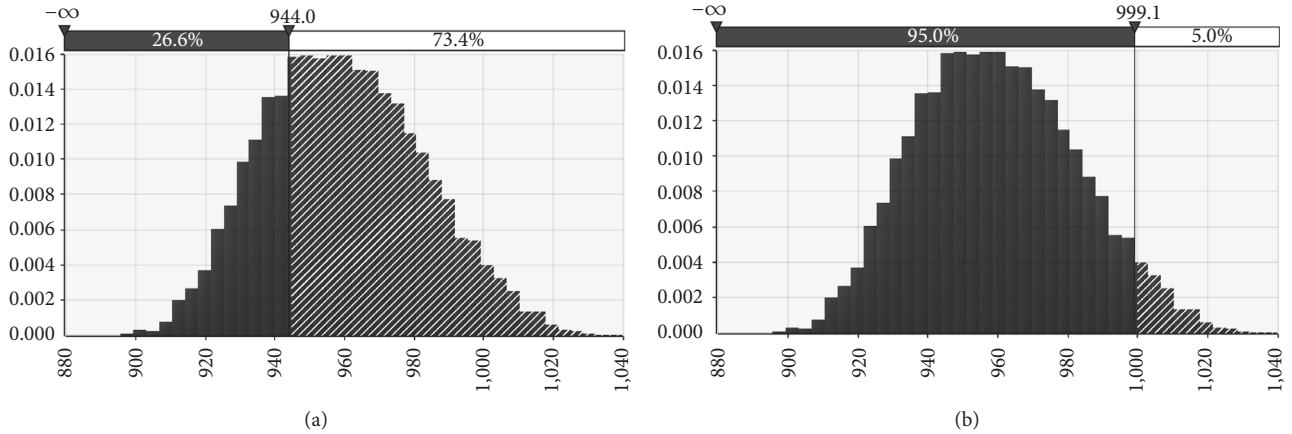



Figure 6: Relative frequency histograms of cost for the Cardenillo project. (a) Cost probability is equal to or less than US $ 944 million (BAC). (b) Cost calculated with 95% probability is equal to US$ 999.1 million.

that a probable forecast accuracy of 100% is achieved in all projects after about 60% of execution time has elapsed and that rate is maintained until the project is concluded.

Turning now to duration forecasting using ES, Figure 9 shows how accuracy performed in the four projects under study. From the execution start-up periods, the probability

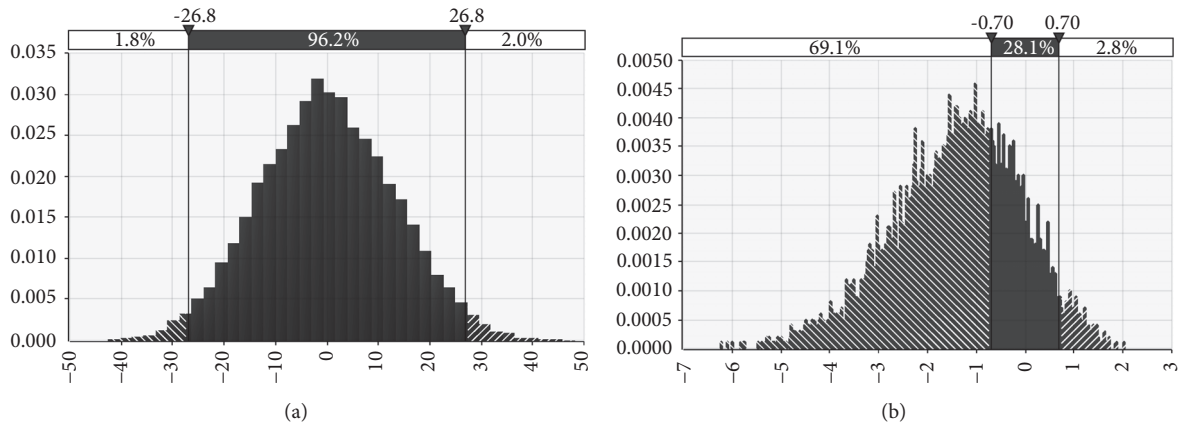



Figure 7: Relative frequency histograms showing the accuracy probability of the forecast produced by the techniques. (a) Histogram of cost accuracy in month 40 for the Santiago project. (b) Histogram of duration accuracy in month 10 for the Mazar Dudas project.

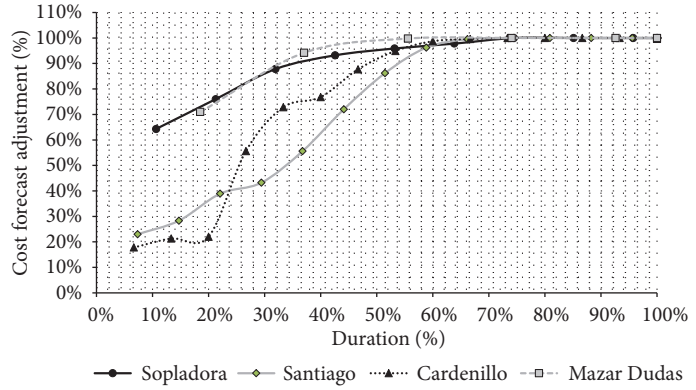



Figure 8: Cost forecasting accuracy with EVM over project execution time.

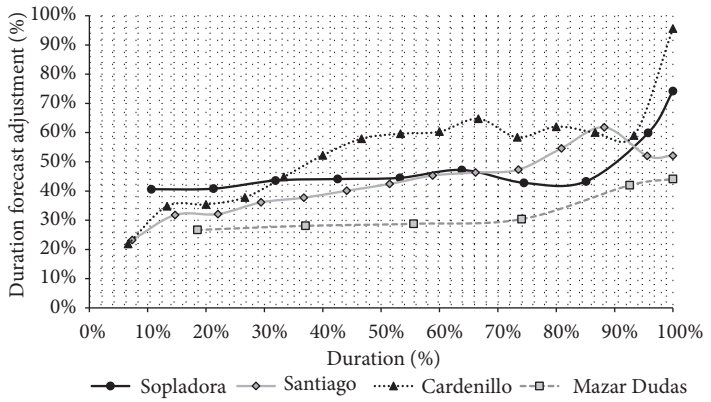



Figure 9: Accuracy of duration forecasts using Earned Schedule over the execution period of the four projects.

of accuracy tended to increase in all projects. However, up until 90% of the duration had elapsed, the improvement in accuracy was minimal, with a maximum of 62% of probable accuracy in the Santiago project and even lower values in the others. After 90% of scheduled duration, the probability of accuracy is seen to improve, although its performance is not the same in all the projects. In fact, this improvement ranges between 44% and 96% at the end of the scheduled duration period.

## 4. Conclusions

From the results shown in Figure 8, it is concluded that using EVM as a tool for predicting the costs of a complex

hydroelectric power production project is seen to produce forecasts whose efficiency improves as project implementation progresses to a point after about 60% of scheduled duration where the cost forecasts it provides match simulated costs with a 100% probability of occurrence.

Complex hydroelectric power production projects are cost intensive and call for capital outlays of many millions of dollars implemented over several years, as shown in Table 2. In this context, EVM is a powerful tool that enables very efficient cost forecasting after approximately 60% of execution time has elapsed, thus allowing sufficient time for decisions to be taken to correct any budget deviations identified.

The ES technique, as an extension of EVM applied to complex hydroelectric power production projects, returned duration forecasts that tend to improve their accuracy as project implementation progresses, as seen in Figure 9. However, these trends are individual and differ among the four projects in our study. Furthermore, until 90% of scheduled duration has elapsed, its forecasting efficiency is seen to be low, as the best duration forecast corresponding to the Santiago project showed an accuracy probability of 62% at 88% of scheduled duration. After 90% of scheduled duration, the tendency is for forecasting efficiency to improve substantially. Nevertheless, the trend varies between the projects, ranging from 44% to 96% accuracy at 100% of the scheduled duration. In light of such performance, it can be concluded that ES is not an accurate technique and does not offer sufficient reliability as a tool for forecasting the duration of complex hydroelectric power plant projects.

While ES indicators are presented in units of time, the S-Curves that the technique uses to project these indicators are based on costs, such as EV and PV. The baseline used by EVM and its ES extension prioritizes the importance of tasks exclusively on the basis of their cost, so that other fundamental variables are not taken into consideration when weighing up the importance of a task, such as duration or the significance of tasks that form part of a critical path. A complex hydroelectric power plant project is characterized by long-lasting tasks and jobs, so it is essential to consider their duration as a principal variable together with the costs.

This paper analyzes the efficiency of EVM and its ES extension as a tool for forecasting cost and duration in complex hydroelectric power production projects and has detected certain inaccuracy in its forecasting of duration. On the basis of this diagnosis and as subsequent work to be done as part of the lead author's Ph.D. thesis, the intention is to develop a new technique as a further extension of EVM that includes duration and criticality of tasks on critical paths as the baselines, in addition to costs. This would entail building a duration forecasting tool of suitable efficiency to afford project managers enough time to make decisions when facing deviations to scheduled deadlines. This initial analysis of how traditional techniques perform when applied to real cases that the lead author has participated in should provide a solid cornerstone in the achievement of such a goal.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that no conflicts of interest exist in regard to the publication of this paper.

## Acknowledgments

This paper is based on the ongoing activities that form part of the lead author's Ph.D. thesis, which he is writing at the International School of Doctorate of the National University of Distance Education; therefore, the authors would like to express their gratitude for the support provided by this institution.

## References

[1] A. Ansar, B. Flyvbjerg, A. Budzier, and D. Lunn, "Should we build more large dams? The actual costs of hydropower megaproject development," *Energy Policy*, vol. 69, pp. 43–56, 2014.

[2] B. Sovacool, A. Gilbert, and D. Nugent, "An international comparative assessment of construction cost overruns for electricity infraestructure," *Energy Research & Social Science*, vol. 3, pp. 152–160, 2014.

[3] O. Awojobi and G. P. Jenkins, "Were the hydro dams financed by the World Bank from 1976 to 2005 worthwhile?" *Energy Policy*, vol. 86, pp. 222–232, 2015.

[4] International Poject Management Association IPMA, *ICB-IPMA Competence Baseline Version 3.0*, Nijkerk, The Netherlands, 2003.

[5] J. Zhu, Q. Shi, P. Wu, Z. Sheng, and X. Wang, "Complexity analysis of prefabrication contractors' dynamic price competition in mega projects with different competition strategies," *Complexity*, vol. 2018, Article ID 5928235, 9 pages, 2018.

[6] S. Kermanshachi, B. Dao, J. Shane, and S. Anderson, "Project complexity indicators and management strategies- a delphi study," *Procedia Engineering*, vol. 145, pp. 587–594, 2016.

[7] J. San Critóbal, "Complexity in project management," *Procedia Computer Science*, vol. 121, pp. 762–766, 2017.

[8] S. W. Azim, *Understanding and Managing Project Complexity*, University of Manchester, 2010.

[9] J. San Cristóbal, "The S-curve enveope as a tool for monitoring and control of projects," *Procedia Computer Science*, vol. 121, pp. 756–761, 2017.

[10] T. M. Bienvenue and R. P. Luc Cassivi, "Project planning and control in social and solidarity economy organizations: a literature review," *Procedia Computer Science*, vol. 121, pp. 692–698, 2017.

[11] J. R. San Cristóbal, L. Carral, E. Diaz, J. A. Fraguela, and G. Iglesias, "Complexity and project management: a general overview," *Complexity*, vol. 2018, Article ID 4891286, 10 pages, 2018.

[12] Y. H. Kwak and F. T. Anbari, "History, practices, and future of earned value management in government: perspectives from NASA," *Project Management Journal*, vol. 43, no. 1, pp. 77–90, 2012.

[13] W. Lipke, O. Zwikael, K. Henderson, and F. Anbari, "Prediction of project outcome The application of statistical methods to earned value management and earned schedule performance indexes," *International Journal of Project Management*, vol. 27, no. 4, pp. 400–407, 2009.

[14] H. Alzraiee, T. Zayed, and O. Moselhi, "Dynamic planning of construction activities using hybrid simulation," *Automation in Construction*, vol. 49, pp. 176–192, 2015.

[15] P. UrgilΘs, J. Claver, M. A. Sebastian, and P. Urgilés, "Analysis of the cost schedule and earned value techniques for the monitoring and control of complex construction projects," in *Proceedings of the in 22nd International Congress on Project Management and Engineering*, Madrid, Spain, 2018.

[16] L. S. Cardona-Meza and G. Olivar-Tost, "Modeling and simulation of project management through the pmbok standar using complex networks," *Complexity*, vol. 2017, Article ID 4791635, 12 pages, 2017.

[17] Mario Vanhoucke, "An overview of recent research results and future research avenues using simulation studies in project management," *ISRN Computational Mathematics*, vol. 2013, Article ID 513549, 19 pages, 2013.

[18] D. F. Muñoz and D. F. Muñoz, "Planeación y control de proyectos con diferentes tipos de precedencias utilizando simulación estocástica," *Información tecnológica*, vol. 21, no. 4, 2010.

[19] A. Garrido and E. M. Conesa, "The Monte Carlo methods for use as criteria generator in construction products quality control," *Informes de la Construcción*, vol. 61, no. 515, pp. 77–85, 2009.

[20] American Association of Cost Engineering AACE International, "Recommended Practice No. 66R-11: Selecting probability distribution functions for use in cost and schedule risk simulation models," AACE International, USA, 2012.

[21] Palisade Corporation, "Palisade," https://www.palisade.com/.

[22] Palisade Corporation, *User's Guide @Risk: Risk Analysis and Simulation Add-In for Microsoft Excel*, Palisade, Ithaca, NY, USA, 2015.

[23] Oracle Company, "Oracle," https://www.oracle.com/applications/crystalball/.

[24] Inc. Real Options Valuation, "Real Option Valuation," http://www.realoptionsvaluation.com/.

[25] Project Management Institute, *A Guide to the Project Management Body of Knowledge PMBOK® Guide*, PMI, Pennsylvania, USA, 6th edition, 2017.

[26] W. Lipke, "Earned schedule contribution to project management," *PM World Journal*, vol. 1, pp. 1–19, 2012.

[27] F. T. Anbari, "Advances in earned schedule and erned value management," in *Proceedings of the Methodology, Time Management, Quality Management*, PMI, Dallas, TX, USA, 2011.

[28] R. G. Solís-Carcaño, G. A. Corona-Suárez, and A. J. García-Ibarra, "The use of project time management processes and the schedule performance of construction projects in Mexico," *Journal of Construction Engineering*, vol. 2015, Article ID 868479, 9 pages, 2015.

[29] A. Ziółkowska and M. Połoński, "Application of the evm method and its extensions in the implementation of construction objects," *Engineering Structures and Technologies*, vol. 7, no. 4, pp. 189–196, 2016.

[30] Ahmad Jrade and Julien Lessard, "An integrated BIM system to track the time and cost of construction projects: a case study," *Journal of Construction Engineering*, vol. 2015, Article ID 579486, 10 pages, 2015.

[31] Y. Jung, B.-S. Moon, Y.-M. Kim, and W. Kim, "Integrated cost and schedule control systems for nuclear power plant construction: Leveraging strategic advantages to owners and EPC firms," *Science and Technology of Nuclear Installations*, vol. 2015, Article ID 190925, 13 pages, 2015.

[32] M. R. Feylizadeh and M. Bagherpour, "Manufacturing performance measurement using fuzzy multi-attribute utility theory and Z-number," *Transactions of FAMENA*, vol. 42, no. 1, pp. 37–49, 2018.

[33] B. Zohoori, A. Verbraeck, M. Bagherpour, and M. Khakdaman, "Monitoring production time and cost performance by combining earned value analysis and adaptive fuzzy control," *Computers & Industrial Engineering*, 33 pages, 2018.

[34] G. H. Martins and M. G. Cleto, "Value stream mapping and earned value analysis: a case study in the paper packaging industry in Brazil," in *Proceedings of the 22nd International Conference on Production Research, ICPR 2013*, Brazil, August 2013.

[35] M. Bagherpour, A. Zareei, S. Noori, and M. Heydari, "Designing a control mechanism using earned value analysis: an application to production environment," *The International Journal of Advanced Manufacturing Technology*, vol. 49, no. 5-8, pp. 419–429, 2010.

[36] Ministerio de Electricidad y Energía Renovable, *Plan Maestro de Electricidad 2016-2025*, MEER, Quito, Ecuador, 2017.

Hindawi
Complexity
Volume 2019, Article ID 9656938, 30 pages
https://doi.org/10.1155/2019/9656938

*Research Article*

# Additive Manufacturing Technologies: An Overview about 3D Printing Methods and Future Prospects

**Mariano Jiménez,[1,2] Luis Romero ,[1] Iris A. Domínguez,[1] María del Mar Espinosa,[1] and Manuel Domínguez[1]**

[1]*Design Engineering Area, Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain*
[2]*Department of Mechanical Engineering, Technical School of Engineering, ICAI, Comillas University, Madrid, Spain*

Correspondence should be addressed to Luis Romero; lromero@ind.uned.es

Received 30 November 2018; Accepted 23 January 2019; Published 19 February 2019

Guest Editor: Jorge Luis García-Alcaraz



The use of conventional manufacturing methods is mainly limited by the size of the production run and the geometrical complexity of the component, and as a result we are occasionally forced to use processes and tools that increase the final cost of the element being produced. Additive manufacturing techniques provide major competitive advantages due to the fact that they adapt to the geometrical complexity and customised design of the part to be manufactured. The following may also be achieved according to field of application: lighter weight products, multimaterial products, ergonomic products, efficient short production runs, fewer assembly errors and, therefore, lower associated costs, lower tool investment costs, a combination of different manufacturing processes, an optimised use of materials, and a more sustainable manufacturing process. Additive manufacturing is seen as being one of the major revolutionary industrial processes of the next few years. Additive manufacturing has several alternatives ranging from simple RepRap machines to complex fused metal deposition systems. This paper will expand upon the structural design of the machines, their history, classification, the alternatives existing today, materials used and their characteristics, the technology limitations, and also the prospects that are opening up for different technologies both in the professional field of innovation and the academic field of research. It is important to say that the choice of technology is directly dependent on the particular application being planned: first the application and then the technology.

## 1. Historical and Current Framework

Many different additive manufacturing (AM) technologies enable the production of prototypes and fully functional artefacts. Although very different in solution, principle, and embodiment, significant functional commonality exists among the technologies.

In order to enter the subject from its origins, before proposing a classification or analyzing the different technologies and their advantages and disadvantages, a chronological analysis of facts will be carried out that will allow us to later base the conclusions. This chronological analysis will be based on dates of publication of works, dates of application for patents, and dates of acceptance of these patents, being aware that in any case the dates in which the developments were reached are always prior to those dates of a public nature.

Without a doubt, the milestone that marked the beginning of additive manufacturing took place on 9th March **1983**, when Charles W. Hull successfully printed a teacup on the first additive manufacturing system: the stereolithography apparatus SLA-1, which he himself built [1].

From then on, there were several advances that paved the way for what is today known as additive manufacturing (Table 1). From a chronological point of view, the most relevant are as follows [2]:

**1986**. Carl R. Deckard, at the University of Texas, develops a "method and apparatus for producing parts by selective sintering", a first step in the development of additive manufacturing by means of selective sintering (SS).

**1988**. Michael Feygin and his team at Helisys, Inc. develop a method for "forming integral objects from

Table 1: Key inventions in additive manufacturing (ordered by publication of patent).

| Technology | Inventors | Patent | Development centre | Request for patent | Publication of patent | Principle of operation |
|---|---|---|---|---|---|---|
| Stereolithography SL | Charles W. Hull | Method and apparatus for production of three-dimensional objects by stereolithography | 3D Systems | 08.08.1984 | 12.02.1986 | Photopolymerization of a photosensitive resin using UV light |
| Selective Sintering SS | Carl R. Deckard | Method and apparatus for producing parts by selective sintering | University of Texas | 17.10.1986 | 21.04.1988 | Selective sintering of powder (fusion – solidification using laser) |
| Material Deposition MD | Scott S. Crump | Apparatus and method for creating three-dimensional objects | Stratasys, Inc. | 30.10.1989 | 01.05.1991 | Deposition of material, using a nozzle, in plastic state (heated by electrical resistance) |
| Jet Prototyping (injection) JP | Emanuel M. Sachs; John S. Haggerty; Michael J. Cima; Paul A. Williams | Three-dimensional printing techniques | Massachusetts Inst. Technology | 08.12.1989 | 09.06.1991 | Injection of binding agent and coloured ink on a bed of powdered material |
| Laminated Manufacturing (cutting) LM | Feygin, Michael; Pak, Sung Sik | Forming integral objects from laminations - Apparatus for forming an integral object from laminations | Helisys, Inc. | 05.10.1988 | 18.04.1996 | Cutting and gluing of laminations with the geometry determined for each layer |

Source: World Intellectual Property Organization (WIPO), https://patentscope.wipo.int.

laminations", an automatic lamination cutting system (laminated manufacturing - LM) that produces layers with the dimensions marked out by the electronic file, layers which will then be bonded to form the final prototype.

**1989**. Scott S. Crump, at the company Stratasys, Inc., develops an "apparatus and method for creating three-dimensional objects", a first step in the development of additive manufacturing by means of fused deposition modelling (FDM).

**1989**. Emanuel M. Sachs and his team, at Massachusetts Inst. Technology, develop "three-dimensional printing techniques", a process of injecting binding agent and coloured ink on a bed of powdered material, using the injectors of a conventional ink-jet printer to do so.

As an evolution of Hull's work based on photopolymerization, other processes have been developed:

(i) Solid Creation System (SCS). Developed by Sony Corporation, JSR Corporation and D-MEC Corporation in 1990.

(ii) Solid Object Ultraviolet Laser Printer (SOUP) Developed by CMET Inc. in 1990.

(iii) Solid Ground Curing (SGC) developed by Cubital Ltd. in 1991

(iv) Inkjet Rapid Prototyping (IRP), the parts are formed by injecting a photopolymer drop by drop which is then cured using ultraviolet light. Developed by Object Geometries Ltd. in 2000, under the name Polyjet.

As an evolution of Deckard's work based on sintering, other processes have been developed [3]:

(i) Direct metal laser sintering (DMLS), where the base material is metal powder and the grains are bonded by sintering, without the grains being fully fused together.

(ii) Selective laser melting (SLM), where metal powder is fully fused together and so the process is not sintering but rather melting.

As an evolution of Crump's work on fused deposition modeling, other processes have been developed:

(i) Metal deposition (MD), where a metal filler material (powder jet or wire) is deposited by a nozzle following the path marked out by the G-code in the .stl or .amf file [4].

(ii) Fused Filament Fabrication (FFF), name from the RepRap community, an open community at RepRap.org, founded by Adrian Bowyer at the University of Bath in 2004 [5].

At this point, hardfacing processes using numerical control (NC) should be mentioned, which are predecessors of fused deposition modeling and prior to the work of Crump, with the difference being that they were not based on electronic files generated by means of solid modeling systems [4]. Although additive manufacturing could date back to automated welding systems (1970s), where a robotic arm controlled by numerical control (G-code) deposited material in welding or hardfacing operations (which may be a similar case), it was not until 1983 that this G-code was used to control a laser that "solidifies" a resin and builds a part using a virtual model (solid model and .stl file).

As an evolution of the work of Sachs and his team on the injection of binding agent or base material, other processes have been developed:

(i) MultiJet Modeling System (MJM) developed by 3D Systems Inc. in 1999, with multiple heads in parallel that move along one axis.

(ii) ModelMaker and Pattern Master, by Solidscape, with one single print head that moves along two axes.

(iii) ProMetal, division of Extrude Hone Corporation, process that binds together steel powder and then infiltrates molten bronze to produce a part that is 40% steel and 60% bronze [6].

Finally, as an evolution of the work of Feygin and his team based on cutting laminations, other processes have been developed:

(i) Selective Deposition Lamination (SDL) Invented in 2003 by MacCormack. The SDL technique works by depositing an adhesive in the area required, both the model and the support, and a blade that cuts the outline of the layer [7].

It is interesting to point out that there are current processes based on more than one of the contributions stated or on integrated processes. This is the case of polyjet modeling (PJM), which is said to be a combination of stereolithography and injection.

Nowadays, the volume of processes, technologies and initialisms is so high that there is no extensive classification system in operation. In Table 2 we can see a list of initialisms used in this field, which is by no means exhaustive, which gives us an idea of how technology is evolving.

In order to highlight the basic pillars of additive manufacturing in Table 2, the abbreviations referring to the technologies discussed at the beginning of this introduction have been highlighted in italic.

Given the large number of initiatives and processes that are developed and patented every day, there is no doubt that additive manufacturing is a technology that will set the standards for many productive processes in the short and medium term. One more proof of this is that the European Union has decided that manufacturing in general and additive manufacturing in particular shall be one of the key tools to tackle some of the European challenges and their subsequent objectives, above all economic growth and the creation of added value and high-quality jobs. This decision is generating the setting up of research and innovation support and promotion programmes aimed at achieving a situation in

Table 2: A sea of initialisms.

| | | | |
|---|---|---|---|
| 3DB | three-dimensional bioplotter | LPD | laser powder deposition |
| 3DP | three-dimensional printing | LPF | laser powder fusion |
| AF | additive fabrication | LPS | liquid-phase sintering |
| ALPD | automated laser powder deposition | LRF | laser rapid forming |
| *AM* | *additive manufacturing* | LS | laser sintering |
| BM | biomanufacturing | $M^3D$ | maskless mesoscale material deposition |
| CAM-LEM | computer-aided manufacturing of laminated engineering materials | MD | metal deposition |
| DCM | direct composite manufacturing | *MD* | *material deposition* |
| DIPC | direct inkjet printing of ceramics | MEM | melted extrusion manufacturing |
| DLC | direct laser casting | MIM | material increase manufacturing |
| DLF | directed light fabrication | MJM | multijet modeling system |
| DLP | digital light processing | MJS | multiphase jet solidification |
| DMD | direct metal deposition | MS | mask sintering |
| DMLS | direct metal laser sintering | M-SL | microstereolithography |
| EBM | electron beam melting | PBF | powder bed fusion |
| EBW | electron beam welding | PJM | poly jet modeling |
| EP | electrophotographic printing | RFP | rapid freeze prototyping |
| ERP | electrophotographic rapid printing | RM | rapid manufacturing |
| FDC | fused deposition of ceramics | RP | rapid prototyping |
| FDM | fused deposition modeling | RPM | rapid prototyping and manufacturing |
| FFEF | freeze-form extrusion fabrication | RT | rapid tooling |
| FFF | fused filament fabrication | RTM | rapid tool maker |
| FLM | fused layer modeling | SALDVI | selective area laser deposition and vapor infiltration |
| FLM | fused layer manufacturing | SCS | solid creation system |
| HSS | high speed sintering | SDL | selective deposition lamination |
| IJP-A | aqueous direct inkjet printing | SDM | shape deposition manufacturing |
| IJP-UV | UV direct inkjet printing | SFC | solid film curing |
| IJP-W | hot-melt direct inkjet printing | SFF | solid free-form fabrication |
| IRP | inkjet rapid prototyping | SGC | solid ground curing |
| *JP* | *jet prototyping* | SHS | selective heat sintering |
| LAM | laser additive manufacturing | *SL* | *stereolithography* |
| LC | laser cladding | SLA | stereolithography apparatus |
| LCVD | laser chemical vapor deposition | SL-C | stereolithography of ceramics |
| LDC | laser direct casting | SLM | selective laser melting |
| LDM | low-temperature deposition manufacturing | SLP | solid laser diode plotter |
| LENS | laser engineered net shaping | SLS | selective laser sintering |
| LFFF | laser free form fabrication | SLSM | selective laser sintering of metals |
| LLM | layer laminated manufacturing | SOUP | solid object ultra-violet laser printer |
| *LM* | *laminated manufacturing* | *SS* | *selective sintering* |
| LMD | laser material deposition | SSM-SFF | semisolid metal solid freeform fabrication |
| LMF | laser metal forming | SSS | solid-state sintering |
| LOM | layered object manufacturing | UC | ultrasonic consolidation |
| LOM | laminated object manufacturing | UOC | ultrasonic object consolidation |

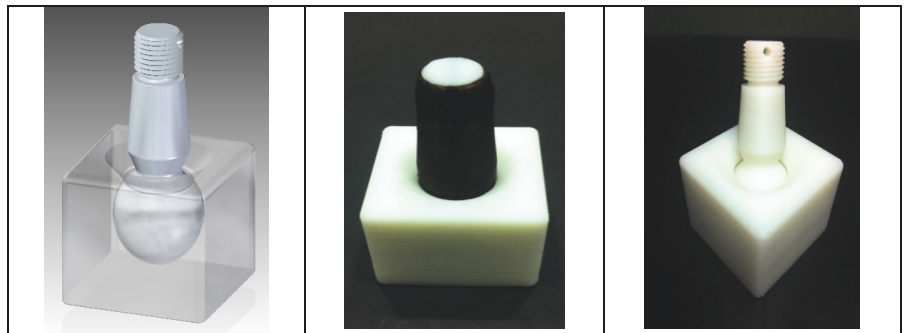



Figure 1: Ball joint mounted in its housing [8].

which additive manufacturing enables the provision of both high-value products and competitive services.

