Complexity

Theories, Methods, and Applications for Performance Assessment in Complex Systems

Lead Guest Editor: Feng Li
Guest Editors: Qingyuan Zhu and Alireza Amirteimoori

Theories, Methods, and Applications for Performance Assessment in Complex Systems

Complexity

Theories, Methods, and Applications for Performance Assessment in Complex Systems

Lead Guest Editor: Feng Li
Guest Editors: Qingyuan Zhu and Alireza Amirteimoori

Contents

Hindawi
Complexity
Volume 2023, Article ID 9848730, 1 page
https://doi.org/10.1155/2023/9848730

Erratum

Erratum to "A Two-Sided Stable Matching Model of Cloud Manufacturing Tasks and Services considering the Nonlinear Relationship between Satisfaction and Expectations"

Yujie Zheng, Meiyan Li, and Jiakun Liu

Shandong University of Science and Technology, Qingdao, China

Correspondence should be addressed to Meiyan Li; limeiyanqdu@163.com

Received 25 November 2022; Accepted 25 November 2022; Published 7 February 2023

In the article titled "A Two-Sided Stable Matching Model of Cloud Manufacturing Tasks and Services considering the Nonlinear Relationship between Satisfaction and Expectations" [1], Dr. Jiakun Liu was incorrectly listed as the corresponding author. The correct corresponding author is Dr. Meiyan Li.

References

[1] Y. Zheng, M. Li, and J. Liu, "A Two-Sided Stable Matching Model of Cloud Manufacturing Tasks and Services considering the Nonlinear Relationship between Satisfaction and Expectations," *Complexity*, vol. 2021, Article ID 6735210, 11 pages, 2021.

Hindawi
Complexity
Volume 2022, Article ID 7801188, 12 pages
https://doi.org/10.1155/2022/7801188

Research Article

EFP-GA: An Extended Fuzzy Programming Model and a Genetic Algorithm for Management of the Integrated Hub Location and Revenue Model under Uncertainty

Yaser Rouzpeykar,[1] **Roya Soltani**,[2] **and Mohammad Ali Afashr Kazemi**[3]

[1]*Department of Industrial Engineering, Qeshm Branch, Islamic Azad University, Qeshm, Iran*
[2]*Department of Industrial Engineering, Faculty of Engineering, Khatam University, Tehran, Iran*
[3]*Department of Industrial and Information Technology Management, Tehran Central Branch, Islamic Azad University, Tehran, Iran*

Correspondence should be addressed to Roya Soltani; r.soltani@khatam.ac.ir

Received 17 November 2021; Revised 8 December 2021; Accepted 19 April 2022; Published 6 July 2022

Academic Editor: Alireza Amirteimoori

The aviation industry is one of the most widely used applications in transportation. Due to the limited capacity of aircraft, revenue management in this industry is of high significance. On the other hand, the hub location problem has been considered to facilitate the demands assignment to hubs. This paper presents an integrated p-hub location and revenue management problem under uncertain demand to maximize net revenue and minimize total cost, including hub establishment and transportation costs. A fuzzy programming model and a genetic algorithm are developed to solve the proposed model with different sizes. The mining and petroleum industry is used for case studies. Results show that the proposed algorithm can obtain a suitable solution in a reasonable amount of time.

1. Introduction

Hub location and revenue management are two research issues in network design that have been considered recently. The hub location model designs the transportation network structure. In contrast, the revenue management model allocates network capacity to customers in various classes based on price sensitivity.

Revenue management determines which products sell to which customers and at what price [1]. On the other hand, it has been widely used in the aviation industry due to the limited number of aircraft seats. Seats are usually offered to various customer classes at different prices [2]. Airlines categorize customers and allocate different capacities according to price to obtain maximum revenue. Capacity control includes several models, algorithms, and policies that allocate seats to maximize expected profits [3]. Hub location problem is related to hub facilities placement and allocation to demand nodes to determine traffic routes between origin and destination pairs. Researchers attract this issue because it significantly reduces the number of network connections and system costs. In the star p-hub network, p nodes are selected. Each node is connected to only one hub, and all hubs are connected to a central hub. A central hub is predefined, while other hubs are determined by the model [4]. There are four types of hub location problems: median, center, covering, and fixed cost hub location. In the p-hub median problem, p nodes are located to minimize the total cost of flows in the network. A number of the hub is predefined in this problem. A p-hub center problem seeks the optimal location of p-hubs. It allocates nonhub nodes to hub nodes where the full path in-network is minimized. The number of hubs is not specified in the hub covering problem, and demands are covered within a certain distance. Minimizing the cost of installing facilities covered by hubs in such problems. In hub location problem with fixed costs number of hubs is not defined at first. Flow and installation hubs costs are minimized in this problem [5]. This research is structured as follows: the literature review is presented in Section 2. The credibility-based fuzzy theory is described in Section 3. Section 4 defines the problem statement along with model formulation. Section 5 describes the proposed solution method. In Section 6, computational results are presented, and finally, in Section 7, conclusions and recommendations for future research studies are presented.

2. Literature Review

Nowadays, the hub location problem is studied to maximize profit considering a revenue-cost trade-off. Revenue management has been considered in several forms in research. It is derived from the transportation flows [6–9]. In another category, it is derived from a combination of pricing and hub location [10]. Finally, integrated revenue management and hub location are considered another category [11, 12]. Hörhammer [13] studied a dynamic multiperiod hub location problem with multiple capacity levels. They considered a nonhub node can be a hub in the next period. Proposed a method that has four main steps called Distribution-Map-Transfer-Combination (DMTC). A quadratic mixed-integer programming model based on flow and route is developed. The aim is to minimize connection costs between a nonhub and hub nodes, transportation costs between a hub and other ones, and installation costs. He et al. [14] proposed a nonlinear mixed-integer programming model for hub location problems considering support hub. Lagrangian relaxation and branch and bound methods were applied to solve the proposed problem. Ebrahimnejad et al. [15] developed a particle swarm optimization algorithm for shortest path problems with mixed fuzzy arc weights. Adibi and Razmi [16] presented a two-stage stochastic model for multiple allocations in the hub location problem. It is assumed that demand and transportation costs are probabilistic. Damgacioglu et al. [17] developed a GA to solve the problem considering uncapacitated allocation. Alumur et al. [18] presented a multiperiod hub location problem for multiple allocations. Installing a new hub and available hub capacity expansion is allowed in the study. A MIP model is developed to minimize shipment, hubs connection, hub installation, and capacity expansion cost. Azizi et al. [19] presented a hub location model under hub failure risk. They considered that a support hub could be applied to supply demand when a hub goes out of order. Grauberger and Kimms [20] investigated an airline revenue management problem considering price competition and limited capacity. He [21] studied the revenue management effect on a hub-to-hub network.

Tikani et al. [3] studied an integrated hub location and revenue management considering several customer classes to maximize profit and minimize costs. To do this end, a two-stage stochastic model is developed to determine hub location. Furthermore, an efficient genetic algorithm is proposed to solve the problem on a large scale. Alumur et al. [22] investigated capacitated single and multiple hub location problems. A direct connection between two nonhubs is considered in this study. A MIP model is developed to minimize transportation and hub installation costs. Hou et al. [12] presented an integrated p-hub location and revenue management problem considering multiple capacities under disruptions. A two-stage stochastic model is developed to maximize net profit in which hub installation cost, shipment cost, and revenue obtained from ticket selling are considered. A robust integrated optimization and stochastic programming to maximize weighted total profit is presented to obtain reasonable solutions. Huo et al. [11] studied an integrated hub location and revenue management problem considering average and worst-case analysis. A p-hub is selected from n nodes while uncertain data and some scenarios are considered in the study. Then, a two-stage stochastic programming model is developed to maximize profit. Ahmadi et al. [23] proposed a unique hybrid strategy for selecting users with Deep-Q-Reinforcement Learning with Federated Learning. Korani et al. [24] proposed a reliable multimodal hub location problem. They developed a Lagrangian method considering the strategic level that causes to achieve accurate solutions. Čvokić and Stanimirović [25] introduced a new uncapacitated single allocation hub location problem under a deterministic and robust approach to maximize net profit. A mixed-integer quadratic model is proposed. Furthermore, a two-phase meta-heuristic algorithm is developed. Rouzpeykar et al. [26] developed a robust optimization model for the integrated hub location and revenue management problem under uncertainty. They applied a case study in Iran to validate the proposed model.

Di Caprio et al. [27] developed an ant colony algorithm under uncertainty for the shortest path method problem. They assumed that the arc weights were fuzzy. The proposed algorithm is compared with GA, PSO, and the artificial bee colony algorithms. Ebrahimnejad et al. [28] developed an artificial bee colony algorithm under uncertainty for the shortest path method problem. They considered mixed interval fuzzy numbers for the arc weights. Sori et al. [29] studied the constrained shortest path problem in location-based online services to find a path with the lowest cost with fuzzy time and cost. The summary of the last works is presented in Table 1.

According to Table 1, many studies have addressed the hub location problem. In contrast, some of them have considered it with revenue management simultaneously. A few numbers types of research have examined this issue under uncertainty. Those research studies have applied stochastic or robust approaches to deal with uncertainty. In this study, a credibility-based fuzzy theory will be used to model uncertainty in an integrated hub location and revenue management for the first time. Through this method, managers can select different levels of confidence based on their experiences. A fuzzy mixed-integer programming model has been developed to deal with the proposed problem with uncertain parameters.

Table 1: Summary of previous research.

Researcher (year)	Problem		Objective		Model			
	Location	Revenue management	Cost	Revenue	Deterministic	Stochastic	Robust	Fuzzy
Hörhammer [13]	*		*		*			
He et al. [14]	*		*		*			
Adibi and Razmi [16]	*		*			*		
Alumur et al. [18]	*		*		*			
Grauberger and Kimms [20]		*		*	*			
He [21]		*		*	*			
Tikani et al. [3]	*	*	*	*		*		
Alumur et al. [22]	*		*		*			
Hou et al. [12]	*	*	*	*		*	*	
Huo et al. [11]	*	*	*	*		*		
Čvokić and Stanimirović [25]	*			*	*		*	
Present study	*	*	*	*				*

3. Credibility-Based Fuzzy Theory

This study uses a fuzzy approach to consider uncertainty [30].

$$\widetilde{A} = \left\{ \left(x, \mu_{\widetilde{A}}(x) \right) | x \in X \right\}, \tag{1}$$

where $\widetilde{A}$ is a fuzzy set and $\mu_{\widetilde{A}}(x)$ is calculated by the following equation:

$$\mu_{\widetilde{A}}(x)\colon X \longrightarrow [0,1]. \tag{2}$$

Different fuzzy numbers such as triangular fuzzy numbers or trapezoidal fuzzy numbers can be used in the fuzzy approach [31]. Due to the nature of the proposed problem, the trapezoidal fuzzy number has been used in this study. In a triangular fuzzy number, only one parameter value gives the maximum amount of confidence. In contrast, in the trapezoidal fuzzy number, the maximum value of a parameter is obtained. In this case, the risk-taking of decision-makers is reduced, and they can accept uncertainty in natural conditions with more confidence [32]. A membership function of a trapezoidal fuzzy number $\xi = (l, m_1, m_2, u)$ is as follows (Figure 1):

$$\mu(x) = \begin{cases} \dfrac{x-l}{m_1-l} & l \le x \le m_1, \\ 1, & m_1 \le x \le m_2, \\ \dfrac{u-x}{u-m_2}, & m_2 \le x \le u, \\ 0, & O.W. \end{cases} \tag{3}$$

A fuzzy credibility model is applied for the integrated proposed model. The credibility measure is defined as [33]

$$Cr\{\xi \le A\} = \frac{1}{2}\{\text{Pos}\{\xi \le A\} + \text{Nec}\{\xi \le A\}\}, \tag{4}$$

where ξ and A are fuzzy variables and real numbers, respectively. Possibility (Pos) and necessity (Nec) measures are defined as (5) and (6), respectively.

$$\text{pos}\{\xi \le A\} = \sup \mu_x(x)$$

$$= \begin{cases} 0, & x \le l, \\ \dfrac{x-l}{m_1-l}, & l \le x \le m_1, \text{Nec}\{\xi \le A\} = 1 - \sup\limits_{\xi \ge A} \mu_x(x) = \begin{cases} 0, & x \le m_2, \\ 1 - \dfrac{u-x}{u-m_2} = \dfrac{x-m_2}{u-m_2}, & m_2 \le x \le u,. \\ 1, & x \ge u. \end{cases} \\ 1, & x \ge m_1, \end{cases} \tag{5}$$

The possibility and necessity measures are also shown in Figure 2.

The following equation calculates the credibility measure shown in Figure 3 [34]:

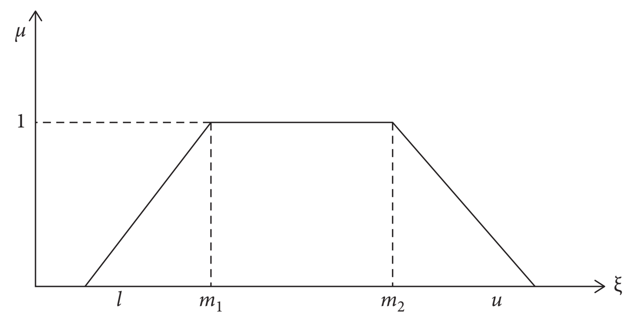


Figure 1: Trapezoidal fuzzy number.

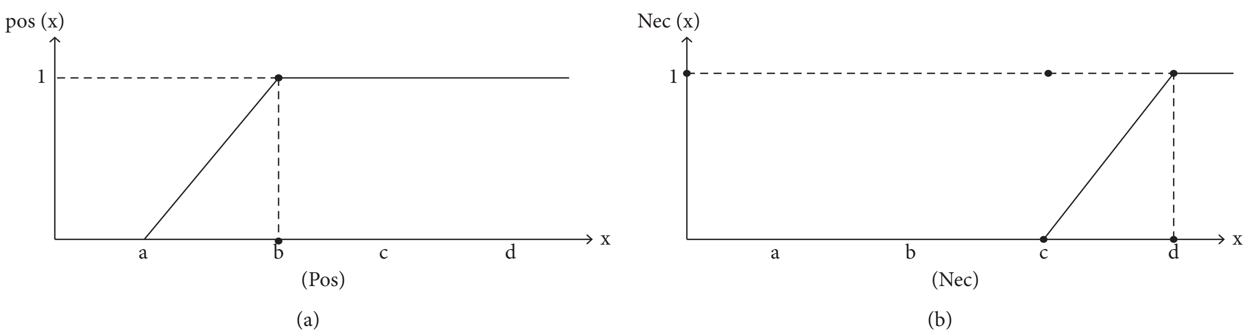


Figure 2: Possibility and necessity measures. (a) (Pos) and (b) (Nec).

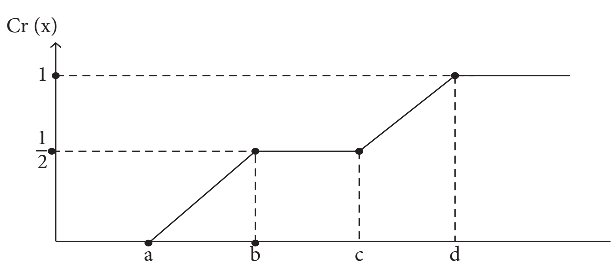


Figure 3: Credibility measures.

$$Cr\{\xi \leq A\} = \frac{1}{2}\{\text{pos}\{\xi \leq A\} + \text{Nec}\{\xi \leq A\}\}$$
$$= \begin{cases} 0, & x \leq l, \\ \dfrac{x-l}{2(m_1 - l)}, & l \leq x \leq m_1, \\ \dfrac{1}{2}, & m_1 \leq x \leq m_2, \\ \dfrac{1}{2}\left(1 + \dfrac{x - m_2}{u - m_2}\right), & m_2 \leq x \leq u, \\ 1, & x \geq u. \end{cases} \tag{6}$$

If ξ is a trapezoidal fuzzy number and $\alpha > 0.5$, then [35]

$$\begin{aligned} Cr\{\xi \leq x\} \geq \alpha &\Longrightarrow x \geq (2-2\alpha)m_2 + (2\alpha - 1)u, \\ Cr\{\xi \geq x\} \geq \alpha &\Longrightarrow x \leq (2\alpha - 1)l + (2-2\alpha)m_1. \end{aligned} \tag{7}$$

4. Problem Statement

In this problem, a central hub is connected to some hub nodes. There are some candidate hub nodes that p of them should be selected. Then, other nonhub nodes are connected to hub ones so that total transportation and installation costs are minimized while the revenue obtained from selling tickets is maximized. Based on their capacity, aircraft determine their route from a hub node to nonhub ones where the maximum required demand is satisfied.

4.1. Model Formulation. The proposed model includes hub location problems and revenue management in the aircraft industry under uncertainty to maximize revenue from network transportation and minimize total cost. It is assumed that all nodes can be selected as a hub and p-hubs have been selected from a set of n nodes connected to a central hub. Other assumptions are presented as follows.

4.2. Assumptions

(i) All origin and destination nodes are candidates to become a hub
(ii) The number of hubs is predefined
(iii) The central hub location is given

(iv) A node is only allocated to one hub
(v) Each two-node is not connected directly while they connected by a hub
(vi) There is not any direct shipment between the two hubs
(vii) The number of flights between nodes (between central and other hubs, as well as between hubs and nonhubs) is limited
(viii) Aircraft which traverse between hubs and nonhubs have different capacity
(ix) The ticket price at each class is predefined
(x) Extra cargo for passengers is allowed
(xi) Goods can be carried

4.3. Notation

4.3.1. Sets and Indices

N: node number
P: hub number
K: flight class number
i, m: node indices $i, m = 1, 2, \ldots, N$
j: hub indices $j = 1, 2, \ldots, P$
k: flight class indices $k = 1, 2, \ldots, K$

4.3.2. Parameters

dis_{j0}: distance from the central hub to hub j
dis_{ij}: distance from hub j to nonhub i
c_{j0k}: unit transfer cost between the central hub and hub j for class k
c_{ijk}: unit transfer cost between hub j and nonhub i for class k
cap_{j0}: available flight number from the central hub to hub j
cap_{ij}: available flight number from hub j and nonhub i
p_{imk}: ticket price from node i to node m for class k
ph_{imk}: extra cargo price per unit from i to m for class k
pg_{im}: goods carrying price per unit from i to m
d_{imk}: demand between i and m for class k
vh_{imk}: amount of extra cargo from i to m for class k
vg_{im}: amount of goods carrying from i to m
cl_1: the capacity of the link between the central hub and other hubs
cl_2: the capacity of the link between hubs and nonhubs
fc_0: fixed cost for establishing a central hub
fc_j: fixed cost for establishing hub j
A: a huge integer
λ_{im}: the confidence level of the decision-maker for the link between i and m

4.3.3. Decision Variables

x_{imk}: number of tickets sold between nodes i and m for class k
y_{imk}: protection level between nodes i and m for class k
z_{ij} 1: if nonhub i is connected to hub j, and 0 otherwise
z_{jj} 1: if node i is selected as a hub, and 0 otherwise
o_{im} 1: if a flight was done between nodes i and m, and 0 otherwise

4.3.4. Mathematical Model. The proposed biobjective model is formulated as follows:

$$\begin{aligned}\max z_1 = &\sum_{i=1}^{N}\sum_{m=1}^{N}\sum_{k=1}^{K} p_{imk} \times x_{imk}\\ &+ \sum_{i=1}^{N}\sum_{m=1}^{N}\sum_{k=1}^{K} ph_{imk} \times vh_{imk} \times o_{\text{im}}\\ &+ \sum_{i=1}^{N}\sum_{m=1}^{N} pg_{\text{im}} \times vg_{\text{im}} \times o_{\text{im}},\end{aligned} \tag{8}$$

$$\begin{aligned}\min z_2 = &\sum_{j=1}^{P}\sum_{k=1}^{K} c_{j0k}\left[\text{dis}_{j0}\left(\sum_{i=1}^{N}\sum_{m=1,m\neq i}^{N}\left(\frac{x_{imk}}{cl_2}\right)z_{ij}\left(1-z_{mj}\right)\right) + \text{dis}_{0j}\left(\sum_{i=1}^{N}\sum_{m=1,m\neq i}^{N}\left(\frac{x_{imk}}{cl_2}\right)z_{ij}\left(1-z_{mj}\right)\right)\right]\\ &+ \sum_{i=1}^{N}\sum_{j=1}^{P}\sum_{k=1}^{K} c_{ijk}\left[\text{dis}_{ij}\left(\sum_{m=1,m\neq i}^{N}\frac{x_{imk}}{cl_1}\right) + \text{dis}_{ji}\left(\sum_{m=1,m\neq i}^{N}\frac{x_{imk}}{cl_1}\right)\right]z_{ij}\\ &+ \sum_{j=1}^{P} fc_j z_{jj} + fc_0,\end{aligned} \tag{9}$$

st

$$\sum_{j=1}^{P} z_{ij} \leq 1 \quad \forall i = 1, 2, \ldots, N, \tag{10}$$

$$\sum_{j=1}^{N} z_{jj} = P, \tag{11}$$

$$z_{ij} \leq z_{jj} \quad \forall i = 1, 2, \ldots, N, \forall j = 1, 2, \ldots, P, \tag{12}$$

$$x_{imk} \leq \tilde{d}_{imk} \quad \forall i, m = 1, 2, \ldots, N, \forall k = 1, 2, \ldots, K, \tag{13}$$

$$x_{imk} \leq y_{imk} \quad \forall i, m = 1, 2, \ldots, N, \forall k = 1, 2, \ldots, K, \tag{14}$$

$$\sum_{k=1}^{K} x_{imk} \geq o_{\text{im}} \quad \forall i, m = 1, 2, \ldots, N, \tag{15}$$

$$\sum_{k=1}^{K} x_{imk} \leq A \times o_{\text{im}} \quad \forall i, m = 1, 2, \ldots, N, \tag{16}$$

$$\begin{aligned} &\sum_{i=1}^{N} \sum_{j=1}^{P} \sum_{k=1}^{K} \left(y_{imk}/cl_2\right) z_{ij}\left(1 - z_{mj}\right) \\ &\quad + \sum_{i=1}^{N} \sum_{j=1}^{P} \sum_{k=1}^{K} \left(y_{\text{mik}}/cl_2\right) z_{ij}\left(1 - z_{mj}\right) \leq \text{cap}_{j0} \times z_{jj} \\ &\quad \forall j = 1, 2, \ldots, P, \end{aligned} \tag{17}$$

$$\begin{aligned} &\sum_{m=1}^{N} \sum_{k=1}^{K} \left(y_{imk}/cl_1\right) + \sum_{m=1}^{N} \sum_{k=1}^{K} \left(y_{mik}/cl_1\right) \leq \sum_{j=1}^{P} \text{cap}_{ij} \times z_{ij} + A \\ &\quad \times z_{ii} \\ &\quad \forall i = 1, 2, \ldots, N, \end{aligned} \tag{18}$$

$$z_{ij}, o_{\text{im}} \in \{0, 1\} \quad \forall i, m = 1, 2, \ldots, N, \forall j = 1, 2, \ldots, P, \tag{19}$$

$$x_{imk}, y_{imk} \geq 0 \quad \forall i, m = 1, 2, \ldots, N, \forall k = 1, 2, \ldots, K. \tag{20}$$

Equation (8) shows the revenue obtained from selling tickets in various classes and carrying extra cargo and goods. Equation (9) calculates the total cost of the network, including the total transportation cost between nodes and the total installation cost of hubs. Total network profit is obtained from the difference between the two objective functions ($z_1 - z_2$). Equation (10) ensures that each nonhub node should be allocated to only one hub. Equation (11) states that there are precisely p-hubs in the network. Equation (12) enforces that a nonhub node has been allocated to a hub node if this node had been selected as a hub. Equations (13) and (14) show that the maximum number of sold tickets equals demand and the protection level, respectively. Equations (15) and (16) indicate that a flight between nodes i and m if tickets had sold for that route. Equation (17) states that the protection level should not exceed the physical capacity between the central hub and other hubs. Equation (18) indicates that the protection level should not exceed the physical capacity between hob and nonhub nodes. Finally, variables of the model are introduced in equations (19) and (20).

4.3.5. Credibility-Based Fuzzy Approach. Generally, the credibility-based chance-constrained programming [31, 36] is a computationally efficient fuzzy mathematical programming approach that relies on solid mathematical concepts and can support different kinds of fuzzy numbers such as triangular and trapezoidal forms as well as enabling the decision-maker to satisfy some chance constraints in at least some given confidence levels. According to equation (18), a trapezoidal fuzzy number is considered for demand between nodes. Based on the credibility approach, equation (13) is reformulated in equation (21), which is equivalent to equation (24).

$$Cr\left\{x_{imk} \leq \tilde{d}_{imk}\right\} \geq \lambda_{\text{im}} \quad \forall i, m = 1, 2, \ldots, N, \forall k = 1, 2, \ldots, K, \tag{21}$$

$$\begin{aligned} x_{imk} &\leq \left(2\lambda_{\text{im}} - 1\right)\tilde{d}_{imk(1)} \\ &\quad + \left(2 - 2\lambda_{\text{im}}\right)\tilde{d}_{imk(2)} \\ &\quad \forall i, m = 1, 2, \ldots, N, \quad \forall k = 1, 2, \ldots, K. \end{aligned} \tag{22}$$

5. Solution Methodology

The proposed model is NP-hard, and its complexity increases by increasing the number of hubs [3]. Thus hub selection and its assignment to other nonhub nodes would be more complex. Therefore, large-sized problems cannot be solved by the exact method in a reasonable time. To deal with this problem, a genetic algorithm, a population-based metaheuristic, is employed in this paper.

5.1. Genetic Algorithm. Few works applied metaheuristic algorithms to integrate revenue management and hub location problem. However, the genetic algorithm is used for this type of problem. The main reason to apply this approach is that it is easier to design the problem by GA. The pseudocode of the proposed genetic algorithm is depicted in Figure 4, where algorithm parameters are firstly set by the Taguchi parameter setting method. Then, initial solutions are created where infeasible solutions are revised until a feasible one is generated. If the feasible solution is not achieved, we use the death penalty as the infeasible solution. After that fitness function of each solution is calculated. After creating initial solutions, a repetitive process involving crossover and mutation operators to generate offsprings and mutated solutions and calculation of fitness functions of solutions and selection of the best solution is made until a predetermined stop condition is satisfied.

1- *Parameter Setting*
2- *Create initial solutions*
3- *Calculate fitness function*
4- *Select the best solution*
5- *Create offspring using:*
6- *Crossover*
7- *Mutation*
8- *Calculate fitness function*
9- *Select the best solution*
10- *Check stop condition*
11- *If it is true, go to the next step while going to step 5*
12- *End*

Figure 4: Pseudocode.

5.1.1. Solution Representation. The solution representation of this algorithm is illustrated in Figure 5 and described as follows: consider a central hub, two hubs, and three nonhubs nodes. Firstly, a random matrix is created equal to the node's number (5 in this example). Then, the maximum values in the diagonal are selected as a hub while exactly p-hubs are obtained. Then, the highest number in each row related to nonhubs nodes at the intersection of columns assigned as hubs determines its allocation.

As shown in Figure 5, nodes 1 and 4 are selected as a hub. In the next step, the maximum value at the intersection of columns selected as a hub in each nonhub row is allocated. According to the allocation structure, the network design is depicted in Figure 6.

5.1.2. Crossover Operator. This study applies a one-cut point crossover to create offspring chromosomes from two randomly selected parents. Two new offspring are obtained using the following equation:

$$\begin{aligned} p_1 &= b\,pfn + (1-b)psn, \\ p_2 &= (1-b)pfn + b\,psn. \end{aligned} \quad (23)$$

Here, b is a matrix with parents size, and *pfn* and *psn* are n dimension matrices of the first and the second parent, respectively.

5.1.3. Mutation Operator. In this operator, a hub node exchanges with a nonhub node. One of the genes representing a hub node is selected randomly. Then, it is changed to a nonhub node so that the nonhub node with a higher random value is selected as a hub based on Figure 5. The following equation is used to mutate each gene of a solution:

$$\text{gen}_j^{\text{new}} = 1 - \text{gen}_j^{\text{old}}. \quad (24)$$

Figure 7 shows the mechanism of the mutation.

6. Computational Results

The proposed model validity is assessed using two problem instances based on Tikani et al. [3]. In the following, sample problems are designed, and the algorithm parameters are tuned.

6.1. Data Generation. Sample problems are presented in various scales, as shown in Table 2. It should be noted that input parameters for medium and large instances are randomly generated using uniform distributions. Also, Table 3 states the problem dimension.

6.2. Parameters Tuning. Taguchi experimental design in MINITAB software is applied to tune the parameters of the proposed GA, including population size, number of iteration, mutation, and crossover rates. Their values are assessed at three levels shown in Table 4.

RPD (relative percentage deviation) shown in the following equation is used as a GAP criterion to analyze the performance of the proposed GA:

$$\text{GAP} = \left(\frac{alg_{\text{sol}} - \text{best}_{\text{sol}}}{\text{best}_{\text{sol}}}\right) \times 100, \quad (25)$$

where alg_{sol} and best_{sol} are the objective function value and the best value of them obtained by the algorithm execution, respectively. An instance (No. 1) is randomly selected to execute for each of the combinations listed in Table 5 and then the GAP measure is calculated and plotted as shown in Figure 5.

Figure 8 indicates the Taguchi method analysis to tune the proposed genetic algorithm parameters. As can be seen, the best values for population size, number of iteration, mutation, and crossover rate are 200, 300, 0.01, and 0.85, respectively.

6.3. Solution Results. As mentioned above, a fuzzy MIP model is proposed for the integrated hub location and revenue management problem under fuzzy demand. A genetic algorithm is developed to solve large-sized problems. To validate the proposed genetic algorithm, small-sized problems are solved by both the GAMS optimization package and the proposed GA. Then, large-sized problems are solved, and the results of the two methods are compared in Table 6.

The comparative results obtained from the two methods revealed that the proposed GA could achieve the same solution as the exact method of GAMS [37]. This indicates the validity of the proposed genetic algorithm. Moreover, the proposed GA can solve the proposed problem on a large scale. The computational time of solving sample problems demonstrates that the problem has high complexity. The average time is increased by increasing the size of the problems, as shown in Figure 9.

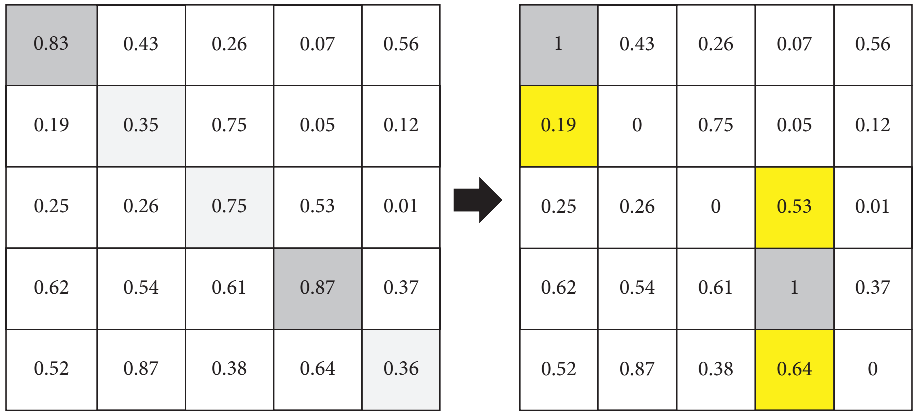


Figure 5: Solution representation.

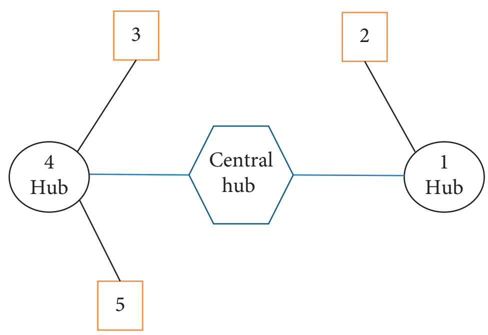


Figure 6: Network design [26].

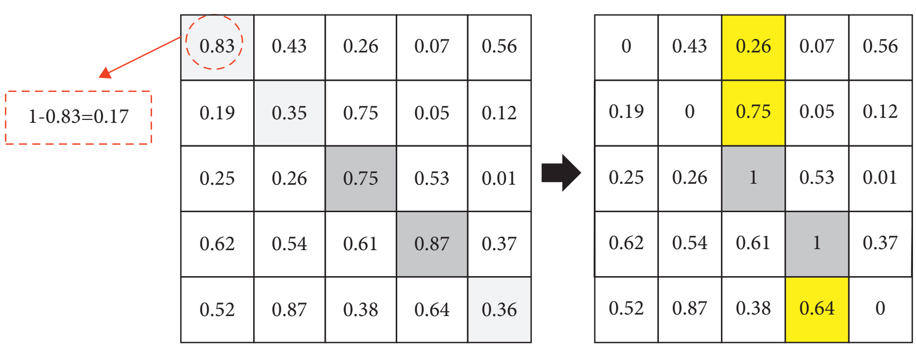


Figure 7: Mutation operator.

Table 2: Sample problems parameters value.

Parameter	Value		Value
dis_{j0}	Uniform (3500, 50000)	dis_{ij}	Uniform (1000, 3500)
c_{j0k}	Uniform (200, 500)	c_{ijk}	Uniform (100, 200)
cap_{j0}	Uniform (10, 20)	cap_{ij}	Uniform (15, 30)
fc_0	Uniform (100000, 200000)	fc_j	Uniform (50000, 80000)
cl_1	Uniform (5000, 10000)	cl_2	Uniform (10000, 15000)
p_{imk}	Uniform (500, 1500)	d_{imk}	Uniform (400, 800)
ph_{imk}	Uniform (20, 100)	pg_{im}	Uniform (50, 150)
vh_{imk}	Uniform (50, 500)	vg_{im}	Uniform (1000, 2000)
A	1000000	λ_{im}	Uniform (0.5, 0.8)

Table 3: Sample problems dimension.

Size	Prob. no.	Nodes no.	Hubs no.	Classes no.
Small	1	5	1	3
	2	6	1	3
Medium	3	10	2	4
	4	12	2	4
Large	5	18	4	5
	6	20	4	5

Table 4: GA parameters value.

Population size	Crossover rate	Mutation rate	No. of iterations
70	0.75	0.006	150
150	0.85	0.009	300
200	0.95	0.01	500

Table 5: Genetic algorithm orthogonal.

State no.	Population size	Crossover rate	Mutation rate	No. of iterations	Value GAP
1	70	0.75	0.006	150	0.5032
2	70	0.85	0.009	300	0.1259
3	70	0.95	0.01	500	0.7419
4	150	0.75	0.009	500	0.6635
5	150	0.85	0.01	150	0.4917
6	150	0.95	0.006	300	0.0045
7	200	0.75	0.01	300	0.7124
8	200	0.85	0.006	500	0.7280
9	200	0.95	0.009	300	0.2942

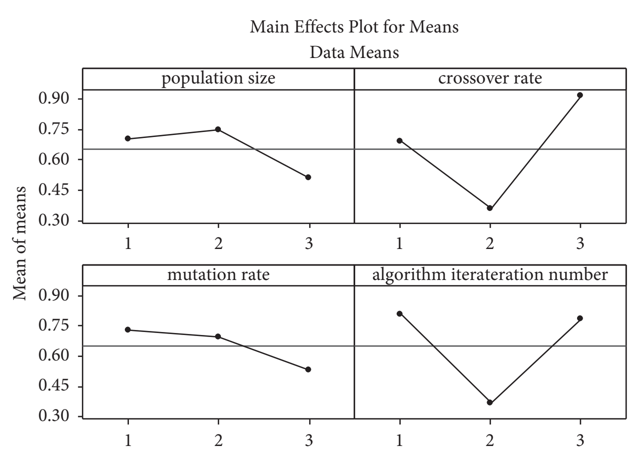


Figure 8: The mean effect of GA.

6.4. Sensitivity Analysis. In this section, the effectiveness of essential parameters is analyzed. Firstly, the impact of demand on objective functions is investigated. As shown in Figure 10, objective function values increased by increasing the amount of demand. Network revenue and total costs increase when the number of passengers increases.

Furthermore, link capacity's effect on objective function is assessed. To this end, we change this capacity from −10% to +10%, as shown in Figure 11. As expected, the capacity link is only affected by the total cost.

Finally, the impact of confidence level (fuzzy membership) on revenue and cost is analyzed. This parameter shows the confidence level of decision-makers for demand. Thus, λ_{im} it is changed from 0.5 to 1 and depicted in Figure 12. It indicates that the higher the confidence level decision-makers adopt, the total cost fluctuation will decrease based.

Table 6: Solution results of sample problems.

Size	Prob. no.	GAMS			GA		
		Z_1	Z_2	CPU time (s)	Z_1	Z_2	CPU time (s)
Small	1	640.000	425.000	74.6	640.000	425.000	43.6
	2	673.500	447.000	91.4	673.500	447.000	76.4
Medium	3	1.098.000	921.500	753.6	1.095.500	923.000	136.1
	4	—	—	>1000	1.154.000	963.500	160.8
Large	5	—	—	>1000	2.041.500	1.714.500	273.7
	6	—	—	>1000	2.327.000	1.908.000	301.2

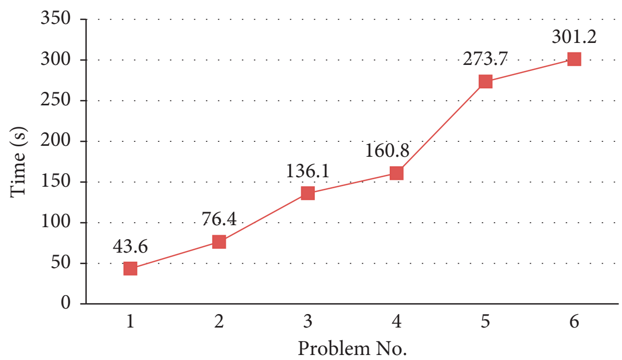


Figure 9: Computational time comparison.

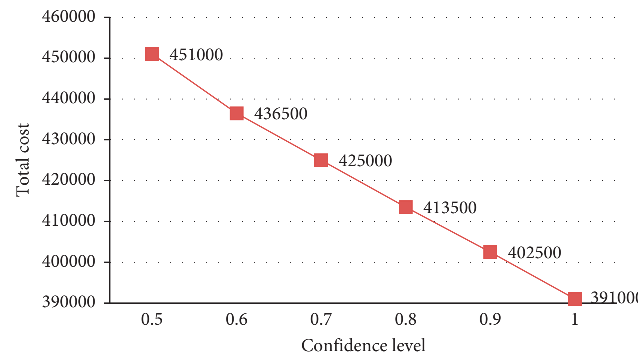


Figure 12: Confidence level effect on total cost.

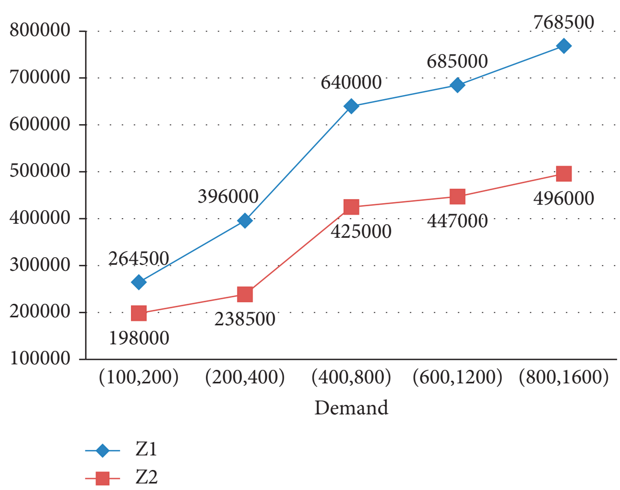


Figure 10: Effect of demand on the objective functions.

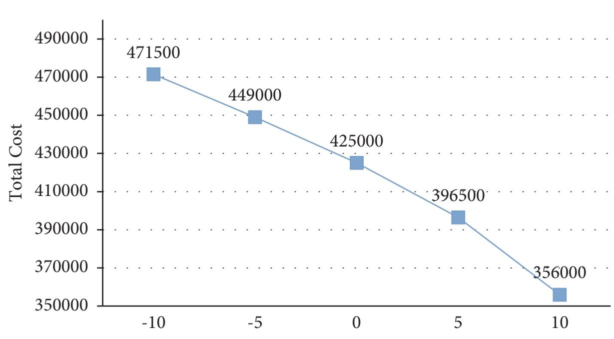


Figure 11: Effect link capacity on the total cost.

7. Conclusions and Future Directions

This paper has developed a fuzzy MIP model for the integrated revenue management and p-hub location problem. The objectives are maximizing network revenue as well as minimizing total costs. A credibility-based fuzzy theory has been used to deal with uncertainty in an integrated problem. In order to evaluate the proposed mathematical model, some problem instances have been used and solved using the CPLEX solver of GAMS software. Furthermore, a genetic algorithm has been developed for large-sized problems. Then, a sensitivity analysis has been performed on crucial inputs of the problem, including demand, link capacity, and confidence level. The revenue will be increased by decreasing the number of hub nodes. Thus, the number of hubs has an essential effect on revenue earned from the network. Therefore, managers should make connections among nonhub nodes so that the minimum number of hub nodes is opened and all nonhub nodes are allocated to the hub ones.

As mentioned above, only the demand for flight is considered uncertain in this study. However, there are different parameters, such as parameters related to cost, that can be considered uncertain, too. A fuzzy approach is applied in this study. At the same time, other methods such as robust optimization and stochastic programming could be used in future works. Moreover, a credibility-based fuzzy approach is applied in this study; however, there are other approaches such as equivalent auxiliary crisp and α-cut level concept. Moreover, artificial intelligence-based algorithms can solve large-sized problems in reasonable run times [38, 39].

Data Availability

Data are available from the corresponding author (r.soltani@khatam.ac.ir) upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] D. D. Sierag, G. M. Koole, R. D. Van der Mei, J. I. Van der Rest, and B. Zwart, "Revenue management under customer choice behaviour with cancellations and overbooking," *European Journal of Operational Research*, vol. 246, no. 1, pp. 170–185, 2015.

[2] D. Çetiner, *Fair revenue sharing mechanisms for strategic passenger airline alliances*, Springer Science & Business Media, Berlin, Germany, 2013.

[3] H. Tikani, M. Honarvar, and Y. Z. Mehrjerdi, "Developing an integrated hub location and revenue management model considering multi-classes of customers in the airline industry," *Computational and Applied Mathematics*, vol. 37, no. 3, pp. 3334–3364, 2018.

[4] H. Yaman and S. Elloumi, "Star p-hub center problem and star p-hub median problem with bounded path lengths," *Computers & Operations Research*, vol. 39, no. 11, pp. 2725–2732, 2012.

[5] M. Mohammadi, R. Tavakkoli-Moghaddam, A. Siadat, and Y. Rahimi, "A game-based meta-heuristic for a fuzzy bi-objective reliable hub location problem," *Engineering Applications of Artificial Intelligence*, vol. 50, pp. 1–19, 2016.

[6] A. Alibeyg, I. Contreras, and E. Fernández, "Hub network design problems with profits," *Transportation Research Part E: Logistics and Transportation Review*, vol. 96, pp. 40–59, 2016.

[7] R. Neamatian Monemi, S. Gelareh, S. Hanafi, and N. Maculan, "A co-opetitive framework for the hub location problems in transportation networks," *Optimization*, vol. 66, no. 12, pp. 2089–2106, 2017.

[8] A. Alibeyg, I. Contreras, and E. Fernández, "Exact solution of hub network design problems with profits," *European Journal of Operational Research*, vol. 266, no. 1, pp. 57–71, 2018.

[9] G. Taherkhani and S. A. Alumur, "Profit maximizing hub location problems," *Omega*, vol. 86, pp. 1–15, 2019.

[10] C.-C. Lin and S.-C. Lee, "Hub network design problem with profit optimization for time-definite LTL freight transportation," *Transportation Research Part E: Logistics and Transportation Review*, vol. 114, pp. 104–120, 2018.

[11] J.-Z. Huo, Y.-T. Hou, F. Chu, and J.-K. He, "A combined average-case and worst-case analysis for an integrated hub location and revenue management problem," *Discrete Dynamics in Nature and Society*, vol. 2019, Article ID 8651728, 13 pages, 2019.

[12] Y. T. Hou, J. Z. Huo, and F. Chu, "An integrated problem of p-hub location and revenue management with multiple capacity levels under disruptions," *Journal of Advanced Transportation*, 2019.

[13] A. M. Hörhammer, "Dynamic hub location problems with single allocation and multiple capacity levels," in *Proceedings of the 2014 47th Hawaii International Conference on System Sciences*, pp. 994–1003, IEEE, Washington, DC, USA, January 2014.

[14] Y. He, T. Wu, C. Zhang, and Z. Liang, "An improved MIP heuristic for the intermodal hub location problem," *Omega*, vol. 57, pp. 203–211, 2015.

[15] A. Ebrahimnejad, Z. Karimnejad, and H. Alrezaamiri, "Particle swarm optimisation algorithm for solving shortest path problems with mixed fuzzy arc weights," *International Journal of Applied Decision Sciences*, vol. 8, no. 2, pp. 203–222, 2015.

[16] A. Adibi and J. Razmi, "2-Stage stochastic programming approach for hub location problem under uncertainty: a case study of air network of Iran," *Journal of Air Transport Management*, vol. 47, pp. 172–178, 2015.

[17] H. Damgacioglu, D. Dinler, N. Evin Ozdemirel, and C. Iyigun, "A genetic algorithm for the uncapacitated single allocation planar hub location problem," *Computers & Operations Research*, vol. 62, pp. 224–236, 2015.

[18] S. A. Alumur, S. Nickel, F. Saldanha-da-Gama, and Y. Seçerdin, "Multi-period hub network design problems with modular capacities," *Annals of Operations Research*, vol. 246, no. 1-2, pp. 289–312, 2016.

[19] N. Azizi, S. Chauhan, S. Salhi, and N. Vidyarthi, "The impact of hub failure in hub-and-spoke networks: mathematical formulations and solution techniques," *Computers & Operations Research*, vol. 65, pp. 174–188, 2016.

[20] W. Grauberger and A. Kimms, "Airline revenue management games with simultaneous price and quantity competition," *Computers & Operations Research*, vol. 75, pp. 64–75, 2016.

[21] H.-Z. He, "Structural properties in a hub-to-hub network revenue management problem," *Journal of the Operations Research Society of China*, vol. 4, no. 4, pp. 503–516, 2016.

[22] S. A. Alumur, S. Nickel, B. Rohrbeck, and F. Saldanha-da-Gama, "Modeling congestion and service time in hub location problems," *Applied Mathematical Modelling*, vol. 55, pp. 13–32, 2018.

[23] M. Ahmadi, A. Taghavirashidizadeh, D. Javaheri, A. Masoumian, S. J. Ghoushchi, and Y. Pourasad, "DQRE-SCnet: A Novel Hybrid Approach for Selecting Users in Federated Learning with Deep-Q-Reinforcement Learning Based on Spectral Clustering," *Journal of King Saud University-Computer and Information Sciences*, 2021.

[24] E. Korani, A. Eydi, and I. Nakhai Kamalabadi, "Reliable hierarchical multimodal hub location problem: models and Lagrangian relaxation algorithm," *Scientia Iranica*, vol. 27, no. 3, pp. 1525–1543, 2020.

[25] D. D. Čvokić and Z. Stanimirović, "A single allocation hub location and pricing problem," *Computational and Applied Mathematics*, vol. 39, no. 1, pp. 40–24, 2020.

[26] Y. Rouzpeykar, R. Soltani, and M. A. A. Kazemi, "A robust optimization model for the hub location and revenue management problem considering uncertainties," *Iranian Journal of Operations Research*, vol. 11, no. 1, pp. 107–121, 2020.

[27] D. Di Caprio, A. Ebrahimnejad, H. Alrezaamiri, and F. J. Santos-Arteaga, "A novel ant colony algorithm for solving shortest path problems with fuzzy arc weights," *Alexandria Engineering Journal*, vol. 61, no. 5, 2021.

[28] A. Ebrahimnejad, M. Enayattabr, H. Motameni, and H. Garg, "Modified artificial bee colony algorithm for solving mixed interval-valued fuzzy shortest path problem," *Complex & Intelligent Systems*, vol. 7, no. 3, pp. 1527–1545, 2021.

[29] A. A. Sori, A. Ebrahimnejad, H. Motameni, and J. L. Verdegay, "Fuzzy constrained shortest path problem for location-based online services," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 29, no. 02, pp. 231–248, 2021.

[30] L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy Sets and Systems*, vol. 1, no. 1, pp. 3–28, 1978.
[31] B. Baoding Liu and Y. K. Yian-Kui Liu, "Expected value of fuzzy variable and fuzzy expected value models," *IEEE Transactions on Fuzzy Systems*, vol. 10, no. 4, pp. 445–450, 2002.
[32] X. Li and B. Liu, "A sufficient and necessary condition for credibility measures," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 14, no. 05, pp. 527–535, 2006.
[33] B. Liu, *Uncertainty Theory*, Springer-Verlag, Berlin, Germany, 2007.
[34] M. S. Pishvaee, S. A. Torabi, and J. Razmi, "Credibility-based fuzzy mathematical programming model for green logistics design under uncertainty," *Computers & Industrial Engineering*, vol. 62, no. 2, pp. 624–632, 2012.
[35] H. Zhu and J. Zhang, "A credibility-based fuzzy programming model for APP problem,"vol. 1, pp. 455–459, in *Proceedings of the 2009 International Conference on Artificial Intelligence and Computational Intelligence*, vol. 1, IEEE, Shanghai, China, November 2009.
[36] B. Liu, *Uncertainty Theory: An Introduction to its Axiomatic Foundations*, Springer-Verlag, Berlin, Germany, 2004.
[37] M. NoParast, M. Hematian, A. Ashrafian, M. J. T. Amiri, and H. AzariJafari, "Development of a non-dominated sorting genetic algorithm for implementing circular economy strategies in the concrete industry," *Sustainable Production and Consumption*, vol. 27, pp. 933–946, 2021.
[38] S. Routroy, A. Bhardwaj, S. K. Sharma, and B. K. Rout, "Analysis of manufacturing supply chain agility performance using Taguchi loss functions and design of experiment," *Benchmarking: An International Journal*, vol. 25, no. 8, pp. 3296–3319, 2018.
[39] M. Varmaghani, F. Farzadfar, F. Sharifi et al., "Prevalence of asthma, COPD, and chronic bronchitis in Iran: a systematic review and meta-analysis," *Iranian Journal of Allergy, Asthma and Immunology*, vol. 15, no. 2, pp. 93–104, 2016.

Hindawi
Complexity
Volume 2022, Article ID 4831006, 11 pages
https://doi.org/10.1155/2022/4831006

Research Article

The Peer Effects of the Usage of Credit Cards in Rural Areas of China: Evidence from Rural China

Dongliang Cai,[1] Jun Ou,[2] Kefei Han ,[1] and Yang Lyu[3]

[1]School of Finance, Southwestern University of Finance and Economics, Chengdu 610074, China
[2]School of Law, Southwestern University of Finance and Economics, Chengdu 610074, China
[3]School of Commerce, The University of Sydney, Camperdown, NSW 2006, Australia

Correspondence should be addressed to Kefei Han; hankefeijinrong@sina.com

Received 16 February 2022; Accepted 11 May 2022; Published 13 June 2022

Academic Editor: Sheng Du

This paper aims to explore whether the usage of credit cards has peer effects in rural areas of China. The results suggest that the usage of credit cards will be affected by the behavior of other farmers; namely, the usage of credit cards has peer effects in rural areas. We also verify that women, older, and low-academic farmers show stronger peer effects. The results emphasize that, compared with the mass farmers and vulnerable farmers, the usage of elite farmers is more likely to affect the same behaviors of other farmers. Our study contributes to prior literature by giving empirical insights into the consequences of peer effects on the credit card usage of farmers, which conduces to a better understanding of the financial decision-making of farmers. The research may provide a practical implication for the optimal design of policy interventions. On the one hand, banks and other financial institutions can promote innovative financial products in rural areas with the help of peer effects; on the other hand, regulators can regulate farmers' wrong financial behaviors by the social multiplier of peer effects to reduce systemic risks in rural finance.

1. Introduction

The imbalance between urban and rural economic development is a structural and historical issue in China. Insufficient rural economic development has restricted sustainable economic development, which is not conducive to the "rural revitalization" and "poverty alleviation" strategy of the Chinese government in rural areas. Various resources should be fully utilized to support rural economic development. Finance can provide strong energy to the rural economy with the functions of financing and risk management. On the one hand, it can provide financing services for rural economic development and farmers to alleviate insufficient capital. On the other hand, farmers can effectively deal with various risks in life. Financial products are the key hub for financial services and financial policies, which also act as the primary role in the rural financial system. At present, insufficient use of financial products and limited financial market participation are the main reasons why rural inclusive finance does not help farmers effectively. It hinders the development of the inclusive financial system and weakens the role finance plays in poverty alleviation in rural areas. Therefore, we explore the peer effect of the farmers' cognition and usage of financial products in China and analyze farmers' how to make a financial decision in the background of low financial literacy and asymmetric information. Our paper provides a new perspective to promote farmers' financial product usage and offers helpful thoughts for deepening the potential of the existing financial system to enhance farmers' well-being. It has important reference value for accelerating regional coordinated development.

Rural areas are villages based on kinship and marriage in China. Farmers in the same village face similar risks and information status with similar cultural backgrounds and strong group consciousness. Farmers' economic decisions will be affected by others in the same village, showing peer effects. The peer effects mean that an individual's economic decision will be affected by the behavior of others. There is a convergence phenomenon in certain spaces, such as the same community [1, 2]. If individuals have similar cultural backgrounds and value orientations, the peer effects will be more significant when they face the situation of lack of information and high risk [3]. At present, most of the research on rural financial inclusion and financial product participation focuses on market analysis frameworks such as revenue and cost. However, the factor of nonmarket interactions among farmers is neglected. Farmers usually have low financial literacy and a lack of rational decision-making ability. Thus, they will consider others' behavior as a useful offset to the irrationality and show peer effects in their decision-making. Therefore, studying the usage of rural financial products from the perspective of peer effects is of great significance. It is good for the development of rural inclusive finance and the promotion of innovative financial products.

The credit card is a financial inclusive tool commonly used in the current economy, whose overdraft function can be regarded as short-term financing for the holders. With flexible repayment methods such as minimum repayment and installment repayment, credit cards can act as a convenient and short-term financing tool. If farmers use credit cards reasonably, it will effectively ease their cash flow pressure and overcome financing difficulties. Compared with most financial products, credit cards are commonplace for urban residents, while they are still unacquainted for most farmers. Few of them can utilize them well, and some even do not hear about it.

Based on the Financial Education Development Foundation survey data on 137 villages in four southwestern provinces of China in 2018, this paper analyzes the usage of credit cards in rural areas from the perspective of peer effects. The probit model is used to explore the peer effects in the usage of financial products and to analyze the peer effects of different types of farmers. In addition, this paper subdivides the peer effects variables (PEV) into elite farmers, mass farmers, and vulnerable farmers to explore the mutual influence mechanism.

2. Literature Review

Since the advent of credit cards, credit cards have been a financial product commonly used by residents, and research on the holding and usage of credit cards (UCC) has also been abundant. Combining the previous research on credit card holding and usage, it is found that, under the framework of market analysis, scholars discuss the factors that affect the use of credit cards from the perspective of personal characteristics and family characteristics, such as resident income, age, education, gender, ethnicity, occupation, marriage, debt situation, financial knowledge, risk appetite, family size, family income, and related wealth accumulation.

Personal characteristics have a greater influence on UCC, and the influence of age and education on credit card holding has been widely recognized by scholars. Most scholars believe that credit card holding is negatively correlated with age [4, 5]. Loke et al. [6] pointed out that credit card holding has an inverted "U" relationship with age, the tendency to hold cards during the 35–56 age and the number of cardholders are higher than those of other age groups, and card holding tendency weakens with age. Existing research shows that there is a significant positive relationship between the education level of residents and credit card holding, which is manifested in two aspects. First, the higher the education level of residents and financial knowledge of credit cards and risk identification and the stronger the coping ability, the higher the number of credit cards held and the frequency of use. Second, the credit management and credit records of highly educated people are generally better, making it easier to obtain credit cards from banks [7]. At present, the impact of residents' income, gender, and occupation on UCC has been controversial in the academic circle. Even though residents' income significantly affects UCC, the agreement has been reached in academia. However, whether the income of residents is positive or negative affects the issue of credit card holding, which is more controversial. Most scholars believe that income will positively affect personal credit card holding and usage [8, 9]. Most literature finds that women have more credit cards than men and UCC more frequently [10]. Tan et al. [11] believe that there is no significant relationship between UCC and individual gender. B. S. Divya [9] pointed out that the gender difference in middle-income groups is greater, and this conclusion eased the differences to some extent.

Some scholars have studied the impact of UCC from a racial perspective and found that race will have a certain impact on UCC. For example, African Americans are more inclined to use credit cards [12]. Studies have shown that residents' financial knowledge, marital status, occupational characteristics, length of work, and job stability all affect personal UCC [5, 13, 14]. Existing literature mainly studies the impact of family characteristics on credit card decision-making from the perspectives of family structure, family population size, family wealth attribution, and asset allocation. Shen et al. [15] pointed out that the increase in family size will significantly reduce the tendency to hold cards, while other scholars agree that the family size is positively related to residents' holding and UCC [16]. Sharpe et al. [14] pointed out that the size of the family population does not have a significant impact on credit card holding decisions, and the impact of household wealth accumulation on credit card holding decisions is not significant.

For the study of credit card decision-making, most of the literature is limited to individual self-characteristics, and it is less related to the influence of other individuals, which is inconsistent with the objective reality. Peer effects originate from the theory of social networks. The theory holds that the exchange of information between individual behaviors will constitute a complex network of social relations, which makes individual behavior decisions not independent but mutually influenced. For example, individual behavior decisions are often influenced by family members, neighbors, friends, or colleagues [17, 18]. Individual decision-making not only is affected by its characteristics but also tends to change due to the behavior of groups with similar characteristics, that is, "same groups" [19–22]. The "peer effects" studied in this paper mean that the interaction between people will directly affect the performance of related actors [2]. In the early days, research on peer effects was mainly distributed in the field of pedagogy. Through different perspectives, scholars have studied peer effects on students' academic performance, noncognitive ability, and preference [23]. At present, research on peer effects extends to the fields of economics, management, and finance, involving company innovation, financial arrangements, and investment and financing behavior [20, 24]. It also includes individual stock market participation, gaming participation, entrepreneurial behavior, and donation behavior [25, 26]. In addition, some scholars have shifted their focus from the general performance of actors to the generation mechanism of violations and put forward suggestions to reduce negative social multipliers, such as personal crimes and corporate violation decisions [2, 27, 28]. Likewise, some scholars have proposed the impact of the peer effect on social structure from a more macroperspective. For example, Margaretic et al. [27] proposed that the peer effect strengthens the identity of groups, especially vulnerable groups, and thus strengthens social segmentation. Lu and Zhang [2] proposed that the peer effect exacerbates the inequities such as the gap between rich and poor, accumulation of human capital, and consumption of public goods.

In the above studies, although there have been more studies on credit cards, less attention has been paid to the peer effects on UCC, that is, whether the increase in peers will lead to an increase in the usage of financial products. At present, most research is to analyze its usage behavior within the framework of market analysis. Few scholars have studied farmers' credit card usage from the perspective of the peer effects. However, rural areas are always showing stronger peer effects of economic decisions. In addition, studies have proved that the behaviors of individuals will affect each other, but few studies have further analyzed which individuals are more susceptible to social interactions and which group behaviors are more likely to affect the behavior of others, thus making research lack reference value in practice. Therefore, to better explain the rural financial phenomenon, it is necessary to consider the mutual influence of farmers in economic decision-making.

3. Hypothesis Development

Studies have shown that when personal information and knowledge are lacking, they tend to take the initiative to seek help from social networks such as relatives, friends, neighbors, and colleagues. On the other hand, social networks contribute to the transmission of information between individuals and companies and affect investor sentiment, cognition, decision-making, and group behavior [29]. According to the Bayesian social learning theory model, social learning is divided into communicative and observational learning [30]. When farmers face an unknown financial product such as a credit card, through the social circle formed by kinship and marriage, farmers can consult and learn related knowledge. Compared with external information channels, farmers trust information obtained from relatives and friends more. The credit cognition of farmers in the same village will overflow through social networks, communication, and observing. That will enhance the credit cognition of farmers.

Otherwise, the UCC of other farmers will increase the positive utility of the UCC of the farmers and reduce the negative utility. There are three specific channels as follows. First, the rural village in China is mainly composed of kinds of social relations based on marriage, blood relationships, and kinship. Farmers have the same cultural background and strong collective concept. Similar behaviors of other farmers in the same village are important indicators for their decision-making. When farmers take the first action, this "alternative" behavior makes farmers need to bear the negative effects of the deviation from the group and collectives as "outstanding birds." So farmers tend to choose similar behaviors in economic decision-making to keep the consistency with other farmers in the same village [3]. Therefore, the more the farmers use credit cards, the less the negative utility the user will feel from the deviation. Second, according to the theory of social learning, farmers can understand the potential benefits and possible losses by observing the UCC of others in the community. Thereby, it will alleviate the suppression of UCC caused by "fuzziness aversion" [28]. Third, by observing and imitating the credit card usage of other farmers in the same village, the farmer can learn the rules, skills, and procedures about credit cards. It will promote UCC in rural areas. In addition, the bank's credit card business has the characteristic of decreasing average costs in rural areas. The more the farmers use credit cards, the lower the cost of bank credit card supply. Thereby, it will increase the willingness of banks to provide credit cards.

Hypothesis 1. There is a peer effect on the UCC of farmers. That is, the more the farmers use credit cards in the village, the more they tend to use credit cards.

The UCC of farmers will be affected by other farmers, but different groups will have different impacts on them. Mainly through the three channels of learning, imitation, and anchoring effect, the economic decision of farmers is influenced by similar economic behaviors of other farmers. First, elite farmers in the same village (i.e., farmers with high education, high income, and high social capital) have stronger social networks and more positive incentives, so UCC of elite farmers has greater reference value and stronger demonstration effect. Therefore, farmers usually prefer UCC of elite farmers as a target to observe and imitate, thereby overcoming the vague aversion caused by lack of knowledge and lack of information. In addition, the economic behavior of elite farmers is usually more frequently concerned and referenced by other farmers. Compared with elite farmers, although the behaviors of mass farmers and vulnerable farmers will also affect the UCC of other farmers through the same channels, their demonstration effects are weaker and less concerned, especially for vulnerable farmers (i.e., farmers with low education, low income, and low social capital), and are usually far from the reference view. So, UCC of elite farmers in the same village can affect the UCC of other farmers more than mass farmers and vulnerable farmers.

Hypothesis 2. There are peer effects on the UCC of farmers, while the UCC of the elite farmers can affect the behavior of other farmers more.

4. Data and Methodology

4.1. Sample Selection and Data. The data comes from the Financial Education Development Foundation survey data on 137 villages in four southwestern provinces of China in 2018. The survey samples were selected in the southwest region, and the three-level stratified sampling method was conducted in cities, counties, and villages in each province in order of high, medium, and low per capita GDP. The selected samples were universal and representative. The survey covered four provinces in Sichuan, Yunnan, Guizhou, and Tibet and 16 cities, including Chengdu, Suining, Guiyang, Zunyi, Chuxiong, Yuxi, and Lhasa, covering 137 villages in 25 districts and counties, and finally obtained 2925 samples. Most notably, this survey inquiry was made face to face by electronic questionnaire loading in Android Pad with a quality control system that can record all the sound of the survey with every respondent and automatically deliver it back to the control terminal in the end. The control terminal can recognize and correct all the misconduct timely, which ensures the objectivity, authenticity, and standard operation of the survey. Otherwise, in the postaudit stage, 20% of the total samples were randomly selected to check by listening to the quality control system recording, and the other 20% were selected to do return tests by telephone. Finally, 90% of the samples passed the test.

4.2. Peer Effects Identification. If we want to identify peer effects from similar nonmarket interactions by empirical methods, the endogenous interactions, contextual interactions, and correlation effects need to be effectively identified [17]. First, endogenous interaction is the peer effects involved in this paper. It is the interpretation of a specific member by the performance of other members within the same group, emphasizing the influence and interaction between members. Most empirical studies identify the peer effects with the peer variables, which is the average of members in the same group. Second, contextual interaction means that a person's behavior is related to the exogenous characteristics of the group. Because they share common external characteristics such as income, education, and age, the economic decisions of individuals and others show convergence. The last is the correlation effect; that is, the members in the group with similar personal characteristics tend to make the same choice. Based on the research of Brock and Durlauf [31], our paper introduces the common external features of sample farmers' income, education, age, and so on into the model as a control variable to solve the situational interaction problem. Meanwhile, we add the peers' gender, age, education, social capital, and income into basic regression to deal with the correlation effects.

In addition, there are usually reflection problems, selection biases, and attrition biases in the empirical estimation of peer effects. We use the IV-probit model to solve the reflection problems referring to the existing literature [2, 26]. And it can also overcome the problems of measurement error, reflectivity, and leakage error. Since China's unique household registration system restricts the free movement of the population, the selection bias caused by self-selection has little effect on the results of the estimation of peer effects.

4.3. Empirical Models. To test the hypothesis, this paper constructs the probit model to investigate whether there are peer effects of UCC of farmers.

$$\text{ccus}_i^c = \beta_0 + \beta_1 \text{pccu}_{-i}^c + \beta_2 X_i^c + \beta_3 Y_{-i}^c + \beta_4 Z_i^c + \varepsilon_i^c, \tag{1}$$

where pccu_{-i}^c is the peer effects variable of credit card usage (PEV); ccus_i^c is the credit card usage dependent variables, which are the UCC of the farmers. The coefficients α_1 are the key indicators to identify whether UCC has peer effects. X_i^c is a group of control variables related to individual characteristics, Y_{-i}^c is a group of peer variables related to the personal characteristics of other interviewed farmers in the same village, and Z_i^c is a group of variables related to the characteristics of villages where farmers are located.

Meanwhile, we are also interested in whether UCC of the farmers from different groups in the same village shows different results; we estimate models (2).

$$\text{ccus}_i^c = \eta_0 + \eta_1 X_1^c + \eta_2 X_2^c + \eta_3 X_3^c + \eta_4 M_i^c + \eta_5 Y_{-i}^c + \eta_6 Z_i^c + \varepsilon_i^c. \tag{2}$$

In model (2), X_1^c, X_2^c, and X_3^c are the main explanatory variables, expressed by pccu_{-i}^{cj} (PEV of the elite farmer's UCC), $\text{pccu}_{-i}^{\text{cm}}$ (PEV of the mass farmer's UCC), and pccu_{-i}^{cw} (PEV of the vulnerable farmer's UCC). M_i^c is the personal characteristics of the farmers, Y_{-i}^c is a group of peer variables related to the personal characteristics of other interviewed farmers in the same village, and Z_i^c is a group of variables related to the characteristics of the village where the farmers are located. In all our tests, we include province-fixed effects.

4.4. Variable Definition and Description. The definition and description of the variables involved are shown in Table 1. Explained variables: UCC is defined as whether farmers use a credit card. The question in the questionnaire is "Have you used a credit card? (1) used; (2) not used." If the rural residents use credit cards, the value is 1; otherwise, the value is 0.

Explanatory variable: the core explanatory variable in this paper is the PEV of UCC of the rural residents. The PEV of UCC is defined as the average value of UCC of other farmers in the same village, except for the sample. We calculate the PEV of UCC as follows:

$$\text{pccu}_{-\text{i}}^{\text{c}} = \sum \frac{\left(\text{ccus}_\text{j}^c - \text{ccus}_\text{i}^c\right)}{(\text{N}-1)}, \tag{3}$$

Table 1: Variable description.