All over the world the additive manufacturing industry is beginning to respond to global, national, and regional standardisation needs via a series of working groups in which the European Union is a key participant: the ISO/TC 261, Additive manufacturing; the ASTM Committee F42 on Additive Manufacturing Technologies; the CEN/TC 438, additive manufacturing; or the AEN/CTN 116, Sistemas industriales automatizados (https://www.aenor.com/) [9].

The ISO standard (ISO 52900) defines additive fabrication as follows [9, 13]:

> "Manufacturing processes which employ an additive technique whereby successive layers or units are built up to form a model."

The terms habitually used in conjunction with additive manufacturing have been evolving at the same pace as the technological developments, and it is convenient to establish a framework of reference that enables an analysis to be carried out of the developments made and of the standardisation required for the future [14]:

(i) "*Desktop manufacturing*", perhaps the first name, in line with the names at the time (1980s) such as desktop computer, desktop design.

(ii) *"Rapid Prototyping"*. This was the first term used to describe the creation of 3D objects by way of the layer-upon-layer method. The technologies that currently exist enable the manufacture of objects that can be considered as being somewhat more than "prototypes".

(iii) *"Rapid tooling". When it became clear that the additive manufacturing system not only enabled us to build prototypes, but also moulds, matrices and tools, this name began to be used to differentiate it from rapid prototyping.*

(iv) *"3D Printing"*. This is the most commonly used term. The term "low-cost 3D printing" is frequently coined when we use printers that domestic or semi-professional users can afford.

(v) *"Freeform Fabrication"*. Is a collection of manufacturing technologies with which parts can be created without the need for part-specific tooling. A computerized model of the part is designed. It is sliced computationally, and layer information is sent to a fabricator that reproduces the layer in a real material.

(vi) *"Additive Manufacturing"*. This is the most recent term applied and it is used to describe the technology in general. It is commonly used when referring to industrial component manufacturing applications and high-performance professional and industrial equipment.

An evolution can be seen in this sequence from obtaining prototypes (with purely aesthetic and geometric goals at the beginning) to simple functional parts, tools, and moulds, to obtain complex functional parts such as those currently obtained via additive manufacturing in the metal industry.

The difference between these techniques is reduced to the application for which the additive manufacturing technology is used. This means that a rapid prototyping technique can be used as a possible technology for the rapid manufacturing of elements, tools, or moulds. It answers the following questions: how would it be possible to manufacture a ball joint fully mounted in its housing without having to manufacture the elements separately prior to their assembly? And is the mass or individual production of this unit possible at an affordable price (Figure 1)?

A characteristic common to the different additive manufacturing techniques is that of the need for a minimum number of phases in the manufacturing process starting with the development of the "idea" by the designer through to obtaining the finished product [15] (Figure 2).

Obviously, the scheme proposed by Yan and his team is a generalization since not in all AM processes G-Codes are created, and preprocessing is not contemplated (necessary in some processes).

In the process described above the designer can perform the entire product manufacturing operation from start to finish. The involvement of another technician is not necessary for carrying out any complementary operations. However, it must be taken into account that, during the process and prior to manufacture, the designer must know the determining

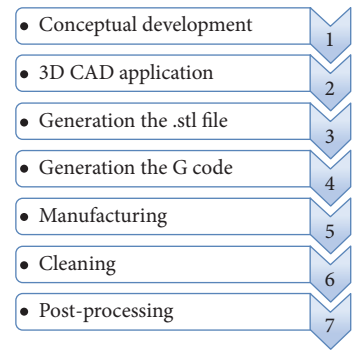
| Detailed process of a model | Conceptual process |
|---|---|
| 1. Conceptual development of the idea.<br>2. Design of the model in a 3D CAD application.<br>3. Generation of an .stl or .amf file to enable the additive manufacturing equipment to interpret the geometrical information (triangulation) modelled in CAD.<br>4. Orientation within the machine and generation of the NC code (G code) by the additive manufacturing equipment.<br>5. Manufacturing of the component.<br>6. Cleaning. Removal of the support material (if the technology uses support material and the component so requires).<br>7. Post-process phase: (improving the finish and hardening. Some technologies do not require this). |  |

Figure 2: Phases of the additive manufacturing process.

factors of the end product in order to be able to select the most suitable manufacturing technique, make the necessary modifications to the geometrical data file (stl or amf file), and review the NC code. The designer must, therefore, have a full overview of and the necessary training in all of the phases of the process [19, 20].

This article contains a full and up-to-date description of the benefits and drawbacks of the most important manufacturing methodologies and processes that exist within the scope of the additive manufacturing concept, of the characteristics of the models manufactured and their associated costs, of the functional models, the applications, and the sectors of influence. At the end, there is a reference to the RepRap community and free software due the importance they acquired last years.

It is important to establish the variables that currently support the implementation of additive manufacturing and its relation to the basic principles of each technology, which allow detecting the advantages and limitations of each of them. In Figure 3, the relationship between the main variables that intervene in the development of additive manufacturing is shown: technologies, materials, type of models, associated costs, and visual appearance. The current scope of these variables and their evolution can be seen in the subdivision that is presented in a circular diagram.

## 2. Processes

*2.1. Additive Manufacturing Technologies.* Conventional component manufacturing processes are based on the use of high-capacity resources combined with control elements to achieve extremely high levels of precision and reliability. The use of computer systems during the design engineering, manufacturing, and simulation phases of a product in combination with other techniques based on mechatronics have succeeded in making production systems highly efficient.

However, a number of limitations still plague manufacturing processes. This is because on occasion we find ourselves being forced to use processes and tools that increase the end cost of the element in accordance with the size of the production run and the geometrical complexity of the component.

Transformation processes currently exist that enable us to extract, shape, fuse, and bind the base material of our component, and for the last few years we have also been able to deposit the material where it is needed; in other words, using a virtual 3D model it is possible to manufacture the component by adding the material in accordance with the solid volume designed into the model.

Current additive-type technologies are based on the dispersion-accumulation principle (Figure 4). The material or additives filler processes are those that involve the solidification of a material the original state of which is solid, liquid, or powder by way of the production of successive layers within a predetermined space using electronic processes [21]. These methods are also known by the acronym MIM (Material Increase Manufacturing).

If we focus our attention on the application of the different manufacturing technologies used for obtaining rapid prototypes, the current technologies can be classified as additive (stereolithography, laser sintering, fused deposition modelling, etc.) and nonadditive (incremental forming, high-speed machining, pressure injection moulding, lost wax, laminating and contouring, etc.).

In Tables 3–8 the most relevant data with respect to the main additive manufacturing technologies are analysed [6, 8, 11, 21, 22].

*2.2. Classification.* The classification of the additive manufacturing processes has been a controversial issue, even bordering on obsession, since the appearance of the first alternatives for obtaining pieces in these technologies. The first classifications took into account whether the starting material was solid, liquid, or powder, posing inconsequential doubts that contributed little to the academic community. The same problem was found in the professional field when the first commercial solutions began to appear, since the


Figure 3: Table of contents.

classifications that were proposed had little or no utility [14, 23–25].

One of the pioneers in the classification of processes, hierarchizing them based on the starting material, is JP Kruth [26], who proposed an interesting classification in 1991. In the literature we can also find other classification alternatives based on the equipment used [27], to the process itself [28] or to the transformation of materials [29]

The classification proposed by Williams and his team [30] is very academic, but of little use from a practical point of view. However, the classification proposed by ASTM has some loopholes and disadvantages, and the same is the case with the classification proposed by ISO.

ASTM (ASTM F2792) also proposes the following groups: binder jetting, directed energy deposition, material extrusion, material jetting, powder bed fusion, sheet lamination, and vat photopolymerization [31].

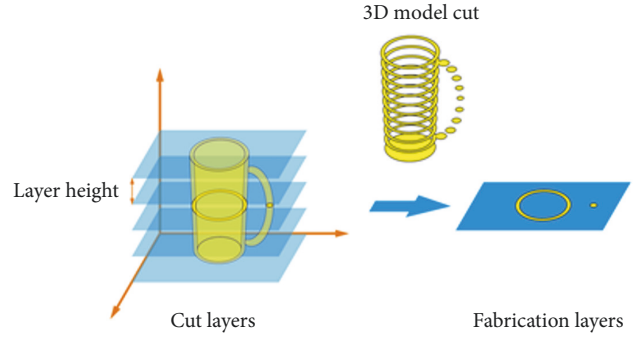



Figure 4: The dispersion-accumulation principle [8].

However, for example, it is not very coherent to separate the photopolymerization processes depending on whether they are carried out in a vat or with another alternative, such

Table 3: Stereolithography (SL).

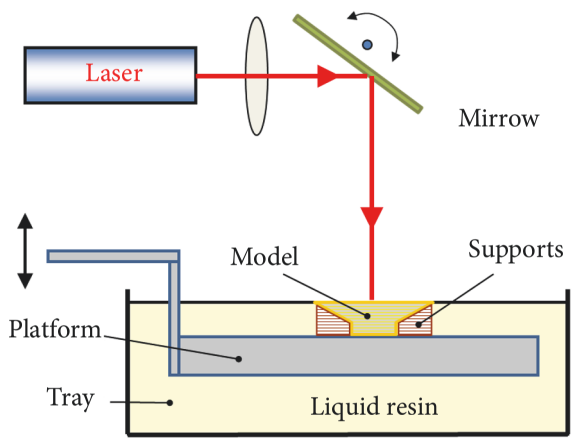
| Resin Solidification - Stereolithography | |
|---|---|
|  | The material undergoes point-to-point solidification due to the photopolymerisation laser being directed onto a 2D cross-section of the model (XY plane). The platform gradually descends (Z plane) in accordance with the height of the layer defined [10].<br>Layer thickness varies from between 0.1 and 0.2 mm. The precision in SL is +/- 0.2% (min. +/- 0.2 mm). Maximum model size: 2100x700x800 mm.<br>The following materials can be used (epoxy-acrylic resins: Poly1500, PP, TuskXC2700T / Tusk2700W, Tusk SolidGrey3000, Flex70B, NeXt, Protogen White, Xtreme, WaterClear [11].<br>The properties of the material NeXt are: tensile modulus: 2370-2490 Mpa; tensile strength: 31-35 Mpa; elongation at rupture: 8-10%; flexural modulus: 2415-2525 Mpa; bending strength: 68-71 Mpa; impact strength: 47-52 J/m; deformation under load temperature: 48-57°C.<br>Advantages and disadvantages<br>The quality and surface finish are good or very good. Great precisions and transparent parts can be obtained.<br>The equipment and materials are medium-high cost. They have problems to obtain pieces with cantilevers or internal holes due to the difficulty of removing the supports. |

Table 4: Selective sintering/melting (SS).

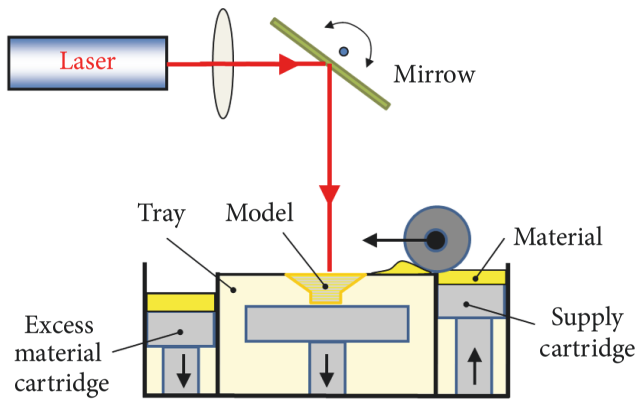
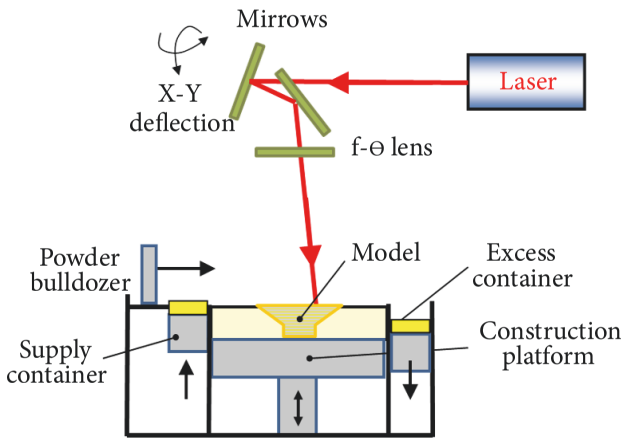
| | |
|---|---|
| Selective Laser Sintering | |
|  | A layer of powder is laid down and a $CO_2$ laser sinters it at the points selected on a 2D cross section of the model (XY plane). The platform gradually descends (Z plane) in accordance with the height of the layer defined [12].<br>Precision is between +/- 0.3% (min. +/- 0.3 mm).<br>The minimum layer thickness is 0.08 mm. Maximum model size 700x380x580 mm.<br>The following materials can be used: Polyamide (PA), Glass filled polyamide (PA-GF), Alumide, PA 2241 FR, TPU 92A-1.<br>Properties of the PA material: tensile modulus: 1650 Mpa; tensile strength: 22 Mpa; elongation at rupture: 20%; flexural modulus: 1500 Mpa; bending strength: - Mpa; impact strength: 53 J/m; deformation under load temperature: 86°C.<br>Advantages and disadvantages<br>Pieces of high quality and precision are obtained. A large quantity of sintering materials is available. They do not present problems to obtain pieces with cantilevers or internal holes because the own dust makes of support.<br>The equipment and materials are medium-high cost. |
| Selective Laser Melting | |
|  | The print nozzle head is fitted with a $CO_2$ laser that is directed via a set of lenses onto the powdered material. The support structures are made of the same material as the model and must undergo a subsequent finishing or even machining process [3].<br>The minimum layer thickness is 0.020 mm.<br>The material can be: stainless steel, Co-Cr, Inconel 625-718, titanium Ti64.<br>Properties of the material Co-Cr: ultimate creep ($R_m$): 1050 Mpa; elongation (E): 14%; Young's modulus: 20 Gpa; Hardness 360 HV.<br>Advantages and disadvantages<br>Pieces of high quality and precision are obtained. There is a large amount of metallic materials to be sintered. Equipment and materials are expensive. They have problems to obtain pieces with cantilevers or internal holes due to the relative difficulty of removing the supports. |

Table 5: Material deposition (MD).

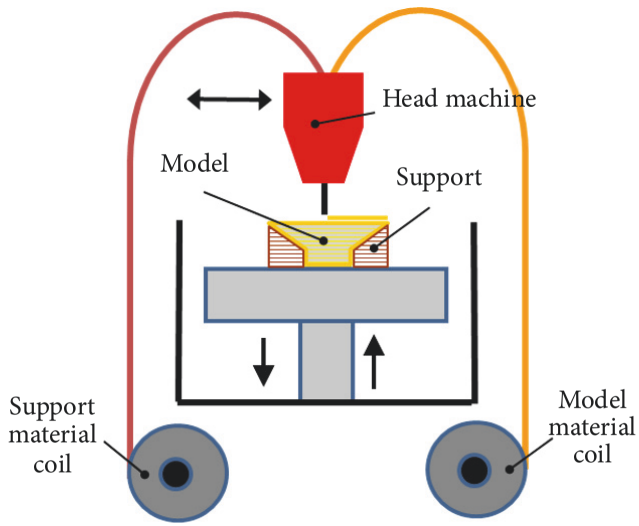
| Fused Deposition Modelling | |
|---|---|
|  | The wire is rolled around a coil and deposited using the thermal nozzle head that moves in accordance with the plane (XY). The platform gradually descends (Z plane) in accordance with the height of the layer defined [16]. The layer thickness is: 0.13 - 0.25mm (for ABS); 0.18 - 0.25 mm (for ABSi); 0.18 - 0.25 mm (for PC); 0.25 mm (for PPSU). The maximum model size is 914x610x914 mm.<br>The following thermoplastic materials can be used: ABS, ABSi, ABS-M30, ABS-ESD7, PC-ABS, PC-ISO and ULTEM 9085. The properties of ABS are: tensile modulus: 1627 Mpa; tensile strength: 22 Mpa; elongation at rupture: 6%; flexural modulus: 1834 Mpa; bending strength: 41 Mpa; impact strength: 107 J/m; deformation under load temperature: 76-90°C.<br>Advantages and disadvantages<br>Low medium cost equipment, available even in domestic environments. Possibility of eliminating supports by dissolution and obtaining very clean pieces. Pieces of quality and precision can be obtained but resorting to high cost equipment. |

Table 6: Jet prototyping (JP).

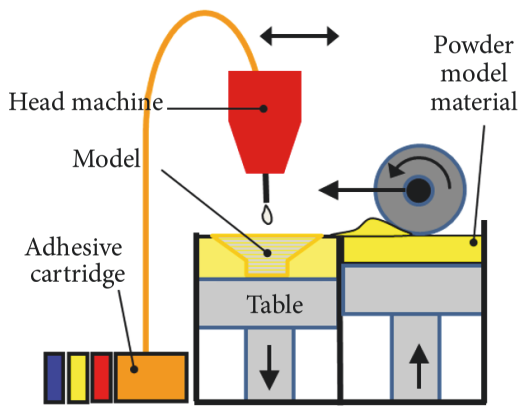
| Three Dimensional Printing – Glue Injection | |
|---|---|
|  | The model is built on a bed full of powdered model material. A nozzle head injects an agglutinate onto the surface of the bed and fuses the powder in accordance with the geometry of the 2D cross-section of the model. The powder is added and levelled using a roller. Once the process has been completed, the excess powder is sucked off the bed leaving the model clean. The model then has to be cured (hardened) using different coatings [17].<br>Minimum layer thickness is between 0.013 and 0.076 mm.<br>The material used can be ceramic, metal and polymers. The properties of the material zp150-Z-Bond are: tensile modulus: - Mpa; tensile strength: 14 Mpa; elongation at rupture: 0.2 %; flexural modulus: 7.2 Mpa; bending strength: 31 Mpa; deformation under load temperature: 112°C.<br>Advantages and disadvantages<br>Pieces of color are obtained, with great aesthetic quality. No supports are needed.<br>It is not easy to obtain functional pieces due to the fragility of the pieces obtained. The cost of the equipment is medium-high. |

Table 7: Laminated manufacturing (LM).

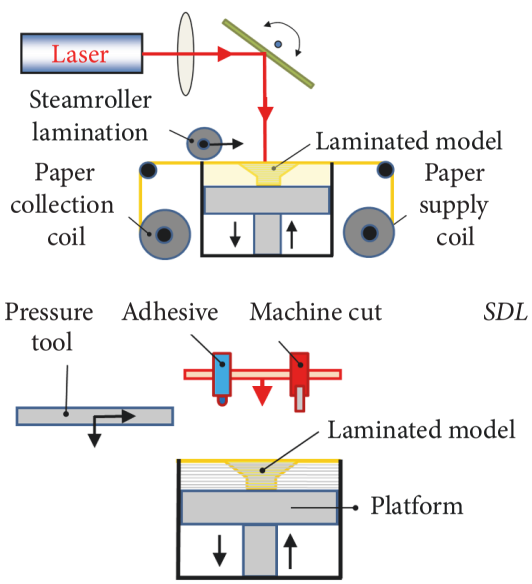
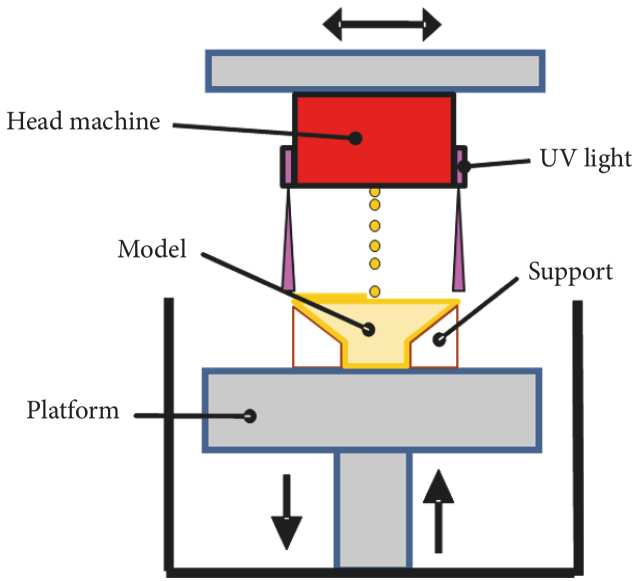
| Selective Deposition Lamination | |
|---|---|
|  | Invented in 2003 by MacCormack, SDL must not be confused with laminated object manufacturing (LOM) technology. LOM uses a laser, laminated paper and an adhesive that fixes the model and support material. The SDL technique works by depositing an adhesive in the area required, both of the model and the support, and a blade that cuts the outline of the layer [7].<br>LOM: The layer thickness is 0.165 mm. The maximum model size is 170x220x145 mm.<br>SDL: Layer thickness corresponds to the thickness of the paper used plus that of the layer of adhesive. Sheets of PVC with the following properties are used: tensile strength: 40-50 Mpa; elongation at rupture: 30-100%; flexural modulus: 1200-200 Mpa; deformation under load temperature: 45-55°C.<br>Advantages and disadvantages<br>Pieces of high quality aesthetic objectives are obtained. The starting materials are low cost. The pieces obtained can be painted. The equipment is of average cost. No functional parts are obtained. |

Table 8: Injection-polymerisation (JP-SL).

| Resin Injection (Projection) and Ultraviolet Light Photopolymerisation | |
|---|---|
|  | A head with thousands of injectors deposits drops of liquid resin that are hardened using two UV ray lights fitted on the sides of the selfsame head. Two materials can be used simultaneously (bi-material pieces) [18].<br>The minimum layer thickness is 0.017 mm.<br>The range of materials is extremely extensive and includes translucent resins, polypropylene, ABS and elastic resins.<br>Properties of the SCI White (PolyJet) material: tensile modulus: 2500 Mpa; tensile strength: 58 Mpa; elongation at rupture: 10-25 %; flexural modulus: 2700 Mpa; bending strength: 93 Mpa; deformation under load temperature: 48°C.<br>Advantages and disadvantages<br>The quality and surface finish are good or very good. Great precisions and transparent parts can be obtained.<br>Equipment and materials are expensive.<br>They have problems to obtain pieces with cantilevers or internal holes due to the difficulty of removing the supports. |

as injection. A difference is made between directed energy deposition and material extrusion when both are, without a doubt, deposition processes.

Moreover, it is not justifiable to distinguish between injection processes when an "adhesive" or a "material" is injected, since, at the end of the day, both materials will end up forming part of the prototype or final part, as is the case of the ProMetal system.

It is also interesting to see that there are two types of injection: (a) when an adhesive is injected (which ends up forming a "material" part of the product) and (b) when a "material" is injected.

Lastly, there are different processes, some very important, that do not fall into any of the groups in this classification, such as in the case of mask sintering or digital light processing, for example.

ISO proposed, in its 2010 working draft, the following ten processes: stereolithography, laser sintering, laser melting, fused layer modeling/manufacturing, multijet modeling, polyjet modeling, 3D printing, layer laminated manufacturing, mask sintering, and digital light processing [32].

There is no doubt that for the simple fact of defining ten processes, other important processes are left out. In this classification, it can be seen that the manufacturers' considerations have more of an influence than logic. In 2015 ISO assumes the ASTM classification with its standard ISO/ASTM 52900:2015 (ASTM F2792).

An additive manufacturing system is in itself a production system and, therefore, for the purposes of classification, the systematics of manufacturing processes should be used. In every manufacturing system, there must be four elements present [4]:

(i) Material

(ii) Energy

(iii) Machine and tool

(iv) Technology (know-how)

From the point of view of the material, we could opt for the classic classification of solids, liquids, powder, etc. [14]; however, for both the engineer who wishes to manufacture the product and their customer, it is more important to expand upon the technical qualities of the material and, thus, we need to begin to classify the processes according to their ability to work with metal materials (with high melting points and which, therefore, require more energy in the process) or with other materials. And it is these technical qualities that we are going to focus on in this article.

From the point of view of energy, it is important to analyse what type of energy is required and how this energy is transmitted. With regards what type of energy is required, this may be as follows:

(i) Heat (electrical resistance, electron beam, etc.)

(ii) UV light (visible or laser)

(iii) Chemical energy (for adhesion processes, chemical reactions, etc.)

With regard to how this energy reaches the material for successful transformation, this may be via the following:

(i) Laser (valid to provide UV light and heat)

(ii) Electrical resistance

(iii) Electron beam

From the point of view of the machine, it is important to analyse the alternatives of smaller machines, suitable for offices, compared to industrial machines and as regards the tools, these shall include the following:

(i) Vats and containers for photosensitive liquids or powder

(ii) Deposition or extrusion nozzles

(iii) Injectors

With regard to technology, the most important variable, it is necessary to know if it is available commercially or if it is only an option available in research centres, as in this case there will likely be a wait involved, although fortunately not long if the technology is valid.

In conclusion to all of this, Table 9 shows the classification system that takes these variables into account: material, energy, machine and tool, and technology.

To specify this classification, which can cover all additive manufacturing processes with simple approaches and that can evolve as technology evolves, we present the classification chain of the five processes analyzed in the introduction, the basic pillars of additive manufacturing (Table 10).

For example, stereolithography (SL) is a process that uses resins as material (2); uses laser as energy (f); is a professional machine (3); uses vats and containers for photosensitive liquids as tools (a); and is available commercially (4). For this reason, stereolithography will have a classification: 2f3a4.

*2.3. Analysis of the Environment.* In order to carry out this study, we worked with leading additive manufacturing companies, both ones that are developing new processes and ones that are part of the market providing service. Contact was also established with researchers in technology centres and universities and their contribution has proved to be highly valuable.

Unsurprisingly, this study entails an exhaustive analysis of the most important literature in the field of additive manufacturing, a study that we cannot include in this paper in full due to the obvious space constraints. Nonetheless, a series of papers must be listed which, due to their special interest in the subject, provide information that has proved fundamental in producing this paper.

In the field of metal additive manufacturing, an analysis has been performed on processes using powder, whether powder injected, powder deposited in layers, or processes using wire [22, 33–35].

As has been discussed, this paper has been approached from both a professional perspective (expanding upon the most interesting innovations in this field) and a research perspective. Therefore, papers that take into account the social impact of additive manufacturing have been analysed

Table 9: Classification matrix of the processes in additive manufacturing.

| | | | | |
|---|---|---|---|---|
| Material | Materials with a low melting point (1) | Resins, transparent materials (2)<br>Flexible materials (3) | Ceramics (4)<br>Powder (5)<br>Sheets (6) | Metals and difficult materials (7) |
| Energy | Heat<br>Electrical resistance (a)<br>Electron beam (b)<br>High power laser (c) | Chemical energy<br>Adhesives (d)<br>Reagents (e) | UV light (visible or laser)<br>Laser (f) or visible light (g) | Mechanical energy<br>Cutting tool (h) |
| Machine and tools | Low cost (such as RepRap) (1)<br>Vats and containers for photosensitive liquids or powder (a) | Office (2)<br>Deposition or extrusion nozzles (b) | Professional (3)<br>Injectors (c) | Industrial (4)<br>Cutting tool (d) |
| Technology | In conceptual and development phase (1) | In experimental phase (2) | Available at research centre (3) | Available commercially (4) |

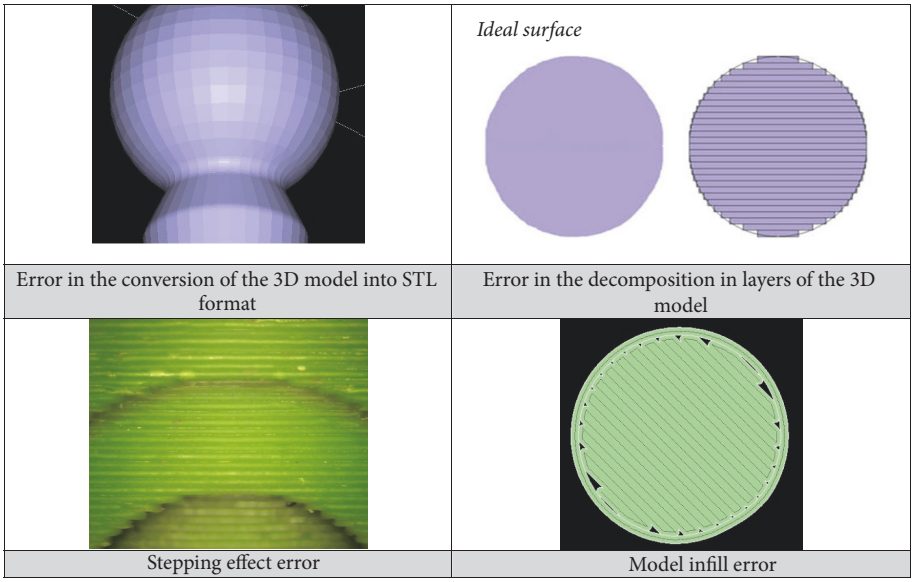



Figure 5: Technology limitations in the rapid prototyping process.