	Abbrev	Variable	Definition
Explained variables	ccus	Usage of credit card	Whether farmers use credit cards, the corresponding question is "Do you use credit card? 1. Not used; 2. Used." If you choose "1," the value is 0; otherwise, it is 1.
Explanatory variables	pccu	Peer variables of usage of credit card	The average usage level of credit cards of other farmers in the same village except the sample farmers. The method is to add and average usage of credit cards of other farmers except the sample farmers.
Individual characteristics	tpay	Total household expenditure	The total household expenditure of the sample farmers, "What is the average total expenditure of your household in the past three years?" Unit: Yuan.
	tincome	Total household income	The household income of the sample farmers, "What is the average total income of your household in the past three years?"
	age	Age	The age of the sample farmers is corresponding to the question "What is your age?"
	sex	Sex	The gender of the sample farmers corresponds to the question "What is your gender? 1. Female; 2. Male." If you select "1," the value is 0; otherwise, it is 1.
	edu	Years of education	The education years of the sample farmers, corresponding to the question "What is your education level?" Calculate the education years of the sample farmers according to their education years.
	sc	Social capital	For the social capital of the sample farmers, the corresponding questions here are "Do you or your family work in government departments? 1. None; 2. I am; 3. Family is in," "you or your family work in financial institutions? 1. None; 2. I am; 3. Family is in," and "you or your family is a village cadre? 1. None; 2. I am; 3. Family is." If you choose "1" for each question, the value is 0. If "2" is selected, the value is 1. If "3" is selected, the value is 2. Add the three question assignments to calculate the social capital of the sample farmers.
	tp	Family size	The household population of the sample farmers corresponds to the question "How much is your family?"
	loandemand	Loan demand	For the loan demand of the sample farmers, the corresponding question is "Do you have a loan demand? 1. No; 2. Yes." If you select "1," the value is 0; otherwise, it is 1.
Neighborhood characteristics	peersex	Peer variables of gender	The average gender status of other farmers in the same village except the sample farmer. The method is to average the gender of other interviewed farmers other than the sample farmer.
	peerage	Peer variables of age	The average age of other farmers in the same village except the sample farmer. The method is to add up and average age of the other farmer other than the sample farmer.
	peersc	Peer variables of social capital	The average social capital of other farmers in the same village except the sample farmer. The method is to add up and average the social capital of other interviewed farmers except the sample farmers.
	peeredu	Peer variables of years of education	The average years of education of other farmers in the same village except the sample farmers. The method is to add up and average the years of education of other interviewed farmers except the sample farmers.
	peertincome	Peer variables of family income	The average total income of other farmers in the same village except the sample farmer. The method is to add up and average the total income of other farmers other than the sample farmer. Unit: Yuan.
Community characteristics	cuntraffic	Village traffic situation	The traffic situation of the village where the sample is located. The corresponding question is the interview of the village committee director in the community questionnaire. The corresponding question is "What is the traffic situation of the village? 1. Not good; 2. Good." If "1" is selected 0, otherwise 1.
	cunwifi	Network status	The wireless signal coverage status of the village where the sample is located. The corresponding question is the interview of the village committee director in the community questionnaire section. The corresponding question is "What is the network status of the village? 1. Not good; 2. Good." If "1" is selected 0, the value is 0; otherwise, it is 1.
	pdistance	Distance from the provincial government	The distance from the village committee where the sample is located to the provincial people's government, measured by Google Maps, unit: km.

where $ccus_i^c$ indicates the usage of credit cards of the sample, C indicates the village sample belongs to, and N indicates the total number of farmers in the village.

Control variables: we introduce three kinds of control variables to more accurately calculate the peer effects on the usage of credit cards. The first kind is the characteristics of the sample, including education, age, gender, social capital, family population, income, expenditure, loan demand, and other variables. The second is neighbors' characteristics, including the average value of education, age, gender, social capital, and income of other farm households except for the sample. These variables are used to deal with situational interaction. The third is village variables. Aiming to control the influence of the correlation effect, we introduce the traffic status, WIFI coverage status, and the distance of the village from the committee location of the provincial government.

Table 2 gives the descriptive statistics of the variables.

5. Peer Effects of Regressions

The baseline regression is listed as follows. Columns (1)–(3) of Table 3 report the regression results of the relationship between PEV and UCC. Column (2) includes individual characteristics, group characteristics, and community characteristics in column (1). Column (3) includes regional characteristics. The results show that the coefficient of pccu in all columns is significantly positive, which indicates that farmers' UCC has peer effects. Therefore, Hypothesis 1 is proved.

In terms of control variables, the gender, years of education, social capital, and family income of rural family policymakers have a significant positive impact on UCC of rural residents. The age and financing experience of family decision-makers have a significant negative impact on UCC of rural residents.

6. Instrumental Variable Regressions

This part will use the instrumental variable method to solve the endogenous problem and better estimate the impact of UCC [26, 32].

We use the average value of third-party payment usage (pthirdpay) of other farmers and the average value of mobile banking usage (pphonebank) of other farmers as instrumental variables. Because third-party payments such as Alipay and mobile banking usually have credit card product introductions, credit card processing ports, and UCC and repayment services in the login interface, thereby increasing UCC of the farmers, they meet the requirements of relevance. pthirdpay and pphonebank will affect the UCC of others in the same village, but it will not affect the sample farmer's UCC. At the same time, it is not related to the current random disturbance item and meets the requirements of exogenous. Therefore, it is reasonable to take pthirdpay and pphonebank as instrumental variables for PEV of UCC to meets the "exogenous" requirement. Columns (1)–(4) of Table 4 report the estimated results using instrumental variables, and columns (2) and (4) add the regional control in columns (1) and (3), respectively. The results show that PEV of UCC is positive and significant, which is consistent with the results in Table 3. The coefficient of the first stage instrumental variable is significantly positive, and the p-value is less than 0.05, thus proving the effectiveness of the instrumental variable; the F-values are all bigger than 15, thus rejecting the assumption that it is a weak instrumental variable.

7. Robustness Checks

In the existing research, some scholars have toke human expenditure as an indicator to indirectly measure the characteristics of the farmer's peer effects [33]. However, this indicator has the following defects. On the one hand, the expenditure does not promote mutual learning and communication between farmers. On the other hand, observation between farmers is an important channel for imitation and learning. Human expenditure cannot accurately measure this behavior. Therefore, this paper uses the questionnaire question "Will your financial decision be affected by the financial decisions of others in the same village? 1. Yes; 2. No" instead of PEV of UCC in the above OLS regression. If the farmer chooses "Yes," it will be assigned a value of 1; otherwise, it will be assigned a value of 0, and this variable is defined as replace. The estimation results in columns (1)-(2) of Table 5 show that the coefficients are still significantly positive, which shows that the impact of PEV of UCC is robust.

Otherwise, we take the simulated sampling test methods to prove the robustness and accuracy of baseline regression [26], which is constructing the virtual community. Then, calculate the false peer effect variables and reestimate the baseline regression. If the coefficient of false peer effect is not significant, it will prove the robustness of the above regression result. The simulation sampling steps are as follows. First, randomly select sample farmers from a different village in the same county. It is the virtual community, and then calculate the "peer variables" of the virtual community according to the above calculation method. That is the false peer effect (fpccu). Second, replace the explanatory variable (pccu) in the baseline regression with the false peer effect variable (fpccu) and regress again. Finally, observe whether the coefficients of the virtual peer variables are significant. Columns (3)-(4) of Table 5 report the simulated sampling test results. Column (3) is the result without the fixed effects, and column (3) is with the fixed effects. Both the coefficients of false peer effect variable are not significant, which shows that the baseline regression is robust.

Finally, we divide all farmers into large-scale family groups and small-scale groups according to the family size. The results are in columns (5)-(6) of Table 5. No matter the large family size group or a small family size group, the coefficients of PEV are all significantly positive, which is consistent with the baseline regression. The result further proves the robustness of the results again.

8. Heterogeneity Analysis

This paper makes group regression according to the gender, education level, and age of the decision-makers in the

Table 2: Variable description.

Variables	Mean	Median	Standard deviation	Minimum	Maximum	Obs
ccus	0.132	0	0.339	0	1	2435
pccu	0.132	0.105	0.099	0	0.538	2435
tpay	37,216	30000	26,376	5000	100000	2435
tincome	54,053	40000	58,346	2500	300000	2435
age	48.90	50	13.13	25	71	2435
sex	0.591	1	0.492	0	1	2435
edu	7.868	9	4.085	0	19	2435
sc	0.962	0	1.456	0	5	2435
tp	4.818	5	2.197	1	12	2435
loandemand	0.721	1	0.448	0	1	2435
peersex	0.590	0.577	0.169	0.0400	1	2435
peerage	48.90	49.60	6.458	23.58	63.27	2435
peersc	0.992	0.950	0.601	0	3.556	2435
peeredu	7.872	7.880	1.616	3.143	12.56	2435
peertincome	57894	52000	29777	9043	240000	2435
cuntraffic	0.824	1	0.381	0	1	2435
cunwifi	0.929	1	0.256	0	1	2435
pdistance	147.9	133.1	106.0	10.90	408	2435

Table 3: Estimation results of the peer effects.

Variables	ccus (1)	ccus (2)	ccus (3)
pccu		4.558^{***} (12.923)	4.466^{***} (10.554)
tpay		≤0.001 (1.093)	≤0.001 (1.042)
tincome		$\leq 0.001^{**}$ (2.404)	$\leq 0.001^{**}$ (2.438)
age		-0.006^{*} (−1.840)	-0.006^{*} (−1.828)
sex		0.131^{*} (1.751)	0.129^{*} (1.724)
edu		0.038^{***} (3.334)	0.039^{***} (3.339)
sc		0.033 (1.301)	0.034 (1.323)
tp		0.029^{*} (1.778)	0.029^{*} (1.809)
loandemand	4.484^{***} (15.522)	-0.162^{**} (−2.175)	-0.164^{**} (−2.215)
peersex		−0.140 (−0.629)	−0.155 (−0.686)
peerage		0.007 (1.089)	0.006 (0.756)
peersc		−0.024 (−0.323)	−0.010 (−0.119)
peeredu		−0.025 (−0.806)	−0.036 (−0.996)
peertincome		-0.001^{**} (−1.983)	−0.001 (−1.483)
cuntraffic		−0.032 (−0.320)	−0.023 (−0.226)
cunwifi		0.028 (0.183)	0.055 (0.356)
pdistance		≤0.001 (1.329)	≤0.001 (0.702)
Constant	-1.819^{***} (−32.390)	-2.137^{***} (−4.392)	-2.024^{***} (−3.228)
Regional control	No	No	Yes
Observations	2435	2435	2435
Pseudo R-squared	0.104	0.150	0.151

* Note: *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The number in parenthesis is T value. The same goes for subsequent tables.

sample families to observe whether the peer effects of farmers' UCC will show differences with gender, education level, and age under different control variables.

First, column (1) of Table 6 is the empirical results of the peer effects of male family decision-makers. Column (2) is the empirical results of the peer effects of female family decision-makers. The results show that, compared with the male family decision-maker, the male family decision-maker has stronger peer effects on UCC. The reason may be because the group effect of credit card use breadth plays a role more through group reference and social comparison, while male farmers have a wider social circle and a stronger tendency to seek common ground in social comparison.

Second, the age of family decision-makers also has a certain influence on UCC. Columns (3) and (4) of Table 6 show the grouping regression results according to the age of family decision-makers. Column (3) is the empirical results of UCC for the elderly group who are older than 45. And column (4) is the empirical results of UCC for the younger age group whose age is below 45. The results show that, compared with the family decision-maker who is a young

Table 4: Instrumental variable estimation results of the peer effects of residents.

Variables	ccus (1)	ccus (2)	ccus (3)	ccus (4)
pccu	3.337*** (4.965)	2.348** (2.320)	3.428*** (4.948)	1.699*** (3.577)
Control variables	Yes	Yes	Yes	Yes
Regional control	No	Yes	No	Yes
Constant	−1.690*** (−15.549)	−1.500*** (−2.726)	−1.703*** (−15.336)	−1.402** (−2.565)
F-value of first stage	743.79	68.29	932.35	64.48
pthirdbank	0.486*** (27.273)	0.410*** (20.476)		
pphonebank			0.42*** (30.534)	0.340*** (18.964)
Observations	2435	2435	2435	2435

Table 5: Robustness testing.

Variables	ccus (1)	ccus (2)	ccus (3)	ccus (4)	ccus (5)	ccus (6)
Replace	0.145** (2.161)	0.132* (1.946)				
fpccu			−0.211 (−0.604)	−0.143 (−0.358)		
pccu					4.110*** (9.094)	5.300*** (9.166)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Regional control	No	Yes	No	Yes	No	Yes
Constant	−1.440*** (−3.315)	−1.073* (−1.828)	−1.265*** (−2.679)	−1.002* (−1.696)	−1.791*** (−2.701)	−2.853*** (−3.728)
Pseudo R-squared	0.066	0.092	0.063	0.091	0.133	0.177
Observations	2435	2435	2435	2435	2435	2435

Table 6: Estimation results of the peer effects of family decision-maker's gender.

Variables	ccus Male (1)	ccus Female (2)	ccus Elder (3)	ccus Young (4)	ccus More education (5)	ccus Less education (6)
pccu	4.757*** (10.646)	4.629*** (7.868)	3.941*** (8.171)	5.148*** (9.818)	3.856*** (8.967)	6.005*** (9.596)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Regional control	Yes	Yes	Yes	Yes	Yes	Yes
Constant	−1.953*** (−3.248)	−2.615*** (−3.091)	−2.393*** (−3.569)	−2.312*** (−2.757)	−2.645*** (−4.053)	−1.669** (−2.083)
Observations	1,439	996	1388	1047	1321	1114
Pseudo R-squared	0.154	0.163	0.131	0.173	0.120	0.202

group, the elderly group shows stronger peer effects. On the one hand, the social tendency and social frequency of younger farmers are stronger. On the other hand, it is because young farmers have less credit card knowledge, have more open hearts, and are more likely to be affected by the emotions of others.

Finally, we grouped the total sample with the education. If the sample has more than 9 years of educational experience, it is classified as the highly educated group. Or it belongs to a low-academic group. Columns (5) and (6) of Table 6 show the grouping regression results according to the education of family decision-makers. Column (5) is the empirical results of the highly educated group, and column (6) is for the low-academic group. Compared with the family decision-maker with the high education, the low-education farmer shows stronger peer effects. It is because farmers in the low-education group have weaker financial knowledge reserve, inductive analysis, and autonomous learning ability, so they prefer to acquire relevant knowledge and information through communication and imitation with farmers in the same village.

9. Peer Effects from Different Groups

From the perspective of practice, our research should be more in-depth. Therefore, this section will subdivide the peer variables of farmers' UCC to explore which groups are more likely to affect other farmers in the same village. Farmers' decisions will be affected by the same or similar behaviors of other farmers. It can be seen that different groups in the village have different influences on the economic decision-making of farmers. Aiming to explore the

Table 7: Estimation results of the peer effects of farmer types.

Variables	ccus (3)	ccus (4)
$pccu_{-i}^{cj}$	3.298*** (3.433)	3.117*** (3.186)
$pccu_{-i}^{cm}$	2.255*** (8.086)	2.274*** (7.592)
$pccu_{-i}^{cw}$	1.385*** (5.642)	1.222*** (4.209)
Control variables	Yes	Yes
Regional control	No	Yes
Constant	−2.101*** (−4.267)	−1.908*** (−2.967)
Observations	2435	2435
Pseudo *R*-squared	0.138	0.139

impact of which farmers on other farmers' UCC is the most, this part decomposes the peer variables into elite farmers' peer variables, mass farmers' peer variables, and vulnerable farmers' peer variables.

Three variables, education, social capital, and income, were selected to construct the elite index by factor analysis. According to the ranking of the index value, the first quarter of sample farmers with financial index scores are defined as elite farmers (j), the last quarter are defined as vulnerable farmers (w), and other farmers are defined as mass farmers (m). The calculation is as follows. If the farmer is a nonelite farmer, the value is $N = n_j$; otherwise, the value is $n_j - 1$, where j represents the sample farmers and n_j is the number of elite farmers in village c. Two variables $ccus_j^c$ represent the UCC of elite farmers, respectively. PEV of UCC of elite farmers in the same village are expressed as

$$pccu_{-i}^{cj} = \sum_{j \neq i}^{n_j} \frac{ccus_j^c}{N}. \quad (4)$$

Similarly, if the farmer is a nonvulnerable farmer, the value is $N = n_w$; otherwise, the value is $n_w - 1$, where w is the sample farmer and n_w is the number of vulnerable farmers in the village c. Two variables $ccus_w^c$ represent the UCC of vulnerable farmers, respectively. PEV of UCC of vulnerable farmers in the same village are expressed as

$$pccu_{-i}^{cw} = \sum_{w \neq i}^{n_w} \frac{ccus_w^c}{N}. \quad (5)$$

If the farmer is a nonmass farmer, the value is $N = n_m$; otherwise, the value is $n_m - 1$, where m is the sample farmers and n_m represents the number of mass farmers in the village. Two variables $ccus_m^c$ represent the UCC of mass farmers, respectively. PEV of UCC of mass farmers in the same village are expressed as

$$pccu_{-i}^{cm} = \sum_{m \neq i}^{n_m} \frac{ccus_m^c}{N}. \quad (6)$$

In Table 7, regression results show that the peer variables of elite farmers, mass farmers, and vulnerable farmers all significantly promote the UCC of farmers. In contrast, the peer variables of elite farmers have the strongest influence in promoting other farmers' UCC. Through the above analysis, Hypothesis 2 is proved.

10. Conclusion

In this paper, we explore whether the usage of financial products has peer effects in rural areas of China. From the perspective of gender, age, and education, we analyze the heterogeneity of the peer effects. According to the grouping of elite farmers, mass farmers, and vulnerable farmers, we report peer effects of farmer type.

Our empirical evidence reveals that farmers' usage of credit cards shows the peer effects; that is, farmers' credit card usage will be significantly affected by other farmers' credit use in the same village. The peer effects of family decision-makers for female farmers are stronger than those of male decision-makers; family coordinators for low-academic farmers are stronger than those of high-education farmers; peer effects of household decision-makers for older farmers are stronger than those of family decision-makers or young farmers. Compared with the mass farmers and vulnerable farmers, the role of elite farmers is more obvious. It is more likely to affect the same behavior as other farmers. To a certain extent, this conclusion can be analogized to the participation of farmers in other financial products in decision-making; that is, there are peer effects in the participation of farmers in financial products.

We will provide a reference for commercial banks and other financial institutions. When promoting electronic inclusive finance and other innovative financial products in rural areas, the behavioral characteristic is taken into consideration. Through the promotion of differentiated key populations, the cost of promotion can be greatly saved, and the effect of supporting agriculture can be rapidly expanded. The government should pay attention to the peer effect of farmers, and the policy effect can be expanded through the social multiplier effect. Financial institutions or related organizations should pay attention to the peer effect and effectively identify vulnerable groups and groups that are easily affected by others, thereby expanding the financial education investment effect. Because the bad financial behavior in rural areas will affect the financial behavior of other farmers through the peer effects, the government and financial institutions can manage the bad financial behavior by cutting off the peer effects of bad behavior in the countryside.

Admittedly, this study has several limitations. First, the data used were from rural areas in west China. Although the data has qualified validity and reliability, future studies would benefit from samples from other regions to better understand the peer effects of farmers in the usage of credit cards. Second, the actual channel of how regional peer effects work is a very interesting yet challenging question. Due to data availability, we fail to further study the mechanism of how the peer effects work. Some mediators should be considered in future research. Besides, peer effects always work through social networks, so it is important to identify the key points of these networks. That will be an interesting and significant topic for further study.

Data Availability

The data used to support the findings of this study have not been made available because the data belong to the People's Bank of China confidential data. According to the Law of the People's Republic of China on the Protection of Secrets and relevant copyright agreements of the People's Republic of China, the authors may use the data for research with the approval of the People's Bank of China but shall not disclose the source data to any third party.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

The authors contributed equally to this work.

References

[1] H. X. Wen, C. Wang, and P. Y. Nie, "Acceleration of rural households' conversion to cleaner cooking fuels: the importance and mechanisms of peer effects," *Energy Policy*, vol. 154, no. 9, Article ID 112301, 2021.

[2] M.. Lu and S. Zhang, "Non-market interaction and sorting: a literature review," *Economics (Quarterly)*, vol. 3, pp. 991–1020, 2007.

[3] F. Gioia, "Peer effects on risk behaviour: the importance of group identity," *Experimental Economics*, vol. 20, no. 1, pp. 100–129, 2016.

[4] W. M. Lim, W. K. Ng, J. H. Chin, and A. W. X. Boo, "Understanding young consumer perceptions on credit card usage implications for responsible consumption," *Contemporary Management Research*, vol. 10, no. 4, pp. 287–302, 2014.

[5] J. M. Lee, "Self-justification, self-control, and credit card usage," *SSRN Electronic Journal*, vol. 59, 2013.

[6] Y. J. Loke, S. T. Yen, and A. K. G. Tan, "Credit card averters and amassers: evidence from Malaysia?" *Asian Economic Journal*, vol. 25, no. 4, pp. 397–412, 2011.

[7] G. Themba and C. B. Tumedi, "Credit card ownership and usage behaviour in Botswana," *International Journal of Business Administration*, vol. 3, no. 6, 2012.

[8] J. Stavins, "Unprepared for financial shocks:Emergency savings and credit card debt," *Contemporary Economic Policy*, vol. 39, no. 1, pp. 59–82, 2021.

[9] B. S. Divya, "A study on credit card usage behaviour in mandya city," *Asian Journal of Management*, vol. 9, no. 3, pp. 1123–1125, 2018.

[10] A. M. Hancock, B. L. Jorgensen, and M. S. Swanson, "College students and credit card use: the role of parents, work experience, financial knowledge, and credit card attitudes," *Journal of Family and Economic Issues*, vol. 34, no. 4, pp. 369–381, 2013.

[11] A. K. G. Tan, S. T. Yen, and Y. J. Loke, "Credit card holders, convenience users and revolvers: a Tobit model with binary selection and ordinal treatment," *Journal of Applied Economics*, vol. 14, no. 2, pp. 225–255, 2011.

[12] P. J. Fisher, "Differences in credit card use between White and Hispanic households," *Journal of Financial Counseling and Planning*, vol. 27, no. 2, pp. 199–211, 2016.

[13] O. Gorbachev and M. J. Luengo-Prado, "The credit card debt puzzle: the role of preferences, credit access risk, and financial literacy," *The Review of Economics and Statistics*, vol. 101, no. 2, pp. 294–309, 2019.

[14] D. L. Sharpe, R. Yao, and L. Liao, "Correlates of credit card Adoption in urban China," *Journal of Family and Economic Issues*, vol. 33, no. 2, pp. 156–166, 2012.

[15] H. B. Shen, H. Huang, and L. Liao, "Study on the cardholder's overdraft behavior in Chinese credit card market," *Statistical Research*, vol. 30, pp. 61–67, 2013.

[16] C. E. C. Gan, D. A. Cohen, B. Hu, M. C. Tran, W. K. Dong, and A. N. Wang, "The relationship between credit card attributes and the demographic characteristics of card users in China," *International Journal of Bank Marketing*, vol. 34, no. 7, pp. 966–984, 2016.

[17] C. F. Manski, "Economic analysis of social interactions," *The Journal of Economic Perspectives*, vol. 14, no. 3, pp. 115–136, 2000.

[18] L. Shi, L. Y. Chen, and Y. S. Li, "Peer effects of regional economic growth: empirical evidence from data of Chinese cities," *Geographical Research*, vol. 4, pp. 853–864, ,2020.

[19] A. Gamba, E. Manzoni, and L. Stanca, "Social comparison and risk taking behavior," *Theory and Decision*, vol. 82, no. 2, pp. 221–248, 2016.

[20] P. J. Gortner and J. J. van der Weele, "Peer effects and risk sharing in experimental asset markets," *European Economic Review*, vol. 116, pp. 129–147, 2019.

[21] S. Kedia, K. Koh, and S. Rajgopal, "Evidence on contagion in earnings management," *The Accounting Review*, vol. 90, no. 6, pp. 2337–2373, 2015.

[22] K. Park, I. Yang, and T. Yang, "The peer-firm effect on firm's investment decisions," *The North American Journal of Economics and Finance*, vol. 40, pp. 178–199, 2017.

[23] H. C. Curtius, S. L. Hille, C. Berger, U. J. Jonas Hahnel, and R. Wüstenhagen, "Shotgun or snowball approach? Accelerating the diffusion of rooftop solar photovoltaics through peer effects and social norms," *Energy Policy*, vol. 118, pp. 596–602, 2018.

[24] T. Foucault and L. Fresard, "Learning from peers' stock prices and corporate investment," *Journal of Financial Economics*, vol. 111, no. 3, pp. 554–577, 2014.

[25] R. Rahmani, M. Abbas, and G. Alahyarizadeh, "The effects of peer scaffolding in problem-based gaming on the frequency of double-loop learning and performance in integrated science process skills," *Procedia-Social and Behavioral Sciences*, vol. 93, no. 1, pp. 1994–1999, 2013.

[26] Y. Y. Yan, J. Y. Deng, and D. Y. Wen, "Peer effects on entrepreneurial activity:Model and empirical study," *Chinese Journal of Management Science,(Chin. Ed.)*, vol. 26, pp. 147–156, 2018.

[27] P. Margaretic, R. Cifuentes, and J. G. Carreño, "Banks' interconnections and peer effects: evidence from Chile," *Research in International Business and Finance*, vol. 58, no. 1, Article ID 101438, 2021.

[28] J. D. Angrist, "The perils of peer effects," *Labour Economics*, vol. 30, pp. 98–108, 2014.

[29] T. Kopp and J. Salecker, "How traders influence their neighbours: modelling social evolutionary processes and peer effects in agricultural trade networks," *Journal of Economic Dynamics and Control*, vol. 117, Article ID 103944, 2020.

[30] C. Frydman, "What drives peer effects in financial decision-making?" *Neural and Behavioral Evidence*, Social Science Electronic Publishing, Rochester, NY, USA, 2015.

[31] W. A. Brock and S. N. Durlauf, "Identification of binary choice models with social interactions," *Journal of Econometrics*, vol. 140, no. 1, pp. 52–75, 2007.
[32] K. Han and J. Tan, "How neighbours influence commercial health insurance purchase: evidence from 2451 rural households in west China," *Journal of Development Effectiveness*, vol. 13, no. 3, pp. 329–341, 2021.
[33] P. Ouimet and G. Tate, "Learning from coworkers:peer effects on individual investment decisions," *The Journal of Finance*, vol. 75, no. 1, pp. 133–172, 2020.

Hindawi
Complexity
Volume 2022, Article ID 8591511, 12 pages
https://doi.org/10.1155/2022/8591511

Research Article

Pricing and Production Decisions for New and Remanufactured Products

Feng Wei ,[1,2] Yan Zhu ,[1,2] Ting Ma,[3,4] Qiaoyan Huang,[5] Zengshan Zhen,[1,2] and Jinhui Chen[6,7]

[1]*School of Economics and Management, Tsinghua University, Beijing 100084, China*
[2]*Institute of Internet Industry, Tsinghua University, Beijing 100085, China*
[3]*Key Laboratory of Regional Sustainable Development Modeling, Chinese Academy of Sciences, Beijing 100101, China*
[4]*Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China*
[5]*Advanced Innovation Center for Future Education, Beijing Normal University, Beijing 102206, China*
[6]*High-Tech Research and Development Center, Ministry of Science and Technology, Beijing 100044, China*
[7]*School of Public Policy and Management, Tsinghua University, Beijing 100084, China*

Correspondence should be addressed to Yan Zhu; zhuyan@sem.tsinghua.edu.cn

Received 9 November 2021; Revised 31 March 2022; Accepted 20 April 2022; Published 31 May 2022

Academic Editor: Feng Li

Remanufacturing widely exists in production activities. Two different game models are involved while considering reverse channels: (1) In Model P, the manufacturer provides new and remanufactured products to two retailers. New products are sold through an online platform (retailer R1), while remanufactured products are sold in offline physical stores (retailer R2) in a decentralized scenario. (2) In Model C, the manufacturer provides new and remanufactured units to only one retailer (retailer R) that operates both online and offline channels in a centralized scenario. This research showed that a manufacturer's profitability and industry profits in Model P were higher than those in Model C from the perspective of economic performance; the sum of the profits of both retailers in Model P was worse than the profits of the retailer in Model C. Moreover, Model P was found to be greener than Model C from the perspective of environmental sustainability. From a social viewpoint, Model P had a higher consumer surplus than Model C; the higher the cost of distributing a remanufactured unit, the more disadvantageous the model to the consumers.

1. Introduction

Recently, the environmental burden has increased because of unsustainable spending and a high reliance on natural resources. Metal and glass packaging (totally, 12 million tons of glass bottles and glass jars were recovered) ranked the second and third with a recovery rate of 78.3% and 74.1%, respectively. If not handled properly, this causes environmental pollution, wastes a considerably high amount of materials, and is harmful to our physical health. The reuse of waste paper, waste iron, computers, mobile phones, and LED energy-saving lamps is a common economic phenomenon in production and operations, which is considered an environmentally friendly method of operation. The reduction of resource consumption accelerates the awareness of product reuse rather than a "one-way" economy.

To protect the environment and product recycling, governments encourage people to enjoy an ecological life. For example, Japan attaches considerable importance to environmental protection, resource conservation, and sustained economic development. In 1994, Japan issued the Programme of Action for the 21st Century, which is an effort to harmonize environmental protection with economic development to minimize the reliance on nonrenewable resources and develop renewable energy. Faced with ecological and environmental problems, the United States formulated a series of sustainable development strategies, such as improving environmental management systems, encouraging companies to recycle packaging materials, and labeling recycled materials. Similarly, China undertook a series of measures including vigorously promoting a cyclic economy and improving the cyclic use of major wastes. In particular, in September 2015, the United Nations Development Summit approved and adopted the "2030 Agenda for Sustainable Development," which proposed 17 sustainable development goals (SDGs) covering three major areas, namely, economy, society, and environment. Whether it is environmental protection or energy conservation, it is necessary to further explore.

Remanufacturing has been recognized as a profitable strategy for firms in a reverse supply chain as it can be beneficial for the recycling of waste products and consumes less energy than the original manufacturing process. Furthermore, it can considerably reduce environmental pollution as it generally requires only 15% of the energy used in production [1]. In October 2012, a report titled "Remanufactured Goods: An Overview of the US and Global Industries, Markets, and Trade" was released, which stated that the US remanufacturing industry is concentrated on aerospace, automotive parts, machinery, medical equipment, electrical equipment, and used tires. From 2009 to 2011, this industry created 180,000 full-time jobs, and the value created by remanufacturing in the US increased by 15%, reaching at least $43 billion. In the 1990s, Japan promoted the implementation of the 3R policy (i.e., reduce, reuse, and resource). By combining remanufacturing with the 3R policy, Japan has achieved the combined goal of reducing pollution emissions and promoting resource conservation.

Although the promotion of environmental sustainability has attracted more attention to "remanufacturing" in the past decades, in practice, remanufacturing involves retailers' marketing related to distribution channels as well as operations management. For example, an internationally well-known company, Hewlett-Packard Development Company (HP, https://hp.com), recycles used computers and printers and resells them after processing. HP sells both new and remanufactured products through independent distributors, similar to Lenovo, Apple, Canon, and Panasonic. Haier Group, a large household appliance brand in China, sells new products such as water heaters, refrigerators, air conditioners, and washing machines through its online website (https://haier.com), as well as second-hand appliances (i.e., remanufactured products) in brick-and-mortar stores. According to the 2017 US "Green Friday" campaign, Amazon's "frustration-free packaging" is expected to save 181,000 tons of packaging materials and 307 million transport containers in the next decade. "Worry-Free Packaging" reduces waste throughout the supply chain with a focus on reducing the environmental impact of packaging and driving the sustainability of packaging. Similarly, according to the emission reduction plan, Walmart is expected to combine energy conservation measures with the use of 50% renewable energy to achieve the target of reducing emissions by 18% by 2025. Walmart stores will be equipped with light-emitting diode (LED) energy-saving lamps, air conditioning energy-saving upgrades, and thermal energy recovery systems. By 2025, Walmart is expected to use 100% recyclable packaging for its own-brand goods. In such a circumstance, how a downstream competition affected all the players' strategic decisions was explored, which in turn affected the environment and society.

In response to these cases, the objective of this study was to compare the economic, environmental, and social effects of decentralized and centralized scenarios and elaborate on them from different perspectives.

In particular, the following fundamental research issues were addressed:

(1) From the viewpoint of economic performance, how does the competition of retailers affect economic performance? Thus, it is important to know how marketing strategies should be formulated in a market.

(2) From the viewpoint of environmental sustainability, how does the competition of retailers affect the environment concerning the marketing of remanufactured products?

(3) From the viewpoint of social performance, will retailers' competition have an impact on consumer surplus?

From an economic perspective, our analysis indicated that when compared with Model C, Model P improved the manufacturer's financial performance and industry profits, while cutting into the retailers' profits. From a social viewpoint, our results showed that in terms of consumer surplus, the decentralized scenario (Model P) was superior to the centralized scenario (Model C). The latter reduced consumer surplus and caused a social loss, which was detrimental to consumers. From the viewpoint of environmental sustainability, Model P was greener than Model C. From the manufacturers' perspective, if they are concerned about economic performance and social development, some measures had to be taken to limit the competition of retailers in marketing.

In response to these issues, a game model for the analysis in the decentralized and centralized contexts is constructed. This article makes three contributions to the existing knowledge of selling concepts in a sustainable marketing strategy. First, although previous scholars have explored the relationship of competition among retailers, there is limited existing literature on the competition of retailers in the remanufacturing industry. To address this gap, an alternative approach to discussing how the competition of retailers affected both parties' profitability and industrial outputs is considered. This is an important novelty for the article. Second, even though the question of retailer's competition has been studied in depth, limited information is known about the effect of the competition of retailers on the economy, environment, and society. We analyzed the effects of competition on these aspects in the remanufacturing industry. We further discussed the impact of manufacturer's operations on the environment and consumer surplus. This is the most important novelty of the article. Finally, the analytical modeling results provided several managerial insights, which would be beneficial for game players to develop reasonable channel strategies and provide a theoretical basis for the supply chain participants.

The rest of this article is organized as follows. Section 2 reviews the existing literature in this field, while Section 3 discusses the model assumptions. The model formulation and the analysis of the two models are presented in Sections 4 and 5, respectively. Sections 6 and 7 provide numerical examples and managerial implications. Section 8 concludes this paper and presents ideas for further research. All the relevant proofs have been moved to an online supplement.

2. Literature Review

The efficient operation of enterprises needs to consider the sustainable development of economy and environment. Sustainability is one of the important issues in operations management, which aims at building harmony and a win-win scenario for society, economy, and nature. Esenduran et al. [2] were concerned about the influence of environmental regulations on the manufacturer's profits and consumer surplus and showed that consumers benefit from buying remanufactured products because of the competition and low cost of remanufacturing in an environmentally regulated marketplace. Cao et al. [3] underscored that environmentally friendly regulation by the governments motivates firms to engage in remanufacturing activities and considered extended producer responsibility (EPR) principles to achieve a breakthrough in the remanufacturing industry. Bittar [4] discussed whether the purchase of remanufactured products by consumers is related to environmental consciousness by using empirical methods. Taleizadeh et al. [5] comprehensively considered carbon reduction and return policy in a supply chain to achieve sustainable development of enterprises. They found that a higher refund price has a positive impact on the supply chain profits and carbon emission reduction. Chetan et al. [6] adopted a two-stage mechanism to analyze the cost functions. The optimal scoring function is designed to maximize its utility. Pazoki et al. [7] addressed environmental performance issues and proposed a decision-making system to set up environmental regulations to recycle as many products as possible. Based on consumer utility, Zhang et al. [8] studied a centralized system for selling short-life-cycle products in which green remanufactured products are remarketed at an appropriate price in the second period. They found that improving the quality and informing consumers of the benefits of remanufactured units are considered environmental benefits. Huang et al. [9] analyzed the impact of government subsidy on the environment and game players' profits. They found that regardless of which subsidy pattern the government adopts, it benefits both manufacturer and collector. It is also beneficial to environmental sustainability. Yang et al. [10] considered the cap-and-trade regulation to construct game models with and without remanufacturing. The research results showed that remanufacturing can increase profit and improve the level of carbon emission reduction. When a manufacturer is subject to strict emission controls, the total amount of carbon emissions under third-party collection mode is always the lowest. Li et al. [11] discussed the impact of low-carbon manufacturing in a closed-loop supply chain (CLSC) and constructed a CLSC model. They further analyzed the impact of emission reduction effort on supply chain performance and stated that recycling benefits the utility of low-carbon consumers but damages the profit of CLSC due to the high investment cost of recycling. Zheng et al. [12] explored the influence of the design for the environment (DfE) on firms' remanufacturing strategies. A theoretical model is constructed to demonstrate the impact on game decisions. Although a DfE can help reduce the environmental impact, a high level of DfE may harm the environment by substantially increasing total sales. Yang et al. [13] studied the environmental impacts of a flexible versus simple trade-in strategy considering carbon tax policies and established a hoteling model. An appropriate carbon tax on green products could lead businesses to adopt a more environmentally friendly trade-in strategy. Giri et al. [14] discussed product quality and return policy under two strategies, which play an increasing role in environmental protection, and assumed that the manufacturer is the Stackelberg leader. The retailer sets the retail price and provides a return policy in the first period only. Mondal et al. [15] investigated greening strategies and pricing for the green supply chain. As the greening cost increases, the greening level and wholesale price decrease. The green level of the products can be improved by a cost-sharing mechanism. Cao et al. [16] analyzed pricing decisions under carbon tax policy (CTP) and remanufacturing subsidy policy (RSP). The social welfare under RSP is greater than that under CTP when the environmental cost coefficient is low. Deng [17] looked at environmental performance and remanufacturing model selection under the influence of consumers' risk aversion. For environmental performance, the supplier tends to choose either the NR (no remanufacturing) model or the SR (remanufacturing is carried out by the supplier) model. Both the environment and supplier will benefit from the SR model. Chung et al. [18] presented a sustainable remanufacturing model in a dynamic supply chain. The disposal of environmental pollutants caused by defective items is considered. Niu et al. [19] discussed the influence of retail link on coordination of social welfare and profit. The social welfare levels are compared in the models of bundled outsourcing and individual outsourcing. A similar approach can be found in Wei et al. [20]. Most of the above studies consider environmental issues related to operation management, but few consider the issue of consumer surplus. This paper makes up for this gap and compares the channel selection strategies of game players.

The second set of literature is related to the remanufacturing industry and focuses on saving resources, reducing energy consumption, reducing product costs, and improving the competitiveness of enterprises. Yan et al. [21] found optimal pricing policies for new and remanufactured units while considering the inventory level, which is inversely proportional to the price of the remanufactured products, but directly proportional to the price difference. Shu et al. [22] investigated game models with three different structures in remanufacturing supply chains. The results show that the dominant player gets higher profits and the players will take the initiative to strive for market leadership. However, there is a cost to pay for market leadership. Wu et al. [23] studied pricing strategy and competitive remanufacturing problems under different scenarios. They analyzed how the contrast effect and assimilation effect influence the pricing strategy of original equipment manufacturers (OEMs). If OEMs sell remanufactured products, this will weaken consumers' perceived value of new products. Zhao et al. [24] developed a decision model considering technology authorizations in a closed-loop supply chain. The results show that fixed technology authorization fees in remanufacturing mode can enable a retailer to improve service levels and improve the recovery rate of the third party. Shi et al. [25] studied divisional conflict and channel choice in remanufacturing the supply chain. The firm sells new and remanufactured products through a direct or indirect channel (through an independent retailer). The study showed that when compared with a direct channel, a decentralized firm achieved higher profit and more consumer demand from an indirect channel. Jin et al. [26] analyzed optimal warranty policy using a game-theoretic model. They argued that the optimal warranty policy depends on the cost structure. Interestingly, higher warranty fees may induce manufacturers to adopt a warranty policy. In some cases, mandatory warranty provisions can be harmful. Xiang et al. [27] discussed the influence of technological innovation and Big Data marketing on the decision making of game players in remanufacturing the supply chain. It is shown that an Internet recycling platform is conducive to manufacturers but hurts suppliers' profits. On the contrary, efficient technological innovation and Big Data marketing weaken the initiative of the manufacturer. Wang et al. [28] analyzed the influence of consumer behavior and the trade-in remanufacturing policy on remanufacturer decisions by using a consumer utility model. The trade-in policy raised brand prices and increased corporate profits and consumer surplus. Jia et al. [29] considered a closed-loop supply chain, including e-retailers platform service and self-operated store, which provide upstream manufacturers with options for selling new and remanufactured products. From an environmental point of view, selling new products through an e-retailer and selling a remanufactured product online are optimal choices. Han et al. [30] discussed the manufacturer's optimal recovery strategy for handling used products. Government subsidies would reduce environmental impact and increase consumer surplus. Interestingly, a manufacturer with higher product quality tends to choose remanufacturing products instead of recycling materials. Xu et al. [31] explored the coordination mechanism of collection rate and pricing in the remanufacturing industry. It has been demonstrated that lower competitive intensity and saving production costs encourage the manufacturer to remanufacture products. A two-part tariff contract can achieve Pareto improvement. Huang et al. [32] studied pricing decisions considering technology licensing and strategic consumers. With an increase in strategic consumers, the demand for remanufactured products increases, and the demand for new products decreases. When a manufacturer authorizes a third party to remanufacture products, it suffers a profit loss. Li et al. [33] explored the issues of remanufacturing construction and demolition waste. They showed that retailer fairness concerns cut into the manufacturers' profits and led to a lower wholesale price of building materials. Rahmani et al. [34] focused on horizontal and vertical cooperation in two reverse supply chains and analyzed quality improvement competition. It has been demonstrated that in decentralized decision making, the remanufacturers who cooperate horizontally will cut down the collector's profits. The multiple-link two-part tariff is applied to coordinate each player. Wang et al. [35] considered a manufacturer who acquired core materials through either outsourcing or self-remanufacturing under yield uncertainty. The distribution bounds determine the manufacturer's strategic choice under a random recovery rate. Kleber et al. [36] considered two-sided competition in both acquisition and sales in remanufacturing. It has been demonstrated that the market advantage is much stronger than the acquisition advantage. Zhang et al. [37] explored a competitive closed-loop supply chain, and authorization mode and outsourcing mode were considered. When per-unit new product production cost is low, the duopoly third-party remanufacturers (TPRs) will select an outsourcing strategy. Raz et al. [38] explored codevelopment at the product level and the influence of outsource manufacturing/process innovation. Outsourcing must include options for codevelopment on specific activities and product innovations. Different from the above literature, this paper makes a comparative analysis of the sales quantity, profit, environmental impact, and consumer surplus, which enriches the theoretical research on channel competition.

Thus far, few researchers have studied the integration of sustainability and remanufacturing. To bridge this gap in this article, instead of outsourcing remanufactured products to third-party manufacturers, it is assumed that all of the used products are recycled by the manufacturer [39]. Moreover, none of the existing studies focused on the effects of retailers' competition on the environment and consumer surplus. On the contrary, the effects of competition on the economy, environment, and society concerning the sale of remanufactured products are demonstrated. The results of this study will be beneficial to game players in developing operations management and marketing strategies, which will enhance economic, environmental, and social sustainability.

3. Model Assumptions and Notations

In this study, how different distribution channels of retailers affect sustainable development in the remanufacturing industry is explored. Consequently, based on business practices, two models have been developed for investigating the effects of a competition involving two retailers: (1) The manufacturer provides new and remanufactured products to two retailers (retailer 1, retailer 2) respectively, in a decentralized scenario (i.e., Model P). (2) In a centralized scenario, the manufacturer provides new and remanufactured units to only one retailer (retailer R) (i.e., Model C). See Table 1 for more detailed notations and explanations

Table 1: Detailed notations and explanations.

Notation	Explanation
w_n^j/w_r^j	The wholesale price of new/remanufactured product in Model j, $j \in (P, C)$
c_n/c_r	Unit cost for distributing new/remanufactured product
p_n^j/p_r^j	New/remanufactured product price in Model j, $j \in (P, C)$
q_n^j/q_r^j	Quantity of new/remanufactured product in Model j, $j \in (P, C)$
c	Unit cost for making a new product
δ	Consumer value discount for remanufactured products
e_n/e_u	Per-unit environmental impact of a new/remanufactured product
π_k^j	Profits of player k in Model j, $j \in (P, C)$, $k \in (R1, R2, R, M, T)$

[40]. Considering the framework involved, these assumptions about the manufacturer, retailer, product, and consumers can be described as follows.

Assumption 1. Game decision making is taken into account in a steady-state period with the following consequences: First, the manufacturer claims the wholesale price for both products (w_n, w_r). Second, the retailer maximizes its profits by responding with the optimal quantity of products (q_n, q_r).

This is a common practice in the existing literature, where new and remanufactured products are repeatedly sold, and each product is sold only once in the business market, which is referred to as the steady-state period. According to the principal-subordinate game theory, the manufacturer is the Stackelberg leader and the retailer is the follower. The equilibrium solution is obtained by backward induction method.

Assumption 2. Compared to the new product, the primary consumers' willingness to pay for the remanufactured product is a ratio of the value discount $\delta \in (0, 1)$.

Here, consumers believe that the value of remanufactured products is lower; that is, $\delta \in (0, 1)$. Similar to existing pieces of literature, Assumption 2 implies a vertical differentiation model with an agreed order for the consumers' valuation; that is, consumers are more inclined to buy new products rather than remanufactured ones. Note that a consumer has a valuation of v for the new product and δv for the remanufactured unit.

Assumption 3. For a new product, the consumers' valuation (v) is heterogeneous and the number of consumers is considered to be constant, which follows a uniform distribution in the market that is normalized to 1; that is, $v \sim U[0, 1]$.

This conforms to the relevant literature, and the market size is normalized to 1. Most notably, the cannibalization problem of both new and remanufactured units should be considered because of the consumer value discount (δ). Note that if $\delta = 0$, consumers will not buy the remanufactured product, which is regarded as a low-quality product, but if $\delta = 1$, consumers believe that the remanufactured product can completely replace the new product and thus pay the same amount for either product. Based on Assumptions 2 and 3, the linear inverse demand functions [41] are obtained as follows:

$$\begin{aligned} p_n &= 1 - q_n - \delta q_r, \\ p_r &= \delta (1 - q_n - q_r). \end{aligned} \tag{1}$$

Assumption 4. The unit cost for distributing a new product is c_n, and the unit cost for distributing a remanufactured product is c_r.

In both models, it is assumed that the distribution costs of new products and remanufactured products are different. According to the relevant pieces of literature (e.g., Taleizadeh et al. [5], Wu et al. [23]), the remanufacturing cost is divided into two parts, namely, the cost of producing and marketing the remanufactured product.

Assumption 5. The unit cost for remanufacturing a used product (c_m) is less than that of manufacturing a new product, c_p (i.e., $c_p = c > c_m = 0$).

To develop a circular economy, used products are recycled to save resources and protect the environment. In order to confirm the rationality of employing a remanufacturing scenario, it is assumed that remanufacturing is lower in cost than production, which is the consensus in the existing literature (e.g., Wu et al. [23], Jia et al. [29]). Without any loss of generality, c_m is normalized to zero $(c_m = 0)$, and it is assumed that $c_p = c > c_m$. In particular, the focus is on the cost of marketing by controlling the cost of producing and remanufacturing, and thus the focus is mainly on the marketing issues related to competition among retailers.

4. Model Formulation and Solution

From the perspective of circular economy, the strategy of online and offline sales channel selection of new products and remanufactured products is discussed, and the impact of unit new products and remanufactured products on the environment is quantitatively described. New products and remanufactured products are sold online and offline by retailer R. It makes up for the lack of combining online and offline channel marketing of new products with remanufactured products in the existing literature. In Model P and Model C, the game order of the events is as follows: The manufacturer first announces the wholesale price for both products (w_n, w_r). Then, to make more profits, the retailer sets the optimal quantity of the two products (q_n, q_r). Note that π_k^j represents the profits of the player k in Model j, where $k \in (R1, R2, R, M, T)$ denote the retailers, the manufacturer, and the total supply chain, respectively, and $j \in (P, C)$ represents Model P and Model C, respectively.

4.1. Decentralized Model (Model P). In this scenario, both products are sold to different retailers. The manufacturer's optimization problem can be expressed as follows:

$$\max_{w_n, w_r} \pi_M^P = (w_n - c)q_n + w_r q_r. \tag{2}$$

From the viewpoint of the wholesale price (w_n^{P*}), retailer 1's optimization problem can be expressed as follows:

$$\max_{q_n} \pi_{R1}^P = (p_n - w_n - c_n)q_n. \tag{3}$$

From the viewpoint of the wholesale price (w_r^{P*}), retailer 2's optimization problem can be expressed as follows:

$$\max_{q_r} \pi_{R2}^P = (p_r - w_r - c_r)q_r. \tag{4}$$

The total profits of the supply chain can be calculated as follows:

$$\pi_T^{P*} = \pi_M^{P*} + \pi_{R1}^{P*} + \pi_{R2}^{P*}. \tag{5}$$

Using the backward induction method, the equilibrium decisions are determined and the important outcomes are summarized as follows.

Lemma 1. *Considering Model P, the quantities, wholesale price, and profits can be calculated as follows:*

$$\begin{aligned}
w_n^{P*} &= \frac{1}{2}(c - c_n + 1),\\
w_r^{P*} &= \frac{1}{2}(\delta - c_r),\\
q_n^{P*} &= \frac{2 - 2c - 2c_n + c_r - \delta}{8 - 2\delta},\\
q_r^{P*} &= \frac{2c_r - \delta(1 + c + c_n)}{2\delta(\delta - 4)},\\
\pi_M^{P*} &= \frac{\delta(c + c_n - 1)^2 + (c_r - \delta)[c_r - \delta(c + c_n)]}{2\delta(4 - \delta)},\\
\pi_{R1}^{P*} &= \frac{(-2 + 2c + 2c_n - c_r + \delta)^2}{4(\delta - 4)^2},\\
\pi_{R2}^{P*} &= \frac{(-2c_r + \delta + c\delta + \delta c_n)^2}{4\delta(\delta - 4)^2},\\
\pi_T^{P*} &= \frac{1}{4\delta(-4 + \delta)^2}\left[\delta(-2 + 2c + 2c_n - c_r + \delta)^2\right.\\
&\quad \left. + (-2c_r + \delta + c\delta + \delta c_n)^2 - 2(\delta - 4)\left((c + c_n - 1)^2\delta + (c_r - \delta)(c_r - c\delta + \delta c_n)\right)\right].
\end{aligned} \tag{6}$$

4.2. Centralized Model (Model C). In this scenario, both products are sold to a retailer R, and the manufacturer's optimization problem can be expressed as follows:

$$\max_{w_n, w_r} \pi_M^M = (w_n - c)q_n + w_r q_r. \tag{7}$$

Given the wholesale price (w_n^{M*} and w_r^{M*}), the downstream end-product market became a monopoly market. In other words, retailers R1 and R2 merged into a more powerful retailer R, who was one of the two retailers when the competition took place. The retailer R's problem could be optimized as follows:

$$\max_{q_n, q_r} \pi_R^M = (p_n - w_n - c_n)q_n + (p_r - w_r - c_r)q_r. \tag{8}$$

The total profits of the supply chain can be calculated as follows:

$$\pi_T^{M*} = \pi_M^{M*} + \pi_R^{M*}. \tag{9}$$

Using the backward induction method, the important outcomes are summarized as follows.

Lemma 2. *Considering Model C, the quantities, wholesale price, and profits can be calculated as follows:*

$$\begin{aligned}
w_n^{M*} &= \frac{1}{2}(1 + c - c_n),\\
w_r^{M*} &= \frac{1}{2}(\delta - c_r),\\
q_n^{M*} &= \frac{-1 + c + c_n - c_r + \delta}{4(\delta - 1)},\\
q_r^{M*} &= \frac{c_r - \delta(c + c_n)}{4\delta(\delta - 1)},\\
\pi_M^{M*} &= \frac{(-1 + c + c_n)^2\delta + (c_r - \delta)(c_r + \delta - 2c\delta - 2\delta c_n)}{8\delta(1 - \delta)},\\
\pi_M^{M*} &= \frac{(-1 + c + c_n)^2\delta + (c_r - \delta)(c_r + \delta - 2c\delta - 2\delta c_n)}{8\delta(1 - \delta)},\\
\pi_T^{M*} &= \frac{3\left[(c + c_n - 1)^2\delta + (c_r - \delta)(c_r + \delta - 2c\delta - 2\delta c_n)\right]}{16\delta(1 - \delta)}.
\end{aligned} \tag{10}$$

To ensure that the players distribute a certain quantity of the two products in the market, the following condition had to be imposed: $q_n > q_r > 0$.

Lemma 3. *Considering both scenarios, the unit cost of distributing a remanufactured product needs to satisfy the following condition:*

$$\frac{-\delta + 3c\delta + 3c_n\delta + \delta^2}{2 + \delta} < c_r < c\delta + c_n\delta. \tag{11}$$

This lemma shows that the manufacturer is engaged in remanufacturing; that is, $q_r > 0$. Therefore, the marketing cost would not be very high: $c_r < c\delta + c_n\delta$. In contrast, if the retailer distributes an adequate quantity of the new product, which is the source of the remanufacturing cores (i.e., $q_n > q_r$), the following is required: $-\delta + 3c\delta + 3c_n\delta + \delta^2/2 + \delta < c_r$.

5. Model Analysis

In this part, the competition between economic performance and green sustainability in the remanufacturing industry is discussed, and some interesting insights are derived. The differences between the two scenarios are discussed, and subsequently the sustainability of the economy, environment, and society based on Lemmas 1 and 2 is discussed to make the following observations.

5.1. Comparison of Economic Sustainability. According to Lemmas 1 and 2, some insightful results were obtained in different scenarios. In particular, we first consider the question posed at the beginning of the article: From the viewpoint of economic performance, how does the competition of retailers affect economic performance? The differences between the decentralized and centralized cases are highlighted to clarify the managerial implications.

Proposition 1. *The manufacturer benefits more in Model P than in Model C; that is,* $\pi_M^{P*} > \pi_M^{C*}$.

Note that a competition of downstream retailers can lead to economic losses for manufacturers. The focus is on the impact when competition occurs, rather than optimal decisions of the supply chain partners in the remanufacturing industry. Further explaining the managerial insight, in Model P, competition among downstream retailers reduces the sales price but increases the quantity, which allows the manufacturer to wholesale more products to obtain more revenue. Therefore, such competition is beneficial to a manufacturer in a decentralized scenario; that is, $\pi_M^{P*} > \pi_M^{C*}$. Furthermore, as it controls the reverse channel, when compared with that in Model C, the number of remanufactured products in Model P is determined by the manufacturer $(q_r^{P*} > q_r^{C*})$. As a result, the quantity of the remanufactured product decreases and the profits for the upstream manufacturer decrease when such competition occurs. In other words, the proposition shows that yield from the remanufactured units in Model P is sufficient to compensate for the loss of new product sales.

Proposition 2. *The sum of the profits of both retailers in Model P is always worse than that in Model C; that is,* $\pi_{R1}^{P*} + \pi_{R2}^{P*} < \pi_R^{C*}$.

Note that the merging retailer earns higher profits because of a higher retail price, which reduces the competitive intensity between the retailers and hurts the upstream manufacturer by lowering its profits (see Proposition 1). The power of the players can be interpreted as follows: according to the economic theory of competition, a downstream competition can enhance the power of downstream retailers by reducing the number of companies in the market, thus hurting the profits of upstream enterprises. Similarly, such competition can reduce the supply chain profits in the remanufacturing market. The competition between downstream firms is beneficial to the merging firm, which is one of them when the competition takes place.

The industry performance of a sustainable supply chain is the driving force for maintaining the well-being of the economy (see Tajbakhsh et al. [42] for more details). Most notably, the focus was on economic significance in market competition according to Lemmas 1 and 2, and the following observation was made.

Proposition 3. *The competition among retailers is always detrimental to the industry; that is,* $\pi_T^{P*} > \pi_T^{C*}$.

Most notably, the equilibrium profits of the industry in the decentralized scenario (Model P) are higher than those in the centralized scenario (Model C), and the merging firm benefits because it alleviates the double marginalization problem (described in Proposition 2). As described in Proposition 1, due to the fierce competition led by mergers and acquisitions in downstream retailers, the manufacturer benefits from the competition, and thus its sales strategies are affected.

5.2. Comparison of Environmental Sustainability. In this section, the focus is on environmental implications in the remanufacturing industry. We answer the second question posed in the introduction of this article: From the viewpoint of environmental sustainability, how does the competition of retailers affect the environment concerning the marketing of remanufactured products?

Environmental sustainability of the decentralized and centralized cases is highlighted to provide managerial inspiration. In these models, it is assumed that the per-unit disposal impact of a new/remanufactured product was e_n/e_u, respectively. Remanufacturing consumes less energy and materials than producing new units in a traditional industry; therefore, the following assumption is necessary.

Assumption 6. The environmental impact per unit of a new product is larger than that of a remanufactured unit with essentially $e_n > e_u$.

Based on this assumption and Lemmas 1 and 2, E^P/E^C indicates the environmental impact for Model P/C, respectively. The difference in environmental sustainability can be summarized as follows.

Proposition 4. *In terms of the environmental impact, Model P is always greener than Model C; that is, $E^P < E^C$.*

Note that the total disposal impact of a new and remanufactured product is $E_n = e_n(q_n - q_r)$ and $E_u = e_u q_r$, respectively. Not only new products but also remanufactured products can have an impact on the environment. As discussed in the optimal quantity comparative analysis, retailer 1 sells fewer units of new products in the decentralized scenario (i.e., $q_n^{P*} < q_n^{C*}$), which indicates the lower environmental impact of new products. In addition, although there are more remanufactured products in Model P (i.e., $q_r^{P*} > q_r^{C*}$), their environmental impact is insufficient to compensate for the environmental damage caused by the new products in Model C. In other words, $q_n^{C*} > q_n^{P*} > q_r^{P*} > q_r^{C*}$. Therefore, Model P is beneficial in terms of environmental sustainability, as stated in Proposition 4.

From a broader viewpoint, if the retailers focus on economic performance, the competitive strategy is conducive to the firms' development; conversely, if they care more about environmental sustainability, some measures should be implemented to prevent competition among retailers in marketing.

5.3. Comparison of Social Sustainability. The focus is now on the consumer surplus of remanufacturing in a market. We can answer the final question posed in the introduction of this article: From the viewpoint of social performance, will retailers' competition have an impact on consumer surplus?

To evaluate the social performance of the retailers' competition strategy in a reverse channel, the following formula is used for calculating the consumer surplus, which included the consumers' willingness to pay for both products:

$$\mathrm{CS} = \int_{1-q_n-q_r}^{1-q_n} (\delta u - p_r)\mathrm{d}u + \int_{1-q_n}^{1} (u - p_n)\mathrm{d}u. \qquad (12)$$

Let CS^P/CS^C indicate the consumer surplus of Model P/C, respectively. The difference in social sustainability can be summarized as follows.

Proposition 5. *In terms of the consumer surplus, Model P always has a higher surplus than Model C with essentially $CS^P > CS^C$.*

Consumer surplus measures the extra benefits that buyers feel they are getting in a particular market. The above proposition implies that a decentralized scenario (i.e., Model P) is more attractive to consumers and brings more utility to them in marketing. In Model C, there is a lower consumer surplus than in Model P because of the higher prices after the competition, which is a common phenomenon in a business market. In contrast, note that retailers are very concerned about the competitive strategy because it leads to better financial performance in Model C (Proposition 2). In general, a downstream competition benefits a merging firm at the expense of the consumers and partner firms. From a social perspective, increasing the consumer surplus and meeting the demand of the consumers to enhance their economic welfare are a central part of ensuring long-term economic growth.

6. Numerical Example

To better present how parameter changes affect sustainable performance, a numerical simulation analysis of the equilibrium decisions and the environmental, economic, and social outputs was conducted.

To illustrate the influence of remarketing costs (c_r) on the supply chain members, the distribution cost of the new product is set as $c_n = 0.4$, and it is reasonable for the marketing cost to be between 20% and 60%. Previous studies have pointed out that the per-unit cost of manufacturing of the manufacturers could not be ignored; therefore, a scenario where the manufacturing cost per unit was $c = 0.1$ was considered. Note that the consumer value discount of the remanufactured product ranged between 45% and 90% (e.g., Esenduran et al. [43]). Thus, $\delta = 0.8$ was set for the numerical analysis. From production to remanufacturing, products are accompanied by energy consumption. Based on Esenduran et al. [43], the environmental impact of disposal of a new/remanufactured product per unit is set to 240 MJ (i.e., $e_n = 240$) and 138 MJ (i.e., $e_u = 138$), respectively. Based on Lemma 3, by using the constraint condition $q_n^{j*} > q_r^{j*} > 0$, the range of the unit cost for distributing a remanufactured product is obtained as $-\delta + 3c\delta + 3c_n\delta + \delta^2/2 + \delta < c_r < c\delta + c_n\delta$, and after substituting the numerical values, it is calculated as $0.371 < c_r < 0.4$. All these figures reflected the extent to which the change c_r affected both models.

First, Figure 1 reports the results of the economic sustainability before and after the competition in a market. From Figures 1(a), 1(b), and 1(c), it can be concluded that the profits of the manufacturer, retailers, and entire system decreased with an increase in the value of c_r. Furthermore, Figures 1(a) and 1(c) imply that the profitability of the manufacturer and industry in the decentralized scenario was higher than that in the centralized scenario. However, from

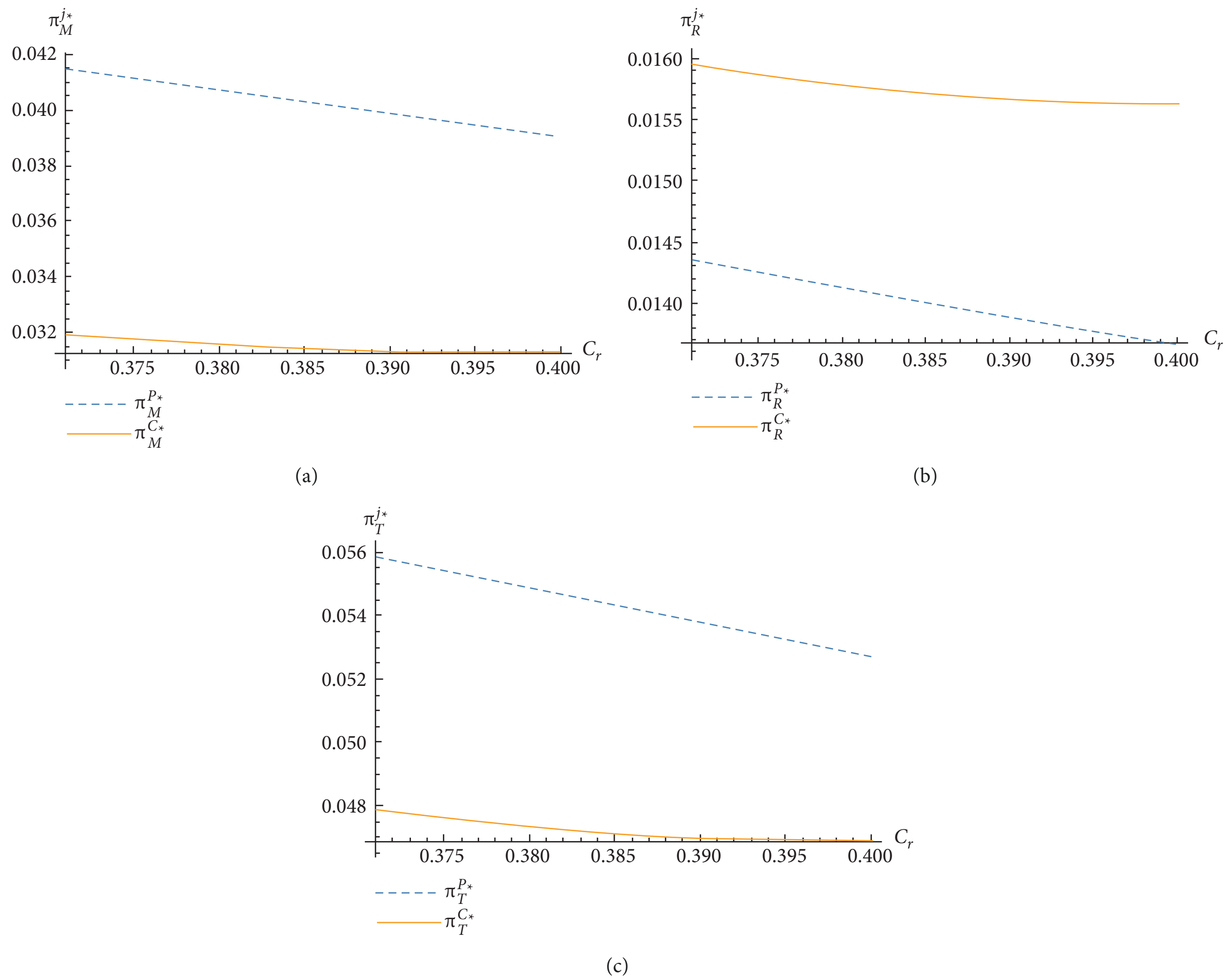


Figure 1: Economic outcomes: (a) effects of c_r on π_M^{j*}; (b) effects of c_r on π_R^{j*}; (c) effects of c_r on π_T^{j*}.

Figure 1(b), it is inferred that the sum of the profits of retailers 1 and 2 in Model P (π_R^{P*}) was always less than the profits of retailer R in Model C (π_R^{C*}). As a result, competition in the retail industry is widespread as they result in a better financial performance of the merging retailer. Figures 1(a), 1(b), and 1(c) show that economic sustainability was in line with the theoretical prediction discussed in Propositions 1–3.

Second, Figure 2 illustrates the impact of both models on environmental sustainability. Model P had a lower destructive impact on the environment than Model C. That is, Model P was greener than Model C (see Proposition 4). In addition, the environmental impact increased with an increase in c_r in both the decentralized and centralized scenarios.

Finally, the focus is now on social sustainability in both models. As shown in Figure 3, consumer surplus in Model P was larger than that in Model C ($CS^P > CS^C$), which was consistent with Proposition 5. Furthermore, consumer surplus in both models decreased as the value c_r increased; in other words, when the cost of remarketing was higher, the corresponding product was more disadvantageous to consumers.

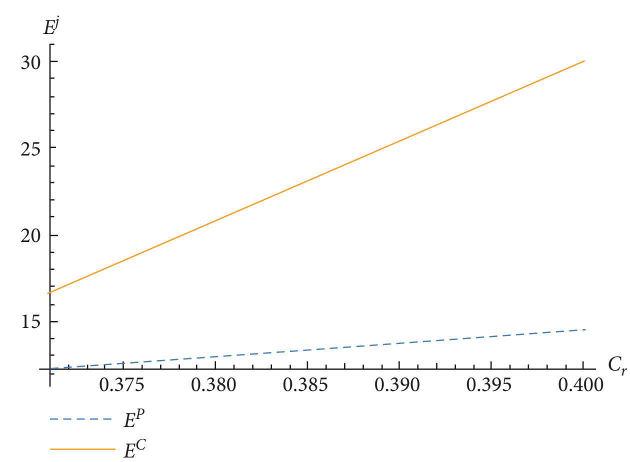


Figure 2: Environmental outcomes: effects of c_r on E^j.

7. Managerial Implications

Based on the above analysis, the managerial insights for marketing remanufacturing products involving a reverse channel were summarized, and some interesting

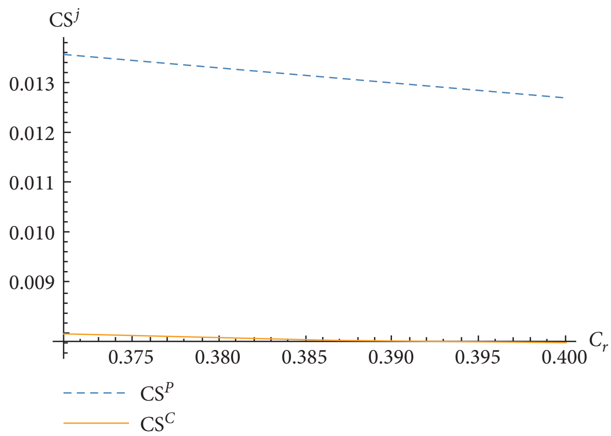


Figure 3: Social outcomes: effects of c_r on CS^j.

observations were derived. Specifically, this summary was divided into the following three aspects:

First, from the viewpoint of economic performance, the manufacturer's profitability and industry profits in the decentralized scenario were always higher than in the centralized scenario. However, the merging retailer earned higher profits than the retailers in Model P. Thus, if manufacturer wants to achieve more performance, it should cooperate with multiple retailers. For retailer, choosing a dual-channel model will bring more performance.

Second, from the viewpoint of environmental sustainability, the decentralized scenario was always greener than the centralized scenario; that is, Model P was beneficial for the environment when the competition took place. For manufacturers, the production process needs to consider environmental pollution and formulate action plans to reduce pollution, for example, adopting technologies to reduce carbon emissions and optimizing production processes.