Table 10: Classification of the original processes in additive manufacturing.

| | |
|---|---|
| Stereolithography (SL) | 2 – f – 3 – a – 4 |
| Selective Sintering (SS) | 1 – c – 3 – a – 4 |
| Material Deposition (MD) | 1 – a – 3 – b – 4 |
| Jet Prototyping (JP) | 5 – d – 3 – c – 4 |
| Laminated Manufacturing (LM) | 6 – h – 3 – d – 4 |

[20, 35–37]. Papers focusing on medical issues have also been analysed [38–40] and papers have been located which expand upon design methodologies for additive manufacturing [6], the involvement of concurrent engineering [41], and the study of nonrigid materials [42]. The fields of architecture [23, 43–45] and the automotive industry have also been addressed [6].

References have also been found relating to functional prototypes [46], which show the use that additive manufacturing still has in this regard, along with papers on very contrasting environments, such as the food industry [47]. In the field of research, papers relating to studies of the processes [48] or medical studies with living cells [36, 38, 49–51] are worth mentioning.

Naturally, this study also considers the RepRap movement [5, 21, 52], as although technologically speaking it does not contribute much, it has undoubtedly played an essential role from both a social perspective and one of technology disseminations, and so it cannot be omitted in any serious study on additive manufacturing.

## 3. Analysis

*3.1. Characteristics of the Models Manufactured Using the Addition of Material Technique.* When the three-dimensional model is obtained using a reverse engineering process and the precision of the final model is determined by the scanning process, the virtual 3D modelling, and the additive manufacturing process. If an inverse engineering process has not been used, the models manufactured are determined by the virtual 3D modelling and the manufacturing process used [53].

We have already seen that during the manufacturing process the model is built by way of the depositing of layers on the x-y plane resulting in solid volume being acquired in the direction of the Z axis. This process is characterised by a volume error between the volume of the virtual 3D model and the volume of material obtained in the model and, therefore, the manufacturing precision is the result of superimposing different errors in the production of the model which affect the surface quality, the dimensional accuracy, and the final weight of the model.

The technology limitations that occur in this process are as follows: an error in the conversion of the 3D model into STL format (triangulation of the geometry), an error in the decomposition in layers of the 3D model (exact division of the thickness), stepping effect error (orthogonal deposition of the material by layers), and, finally, model infill error (Figure 5).

There are studies which show that the models obtained using rapid prototyping techniques have an average error in the majority of the 0.05 mm dimensions with respect to the original model or of modelling control in 3D [42].

The additive manufacturing of three-dimensional models with an aesthetic (visual) or assembly objective is achieved using techniques that involve the layer-upon-layer addition of plastic materials, while functional models or those capable of withstanding mechanical testing must be manufactured mainly in metal and, in some cases, in a polymeric material that is subjected to a postproduction hardening process.

Studies undertaken by different research institutes show that those products manufactured in metal using additive technologies provide the same or better mechanical performances than the same products manufactured using conventional processes [22]. The resistance to corrosion of

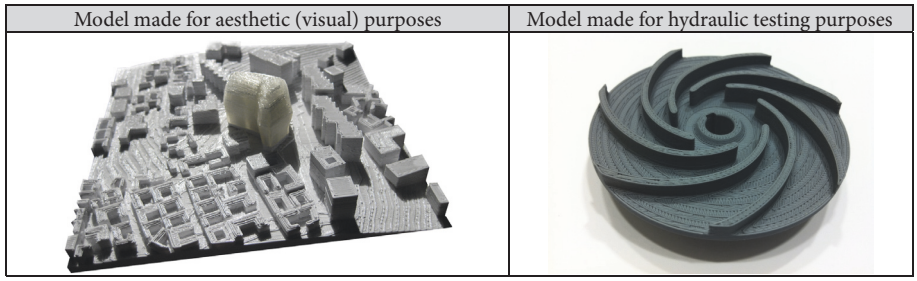



Figure 6: Function prototypes for testing.

products manufactured using additive technologies is similar where the same level of surface finish is involved.

One objective during research is that of obtaining functional prototypes using polymeric materials capable of withstanding mechanical testing using rapid prototyping technologies. The advances made in the field of deposition materials and the subsequent finishing of the model may lead to functional prototypes that do not require the use of techniques based on rapid manufacturing (RM), thereby avoiding tooling costs (see Figure 6) [23, 46].

*3.2. Technologies and Decision Variables.* Some of the technologies described above require the use of a material, known by the name of support material, the purpose of which is to hold any overhanging designed parts in place. Once the deposition process has been completed, the support material must be removed during an operation carried out subsequently to manufacture (postprocessing), and the technique used to remove it shall depend on the support material in question and, therefore, on the additive manufacturing technology employed. In some additive technologies there is no support material as the support function that is performed by the material that has not hardened [18].

With respect to the mechanical properties of the prototypes obtained via the addition of material, these are determined by the quality of the result of the fusion between layers and the properties of the material. The following parameters (DIN EN ISO 178/179/180/527/2039) must be established in order to analyse the mechanical properties of the materials used in the different additive manufacturing methods: elastic modulus, breaking stress, elongation, flexural modulus, impact strength, compressive strength, and melting point (Table 11) [11, 54–56].

The choice of the most suitable technique for each type of prototype is based on the definition of the objective behind the production of the prototype: aesthetic, functional, investigational, or visual if the purpose is only to check the external appearance of the item designed [19, 39, 42, 57–60]. When it comes to making this choice an analysis may be planned based on a study of the possible variables: technology, resolution and precision, materials, software, the mechanical properties of the material (traction, compression, impact, softening, and density), surface finish, production time, cost, maximum dimensions of the item or model, posthardening requirements, guarantee, noise, CE certification, operational temperature, electrical connections and consumption, interface (network, hardware, software and exchange formats), weight, and spares and consumables (Table 12) [23].

Of all the aforementioned study variables, those habitually taken into account when choosing the prototyping technology are *resolution-precision, the mechanical and thermal properties of the material, surface finish, production time, and the cost of the prototype.*

With respect to the evaluation of the different prototyping technologies in accordance with the cost of the prototype, the fact that the comparison of technologies is restricted by the type of machine used must be taken into account. It is commonly thought that those prototyping machines based on stereolithography (SL) and selective laser sintering (SS) can be applied to the industrial production of prototypes, while all the others are seen as being machines that can be used professionally, but not in situations where the main objective is production. The manufacturers are currently offering domestic or desktop, professional, and industrial rapid prototyping machines, and therefore the costs incurred by using these machines must be offset by the performance levels shown above and by the production levels that can be obtained [39].

However, to calculate the price of an element manufactured using additive technologies the following general model can be followed in which the final manufacturing cost of the prototype ($C_p$) of the 3D model has been calculated in accordance with the following equation [8]:

$$C_p = C_e + C_m + C_t + C_a \quad (1)$$

where

$C_e$ is production cost (machine depreciation data)

$C_m$ is cost of material

$C_t$ is the processing cost of the 3D model and labour cost

$C_a$ is finishing (post-processing) cost

If this calculation model is transferred to a specific case, observe the following example of the cost of a prototype generated using the MD technique in a professional grade machine that can be easily adapted to any additive technology. In this case, as can be seen, there is no finishing cost (Table 13).

Table 11: The principal mechanical and thermal properties of the functional materials used in AM.

| PROPERTY | STANDARD | PROTOTYPING TECHNOLOGIES – HABITUAL MATERIALS | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | SL | SS | MD | | JP | | SL-JP | |
| | | Next | PA 12 | ABS | ABS+ | VisiJet M3 X | zp150-Z-bond | Digital ABS | PolyJet White |
| Tensile modulus (MPa) | ASTM D638M2 DIN EN ISO 527 | 2370-2490 | 1650 | 1627 | 1915 | 2168 | - | 2600-3000 | 2500 |
| Tensile strength (MPa) | ASTM D638M DIN EN ISO 527 | 31-35 | 48 | 22 | 37 | 49 | 14 | 55-60 | 58 |
| Elongation at rupture (%) | ASTM D638M DIN EN ISO 527 | 8-10 | 20 | 6 | 4,4 | 8.3 | 0.2 | 25-40 | 10-25 |
| Flexural modulus (MPa) | ASTM D790M DIN EN ISO 178 | 2415-2525 | 1500 | 1834 | 1917 | - | 7.2 | 1700-2200 | 2700 |
| Bending strength (MPa) | ASTM D790M DIN EN ISO 178 | 68-71 | - | 41 | 62 | 65 | 31 | 65-75 | 93 |
| Impact strength (J/m) | ASTM D256 DIN EN ISO 180 | 47-52 | 53 | 107 | 96.4 | - | - | 65-80 | - |
| Deformation under load temperature (°C) | ASTM D648 | 48-57 | 86 | 76-90 | 73-86 | 88 | 112 | 58-90 | 48 |

Table 12: Decision variables when choosing a prototyping technology.

| CHARACTERISTICS | SPECIFICATIONS TO BE CONSIDERED |
|---|---|
| COSTS | Price of the machine (including post-production and maintenance) |
| | Unitary model cost |
| | Cost of training qualified operators |
| | Control and modelling software |
| | Annual maintenance cost |
| DIMENSIONS | Workspace |
| | Dimensions of the machine |
| | Weight of the machine |
| | Noise level |
| | Mandatory accessories |
| WORK MATERIAL | Colour or number of colours, transparency |
| | Possibility of recycling material |
| | Technical characteristics of the material |
| | Working temperature |
| PRECISION | Precision |
| | Height/Thickness of layer |
| | Minimum detail size |
| | Resolution |
| | Minimum wall thickness |
| OTHERS | Vertical working speed |
| | Network/On-line connection |
| | Files supported, scope of the software associated with the machine |
| | Adaptability to accessories |
| | User friendliness (ease of handling and maintenance) |

*3.3. The Benefits and Disadvantages of Additive Manufacturing.* The processes used to manufacture conventional parts and components are influenced by a series of limitations related to the obtaining of certain shapes, such as curved holes, mould release angles, or preventing tools from coming into contact with geometrically complex pieces. And then there is the fact that some manufacturing processes do not comply with a company's commitment to a sustainable production process by involving the residues related with the use of cooling liquids.

Two characteristics comprise the main difference between the additive manufacturing techniques and their conventional counterparts. These not only provide significant competitive advantages, but also do not make the manufacturing process more expensive:

(1) *The geometrical complexity of the part to be manufactured.* Elegant geometrical forms, hollow interiors, internal channels, variable thicknesses, irregular shapes, etc. can easily be reproduced based on the geometrical template obtained from a 3D CAD.

(2) *The customisation of the part to be manufactured.* Products that are exactly identical or completely different can be obtained without any notable influence on the process and without additional costs. This customisation represents one of the main current trends in the development of products with a high added value, and the mass application thereof is one of the paradigms pursued by the industrial sectors in developed countries, which see it as being the key to their sustainability.

These two characteristics can provide massive benefits in different industrial sectors:

(i) *Lightweight Products.* They enable the manufacture of products designed for a specific function and with made-to-measure features, e.g.: lighter for reasons of weight savings, strength or costs. Some of the additive manufacturing techniques are capable of filling a model with different degrees of porosity without a change of material.

(ii) *Multimaterial Products.* They make it possible to manufacture a product using several materials simultaneously in the same solid. This means that the technique overcomes one of the current limitations with respect to the weight/mechanical strength ratio by the introduction of new functionalities or the lowering of production costs [61, 62].

(iii) *Ergonomic Products.* The design of the components can achieve a greater degree of interaction with the

Table 13: Calculation of the prototyping cost using material deposition (MD) technique.

| COSTS ANALYSIS FOR THE PROTOTYPING OF ITEMS IN A 3D PRINTER | | |
|---|---|---|
| MACHINE DEPRECIATION DATA Ce | | |
| Price of Machine (€) | | 25,000 |
| Yearly maintenance cost (€) | | 2,900 |
| Years of depreciation | | 4 |
| Depreciation (h/year) - 223 days-year / 8 hours-day | | 1,784 |
| Machine-depreciation price per hour (€/h) | | **4.72** |
| Retail sale price per hour (€/h) | | **4.72** |
| COST OF MATERIAL DATA Cm | | |
| Cost of model material: ABS filament (€/cc) (€271-950 cc) | | 0.23 |
| Cost of support material: acrylic filament (€/cc) (€271-950 cc) | | 0.23 |
| Cost of tray material (€/unit): (€100-24 units.) | | 4.17 |
| COST OF TECHNICAL ANALYSIS DATA Ct | | |
| Cost of technical model analysis – including release from mould (€/h) | | 20 |
| TECHNICAL DEPOSITION DATA | | |
| *Model type* | ***Mesh*** | ***Solid*** |
| Deposition rate (cc/h) | 11 | 16 |
| Machine-deposition price per hour (€/h) | | |
| ITEM TEST Model - Positioning: HORIZONTAL | | |
| CONCEPTS BUDGETED | Solid Interior | Partial costs |
| Model material (cc) | 17.32 | €3.98 /unit |
| Model support (cc) | 2.44 | €0.56 /unit |
| Model time (h) | 1.30 | €6.14 /unit |
| Items per tray (unit) | 1.00 | €4.17/unit |
| Technical-analysis time (h) | 0,30 | €6.00 |
| Number of items | 1.00 | |
| **Unitary cost (€) + VAT** | **€20.85 /unit** | |
| **Total cost (€) + VAT** | **€** | |

user by adapting to the exact anthropometric characteristics of each individual (prostheses) without necessarily affecting the manufacturing costs.

(iv) *Integrated Mechanisms*. They make it possible to manufacture a mechanism that is totally embedded in the finished item without the need for subsequent assembly and adjustments, e.g., a journal bearing, a roller bearing, a spring and its support, and a screwed-on worm gear.

As far as the production of industrial components is concerned, the following must be highlighted as obvious benefits:

(i) *A reduction of the time it takes new designs to reach the market*: when additive manufacturing is used as a manufacturing technique of the end product and not only in the production of prototypes, many of the current launch and validation phases can be drastically shortened. Another advantage is that it provides great flexibility when it comes to responding to the continuous changes in market demand.

(ii) *Short production runs*: the size of the production run can be minimal to the extent of being on a per unit basis while hardly influencing manufacturing costs (if and when the depreciation of the equipment is not considered). One of the characteristics that make this possible is the lack of a need for tooling, which represents a considerable advantage with respect to the conventional manufacturing methods.

(iii) *A reduction of assembly errors and their associated costs:* ready assembled components can be obtained with the only subsequent operation being the quality control inspection.

(iv) *A reduction of tool investment costs*: tools do not form part of the additive manufacturing process. This represents a great deal of flexibility as regards adapting to the market and a reduction, or even elimination, of the associated costs (toolmaking, stoppages due to referred changes, maintenance, and inspection).

(v) *Hybrid processes:* it is always possible to combine different manufacturing processes. In this case combining additive manufacturing processes with conventional processes might be interesting to make the most of the advantages offered by both. For example, it might be extremely beneficial to combine additive manufacturing technology with mechanised material removal in order to improve surface quality via a reduction of the "stepping effect" produced by the additive manufacturing technologies. Hybridisation can also occur in the opposite direction, in other words manufacturing using subtractive methods starting with a block before adding, by way of additive manufacturing, those especially complicated characteristics which generate high value.

(vi) *Optimum usage of materials*: material wastage is reduced to a minimum. Any waste material can be easily recycled.

(vii) *A more sustainable manufacturing process*: toxic chemical products are not directly used in appreciable quantities.

However, additive manufacturing technologies do have a number of drawbacks which must be borne in mind when choosing the technology best adapted to the requirements of the product to be manufactured.

(i) Additive layer manufacturing produces what is known as the stepping effect. The disadvantages of this phenomenon include complicating the shaping of geometrical curves and an extremely rough surface finish. This effect means that shafts and holes must typically be manufactured with their circular cross-section in plan. If they are not, the roundness of the piece would not be acceptable. On the other hand, and putting roundness to one side, positioning the piece in another way might be useful depending on the application in question; it would be interesting to manufacture an overturned sliding axis in such a way that no 'interlocking' occurs.

(ii) With respect to some technologies, the manufacturing operation itself can be slow, thereby making it particularly suitable for small production runs. When the production run reaches a certain size it may well be appropriate to use a conventional technology despite the fact that, as has been seen above, these technologies have a number of limitations, especially geometrically speaking.

(iii) The materials used in some of the technologies might not be suitable for the product to be manufactured.

(iv) The depositing of layers produces anisotropic materials. Given the fact that many industrial components are subjected to forces that put the material under stress and that they are so sized as to use the minimum amount of material, it is possible that the performance of the components with respect to the forces they must withstand while in service results inadequate.

(v) The tolerances obtained using the majority of the additive manufacturing methods are still higher than those achieved using other manufacturing methods such as those based on the removal of material.

*3.4. By Sector Innovation with Additive Manufacturing.* Both in innovation and in research, advances are going to be defined by acquiring new materials, more precise, and less costly equipment and also by seeking out new sectors for 3D printing.

An interesting proposal in this field is that presented by Wong and his team in 2012 [6]. Six sectors are analysed: lightweight machines, architectural modelling, medical applications, improving the manufacturing of fuel cells, and additive manufacturing for hobbyist and additive manufacturing in art. We have no doubt that these were the sectors of innovation five years ago. However, our analyses show us that nowadays the additive manufacturing sectors where innovation can really be seen are as follows: consumer products,

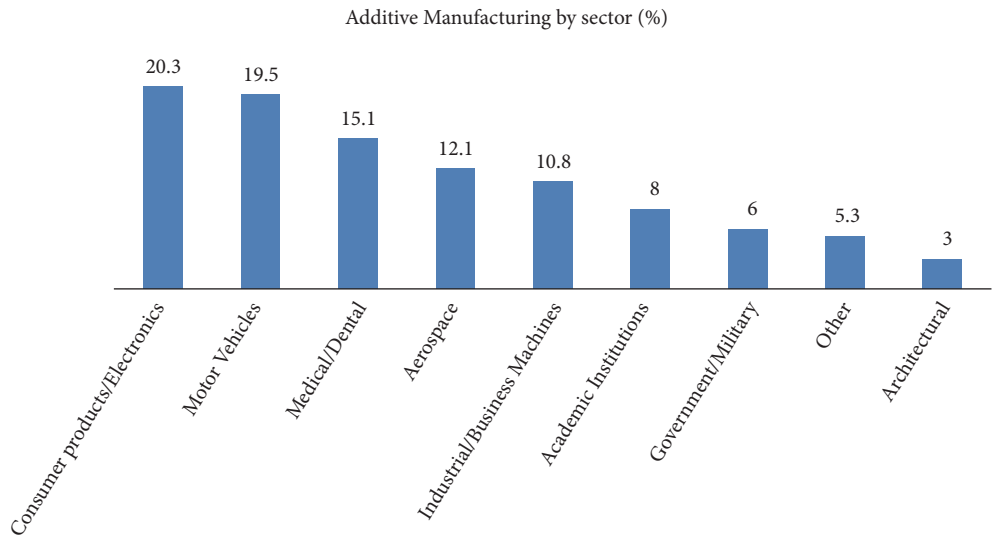



Figure 7: The use of additive manufacturing in the different sectors (Wohlers Report 2013).

automotive industry, medicine and medical engineering, aviation industry, architecture, construction, and food.

The degree to which additive manufacturing is used in different sectors is shown in Figure 7.

A review of the sectors in which additive manufacturing is currently used is presented as follows.

*3.4.1. Consumer-Electronic Products.* This sector uses additive manufacturing to obtain prototypes and models of a multitude of articles for the home, sports equipment, toys, etc. It is the number one customer of those additive manufacturing technologies that enable the direct digital manufacture of finished components of high geometrical complexity and that require customisation.

As soon as materials that are both flexible and strong even when thin become available, it shall be possible to manufacture consumer products such as clothes and footwear using additive techniques. The deposition of conductive materials via the printing of passive circuit components such as resisters, condensers and coils, diodes, organic light emitting diodes (OLEDs) and circuit interconnections can only benefit the production of electronic devices and components.

*3.4.2. Motor Vehicles.* In this sector additive manufacturing is being used to create prototypes that enable the validation of engineering processes and, above all, functional and aesthetic component design processes. The production of finished parts is not yet a reality, with the technique only being used in the customising of certain elements in one-off vehicles. The hope is that the development of new materials and their application of large, high-speed machines will favour the use of additive manufacturing in conjunction with the highly demanding production criteria inherent to this sector.

*3.4.3. Medical/Dental.* The application of additive manufacturing in the medical/dental sector enables physical 3D models to be obtained from processed medical images (3D scans, TAC) for application in different specialist areas.

The use of additive rapid prototyping technologies enables preoperative planning processes, the production of prostheses, and the preparation of surgical templates and guides to be carried out with a higher diagnostic quality and greater surgical safety in less time and more cheaply than is possible using conventional manufacturing techniques. In the case of specific and customised implants optimum planning of the surgical process and a reduction of operating times has already been achieved.

We must not omit today's 3D printing of living cells, bioprinting, in which a lot of resources are being invested and which we trust will soon present some very interesting results.

*3.4.4. Aerospace.* This market requires additive manufacturing to respond to high mechanical and thermal performance demands, weight reduction, and minimum losses of material as regards certain components with respect to both polymeric and metallic materials, primary titanium, and nickel alloys. The selective sintering of powdered metals has become a manufacturing, repair, and maintenance solution for certain components, e.g., turbine blades, as well as for the manufacture of high added value aeronautical tooling.

*3.4.5. Architecture.* The manufacture of mock-ups and prototypes within the architecture and construction sector was, and still is carried out on a significant handicraft basis. The development of assisted design systems, with its resulting progress towards solid modelling systems and the current BIM systems with respect to building, has enabled the production of highly attractive quality digital mock-ups, infographics, and virtual animation of plans and projects. However, the same cannot yet be said about the physical mock-ups obtained from that digital model of the plan using additive mock-up and prototype construction machines. 3D printing could well become an essential piece of equipment in the studios of architects and designers. In Figures 8 and 9, we can see a number of examples of how these techniques are

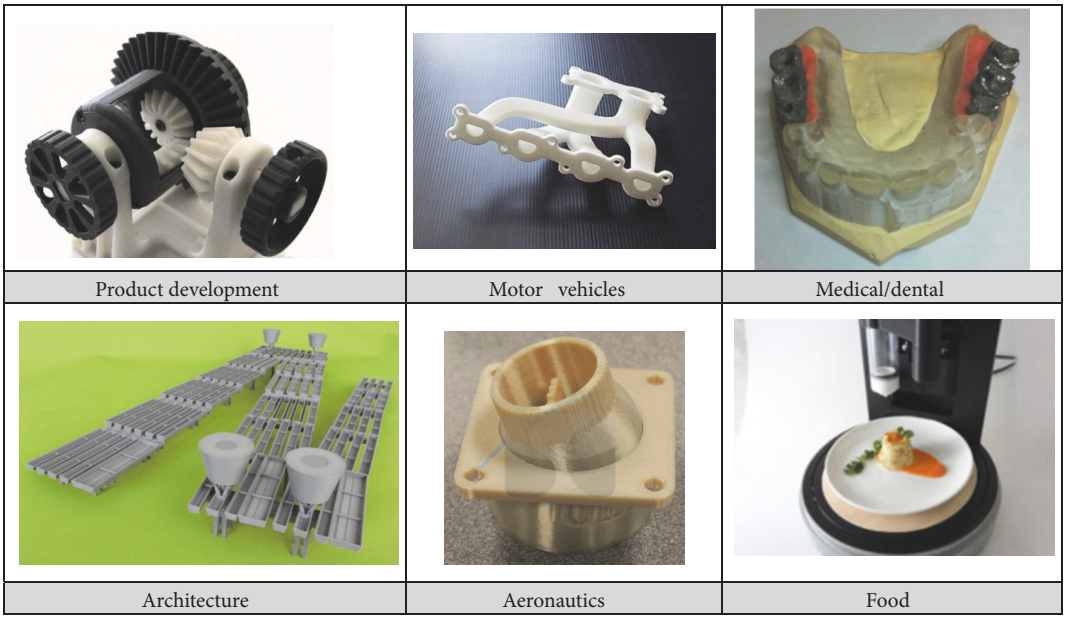



Figure 8: AM applications.

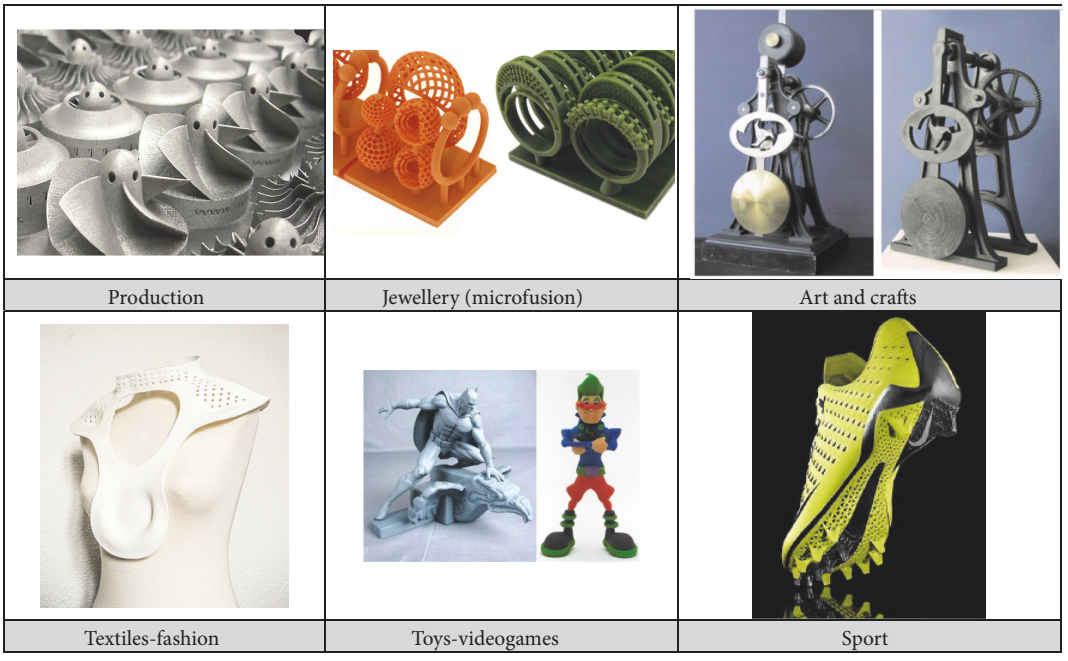



Figure 9: Another AM applications.

applied and they clearly show the great potential of additive manufacturing.

*3.4.6. Food.* Even when the catering industry is incorporating new 3D printing techniques for food, perhaps it is a good idea to discuss the advances being made in this field with regard to the food-health combination, that is, when 3D printing systems are used to measure out food and structure patients' diets.

However, naturally, in the field of catering and food in general, major advances are being achieved, which are already in operation in prestigious restaurants or in the production of desserts and sweets.

*3.5. By Sector Investigation with Additive Manufacturing.* In the field of research, that is, in the field of the work that is being carried out today in research centres and universities, both public and private, there are three approaches

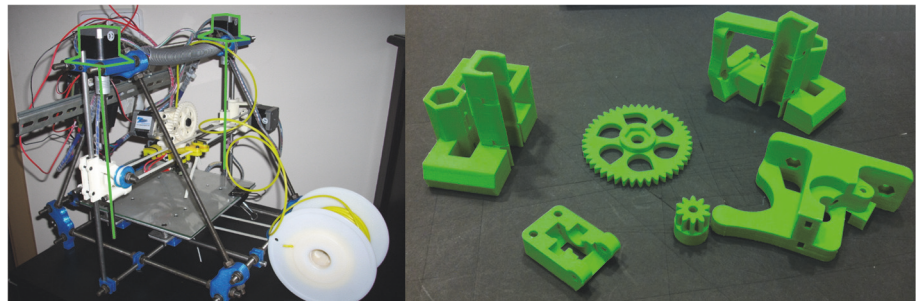


Figure 10: Example of RepRap machines and parts for replication.

that stand out: materials, equipment, and new fields of applicability.