Last, from the viewpoint of social performance, Model P was always better than Model C in terms of the consumer surplus. Consumer surplus, which measures the extra benefits that buyers feel they are getting, is an important factor influencing which business model to adopt. Sustainable operations management needs to maintain a balance of economy, society, and environment.

In brief, our findings provide some practical implications for business managers. The most significant contribution is that our research helps to develop channel selection strategies and environmental sustainability in the remanufacturing industry.

8. Conclusion

With the maturity of remanufacturing technology, an increasing number of manufacturers are motivated to produce new and remanufactured products simultaneously. The end-of-life products from consumers are collected by manufacturers for remarketing, which forms a reverse supply chain that reduces environmental pollution. Developing a circular economy in the future has become one of the objectives of modern enterprise operations and management. Although a considerable amount of literature has investigated competitive strategy and channel selection in the remanufacturing industry, to the best of our knowledge, closed-loop supply chains in the decentralized and centralized contexts have been rarely studied, which represents a gap in the theory. However, in practice, the importance of both marketing strategies and environmental sustainability of the supply chain has been recognized in recent years, and the selection of the appropriate environment for an enterprise ecosystem is more complicated from the viewpoint of operations management. Moreover, none of these studies focused on how retailers' competition affects the environment and consumer surplus. To address this gap, two theoretical models were constructed where a manufacturer collected the used products to produce the remanufactured products, which can be viewed as green products: (1) Model P, which is in a decentralized scenario; (2) Model C, which is in a centralized scenario.

The following three results are beneficial for managers in developing pricing and channel selection strategies. First, the economic benefit of the manufacturer in Model P was higher than that in Model C. The sum of profits of retailer R1 and retailer R2 in Model P was lower than that of retailer R in Model C. Second, for environmental protection, Model P was found to be greener than Model C. Last, from a social viewpoint, Model P had a higher consumer surplus than Model C, because of the higher prices when such a competition occurs.

In this paper, some valuable managerial insights were presented, but it is also acknowledged that some limitations deserve further study. First, the competition of downstream retailers was analyzed without considering the vertical competition between the upstream and downstream retailers. A vertical competition will have different effects on the firm's business strategy, which is one of the directions that can be investigated in the future. Second, it is assumed that there was a monopolistic manufacturer in the model. In reality, remanufacturing can be outsourced to a third-party remanufacturer or other agents such as retailers. Third, to pay close attention to sustainability, other factors, such as information asymmetry, network externality, and used product quality, were abstracted, which can potentially impact sustainable operations in a supply chain where such competition occurs.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 92046004). The authors are grateful to the participants for their helpful suggestions. Special thanks are due to the Institute of Internet Industry, Tsinghua University, for providing the article's framework.

References

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

[2] G. Esenduran, E. Kemahlıoğlu-Ziya, and J. M. Swaminathan, "Impact of take-back regulation on the remanufacturing industr," *Production and Operations Management*, vol. 26, no. 5, pp. 924–944, 2017.

[3] J. Cao, X. Zhang, L. Hu et al., "EPR regulation and reverse supply chain strategy on remanufacturing," *Computers & Industrial Engineering*, vol. 125, pp. 279–297, 2018.

[4] A. d. V. Bittar, "Selling remanufactured products: does consumer environmental consciousness matter," *Journal of Cleaner Production*, vol. 181, pp. 527–536, 2018.

[5] A. A. Taleizadeh, N. Alizadeh-Basban, and S. T. A. Niaki, "A closed-loop supply chain considering carbon reduction, quality improvement effort, and return policy under two remanufacturing scenarios," *Journal of Cleaner Production*, vol. 232, pp. 1230–1250, 2019.

[6] T. G. Chetan, M. Jenamani, and S. P. Sarmah, "Two-stage multi-attribute auction mechanism for price discovery and winner determination," *IEEE Transactions on Engineering Management*, vol. 66, no. 1, pp. 112–126, 2019.

[7] M. Pazoki and G. Zaccour, "A mechanism to promote product recovery and environmental performance," *European Journal of Operational Research*, vol. 274, no. 2, pp. 601–614, 2019.

[8] W. Zhang and Y. He, "Optimal policies for new and green remanufactured short-life-cycle products considering consumer behavior," *Journal of Cleaner Production*, vol. 214, pp. 483–505, 2019.

[9] Y. Huang, B. Zheng, and Z. Wang, "Advertisement vs. Monetary subsidy: which is better for remanufacturing?" *Journal of Systems Science and Systems Engineering*, vol. 29, no. 3, pp. 344–359, 2020.

[10] L. Yang, Y. Hu, and L. Huang, "Collecting mode selection in a remanufacturing supply chain under cap-and-trade regulation," *European Journal of Operational Research*, vol. 287, no. 2, pp. 480–496, 2020.

[11] J. Li and S. Gong, "Coordination of closed loop supply chain with dual-source supply and low-carbon concern," *Complexity*, vol. 2020, Article ID 7506791, 14 pages, 2020.

[12] X. Zheng, K. Govindan, Q. Deng, and L. Feng, "Effects of design for the environment on firms' production and remanufacturing strategies," *International Journal of Production Economics*, vol. 213, pp. 217–228, 2019.

[13] Z. Yang, X. Hu, J. Sun, and Y. Zhang, "Flexible versus simple trade-in strategy for remanufacturing," *Journal of the Operational Research Society*, vol. 72, no. 11, pp. 2472–2489, 2020.

[14] B. C. Giri, C. Mondal, and T. Maiti, "Optimal product quality and pricing strategy for a two-period closed-loop supply chain with retailer variable markup," *RAIRO - Operations Research*, vol. 53, no. 2, pp. 609–626, 2019.

[15] C. Mondal, B. C. Giri, and T. Maiti, "Pricing and greening strategies for a dual-channel closed-loop green supply chain," *Flexible Services and Manufacturing Journal*, vol. 32, no. 3, pp. 724–761, 2019.

[16] K. Cao, P. He, and Z. Liu, "Production and pricing decisions in a dual-channel supply chain under remanufacturing subsidy policy and carbon tax policy," *Journal of the Operational Research Society*, vol. 71, no. 8, pp. 1199–1215, 2019.

[17] W. Deng, "Sustainable development: impacts of consumers' risk aversion on remanufacturing model selection and environmental performance," *Sustainable Development*, vol. 28, no. 6, pp. 1564–1574, 2020.

[18] S. H. Chung, R. D. Weaver, and H. W. Jeon, "Sustainable management of remanufacturing in dynamic supply chains," *Networks and Spatial Economics*, vol. 20, no. 3, pp. 703–731, 2020.

[19] B. Niu, Z. Mu, and Z. Shen, "The role of retail information link in MNF's coordination of profit and social welfare via procurement outsourcing," *IEEE Transactions on Engineering Management*, pp. 1–14, 2021.

[20] F. Wei and H. Chen, "Independent sales or bundling? Decisions under different market-dominant powers," *Journal of Industrial and Management Optimization*, vol. 17, no. 4, pp. 1593–1612, 2021.

[21] X. Yan, X. Chao, Y. Lu, and S. X. Zhou, "Optimal policies for selling new and remanufactured products," *Production and Operations Management*, vol. 26, no. 9, pp. 1746–1759, 2017.

[22] T. Shu, Y. Wang, S. Chen, S. Wang, K. K. Lai, and Y. Yang, "Analysis of evolutionary game in structural formation of market power in remanufacturing supply chains," *Applied Economics*, vol. 51, no. 20, pp. 2195–2220, 2018.

[23] L. Wu, L. Liu, and Z. Wang, "Competitive remanufacturing and pricing strategy with contrast effect and assimilation effect," *Journal of Cleaner Production*, vol. 257, Article ID 120333, 2020.

[24] J. Zhao, C. Wang, and L. Xu, "Decision for pricing, service, and recycling of closed-loop supply chains considering different remanufacturing roles and technology authorizations," *Computers & Industrial Engineering*, vol. 132, pp. 59–73, 2019.

[25] T. Shi, D. Chhajed, Z. Wan, and Y. Liu, "distribution channel choice and divisional conflict in remanufacturing operations," *Production and Operations Management*, vol. 29, no. 7, pp. 1702–1719, 2020.

[26] M. Jin and Y. Zhou, "Does the remanufactured product deserve the same warranty as the new one in a closed-loop supply chain," *Journal of Cleaner Production*, vol. 262, Article ID 121430, 2020.

[27] Z. Xiang and M. Xu, "Dynamic game strategies of a two-stage remanufacturing closed-loop supply chain considering Big Data marketing, technological innovation and overconfidence," *Computers & Industrial Engineering*, vol. 145, Article ID 106538, 2020.

[28] Z. Wang, Y. Duan, and J. Huo, "Impact of trade-in remanufacturing policy and consumer behavior on remanufacturer decisions," *Sustainability*, vol. 12, no. 15, p. 5980, 2020.

[29] D. Jia and S. Li, "Optimal decisions and distribution channel choice of closed-loop supply chain when e-retailer offers online marketplace," *Journal of Cleaner Production*, vol. 265, Article ID 121767, 2020.

[30] X. Han, Y. Shen, and Y. Bian, "Optimal recovery strategy of manufacturers: remanufacturing products or recycling materials," *Annals of Operations Research*, vol. 290, no. 1-2, pp. 463–489, 2018.

[31] L. Xu, J. Shi, and J. Chen, "Pricing and collection rate for remanufacturing industry considering capacity constraint in recycling channels," *Complexity*, vol. 2020, Article ID 8391252, 13 pages, 2020.

[32] Y. Huang and Z. Wang, "Pricing and production decisions in a closed-loop supply chain considering strategic consumers and technology licensing," *International Journal of Production Research*, vol. 57, no. 9, pp. 2847–2866, 2018.

[33] D. Li, Y. Peng, C. Guo, and R. Tan, "Pricing strategy of construction and demolition waste considering retailer fairness concerns under a governmental regulation

environment," *International Journal of Environmental Research and Public Health*, vol. 16, no. 20, p. 3896, 2019.
[34] S. Rahmani, A. Haeri, and S. M. Hosseini-Motlagh, "Proposing channel coordination and horizontal cooperation in two competitive three-echelon reverse supply chains," *International Transactions in Operational Research*, vol. 27, no. 3, pp. 1447–1477, 2019.
[35] M. Wang, T. Tian, and X. Zhu, "Self-remanufacturing or outsourcing? Hybrid manufacturing system with remanufacturing options under yield uncertainty," *IEEE Access*, vol. 7, pp. 150642–150656, 2019.
[36] R. Kleber, M. Reimann, G. C. Souza, and W. Zhang, "Two-sided competition with vertical differentiation in both acquisition and sales in remanufacturing," *European Journal of Operational Research*, vol. 284, no. 2, pp. 572–587, 2020.
[37] Y. Zhang, W. Chen, and Y. Mi, "Third-party remanufacturing mode selection for competitive closed-loop supply chain based on evolutionary game theory," *Journal of Cleaner Production*, vol. 263, Article ID 121305, 2020.
[38] G. Raz, C. Druehl, and H. Pun, "Codevelopment versus outsourcing: who should innovate in supply chains," *IEEE Transactions on Engineering Management*, pp. 1–16, 2021.
[39] Z.-B. Wang, Y.-Y. Wang, and J.-C. Wang, "Optimal distribution channel strategy for new and remanufactured products," *Electronic Commerce Research*, vol. 16, no. 2, pp. 269–295, 2016.
[40] P. He, Y. He, and H. Xu, "Product variety and recovery strategies for a manufacturer in a personalised and sustainable consumption era," *International Journal of Production Research*, vol. 60, no. 7, pp. 2086–2102, 2021.
[41] P. He, Y. He, and H. Xu, "Channel structure and pricing in a dual-channel closed-loop supply chain with government subsidy," *International Journal of Production Economics*, vol. 213, pp. 108–123, 2019.
[42] A. Tajbakhsh and E. Hassini, "Performance measurement of sustainable supply chains: a review and research questions," *International Journal of Productivity and Performance Management*, vol. 64, no. 6, pp. 744–783, 2015.
[43] G. Esenduran, E. Kemahlıoğlu-Ziya, and J. M. Swaminathan, "Take-back legislation: consequences for remanufacturing and environment," *Decision Sciences*, vol. 47, no. 2, pp. 219–256, 2016.

Hindawi
Complexity
Volume 2022, Article ID 4918640, 12 pages
https://doi.org/10.1155/2022/4918640

Research Article

Research on the Flattening Method of an Energy-Based Fabric Deformation Model in Garment Design

Yuxiang Zhu ,[1,2] Gangqiang Li,[1,2] Yanjun Peng,[3] Mingmin Zhang,[4] Yingran Ma,[3] and Yuanhong Wang[3]

[1]*College of Information Engineering, Huanghuai University, Zhumadian, Henan 463000, China*
[2]*Henan Provincial Key Laboratory of Smart Lighting, Huanghuai University, Zhumadian, Henan 463000, China*
[3]*College of Computer Science and Technology, Shandong University of Science and Technology, Qingdao, Shandong 266590, China*
[4]*College of Computer Science and Engineering, Zhejiang University, Hangzhou, Zhejiang 310007, China*

Correspondence should be addressed to Yuxiang Zhu; zhuyuxiangcn@163.com

Received 30 October 2021; Revised 28 December 2021; Accepted 1 April 2022; Published 16 April 2022

Academic Editor: Feng Li

Surface flattening plays an important role in the whole process of garment design. We proposed a novel method by using three-dimensional triangle mesh flattening in this study. First, the three-dimensional triangle mesh is flattened to a two-dimensional plane to approximate the original surface. The initial flattening results are then used as preliminary guesses for subsequent optimizations. Considering that the deformation energy in the real woven fabric is related to tensile or shear deformation, a simplified fabric deformation model based on energy is proposed to update the energy distribution to determine the best two-dimensional pattern. An innovative unified axis system process is proposed to obtain the deformation energy, and energy relaxation in local flattening is proposed to release the distortion of flattening. Finally, the experimental results show that complex surfaces such as garments could achieve better flattening results. Compared with other energy-based methods in garment design, our proposed methods are more flexible and practical.

1. Introduction

Three-dimensional computer-aided design(CAD) tools have developed rapidly in the fashion design industry [1, 2] and achieved better results in recent years. However, what garment industry need is a two-dimensional design rather than three-dimensional model. So, surface flattening [3–7] plays an important role in the whole garment design process. Almost all garments are nondevelopable [8], and the flattening of a nondevelopable is a complicated process and local distortions are inevitable. While, the existing surface flattening methods for triangulated surfaces have several limitations. A major problem is the existence of gaps and overlaps in the resulting mesh.

For this phenomenon, a flattening method of energy-based fabric deformation model is proposed in this study. The area representation method is more likely to prevent the triangle area from being reduced to zero during flattening than models that only consider edge deformation energy. Therefore, the fabric energy acquisition method based on area representation is adopted. Traditional energy-based fabric deformation models [9] need to be converted to weft and warp axes before energy can be obtained, which can be cumbersome. In this article, a simplified model of fabric deformation based on energy is proposed. A simplified energy-based fabric deformation model simplifies the calculations; at the same time, the energy relaxation in local flattening eliminates local deformation.

The properties of the materials need to be considered when flattening, so energy-based or physics-based methods are generally used in the flattening process. In the field of computer graphics, flattening is mainly used to expand a three-dimensional grid into a two-dimensional grid to facilitate texture mapping. This process is also called grid parameterization. To obtain good texture mapping quality, the mesh should be prevented from being deformed as much as possible, and the flattening method based on the problem of minimizing the deformation is often used. There are many flattening methods [10–15], and we briefly summarize the development of flattening methods over the past few decades.

Levy [16] proposed a quasi-conformal parameterization method based on least squares approximation. This method uses an objective function to minimize the angular deformation. First, the complex surface is decomposed into a series of developable surface elements, and then these developed surfaces are decomposed into another series of developable surface elements. After the elements are expanded, they are combined to obtain the final parameterized result. When the elements are expanded, the original orientation of the triangles can be maintained, so there will be no triangle inversion. However, when the boundaries of the elements intersect in the texture space, they may overlap. In this case, we could subdivide the face element until there is no overlap. Wang et al. [17] proposed a new wire warping method for flattening surfaces using feature curves retained by length. In addition, the progressive deformation scheme for local shape control and the global deformation scheme for highly curved surfaces were introduced. Since the invariant length of feature curves were used to generate the 2D patterns, these strictly controlled lengths were inappropriate when highly nondevelopable 3D surfaces were required. Zhang et al. [18] improved Wang's wire warping algorithm by replacing the constant length of the feature curve with an elastic feature curve. We could use this new feature curve to control the shape of a flattening 2D pattern.

Energy-based or physics-based methods are often used in surface flattening. There are many methods to obtain energy [19, 20]. In [21], Afzal proposed a neutrosophic statistical approach to analyze the resistance of conducting material for big data. In [22], Afzal adopted neutrosophic statistical methods; they were more informative, flexible, and adequate than classical statistics for analyzing the measured values of data. McCartney et al. [23] proposed a flattening algorithm to obtain a planar development by minimizing strain energy. The algorithm uses an energy model to deform the edges of the triangle mesh. In addition, darts or gussets can be handled in a triangular mesh to assist in the fitting process. Based on McCartney's algorithm, Wang et al. [24] introduced a spring-mass system from a geometric point of view, in which the forces generated by the elastic deformation energy stored in the spring-mass system could be used to move these points. The Lagrange equation was applied to release energy and the penalty function was used to prevent overlap. In addition, the energy distribution is displayed using the interpolation function, which is determined by the surface cutting line. Li et al. [25] introduced cross-springs to reduce the deformation of the final surface. The central triangle improves the levelling efficiency and proposes a local correction method to overcome the overlapping problem. Zhong et al. [26] introduced a novel method of surface flattening by opening the curved configuration of each winged triangle pair. The 2D final pattern is generated on the collision plane, and the strain control mechanism is introduced to preserve the area and size of the original 3D surface.

2. The Flattening Process of the Energy-Based Fabric Deformation Model

2.1. A Simplified Energy-Based Fabric Deformation Model. Fabric deformation produces energy. Because of shearing, bending, gravity, and tension, a piece of fabric is thought to have an energy component. Each energy component is represented by one term in the energy equation [24]:

$$E_{\text{total}} = E_{\text{gravitional}} + E_{\text{tensile}} + E_{\text{shear}} + E_{\text{bending}}, \quad (1)$$

where E_{total} is the energy of the entire fabric, $E_{\text{gravitional}}$ is the potential energy of the fabric, E_{tensile} represents the energy produced by tensile energy or compression of the fabric, E_{shear} represents the energy produced by shear, and E_{bending} is the energy produced by bending of the fabric.

In the process of flattening, a single triangle is flattened at a time. The triangle of the 3D shape is designed to be fixed, and its corresponding 2D triangle is variable during the flattening process. Therefore, the energy model will calculate the energy generated by transforming the triangle of the 2D shape to its corresponding 3D triangle. The summation of the shear energy and tensile strain energy indicates the total internal energy of a triangle [24]. Therefore, this fabric model only considers the variation associated with shear and tensile stress. At the same time, it is assumed that the energy generated by the bending is a constant. Tensile energy E_{tensile} is described as follows [24]:

$$E_{\text{tensile}} = \iint \frac{1}{2} K_{su} \left(S_u - 1\right)^2 \mathrm{d}u\mathrm{d}v + \iint \frac{1}{2} K_{sv} \left(S_v - 1\right)^2 \mathrm{d}u\mathrm{d}v = \frac{1}{2} A \left\{ K_{su} \left(S_u - 1\right)^2 + K_{sv} \left(S_v - 1\right)^2 \right\}, \quad (2)$$

where A is the area of the 2D triangle, K_{su} and K_{sv} are the tensile constants for the U and V directions, respectively, and S_u and S_v (see Section 4.1) are strains. Shear energy E_{shear} is described as follows [24]:

$$E_{\text{shear}} = \iint \left(\frac{1}{2} K_r \phi_v^2 \right) \mathrm{d}u\mathrm{d}v = \frac{1}{2} A K_r \phi_v^2. \quad (3)$$

Similar to (2), A is the area of the 2D triangle, K_r is the shear constant, and ϕ_v (see Section 4.1) is the strain.

To obtain the energy based on fabric deformation, a transition stage of converting 2D pattern coordinates to weft and warp coordinates is necessary in the traditional model. In addition, tedious calculations could take a considerable amount of time. To reduce the tedious work, we propose a novel simplified model that instead uses the variable R_{ratio} during this stage. Based on this simplified energy model, we propose an innovative unified axis system in Section 4.1 that will simplify computation and innovate energy relaxation in Section 4.2, which will increase the convergence speed. While energy data come from uncertain energy, we adopt some neutrosophic statistical methods [27, 28].

2.2. Flattening Process. Given a triangular mesh $\Omega(K, P)$, where $K = \{K_1, K_2, \ldots, K_n\}$ is the spatial vertex set, $K_i \in \mathfrak{R}^3$, $P = \{T_1, T_2, \ldots, T_n\}$ is the topological connection information of the mesh, and $T = \{v_i, v_j, v_k\}$ is the triangle of the surface mesh. The goal of mesh flattening is to obtain another triangular mesh $\Omega'(K', P)$, where $K' = \{K'_1, K'_2, \ldots, K'\}_n$ is a 2D plane point set, $K'_i \in \mathfrak{R}^2$, and K_i corresponds to K'_i. The area of each triangle and the length of each side remain unchanged after flattening. In addition, the entire algorithm flow is shown in Figure 1.

To be flattened, triangle list V, active triangle list A, flattened triangle list F, and initial flattening mesh $\Theta(K'', P)$ are introduced in an initial flattening algorithm.

Initial flattening algorithm (its aim is to obtain the initial flattening mesh Θ of Ω) is as follows:

(i) Step 1: Initially, add all triangles in Ω to V, and set A and F as empty.

(ii) Step 2: Seed triangle T_s is searched in list V and then T_s is flattened. Then, T_s is added to the active list A and removed from the available list V. The selection of seed triangle is detailed in Section 3.1.

(iii) Step 3: Perform a search in list V to find all the triangles that share an edge with T_s and insert them into the tail of list A.

(iv) Step 4: If the list A is empty, stop processing. Otherwise, obtain the next triangle T from the rest of list A.

(v) Step 5: Insert triangle T into the tail of list F and then mark T as flattened. A search is performed in list V to find all the triangles that share an edge with T and are not marked as flattened, and then they are inserted into the tail of list A. Return to Step 4.

To optimize vertex V_k and optimize triangle set T_{set}, initial vertex set K'' is introduced in the optimal flattening algorithm.

Optimal flattening algorithm (its aim is to obtain the optimal flattening mesh Ω' of Ω) is as follows:

(i) Step 1: Obtain the first vertex V_k from K''.

(ii) Step 2: A search is performed in list F to find the triangle set T_{set} that contains vertex V_k.

(iii) Step 3: Each 2D triangle of triangle set T_{set} proceeds with its corresponding 3D triangle of Ω in the unified axis system and then obtains the energy of vertex V_k.

(iv) Step 4: Calculate the minimum value of the energy at the original position of vertex V_k and its eight directions position. Thereafter, vertex V_k is moved to the minimum energy position.

(v) Step 5: If the vertices in K'' are all traversed, go to Step 6. Otherwise, take the next vertex V_k. Return to Step 2.

(vi) Step 6: Iterate Steps 1–5, until all vertex positions are unchanged.

3. Key Procedure for Initial Flattening

3.1. Selection of Seed Triangle. Before flattening, selecting the seed triangle is an important process. The order in which triangles are flattened depends on the selected seed triangle, so selecting the seed triangle can effectively reduce the number of iterations and running time. Li et al. [25] used integer indices to mark all vertices in the mesh. Also, Liu et al. [29] applied the integer index to label all triangular faces in the mesh. However, we use a novel labelling standard that relies on the triangular edges. In addition, the selection algorithm is detailed as follows:

(1) Set the triangle mesh edge index on the boundary to 0.

(2) Search all unlabelled triangle mesh edges that have the same vertices as the labelled edge and increase the unlabelled edge index number by 1.

(3) Repeat Step 2 until all triangular mesh edges are fully labelled.

(4) Select the largest sum of the three edge indices as the seed triangle. If there is more than one, choose one of them.

3.2. Unconstrained Triangle Flattening. In the phase of surface flattening, there are two kinds of flattening methods: the unconstrained triangle flattening method and the constrained triangle flattening method. When one edge of a triangle has been flattened, the third vertex can be located using the unconstrained triangle flattening method. As shown in Figure 2, $T(v_0, v_1, v_2)$ is a 3D triangle and the vertices are arranged in reverse order. Edge v_0v_1 has been flattened to edge $v'_0v'_1$ and then vertex v_2 is flattened. The angle between the vector $\overrightarrow{v_0v_1}$ and the vector $\overrightarrow{v_0v_2}$ is calculated as follows:

$$\theta = \arccos\left(\overrightarrow{v_0v_1}, \overrightarrow{v_0v_2}\right). \quad (4)$$

Vector τ is obtained by rotating vector $\overrightarrow{v'_0v'_1}$ counterclockwise by angle θ. Then, vertex v'_2 is located using the following expression:

$$v'_2 = v'_0 + \left|\overrightarrow{v_0v_2}\right|\tau. \quad (5)$$

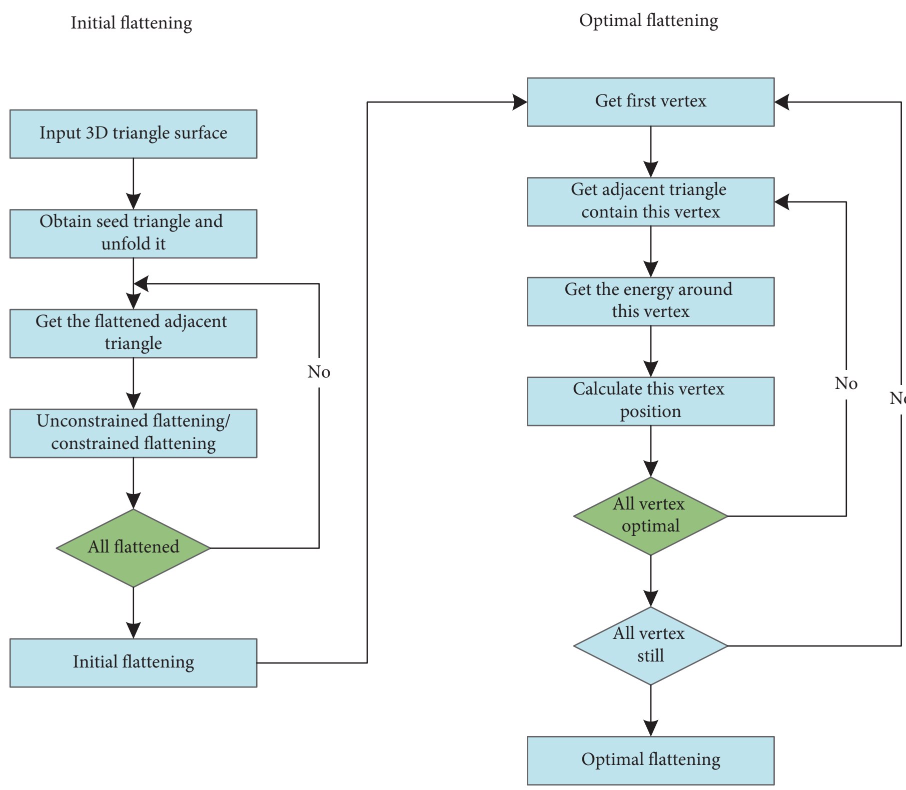


Figure 1: Algorithm flow chart.

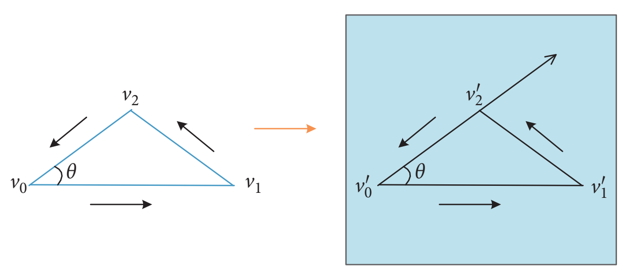


Figure 2: Unconstrained triangle flattening.

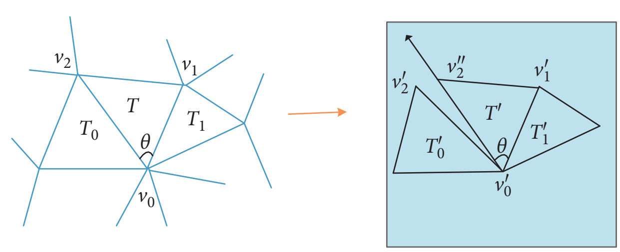


Figure 3: Novel constrained triangle flattening.

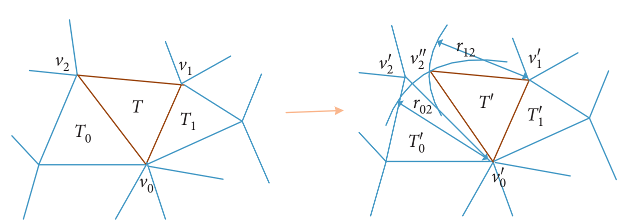


Figure 4: Traditional constrained triangle flattening.

As a result of flattening, the length of the triangle edge has no change without producing elastic potential energy, which is known as unconstrained triangle flattening.

3.3. Constrained Triangle Flattening. Figures 3 and 4 show two different constrained flattening methods, Figure 4 is a classic traditional method [25] of calculating the intersection point of two circles with two central points and two radii. These circles are centered at v_0' and v_1', and their radii are r_{02} and r_{12}, respectively, as shown in Figure 4. However, this method will produce two intersections. A point within the flattening range needs to be discarded. We use another method to find the third point, as shown in Figure 3. Triangles T_0 and T_1 have been flattened, and our aim is to flatten triangle $T(v_0, v_1, v_2)$. We choose edge $v_0'v_1'$, and vertex v_2 is flattened to v_2'' by using the unconstrained method in Section 3.2. Then we calculate their average coordinate as the flattened vertex, that is $v_2' = 1/2(v_2' + v_0'')$. The side length of the flattened triangle $T'(v_0', v_1', v_2')$ has been changed with respect to T, which will produce elastic potential energy. Therefore, after the initial flattening, optimized flattening is required to release the elastic potential energy.

4. Optimized Flattening

4.1. Process of Obtaining a Unified Axis System. The detailed process of obtaining a unified axis system is shown in

Figure 5. P is a 2D triangle and P' is its corresponding 3D triangle. First, the 2D woven triangle without distortion is placed on a uv axis system. The corresponding 3D deformed triangle is then superimposed on the system under the same conditions. The 2D fabric undeformed triangle P without energy, and its corresponding 3D fabric triangle P' with energy is generated after geometric deformation. The strain variables S_u and S_v, as well as the shear variable ϕ_v, transform position v_1 to v_1' and v_2 to v_2', respectively.

As shown in the right part of Figure 5, v_0 is located at the origin of the coordinates. The three-side length of the 2D triangle P is known, and the coordinates of v_1 and v_2 can be obtained using (6) and (7), where θ is the angle between axis u and edge v_0v_2. ∂ is the angle between edge v_0v_1 and edge v_0v_2. By setting the value of R_{ratio} as 1/2, the flattening could achieve a better effect. The v_1' and v_2' coordinates of the 3D triangles can be obtained in the same way.

$$R_{\text{ratio}} = \frac{|v_0v_2|\sin\theta}{|v_0v_2|\sin\theta + |v_0v_1|\sin(\partial-\theta)}, \tag{6}$$

$$\partial = \arccos\left(\frac{|v_0v_1|^2 + |v_0v_2|^2 - |v_1v_2|^2}{2|v_0v_1||v_0v_2|}\right). \tag{7}$$

The unifying process completes the specification geometric transformation that a representative rectangle must undergo. This geometric transformation can also deform the triangle P to P'. The total energy required for the distortion can be calculated by determining the energy to distort the representative weave element. Thereafter, considering it in its entirety, we utilized the same transformation for the triangle. The distortion of the representative woven element can be represented by an affine transformation. The scaling factor and shear angle can be calculated from affine transformation [30]. The general representation of an affine transformation is as follows:

$$[u'v'1] = [uv1]\begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & 0 \end{pmatrix}. \tag{8}$$

For a particular affine transformation, the affine transformation matrix will now be referred to as M, where this matrix is comprised as follows:

$$[u'v'1] = [uv1]M. \tag{9}$$

The shape of a triangular fabric is deformed first by stretching and then by shearing; M is therefore represented by a matrix:

$$M = \begin{pmatrix} S_u & 0 & 0 \\ 0 & S_v & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ \sin\phi_v & \cos\phi_v & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} S_u & 0 & 0 \\ \sin\phi_v & S_v\cos\phi_v & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{10}$$

For this special change, coordinate v_1 transforms to coordinate v_1' and coordinate v_2 transforms to v_2':

$$(v_{11}, v_{12}) \longrightarrow (v_{11}', v_{12}'), (v_{21}, v_{22}) \longrightarrow (v_{21}', v_{22}'). \tag{11}$$

The equation can be rewritten again as shown below:

$$\begin{pmatrix} 0 & 0 & 1 \\ v_{11}' & v_{12}' & 1 \\ v_{21} & v_{22}' & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ v_{11} & v_{12} & 1 \\ v_{21} & v_{22} & 1 \end{pmatrix}\begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & 1 \end{pmatrix}, \tag{12}$$

or $U' = UA$, and this results in:

$$A = \frac{1}{v_{11}v_{22} - v_{21}v_{12}} \times \begin{pmatrix} (v_{22}v_{11}' - v_{12}v_{21}') & (v_{22}v_{12}' - v_{12}v_{22}') & 0 \\ (v_{11}v_{21}' - v_{21}v_{11}') & (v_{11}v_{22}' - v_{21}v_{12}') & 0 \\ 0 & 0 & (v_{11}v_{22} - v_{21}v_{12}) \end{pmatrix}. \tag{13}$$

By comparing (10) and (13), the strain variables S_u and S_v and the shear variable ϕ_v are as follows:

$$S_u = \frac{(v_{22}v_{11}' - v_{12}v_{21}')}{(v_{11}v_{22} - v_{21}v_{12})}, \tag{14}$$

$$\phi_v = \tan^{-1}\frac{(v_{11}v_{21}' - v_{21}v_{11}')}{(v_{11}v_{22}' - v_{21}v_{12}')}, \tag{15}$$

$$S_v = \frac{(v_{11}v_{21}' - v_{21}v_{11}')}{(v_{11}v_{22} - v_{21}v_{12})\sin\phi_v}. \tag{16}$$

The steps for the energy of the triangular fabric deformation are given as follows:

(1) Obtain 3D coordinates of the triangle

(2) Obtain the 2D coordinates of the triangle

(3) Unify the 3D and 2D triangles into the (U, V) axes with equations (6) and (7)

(4) According to the two unified triangular coordinates, acquire the strain variables S_u and S_v and the shear variable ϕ_v by applying equations (14)–(16)

(5) Use equations (2) and (3) to obtain tensile energy E_{tensile} and shear energy E_{shear} respectively

4.2. Energy Relaxation Process in Partial Flattening. Energy relaxation plays an important role through the flattening process, which determines the efficiency of

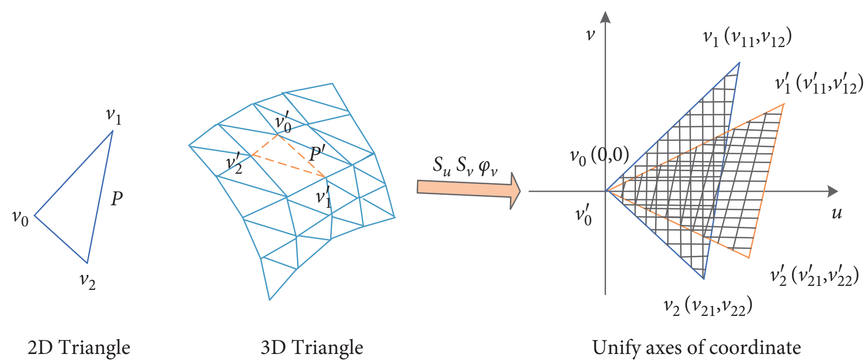


FIGURE 5: Process of unify axis coordinate.

flattening and the effect of flattening. In previous studies, most works have been conducted to determine how to use the energy released to achieve a satisfactory flattening result. Wang et al. [24] used the spring force obtained by the Lagrange equation to move the vertex, and a better result could be achieved after many iterations, but the obvious drawback is the time-consuming process. However, we propose an innovative method of simplifying fabric deformation based on energy in Section 2.1. As this method adds four orthogonal directions in the specific node, it shows the weighted optimization efficiency and optimization effect. The original position of the node is considered when calculating the next position of the node; the details of the process are described below.