In the field of new materials, a lot of progress is being made on biocompatible materials, compound materials, and metals, each one within a clearly identified field of development. In these fields, materials with good mechanical properties are also being sought; however, at the same time, progress is being made on the development of elastic materials, which open up a whole range of possibilities for 3D printing that is yet to be quantified.

In terms of new equipment, greater precision and lower cost are clearly what is being sought after, in order to make this technology competitive with regard to other conventional technologies.

With regard to the new fields of applicability in which, without a doubt, great progress will be made over the coming years in terms of printing metal materials, we resolve the problem of the dispersion of metallic liquid when it reaches molten state, and in terms of bioprinting, we develop new applications with regard to living tissues, not only in bones and cartilage, where advances have already been made, but also in other tissues of the human body, including the viscera and muscles.

*3.6. The RepRap Community and Free Software.* In 2004 Adrian Bowyer founded RepRap [5], an open-code initiative for building a 3D printer capable of printing out most of its own components, at Bath University. The vision of this project is to make the manufacture of low-cost distribution units available to people all over the world, thereby enabling them to create their own products on a daily basis (see www.RepRap.org).

As the term 'Fused Deposition Modelling' had already been registered by Stratasys, the RepRap Community has coined the term 'Fused Filament Fabrication (FFF)', which can be used by anybody without any restriction whatever (under a version 2 GPL licence). Under these terms and conditions anybody can distribute and modify the RepRap machine, but they must respect the modifications made under this licence [21]. In other words, all changes must remain in the public domain. As the machine is both free and open-code; anybody can, without having to pay any fees whatever, build an unlimited number of copies for themselves or for anyone else using the selfsame RepRap machines to manufacture the plastic parts of the copies (thereby making it self-replicating, Figure 10).

Although we know that the RepRap movement has played and continues to play a very important role in the development of additive manufacturing, it is interesting to discuss here that it was not initially accepted by the academic community, as it was considered a subject of minor importance. Only a few years later the subject was accepted, however not as a phenomenon in itself, rather under the consideration of a machine that could replicate itself, something that every manufacturing professional knows has been possible since the 19th century, with milling machines.

There is an increasing number of meetings of "experts" who analyse what the 'printing' of physical objects using 3D printers will represent, which is now being seen as one of the great industrial revolutions of the next few years, and there has even been talk of the Third Industrial Revolution. The #Redada sessions are a meeting that allows users and professionals to exchange ideas and analyse the possibilities open to them, as in the case of Video 34, "#Redada 18 Madrid: The Challenges of 3D Printing", which features a debate about the social trends and the aspects related to culture, civil rights, and technology.

*3.7. User and Exchange Communities.* The social importance of this technology has been enormous and, as has already been said, it is developing in leaps and bounds, thereby enabling engineering students and professionals who specialise in these particular techniques from all over the world to experiment with their creations and perfect them prior to the production of full scale versions.

Accessibility to the technology links a series of communities of users and developers who exchange their know-how and experiences in order to continue perfecting the printing system and open up new and previously unimaginable fields in the process.

Alongside this, these communities of users have developed a series of platforms for exchanging existing 3D models for downloading and printing, thus broaching a far from ludicrous idea that involves manufacturers making 3D models of ex-catalogue parts of their products available to users. To a certain extent Google Earth has already done this by allowing the community of Google Sketch Up users to upload models of buildings from all over the world in their exact

location for everybody to enjoy. The fact that the future holds countless possibilities cannot be too highly stressed. There are model upload and download banks such as the English language platform Thingiverse and its Spanish counterpart Rascomras. Many more exist, and their number increases by the day, some more creative than others, such as The Pirate Bay (a community for the downloading of all types of audiovisual material which has incorporated a new section for 3D models).

The Clone Wars Projects seek to spread the word about RepRap technology while at the same time contributing new designs and innovative research channels, but not so much along self-replication lines.

A series of workshops have existed all over the world for a number of years now. Known as FabLabs (Fabrication Laboratories), they are being promoted by the Center of Bits and Atoms (CBA) of the Massachusetts Institute of Technology (MIT), at which a lot of hard work is being done on this technological revolution in light of the social changes it is bringing about. They are equipped with a series of computer-controlled machines "for building (almost) anything": 3D printers, laser cutters, CNC (computerized numerical control) routers and an electronics laboratory (among many others that vary from workshop to workshop). Worldwide, with respect to the number of Fablabs Spain ranks fourth with 7-8 behind the USA, with more than thirty, The Netherlands (9), France (8), and ahead of Germany (6).

## 4. Discussion

As has been discussed throughout this paper, additive manufacturing (AM) processes are considered in many applications as a new industrial revolution. This article conducts an exhaustive study of the current state of additive manufacturing. As is shown in Table 2, the technologies and processes that currently exist are very diverse and, therefore, producing a classification that unites and differentiates all of them being truly complex. Thus, this paper proposes several types of classifications.

Over recent years, many names have arisen to encompass these technologies, such as "rapid prototyping", "rapid tooling", "3D printing", and "freeform fabrication". All of these are commonly accepted; however, "additive manufacturing" is probably that which best brings them all together.

The main advantages associated with these technologies are the high precision, the possibility of using different materials, and the ability to obtain impossible prototypes using conventional means. The current limitations include the high cost of the processes, the time required to obtain the prototypes, and perhaps the lower resistance they have. Active work is underway to improve these limitations in order for additive manufacturing to be competitive with regard to other more conventional means.

It is important to note that within additive manufacturing there is no perfect technology for all purposes, rather what we need to do is to determine the most suitable technologies for a specific use.

For example, in the dental sector, as is discussed by Jiménez et al. (2015) [39], in order to manufacture the models used in the thermoforming of correction splints, technologies based on printing by injecting resin (IJP-UV), digital light processing (DLP), and fused deposition modeling (FDM) are the most suitable as they offer the best price-quality ratio of the model for thermoforming.

Among the agents in the aviation market, metal material additive manufacturing technologies, such as 'Electron Beam Melting' (EBM), 'Sintering Laser Melting' (SLM), or 'Laser Cladding' (LC), are those that attract most interest, in particular, for part manufacturing, case of the OEM or Tier1; or for part repair, case of the 'Maintenance and Repairing Overhaul' (MRO). These manufacturing technologies provide many advantages in comparison with other conventional metal transformation processes (http://www.ctsolutions.es).

Changing sector, 3D printing of architectural models will lead to a reduction in the number of steps, an improved design timeframe, and the preservation of the finer details of the final architectural design, and therefore its market niche is on the rise. As discussed by Domínguez et al. (2013) [23], fused deposition modeling machines appear among the most suitable for obtaining working models, given their low cost (especially in the case of the RepRap models), the speed of the process, and the possibility of recycling the material. Machines projecting binding agent would also be suitable for obtaining models for the client, thanks to their competitive prices, good surface finish, wide range of colours, and lack of support fixtures, among other qualities. Although, perhaps, the method most used for architecture is printing via the sintering of composite powder, this material requires a postprocess to harden it and give it the necessary consistency and finish for an optimum result (http://sicnova3d.com). In any case, it is certainly true that, right now, no technology fully meets all of the requirements of the work specifications in the field of architecture and construction. Thus, this sector still has quite a long way to go.

PolyJet technology produces ultradetailed prototypes, moulds, and even final parts that incorporate smooth rigid, transparent, and flexible materials, which is why it has been the technology most used in the jewellery sector in recent years. Multimaterial 3D printers produce lifelike models with a variety of properties on a single build tray (www.stratasys.com). Regarding jewellery, one of the advantages of using 3D printers is speed. The plastic parts take 7 to 10 days to be made, whereas metal parts take 10 to 15 days. Other positive points include the cost saving and the fact that it is possible to retouch the jewellery while it is being printed.

*4.1. Immediate Future of Additive Manufacturing.* In production lines, one of the main focal points for improvement in additive manufacturing consists of optimizing its features in order to be competitive with regard to conventional manufacturing processes in different production lines. In comparison with the traditional means, the use of additive manufacturing technologies continues to be too costly.

An important niche for saving in the industrial sector would be the so-called virtual libraries. There is a large number of fixed assets in all industrial platforms within what is known as physical replacement parts, spare parts, etc. Many of these items could be saved by means

of a virtual parts library (https://www.thingiverse.com/, http://www.3dprintfilemarket.com/), which could print suitable parts or components as and when required.

Another important section is that of the study of new materials. Cellulose, the plant material we have used for centuries to make paper, has emerged as a new resource for better, faster, and cheaper three-dimensional printing, in addition to providing an alternative that is recyclable and biodegradable by nature, according to new research by the MIT, published in Advanced Materials Technologies. At present, the key raw material for 3D printing is polymers, compounds that are largely synthetic and which use inks to create three-dimensional objects in accordance with the models via a computer used to execute the three-dimensional printing (www.imprimalia3D.com).

A particularly interesting field and one for study is that of the space sector, where additive manufacturing should also play an important role. The National Aeronautics and Space Administration of the US (NASA) is seeking a habitat design built using a 3D printer that can be used as a base to build houses on the surface of Mars. The final objective is to achieve a space design that allows astronauts to stay on the red planet for long periods at a time. Different projects are being carried out to conduct research on materials and explore the possibilities of 3D technology, which would mean many of the necessary infrastructures could be directly built on the Moon using, moreover, resources that are already there. This would speed up this large undertaking, as it would notably reduce the amount of parts that would need to be taken to the Moon and then later to Mars. 3D technology and the use of resources may help reduce costs both in the long and in the short term.

*4.2. Connected Industry: Future Prospects of Additive Manufacturing in the 4.0 Environment—A Study Is Conducted on the Possibility of Positioning Additive Manufacturing in a Service Environment.* The term Industry 4.0 was coined to describe the smart factory, a vision of computer-aided manufacture with all of the processes interconnected through the Internet of things (IOT). It is what we know as the industrial Internet of things, I2OT.

It is hoped that the new concept of industry 4.0 will be able to drive forward fundamental changes on the same level as the steam-powered first industrial revolution, the mass production of the second, and the electronics and proliferation of information technology that characterised the third.

According to Mark Watson, associate director for the industrial automation of IHS, "The challenge for the fourth industrial revolution is the development of software and analytical systems that turn the deluge of data produced by intelligent factories into useful and valuable information."

Factories with fully computerized production processes are better prepared to respond faster to changes in the market, as they have integrated greater flexibility and individualization in their manufacturing processes.

However, in order to obtain real improvements in manufacturing efficiency and flexibility, manufacturers must be able to manage and analyse large amounts of data, the biggest challenge of which will be regarding software. Companies should implement Big Data systems capable of managing large amounts of data from the manufacturing environment and conducting an intelligent real-time analysis, providing valuable information for decision-making, thus optimizing the processes and increasing business intelligence.

Industrial companies have to take the technological leap to Industry 4.0, in the field of additive manufacturing. The concepts and experiences being accumulated in companies and research institutes need to be passed on to companies by means of guided visits to leading facilities, induction and practical training, guided training, diagnosis, specific advice, testing, and prototypes.

Why should additive manufacturing be introduced in companies in a connected industry setting? The answer is simple: because the following can be achieved:

(i) Creative product design

(ii) More customised products with high quality and performance

(iii) DDM (demand driven manufacturing) with less waste generated and efficient use of energy

(iv) Internet (EICTs in general) as a tool with a high potential to support new supply chain models

(v) The consumer as designer and "customiser"

(vi) Additive manufacturing as enabling technology

*4.3. Additive Manufacturing and B2C/B2B.* The majority of the work on systematizing and disseminating sales experience and on the techniques, methods, texts, and sales and marketing courses focuses on selling to the consumer, what we call B2C (Business-to-Consumer). However, the volume of business generated by sales to other companies, B2B (Business-to-Business), is much higher, not to mention complex and different.

Additive manufacturing can take the shape of a service activity and, therefore, it needs to adapt to ensure that the companies that are willing to provide this service obtain higher growth and profitability on sales to other companies and organizations and also on sales to the consumer, taking into account the final destination of the additive manufacturing service.

The current problem is that there is no vertical platform that organises the advanced manufacturing technologies and customised manufacturing based on additive manufacturing (3DP.)

Current service providers partially facilitate the transaction for 3DP solution companies. It must be ensured that additive manufacturing is on the cloud as a reliable service based on the fact that clients manage the details of their manufacturing project in real time: material, size, delivery time, quality, price, location, and catalogue payment (direct e-commerce selection), by means of a quotation or offer.

A global platform must be configured as a network of 3DP services: marketplace for B2B vertical offering (Figure 1). The service portfolio should be built around the following activities:

Business model: marketplace, all agents, all orders (large series focus), matrix selection, web location, and real-time all-agents capacity management.

(i) Market: B2B focus (could do B2C), KPI-customised orders.

(ii) Value Chain: global E2E services to B2B vertical clients.

(iii) Technology: E2E open API connecting in real-time clients and provider's ERP.

(iv) Product: Manufacturing and high quality / high definition.

(v) Expertise: Institute 3DP: industry, research centres, universities ecosystem.

However, the development of additive manufacturing does not stop there; the following step was what is now called "bioprinting", which is the printing of cells and living tissues [38, 50]. The interest in this is mainly due to the shortage of organs available for transplant and the possibility of avoiding rejection if the organ required can be successfully printed using the individual's own cells. Nevertheless, its use is very important in research on new drugs in order to reduce the use of laboratory animals.

3D printing of living cells usually requires the deposition of cells and also the deposit of the support element or matrix. This matrix, the place where the cells are going to grow, where nourishment will be found and on which the structure is to be formed, may be liquid, usually called bioink, paste-like, rigid solid, or elastic solid. Also, solid materials in particular may or may not be biodegradable [51].

The issue of biodegradation is an added factor in this technology, as if this biodegradation occurs too quickly; we will not obtain the desired results; however, if it occurs too slowly, there is a possibility that the structure will prevent the development of the cells and so, in the end, we will not obtain the desired results this way either. This is why bioprinting is currently the focal point of the majority of research projects: it is anticipated that bone regeneration, printing different vital organs, printing human skin, etc., will be functionally viable in the near future, as prostheses currently are.

## 5. Conclusions

(i) Conventional manufacturing is mainly limited by production run size and the geometrical complexity of the component, and we are occasionally forced to use processes and tools that raise the final cost of the element. What is more, some manufacturing processes do not comply with a commitment to sustainable manufacturing (contamination, recycling, etc.).

(ii) Additive manufacturing is one of the key tools for tackling the growth and the creation of added value and high quality employment.

(iii) Conceptually, the term 'additive manufacturing' describes the technology in general and it is used when referring to industrial component manufacturing applications and high-performance professional and industrial equipment. Other terms exist, with the best known being "*rapid prototyping and 3D printing*", in accordance with the scope of the model and the type of additive machine used.

(iv) Additive manufacturing techniques provide huge competitive advantages because they adapt so well to the *geometrical complexity and the customisation of the design of the part to be manufactured.* The following can also be achieved in accordance with the sectors of application: *lighter weight products, multi-material products, ergonomic products, short production runs, fewer assembly errors resulting in lower associated costs, lower tooling investment costs, a combination of different manufacturing processes, optimum use of material, more sustainable manufacture.*

(v) Even though this type of process began as a new independent technology, nowadays additive manufacturing is a manufacturing system more, comparable to others such as subtractive manufacturing or foundry. Therefore, the protocols for the classification of additive manufacturing processes do not have to be different from those applicable to other manufacturing systems.

(vi) The drawbacks are: *the finish of complex surfaces can be extremely rough, long production times, materials with limited mechanical and thermal properties which restrict performance under stress, higher tolerances than with other manufacturing methods such as those based on material removal.*

(vii) The study variables habitually taken into account when choosing the prototyping technology are: *resolution-precision, the mechanical and thermal properties of the material, surface finish, production time and the cost of the prototype.*

(viii) The manufacturing precision is the result of superimposing different errors in the production of the model which affect the surface quality, the dimensional accuracy and the final weight of the model. The errors that occur in this process are: an error in the conversion of the 3D model into STL format (triangulation of the geometry), an error in the decomposition in layers of the 3D model (exact division of the thickness), stepping effect error (orthogonal deposition of the material by layers) and, finally, model infill error.

(ix) Three-dimensional models with an aesthetic (visual) or assembly objective and functional models capable of withstanding mechanical testing can be achieved.

(x) Additive manufacturing can be applied across many sectors and it can be easily adapted to the demands of each of them.

(xi) Design and printing using 3D printers is seen as being one of the major industrial revolutions of the nest few years. Proposals exist for making the manufacture

of low-cost RepRap distribution units available to people all over the world via communities of users and developers who exchange 3D models, know-how and experiences for optimising the manufacturing performance of a self-replicating 3D printer.

(xii) It still seems puzzling that the first scientific publications relating to the important movement that is additive manufacturing came to light several years after the development of the first inventions, the first patents and even the first commercial communications on the advance.

(xiii) It is also strange that the RepRap movement did not gain ground in academic environments at the beginning, and only found its niche when the movement was justified as "machines that can replicate themselves" when, as is known in technical environments, there were already milling machines available more than one hundred years earlier that were able to self-replicate.

(xiv) The choice of technology is directly dependent on the particular application being planned: first the application, then the technology. Laser systems are being increasingly used, especially in the field of finished part production. In the future, the use of print technology systems is going to increase by the day.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

The authors would like to express their gratitude for their support to the Escuela Técnica Superior de Ingenieros Industriales of UNED, in the frame of the proyects ICF06-2018 and ICF08-2018, and to the Instituto Universitario de Educación a Distancia of the UNED, via the Proyecto de Innovación Educativa GID2016-35. In addition, it is necessary to mention the additive manufacturing laboratory of ETSI-ICAI, where some of the tests have been carried out.

## References

[1] 2017, https://es.3dsystems.com/our-story.

[2] 2018, http://www.wipo.int/portal/en/index.html.

[3] J. Delgado, J. Ciurana, and C. A. Rodríguez, "Influence of process parameters on part quality and mechanical properties for DMLS and SLM with iron-based materials," *The International Journal of Advanced Manufacturing Technology*, vol. 60, no. 5-8, pp. 601–610, 2012.

[4] M. M. Espinosa, *Introducción a Los Procesos de Fabricación*, UNED, Madrid, Spain, 2000.

[5] E. Sells, S. Bailard, Z. Smith, A. Bowyer, and V. Olliver, "RepRap: The replicating rapid prototyper: Maximizing customizability by breeding the means of production," in *Proceedings of the 2007 World Conference on Mass Customization & Personalization (MCP)*.

[6] K. V. Wong and A. Hernandez, "A Review of additive manufacturing," *ISRN Mechanical Engineering*, vol. 2012, Article ID 208760, 10 pages, 2012.

[7] J. Nylund, *Utskrift av Tredimensionell Arkitekturmodell [Bachelor's Thesis]*, Construction Engineering, Vaasa, 2015.

[8] M. Jiménez, J. Porras, I. A. Domínguez, L. Romero, and M. M. Espinosa, "La fabricación aditiva. La evidencia de una necesidad," *Interempresas Industria Metalmecanica*, vol. 235, no. 1047, pp. 74–82, 2013.

[9] 2018, https://www.iso.org/committee/629086.html.

[10] P. F. Jacobs, "Rapid Prototyping Manufacturing: Fundamentals of Stereolithography," *Society of Manufacturing Engineers*, 1992.

[11] 2018, https://www.materialise.com/es/manufacturing/materiales.

[12] U.S. Department of Energy, *Materials Development and Evaluation of Selective Laser Sintering Manufacturing Applications*, 1997.

[13] 2018, https://www.aenor.com/.

[14] E. L. Cañedo-Argüelles and M. Domínguez, "Estado actual del prototipado rápido y futuro de éste," *Actas del XI Congreso Internacional de Ingeniería Gráfica*, vol. 3, pp. 1242–1255, 1999.

[15] Y. Yan, S. Li, R. Zhang et al., "Rapid prototyping and manufacturing technology: principle, representative technics, applications, and development trends," *Tsinghua Science and Technology*, vol. 14, no. 1, pp. 1–12, 2009.

[16] J. Russell and R. Cohn, *Fused Deposition Modeling*, Book on Demand, 2012.

[17] R. Noorani, *3D Printing: Technology, Applications, and Selection*, CRC Press, 2017.

[18] 2018, http://www.stratasys.com/mx/impresoras-3d/technologies/polyjet-technology.

[19] J. A. Oriozabala Brit, M. D. Espinosa Escudero, and M. Dominguez Somonte, "Additive manufacturing opportunities to optimize product design: oportunidades de la fabricación aditiva para optimizar el diseño de productos," *Dyna Ingenieria e Industria*, vol. 91, no. 3, pp. 263–271, 2016.

[20] N. De la Torre, M. M. Espinosa, and M. Domínguez, "Rapid prototyping in humanitarian aid to manufacture last mile vehicles spare parts: an implementation plan," *Human Factors and Ergonomics in Manufacturing & Service Industries*, vol. 26, no. 5, pp. 533–540, 2016.

[21] A. Guerrero de Mier and M. D. Espinosa Escudero, "Progress in RepRap: open source 3D printing-avances en RepRap: impresión 3d de código abierto," *Dyna Ingenieria e Industria*, vol. 89, no. 1, pp. 34–38, 2014.

[22] W. E. Frazier, "Metal additive manufacturing: a review," *Journal of Materials Engineering and Performance*, vol. 23, no. 6, pp. 1917–1928, 2014.

[23] I. Domínguez, L. Romero, M. Espinosa, and M. Domínguez, "Impresión 3D de maquetas y prototipos en arquitectura y construcción," *Revista de la construcción*, vol. 12, no. 2, pp. 39–53, 2013.

[24] G. N. Levy, R. Schindel, and J. P. Kruth, "Rapid manufacturing and rapid tooling with layer manufacturing (LM) technologies,

state of the art and future perspectives," *CIRP Annals - Manufacturing Technology*, vol. 52, no. 2, pp. 589–609, 2003.

[25] D. T. Pham and R. S. Gault, "A comparison of rapid prototyping technologies," *The International Journal of Machine Tools and Manufacture*, vol. 38, no. 10-11, pp. 1257–1287, 1998.

[26] J. P. Kruth, "Material increase manufacturing by rapid prototyping techniques," *CIRP Journal of Manufacturing Science and Technology*, vol. 40, no. 2, pp. 577–639, 1991.

[27] M. Greul, F. Petzoldt, M. Greulich, and J. Wunder, "Rapid prototyping moves on metal powders," *Metal Powder Report*, vol. 52, no. 10, pp. 24–27, 1997.

[28] B. Derby and N. Reis, "Inkjet printing of highly loaded particulate suspensions," *MRS Bulletin*, vol. 28, no. 11, pp. 815–818, 2003.

[29] I. Gibson, D. W. Rosen, and B. Stucker, *Additive Manufacturing Technologies: Rapid Prototyping to Direct Digital Manufacturing*, Springer, New York, NY, USA, 2009.

[30] C. B. Williams, F. Mistree, and D. W. Rosen, "A functional classification framework for the conceptual design of additive manufacturing technologies," *Journal of Mechanical Design*, vol. 133, no. 12, p. 121002, 2011.

[31] 2018, https://www.astm.org/.

[32] 2018, https://www.iso.org/home.html.

[33] A. S. Wu, D. W. Brown, M. Kumar, G. F. Gallegos, and W. E. King, "An Experimental Investigation into Additive Manufacturing-Induced Residual Stresses in 316L Stainless Steel," *Metallurgical and Materials Transactions A: Physical Metallurgy and Materials Science*, vol. 45, no. 13, pp. 6260–6270, 2014.

[34] M. Xia, D. Gu, G. Yu, D. Dai, H. Chen, and Q. Shi, "Selective laser melting 3D printing of Ni-based superalloy: understanding thermodynamic mechanisms," *Chinese Science Bulletin*, vol. 61, no. 13, pp. 1013–1022, 2016.

[35] S. E. Zeltmann, N. Gupta, N. G. Tsoutsos, M. Maniatakos, J. Rajendran, and R. Karri, "Manufacturing and security challenges in 3D printing," *JOM: The Journal of The Minerals, Metals & Materials Society (TMS)*, vol. 68, no. 7, pp. 1872–1881, 2016.

[36] B. C. Gross, J. L. Erkal, S. Y. Lockwood, C. Chen, and D. M. Spence, "Evaluation of 3D printing and its potential impact on biotechnology and the chemical sciences," *Analytical Chemistry*, vol. 86, no. 7, pp. 3240–3253, 2014.

[37] S. H. Huang, P. Liu, A. Mokasdar, and L. Hou, "Additive manufacturing and its societal impact: a literature review," *The International Journal of Advanced Manufacturing Technology*, vol. 67, no. 5-8, pp. 1191–1203, 2013.

[38] W. Zhu, X. Ma, M. Gou, D. Mei, K. Zhang, and S. Chen, "3D printing of functional biomaterials for tissue engineering," *Current Opinion in Biotechnology*, vol. 40, pp. 103–112, 2016.

[39] M. Jiménez, L. Romero, M. Domínguez, and M. M. Espinosa, "Rapid prototyping model for the manufacturing by thermoforming of occlusal splints," *Rapid Prototyping Journal*, vol. 21, no. 1, pp. 56–69, 2015.

[40] L. Romero, M. Jiménez, M. D. M. Espinosa, and M. Domínguez, "New design for rapid prototyping of digital master casts for multiple dental implant restorations," *PLoS ONE*, vol. 10, no. 12, pp. 145253–145313, 2015.

[41] M. M. Espinosa and M. Domínguez, "La ingeniería concurrente, una filosofía actual con plenas perspectivas de futuro," *MetalUnivers*, vol. 16, pp. 16–20, 2003.

[42] L. Rodriguez Parada, M. Dominguez Somonte, and L. Romero Cuadrado, "Analysis and validation model of design proposals through flexible prototypes: Modelo de análisis y validación de propuestas de diseño mediante prototipos flexibles," *Dyna Ingenieria e Industria*, vol. 91, no. 5, pp. 502–506, 2016.

[43] I. Farina, F. Fabbrocino, G. Carpentieri et al., "On the reinforcement of cement mortars through 3D printed polymeric and metallic fibers," *Composites Part B: Engineering*, vol. 90, pp. 76–85, 2016.

[44] C. Gosselin, R. Duballet, P. Roux, N. Gaudillière, J. Dirrenberger, and P. Morel, "Large-scale 3D printing of ultra-high performance concrete - a new processing route for architects and builders," *Materials & Design*, vol. 100, pp. 102–109, 2016.

[45] P. Wu, J. Wang, and X. Wang, "A critical review of the use of 3-D printing in the construction industry," *Automation in Construction*, vol. 68, pp. 21–31, 2016.

[46] S. Fernández, M. Jiménez, J. Porras, L. Romero, M. M. Espinosa, and M. Domínguez, "Additive manufacturing and performance of functional hydraulic pump impellers in fused deposition modeling technology," *Journal of Mechanical Design*, vol. 138, no. 2, pp. 24501–24504, 2016.

[47] J. Sun, W. Zhou, D. Huang, J. Y. H. Fuh, and G. S. Hong, "An overview of 3D printing technologies for food fabrication," *Food and Bioprocess Technology*, vol. 8, no. 8, pp. 1605–1615, 2015.

[48] A. Guerrero-De-Mier, M. M. Espinosa, and M. Domínguez, "Bricking: A new slicing method to reduce warping," *Procedia Engineering*, vol. 132, pp. 126–131, 2015.

[49] J. Zhang and Y. G. Jung, *Additive Manufacturing: Materials, Processes, Quantifications and Applications*, Butterworth-Heinemann, 2018.

[50] S. V. Murphy and A. Atala, " 3D bioprinting of tissues and organs," *Nature Biotechnology*, vol. 32, no. 8, pp. 773–785, 2014.

[51] R. Domínguez, M. M. Espinosa, L. Romero, and M. Domínguez, *Impresión 3D en Ingeniería Médica*, VI Encuentro de Investigación - IMIENS, Madrid, Spain, 2016.

[52] E. Sells, S. Bailard, Z. Smith, A. Bowyer, and V. Olliver, "RepRap: The replicating rapid prototyper: maximizing customizability by breeding the means of production," *SSRN eLibrary*, 2010.

[53] M. Paulic, T. Irgolic, J. Balic et al., "Reverse engineering of parts with optical scanning and additive manufacturing," *Procedia Engineering*, vol. 69, pp. 795–803, 2014.

[54] 2018, https://www.sculpteo.com/es/servicios/fabricacion-aditiva/.

[55] Fundación COTEC para la Innovación Tecnológica, "Fabricación aditiva," Documentos COTEC sobre Oportunidades Tecnológicas, 2011, http://informecotec.es/media/N30_Fabric_Aditiva.pdf.

[56] M. Fernandez-Vicente, W. Calle, S. Ferrandiz, and A. Conejero, "Effect of infill parameters on tensile mechanical behavior in desktop 3D printing," *3D Printing and Additive Manufacturing*, vol. 3, no. 3, pp. 183–192, 2016.