As shown in Figure 6, we assumed that there are 2D vertices P and adjacent vertices $P_1, P_2, \ldots, P_5$. The triangles formed with P are $T_0, T_1, \ldots, T_4$. The energy of the vertex P is obtained by the following expression, where $E_{\text{energy}}(i)$ is the energy of the ith triangle connected to P and n is the number of triangles connected to vertex P.

$$E_P = \sum_{i=0}^{n-1} E_{\text{energy}}(i). \tag{17}$$

The eight directions of optimized flattening are $\overrightarrow{n}, \overrightarrow{n+1}, \ldots, \overrightarrow{n+7}$, and the movement increment is Δd. The next position of P will slip to the minimum energy value direction. The minimum energy value $E_{\min}$ can be obtained as follows:

$$E_{\min} = \min\left\{E_p, E_{\overrightarrow{op+n}}, E_{\overrightarrow{op+n+1}}, E_{\overrightarrow{op+n+2}}, \ldots E_{\overrightarrow{op+n+7}}\right\}. \tag{18}$$

5. Realization of the Virtual Try-On Process

The realization of the virtual stitching process is an important step of the virtual try-on. Before virtual garment is flattened, the virtual human body needs to be tried-on.

5.1. Virtual Stitching. When user selects the stitching edge, if the start point and the end point of the two corresponding edges are selected in the same direction (clockwise or counterclockwise), the corresponding relationship between the points on the two edges is correct, as shown in Figure 7(a). If different directions are chosen, that is, one side selects the start and end points in a clockwise direction, and the other side selects the start and end points in a counterclockwise direction, the corresponding relationship between the points on the two sides will be wrong, as shown in Figure 7(b).

The specific representation of stitching is shown in Figure 7(c). The brown dots represent discrete mass points of the garment piece, and Dis represents the distance between the two mass points. Among them, point A is applied with a stitching force in the direction of $\overrightarrow{AB}$, point B is applied with a stitching force in the direction of $\overrightarrow{BA}$, point A and point B will gradually approach each other. The stitching force is defined as a linear function corresponding to the distance between stitching points. For two stitching points A and B, the stitching force can be calculated as follows:

$$\overrightarrow{F_{\text{stitching force}}} = -k \bullet \overrightarrow{l}. \tag{19}$$

Among them, k is the stitching force coefficient, which is related to the stitching performance of the fabric. Generally, a larger stitching force coefficient is used for fabrics that are more difficult to deform; $\overrightarrow{l}$ represents the distance direction vector from the stitching point A to B. The closer the distance between the stitching points, the smaller the stitching force.

5.2. Virtual Try-On. The complete virtual stitching process is shown in Figure 8. First, seam lines are explicitly specified by choosing pairs of panel boundary edges. Then, the designed coats are assembled and linked by seaming lines to simulate clothing behavior on the 3D mannequin. By applying elastic force on the seam lines, the garment patterns can be connected to each other during the stitching process, as shown in Figure 8(a). After stitching, the virtual human completes the 3D virtual garment try-on, as shown in Figure 8(b).

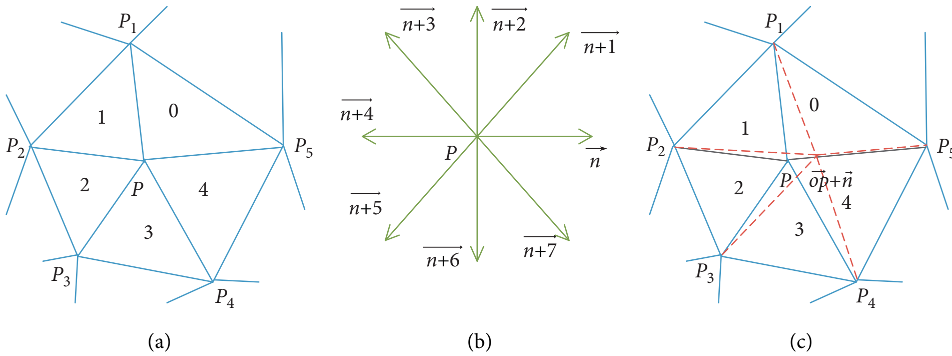


Figure 6: Energy relaxation procedure.

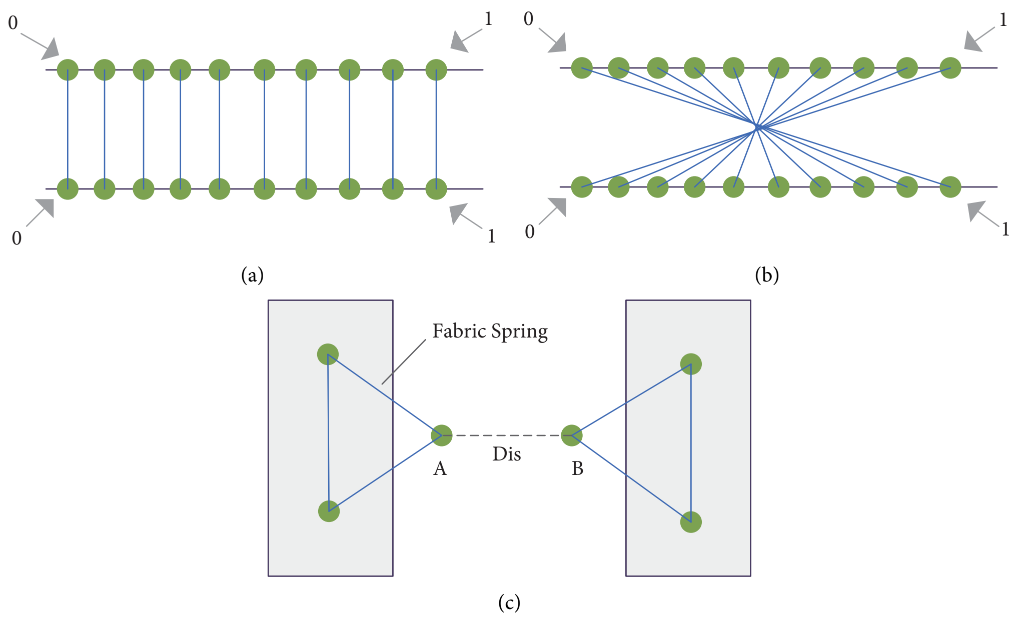


Figure 7: Diagram of virtual stitching information: (a)correct correspondence, (b)incorrect correspondence, and (c)diagram of stitching.

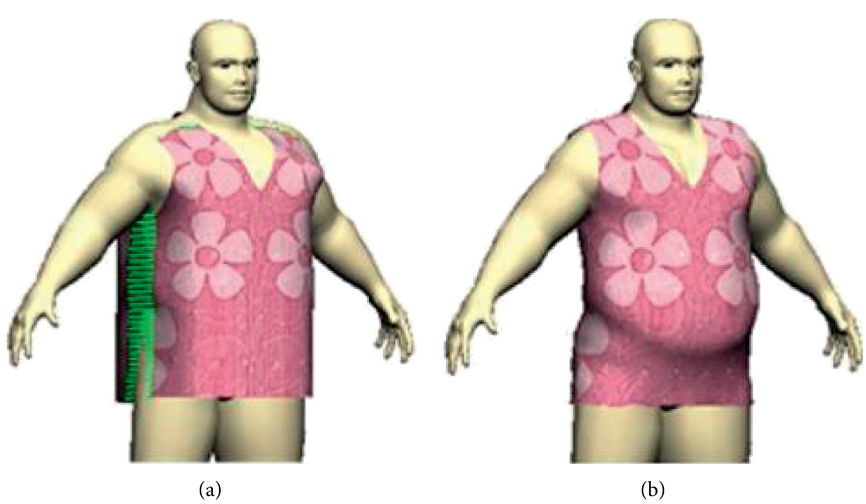


Figure 8: Diagram of virtual dressing process: (a) virtual stitching in progress and (b) virtual stitching completed.

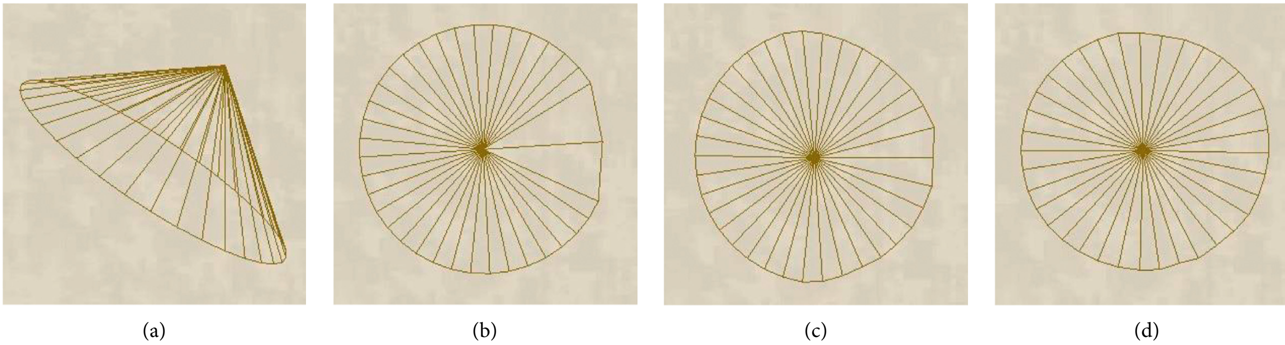


Figure 9: Optimized flattening for the 36-sided cone model: (a) 3D original Mesh, (b) initial flattening, (c) optimized iteration after 170 times, and (d) final flattening result.

6. Experiments and Discussion

This article introduces an energy-based fabric deformation algorithm, which is implemented by the object-oriented Visual C++ language. It is implemented on a PC with a Core i5 CPU and 4 GB RAM. A series of triangular mesh object surfaces are presented to measure the performance of our algorithm. As classical geometrically shaped parts of a torus, the garment shapes of coats and pants are selected as our example.

We used the typical 36-sided cone model as an example to test our optimized flattening method. The 3D model is shown in Figure 9(a). The initial flattening result is shown in Figure 9(b) and Figure 9(c) shows the optimized flattening results after 170 iterations. The final flattening result is shown in Figure 9(d).

Three criteria are used to evaluate the surface accuracy after flattening, including area accuracy, shape accuracy, and energy accuracy. The area accuracy and shape accuracy were proposed by Wang et al. [24]. The energy accuracy is presented in this study based on an energy-based fabric deformation model.

Area accuracy:

$$E_{\text{area}} = \frac{\sum_{i=0}^{n-1} \left| a_i - a_i' \right|}{\sum_{i=0}^{n-1} a_i}, \tag{20}$$

where E_{area} is the accuracy of the area after flattening the entire surface; a_i is the ith triangle area before flattening and a_i' is its corresponding triangle after flattening; and E_{energy} is the number of triangles.

Shape accuracy:

$$E_{\text{edge}} = \frac{\sum_{i=0}^{m=1} \left| e_i - e_i' \right|}{\sum_{i=0}^{m=1} e_i}, \tag{21}$$

where E_{edge} is the accuracy of the shape after flattening the entire surface; n is the ith edge length before flattening and e_i' is the length of its corresponding edge after flattening; and m is the number of triangular mesh edges.

Energy accuracy:

$$E_{\text{energy}} = \sum_{i=0}^{n-1} E_{\text{energy}}(i), \tag{22}$$

where E_{energy} is the energy accuracy of the entire surface after flattening; $E_{\text{energy}}(i)$ is the ith triangle energy after flattening, and the corresponding triangle energy before flattening is zero; and n is the number of triangles on the entire surface.

Table 1 shows the 36-sided cone measurement accuracy data using the three kinds of surface accuracy methods. As we can see from Table 1, the three measurement accuracies will decrease during the optimized flattening process, and the minimum value is achieved in the final process. The value of area accuracy, shape accuracy, and energy accuracy in the final process is 0.397%, 1.053%, and 5.1, respectively.

We compare our optimized method with the classical mass-spring energy relaxation method, and Figures 10(a)–10(d) show the flattening result of the partial torus. Figure 10(a) is the original 3D mesh model of a part of the torus. As shown in Figure 10(b), the initial flattening result has great distortion. Figure 10(c) is the result of optimization based on the fabric deformation model proposed in this study, and the result of mass-spring energy relaxation is shown in Figure 10(d). From the appearance point of view, the optimization result in Figure 10(c) is slightly better than that in Figure 10(d).

The performance accuracy is listed in Table 2. In our experiment, the area accuracy, shape accuracy, and energy accuracy after mass-spring energy relaxation are 1.201%, 1.533%, and 20.2 Nmm, respectively. The value of the area accuracy, shape accuracy, and energy accuracy after the woven fabric-based method is 1.073%, 1.342%, and 10.3Nmm, respectively.

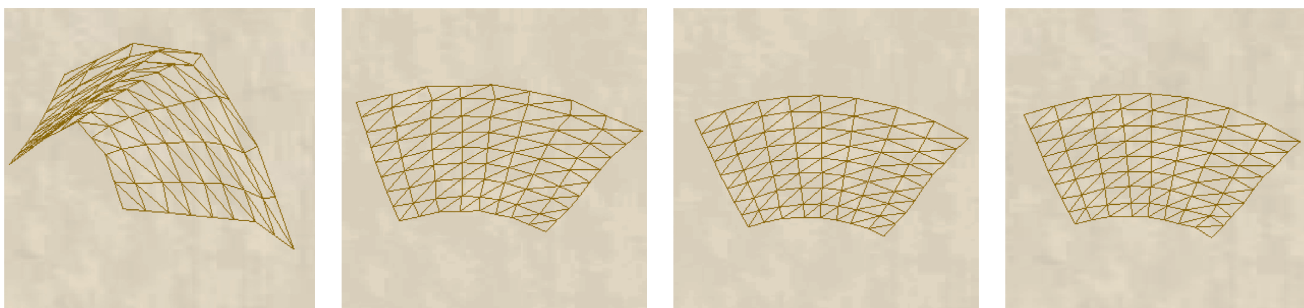
The following two groups of experiments are applied in the garment design process. Figure 11(a) shows the 3D triangular mesh surface of a skirt, with 3414 triangular faces and 1851 vertices. Figure 11(b) shows the flattening result after optimization.

Figure 12(a) shows the 3D triangular mesh surface of a coat, with 3492 triangular faces and 1965 vertices. Figure 12(b) is the result of its optimized flattening.

To illustrate the flattening effect of our method, we added two group experiments of pant and vest in the garment design process. Figure 13(a) shows the 3D triangular mesh surface of a pant, with 3852 triangular faces and 1963 vertices. Figure 13(b) shows the flattening result after optimization.

Table 1: Accuracy statistic of the 36-sided cone.

	Initial	Optimized (1)	Optimized (2)	Final
Area (%)	0.897	0.682	0.564	0.397
Shape (%)	1.642	1.435	1.211	1.053
Energy (Nmm)	23.1	12.2	7.9	5.1



(a) (b) (c) (d)

Figure 10: Comparison of optimization with the woven fabric method and mass-spring method: (a) original 3D mesh, (b) initial flattening, (c) optimized with woven fabric method, and (d) optimized with mass-spring method.

Table 2: Accuracy statistics of a torus.

	Initial	Woven fabric	Mass-spring
Area (%)	1.932	1.073	1.201
Shape (%)	2.53	1.342	1.533
Energy (Nmm)	62.1	10.3	20.2

(a) (b)

Figure 11: Energy-based flattening method for a skirt: (a) original 3D mesh of a skirt and (b) final 2D flattening pattern of a skirt.

(a) (b)

Figure 12: Energy-based flattening method for a coat: (a) original 3D mesh of a coat and (b) final 2D flattening of a coat.

(a) (b)

Figure 13: Energy-based flattening method for a pant: (a) original 3D mesh of a pant and (b) final 2D flattening pattern of a pant.

(a) (b)

Figure 14: Energy-based flattening method for a vest: (a) original 3D mesh of a vest and (b) final 2D flattening pattern of a vest.

Figure 14(a) shows the 3D triangular mesh surface of a vest, with 3752 triangular faces and 1857 vertices. Figure 14(b) is the result of its optimized flattening.

7. Conclusion and Future Work

In this article, an optimized flattening method based on the fabric energy deformation model is proposed. First, the 3D triangle mesh is flattened. The selection of seed triangles can effectively reduce the number of iterations and runs. A new markup algorithm using the edges of triangles is adopted, and the constraint triangle flattening method generates the elastic potential energy and then optimizes the initial flattening results. Second, in the process of optimizing flattening, a new unified axis system process is proposed to obtain the deformation energy, and an energy relaxation method is used to release energy to eliminate the local deformation caused by initial flattening. Finally, based on the model of fabric deformation, which is determined by the energy, a new energy precision is proposed in this study to evaluate the surface accuracy after flattening. The experimental results show that the complex surface of clothing can obtain a better flattening effect. Our approach is more flexible and practical in clothing design than other energy-based methods. In the future, we will consider different materials and other physical parameters of the material in the optimized flattening process for better flattening results.

Data Availability

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

Disclosure

The authors have presented the manuscript in "2017 International Conference on Virtual Reality and Visualization (ICVRV)."

Conflicts of Interest

No conflicts of interest exist in the submission of this manuscript.

Acknowledgments

The authors thank Prof.Yanjun Peng at the Shandong University of Science and Technology for his valuable comments, Dr. Yingran Ma at the Shandong University of Science and Technology for his data collection, Mrs. Yuanhong Wang at the Shandong University of Science and Technology for her literation collection, and Assoc Prof. Mingmin Zhang at the Zhejiang University for discussion. This work was supported by the Key Science and Technology Research of Henan Province under Grant nos. 212102210516 and 212102210255, the National Natural Science Foundation of China under Grant no. 61976126, and the Natural Science Foundation of Shandong Province under Grant no. ZR2019MF003.

References

[1] N. Umetani, D. M. Kaufman, T. Igarashi, and E. Grinspun, "Sensitive couture for interactive garment modelling and editing," *ACM Transactions on Graphics*, vol. 30, pp. 1–11, 2011.

[2] A. Bartle, A. Sheffer, V. G. Kim, D. M. Kaufman, and N. Vining, "Physics-driven pattern adjustment for direct 3D garment editing," *ACM Transactions on Graphics*, vol. 35, no. 4, pp. 1–11, 2016.

[3] S. Lu, P. Y. Mok, and X. Jin, "A new design concept: 3D to 2D textile pattern design for garments," *Computer-Aided Design*, vol. 89, pp. 35–49, 2017.

[4] H. Q. Huang, P. Y. Mok, Y. L. Kwok, and J. S. Au, "Block pattern generation: from parameterizing human bodies to fit feature-aligned and flattenable 3D garments," *Computers in Industry*, vol. 63, pp. 680–691, 2012.

[5] Y. J. Liu, D. L. Zhang, and M. M. F. Yuen, "A survey on CAD methods in 3D garment design," *Computers in Industry*, vol. 61, pp. 576–593, 2010.

[6] C. K. Au and Y. X. Ma, "Garment pattern definition, development and application with associative feature approach," *Computers in Industry*, vol. 61, no. 6, pp. 524–531, 2010.

[7] S. W. Hsiao, R. Q. Chen, and C. Y. Lin, "Surface flattening assisted with 3D mannequin based on minimum energy," *World Academy of Science, Engineering and Technology*, vol. 8, pp. 1444–1449, 2014.

[8] R. Goldman, "Curvature formulas for implicit curves and surfaces," *Computer Aided Geometric Design*, vol. 22, pp. 632–658, 2005.

[9] J. McCartney, B. Hinds, and K. W. Chong, "Pattern flattening for orthotropic materials," *Computer Aided Design*, vol. 37, pp. 631–644, 2005.

[10] A. Sheffer, B. Lévy, M. Mogilnitsky, and A. Bogomyakov, "ABF++: fast and robust angle based flattening," *ACM Transactions on Graphics*, vol. 24, no. 2, pp. 311–330, 2005.

[11] Y. C. Yang and W. Y. Zhang, "Prototype garment pattern flattening based on individual 3D virtual dummy," *International Journal of Clothing Science and Technology*, vol. 19, pp. 334–348, 2007.

[12] Y. Yang, W. Zhang, and C. Shan, "Investigating the development of digital patterns for customized apparel," *International Journal of Clothing Science and Technology*, vol. 19, pp. 167–177, 2007.

[13] A. Jung, S. Hahmann, D. Rohmer, A. Begault, L. Boissieux, and M. P. Cani, "Sketching folds: developable surfaces from non-planar silhouettes," *ACM Transactions on Graphics*, vol. 34, pp. 1–12, 2015.

[14] C. C. L. Wang, "Towards flattenable mesh surfaces," *Computer Aided Design*, vol. 40, pp. 109–122, 2008.

[15] J. Hoschek, "Approximation of surfaces of revolution by developable surfaces," *Computer Aided Design*, vol. 30, pp. 757–763, 1998.

[16] B. Lévy, S. Petitjean, N. Ray, and J. Maillot, "Least squares conformal maps for automatic texture atlas generation," *ACM Transactions on Graphics*, vol. 21, pp. 362–371, 2002.

[17] C. C. L. W. W. Wang, "A fast surface flattening approach with length-preserved feature curves," *Computer Aided Design*, vol. 40, pp. 381–395, 2008.

[18] Y. B. Zhang and C. C. L. Wang, "WireWarping++: robust and flexible surface flattening with length control," *IEEE Transactions on Automation Science and Engineering*, vol. 8, pp. 205–215, 2011.

[19] M. A. U. Haq, S. R. Gadde, M. Albassam, and M. Aslam, "Marshall-olkin power lomax distribution for modeling of wind speed data," *Energy Reports*, vol. 6, pp. 1118–1123, 2021.
[20] M. Aslam and A. Algarni, "Analyzing the solar energy data using a new Anderson-Darling test under indeterminacy," *International Journal of Photoenergy*, vol. 2020, Article ID 6662389, 6 pages, 2020.
[21] U. Afzal, H. Alrweili, N. Ahamd, and M. Aslam, "Neutrosophic statistical analysis of resistance depending on the temperature variance of conducting material," *Scientific Reports*, vol. 11, no. 1, Article ID 23939, 2021.
[22] U. Afzal, N. Ahmad, Q. Zafar, and M. Aslam, "Fabrication of a surface type humidity sensor based on methyl green thin film, with the analysis of capacitance and resistance through neutrosophic statistics," *RSC Advances*, vol. 11, pp. 38674–38682, 2021.
[23] J. McCartney, B. K. Hinds, and B. L. Seow, "The flattening of triangulated surfaces incorporating darts and gussets," *Computer Aided Design*, vol. 31, pp. 249–260, 1999.
[24] C. C. L. Wang, S. S. F. Smith, and M. M. F. Yuen, "Surface flattening based on energy model," *Computer Aided Design*, vol. 34, pp. 823–833, 2002.
[25] J. Li, D. Zhang, G. Lu, Y. Peng, X. Wen, and Y. Sakaguti, "Flattening triangulated surfaces using a mass-spring model," *The International Journal of Advanced Manufacturing Technology*, vol. 25, no. 1-2, pp. 108–117, 2005.
[26] Y. Zhong and B. Xu, "A physically based method for triangulated surface flattening," *Computer-Aided Design*, vol. 38, no. 10, pp. 1062–1073, 2006.
[27] M. Aslam, "Analyzing wind power data using analysis of means under neutrosophic statistics," *Soft Computing*, vol. 25, no. 10, pp. 7087–7093, 2021.
[28] M. Aslam, M. Azam, and M. Albassam, "Sampling plan using process loss index using multiple dependent state sampling under neutrosophic statistics," *IEEE Access*, vol. 7, pp. 38568–38576, 2019.
[29] Q. S. Liu, J. T. Xi, and Z. Q. Wu, "An energy-based surface flattening method for flat pattern development of sheet metal components," *International Journal of Advanced Manufacturing Technology*, vol. 68, pp. 1155–1166, 2013.
[30] P. J. Schneider and D. H. Eberly, *Geometric Tools for Computer Graphics,*pp. 98-99, Morgan Kaufmann Publisher, Burlington, MA, USA, 2002.

Hindawi
Complexity
Volume 2022, Article ID 9986137, 20 pages
https://doi.org/10.1155/2022/9986137

Research Article

Comparative Analysis of TOPSIS and TODIM for the Performance Evaluation of Foreign Players in Indian Premier League

Vaishnudebi Dutta,[1] **Subhomoy Haldar**,[1] **Prabjot Kaur**,[1] **and Yuvraj Gajpal**[2]

[1]*Department of Mathematics, Birla Institute of Technology, Mesra, Ranchi, Jharkhand, India*
[2]*Supply Chain Management, Asper School of Business, University of Manitoba, Winnipeg, Canada*

Correspondence should be addressed to Prabjot Kaur; prabjotkaur@bitmesra.ac.in

Received 7 July 2021; Revised 29 October 2021; Accepted 15 March 2022; Published 10 April 2022

Academic Editor: Qingyuan Zhu

Sports officials, players, and fans are concerned about overseas player rankings for the IPL auction. These rankings are becoming progressively essential to investors when premium leagues are commercialized. The decision-makers of the Indian Premier League choose cricketers based on their own experience in sports and based on performance statistics on several criteria. This paper presents a scientific way to rank the players. Our research examines and contrasts different multicriteria decision-making algorithms for ranking foreign players under various criteria to assess their performance and efficiency. The paper uses three MCDM algorithms, TOPSIS, TODIM, and NR-TOPSIS, for foreign players ranking in the Indian Premier League. Our analysis is limited to the batsmen and bowlers only. We perform the analysis using Python language, a popular high-level programming language. Finally, we perform a sensitivity analysis to determine the stability of each method when the weights of the criterion or the value of a parameter was changed. Our analysis exhibits the superiority of TODIM over TOPSIS and NR-TOPSIS.

1. Introduction

The Indian Premier League (IPL), established in 2007 by the BCCI (Board of Control for Cricket in India), is a professional twenty-team cricket league in India, with eight teams from eight different cities. Each team in the Indian Premier League has only four foreign players in its starting eleven for any match according to the IPL Player Regulations [1], as well as a maximum of eight overseas players in the entire team [2]. A team can get players in one of three ways: through the annual player auction, exchanging players with other teams during trading periods, or signing substitutes for players who are unavailable. Players sign up for the auction and establish their starting price, following which the franchise with the highest offer buys them. For proper bidding, each player's price is determined by their individual outcomes; franchisee owners have access to all the statistics. They invest in an appropriate team of players, aiming to earn a profit through the cricket match as prize money, sponsorships, and other forms of income. The decision-maker's job of selecting the best players in a conflicting situation is often challenging, since many qualitative and quantitative qualities must be included in the player selection process. Several sporting activities have been commercialized for decades, yet IPL is India's richest sport, and it is only getting richer. We considered the selection of overseas players for a few reasons. First, there might be a regional preference for domestic players making their debut. These preferences complicate the selection process. Second, comparing overseas players mitigates internal biases of the authors. Our approach assumes no preference for any team for any of the players being ranked. At the auction, it is entirely up to the decision-makers and team owners.

In this paper, we have studied three different multi-criteria decision-making (MCDM) algorithms, namely, TOPSIS (Technique for Order of Preference by Similarity to

Ideal Solution), an improvement on it called NR-TOPSIS, and TODIM (*TOmada de DecisãoInterativa e Multicritério*, Portuguese acronym for Interactive multicriteria decision-making). There are various other methods available such as *ELimination Et Choix Traduisant la REalité* (ELECTRE), Preference Ranking Organization Method for Enrichment of Evaluations (PROMETHEE), and *VIseKriterijumskaOptimizacija I KOmpromisnoResenje* (VIKOR), whose comparative analysis can be performed later.

The principles of the two MCDM algorithms are different, as TODIM employs Prospect Theory, whereas TOPSIS makes use of normalised n-dimensional Euclidean distance. We proved the two approaches to be contrasting enough to make an interesting comparison. TOPSIS uses a compensatory approach that allows trade-offs between criteria, where a bad outcome in one criterion may be offset by a favourable outcome in another as studied here [3]. It is preferred over ELECTRE I because of better consistency in the results [4]. Although TOPSIS gives various benefits [5], including simplicity, logic, comprehensibility, computing efficiency, and the ability to measure relative performance for each alternative in a simple mathematical form, in our study, we employed a crisp and accurate dataset. Consequently, we did not make use of fuzzy TOPSIS, which is a superior strategy for imprecise or ambiguous performance assessments [6].

Our article shows how TODIM outperforms TOPSIS in multicriteria decision-making. This approach may be used by decision-makers of all levels of experience over TOPSIS whose applications in IPL performance analysis were made by various authors. The sensitivity analysis of TODIM provided in this article does not depict any major change in rank of the foreign players when the weights of the criteria (or preferences of criteria) are shuffled, which makes the algorithm more feasible than TOPSIS. For robustness and stability, we have proved that TODIM is a better and more trustworthy MCDM approach.

Compared to the other methodologies, TODIM, which was formed by Gomes and Lima in 1992 [7], has the benefit of considering the subjectivity of decision-makers' (DM) actions and giving the dominance of one alternative over others with specific operation formulae. TODIM is claimed to be more rational and scientific in the application of MCDM problems and various extensions of TODIM have been studied in fuzzy environment [8, 9]. To the best of our knowledge, this is the first study to employ TODIM for evaluating the performance of international players in the IPL and prove the superiority of the algorithm over TOPSIS.

In our article, we compared TOPSIS, which is a more popular MCDM algorithm than TODIM. We claim so with the help of "Google Trends," which allows us to compare various keywords searched over the Internet [10]. We inferred from the findings that TODIM applications have been examined far less than TOPSIS since 2004. For a crisp dataset, however, our study clearly shows that TODIM outperforms TOPSIS for evaluating alternatives based on many criteria. Other than that, we have also explored several other existing articles where research scholars have studied TOPSIS and TODIM from various aspects such as Choquet based TOPSIS and TODIM [11], TOPSIS and TODIM with Z-numbers [12], and fuzzy TOPSIS-TODIM hybrid method [13]. Some also proved in their article that TOPSIS, TODIM, and PROMETHEE are methods that have higher utility in the selection of mining method in the coal mining industry [14]. Comparative studies have also been made for fuzzy AHP (TOPSIS is an additive variant of AHP) and fuzzy TODIM [15].

As both TOPSIS and TODIM are applied at their most basic stages in this paper, without implementation of any extension, decision-makers can readily understand them. Future studies and sensitivity analyses of other MCDM approaches can be conducted to determine the superiority of other methods over TODIM. One fact to note is that both TOPSIS and TODIM are said to have rank reversal problem, which is a shift in the rank ordering of the preferability of alternative feasible decisions. Therefore, we tried to implement NR-TOPSIS [16], which in the article is proved to solve the rank reversal problem in TOPSIS. However, because we have numerous alternatives and criteria and a different approach to calculate criteria weights, NR-TOPSIS does not address the problem in our situation. Instead, we observed that the approach is significantly less stable than both TOPSIS and TODIM. All calculations and results in this paper are done by implementing the algorithm in Python, which is provided with this article so that other authors can use them as well. Finally, while TOPSIS has historically been used to rank players in the IPL, TODIM is not employed as frequently. As a result, we picked these two approaches and used sensitivity analysis to show how TODIM is superior to both TOPSIS and NR-TOPSIS.

We employed rank-sum method for generating weights for the criteria. Because of this, the weights vary significantly based on the preferences of the decision-maker. It is also simpler for a regular user to rank the criteria in order of their preference rather than calculating the numeric value of the weights manually. Owing to this significant variance, the results generated by TOPSIS do, in fact, vary based on subjective preferences. The same holds true for TODIM. Personal preferences are not included in TOPSIS, but they are in TODIM, since it employs the idea of prospect theory. However, as we used the rank-sum weight approach that relies on the preferences of decision-makers, as a result, the TOPSIS outcome will be influenced, exposing it to subjective preferences such as TODIM.

The following is how the rest of the paper is organised: Section 2 contains a review of the literature on the MCDM methodologies used in this work, and Section 3 describes the methodology, which explains the processing done in a step-by-step fashion. The application of the methodology explained in Section 3 is explained through the case study in Section 4. In Section 5, we discuss results we had obtained from the case study. The next section, that is, Section 6, is dedicated to discussion and interpretation of the results, followed by future perspectives, and we conclude the paper in Section 7. At last, we also provide Supplementary Materials, containing Python code and intermediate steps for the calculations performed, followed by a list of references used in this paper.