[57] B. P. Conner, G. P. Manogharan, A. N. Martof et al., "Making sense of 3-D printing: Creating a map of additive manufacturing products and services," *Additive Manufacturing*, vol. 1-4, pp. 64–76, 2014.

[58] R. V. Rao, *Decision Making in the Manufacturing Environment: Using Graph Theory and Fuzzy Multiple Attribute Decision Making Methods*, Springer Science & Business Media, 2007.

[59] S. P. Sethi and Q. Zhang, *Hierarchical Decision Making in Stochastic Manufacturing Systems*, Springer Science & Business Media, 2012.

[60] S. H. Khajavi, J. Partanen, and J. Holmström, "Additive manufacturing in the spare parts supply chain," *Computers in Industry*, vol. 65, no. 1, pp. 50–63, 2014.

[61] S. Meyers, L. De Leersnijder, J. Vleugels, and J.-P. Kruth, "Direct laser sintering of reaction bonded silicon carbide with low residual silicon content," *Journal of the European Ceramic Society*, vol. 38, no. 11, pp. 3709–3717, 2018.

[62] X. Wang, M. Speirs, S. Kustov et al., "Selective laser melting produced layer-structured NiTi shape memory alloys with high damping properties and Elinvar effect," *Scripta Materialia*, vol. 146, pp. 246–250, 2018.

Hindawi
Complexity
Volume 2019, Article ID 6048195, 23 pages
https://doi.org/10.1155/2019/6048195

*Research Article*

# STEAM-ME: A Novel Model for Successful Kaizen Implementation and Sustainable Performance of SMEs in Vietnam

**Thanh-Lam Nguyen**

*Office of International Affairs, Lac Hong University, Dong Nai, Vietnam*

Correspondence should be addressed to Thanh-Lam Nguyen; green4rest.vn@gmail.com

Received 12 November 2018; Revised 14 January 2019; Accepted 23 January 2019; Published 7 February 2019

Guest Editor: Jorge Luis García-Alcaraz



The current trend of international integration urges every business organizations to continuously improve their competitive advantage for their survival and sustainable growth. And Kaizen has been a preferable approach in practice. Due to the special role of SMEs in the Vietnam economy, improving their competitiveness is critical. Thus, this study is aimed at identifying determinants of the successful Kaizen implementation and sustainable performance so that SMEs can have proper actions and prioritize their operations within their available resources. Through a formal survey of 213 participants from 62 SMEs which have been successful in implementing Kaizen and appropriate statistical analyses, seven important determinants have been identified, namely, (1) supports from senior management; (2) training; (3) environment; (4) assessment; (5) motivation; (6) mindset; and (7) engagement of all members in the organization. Among them, "mindset" is newly proposed in this study through a qualitative research and found as crucial component in the model. The finding obviously fulfills the existing literature. Moreover, the first letters of the identified factors are orderly congregated as "STEAM-ME" which is a novel model for the successful Kaizen implementation and the sustainable performance of SMEs in Vietnam. "STEAM-ME" implies that organizations need to have a new airflow as "steam" to make all of its members refreshed and brimful of energy to gain significant success in implementing Kaizen, and improve their business performance as well as competitive advantage for their sustainable development. Notably, the novel model can efficiently demonstrate organic relationships among its components which all have positive and significant impacts on the successful Kaizen implementation and sustainable performance of SMEs in Vietnam.

## 1. Introduction

Nowadays, the inevitable globalization has offered several opportunities and many challenges to almost every business organization. Thus, being competitive on the marketplace is critical for their survival and sustainable growth [1]. To improve their competitiveness, different businesses may have different strategies; among them, continuous improvement for operational excellence has been preferably used in practice [2, 3]. However, applying the Kaizen concept for continuous improvement has been an attractive choice [4–6] because it significantly helps to increase quality, improve level of efficiency, and reduce waste and production cost for business excellence [7]. Thus, Kaizen is one of the most common "Japanese business terms" [8]. The Kaizen approach has been successfully implemented in different industries in several countries regardless of business sectors. Homma [9] and Costa & Filho [10] pointed out that Kaizen can be effectively used not only for industrial development but also for productivity improvement in public services and utility management such as energy or healthcare; or non-firm-related use; or even the improvement of environmental performance [11–13]. The applicability of Kaizen and its practical benefits in terms of inventory reduction, customer satisfaction, lead time, and waste reduction, etc. have been validated by different researchers worldwide, such as Chahal et al. [14], Marodin et al. [15], Gupta et al. [16], Belekoukias et al. [17], Fullerton et al. [18], Ingelsson & Mårtensson [19], Prashar [20], Teehan & Tucker [21], and Dora et al. [22]. Consequently, Lozano et al. [8] concluded that Kaizen has economic and environmental implications because it closely relates to organizational systems and business strategies

Table 1: Criteria in identifying types of SMEs in Vietnam.

| Areas | Criteria | SME types under Decree 39/2018 | | | SME types under Decree 56/2009 | | |
|---|---|---|---|---|---|---|---|
| | | *Micro* | *Small* | *Medium* | *Micro* | *Small* | *Medium* |
| Agriculture, forestry, aquaculture, industry and construction | No. of employees *(e)*∗ | ≤ 10 | ≤ 100 | ≤ 200 | ≤ 10 | ≤ 200 | ≤ 300 |
| | Total capital *(BV)*∗∗ | ≤ 3 | ≤ 20 | ≤ 100 | | ≤ 20 | ≤ 100 |
| | Annual revenue *(BV)*∗∗ | ≤ 3 | ≤ 50 | ≤ 200 | | | |
| Trading and Services | No. of employees *(e)*∗ | ≤ 10 | ≤ 50 | ≤ 100 | ≤ 10 | ≤ 50 | ≤ 100 |
| | Total capital *(BV)*∗∗ | ≤ 3 | ≤ 50 | ≤ 100 | | ≤ 10 | ≤ 50 |
| | Annual revenue *(BV)*∗∗ | ≤ 10 | ≤ 100 | ≤ 300 | | | |

Notes: ∗ employees; ∗∗ Billion Vietnam Dong.

by engaging all levels of management and employees for continuous improvement.

On the other hand, the important role of small and medium enterprises (SMEs) in most socioeconomic activities has been well recognized globally; thus, it is one of the common topics discussed in multilateral cooperation forums and meetings, such as Asia-Pacific Economic Cooperation (APEC), Organisation for Economic Cooperation and Development (OECD), Asia–Europe Meeting (ASEM), and Association of Southeast Asian Nations (ASEAN) [23]. Especially, its importance is further affirmed in APEC 2017 as it is one of their four key priorities, *"Strengthening Micro, SMEs' Competitiveness and Innovation in the Digital Age"*.

In Vietnam, the number of SMEs accounts for about 97.5% of 561,064 enterprises of all types operating in Vietnam [24]. The new definition of SMEs has been issued in Article 6 of Decree No. 39/2018/ND-CP by the Government and comes into effect since March 11th, 2018. SMEs can be classified into three categories depending on two criteria: (1) annual average number of employees contributing Social Insurance (No. of employees) and (2) annual revenue or total capital registered. These criteria are somehow different from previous Decree 56/2009/ND-CP. Table 1 briefly presents details of these categories mentioned in the two Decrees.

Practically, SMEs not only contribute over 40% of national GDP and 17.26% of the annual national budget but also employ more than 50% workforce [23]; consequently, SMEs are an important contributor to the development of Vietnam economy. Comparing between 2017 and 2016, we found that the number of medium enterprises increased by 23.6%, small ones increased by 21.2%, and micro ones increased by 65.5%. But there were also more than 60,660 enterprises bankrupted in 2016 [23], indicating that SMEs are vulnerable in the competitive marketplace and current economic context due to their limited resources and capacity [25].

However, with the small and medium business scale, SMEs have their advantages in flexibly renovating themselves and adopting new management approaches as well as easily adapting to the changes in their business environment. Therefore, when advanced management approaches such as Kaizen and 5S are introduced, they are always willing to learn and apply as much as they can to improve their operational efficiency, effectiveness, and productivity [25]. In the current context of international integration, the improvement has become notably mandatory since the introduction of ASEAN Economics Community (AEC) in 2015 because the free movement of goods, services, and investments as well as freer flow of capital and skills among the ASEAN countries results in more intensive competition on the marketplace. In such competitive environment, providing good products/services at reasonable prices becomes critical to the survival and growth of the enterprises. And Kaizen has been considered as an effective tool to improve the productivity, cost-effectiveness, profitability, efficient use of capital, reduction of operating time, and competitive advantage [26].

Kaizen has been well transferred to Vietnam since early 1990s. Over the years, more and more companies located throughout Vietnam are trying their best to implement Kaizen in their operations. Though there are some differences in the practical implementation of Kaizen among Japanese-owned companies, Japanese-joint companies, and foreign and local ones, many of them have well recognized the importance of Kaizen for their development. From the training workshops on Kaizen organized in Vietnam, practitioners find that Kaizen approach is suitable to be widely applied across the industrial enterprises in Vietnam because it is considered simple and inexpensive. However, its practical implementation is actually more complex than expected. Consequently, some of them fail to implement Kaizen in their companies but some with successful implementation have gained significant benefits in terms of increased efficiency and productivity. Therefore, this study is aimed at identifying key determinants of the successful Kaizen implementation and their impacts on the sustainable performance to encourage more and more SMEs in Vietnam to effectively deploy Kaizen approach to improve their competitiveness.

The rest of this paper is organized as follows: Section 2 reviews relevant literature about Kaizen and organizational performance as well as key factors affecting them before research hypotheses and model are proposed in this study.

Methods used for data collection and data analysis are explicitly presented in Section 3 while empirical results are explained in Section 4. Section 5 provides detailed discussions and managerial implications departed from the obtained results. Conclusions make up the last section.

## 2. Literature Review

To achieve the above-mentioned research objectives and support the following analyses and discussions, this section will present some key terminologies, such as "Kaizen" and "sustainable performance", and cover some fundamental literature about (1) sustainable performance of an organization; (2) briefs about Kaizen; (3) Kaizen implementation and measures of successful Kaizen implementation; (4) relationship between Kaizen implementation and organizational performance; and (5) factors affecting the success of Kaizen implementation. Through such presentation, research hypotheses and research model investigated in this study are accordingly proposed.

*2.1. Sustainable Performance.* Organizational performance refers to the extent to which an organization succeeds or achieves its objectives and strategies [27]. Proper management of performances helps organizations to effectively capture their current situation, monitor their progress in achieving their goals, and identify latent causes obstructing their success [28]. Current context of fierce competitive marketplace urges organizations to strive for their long-term development through "sustainable performance" which is differently defined by different scholars. For example, Artiach et al. [29] defined it as the degree to which an organization incorporates its concerns in terms of profit, environment, people, and governance into its operations for ultimate impacts on the organization and society, whereas, Stanciu et al. [30] defined it as the ability of organizations to satisfy the needs and expectations of their stakeholders based on long-term, balanced, and effective management with proper awareness of their staffs through their learning and applying of improvements and innovations; UBS [31] claimed that sustainable performance focuses on long-term and consistent benefits to stakeholders.

Literally, sustainable performance and sustainability have been interesting topics in different research areas as found in [32–38]. Several researchers, such as Long & Nguyen [39], Norazlan et al. [40], Moldan et al. [41], and Schoenherr [42], agreed that the sustainable performance is measured with three dimensions, namely, (1) economic performance, defined as the extent to which an organization improves its operations, market, and financial results; (2) environment performance, defined as the extent to which an organization improves its control of pollution and its resource efficiency; and (3) social performance, defined as the extent to which an organization improves its practical outcomes related to its employees and community. Considered as the key pillars of triple bottom line theory, balancing these dimensions is critical to improve organizational competitive advantages [43].

*2.2. Briefs about Kaizen.* As human always wants to become better and better, consistent improvement is a fundamental need. Searching for ways to improve business operations led to the term "Kaizen" which combines two separate words: "Kai" (change) and "Zen" (good/better). Thus, "Kaizen" is commonly understood as "change for the better" or "continuous improvement" [26, 44], "a philosophy guiding individuals and organizations to do better achievements in the long term" [45] or "self-sacrifice for everyone's betterment". Over the last 30 years, the term "Kaizen" has become a popular management concept in the 21st century [45–47]. Kaizen can be used in all aspects of life, including business organizations [48, 49]. Nowadays, Kaizen is considered as grand-scale, companywide, daily, and everywhere improvement made by everyone. Fundamentally, Kaizen is aimed at transforming work area and developing employees for specific targets in an escalated timeframe [49–51] by using cross-functional teams, training employees, and rotating jobs [51, 52] so that the workforce can be subtly controlled to avoid latent conflicts with the management [50]. According to Lemma [26], Kaizen is a firm-level process working as a strategic tool to improve the productivity in manufacturing firms. It is actually the core of "monozukuri" which means "making things" to satisfy customers. By focusing on three areas for improvement, namely, Muda (waste), Mura (discrepancy), and Muri (strain), if implemented correctly, Kaizen is a donor to make employees have more positive attitude towards their work and enhance the self-esteem and the awareness of their responsibilities towards their workplace, their working processes, and ways to improve them because they are always encouraged to share their ideas to make the existing standards better [52].

Practically, Kaizen is a process-oriented method to make small, immediate, and incremental improvements in work standards generated repeatedly by workers [44]. Thus, Kaizen mainly asks for the engagement of all members in the improvement effort [48, 53], and there is no need for a huge capital investment nor an enormous preparation at one time. According to Lozano et al. [8], Kaizen philosophy is based on three pillars: (1) preventing waste, (2) organizing workplace, and (3) making things standardized. Therefore, according to Jurburg et al. [54], Kaizen is an effective tool to (1) cheaply abolish or lessen hidden costs resulting from undue waste; (2) improve operational performance in terms of high-quality products, low production cost, and short service time; (3) optimize operations with minimum downtime which is irrecoverable [55]; and others. Consequently, Kaizen is considered as a good strategy for any organization to improve its competitive advantages.

*2.3. Kaizen Implementation and Measures of Its Success*

*2.3.1. Kaizen Implementation.* Kaizen is a companywide process which involves all people from high-level management to front-line employees. The former provides commitment and supports to motivate the latter who directly performs the "continuous improvement". In implementing Kaizen in practice, a Plan-Do-Check-Act (PDCA) cycle is usually

used to deal with not only unit-functional but also cross-functional problems in their operations. Specifically, areas for improvement must be firstly identified (planning phase) before corrective actions are taken (doing phase). In the doing phase, also called the Kaizen implementation, several techniques such as 5 Whys [56] and Value Stream Mapping (VSM) [57–59] can be used to fully capture the root causes of the problems, for example, the quality level, scrap/rework rate, layout performance, and amount of certain resources used in each stage of the process. From the identified causes, proper improvement solutions should be considered and accordingly implemented.

In the checking phase, we need to closely monitor the impacts of the Kaizen solutions on the detected problems and determine whether positive results can be observed as expected. If the solutions are satisfactory, in the acting phase, we should formally set the Kaizen activities as new standards and move forward; otherwise, an adjustment in terms of solutions, implementing methods, etc. should be reconsidered in the next cycle. Once Kaizen is successfully implemented in an organization, innovation becomes its cutting-edges in strengthening its competitiveness, and the Kaizen activities should be standardized and turned into permanent tasks in their processes [26].

Literally, Kaizen is a slow and long-term process of changes rather than a sudden intervention [60]. Implementing Kaizen should first begin with reviewing the existing processes and identifying areas for improvement before providing proper training, tools, and structure to employees. Then, employees are encouraged to become aware of all possible problems in their daily operations and think about feasible improvement solutions. Gradually, they likely take their mental ownership of their individual processes; finally, they consider improving the processes as a critical part of their responsibility.

Though the Kaizen principles are quite easy to be fully understood, there are still several challenges in its implementation in practice due to the difficulties in managing Kaizen activities [61–64]. Several obstacles have been found, such as resistance to change among mature workers, the abstraction of "continuous improvement" concepts [65], the absence of compensation or reward, lack of proper training for employees and long delays in getting suggestions processed [66], lack of resources to run Kaizen activities, lack of focus due to business pressure and lack of understanding of the need to change [64], lack of knowledge, and poor employee participation [22]. Thus, innovation and education are key components in Kaizen implementation [12].

*2.3.2. Measures of Successful Kaizen Implementation.* Though there have been several studies in identifying factors affecting the success of Kaizen implementation, there are a few effective approaches to measure the overall success. For instance, "Overall Equipment Effectiveness" (OEE) proposed by Nakajima [67] focused on equipment utilization while Domingo & Aguado [68] proposed a more comprehensive metric, "Overall Environmental Equipment Effectiveness" (OEEE). However, through group discussions with leaders from six SMEs successfully implementing Kaizen, they failed to deploy OEEE in measuring the success. Thus, further discussions were conducted to explore what measures should be used. Based on the qualitative research, there are four measures suggested: (1) effective usage of existing resources (including space utilization) for incremental and continuous improvement; (2) increased efficiency by optimizing operations and processes with properly arranged layouts of work area and work flows to minimize superfluous movement or operations as well as production costs; (3) safer, cleaner, and better-organized working environment perceived by relevant stakeholders; and (4) positive mindset of "continuous improvement" among employees. The improvement level of these measures is evaluated in 5-Likert scale as explained in Section 3.

*2.4. Relationship between Kaizen and Sustainable Performance.* Several scholars worldwide have made special efforts to promote the benefits of Kaizen across different countries. Existing researches from different industries clearly show that successful implementation of Kaizen brings several benefits, including reducing scraps, reworks, inventory, unnecessary movement, production lead time, and failures in tools/machinery and improving product quality, productivity, delivery, floor security and safety, employees' motivation, responsibility, cross-communication, and teamwork, among others [69, 70]. Therefore, Kaizen helps a business firm to satisfy its stringent customers' requirements and expectations, gaining more trusts from its stakeholders, and boosting its competitive advantages through the increase in customer satisfaction, employee satisfaction, productivity, and financial performance [71]. Moreover, as public are paying more and more attention to environmental protection and social impacts, successfully implementing Kaizen will help organizations to achieve "green attributes" which were found to have positive and direct influence on business performance of industrial manufacturers [72]. Consequently, successful Kaizen implementation helps to sustain organizational performance [40, 73].

*2.5. Factors Affecting the Success of Kaizen Implementation.* Existing literature shows that there are a number of factors affecting the success of Kaizen implementation. For example, an open working environment that allows effective cross-communication and encourages innovation is critical for a better understanding between management bodies and their employees as well as the sharing of improvement ideas for easier and faster processes based on their practical experience [74–76]. In addition, strong commitments from top management in implementing Kaizen with clear approaches, strategies, policies, and targets also play significant roles in sustaining improvement actions [64, 74] and building Kaizen culture because they help to effectively support, direct, and allocate relevant resources [77]. In particular, this study conducted a thorough search of more than 200 research reports published in the last two decades on key databases such as ScienceDirect, Elsevier, EBSCOhost, Springer, and Emerald. For brevity, only some reports cited in main texts

are listed in the References while many others are listed in Appendix I. The search well gives the rational validation to the six key affecting factors presented in Table II.1 (Appendix II). Similar approaches can be found in [70, 78–81]. The identified determinants are clarified in the following subsections.

*2.5.1. Supports from Senior Management.* As continuous improvement is the core of Kaizen, senior leaders must act as the most vital driving force to make the improvement process effectively implemented with their strong supports to ensure the full and active participation of every member [54, 82]. Such supports, including spiritual and physical ones as well as necessary resources allocated, can be expressed in verbal or written commitments, statements, policies, plans, or even direct involvement in following up the progress of Kaizen and related practical activities [47, 74]. The supports and commitments should be well formulated and effectively articulated as a motivational factor for employees to perform better [83] and more engage in the continuous improvement [51]. Further evidence of this factor can be found in [81, 84–86]. It is found that such involvement from senior leaders is the most fundamental factor affecting the success of continuous improvement programs [87–89].

With this factor, the following hypotheses will be investigated:

(i) H1: Support from senior management has positive impacts on the successful Kaizen implementation.

(ii) H2: Support from senior management has positive impacts on the sustainable performance.

*2.5.2. Training.* Literally, the importance of training and education for the success of Kaizen has been well validated by several scholars worldwide [52, 74, 90, 91] because it is critical for not only providing "need-to-know" basis but also consolidating human development and changing the employees' mindset [92]. According to Soltero & Waldrip [93], Kaizen training should be first provided to managers/supervisors/leaders of all levels because they not only focus on soliciting proposals but also act as "bellwethers" in the journey for successful Kaizen implementation. Therefore, such training helps them to (1) clearly understand the philosophy; (2) realize positive outcomes of Kaizen implementation for their better reinforcement and engagement; (3) know how to motivate and elicit active participation of their employees; and (4) lead the whole process of continuous improvement.

Moreover, through on-the-job/off-the-job training and proper schemes for job rotation or relocation, organizations gain certain benefits from innovative suggestions/ideas of their employees [94, 95]. Importantly, the training not only equips the employees with new skills and updated knowledge but also raises their awareness of continuous improvement [96] and sense of belonging [97]. In addition, there is a statistically significant relationship between employee training and employee motivation [98, 99] as well as employee engagement [100–103].

With this factor, the following hypotheses will be investigated:

(i) H3: Training has positive impacts on the successful Kaizen implementation.

(ii) H4: Training has positive impacts on the sustainable performance.

*2.5.3. Environment.* Realyvásquez et al. [104] also pointed out that environmental elements such as air quality, humidity, temperature, noise, lighting have significant impacts on workers' psychological characteristics and their performance whereas Day & Randell [97] claimed that a healthy working environment is one of the cores of Kaizen philosophy because it positively results in significant increase in employees' commitment, retention, stakeholders' satisfaction, and firms' financial performance. In addition, working environment strongly affects organizational productivity [105] and employee satisfaction [106–108], leading to an increase in overall performance. Hence, a good working environment in terms of openness, cleanliness, tidiness, social interaction, interpersonal relationship, group norms and values, organizational structure, etc. makes employees self-motivated and concentrated to their work with better behavior, attitude, and productivity [109].

Similarly, Liker & Franz [110] and Soltero & Waldri [93] pointed out that Kaizen implementation needs a democratic working environment in which open communication, creativity, innovation, and improvement proposals among employees are appreciated and encouraged. Aguado et al. [111] claimed that innovation is the best approach to efficiency and sustainability. As such, Stadnicka & Sakano [112] suggested that organizations should create a friendly working environment and build their culture of continuous improvement for their successful Kaizen implementation.

With this factor, the following hypotheses will be investigated:

(i) H5: Environment has positive impacts on the successful Kaizen implementation.

(ii) H6: Environment has positive impacts on the sustainable performance.

*2.5.4. Assessment.* As discussed above, training is mainly aimed at changing people's behavior. To have an effective training program, Gravells [113] proposed a training cycle with five stages: identifying needs, planning and designing, delivering, assessing, and evaluating. Among them, assessing training needs and effectiveness of training program as well as increase in employee performance/ability/skills/attitudes in their work is a critical task [114–116]. Therefore, employee assessment must be done before the training, in the training, and after the training so that we can have necessary actions to improve the performance of the whole system. Importantly, such assessment provides useful information to evaluate the effectiveness of the training program and to design future ones better.

Nonetheless, in order to ensure the success of Kaizen implementation, regularly assessing the improvement of work ergonomics (employee productivity, efficiency, attitude, etc.) and working environment (vibrations, noise, internal

air pollution, microclimate, radiation, dustiness or energy expenditure of the worker, etc.) is critical [117]. Such regular activity is of great help in taking prompt corrective actions if needed to properly adjust relevant processes and/or approaches to achieve certain specific targets.

With this factor, the following hypotheses will be investigated:

(i) H7: Assessment has positive impacts on the successful Kaizen implementation.

(ii) H8: Assessment has positive impacts on the sustainable performance.

*2.5.5. Motivation.* In the field of organizational behavior, there are two key components of job motivation: intrinsic motivation and extrinsic motivation [118, 119], which urge employees to accomplish their personal and organizational goals [120–124]. And there are several motivation approaches, including salary and benefits [125–128], rewards and recognition [129–131], career promotion [132–137], and empowerment [129, 138–141]. Motivation approaches should be carefully considered and selected in line with required improvements [142, 143].

In the current context of fierce competition on the marketplace, motivated and engaged employees are usually considered as invaluable asset and competitive advantage of an organization [144]. And, employee motivation is a key determinant of organizational success [145] because motivated employees tend to foster a creative working environment [146–148] and accept changes for better [146], resulting in increased profitability [149], higher customer satisfaction and loyalty due to better customer service [150, 151], and improved organizational competitiveness [152]. Besides, it is also found that motivated workforce usually (1) think creatively and proactively [153, 154]; (2) have higher job satisfaction [155–157]; (3) perform better [151, 158, 159]; (4) have higher life satisfaction [160, 161]; (5) have higher productivity [150, 162]; and (6) are more diligent and loyal [163–165]. As such, employee motivation is one of the key determinants for the success of Kaizen implementation [63, 86, 166–169].

With this factor, the following hypotheses will be investigated:

(i) H9: Motivation has positive impacts on the successful Kaizen implementation.

(ii) H10: Motivation has positive impacts on the sustainable performance.

*2.5.6. Mindset.* This factor is newly proposed in this study through a formal qualitative research as presented in Section 3. In this study, the term "mindset" refers to that of all management levels and employees. Literally, Dweck [170] defined a mindset as the views a person adopts for himself/herself. Such views, including personal assumptions and expectations, significantly affect his/her usual behaviors and relevant responses to his/her daily affairs. Besides, Thomas et al. [171] defined employees' mindset as their attitudes, behaviors, and practices which shape the way an organization approaches and executes its strategies. There are two major types of mindset: fixed mindset and growth mindset [170]. The growth mindset is more important because it provides more benefits in terms of creating resilience [172–174], tenacity [172], improving collaboration, communication and engagement [174], and increasing motivation for learning and developing [175]. However, relationships between mindset and successful Kaizen implementation as well as sustainable performance are left unsolved in the current literature. Therefore, investigating its impacts is one of the key contributions presented in this study.

With this factor, the following hypotheses will be investigated:

(i) H11: Mindset has positive impacts on the successful Kaizen implementation.

(ii) H12: Mindset has positive impacts on the sustainable performance.

*2.5.7. Engagement.* To ensure the success of Kaizen implementation, several studies have claimed that all management levels and employees should proactively engage in the journey towards operational excellence through continuous improvement. The engagement from management levels closely relates to their supports and commitments. And that from employees should be further examined. According to Takeuchi et al. [176], employees in Toyota are appreciated as a source of knowledge and wisdom of experience; thus, they should engage in the continuous improvement process. Practically, there have been several different definitions of employee engagement in the field of organizational behavior, but generally it is all about how employees stay either emotionally, cognitively, or physically connected with their organizations [177–179]. Anitha [102] claimed that employee engagement is critical for an organization to gain not only useful business performance results but also competitive advantages over its rivals. It is because engaged employees help organizations serve customers better in terms of satisfaction, loyalty, productivity, and profit [180]. Moreover, they tend to be more satisfied with their jobs, committed, and loyal to their organizations [181] because they believe that they constitute a part of the organization [182]. Siddhanta & Roy [183] found that engagement makes employees more motivated and committed; thus, it positively affects organizational performance [182, 184–193]. Hence, engaged employees tend to proactively and enthusiastically participate in assigned activities with their full responsibilities.

To improve employee engagement, Marinova et al. [194] suggested that companies build different incentive systems and continuous improvement programs so that employees become satisfied and motivated with their jobs. Stadnicka & Sakano [112] claimed that active participation of all members, including management and employees, is critical for the success of continuous improvement/Kaizen implementation of an organization.

With this factor, the following hypotheses will be investigated:

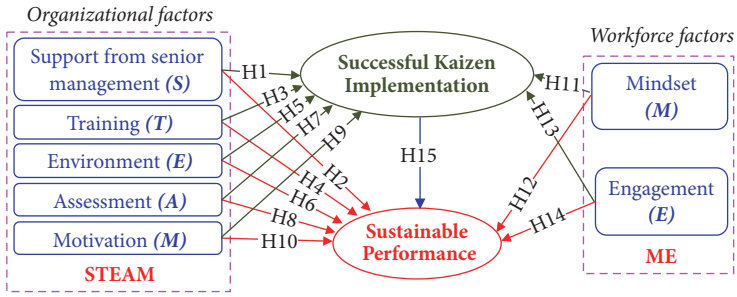



Figure 1: Proposed research model.

(i) H13: Engagement has positive impacts on the successful Kaizen implementation.

(ii) H14: Engagement has positive impacts on the sustainable performance.

Moreover, with the relationship between Kaizen and sustainable performance of organizations presented in Section 2.4, this study will also investigate the following hypothesis:

(i) H15: Successful Kaizen implementation has positive impacts on the sustainable performance.