2. Literature Review

The fundamental concept for selecting a foreign player for any club originates from an article by Xing [17], in which a vast amount of data demonstrates that different decision-makers base their selections on prior competition scores and games played. The technical data supplied is frequently used to judge several international players. Multicriteria decision-making (MCDM) is a common method in performance analysis. Thus, several studies on various MCDM methods to analyse IPL team performance analysis have been conducted. Multicriteria group decision method studied in [18] gives a practical way to evaluate a team's success while producing consistent performance ratings. Multicriteria decision tree approach [19] can classify all-rounders in the Indian Premier League for accurately and efficiently classifying data based on the output of players. It is also shown how multicriteria decision tree method provided a good image of the players in several categories, including performer, batting all-rounder, bowling all-rounder, and underperformer, aiding franchisee owners in picking all-rounders in the auction and compensating them depending on their performance.

Application of MGDA (Modified Group Decision Algorithm) is studied further [20] to analyse batsman, fast bowler, and spin bowler statistics from IPL sessions IV, V, and VI separately, and it consistently produced reliable results. The consistency testing property of AHP is used to calculate the weights of the criterion, proving its correctness. The hybridization of two well-known multicriteria decision-making (MCDM) approaches with their classifications and characteristics has been explored [21]. The goal was divided into two parts: First, prioritize the weight of selected parameters for evaluating the players' output using Analytic Hierarchy Process (AHP). Second, TOPSIS is used to select the best alternative strategy for ranking the teams. Performance-based analysis is studied among IPL batsmen in seasons I (2008), II (2009), and III (2010) using statistical multicriteria decision-making [22]. They concluded that international players performed well in season II, but their performance varied from season to season. In case of IPL, a team has the liberty to retain their previously auctioned players, but it lowers the funds for the owner to enter the auction. A decision tree is made use of to compute the "most valuable player" for a team by player's batting and bowling points and experiences [23]. Prediction of match is calculated through various match statistics using TOPSIS in multiple studies based on World Cup 2019 [24].

Other related recent works in sports include the applications of Bayesian BWM (Best Worst Method) and rough DEMATEL (Decision-MAking Trial and Evaluation Laboratory), which is a type of MCDM algorithm, investigating the impact of Sustainable Sports Tourism Criteria in Taichung City [25]. Applications of fuzzy-ANP and DEMATEL [26], ANP-DEMATEL, and VIKOR (*VIseKriterijumskaOptimizacija I KOmpromisnoResenje*) [27] are also studied in various sports business applications and helped in creating a management strategy.

Another outranking approach known as TODIM (*TOmada de DecisãoInterativa e Multicritério*, an acronym in Portuguese for iterative multicriteria decision-making) is studied in this article. We may use this MCDM technique to discover foreign player rankings in IPL 2019 as it uses prospect theory studied by two Israeli psychologists in 1979 [28, 29] as well to create a multiattribute value function. The purpose of their study was to evaluate human behaviour during decision-making and in high-risk situations. The two psychologists then discovered that, in scenarios involving advantages, humans have a propensity to be more conservative in terms of risk; that is, individuals prefer to select for a smaller, more secure gain than taking a risk to earn a larger benefit.

The systematic flowchart diagram and highlighted formulae in article [30] were used to determine the weight, four normalization techniques, and other details for TODIM, making the algorithm clearer. The comparison between TODIM and modified TODIM on recycled water alternatives based on a range of parameters were mainly studied. The formulae and tabular procedures are likewise based on research work mentioned in article [31]. The weights of the criteria and an assessment matrix that aided us in our implementation were explained. We also referred to another article [32] where applications of TODIM are shown to evaluate broadband Internet services. For TODIM sensitivity analysis, we used the study from article [33], where the authors looked at different values of θ (loss of attenuation factor) in the space curved surfaces of distributor dominance in four distance equations. As a result, how changes in the value of θ have an impact on the alternatives was evaluated. For TOPSIS, we have shuffled the ranks of the criteria, thereby changing the weight for which we referred to related article [34].

Although TODIM is an emerging method of MCDM which is not as popularly used as TOPSIS and its extensions, there are recent applications outside the sports world [35–38], where the implementation is based on MCGDM, stock investment selection and assessment of hydro energy storage plant, and green supplier selection problems. Other studies include applications of failure mode and effect analysis (FMEA) and TODIM to demonstrate risk ranking of wind turbine systems [39], extended fuzzy TODIM with dual connection numbers [32], and Pythagorean fuzzy TODIM based on cumulative prospect theory to assess the risk in science and technology project [40]. Progressive studies have also been made on development of TODIM with different fuzzy sets [41]. But there are a handful of research articles where TODIM is implemented to analyse the performance, directly based on the scores achieved by the players in any sport. In our article, we studied the IPL 2019 dataset, but similar applications can also be made in other areas of sports, such as football clubs, golf clubs, and gymnastics.

When a choice alternative is added or removed, rank reversal occurs, which implies that the relative ranks of two decision alternatives can be reversed. Belton and Gear were the first to point out such a phenomenon [42]. Although the rank reversal problem may not always have a detrimental influence on many datasets, the decision-makers in our situation may not like to have such an issue when removing, adding, or replacing a player can affect the entire rank list of players. As a result, we examined various recent and previous studies in order to improve the MCDM technique and avoid the problem of rank reversal. Several methods exist such as a combination of Characteristic Objects Method (COMET) and TOPSIS or PROMETHEE II [43], improvement of VIKOR method using R-VIKOR [44], and G-TOPSIS (Gaussian TOPSIS) method for rank reversal problem [45]; analytical studies have also been made [16, 46–49] to find a solution to the problem for various applications. Rank reversal problem for TODIM was first discussed in 1990 [50], where the author compared the problem with AHP and tried to provide a solution for the problem. But few modifications or developments have been made on TODIM rank reversal problem.

In this paper, we tried to implement the work of Yang, where R-TOPSIS was modified [51] and the new method was named NR-TOPSIS [52] to remove the rank reversal problem of TOPSIS and compare the results with TODIM. Unfortunately, the approach did not work, and we were still dealing with rank reversal. Instead, we could show and infer that when players are added or deleted from the list, TODIM's ranking changes relatively little. As a result, the MCDM method becomes more appealing, demonstrating the necessity of focusing on solving the rank reversal problem in TODIM in order to get supremum outcomes.

3. Methodology

3.1. Weighting through Rank-Sum Weight Method. This subjective method establishes weights solely based on the decision-makers' considerations or judgments [53]. It may be easier to rank order the relevance of criteria than it is to describe other inaccurate weights, such as bounded weights, for example, in instances involving time constraints, the nature of the criteria, a lack of expertise, imprecise or partial information, and the decision-maker's limited attention or information processing skills. Because a group of decision-makers may not agree on a set of exact weights, it may be reasonable to assume agreement merely on a ranking of weights in such a case as stated by article [54]. This rank order weight approach entails two steps: first, ranking the criteria by significance and then weighting the criteria using the formula. In this paper, we have used the rank-sum weight method which was proposed by Stillwell [55]. In this method, individual ranks are normalised by dividing the sum of the rankings in the rank-sum (RS) technique. The weights are calculated using the following formula:

$$w_j(\mathrm{RS}) = \frac{2\left(n+1-r_j\right)}{n(n+1)}. \tag{1}$$

Here, r_j is the rank of the jth criteria, $j = 1, 2, \ldots, n$.

If there are multiple decision-makers, the ranks can be determined through discussion. Alternatively, rankings of the individual decision-makers can be added together and averaged to get the final rank values. It is not strictly necessary for the "ranks" to be integers, but it keeps the process easily understandable.

3.2. TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution). TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) is a multicriteria decision-making approach created by [56]. It is founded on the principle that the best option should be the one with the least geometric distance from the positive ideal solution and the greatest geometric distance from the negative ideal solution. The steps for this method are listed as follows:

Step 1: create an evaluation matrix with m alternatives, with x_{ij} representing the score for the ith candidate in the jth criteria in an $m \times n$ matrix of the form given below as decision matrix D.

$$D = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix}. \tag{2}$$

Step 2: calculate the normalised matrix using the following formula:

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}}. \tag{3}$$

In the previous formula, $i = 1, 2, \ldots, m$ and $j = 1, 2 \ldots, n$.

Step 3: construct the weighted normalised matrix v_{ij}. This is calculated by multiplying the weight w_j we had calculated before by each column of matrix r_{ij}. The formula is

$$v_{ij} = w_j r_{ij}. \tag{4}$$

Step 4: identify the Positive Ideal Solution (PIS) (A^*) and Negative Ideal Solution (NIS) (A^-), where

$$\begin{aligned} A^* &= v_1^*, v_2^*, \ldots, v_n^*, \\ A^- &= v_1^-, v_2^-, \ldots, v_n^-. \end{aligned} \tag{5}$$

We define

$$\begin{aligned} v_j^* &= \begin{cases} \max\left(v_{ij}\right), & \text{if } j\epsilon J_1 \\ \min\left(v_{ij}\right), & \text{if } j\epsilon J_2 \end{cases}, \\ v_j^- &= \begin{cases} \min\left(v_{ij}\right), & \text{if } j\epsilon J_1 \\ \max\left(v_{ij}\right), & \text{if } j\epsilon J_2 \end{cases}. \end{aligned} \tag{6}$$

In the previous formula, J_1 and J_2 are attributes related to criteria that have a positive impact and attributes related to criteria that have a negative impact, respectively.

Steps 5 and 6: the n-dimensional Euclidean distance can be used to calculate the separation or distance between the alternatives. The separations from the PIS A^* and NIS A^- are S^* and S^-, respectively.

$$S_i^* = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^*\right)^2},$$
$$S_i^- = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^-\right)^2}. \quad (7)$$

In the previous formula, $i = 1, 2 \ldots, m$ and $j = 1, 2 \ldots, n$. After this, in Step 3.2.5, we calculate the similarity using

$$C_i^* = \frac{S_i^-}{S_i^* + S_i^-}. \quad (8)$$

The terms have usual meaning mentioned previously.

Step 7: we choose the candidate with the maximum C_i^* or rank all the alternatives in a descending order according to their C_i^* values. This process is done for both the S_i^* and S_i^- values. The flowchart of these steps is given in Figure 1.

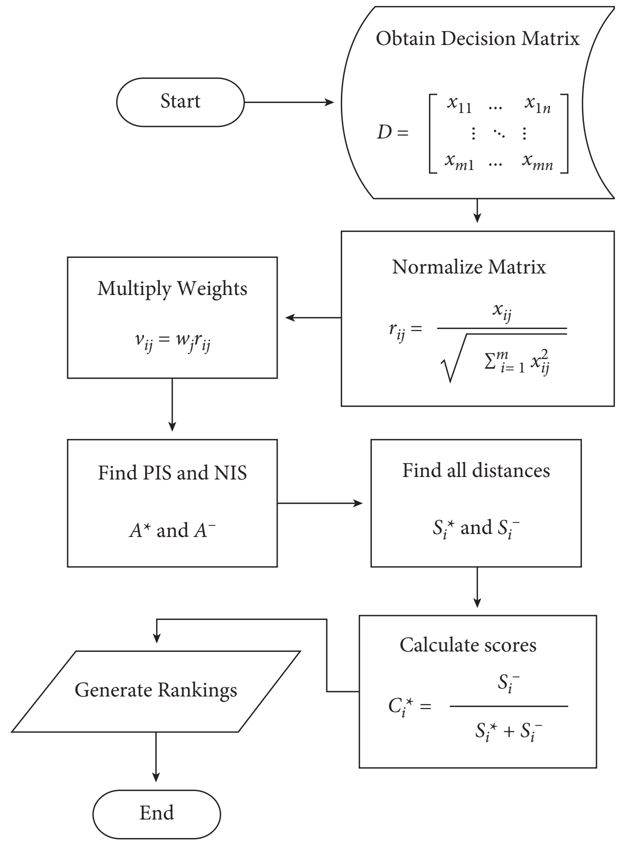


Figure 1: Flowchart of TOPSIS.

3.3. NR-TOPSIS (Improved R-TOPSIS). This updated version of TOPSIS method is abbreviated as NR-TOPSIS method because it can (supposedly) handle rank reversal problems based on the historical maximum value of indicator data. In article [52], the author claimed and proved the method to be effective on the dataset, which had 4–7 alternatives and removed the rank reversal problem in original TOPSIS method. The steps of the algorithm are as follows:

Step 1: find the minimum value m_j and maximum value M_j of each attribute C_j. So, for any attribute x_{ij}, the condition $m_j \le x_{ij} \le M_j$ is satisfied. Also, the condition $m_j \le x_{ij} \le M_j$ must be satisfied when the scheme is increased, decreased, or replaced.

Step 2: the original decision-making matrix $X = (x_{ij})_{m \times n}$ is standardized and transformed to generate standardized decision-making matrix $Y = (y_{ij})_{m \times n}$, where y_{ij} are normalised attribute values. This will eliminate the influence of dimension on data decision-making.

If C_j is benefit attribute, then

$$y_{ij} = \frac{x_{ij} - m_j}{M_j - m_j}. \quad (9)$$

If C_j is cost attribute, then

$$y_{ij} = \frac{M_j - x_{ij}}{M_j - m_j}. \quad (10)$$

Step 3: calculate the weighted normalised decision-making matrix $= (v_{ij})_{m \times n}$. The weighted normalised attribute value v_{ij} has the same calculation as that of TOPSIS, $v_{ij} = w_j y_{ij}$, $i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$.

Step 4: identify the Positive Ideal Solution PIS (A^*) and Negative Ideal Solution NIS (A^-), where

$$A^* = v_1^*, v_2^*, \ldots, v_n^*,$$
$$A^- = v_1^-, v_2^-, \ldots, v_n^- \quad (11)$$

and

$$v_j^* = w_j,$$
$$v_j^- = 0. \quad (12)$$

$j = 1, 2, \ldots, n$.

Step 5: compute the Euclidean distances S_i^* and S_i^- for every alternative A_i between the positive ideal solution and the negative ideal solution, respectively.

$$S_i^* = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^*\right)^2},$$
$$S_i^- = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^-\right)^{22}}. \quad (13)$$

In the previous formula, $i = 1, 2 \ldots, m$ and $j = 1, 2 \ldots, n$. After this, we calculate the similarity using

$$C_i^* = \frac{S_i^-}{S_i^* + S_i^-}. \quad (14)$$

The terms have usual meaning mentioned previously. The logic of the two instances in which the indicator is a benefit type or a cost type is examined as follows:

(1) Two extreme cases are considered when C_j is a benefit attribute: $x_{ij} = m_j$ and $x_{ij} = M_j$, which means if the value of an attribute is the same as the maximum or minimum value in that column. When $x_{ij} = m_j$ consider $y_{ij} = m_j - m_j/M_j - m_j = 0$ and $v_j^- = v_{ij} = 0$, and when $x_{ij} = M_j$ consider $y_{ij} = M_j - m_j/M_j - m_j = 1$, where $v_j^+ = v_{ij} = w_j$.

(2) In case of cost attribute C_j, again two extreme cases can occur. When $x_{ij} = m_j$ consider $y_{ij} = M_j - m_j/M_j - m_j = 1$, where $v_j^+ = v_{ij} = w_j$, and when $x_{ij} = M_j$ consider $y_{ij} = M_j - M_j/M_j - m_j = 0$ and $v_j^- = v_{ij} = 0$.

3.4. TODIM (TOmada de DecisãoInterativa e Multicritério, an Acronym in Portuguese for Interactive Multicriteria Decision-Making). The TODIM technique we studied from [9] is based on PT (prospect theory), as its value gain/loss function is defined in [29]. Gains and losses are always calculated in relation to a reference point in this situation. As a result, while this technique acknowledges the possibility of decision-makers, it does not incorporate their actual involvement. The following are the stages that a TODIM application would take in algorithmic form:

Step 1: we will reuse the initial evaluation matrix considered in the previous section. We directly normalize the ratings and weights using the following formula:

$$P_{ij} = \begin{cases} \dfrac{x_{ij}}{\sum_{k=1}^{m} x_{kj}} \text{if } j \in J_1, \\ \\ \dfrac{1/x_{ij}}{\sum_{k=1}^{m} 1/x_{kj}} \text{if } j \in J_2. \end{cases} \quad (15)$$

For the weighting factor or trade-off rate between the reference criteria r and the generic criteria c, here, w_r determines the most relevant reference criterion for the decision-maker. Often, it is the maximum weight. In general, any criterion can be used as the reference criterion, and this decision has no effect on the final findings. So, the formula we have is

$$w_{rc} = \frac{w_c}{w_r}, \quad (16)$$

where $w_r = \max\{w_c | c = 1, 2, \ldots, n\}$ and $i = 1, 2 \ldots, m$; $j = 1, 2 \ldots, n$.

Step 2: for calculating the dominance degree, we need to first check the contribution of each criterion using a formula, where φ_c is the contribution of criterion c to the function $\delta(A_i, A_j)$ and θ is the loss of attenuation factor whose value we considered as 1. In our case,

$$\varphi_c(A_i, A_j) = \begin{cases} \sqrt{\dfrac{(P_{ic} - P_{jc}) w_{rc}}{\sum_{c=1}^{n} w_{rc}}}, \text{if } P_{ic} - P_{jc} \geq 0 \\ \\ -\dfrac{1}{\theta} \sqrt{\dfrac{(\sum_{c=1}^{n} w_{rc})(P_{jc} - P_{ic})}{w_{rc}}}, \text{if } P_{ic} - P_{jc} < 0 \end{cases}, \quad (17)$$

and combining all contributions, we get the dominance degrees from the measurement of dominance $\delta(A_i, A_j)$ as

$$\delta(A_i, A_j) = \sum_{c=1}^{n} \varphi_c(A_i, A_j), \quad (18)$$

where $i, j = 1, 2 \ldots, m$; $c = 1, 2 \ldots, n$.

Step 3: finally, compute the values of ξ_i, which are the normalised global performances of alternatives compared to others, such that the largest value is picked as more significant than the other alternative's values:

$$\xi_i = \frac{\sum_{j=1}^{n} (A_i, A_j)\delta - \min_i \sum_{j=1}^{n} \delta(A_i, A_j)}{\max_i \sum_{j=1}^{n} (A_i, A_j) - \min_i \sum_{j=1}^{n} \delta(A_i, A_j)}. \quad (19)$$

In the previous formula, $i = 1, 2, \ldots, n$. The flowchart of these steps is in Figure 2.

4. Case Study

We used the IPL 2019 dataset and rated 15–16 randomly picked cricketers. In this section, we will first calculate the weights for the 12 and 9 criteria selected to rank batsman and bowlers accordingly. Then we will implement them in TOPSIS, NR-TOPSIS, and TODIM to check the rank generated by these algorithms, respectively. We have highlighted the important calculations only and thus the intermediate steps can be found in the Supplementary Materials provided with this paper.

4.1. Criteria Selection. We used the rank-sum weight approach to determine the weights of the criteria specified for batsmen and bowlers, making it easier for decision-makers to grasp and rank the criteria based on their preferences. The criteria for batsmen are described in Table 1, and those for bowlers are stated in Table 2. It should be noted that the criteria rankings are adjusted for the T20 format and are subject to change based on the decision-makers' preferences. The batsmen and bowlers chosen here are randomly selected with no regional preferences from the list given in the two following links: https://www.iplt20.com/stats/2019/most-runs and https://www.iplt20.com/stats/2019/most-wickets. As we can see, our data is collected from a trustworthy and

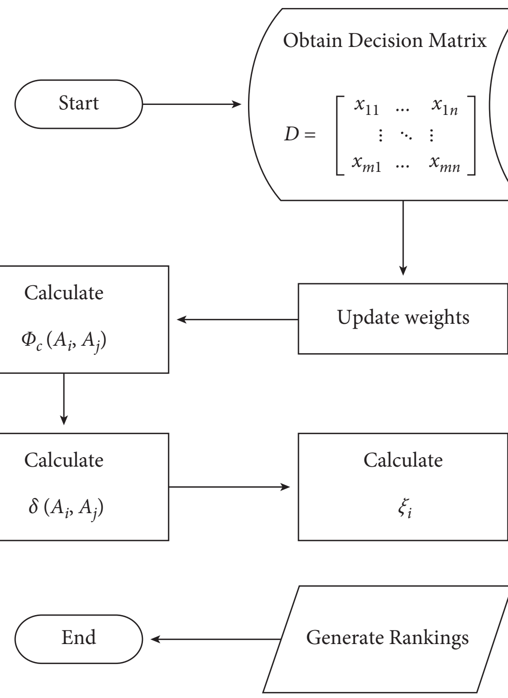


Figure 2: Flowchart of TODIM.

Table 1: Batting performance criteria.

Batting criteria	Symbol	Definition	Rank	Weights	Ideally
Strike rate	SR	SR = runs/BF × 100	1	0.1538462	Higher
Batting average	Avg	Avg = runs/outs	2	0.1410256	Higher
Runs	Runs	The total runs scored in the season	3	0.1282051	Higher
Innings	Inn	The number of innings played in the season	4	0.1153846	Higher
Not outs	NO	The number of innings undefeated	5	0.1025641	Higher
6s	6s	The number of sixes hit	6	0.0897436	Higher
4s	4s	The number of fours bit	7	0.0769231	Higher
100s	100s	The number of centuries scored	8	0.0641026	Higher
50s	50s	The number of half-centuries scored	9	0.0512821	Higher
Matches	Mat	The number of matches played, might be more than the number of innings played	10	0.0384615	Higher
Highest score	HS	The highest individual score	11	0.025641	Higher
Balls faced	BF	The number of balls faced in the tournament	12	0.0128205	Higher

Table 2: Bowling performance criteria.

Bowling criteria	Symbol	Definition	Rank	Weights	Ideally
Strike rate	SR	SR = BB/wkts	1	0.200000	Lower
Economy	Econ	Econ = runs/overs = runs/BB × 6	2	0.177778	Lower
Bowling average	Avg	Avg = runs/wkts	3	0.155556	Lower
Wickets	Wkts	The number of wickets taken in the season	4	0.133333	Higher
Runs	Runs	Total runs conceded	5	0.111111	Lower
Innings	Inns	The number of innings played	6	0.088889	Higher
Balls bowled	BB	Total number of balls bowled	7	0.066667	Higher
4 wicket hauls	4w	Frequency of times 4 wickets were taken	8	0.044444	Higher
Matches	Mat	The number of matched played	9	0.022222	Higher

easily available to public source; thus we do not need to implement a fuzzy algorithm here.

The ranking of criteria is subjective, and the decision-maker is expected to arrange them according to their preferences. Keeping in mind the T20 format, we assume that strike rate is more important than batting average for batsmen courtesy of the limited playtime. Similarly, a good batsman may have scored fewer runs overall but may have helped finish more run chases successfully. In case of bowlers, both strike rate and economy are important. However, bowling out all batsmen ends the game quicker, so strike rate takes priority. This form of reasoning applies to ranking the remaining criteria as well. The rankings we present here are not authoritative, but they are certainly close to what would be used in practice. We illustrate an example that differs from the current context: in One-Day Internationals (ODIs), strike rate may have a lower preference than batting average and total runs scored for batsmen. Similarly, the number of balls faced becomes more important in a test match. Therefore, the decision-makers need to rank the criteria based on the format of the game being played.

4.2. Application of TOPSIS. After computing the weights of the criteria, we will rank the alternatives, which are batsmen and bowlers, using TOPSIS whose algorithm is described in the previous section. The evaluation matrix is the same for all algorithms and we will be dealing with the crisp dataset.

Step 1: the evaluation matrix for Foreign Batsmen of IPL 2019 (retrieved from the official website mentioned at the beginning of the section) is represented in Table 3.

The same for bowlers is presented in Table 4 and all the data are taken form the website mentioned at the beginning of the section.

Step 2: the normalised matrix is calculated.

Step 3: the weighted normalised matrix is calculated.

Step 4: the Positive Ideal Solution (PIS) (A^*) and Negative Ideal Solution (NIS) (A^-) are identified.

Step 5: S_i^*, S_i^-, and C_{i^*} are calculated for all players (batsmen and bowlers).

Step 6: S_i^*, S_i^-, and C_{i^*} are calculated for all players (batsmen and bowlers).

Then, we proceed to rank the candidates according to the distance values. These values for the batsmen and bowlers are provided in Tables 5 and 6.

Step 7: we choose the candidate with the maximum C_i^* or rank all the alternatives in a descending order according to their C_i^* values. This process is done for both the S_i^* and S_i^- values.

4.3. Application of NR-TOPSIS. We will utilize the same evaluation matrix and weights for the criteria as calculated previously. Here, M_j and m_j are the maximum and minimum values selected from each column of Tables 3 and 4, respectively:

Step 1: the evaluation matrix for Foreign Batsmen of IPL 2019 (obtained from the official website mentioned at the beginning of the section) is represented in Table 3 and that for bowlers is presented in Table 4 and all the data are taken form the website mentioned at the beginning of the section.

Step 2: we calculate the normalised matrix.

Step 3: the weighted normalised matrix is then calculated using a formula. These results are all available in the Supplementary Materials.

Step 4: the Positive Ideal Solution (PIS) (A^*) and Negative Ideal Solution (NIS) (A^-) are identified.

Step 5: S_i^*, S_i^-, and C_{i^*} are calculated for all players (batsmen and bowlers).

Step 6: S_i^*, S_i^-, and C_{i^*} are calculated for all players (batsmen and bowlers).

Then we proceed to rank the candidates according to the distance values. These values for the batsmen and bowlers are provided in Tables 7 and 8.

4.4. Application of TODIM. In this part, we will be evaluating the rank generated using TODIM. We have considered the value of attenuation factor θ as 1. However, decision-makers can change the value accordingly. Changing this value might affect the final ranking and is therefore discussed in the Sensitivity Analysis section.

Step 1: we can determine the trade-off rate for batsmen and bowlers using the list of weights we calculated before. They are presented in Table 9.

Step 2: we calculate the matrix $\varphi_c(A_i, A_j)$ and then the matrix $\delta(A_i, A_j)$ for $\theta = 1$.

Step 3: then, we compute the values of ξ_i. The values of ξ_i are calculated in Table 10.

4.5. Results. The rankings generated by TOPSIS for batsmen and bowlers, who are international players in the IPL 2019, are tabulated in a descending order of preference and summarised in Table 11. Similarly, the ranks generated by NR-TOPSIS and TODIM are summarised in Tables 12 and 13, respectively.

From the tables, we see that David Warner is ranked as the best batsman (according to our criteria) considering the criteria strike rate, batting average, runs, and so forth in TOPSIS and TODIM. Meanwhile, in case of NR-TOPSIS, Andre Russell is the best batsman. For bowlers, Kagiso Rabada and Imran Tahir are the top two choices for all the three approaches based on criteria like strike rate, bowling average, and wickets. However, their relative ranks are swapped in case of TODIM. For the lowest performers (out of the chosen candidates), we find that Kane Williamson and Ben Stokes appear near the end of the list for batsmen, and Trent Boult is the analogue for bowlers. These rankings are relative to the number of candidates chosen. The rankings of the players in the middle of the lists vary slightly. We elaborate on the application of these findings in the Conclusion.

The rankings for both batsmen and bowlers are similar enough in the middle of the list as is evident from our Spearman Rank Correlation calculations. It is generally the top ranked players that shuffle their positions greatly across the methods. We tried to minimize the effect of rank reversal using NR-TOPSIS over TOPSIS but failed to see any improvement at all. TODIM performed the best in this regard out of the three methods by producing the least number of rank reversals during the sensitivity analysis.

Table 3: Batsmen statistics.

Name	Nationality	Mat	Inn	NO	Runs	HS	Avg	BF	SR	100s	50s	4s	6s
AB de Villiers	South African	13	13	3	442	82	44.20	287	154.00	0	5	31	26
Andre Russell	West Indian	14	13	4	510	80	56.67	249	204.81	0	4	31	52
Ben Stokes	English	9	9	3	123	46	20.50	99	124.24	0	0	8	4
Chris Gayle	West Indian	13	13	1	490	99	40.83	319	153.60	0	4	45	34
Chris Lynn	Australian	13	13	0	405	82	31.15	290	139.65	0	4	41	22
David Warner	South African	12	12	2	692	100	69.20	481	143.86	1	8	57	21
Faf du Plessis	South African	12	12	1	396	96	36.00	321	123.36	0	3	36	15
Jonny Bairstow	English	10	10	2	445	114	55.63	283	157.24	1	2	48	18
Jos Buttler	English	8	8	0	311	89	38.88	205	151.70	0	3	38	14
Kane Williamson	New Zealander	9	9	2	156	70	22.29	130	120.00	0	1	12	5
Kieron Pollard	West Indian	16	14	6	279	83	34.88	178	156.74	0	1	14	22
Marcus Stoinis	Australian	10	10	6	211	46	52.75	156	135.25	0	0	14	10
Moeen Ali	English	11	10	2	220	66	27.50	133	165.41	0	2	16	17
Quinton de Kock	South African	16	16	1	529	81	35.27	398	132.91	0	4	45	25
Shane Watson	Australian	17	17	0	398	96	23.41	312	127.56	0	3	42	20
Steve Smith	Australian	12	10	2	319	73	39.88	275	116.00	0	3	30	4

Table 4: Bowlers statistics.

Name	Nationality	Mat	Inn	Ov	BB	Runs	Wkts	Avg	Econ	SR	4w
Andre Russell	West Indian	14	12	30.1	181	287	11	26.09	9.51	16.45	0
Ben Stokes	English	9	6	16.5	101	189	6	31.5	11.23	16.83	0
Chris Morris	South African	9	9	33.0	198	306	13	23.54	9.27	15.23	0
Dwayne Bravo	West Indian	12	12	41.1	247	330	11	30	8.02	22.45	0
Imran Tahir	South African	17	17	64.2	386	431	26	16.58	6.7	14.85	2
Jofra Archer	English	11	11	43.0	258	291	11	26.45	6.77	23.45	0
Kagiso Rabada	South African	12	12	47.0	282	368	25	14.72	7.83	11.28	2
Keemo Paul	West Indian	8	8	27.1	163	237	9	26.33	8.72	18.11	0
Lasith Malinga	Sri Lankan	12	12	44.5	269	438	16	27.38	9.77	16.81	2
Moeen Ali	English	11	9	25.0	150	169	6	28.17	6.76	25	0
Mohammad Nabi	Afghan	8	8	29.1	175	194	8	24.25	6.65	21.88	1
Rashid Khan	Afghan	15	15	60.0	360	377	17	22.18	6.28	21.18	0
Sam Curran	English	9	9	33.0	198	323	10	32.3	9.79	19.8	1
Sunil Narine	West Indian	12	12	44.2	266	347	10	34.7	7.83	26.6	0
Trent Boult	New Zealander	5	5	19.0	114	163	5	32.6	8.58	22.8	0

Table 5: Distance values for batsmen.

Batsmen	$\mathbf{S_i^*}$	$\mathbf{S_i^-}$	$\mathbf{C_i^*}$
AB de Villiers	0.070196	0.055112	0.439813
Andre Russell	0.056888	0.081165	0.587927
Ben Stokes	0.106526	0.027294	0.203959
Chris Gayle	0.076169	0.055195	0.420169
Chris Lynn	0.090296	0.040840	0.311430
David Warner	0.051679	0.090778	0.637231
Faf du Plessis	0.088862	0.036606	0.291756
Jonny Bairstow	0.062814	0.069648	0.525796
Jos Buttler	0.095810	0.032634	0.254072
Kane Williamson	0.105748	0.019120	0.153120
Kieron Pollard	0.079607	0.062885	0.441323
Marcus Stoinis	0.087082	0.061924	0.415581
Moeen Ali	0.093227	0.029365	0.239533
Quinton de Kock	0.080278	0.053022	0.397761
Shane Watson	0.094307	0.041946	0.307855
Steve Smith	0.092551	0.033953	0.268397

4.6. Spearman Rank Correlation Coefficient Test. The Spearman Rank Correlation Coefficient [57] is a nonparametric measure of the consistency and control of the relationship between two ranking variables. There are several research articles where authors have evaluated various ranking methods, such as ELECTRE and TOPSIS [58],

Table 6: Distance values for bowlers.

Bowlers	S_i^*	S_i^-	C_i^*
Andre Russell	0.05682	0.04047	0.41596
Ben Stokes	0.07509	0.03480	0.31667
Chris Morris	0.05326	0.04399	0.45232
Dwayne Bravo	0.06233	0.03381	0.35169
Imran Tahir	0.02677	0.08195	0.75376
Jofra Archer	0.06068	0.03907	0.39169
Kagiso Rabada	0.02479	0.07958	0.76251
Keemo Paul	0.06216	0.03659	0.37049
Lasith Malinga	0.04895	0.05030	0.50679
Moeen Ali	0.07308	0.03787	0.34136
Mohammad Nabi	0.06135	0.04246	0.40905
Rashid Khan	0.04766	0.05515	0.53646
Sam Curran	0.06326	0.03036	0.32433
Sunil Narine	0.07204	0.03081	0.29958
Trent Boult	0.07851	0.03114	0.28400

Table 7: Distance values for batsmen.

Batsmen	S_i^*	S_i^-	C_i^*
AB de Villiers	0.174382	0.160163	0.478748
Andre Russell	0.116700	0.248776	0.680691
Ben Stokes	0.302746	0.054920	0.153552
Chris Gayle	0.182300	0.164425	0.474222
Chris Lynn	0.226352	0.126221	0.357999
David Warner	0.154195	0.238726	0.607567
Faf du Plessis	0.237810	0.110955	0.318137
Jonny Bairstow	0.175557	0.179480	0.505525
Jos Buttler	0.245161	0.108004	0.305817
Kane Williamson	0.300392	0.040603	0.119072
Kieron Pollard	0.202568	0.164377	0.447960
Marcus Stoinis	0.236055	0.147292	0.384226
Moeen Ali	0.241691	0.105719	0.304306
Quinton de Kock	0.204539	0.169953	0.453822
Shane Watson	0.244431	0.153702	0.386057
Steve Smith	0.257682	0.094443	0.268208

Table 8: Distance values for bowlers.