Thus, the research model proposed in this study is visually presented in Figure 1.

## 3. Research Method

This research is conducted in three main phases as explained in the followings.

*3.1. Phase 1: Questionnaire Design.* This initial phase is aimed at constructing a complete questionnaire for a formal survey. From the exhaustive literature review mentioned in Section 2.5, a list of six determinants, namely, support from senior management, training, environment, assessment, motivation, and engagement, is created and then used to conduct a qualitative research to validate the relevance of the factors and explore other prospective ones. The qualitative research invited seven experts from two companies which have successfully implemented Kaizen in Dong Nai and Binh Duong. Among the seven, two are working as director and vice director, three working as managers of their warehouses and production departments, and two working as Kaizen leaders. Their practical experiences from such positions would provide clear insights into these factors as well as suggesting possible measures for the success of Kaizen implementation in their cases.

From the initial interviews, they not only agreed about the relevance of the six listed factors but also proposed a new factor named "mindset of all personnel in an organization" to be considered in this study. The importance of this newly added factor has already been discussed in Section 2.5. Moreover, they also provided some key measures of a successful Kaizen implementation as discussed in Section 2.3.2 above. These inputs were carefully considered in the design of primary survey questionnaire which was then used in a pilot test to evaluate the lucidity of each surveyed statement in terms of meaning and word usage. Four participants from top management levels of other two companies located in Ho Chi Minh City joined the pilot test. Their feedback was carefully checked and integrated to refine the questionnaire for an official survey. The final version consists of three major parts:

(1) Seven independent factors are composed of 34 observed items. The participants were asked to evaluate the importance level of each item on a 5-Likert scale towards the success of Kaizen implementation in their organizations, where 1 indicates the least important level and 5 indicates the most important level.

(2) Successful Kaizen implementation is composed of 6 observed items whose success levels are evaluated on a 5-Likert scale where 1 indicates lowest level and 5 indicates highest level.

(3) Organizational performance consists of 6 items reflecting the economic performance, environment performance, and social performance. The participants were asked to evaluate the current performance of these items on a 5-Likert scale (1- "unacceptable", 2- "inconsistent", 3- "rather effective", 4- "effective", 5- "exceptional").

For brevity, full contents of these constructs and detailed items will be supplemented on request.

*3.2. Phase 2: Survey and Data Capture.* The official survey was conducted from March 15, 2018, to June 20, 2018. First, from personal network with other trainees participating in previous workshops on Kaizen, this study lists 62 SMEs which have successfully implemented Kaizen; among them, 34 SMEs are located in the South, 21 SMEs are in the north, and the rest are in the middle of Vietnam. Then, 254 hard copies of the final questionnaire were directly delivered to 254 people working as directors, vice directors, department managers, or Kaizen leaders in the selected SMEs. Because the objectives of this study were effectively communicated, most of them actively took part in the survey. Therefore, 237 out of 254 pieces of completed questionnaires were collected. Among them, there were 24 pieces invalid, so, data from 213 valid observations were finally analyzed in this study. Prior to

Table 2: Codes of investigated constructs and observed items.

| *Constructs* | *No. of items* | *Codes* |
|---|---|---|
| Supports from senior management *(SUP)* | 6 | SUP1 ⟶ SUP6 |
| Training *(TRA)* | 4 | TRA1 ⟶ TRA4 |
| Environment *(ENV)* | 4 | ENV1 ⟶ ENV4 |
| Assessment *(AST)* | 5 | AST1 ⟶ AST5 |
| Motivation *(MOT)* | 5 | MOT1 ⟶ MOT5 |
| Mindset *(MIN)* | 6 | MIN1 ⟶ MIN6 |
| Engagement *(ENG)* | 4 | ENG1 ⟶ ENG4 |
| Successful Kaizen implementation *(SUC)* | 6 | SUC1 ⟶ SUC6 |
| Organizational performance *(PER)* | 6 | PER1 ⟶ PER6 |

Table 3: Descriptive statistics of respondents.

| *Demographic Characteristics* | | *Frequency* | *Percent (%)* |
|---|---|---|---|
| *Working Position* | Kaizen leader | 62 | 29.1 |
| | Department Manager | 107 | 50.2 |
| | Director/Vice Director | 44 | 20.7 |
| *Enterprise Location* | South of Vietnam | 172 | 80.8 |
| | Middle of Vietnam | 7 | 3.2 |
| | North of Vietnam | 34 | 16.0 |
| *Enterprise Size* | Micro | 14 | 6.6 |
| | Small | 84 | 39.4 |
| | Medium | 115 | 54.0 |
| *Ownership Type* | State-owned enterprise | 9 | 4.2 |
| | Private enterprise | 37 | 17.4 |
| | Joint-venture enterprise | 79 | 37.1 |
| | Foreign-owned enterprise | 88 | 41.3 |

the analysis, the investigated constructs and their observed items are accordingly coded as shown in Table 2.

*3.3. Phase 3: Data Analysis.* In this phase, the collected data were first screened. Some data analysis approaches like exploratory factor analysis (EFA) and scale reliability analysis with Cronbach's Alpha ($\alpha$) coefficients were deployed with IBM SPSS V.22. Fundamentally, EFA is considered appropriate if its parameters well satisfy the following criteria: (1) eigenvalue $\geq 1$; (2) total variance explained $\geq 50\%$; (3) KMO $\geq 0.5$; (4) significance (Sig.) coefficient of KMO test $\leq 0.05$; (5) factor loadings of all observed variables $\geq 0.4$ as there are 213 observations in the sample; and (6) weight difference between the loadings of two factors $> 0.3$ [195]. And, key criteria to judge if a scale is considered reliable include the following: (1) all corrected item-total correlations of its components are $> 0.3$; (2) its $\alpha$ coefficient $\geq 0.7$ [196].

After EFA and scale reliability analysis, the extracted factors are further analyzed with (1) confirmatory factor analysis (CFA) to affirm their unidirectionality, internal consistency, convergence value, and distinguishing value; (2) structural equation modelling (SEM) to test the validity of the proposed research model and stated hypotheses [39, 43]. According to Hair et al. [197] and Steenkamp & Trijp [198], these two analyses are considered appropriate if the following criteria are satisfied: (1) the significance value (p-value) of the Chi-square test $\leq 0.05$; (2) ratio of Chi-square (CMIN) over the degree of freedom (df), CMIN/df $\leq 2.00$ (in some cases, CMIN/df $\leq 3.00$ is also acceptable); (3) the goodness of fit index (GFI), Tucker–Lewis index (TLI), and comparative fit index (CFI) $\geq 0.90$; (4) root mean square error of approximation (RMSEA) $\leq 0.08$; (5) overall reliability $\geq 0.6$; and (6) extracted variance $\geq 0.5$.

## 4. Empirical Results

*4.1. Descriptive Statistics.* Some key characteristics of the 213 respondents are briefly shown in Table 3. Particularly, among the 213 valid observations, there were 172 people, accounting for 80.8%, from 34 SMEs located in the South because most of existing joint-venture and foreign-owned enterprises are located in the South due to special calls for investment and attractive policies by the local authorities to create dynamic business environment.

Table 4: EFA rotated matrix of independent variables and reliability analysis.

| | *Component*[a] | | | | | | | *α* | *CITC*[b] | *α if item deleted* |
|---|---|---|---|---|---|---|---|---|---|---|
| | *1* | *2* | *3* | *4* | *5* | *6* | *7* | | | |
| MIN1 | 0.938 | | | | | | | 0.845 | 0.905 | 0.890 |
| MIN3 | 0.853 | | | | | | | | 0.791 | 0.805 |
| MIN5 | 0.844 | | | | | | | | 0.781 | 0.807 |
| MIN6 | 0.828 | | | | | | | | 0.748 | 0.811 |
| MIN4 | 0.809 | | | | | | | | 0.726 | 0.814 |
| MIN2 | 0.784 | | | | | | | | 0.707 | 0.816 |
| SUP1 | | 0.916 | | | | | | 0.832 | 0.863 | 0.892 |
| SUP3 | | 0.850 | | | | | | | 0.791 | 0.803 |
| SUP5 | | 0.849 | | | | | | | 0.785 | 0.804 |
| SUP2 | | 0.831 | | | | | | | 0.748 | 0.809 |
| SUP6 | | 0.811 | | | | | | | 0.735 | 0.810 |
| SUP4 | | 0.795 | | | | | | | 0.708 | 0.814 |
| AST2 | | | 0.899 | | | | | 0.851 | 0.838 | 0.818 |
| AST1 | | | 0.876 | | | | | | 0.795 | 0.827 |
| AST4 | | | 0.856 | | | | | | 0.778 | 0.831 |
| AST3 | | | 0.854 | | | | | | 0.778 | 0.831 |
| AST5 | | | 0.775 | | | | | | 0.686 | 0.750 |
| ENV1 | | | | 0.891 | | | | 0.865 | 0.809 | 0.785 |
| ENV3 | | | | 0.858 | | | | | 0.721 | 0.823 |
| ENV4 | | | | 0.806 | | | | | 0.668 | 0.845 |
| ENV2 | | | | 0.783 | | | | | 0.655 | 0.849 |
| MOT1 | | | | | 0.885 | | | 0.811 | 0.787 | 0.735 |
| MOT3 | | | | | 0.831 | | | | 0.681 | 0.784 |
| MOT4 | | | | | 0.765 | | | | 0.609 | 0.816 |
| MOT2 | | | | | 0.738 | | | | 0.590 | 0.823 |
| ENG3 | | | | | | 0.795 | | 0.773 | 0.633 | 0.718 |
| ENG4 | | | | | | 0.784 | | | 0.618 | 0.726 |
| ENG1 | | | | | | 0.763 | | | 0.582 | 0.744 |
| ENG2 | | | | | | 0.761 | | | 0.554 | 0.758 |
| TRA1 | | | | | | | 0.795 | 0.765 | 0.599 | 0.694 |
| TRA3 | | | | | | | 0.791 | | 0.605 | 0.691 |
| TRA2 | | | | | | | 0.755 | | 0.552 | 0.719 |
| TRA4 | | | | | | | 0.706 | | 0.514 | 0.740 |

Extraction method: Principal Component Analysis.
Rotation method: Varimax with Kaiser Normalization.
(a) Rotation converged in 6 iterations.
(b) Corrected item-total correlation.

Moreover, more than 50% of the participants are working as department managers and about 30% working as Kaizen leaders in the investigated enterprises; generally, about 80% of the respondents are from joint-venture and foreign-owned enterprises. In addition, 54% and about 40% of the participants are from medium size and small size enterprises, respectively.

*4.2. Exploratory Factor Analysis.* The latent relationships among the 34 observed variables of seven key factors are first investigated with EFA approach. Results from the first analysis showed that MOT5 failed to satisfy the required criterion of discrimination in its loadings among two extracted factors; thus, it was dropped out from the list of variables. The second analysis of 33 items resulted in seven factors extracted as shown in Table 4. With the obtained KMO = 0.792, the significance of Bartlett's test p-value ≤ 0.001, and the satisfactory factor loadings of the components, EFA analysis used in this study is considered appropriate.

*4.3. Scale Reliability Analysis.* These extracted scales were then tested for their internal consistency with scale reliability

Table 5: EFA rotated matrix of dependent variables and reliability analysis.

| | Component[a] | | α | CITC[b] | α if item deleted |
|---|---|---|---|---|---|
| | 1 | 2 | | | |
| PER1 | 0.908 | | 0.875 | 0.853 | 0.891 |
| PER2 | 0.874 | | | 0.808 | 0.898 |
| PER6 | 0.842 | | | 0.765 | 0.904 |
| PER3 | 0.837 | | | 0.758 | 0.905 |
| PER5 | 0.826 | | | 0.748 | 0.906 |
| PER4 | 0.770 | | | 0.679 | 0.916 |
| SUC4 | | 0.884 | 0.824 | 0.816 | 0.860 |
| SUC1 | | 0.862 | | 0.786 | 0.864 |
| SUC3 | | 0.805 | | 0.708 | 0.877 |
| SUC6 | | 0.780 | | 0.676 | 0.882 |
| SUC2 | | 0.761 | | 0.659 | 0.885 |
| SUC5 | | 0.759 | | 0.657 | 0.885 |

Extraction method: Principal Component Analysis.
Rotation method: Varimax with Kaiser Normalization.
(a) Rotation converged in 3 iterations.
(b) Corrected item-total correlation.

Table 6: Confirmatory factor analysis.

| Term | Scale | No. of Observed variables | Reliability test | |
|---|---|---|---|---|
| | | | Cronbach's α | Composite α |
| Determinants of successful Kaizen implementation and sustainable performance of SMEs in Vietnam | Support from senior management (SUP) | 6 | 0.832 | 0.835 |
| | Training (TRA) | 4 | 0.765 | 0.769 |
| | Environment (ENV) | 4 | 0.864 | 0.867 |
| | Assessment (AST) | 5 | 0.851 | 0.858 |
| | Motivation (MOT) | 4 | 0.811 | 0.840 |
| | Mindset (MIN) | 6 | 0.845 | 0.859 |
| | Engagement (ENG) | 4 | 0.773 | 0.789 |
| Successful Kaizen implementation (SUC) | | 6 | 0.824 | 0.866 |
| Sustainable performance (PER) | | 6 | 0.875 | 0.896 |

analysis. Their results are shown in columns "$\alpha$" and "CITC" of Table 4.

The high values of $\alpha$ coefficients (ranging from 0.773 to 0.865) and all corrected item-total correlations (CITC) larger than 0.3 indicate that the extracted scales have high internal consistency because they well satisfy the required criteria for scale reliability analysis mentioned in Section 3.3; hence, these extracted scales are considered reliable for further analysis, such as CFA and SEM.

With the same token, EFA approach was also used to explore the structure of the dependent factors "successful Kaizen implementation" and "organizational performance". Table 5 clearly shows that the use of EFA approach for these two scales is also appropriate because its KMO is 0.887, the significance of Bartlett's test is p-value ≤ 0.001, and the factor loadings of the components are all larger than 0.4.

*4.4. Confirmatory Factor Analysis.* Table 6 briefly shows the composite reliability of the investigated factors and the two dependent scales denoted by SUC and PER. And Figure 2 displays estimated standardized results of saturated model in CFA, including CMIN=1253.360, df= 909, p-value≤ 0.001, CMIN/df = 1.378< 2.00, GFI= 0.914, TLI = 0.932, CFI = 0.928, RMSEA = 0.042 < 0.08. As these figures well satisfy the required criteria for CFA in terms of (1) unidimensionality, (2) scale reliability, (3) convergent validity, and (4) discriminant validity presented in Section 3.3, it can be concluded that the research model fits market data.

### 4.5. Structural Equation Modelling

*4.5.1. Model of Successful Kaizen Implementation.* Figure 3 briefly shows the analysis results of SEM model of the determinants of the successful Kaizen implementation of SMEs in Vietnam. The estimated standardized parameters of the saturated model, such as CMIN= 953.090, df= 674, p-value ≤ 0.001, CMIN/df = 1.414 < 2.00, GFI=0.916, TLI=0.933, CFI=0.939, RMSEA=0.044 < 0.08, well satisfy the required criteria for SEM as presented in Section 3.3; thus, the

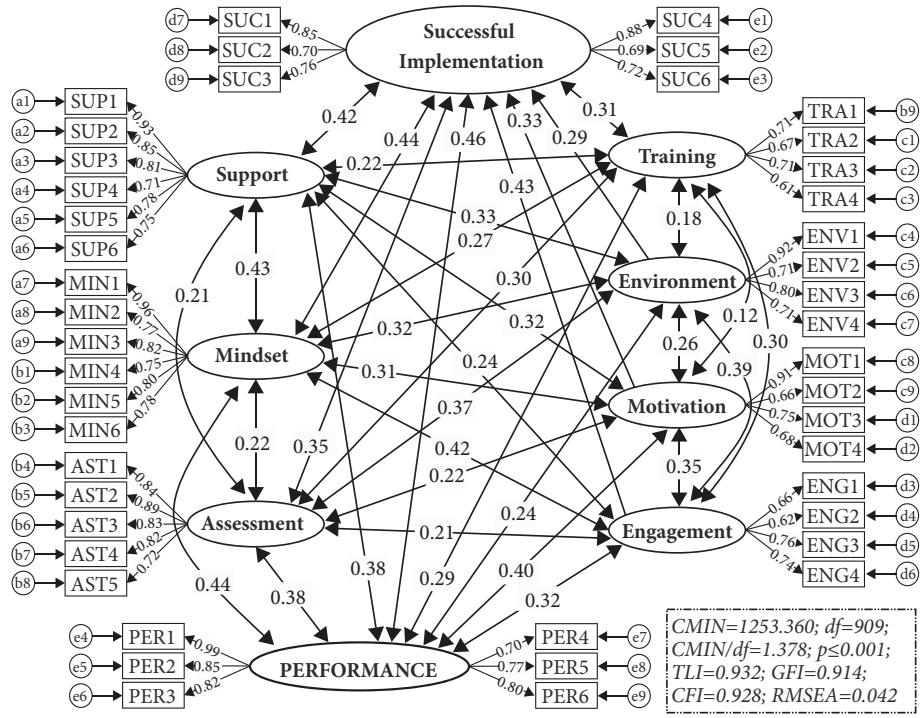



Figure 2: Confirmatory factor analysis.

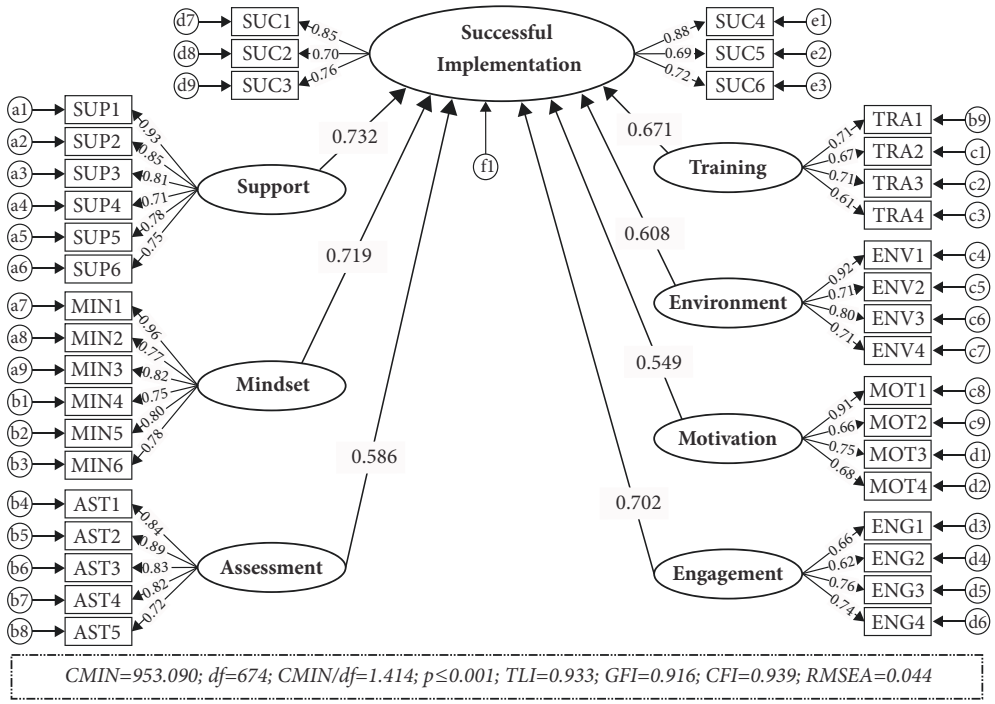



Figure 3: Standardized SEM model of successful Kaizen implementation.

proposed model is considered fit for the actual data. In addition, the bias of the model estimation obtained from bootstrapping 500 times was found insignificant. Therefore, it can be concluded that the estimates obtained in the model are reliable.

*4.5.2. Model of Sustainable Performance.* With the same token, Figure 4 displays the analysis results of the determinants of sustainable performance of SMEs in Vietnam. The estimated standardized parameters, such as CMIN= 1253.360, df= 909, p-value ≤ 0.001, CMIN/df = 1.378 < 2.00, GFI =

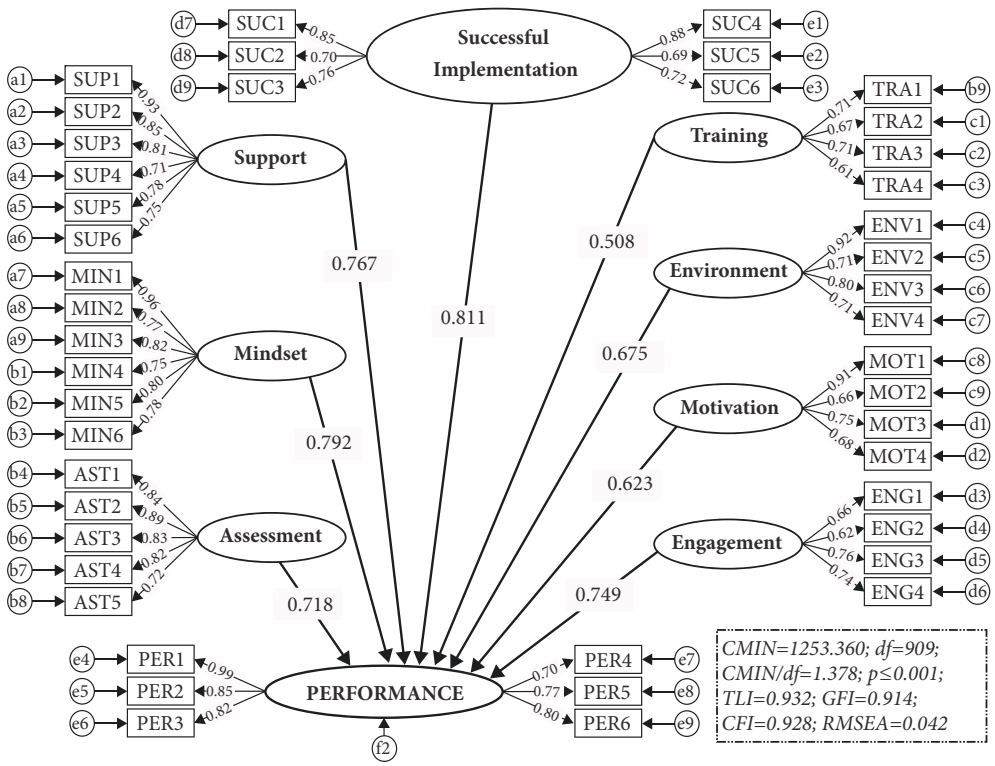



Figure 4: Standardized SEM model of sustainable performance of SMEs in Vietnam.

0.914, TLI = 0.932, CFI = 0.928, RMSEA = 0.042 < 0.08, well satisfy the required criteria for SEM as presented in Section 3.3; thus, the proposed model is considered fit for the actual data. Moreover, analysis results obtained from 500-time bootstrapping approach show that there is insignificant bias in the model estimation parameters, indicating that the obtained model estimates are reliable.

*4.6. Hypothesis Tests with SEM.* The results of the model estimation and bootstrapping in SEM shown in Table 7 clearly indicate that all of the proposed hypotheses (H1 ⟶ H15) are statistically supported as the p-values of related coefficients are less than 0.05.

*4.7. Tests of the Impacts of Demographic Characteristics.* This study used one-way ANOVA test to investigate the impacts of demographic characteristics such as location, size, ownership type of the enterprise, and the working position of the respondents on the evaluation of the two dependent factors, "successful Kaizen implementation" and "organizational performance". In order to achieve the objective, two new variables coded as "SUCC" and "PERF" were created by taking averages of the six components of each dependent factor, respectively.

Table 8 briefly presents the analysis results from tests of homogeneity of variances among the groups within each characteristic. With the given significance level of 5% used in this study, Table 8 clearly shows the different variances of SUCC and PERF among respondents' groups based on the ownership type and the enterprise location. In addition, the variances of PERF among respondents' groups based on the enterprise size are also different. The results in Table 8 provide important information to further test the equality of means of SUCC and PERF among the groups within each characteristic as shown in Table 9.

The figures in Table 9 clearly show that there are certain differences in the evaluation of SUCC and PERF among groups based on the working position, ownership type, and enterprise size. From the results in Table 8 and Table 9, post hoc tests were conducted to investigate which groups are different from others.

(1) In terms of working positions, Kaizen leaders and department managers have similar evaluations which are higher than those of directors/vice directors. It was found that Kaizen leaders and department managers are the ones directly involving in the Kaizen implementation and monitoring the improvement from the shop floors; thus, they tend to be satisfied with the success and the organizational performance. However, as directors and vice directors more concerned about the overall performance and general targets, they always expect to have better gains.

(2) In terms of size, it was found that medium enterprises have better success and higher performance than the micro and small ones because they usually pay more attention to the improvement of their operational effectiveness and efficiency to increase their competitive advantages.

Table 7: Coefficients from the SEM model.

| *Relationships* | *Coefficients* | *Std. Coefs.*[a] | *S.E.*[b] | *C.R.*[c] | *p-value* | *Conclusion* |
|---|---|---|---|---|---|---|
| SUC ⟵ SUP | 0.729 | 0.732 | 0.089 | 8.191 | * | H1 supported |
| SUC ⟵ MIN | 0.712 | 0.719 | 0.081 | 8.790 | * | H11 supported |
| SUC ⟵ ENG | 0.716 | 0.702 | 0.079 | 9.063 | * | H13 supported |
| SUC ⟵ TRA | 0.693 | 0.671 | 0.079 | 8.772 | * | H3 supported |
| SUC ⟵ ENV | 0.591 | 0.608 | 0.053 | 11.151 | * | H5 supported |
| SUC ⟵ AST | 0.578 | 0.586 | 0.085 | 6.800 | * | H7 supported |
| SUC ⟵ MOT | 0.557 | 0.549 | 0.072 | 7.736 | * | H9 supported |
| PER ⟵ SUC | 0.802 | 0.811 | 0.067 | 11.970 | * | H15 supported |
| PER ⟵ MIN | 0.785 | 0.792 | 0.081 | 9.691 | * | H12 supported |
| PER ⟵ SUP | 0.791 | 0.767 | 0.061 | 12.967 | * | H2 supported |
| PER ⟵ ENG | 0.751 | 0.749 | 0.079 | 9.506 | * | H14 supported |
| PER ⟵ AST | 0.722 | 0.718 | 0.076 | 9.500 | * | H8 supported |
| PER ⟵ ENV | 0.659 | 0.675 | 0.053 | 12.434 | * | H6 supported |
| PER ⟵ MOT | 0.642 | 0.623 | 0.071 | 9.042 | * | H10 supported |
| PER ⟵ TRA | 0.504 | 0.508 | 0.075 | 6.720 | * | H4 supported |

Notes: [a] standardized coefficients; [b] standard error; [c] critical ratio; * less than 0.1%.

Table 8: Tests of homogeneity of variances.

| *Characteristic* | *Factor* | *Levene Statistic* | *df1* | *df2* | *Sig.* |
|---|---|---|---|---|---|
| Ownership type | SUCC | 3.4894 | 2 | 210 | 0.032 |
| | PERF | 3.1752 | 2 | 210 | 0.044 |
| Enterprise location | SUCC | 3.9012 | 2 | 210 | 0.022 |
| | PERF | 3.2636 | 2 | 210 | 0.040 |
| Enterprise size | SUCC | 1.9781 | 2 | 210 | 0.141 |
| | PERF | 1.2796 | 2 | 210 | 0.280 |
| Working position | SUCC | 1.1278 | 2 | 210 | 0.326 |
| | PERF | 0.6910 | 2 | 210 | 0.502 |

(3) In terms of ownership types, it was found that there is no difference in the evaluations of SUCC and PERF between the state-owned enterprises and local private ones, and between the joint-venture enterprises and foreign-owned ones. However, the joint-venture and foreign-owned enterprises, especially Japan-based ones, were found more successful than others because they better recognize the importance of Kaizen in their business operations and invest more resources to implement it in practice.

(4) In terms of location, it was found that the location of enterprises fails to have significant impacts on the evaluations of SUCC and PERF. This indicates that once Kaizen is carefully understood and implemented, it would result in similar success and performance.

## 5. Discussions and Managerial Implications

*5.1. Discussions.* As shown in Table 7, all research hypotheses proposed in this study are statistically supported, meaning that the success of Kaizen implementation and the sustainable performance of SMEs in Vietnam are affected by several factors, including (1) supports from senior management; (2) training; (3) working environment; (4) assessment; (5) motivation; (6) mindset; and (7) engagement of all leaders and employees in the enterprises. Among them, the support from senior management ($\beta$=0.732) plays the most important role in the successful Kaizen implementation. This finding further agrees with those by Goodridge et al. [87], García et al. [81], Al-Najem et al. [88], Imai [47], Suárez-Barraza et al. [74], and Crute et al. [89]. Though the support is ranked as the 3rd important factor directly affecting the sustainable performance, it is also considered crucial because the successful Kaizen implementation has the strongest impact on their sustainable performance ($\beta$=0.811). Consequently, senior management should formulate and effectively articulate their supports in terms of commitments, statements, policies, plans, resources, or even direct involvement, etc. SMEs should consider this as their top prioritized factor because it works as the cornerstone for other factors and activities.

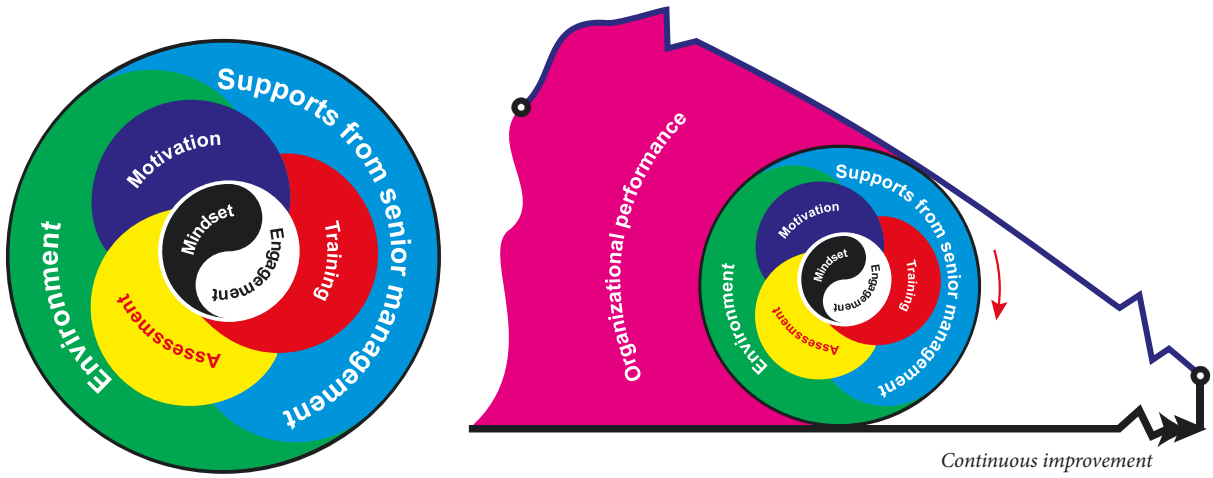
Table 9: ANOVA.

| *Characteristic* | *Factor* | | *Sum of Squares* | *df* | *Mean Square* | *F* | *Sig.* |
|---|---|---|---|---|---|---|---|
| Ownership type | SUCC | Between Groups | 2.159 | 2 | 1.080 | 3.797 | 0.024 |
| | | Within Groups | 59.707 | 210 | 0.284 | | |
| | | Total | 61.866 | 212 | | | |
| | PERF | Between Groups | 2.611 | 2 | 1.306 | 4.217 | 0.016 |
| | | Within Groups | 65.007 | 210 | 0.310 | | |
| | | Total | 67.618 | 212 | | | |
| Enterprise location | SUCC | Between Groups | 0.564 | 2 | 0.282 | 0.996 | 0.371 |
| | | Within Groups | 59.436 | 210 | 0.283 | | |
| | | Total | 60.000 | 212 | | | |
| | PERF | Between Groups | 0.828 | 2 | 0.414 | 1.344 | 0.263 |
| | | Within Groups | 64.751 | 210 | 0.308 | | |
| | | Total | 65.579 | 212 | | | |
| Enterprise size | SUCC | Between Groups | 2.310 | 2 | 1.155 | 4.096 | 0.018 |
| | | Within Groups | 59.152 | 210 | 0.282 | | |
| | | Total | 61.462 | 212 | | | |
| | PERF | Between Groups | 2.011 | 2 | 1.006 | 3.244 | 0.041 |
| | | Within Groups | 65.095 | 210 | 0.310 | | |
| | | Total | 67.106 | 212 | | | |
| Working position | SUCC | Between Groups | 1.992 | 2 | 0.996 | 3.532 | 0.031 |
| | | Within Groups | 59.148 | 210 | 0.282 | | |
| | | Total | 61.140 | 212 | | | |
| | PERF | Between Groups | 2.175 | 2 | 1.088 | 3.601 | 0.029 |
| | | Within Groups | 63.428 | 210 | 0.302 | | |
| | | Total | 65.603 | 212 | | | |

Moreover, mindset of all leaders and employees is ranked as the second important factor determining the success of Kaizen implementation and the sustainable performance of an enterprise, respectively, taking $\beta$=0.719 and 0.792. This finding further strengthens that of Thomas et al. [171] who claimed that employees' mindset is critical to organizational achievements and sustainability of their high performance because it greatly affects the productivity, innovation, and persistence of the workforce. Positive mindset should be translated into organizational practices to create a good culture for better performance [171] because the good culture helps to hoard habitual changes and support continuous improvement [48, 90]. Consequently, SMEs should have proper policies to foster and cultivate growth mindset in quality culture and continuous improvement practices; meanwhile fixed mindset should be gradually redirected and changed. However, changing the mindset of a person is always a difficult task in practice. Thus, this study proposes some typical implications to deal with it. It is noteworthy that mindset is a newly proposed factor discovered from the qualitative research; thus, it is considered as one of the key contributions of this study.

Along with the mindset, every member in an enterprise should actively and fully participate in the improvement process. Therefore, the engagement is ranked as the third significant factor affecting the success of Kaizen implementation ($\beta$=0.811) which is similar to the finding by Stadnicka & Sakano [112]. It is also ranked the fourth in affecting the sustainable performance ($\beta$=0.811), further agreeing with [182, 184–193]. Basically, the engagement from management levels can refer to their supports and commitments, whereas the engagement from employees refers to their participation in relevant activities with their responsibility.

In this study, among the seven independent factors, training is found as the fourth important factor affecting the successful Kaizen implementation in the SMEs in Vietnam. Its importance was also previously identified by [52, 74, 90, 91, 96]. As presented in Section 4.4, the training positively helps to change the mindset (r=0.27) and improve employee motivation (r=0.12) as well as employee engagement (r=0.30). Similar findings were found by Alvarado-Ramirez et al. [92]. However, the training has the lowest impact on the sustainable performance. This is explained by the fact that it has significant impacts on other factors such as mindset, engagement, motivation, and success of Kaizen implementation, while these factors have more direct relationships to the organizational performance. Therefore, in general, training also plays crucial role in improving the sustainable performance of the SMEs.




(a) STEAM-ME model

(b) STEAM-ME and Kaizen, sustainable performance

Figure 5: STEAM-ME model.

Besides, environment also has positive impacts on the successful Kaizen implementation and the performance of an enterprise. Specifically, its importance is ranked the fifth among the seven factors affecting the success ($\beta$=0.608) and the sixth among the eight factors affecting the performance ($\beta$=0.675). This finding is similar to those by [97, 104–108]. Consequently, creating a friendly working environment and a good culture of quality and continuous improvement is also crucial to be considered by the SMEs in Vietnam.

Practically, this study also finds that regular assessment of work ergonomics (employee productivity, efficiency, attitude, etc.) and working environment (vibrations, noise, internal air pollution, microclimate, radiation, dustiness or energy expenditure of the worker, etc.) has positive impacts on the success of Kaizen implementation and sustainable performance of SMEs because it can help to effectively trace the current progress and lead to reasonable actions to achieve organizational targets. This finding is further validated by Glover et al. [117]. An effective assessment also helps to improve organizational performance.

Lastly, organizations should have good policies and approaches to motivate their employees because the motivation is also a significant factor affecting the successful of Kaizen implementation ($\beta$=0.549) and the organizational performance ($\beta$=0.623). It is further supported by [63, 86, 144, 150, 151, 167–169].

In short, seven determinants of the successful Kaizen implementation and the sustainable performance of SMEs in Vietnam are (1) *S*upports from senior management; (2) *T*raining; (3) *E*nvironment; (4) *A*ssessment; (5) *M*otivation; (6) *M*indset; and (7) *E*ngagement. The first letters of these factors are orderly congregated as *"STEAM-ME"* which is considered as a novel model for the successful Kaizen implementation and the sustainable performance of SMEs in Vietnam. The name of the model also implies that an organization needs to have a new airflow with energy as "steam" to firstly make gradual changes to start its journey towards significant success in implementing Kaizen and sustaining organizational performance. The "steam" will make all of its members refreshed and brimful of energy to improve their minds, attitudes, behaviors, engagement, productivity, and responsibilities which will result in substantial increase in both personal and organizational performance.

Especially, Figure 5 visually presents the components of STEAM-ME model and their positive correlations as well as their impacts on the success of Kaizen implementation and organizational performance. Mindset and engagement are placed in the center of the model due to their critical roles as discussed above. Nonetheless, related activities in terms of motivation, training, and assessment taking place help to positively change the mindset and improve the engagement of all members in an organization whereas the supports from senior management and environment provide foundations for the activities.

With the strong correlations identified in Figure 2, no clear boundary exists among these factors as shown in Figure 5(a). They are all flexibly and continuously transformed from one state to others in a spiral endless-circle. Though the model looks like the traditional yin-yang circle, it only presents the mutual relationships and organic transformation among the factors; it does not mean "opposite" as of the yin-yang theory. In addition, the positive impacts of the identified factors on the successful Kaizen implementation and sustainable performance indicate that the more the factors are improved, the more success and the better performance an organization will have. Thus, if the STEAM-ME circle moves forwards, the organization will have better improvement and greater performance. This mechanism is demonstrated in Figure 5(b).

*5.2. Managerial Implications.* The existing literature clearly shows that successfully implementing Kaizen is a long and complex mission which should be integrated into strategic management instead of being considered as a particular project. The insights of the mutual relationships among the seven affecting factors proposed in the novel STEAM-ME

model greatly help business organizations, especially SMEs, to create proper strategies for their continuous improvement and sustainable performance.

Firstly, to effectively cultivate growth mindsets within the organizations, top executives and department managers should be the first ones to refresh their mindsets by taking Kaizen training workshops so that they fully capture the Kaizen philosophy as well as potential benefits they will gain once Kaizen is successfully implemented. This is really important to start the first cycle because such new mindsets not only urge them to set and patiently pursuit Kaizen as a strategic goal but also make them willing to provide sufficient supports and create good environment for their employees. After that, they should either send more staffs to join similar workshops or organize some internal training by either Kaizen experts or the trained executives/managers because the staffs will be the ones directly participating in the continuous improvement process. With encouraging and open environment, they can quickly employ the knowledge and experiences learnt from the training; hence, we can observe immediate improvements. From such training, all members will shape their own Kaizen mindsets which drive them to (1) consider continuous improvement as a permanent need in every daily operation; (2) always welcome suggestions for improvement; (3) always strive for better productivity and quality because there are several areas for improvement; (4) appreciate teamwork and constructive contributions; and (5) always consider "sustainability" in every solutions or activities for long-term achievements. Such Kaizen mindsets will steadily transform into organizational culture of continuous improvement and sustainable development.

Secondly, with the positive mindsets, they will actively engage in improvement processes, and more innovative solutions for improvement will be proposed. Therefore, the SMEs should have right motivation approaches to encourage their engagement and increase their overall performance.

Thirdly, SMEs should have proper tools and measures to incessantly monitor and assess their actual performance and benchmark with their expected outcomes to take corrective actions if needed. Importantly, the tools and measures should incorporate three critical pillars for sustainable performance: people, planet, and profit.

Finally, the findings in Section 4.7 urge the state-owned enterprises and the private ones to pay more attention to the understanding and implementing of Kaizen philosophy in their business operations. They should send more senior leaders/staffs to Kaizen training workshops to fully capture the philosophy and learn the practical experiences from the sharing of their peers. This is really important to improve their competitive advantages against the joint-venture and foreign-owned enterprises to assure their sustainable development in the current trend of regional and international integration. Practically, joint-venture and foreign-owned enterprises tend to implement Kaizen easier because they have better management system with stronger quality culture. Moreover, the micro and small enterprises should also make more efforts to implement Kaizen to improve their performance and their productivity before they can enlarge their business.

## 6. Conclusion

Over the past few decades, Kaizen has been successfully implemented across different industries in many countries worldwide and brought significant benefits towards relevant organizations, including SMEs. SMEs in Vietnam play an important role in developing the national economy. However, the recent trend in international integration urges them to improve their competitive advantages for their survival and sustainable growth. Therefore, this study is aimed at identifying determinants of the successful Kaizen implementation and sustainable performance of SMEs in Vietnam so that others can have proper actions and prioritize their operations in accordance with their available resources. Specifically, through a formal survey of 213 participants from 62 SMEs successfully implementing Kaizen in the North, Middle, and South of Vietnam and appropriate statistical approaches such as exploratory factor analysis (EFA), scale reliability analysis, confirmatory factor analysis (CFA), and structural equation modelling (SEM), seven important determinants have been identified: (1) supports from senior management; (2) training; (3) working environment; (4) assessment; (5) motivation; (6) mindset; and (7) engagement of all members in the enterprises. These seven factors perfectly form a new model named as *"STEAM-ME"*, implying that organizations need to have a new airflow as "steam" to make all of its members refreshed and brimful of energy to foster their growth minds, positive attitudes, behaviors, engagement, productivity, and responsibilities and improve their performance so that the organizations can (1) gain significant success in implementing Kaizen and (2) improve their business performance and competitive advantage for their sustainable development.

In particular, among the seven identified factors, "mindset" is newly proposed in this study. It was identified from the qualitative research and has significant impacts on the success of Kaizen implementation and sustainable performance. The finding obviously adds a new affecting factor to fulfill research gap in the existing literature. In addition, the quantitative relationships among the identified factors help to create an innovative STEAM-ME model whose components positively and crucially affect the successful Kaizen implementation and sustainable performance of SMEs in Vietnam.

As this study focuses on SMEs only, future research should investigate if similar determinants exist in the cases of large enterprises and multinational corporations. Comparative analysis of the success and organizational performance among enterprises of all sizes will deepen our understanding of how Kaizen can be successfully implemented across the enterprise sizes.

## Data Availability

The data used to support the findings of this study are available from previously reported studies and datasets, which have been cited. In addition, the official survey and the data will be supplemented by the author upon request.

## Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This study is funded by Lac Hong University under the Decision No. 879/QĐ-ĐHLH dated October 24, 2018, by the Rector.

## Supplementary Materials

Appendix I provides a full list of references supporting the rational validation of the six identified factors presented in the main text while Appendix II provides a table mapping each factor with its reference sources. *(Supplementary Materials)*

## References

[1] A. K. Arya and S. Choudhary, "Assessing the application of Kaizen principles in Indian small-scale industry," *International Journal of Lean Six Sigma*, vol. 6, no. 4, pp. 369–396, 2015.

[2] H. Iberahim, H. Mazlinda, M. Marhainie, and A. N. Hidayah, "Determinants of sustainable continuous improvement practices in mail processing service operations," *Procedia - Social and Behavioral Sciences*, vol. 219, pp. 330–337, 2016.

[3] B. Kamińska, "Kaizen as a method of management improvement in small production companies," *Entrepreneurship and Management*, vol. 16, no. 2, pp. 157–170, 2015.

[4] M. Oropesa Vento, J. L. García Alcaraz, A. A. Maldonado Macías, and V. Martínez Loya, "The impact of managerial commitment and Kaizen benefits on companies," *Journal of Manufacturing Technology Management*, vol. 27, no. 5, pp. 692–712, 2016.

[5] C. Topuz and Z. Arasan, "Kaizen-educational: An awareness-raising and motivational-enhancement group counseling model," *Procedia - Social and Behavioral Sciences*, vol. 84, pp. 1356–1360, 2013.

[6] D. J. Teece, "Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance," *Strategic Management Journal*, vol. 28, no. 13, pp. 1319–1350, 2007.

[7] W. G. Macpherson, J. C. Lockhart, H. Kavan, and A. L. Iaquinto, "Kaizen: a Japanese philosophy and system for business excellence," *Journal of Business Strategy*, vol. 36, no. 5, pp. 3–9, 2015.

[8] R. Lozano, M. Suzuki, A. Carpenter, and O. Tyunina, "An analysis of the contribution of Japanese business terms to corporate sustainability: learnings from the "looking-glass" of the east," *Sustainability*, vol. 9, no. 2, article no 188, 2017.

[9] T. Homma, "JICA's industrial cooperation in africa," in *Proceedings of the GRIPS Development Forum International Seminar on African Manufacturing*, Tokyo, 2014.

[10] L. B. M. Costa and M. G. Filho, "Lean healthcare: Review, classification and analysis of literature," *Production Planning & Control*, vol. 27, no. 10, pp. 823–836, 2016.

[11] S. Duarte and V. Cruz-Machado, "Modelling lean and green: a review from business models," *International Journal of Lean Six Sigma*, vol. 4, no. 3, pp. 228–250, 2013.

[12] A. Chiarini, "Sustainable manufacturing-greening processes using specific lean production tools: An empirical observation from european motorcycle component manufacturers," *Journal of Cleaner Production*, vol. 85, no. 4, pp. 226–233, 2014.

[13] J. A. Garza-Reyes, "Lean and green-a systematic review of the state of the art literature," *Journal of Cleaner Production*, vol. 102, no. 8, pp. 18–29, 2015.

[14] V. Chahal, N. Grover, N. Kumar, and M. T. Pardeep, "Impact of lean strategies on different industrial lean wastes," *International Journal of Theoretical and Applied Mechanics*, vol. 12, no. 2, pp. 275–286, 2017.

[15] G. A. Marodin, A. G. Frank, G. L. Tortorella, and D. C. Fetterman, "Lean production and operational performance in the Brazilian automotive supply chain," *Total Quality Management & Business Excellence*, vol. 30, no. 3-4, pp. 370–385, 2017.

[16] S. Gupta, M. Sharma, and V. Sunder M, "Lean services: a systematic review," *International Journal of Productivity and Performance Management*, vol. 65, no. 8, pp. 1025–1056, 2016.

[17] I. Belekoukias, J. A. Garza-Reyes, and V. Kumar, "The impact of lean methods and tools on the operational performance of manufacturing organisations," *International Journal of Production Research*, vol. 52, no. 18, pp. 5346–5366, 2014.

[18] R. R. Fullerton, F. A. Kennedy, and S. K. Widener, "Lean manufacturing and firm performance: The incremental contribution of lean management accounting practices," *Journal of Operations Management*, vol. 32, no. 7-8, pp. 414–428, 2014.

[19] P. Ingelsson and A. Mårtensson, "Measuring the importance and practices of Lean values," *TQM Journal*, vol. 26, no. 5, pp. 463–474, 2014.

[20] A. Prashar, "Redesigning an assembly line through Lean-Kaizen: An Indian case," *TQM Journal*, vol. 26, no. 5, pp. 475–498, 2014.

[21] R. Teehan and W. Tucker, "Service quality Kaizen blitz: The road to improving customer satisfaction," *Sinergie Italian Journal of Management*, vol. 94, no. 1, pp. 233–241, 2014.

[22] M. Dora, M. Kumar, D. Van Goubergen, A. Molnar, and X. Gellynck, "Operational performance and critical success factors of lean manufacturing in European food processing SMEs," *Trends in Food Science & Technology*, vol. 31, no. 2, pp. 156–164, 2013.

[23] AFED - Agency for Enterprise Development, "White paper - Small and medium enterprises in vietnam, ministry of planning and investment," 2017, http://business.gov.vn/Portals/0/2018/ST20DNNVV202017_final1.pdf.

[24] VGP- Vietnam Government Portal, "Doanh nghiệp Việt Nam càng ngày càng nhỏ đi?" 2018, http://baochinhphu.vn/Kinh-te/Doanh-nghiep-Viet-Nam-cang-ngay-cang-nho-di/328552.vgp.

[25] N. D. Minh, D. T. Cuc, T. T. H. Giang, and H. T. T. Ha, "Application of 5S in Vietnam small and medium manufacturing enterprises current situation and recommendations," *Journal of Science of Vietnam National University*, vol. 29, no. 1, pp. 23–31, 2013.

[26] A. F. Lemma, "The role of Kaizen in economic transformation, working paper 523, overseas development institute," 2018, http://www.odi.org/sites/odi.org.uk/files/resource-documents/12110.pdf.

[27] Sebhatu S. P., "The challenges and opportunities in creating sustainable shared values at the base of the Pyramid- Cases from sub-Saharan Africa," in *Sustainability Challenges and Solutions at the Base-of-the-Pyramid: Business, Technology and the Poor*,

P. Kandachar and M. Halme, Eds., pp. 146–162, Green Leaf Publishing, Sheffield, UK, 2017.

[28] A. N. Norazlan, N. F. Habidin, M. H. Roslan, and M. Z. Zainudin, "The development of sustainable supply chain management and sustainable performance in Malaysian healthcare industry," *International Journal of Ethics in Engineering and Management Education*, vol. 1, no. 2, pp. 51–55, 2014.

[29] T. Artiach, D. Lee, D. Nelson, and J. Walker, "The determinants of corporate sustainability performance," *Accounting & Finance*, vol. 50, no. 1, pp. 31–51, 2010.

[30] A. Stanciu, M. Constandache, and E. Condrea, "Concerns about the sustainable performance of firm in the context of quality management systems implementation," *Procedia - Social and Behavioral Sciences*, vol. 131, pp. 340–344, 2014.

[31] UBS, "Achieving sustainable performance- Integrated Reporting 2017," 2017, http://www.ubs.com/global/en/about_.../integrated-report-2017-en.pdf.

[32] J. É. Corrêa, J. B. Turrioni, A. P. D. Paiva et al., "The influence of accreditation on the sustainability of organizations with the Brazilian accreditation methodology," *Journal of Healthcare Engineering*, vol. 2018, Article ID 1393585, 11 pages, 2018.

[33] Q. Feng, X. Liu, L. Tang, L. Shi, J. Jiang, and X. Su, "Research on a connotation and assessment index system of eco-communities," *International Journal of Sustainable Development & World Ecology*, vol. 24, no. 6, pp. 524–531, 2017.

[34] M. Yang, M. Movahedipour, J. Zeng, Z. Xiaoguang, and L. Wang, "Analysis of success factors to implement sustainable supply chain management using interpretive structural modeling technique: A real case perspective," in *Mathematical Problems in Engineering*, vol. 2017, p. 14, 2017.

[35] L. Shen, C. Shuai, L. Jiao, Y. Tan, and X. Song, "A global perspective on the sustainable performance of urbanization," *Sustainability*, vol. 8, no. 8, article no 783, 2016.

[36] S. K. Chaharsooghi and M. Ashrafi, "Sustainable supplier performance evaluation and selection with Neofuzzy TOPSIS Method," *International Scholarly Research Notices*, vol. 2014, Article ID 434168, 10 pages, 2014.

[37] S. M. Masoumik, S. H. Abdul-Rashid, E. U. Olugu, and R. A. Raja Ghazilla, "Sustainable supply chain design: A configurational approach," *The Scientific World Journal*, vol. 2014, Article ID 897121, 16 pages, 2014.

[38] W. C. Huang, C. H. Jhong, and J. F. Ding, "Key factors influencing sustainable development of a green energy industry in Taiwan," in *Mathematical Problems in Engineering*, vol. 2013, p. 10, 2013.

[39] N. Long and T. Nguyen, "Sustainable development of rural tourism in an Giang Province, Vietnam," *Sustainability*, vol. 10, no. 4, article no 953, 2018.

[40] A. N. Norazlan, N. F. Habidin, M. H. Roslan, and M. Z. Zainudin, "Investigation of kaizen blitz and sustainable performance for Malaysian healthcare industry," *International Journal of Quality and Innovation*, vol. 2, no. 3/4, p. 272, 2014.

[41] B. Moldan, S. Janoušková, and T. Hák, "How to understand and measure environmental sustainability: Indicators and targets," *Ecological Indicators*, vol. 17, pp. 4–13, 2012.

[42] T. Schoenherr, "The role of environmental management in sustainable business development: a multicounty investigation," *International Journal Production Economics*, vol. 140, no. 1, pp. 116–128, 2011.

[43] T. Q. Nguyen, N. T. Long, and T. Nguyen, "Impacts of corporate social responsibility on the competitiveness of tourist enterprises," *Tourism Economics*, 2018.

[44] S. Iwao, "Revisiting the existing notion of continuous improvement (Kaizen): literature review and field research of Toyota from a perspective of innovation," *Evolutionary and Institutional Economics Review*, vol. 14, no. 1, pp. 29–59, 2017.

[45] J. Miller, M. Wroblewski, and J. Villafuerte, *Creating a Kaizen Culture*, McGraw Hill, NY, USA, 2014.

[46] D. Carnerud, C. Jaca, and I. Bäckström, "Kaizen and continuous improvement – trends and patterns over 30 years," *The TQM Journal*, vol. 30, no. 4, pp. 371–390, 2018.

[47] M. Imai, *Gemba Kaizen: A Common Sense Approach to a Continuous Improvement Strategy*, McGraw-Hill Education, New York, NY, USA, 2nd edition, 2012.

[48] J. Singh and H. Singh, "Continuous improvement philosophy – literature review and directions," *Benchmarking: An International Journal*, vol. 22, no. 1, pp. 75–119, 2015.

[49] S. Isenberg, "Merging education and business models to create and sustain transformational change," *International Journal of Adult Vocational Education and Technology*, vol. 1, no. 4, pp. 31–47, 2010.

[50] A. Styhre, "Kaizen, ethics, and care of the operations: management after empowerment," *Journal of Management Studies*, vol. 38, no. 6, pp. 795–810, 2001.

[51] J. A. Farris, E. M. Van Aken, T. L. Doolen, and J. Worley, "Critical success factors for human resource outcomes in Kaizen events: An empirical study," *International Journal of Production Economics*, vol. 117, no. 1, pp. 42–65, 2009.

[52] J. Ma, Z. Lin, and C. K. Lau, "Prioritising the enablers for the successful implementation of Kaizen in China," *International Journal of Quality & Reliability Management*, vol. 34, no. 4, pp. 549–568, 2017.

[53] M. F. Suárez-Barraza and J. Ramis-Pujol, "Implementation of Lean-Kaizen in the human resource service process: A case study in a Mexican public service organisation," *Journal of Manufacturing Technology Management*, vol. 21, no. 3, pp. 388–410, 2010.

[54] D. Jurburg, E. Viles, M. Tanco, and R. Mateo, "What motivates employees to participate in continuous improvement activities?" *Total Quality Management & Business Excellence*, vol. 28, no. 13-14, pp. 1469–1488, 2017.

[55] N. Rodríguez-Padial, M. Marín, and R. Domingo, "An approach to integrating tactical decision-making in industrial maintenance balance scorecards using principal components analysis and machine learning," *Complexity*, vol. 2017, Article ID 3759514, 15 pages, 2017.

[56] P. Alexander and J. B. Fadden, "A value-stream mapping success story: mba recruiting process improvements," in *Proceedings of the 4th International Conference on Lean Six Sigma for Higher Education*, pp. 40–49, 2017.

[57] B. K. Jeong and T. E. Yoon, "Improving IT process management through value stream mapping approach: A case study," *Journal of Information Systems and Technology Management*, vol. 13, no. 3, pp. 389–404, 2016.

[58] F. E. Ciarapica, M. Bevilacqua, and G. Mazzuto, "Performance analysis of new product development projects," *International Journal of Productivity and Performance Management*, vol. 65, no. 2, pp. 177–206, 2016.

[59] A. Kuiper, R. van de Hoef, M. Wesseling, B. A. Lameijer, and R. J. Does, "Quality quandaries: Improving a customer value stream at a financial service provider," *Quality Engineering*, vol. 28, no. 1, pp. 155–163, 2016.

[60] M. A. Lewis, "Lean production and sustainable competitive advantage," *International Journal of Operations and Production Management*, vol. 20, no. 8, pp. 959–978, 2000.

[61] M. A. Idris and M. Zairi, "Sustaining TQM: A synthesis of literature and proposed research framework," *Total Quality Management & Business Excellence*, vol. 17, no. 9, pp. 1245–1260, 2006.

[62] J. Pullin, "Room for improvement," *Professional Engineering*, vol. 18, no. 15, pp. 38–138, 2005.

[63] D. I. Prajogo and A. S. Sohal, "The sustainability and evolution of quality improvement programmes - An Australian case study," *Total Quality Management & Business Excellence*, vol. 15, no. 2, pp. 205–220, 2004.

[64] N. Bateman and N. Rich, "Companies perceptions of inhibitors and enablers for process improvement activities," *International Journal of Operations & Production Management*, vol. 23, no. 2, pp. 185–199, 2003.

[65] J. J. Garcia-Sabater and J. A. Marin-Garcia, "Can we still talk about continuous improvement? Rethinking enablers and inhibitors for successful implementation," *International Journal of Technology Management*, vol. 55, no. 1-2, pp. 28–42, 2011.

[66] A. G. Robinson and D. M. Schroeder, *Ideas Are Free: How The Idea Revolution Is Liberating People and Transforming Organizations*, Berrett-Koehler Publishers, USA, 2004.

[67] S. Nakajima, *Introduction to TPM Total Productive Maintenance*, Massachusetts Productivity Press, Cambridge, Mass, USA, 1988.

[68] R. Domingo and S. Aguado, "Overall environmental equipment effectiveness as a metric of a lean and green manufacturing system," *Sustainability*, vol. 7, no. 7, pp. 9031–9047, 2015.

[69] S. Kumar, A. K. Dhingra, and B. Singh, "Kaizen selection for continuous improvement through VSM-FUZZY-TOPSIS in small-scale enterprises: An Indian case study," in *Advances in Fuzzy Systems*, vol. 2018, p. 10, 2018.

[70] J. L. García, A. A. Maldonado, A. Alvarado, and D. G. Rivera, "Human critical success factors for kaizen and its impacts in industrial performance," *The International Journal of Advanced Manufacturing Technology*, vol. 70, no. 9-12, pp. 2187–2198, 2014.

[71] Y. F. Chen and D. Tjosvold, "Participative leadership by American and Chinese managers in China: The role of relationships," *Journal of Management Studies*, vol. 43, no. 8, pp. 1727–1752, 2006.

[72] J. Mendoza-Fong, J. García-Alcaraz, J. Díaz-Reza, J. Sáenz Diez Muro, and J. Blanco Fernández, "The role of green and traditional supplier attributes on business performance," *Sustainability*, vol. 9, no. 9, article no 1520, 2017.

[73] M. E. Pullman, M. J. Maloni, and C. R. Carter, "Food for thought: Social versus environmental sustainability practices and performance outcomes," *Journal of Supply Chain Management*, vol. 45, no. 4, pp. 38–54, 2009.

[74] M. F. Suárez-Barraza, J. Ramis-Pujol, and L. Kerbache, "Thoughts on kaizen and its evolution: Three different perspectives and guiding principles," *International Journal of Lean Six Sigma*, vol. 2, no. 4, pp. 288–308, 2011.

[75] J. Womack, D. Jones, and D. Roos, *The Machine That Changed the World Published*, Simon & Schuster, New York, NY, USA, 2007.

[76] A. Hiam, *Motivational Management: Inspiring Your People for Maximum Performance*, American Management Association, New York, NY, USA, 2003.

[77] M. G. Maarof and F. Mahmud, "A review of contributing factors and challenges in implementing kaizen in small and medium enterprises," *Procedia Economics and Finance*, vol. 35, pp. 522–531, 2016.

[78] J. L. García-Alcaraz, M. Oropesa-Vento, and A. A. MMaldonado-Macías, "Literature review," in *Kaizen Planning, Implementing and Controlling*, Management and Industrial Engineering, pp. 23–31, Springer International Publishing, 2017.

[79] J. L. García-Alcaraz, M. Oropesa-Vento, and A. A. Maldonado-Macías, "Methodology," in *Kaizen Planning, Implementing and Controlling*, Management and Industrial Engineering, pp. 59–78, Springer International Publishing, 2017.

[80] L. Avelar-Sosa, J. García-Alcaraz, and J. Castrellón-Torres, "The effects of some risk factors in the supply chains performance: A case of study," *Journal of Applied Research and Technology*, vol. 12, no. 5, pp. 958–968, 2014.

[81] J. L. García, D. G. Rivera, and A. A. Iniesta, "Critical success factors for Kaizen implementation in manufacturing industries in Mexico," *The International Journal of Advanced Manufacturing Technology*, vol. 68, no. 1-4, pp. 537–545, 2013.

[82] M. Oropesa-Vento, J. L. García-Alcaraz, L. Rivera, and D. F. Manotas, "Effects of management commitment and organization of work teams on the benefits of Kaizen: Planning stage," *DYNA*, vol. 82, no. 191, pp. 76–84, 2015.

[83] J. Díaz-Reza, J. García-Alcaraz, L. Avelar-Sosa, J. Mendoza-Fong, J. Sáenz Diez-Muro, and J. Blanco-Fernández, "The role of managerial commitment and TPM implementation strategies in productivity benefits," *Applied Sciences*, vol. 8, no. 7, article no 1153, 2018.

[84] N. Bateman, "Sustainability: The elusive element of process improvement," *International Journal of Operations and Production Management*, vol. 25, no. 3, pp. 261–276, 2005.

[85] R. Cooney and A. Sohal, "Teamwork and total quality management: A durable partnership," *Total Quality Management & Business Excellence*, vol. 15, no. 8, pp. 1131–1142, 2010.

[86] C. Rapp and J. Eklund, "Sustainable development of improvement activities–the long-term operation of a suggestion scheme in a Swedish company," *Total Quality Management*, vol. 13, no. 7, pp. 945–969, 2010.

[87] D. Goodridge, G. Westhorp, T. Rotter, R. Dobson, and B. Bath, "Lean and leadership practices: development of an initial realist program theory," *BMC Health Services Research*, vol. 15, no. 1, 2015.

[88] M. Al-Najem, H. Dhakal, and N. Bennett, "The role of culture and leadership in lean transformation: A review and assessment model," *International Journal of Lean Thinking*, vol. 3, no. 1, pp. 119–138, 2012.

[89] V. Crute, Y. Ward, S. Brown, and A. Graves, "Implementing Lean in aerospace - Challenging the assumptions and understanding the challenges," *Technovation*, vol. 23, no. 12, pp. 917–928, 2003.

[90] K. J. Fryer, J. Antony, and A. Douglas, "Critical success factors of continuous improvement in the public sector: A literature review and some key findings," *The TQM Magazine*, vol. 19, no. 5, pp. 497–517, 2007.

[91] A. Trostel and A. Light, "Carrier Mexico S.A. De C.V.," *Journal of Business Research*, vol. 50, no. 1, pp. 97–110, 2000.

[92] K. M. Alvarado-Ramírez, V. H. Pumisacho-Álvaro, J. Á. Miguel-Davila, and M. F. Suárez Barraza, "Kaizen, a continuous improvement practice in organizations," *The TQM Journal*, vol. 30, no. 4, pp. 255–268, 2018.

[93] C. Soltero and G. Waldrip, "Using Kaizen to reduce waste and prevent pollution," *Environmental Quality Management*, vol. 11, no. 3, pp. 23–38, 2002.

[94] U. Kumar, V. Kumar, D. de Grosbois, and F. Choisne, "Continuous improvement of performance measurement by TQM adopters," *Total Quality Management & Business Excellence*, vol. 20, no. 6, pp. 603–616, 2009.

[95] 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.

[96] K. Ariga, M. Kurosawa, F. Ohtake, M. Sasaki, and S. Yamane, "Organization adjustments, job training and productivity: Evidence from Japanese automobile makers," *Journal of the Japanese and International Economies*, vol. 27, no. 1, pp. 1–34, 2013.

[97] A. Day and K. D. Randell, "Building a foundation for physically healthy workplaces and well-being," in *Workplace Well-Being: How to Build Psychologically Healthy Workplaces*, A. Day, E. K. Kelloway, and J. J. Hurrell, Eds., pp. 3–26, John Wiley & Sons, Ltd., Chichester, 2014.

[98] I. Beltrán-Martín and J. C. Bou-Llusar, "Examining the intermediate role of employee abilities, motivation and opportunities to participate in the relationship between HR bundles and employee performance," *BRQ Business Research Quarterly*, vol. 21, no. 2, pp. 99–110, 2018.

[99] A. M. Sharma and A. Shirsath, "Training –A motivational tool," *IOSR Journal of Business and Management*, vol. 16, no. 3, pp. 27–35, 2014.

[100] T. P. Sung, G. C. S. Yee, A. Bahron, and I. H. A. Rahim, "The influence of training , employee engagement and performance appraisal on turnover intention among lecturers in Sabah private higher education institutions," *Journal of Global Business and Social Entrepreneurship (GBSE)*, vol. 1, no. 3, pp. 89–98, 2017.

[101] F. A. Malik and Y. Rubina, "Role of human resource practices on employee performance: Mediating role of employee engagement," *Science International*, vol. 27, no. 6, pp. 6403–6412, 2015.

[102] A. J., "Determinants of employee engagement and their impact on employee performance," *International Journal of Productivity and Performance Management*, vol. 63, no. 3, pp. 308–323, 2014.

[103] A. Paradise, "Influences engagement," *ASTD Training Development*, vol. 62, no. 1, pp. 54–59, 2008.

[104] A. Realyvásquez, A. A. Maldonado-Macías, J. García-Alcaraz, G. Cortés-Robles, and J. Blanco-Fernández, "Structural model for the effects of environmental elements on the psychological characteristics and performance of the employees of manufacturing systems," *International Journal of Environmental Research and Public Health*, vol. 13, no. 1, article no. 104, 2016.

[105] M. A. Quddus and A. M. M. Nazmul Ahsan, "A shop-floor kaizen breakthrough approach to improve working environment and productivity of a sewing floor in RMG industry," *Journal of Textile and Apparel, Technology and Management*, vol. 8, no. 4, pp. 1–12, 2014.

[106] A. Skalli, I. Theodossiou, and E. Vasileiou, "Jobs as Lancaster goods: Facets of job satisfaction and overall job satisfaction," *Journal of Socio-Economics*, vol. 37, no. 5, pp. 1906–1920, 2008.

[107] S. Gazioglu and A. Tansel, "Job satisfaction in Britain: Individual and job related factors," *Applied Economics*, vol. 38, no. 10, pp. 1163–1171, 2006.

[108] A. Sousa-Poza and A. A. Sousa-Poza, "Well-being at work: A cross-national analysis of the levels and determinants of job satisfaction," *Journal of Socio-Economics*, vol. 29, no. 6, pp. 517–538, 2000.

[109] H. Zareh, M. Golverdi, A. H. S. Nasab, and A. A. Rashid, "Engagement at work: Approaches, benefits and guidelines, applied mathematics in engineering," *Management and Technology*, vol. 2, no. 4, pp. 83–92, 2014.

[110] J. Liker and J. Franz, "The Toyota way: Helping others help themselves," *Manufacturing Engineering*, vol. 149, no. 5, pp. 87–95, 2012.

[111] S. Aguado, R. Alvarez, and R. Domingo, "Model of efficient and sustainable improvements in a lean production system through processes of environmental innovation," *Journal of Cleaner Production*, vol. 47, pp. 141–148, 2013.

[112] D. Stadnicka and K. Sakano, "Employees motivation and openness for continuous improvement: Comparative study in polish and japanese companies," *Management and Production Engineering Review*, vol. 8, no. 3, pp. 70–86, 2017.

[113] A. Gravells, *Principles and Practices of Teaching and Training: A Guide for Teachers and Trainers in The FE and Skills Sector*, Learning Matters, Exeter, UK, 2017.

[114] T. Ferdous and B. Razzak, "Importance of Training needs assessment in the banking sector of Bangladesh: A case study on national bank limited (nbl)," *International Journal of Business and Management*, vol. 7, no. 10, pp. 63–73, 2012.

[115] J. Carlisle, R. Bhanugopan, and A. Fish, "Training needs of nurses in public hospitals in Australia: Review of current practices and future research agenda," *Journal of European Industrial Training*, vol. 35, no. 7, pp. 687–701, 2011.

[116] A. N. Abdelhafiz Elbadri, "Training practices of Polish companies: An appraisal and agenda for improvement," *Journal of European Industrial Training*, vol. 25, no. 2, pp. 69–79, 2001.

[117] W. J. Glover, J. A. Farris, E. M. Van Aken, and T. L. Doolen, "Critical success factors for the sustainability of Kaizen event human resource outcomes: An empirical study," *International Journal of Production Economics*, vol. 132, no. 2, pp. 197–213, 2011.

[118] J. L. Arquero, C. Fernández-Polvillo, T. Hassall, and J. Joyce, "Vocation, motivation and approaches to learning: a comparative study," *Education + Training*, vol. 57, no. 1, pp. 13–30, 2015.

[119] C. Stringer, J. Didham, and P. Theivananthampillai, "Motivation, pay satisfaction, and job satisfaction of front-line employees," *Qualitative Research in Accounting & Management*, vol. 8, no. 2, pp. 161–179, 2011.

[120] D. Conrad, A. Ghosh, and M. Isaacson, "Employee motivation factors," *International Journal of Public Leadership*, vol. 11, no. 2, pp. 92–106, 2015.

[121] S. Organ, D. Proverbs, and G. Squires, "Motivations for energy efficiency refurbishment in owner-occupied housing," *Structural Survey*, vol. 31, no. 2, pp. 101–120, 2013.

[122] A. Keshwar Seebaluck and T. Devi Seegum, "Motivation among public primary school teachers in Mauritius," *International Journal of Educational Management*, vol. 27, no. 4, pp. 446–464, 2013.

[123] M. Mozes, Z. Josman, and E. Yaniv, "Corporate social responsibility organizational identification and motivation," *Social Responsibility Journal*, vol. 7, no. 2, pp. 310–325, 2011.

[124] A. Furnham, A. Eracleous, and T. Chamorro-Premuzic, "Personality, motivation and job satisfaction: Hertzberg meets the Big Five," *Journal of Managerial Psychology*, vol. 24, no. 8, pp. 765–779, 2009.

[125] A. Ismail and M. R. Abd Razak, "A study on job satisfaction as a determinant of job motivation," *Acta Universitatis Danabius*, vol. 12, pp. 30–44, 2016.

[126] A. Tella, C. O. Ayeni, and S. O. Popoola, "Work motivation, job satisfaction, and organisational commitment of library personnel in academic and research libraries in Oyo State, Nigeria," *Library Philosophy and Practice*, vol. 2007, no. 118, pp. 1–16, 2007.

[127] B. A. Hennessey and T. M. Amabile, "Extrinsic and intrinsic motivation," in *Organizational Behavior*, N. Nicholson, P. Audia, and M. Pillutla, Eds., Blackwell Publishing, Malden, Mass, USA, 2005.

[128] A. Nelson and G. Quick, "The effects of contingent and non-contingent rewards and controls on intrinsic motivation," *Organizational Behavior & Human Performance*, vol. 8, no. 2, pp. 217–229, 2005.

[129] R. Yasothai, J. Jauhar, and A. G. Bashawir, "A study on the impact of employee performance: The mediating role of appraisal," *International Journal of Humanities and Social Science*, vol. 3, no. 1, pp. 92–104, 2015.

[130] O. P. Salau, H. O. Falola, and J. O. Akinbode, "Induction and staff attitude towards retention and organizational effectiveness," *IOSR Journal of Business and Management (IOSR-JBM)*, vol. 16, no. 4, pp. 47–52, 2014.

[131] P. M. Muchinsky, *Psychology Applied to Work*, Thomson Higher Education, Belmont, Nashville, Tennessee, USA, 9th edition, 2006.

[132] L. G. Bolman and T. E. Deal, *Reframing Organizations: Artistry, Choice, and Leadership*, Jossey-Bass, NJ, USA, 6th edition, 2017.

[133] A. Erbasi and T. Arat, "The effect of financial and non-financial incentives on job satisfaction: An Examination of food chain premises in Turkey," *International Business Research*, vol. 5, no. 10, pp. 136–145, 2012.

[134] R. Russell-Bennett, J. R. McColl-Kennedy, and L. V. Coote, "The relative importance of involvement and satisfaction on brand loyalty in a small business services setting," *Journal of Business Research*, vol. 60, no. 12, pp. 1253–1260, 2007.

[135] R. D. Stueart and B. B. Moran, *Library and Information Center Management*, Libraries Unlimited, Westport, USA, 2007.

[136] G. Von Dran, "Human resources and leadership strategies for libraries in transition," *Library Administration and Management*, vol. 19, no. 4, pp. 177–184, 2005.

[137] J. Cook and A. Crossman, "Satisfaction with performance appraisal systems: A study of role perceptions," *Journal of Managerial Psychology*, vol. 19, no. 5, pp. 526–541, 2004.

[138] H. Ganjinia, S. Gilaninia, and R. P. Sharami, "Overview of employees empowerment in organizations," *Oman Chapter of Arabian Journal of Business and Management Review (Oman Chapter)*, vol. 3, no. 2, pp. 38–43, 2013.

[139] M. S. Kahreh, H. Ahmadi, and A. Hashemi, "Achieving competitive advantage through empowering employees: An empirical study," *Far East Journal of Psychology and Business*, vol. 3, no. 2, pp. 26–37, 2011.

[140] N. Karakoc and A. K. Yilmaz, "Employee empowerment and differentiation in companies: A literature review and research agenda," *Enterprise Risk Management*, vol. 1, no. 2, 12 pages, 2009.

[141] R. Wagner and J. K. Harter, *12: The Elements of Great Managing*, Gallup Press, Canada, 2006.

[142] W. H. Knol, J. Slomp, R. L. Schouteten, and K. Lauche, "Implementing lean practices in manufacturing SMEs: testing 'critical success factors' using Necessary Condition Analysis," *International Journal of Production Research*, vol. 56, no. 11, pp. 3955–3973, 2018.

[143] M. Dora, M. Kumar, and X. Gellynck, "Determinants and barriers to lean implementation in food-processing SMEs – a multiple case analysis," *Production Planning and Control*, vol. 27, no. 1, pp. 1–23, 2015.

[144] M. Salanova and S. Llorens, "Employee empowerment and engagement," in *Workplace Well-Being: How to Build Psychologically Healthy Workplaces*, A. Day, E. K. Kelloway, and J. J. Hurrell, Eds., pp. 117–141, John Wiley & Sons, Ltd., Chichester, UK, 2014.

[145] J. Barrs, "Factors contributed by community organizations to the motivation of teachers in rural Punjab, Pakistan, and implications for the quality of teaching," *International Journal of Educational Development*, vol. 25, no. 3, pp. 333–348, 2005.

[146] W. W. Burke, *Organization Change: Theory and Practice*, SAGE Publications, Calif, USA, 5th edition, 2017.

[147] U. A. Agarwal, "Examining the impact of social exchange relationships on innovative work behaviour: Role of work engagement," *Team Performance Management*, vol. 20, no. 3-4, pp. 102–120, 2014.

[148] U. A. Agarwal, "Linking justice, trust and innovative work behaviour to work engagement," *Personnel Review*, vol. 43, no. 1, pp. 41–73, 2014.

[149] U. A. Agarwal, S. Datta, S. Blake-Beard, and S. Bhargava, "Linking LMX, innovative work behaviour and turnover intentions: The mediating role of work engagement," *Career Development International*, vol. 17, no. 3, pp. 208–230, 2012.

[150] M. Banihani, P. Lewis, and J. Syed, "Is work engagement gendered?" *Gender in Management: An International Journal*, vol. 28, no. 7, pp. 400–423, 2013.

[151] A. A. Chughtai and F. Buckley, "Work engagement: Antecedents, the mediating role of learning goal orientation and job performance," *Career Development International*, vol. 16, no. 7, pp. 684–705, 2011.

[152] S. E. Fawcett, G. K. Rhoads, and P. Burnah, "People as the bridge to competitiveness," *Benchmarking: An International Journal*, vol. 11, no. 4, pp. 346–360, 2004.

[153] Y. K. Park, J. H. Song, S. W. Yoon, and J. Kim, "Learning organization and innovative behaviour- The mediating effect of work engagement," *European Journal of Training and Development*, vol. 38, no. 1, pp. 75–94, 2013.

[154] A. B. Bakker and E. Demerouti, "Towards a model of work engagement," *Career Development International*, vol. 13, no. 3, pp. 209–223, 2008.

[155] C. Timms and P. Brough, ""I like being a teacher": Career satisfaction, the work environment and work engagement," *Journal of Educational Administration*, vol. 51, no. 6, pp. 768–789, 2013.

[156] R. J. Aldag and L. W. Kuzuhara, *Organizational Behaviour and Management: An Integrated Skills Approach*, Thomson Learning, South Western, UK, 2002.

[157] E. A. Locke and G. P. Latham, "What should we do about motivation theory? Six recommendations for the twenty-first century," *Academy of Management Review (AMR)*, vol. 29, no. 3, pp. 388–403, 2004.

[158] J. A. Gruman and A. M. Saks, "Performance management and employee engagement," *Human Resource Management Review*, vol. 21, no. 2, pp. 123–136, 2011.

[159] A. Wefald and R. Downey, "Construct dimensionality of engagement and its relation with satisfaction," *The Journal of Psychology: Interdisciplinary and Applied*, vol. 143, no. 1, pp. 91–111, 2009.

[160] O. M. Karatepe and G. Karadas, "Do psychological capital and work engagement foster frontline employees' satisfaction?: a study in the hotel industry," *International Journal of Contemporary Hospitality Management*, vol. 27, no. 6, pp. 1254–1278, 2015.

[161] A. B. Bakker, A. Shimazu, E. Demerouti, K. Shimada, and N. Kawakami, "Work engagement versus workaholism: A test of the spillover-crossover model," *Journal of Managerial Psychology*, vol. 29, no. 1, pp. 63–80, 2014.

[162] S. Abraham, "Development of employee engagement programme on the basis of employee satisfaction survey," *Journal of Economic Development, Management, IT, Finance and Marketing*, vol. 4, no. 1, pp. 27–37, 2012.

[163] M. Ibrahim and S. Al Falasi, "Employee loyalty and engagement in uae public sector," *Employee Relations*, vol. 36, no. 5, pp. 562–582, 2014.

[164] S. Biswas and J. Bhatnagar, "Mediator analysis of employee engagement: Role of perceived organizational support, p-o fit, organizational commitment and job satisfaction," *Vikalpa: The Journal for Decision Makers*, vol. 38, no. 1, pp. 27–40, 2013.

[165] Y. Brunetto, S. T. T. Teo, K. Shacklock, and R. Farr-Wharton, "Emotional intelligence, job satisfaction, well-being and engagement: Explaining organisational commitment and turnover intentions in policing," *Human Resource Management Journal*, vol. 22, no. 4, pp. 428–441, 2012.

[166] D. Swartling and B. Poksinska, "Management initiation of continuous improvement from a motivational perspective," *Journal of Applied Economics and Business Research*, vol. 3, no. 2, pp. 81–94, 2013.

[167] S. Bisgaard, "Quality management and Juran's legacy," *Quality and Reliability Engineering International*, vol. 23, no. 6, pp. 665–677, 2007.

[168] J. Readman and J. Bessant, "What challenges lie ahead for improvement programmes in the UK? Lessons from the CINet Continuous Improvement Survey 2003," *International Journal of Technology Management*, vol. 37, no. 3/4, article no 290, 2007.

[169] F. Jørgensen, H. Boer, and F. Gertsen, "Development of a team-based framework for conducting self-assessment of continuous improvement," *Journal of Manufacturing Technology Management*, vol. 15, no. 4, pp. 343–349, 2004.

[170] C. S. Dweck, *Mindset. The New Psychology of Success*, Random House Publishing, NY, USA, 2007.

[171] R. J. Thomas, F. Harburg, and A. Dutra, "How employee mindsets can be assessed to improve business performance," *Outlook- Accenture*, vol. 2, pp. 1–6, 2007.

[172] C. S. Dweck, G. M. Walton, and G. L. Cohen, *Academic tenacity: Mindsets and Skills that Promote Long-Term Learning*, Bill & Melinda Gates Foundation, Seattle, Wash, USA, 2014.

[173] D. B. Miele, L. K. Son, and J. Metcalfe, "Children's naive theories of intelligence influence their metacognitive judgments," *Child Development*, vol. 84, no. 6, pp. 1879–1886, 2013.

[174] A. Nolan, A. Taket, and K. Stagnitti, "Supporting resilience in early years classrooms: The role of the teacher," *Teachers and Teaching: Theory and Practice*, vol. 20, no. 5, pp. 595–608, 2014.

[175] K. Haimovitz, S. V. Wormington, and J. H. Corpus, "Dangerous mindsets: How beliefs about intelligence predict motivational change," *Learning and Individual Differences*, vol. 21, no. 6, pp. 747–752, 2011.

[176] H. Takeuchi, E. Osono, and N. Shimizu, "The contradictions that drive Toyota's success," *Harvard Business Review*, vol. 86, no. 6, pp. 96–141, 2008.

[177] N. A. Mehrzi and S. K. Singh, "Competing through employee engagement: A proposed framework," *International Journal of Productivity and Performance Management*, vol. 65, no. 6, pp. 831–843, 2016.

[178] R. Wellins and J. Concelman, "Creating a culture for engagement," *Workforce Performance Solutions*, vol. 4, pp. 1–4, 2005.

[179] B. Catlette and R. Hadden, *Contented Cows Give Better Milk: The Plain Truth about Employee Relations and Your Bottom Line*, Saltillo Publishing, Germantown, Md, USA, 2001.

[180] J. K. Harter, F. L. Schmidt, and T. L. Hayes, "Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: A meta-analysis," *Journal of Applied Psychology*, vol. 87, no. 2, pp. 268–279, 2002.

[181] D. A. Ortiz, W. K. Lau, and H. Qin, "Quantitative analysis of impacts of employee engagement on continuance and normative commitment," *International Journal of Services and Standards*, vol. 8, no. 4, article no 315, 2013.

[182] C. B. Agyemang and S. B. Ofei, "Employee work engagement and organisational commitment: A comparative study of private and public sector organisations in Ghana," *European Journal of Innovation and Research*, vol. 1, no. 4, pp. 20–33, 2013.

[183] A. Siddhanta and D. Roy, "Employee engagement: Engaging the 21st century workforce," *Asian Journal of Management Research*, vol. 3, pp. 2229–3795, 2010.

[184] S. G. Cheche, S. M. Muathe, and S. M. Maina, "Employee engagement, organisational commitment and performance of selected state corporations in Kenya," *European Scientific Journal*, vol. 13, no. 31, pp. 317–327, 2017.

[185] S. Devi, "Impact of employee engagement on organizational performance: A study of select private sector," *IMS Business School Presents Doctoral Colloquium*, pp. 10–13, 2017.

[186] E. M. Mone and M. London, *Employee Engagement- through Effective Performance Management- A Practical Guide for Managers*, Routledge, NY, USA, 2nd edition, 2017.

[187] P. Kazimoto, "Employee engagement and organizational performance of retails enterprises," *American Journal of Industrial and Business Management*, vol. 6, no. 4, pp. 516–525, 2016.

[188] M. Alagaraja and B. Shuck, "Exploring organizational alignment-employee engagement linkages and impact on individual performance," *Human Resource Development Review*, vol. 14, no. 1, pp. 17–37, 2015.

[189] M. A. Z. Dajani, "The impact of employee engagement on job performance and organisational commitment in the Egyptian banking sector," *Journal of Business and Management Sciences*, vol. 3, no. 5, pp. 138–147, 2015.

[190] A. Khalid and S. Khalid, "Relationship between organizational commitments, employee engagement and career satisfaction a case of University of Gujrat, Pakistan," *Journal of South Asian Studies*, vol. 3, no. 3, pp. 323–330, 2015.

[191] M. Geldenhuys, K. Łaba, and C. M. Venter, "Meaningful work, work engagement and organisational commitment," *SA Journal of Industrial Psychology*, vol. 40, no. 1, 2014.

[192] A. Imam and M. Shafique, "Impact of employee engagement in retaining employees through mediating effect of job satisfaction and organizational commitment and moderating effect of job stress: A Corporate banking sector study of Pakistan," *Journal of Applied Environmental and Biological Sciences*, vol. 4, no. 12, pp. 1–15, 2014.

[193] M. Shoko and A. Z. Zinyemba, "Impact of employee engagement on organizational commitment in national institutions of higher learning in Zimbabwe," *International Journal of*

*Advanced Research in Management and Social Sciences*, vol. 3, no. 9, pp. 255–268, 2014.

[194] S. V. Marinova, C. Peng, N. Lorinkova, L. Van Dyne, and D. Chiaburu, "Change-oriented behavior: A meta-analysis of individual and job design predictors," *Journal of Vocational Behavior*, vol. 88, pp. 104–120, 2015.

[195] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis*, Pearson, Hoboken, NJ, USA, 2014.

[196] J. C. Nunnally and I. H. Bernstein, *Psychometric Theory*, McGraw-Hill, New York, NY, USA, 1994.

[197] J. F. Hair, R. E. Anderson, R. L. Tatham, and W. C. Black, *Multivariate Data Analysis with Readings*, Prentice-Hall, Upper Saddle River, NJ, USA, 1998.

[198] J.-B. E. M. Steenkamp and H. C. M. van Trijp, "The use of lisrel in validating marketing constructs," *International Journal of Research in Marketing*, vol. 8, no. 4, pp. 283–299, 1991.