Bowlers	S_i^*	S_i^-	C_i^*
Andre Russell	0.207621	0.185358	0.471674
Ben Stokes	0.288852	0.164804	0.363279
Chris Morris	0.191566	0.203492	0.515094
Dwayne Bravo	0.239656	0.157935	0.39723
Imran Tahir	0.119727	0.320159	0.727823
Jofra Archer	0.224961	0.199751	0.470322
Kagiso Rabada	0.109768	0.320403	0.744827
Keemo Paul	0.213923	0.180645	0.457831
Lasith Malinga	0.221858	0.183265	0.452369
Moeen Ali	0.261022	0.2044	0.439172
Mohammad Nabi	0.213779	0.221073	0.508387
Rashid Khan	0.181835	0.24905	0.577996
Sam Curran	0.260104	0.126541	0.327279
Sunil Narine	0.295574	0.147027	0.332188
Trent Boult	0.284633	0.155375	0.353119
Andre Russell	0.207621	0.185358	0.471674

Table 9: Table of updated weights.

Batting criteria	w_{rc}	Bowling criteria	w_{rc}
SR	1	**SR**	1
Avg	0.916667	**Econ**	0.888889
Runs	0.833333	**Avg**	0.777778
Inn	0.75	**Wkts**	0.666667
NO	0.666667	**Runs**	0.555556
6s	0.583333	**Inns**	0.444444
4s	0.5	**BB**	0.333333
100s	0.416667	**4w**	0.222222
50s	0.333333	**Mat**	0.111111
Mat	0.25		
HS	0.166667		

Table 10: ξ_i for batsmen and bowlers.

Batsmen	ξ_i	Bowlers	ξ_i
AB de Villiers	0.804597	**Andre Russell**	0.561531
Andre Russell	0.818468	**Ben Stokes**	0.155654
Ben Stokes	0	**Chris Morris**	0.426346
Chris Gayle	0.848149	**Dwayne Bravo**	0.524013
Chris Lynn	0.687784	**Imran Tahir**	1
David Warner	1	**Jofra Archer**	0.525132
Faf du Plessis	0.667418	**Kagiso Rabada**	0.91773
Jonny Bairstow	0.777412	**Keemo Paul**	0.279925
Jos Buttler	0.427361	**Lasith Malinga**	0.738363
Kane Williamson	0.130708	**Moeen Ali**	0.374627
Kieron Pollard	0.552321	**Mohammad Nabi**	0.428831
Marcus Stoinis	0.247835	**Rashid Khan**	0.719623
Moeen Ali	0.320078	**Sam Curran**	0.398341
Quinton de Kock	0.823886	**Sunil Narine**	0.472208
Shane Watson	0.703334	**Trent Boult**	0
Steve Smith	0.508533		

Table 11: Ranks generated by TOPSIS.

Rank	Batsmen	Bowlers
1	David Warner	Kagiso Rabada
2	Andre Russell	Imran Tahir
3	Jonny Bairstow	Rashid Khan
4	Kieron Pollard	Lasith Malinga
5	AB de Villiers	Chris Morris
6	Chris Gayle	Andre Russell
7	Marcus Stoinis	Mohammad Nabi
8	Quinton de Kock	Jofra Archer
9	Chris Lynn	Keemo Paul
10	Shane Watson	Dwayne Bravo
11	Faf du Plessis	Moeen Ali
12	Steve Smith	Sam Curran
13	Jos Buttler	Ben Stokes
14	Moeen Ali	Sunil Narine
15	Ben Stokes	Trent Boult
16	Kane Williamson	

Table 12: Ranks generated by NR-TOPSIS.

Rank	Batsmen	Bowlers
1	Andre Russell	Kagiso Rabada
2	David Warner	Imran Tahir
3	Jonny Bairstow	Rashid Khan
4	Ab de Villiers	Chris Morris
5	Chris Gayle	Mohammad Nabi
6	Quinton de Kock	Andre Russell
7	Kieron Pollard	Jofra Archer
8	Shane Watson	Keemo Paul
9	Marcus Stoinis	Lasith Malinga
10	Chris Lynn	Moeen Ali
11	Faf du Plessis	Dwayne Bravo
12	Jos Buttler	Ben Stokes
13	Moeen Ali	Trent Boult
14	Steve Smith	Sunil Narine
15	Ben Stokes	Sam Curran
16	Kane Williamson	

Table 13: Ranks generated by TODIM.

Rank	Batsmen	Bowlers
1	David Warner	Imran Tahir
2	Chris Gayle	Kagiso Rabada
3	Quinton de Kock	Lasith Malinga
4	Andre Russell	Rashid Khan
5	AB de Villiers	Andre Russell
6	Jonny Bairstow	Jofra Archer
7	Shane Watson	Dwayne Bravo
8	Chris Lynn	Sunil Narine
9	Faf du Plessis	Mohammad Nabi
10	Kieron Pollard	Chris Morris
11	Steve Smith	Sam Curran
12	Jos Buttler	Moeen Ali
13	Moeen Ali	Keemo Paul
14	Marcus Stoinis	Ben Stokes
15	Kane Williamson	Trent Boult
16	Ben Stokes	

TOPSIS and modified TOPSIS [59], and MOTV algorithms [60]. In our article, we have implemented the approach to study the correlation between TOPSIS and NR-TOPSIS, TOPSIS and TODIM, and TODIM and NR-TOPSIS. The formula used is

$$\rho = 1 - \frac{6\sum_{i=1}^{m} d_i^2}{n\left(n^2 - 1\right)}, -1 \leq \rho \leq 1, \tag{20}$$

where ρ is the spearman rank correlation coefficient, d_i is the difference between the two ranks of each method we used, and n is the number of alternatives or players. The values of the correlation can vary from −1 to +1 which are further categorized as follows:

(i) If $\rho = 1$, then correlation is absolute
(ii) If $0.8 \leq \rho 0.999$, then correlation is very strong
(iii) If $0.6 \leq \rho \leq 0.79$, then correlation is strong
(iv) If $0.4 \leq \rho \leq 0.59$, then correlation is moderate
(v) If $0.2 \leq \rho \leq 0.39$, then correlation is weak
(vi) If $0.19 \geq \rho$, then correlation is very weak

Therefore, the values of ρ we obtained from results in Tables 11–13 are given in Table 14, and as we can see TOPSIS and TODIM are strongly positively correlated, whereas TOPSIS and NR-TOPSIS or TODIM and NR-TOPSIS are very strongly positively correlated.

4.7. Rank Reversal. Rank reversal is a phenomenon that occurs when a decision-maker is choosing an alternative from a group of options, and they are presented with additional options that were not considered when the selection process began. We proved rank reversal by deleting one and two alternatives from the list of batsmen in Table 3 and then performing the ranking. Among the alternatives, we have seen substantial rank reversals. When we compared the three approaches in this study, we discovered that TODIM had the least amount of variation in the ranking. TOPSIS and NR-TOPSIS exhibit a substantial number of reversals. Table 15 demonstrates the comparison when one alternative is

Table 14: Spearman Rank Correlation Coefficient results of TOPSIS, TODIM, and NR-TOPSIS.

TOPSIS and TODIM, TOPSIS and NR-TOPSIS, and TODIM and NR-TOPSIS
ρ for batsmen: 0.7676, 0.9529, and 0.8647
ρ for bowlers: 0.8179, 0.9143, and 0.6464

removed and Table 16 demonstrates the comparison when two of the alternatives are removed.

In case of NR-TOPSIS, when we removed the batsman who was ranked 2 (Andre Russell) in Table 3, we saw a change in ranking of alternatives present in ranks 3 (Ben Stokes), 4 (Chris Gayle), 5 (Chris Gayle), and 6 (David Warner). They are not moving positions by one spot owing to the elimination of one player; rather they are switching positions with one another. NR-TOPSIS was not meant to operate in this way, as it was not supposed to have a rank reversal problem. In case of TOPSIS, however, when we again removed the batsman present in position 2 (Andre Russell), others in positions 5 (Chris Lynn), 6 (David Warner), 7 (Faf du Plessis), 8 (Jonny Bairstow), 9 (Jos Buttler), 10 (Kane Williamson), and 12 (Marcus Stoinis) changed their respective ranks as well. Similar cases were observed when we removed two alternatives. Finally, we noticed that, in TODIM, only two players' rankings are switched. When one alternative (3, 9, 10, or 12) or two alternatives ((3, 1), (6, 5), (7, 6), or (9, 1)) are removed, the 4th (Chris Gayle) and 5th (Chris Gayle) players, or 2nd (Andre Russell) and 3rd (Ben Stokes), exchanged positions.

We added an alternative mentioned in Table 17 to the existing list in Table 3 to observe the problem of rank reversal in all the three methods. Because the new alternative was ranked last in all the three methods, we removed one player from the original list in Table 3 and checked how the rank reversal occurs. With TODIM, we discovered an exceptional phenomenon. Only when the player from the 6th position is removed is there an exchange of position between 2nd and 3rd. But, for all other players, when removed, we see no occurrence of rank reversal issue. For the other two methods, namely, TOPSIS and NR-TOPSIS, we see usual occurrence of rank reversal as we have observed previously.

5. Sensitivity Analysis

The impact and stability of the weights (for both batting and bowling criteria) produced by rank-sum weight method are revealed in the sensitivity analysis for TOPSIS and TODIM. We also tweaked the TODIM attenuation factor to see how sensitive the system is for the purpose of selecting international players. For our analysis, we altered the weights for each criterion by shuffling the rank of the criteria. Because decision-makers can select how the criteria are ranked, we looked at how weight changes can affect TOPSIS and TODIM rankings. The sensitivity of the approaches is displayed using the heat map representation method. The ranks that are seen more frequently in this scheme have a darker colour than the ranks that are seen less often, which have a lighter colour. In this paper, the darkest hue is black, and the lightest colour is white.

5.1. Criteria Shuffling or Weight Shuffling for TOPSIS. As there are twelve criteria for batsmen and nine criteria for bowlers, we end up with 12! permutations for batting performance criteria and 9! for bowling criteria. In order to see how changing the ranking affects the ranking of alternatives, we shuffle the rankings of the criteria (and hence their weights). We uniformly selected 1000 random shuffles for the weights using a Monte Carlo approach. The ranking of international players using TOPSIS is represented using heat map style point graphs in Figures 3 and 4.

The darker (black) dots indicate those ranks which are more frequent in the 1000 samples chosen. Grey dots are documented, though they appear less frequently. The ranks that all the players have had for various combinations of rank of the criterion or various values of the weight may be seen on this point graph. For David Warner (batsman), for example, in Figure 3, we may observe two points. The colour of one point differs from that of the other one. The darker dot represents the rank that appeared more frequently than the other. As a result, regardless of changes in weights or criteria, we can show that he has consistently ranked first in most cases. We may also say the same thing about Imran Tahir (bowler) in Figure 4. Kieron Pollard (batsman) has had a rank of 2–13, with 4 being the most common because it is the darkest of all the dots and 2 being the least often because it is the lightest. Even for Mohammad Nabi (bowler) his rank varies between 4 and 12, with the darkest point at rank 5. Players like Lasith Malinga (bowler) and Rashid Khan (bowler) or Andre Russell (batsman) and Ben Stokes (batsman) have a ranking plot with two equal dark spots, indicating that both rankings occur frequently.

5.2. Criteria Shuffling or Weight Shuffling for NR-TOPSIS. In the instance of NR-TOPSIS, we used the same definition as in TOPSIS, but the outcome is different. Figure 5 represents the sensitivity analysis of NR-TOPSIS for batsmen when the weights are shuffled. With Ben Stokes, Kane Williamson, and David Warner, we observe somewhat consistent results, with their rankings not deviating significantly with the change in weight. In case of TOPSIS, we saw that these players had more grey or lighter coloured dots surrounding the black ones, indicating that they had those ranks for a specific weight value. Other rankings of batsmen like Kieron Pollard, Jonny Bairstow, Shane Watson, and Marcus Stoinis are highly unstable, similar to TOPSIS. There is not a lot of consistency to be seen.

In Figure 6, we see the sensitivity analysis of bowlers. The stability is quite feeble overall. Only Kagiso Rabada, Imran Tahir, and Rashid Khan have a reasonable level of consistency. When the weights are shuffled or altered, all the other bowlers exhibit a considerable deal of inconstancy in their rankings. Even though the dark-coloured dots most of the time reflect the incidence of that rank when the weights are shuffled in the player's direction, the existence of

Table 15: Comparison of rank reversal results of NR-TOPSIS, TOPSIS, and TODIM when one alternative is removed from the original rank list of batsmen.

Removed alternatives	Rank reversals for NR-TOPSIS	Removed alternatives	Rank reversals for TOPSIS	Removed alternatives	Rank reversals for TODIM
2	3, 4, 5, 6	2	5, 6, 7, 8, 9, 10, 12	3	4, 5
3	6, 7, 13, 14	3	12, 13	9	4, 5
4	4, 5	6	7, 8, 12, 13	10	4, 5
5	4, 5	15	11, 12	12	4, 5

Table 16: Comparison of rank reversal results of NR-TOPSIS, TOPSIS, and TODIM when two alternatives are removed from the original rank list of batsmen.

Removed alternatives	Rank reversals for NR-TOPSIS	Removed alternatives	Rank reversals for TOPSIS	Removed alternatives	Rank reversals for TODIM
6, 1	6, 7, 11, 12	3, 1	5, 6, 12, 13	3, 1	4, 5
7, 2	5, 6, 7, 8, 9, 10, 11	6, 4	1, 2, 3, 4, 5, 6	6, 5	2, 3
5, 3	11, 12	10, 3	6, 7, 9, 10, 13, 14	7, 6	2, 3
10, 2	5, 6, 7, 8, 9, 10, 12	14, 2	6, 7, 8, 9, 11	9, 1	4, 5

Table 17: New batsman to be added.

Name	Nationality	Mat	Inn	NO	Runs	HS	Avg	BF	SR	100s	50s	4s	6s
Shimron Hetmyer	Guyanese	5	5	0	90	75	18.00	73	123.28	0	1	4	7

Table 18: Comparison of rank reversal results of NR-TOPSIS, TOPSIS, and TODIM when one alternative is removed from the new list of batsmen.

Added alternative	Rank reversals for NR-TOPSIS	Added alternative	Rank reversals for TOPSIS	Added alternative	Rank reversals for TODIM
2	5, 6, 7, 8, 9, 10, 12	6	1, 2, 3, 4	3	No change
3	12, 13	7	7, 8	9	No change
6	7, 8, 12, 13	9	7, 8	6	2, 3
15	11, 12	11	5, 6, 12, 13	14	No change

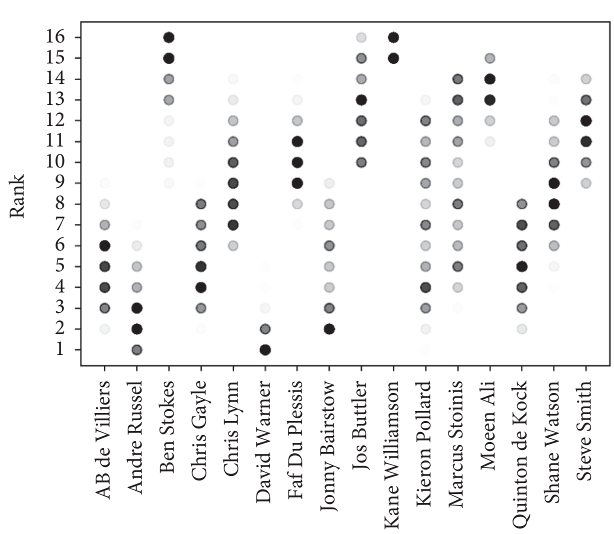


Figure 3: Heat map for batsmen ranks generated by TOPSIS.

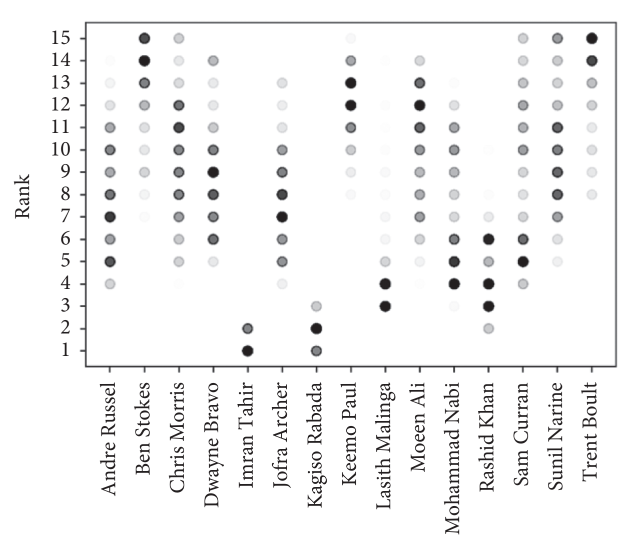


Figure 4: Heat map for bowler ranks generated by TOPSIS.

comparable black or lighter coloured dots shows that such positions were likewise achieved for certain sets of criterion weights.

5.3. Criteria Shuffling or Weight Shuffling for TODIM. We repeat the same procedure for generating a heat map of the rankings generated by TODIM when the weights are shuffled. The results are shown in Figures 7 and 8.

Unlike TOPSIS, where players like Chris Morris (bowler) and Kieron Pollard (batsman) have a wide range of ranks that shift dramatically with weight changes, the TODIM rankings are stable ranks differing by four positions at most. Even for Sam Curran (bowler), who had the highest rating range of 6–12, the darkest point is at rank 8. We can also notice that several alternatives exhibit stable behaviour, such as the first, second, or last and second last rankers of bowlers and batsmen, who keep their rank regardless of the weight or criteria ranking.

5.4. Change in Attenuation Factor (θ) for TODIM. The sensitivity analysis studied from [34] is formulated by taking six randomly selected different values of attenuation factor. In our case, we observe the effects of change in the attenuation factor when it is varied from 1 to 101 with increments of 0.1, providing us around 1000 observations. If the value of θ is too large, the penalty for being inferior in a category becomes insignificant. The results are graphed in Figures 9 and 10.

We can see the ranks that all the players have had for various values of θ, the attenuation factor. David Warner (batsman) and Imran Tahir (bowler), for example, had rank 1 for all values of θ. As a result, there are no grey dots areas for them. Pollard (batsman), on the other hand, had a rank system that ranges from 7 to 9, with 7 being the most common and 9 being the least common. Similarly, for all values of θ, Ben Stokes (batsman) and Kane Williamson (batsman) have been last or second last.

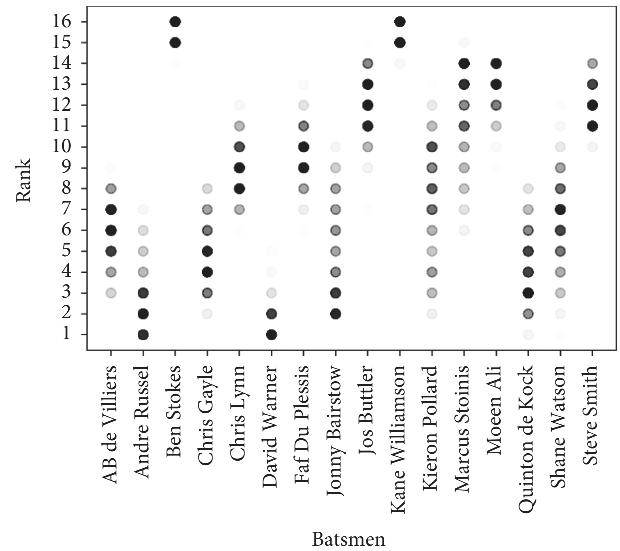


Figure 5: Heat map for batsmen ranks generated by NR-TOPSIS.

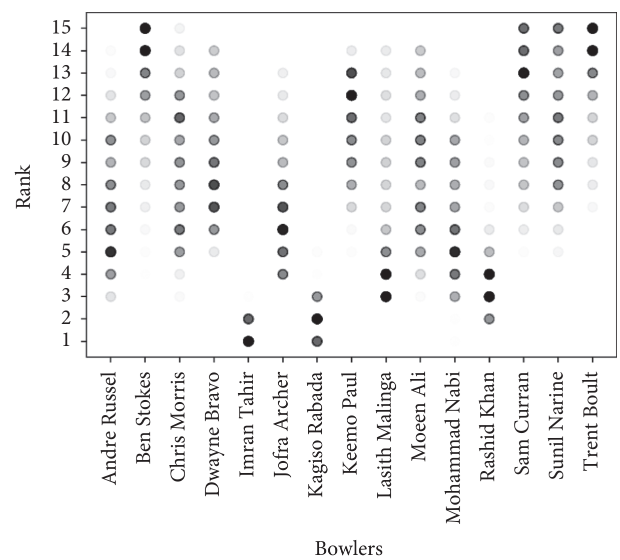


Figure 6: Heat map for bowlers ranks generated by NR-TOPSIS.

6. Discussion and Future Scope

We do not need to apply our models on a fuzzy environment here because our data is obtained from a reliable and publicly accessible source. When it came to choosing the players, no geographical preferences were made. This is one of the reasons why we chose international players rather than Indian ones. TOPSIS and TODIM were used in this research to analyse the performance of international players during the selection of persons through auction in the IPL, and we could obtain conclusive results. But rank reversal is an issue with both methods. As a result, we implemented NR-TOPSIS [52] to address the problem, which is a modified version of TOPSIS. We did, however, run across a rank reversal issue with NR-TOPSIS. One of the probable reasons might be the selection of weighting method. Although NR-TOPSIS and TODIM are strongly correlated, there is no advantage of implementing NR-TOPSIS as it is more unstable than TODIM. Hence, for ranking, we have proven that TODIM is better than both TOPSIS and NR-TOPSIS. We chose rank-sum weight method for our models instead of implementing our own randomly selected values. We present the differences between the two basic methods TOPSIS and TODIM in Table 19.

In this article, our findings proved that, whether players were added or withdrawn, TODIM saw a significantly lower change in rank. After sensitivity analysis of the rankings, we observed that the rankings generated by TOPSIS and NR-TOPSIS are both more susceptible to change of weights than the rankings generated by TODIM. In article [34], it is stated that the final rankings from TOPSIS will vary if any single weight varies in proportion by ±20% to ±50%. Since we are only allowing decision-makers to set the rankings of the judgement criteria and using these rankings to calculate criteria weights, the changes in proportion are larger than 20%, thereby affecting the final rankings and making the outcome unstable. In case of TODIM, the rankings of the

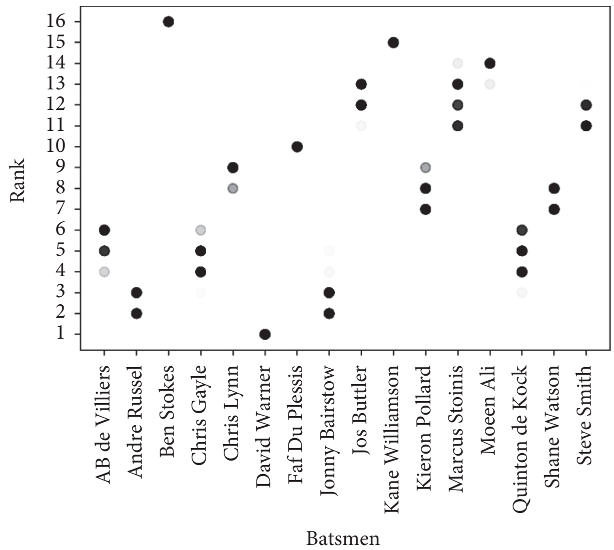


FIGURE 7: Heat map for batsmen ranks by TODIM through shuffled weights.

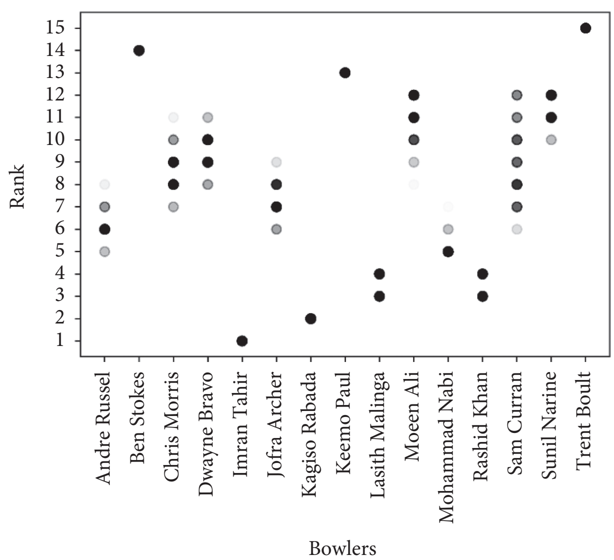


FIGURE 8: Heat map for bowler ranks by TODIM through shuffled weights.

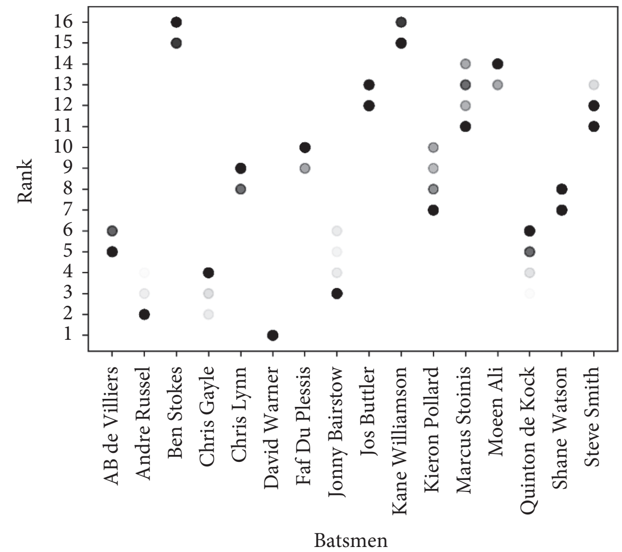


FIGURE 9: Heat map for batsmen ranks by TODIM through varying θ.

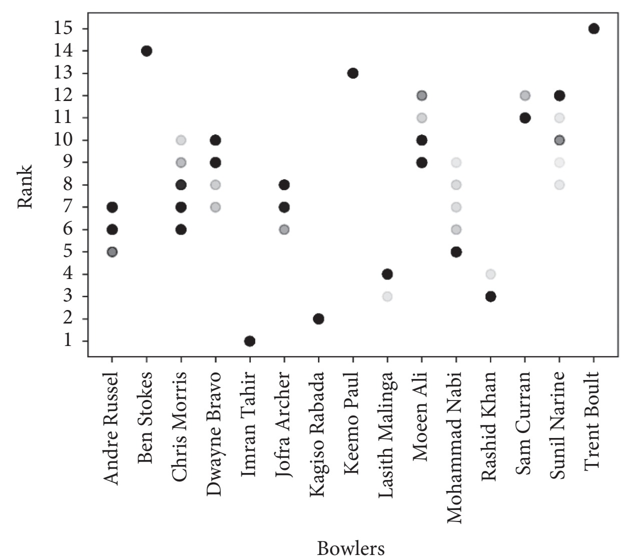


FIGURE 10: Heat map for bowlers ranks by TODIM through varying θ.

criteria are not as important because the final ranks for the candidates change little. The decision-makers can choose any parameter values they like, and the TODIM model will give stable rankings anyway. Sensitivity analysis done for a range of values of attenuator factor θ and weight shuffling for criteria proved the stability of TODIM. Therefore, this method is recommended over TOPSIS and NR-TOPSIS for any level of decision-makers, from experts to less experienced. The ranks for the best and worst performers remain relatively the same for both approaches, with the candidates in the middle hovering near their general regions.

Amos Tversky and Daniel Kahneman proposed cumulative prospect theory (CPT) in 1992 [61] as a model for describing judgments under risk and uncertainty. The main difference between prospect theory (PT) and cumulative prospect theory is that, rather than the probabilities themselves, the cumulative probabilities are transformed in CPT. Thereby, CPT removes the limitation of PT which violates first-order stochastic dominance. We did not use CPT-TODIM in our paper since we wanted to demonstrate the superiority of TODIM over TOPSIS for rank alternatives or IPL foreign player performance analysis. We have therefore demonstrated that TODIM is better than TOPSIS in terms of sensitivity and accuracy without having to use any modified version of TODIM. However, the following questions may arise: What if the source we used is deemed untrustworthy by people? What if a circumstance arises in which player performance must be analysed based on more than simply their scores? In such scenario, scholars may want to employ CPT-TODIM in a fuzzy environment.

Table 19: Difference between TOPSIS and TODIM.

	TOPSIS	TODIM
1.	The essential premise is that the chosen alternative should be closest to the Positive Ideal Solution and farthest from the Negative Ideal Solution. The metric most used is Euclidean distance.	The chosen alternative should have better scores for most criteria against all alternatives. Having a lesser score is penalised by an amount determined by θ.
2.	It has a lower time complexity for comparison, $\mathbf{O}\,(\mathbf{mn})$, where $\mathbf{m}$ is the number of criteria and $\mathbf{n}$ is the number of candidates.	It has a higher time complexity for comparison, $\mathbf{O}\,(\mathbf{mn}^2)$, where $\mathbf{m}$ is the number of criteria and $\mathbf{n}$ is the number of candidates.

Researchers may also want to study the risk of the economic side of the IPL auction, and in such cases as well CPT-TODIM in a fuzzy environment can be used. Some articles [40, 62, 63] presented implementation of fuzzy CPT-TODIM for diverse decision-making applications in recent studies. Future research can be conducted to compare the findings of TODIM with those of CPT-TODIM or in a fuzzy environment for the IPL dataset. DEA is another quantitative tool that can be used for foreign player ranking in the Indian Premier League. In 1978, Charnes, Cooper, and Rhodes proposed Data Envelopment Analysis (DEA) as a performance evaluation method. It is sometimes known as frontier analysis. It is a strategy for assessing the relative efficiency of decision-making units (DMUs) in organizations. A DMU is a discrete unit inside an organization that has flexibility in some of its decisions but not total freedom in others. In the case of IPL, runs per innings, boundaries per innings, and so on can be commonly used to determine efficiency. The relative efficiencies are then determined using the highest ratio as a reference. Future research might use DEA [64–66] to assess the player's efficiency in Indian Premier League.

We did not pick any additional MCDM techniques for our comparative analysis because of practicality and ease of comprehending the approaches. We proved through this research work that TODIM is superior to TOPSIS and its variations. The simplicity and high usage of the algorithm are the reasons why TOPSIS was chosen over other MCDM approaches in this study. TODIM was chosen because it is the only approach based on prospect theory, the result of which the decision-makers can use to auction overseas players. We have therefore proved in our article that TODIM is better than TOPSIS and NR-TOPSIS through the sensitivity analysis. The ease with which it can be computed in challenging scenarios and its low sensitivity are some other reasons for selecting this method.

We had also attempted to alter TODIM by applying the same modification used in NR-TOPSIS, that is, a change in the normalization procedure formulation. However, no meaningful results were obtained, and the rank reversal problem persisted. In fact, the modified TODIM was less stable than the original TODIM and was more sensitive to changes in weight or attenuation factor. Rank reversal may occur in TODIM, as shown in our work; as a future scope, it should be explored further, so that the problem may be eradicated from the approach and the application can be applied on a wide scale. Further, CPT-TODIM and modified variants of TODIM can be applied to the IPL dataset in the future to compare international player rankings for the IPL auction. Lastly, all the player statistics were obtained from the official website (http://www.iplt20.com). We did not take into consideration the statistics from IPL 2020 or IPL 2021 because the matches were played during the COVID-19 pandemic, which could have affected the players' general confidence and mental health, preventing them from performing as well as expected. 2019 was the latest year unaffected by the effects of the pandemic and, consequently, our year of choice.

7. Conclusion

IPL cricket players' salaries are determined through an auction process. The same rule applies to foreign players as well. As a result, the team owners must make a judgement based on the performance of the players in previous matches of IPL to determine which player to bid for and at what price. So, the models proposed in this paper might assist a franchisee in selecting the right players available. The IPL management releases a list of international players who may draw a little attention for a maximum number of slots available during the auction. For example, according to one article [67], the IPL management had released a list of 292 international players for sixty-one places available for the 2021 auction. We ranked players in this paper based only on statistics from the 2019 season of IPL as an example. During the real-life player enlistment process, decision-makers can change these criteria at their will.

David Warner and Andre Russell have a shared first rank among batsmen, while Imran Tahir and Kagiso share the top spot among bowlers, according to our TOPSIS, NR-TOPSIS, and TODIM approaches. Hence, these players can establish higher starting prices during the IPL auction. Since the team owners may need to work on several lists of players simultaneously, our recommended approach TODIM is basic and straightforward to execute. To compute the value of the weights, we use the rank-sum weight approach, which comprises rating the criteria. As a result, it is up to the decision-makers to prioritize the criteria. In this paper, we performed sensitivity analysis on our algorithms to examine how the rank may vary owing to weight changes. TOPSIS and NR-TOPSIS are definitely affected by this weighting scheme, and the subjective preferences of the decision-maker are reflected properly. However, TODIM allows retention of subjectivity along with improved stability. If the decision-makers cannot rank the criteria in order of importance, they can give the same rank to conflicting criteria. Even when the attenuation factor or weight values were changed, the model remained quite stable. The fact that TODIM's higher time complexity might be viewed as a detractor is overshadowed by the accuracy of the method. As

TODIM was never used before to rank players in IPL or any sports, we have demonstrated how this approach can be useful. We also looked at the differences between two well-known MCDM techniques (and a modification) and how their behaviour affects the ranking of alternatives. It is therefore recommended that decision-makers should select TODIM over TOPSIS and NR-TOPSIS based on their requirements.

However, while utilising NR-TOPSIS to solve the heavy rank reversal problem of TOPSIS, we observed that the problem still exists. This showed TODIM's superiority, making it the preferred technique for ranking players. We have included the Python code for TOPSIS, NR-TOPSIS, and TODIM in the Supplementary Materials section of this paper. Therefore, researchers may use the comparative analysis for additional scenarios in the future, not simply rating players, to aid in practicality.

Data Availability

The data are available at https://resources.platform.iplt20.com/IPL/document/2021/04/04/35d7aa60-14d1-4260-bcc1-72b8d1e461ba/IPL-2021-Match-Playing-Conditions.pdfURL, retrieved: 2021-05-22.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Supplementary Materials

The Python source code and the intermediate steps are available at the following repository: https://github.com/hungrybluedev/MCDM_IPLWe also provide the output generated by the Python code formatted into a single PDF file that contains all the intermediate steps of the calculations. The order in which the code and results are presented in the documentis as follows:

(1) TOPSIS, batsmen ranking
(2) TOPSIS, bowlers ranking
(3) NR-TOPSIS, batsmen ranking
(4) NR-TOPSIS, bowlers ranking
(5) TODIM, batsmen ranking
(6) TODIM, bowlers ranking (*Supplementary Materials*)

References

[1] Iplt20, "IPL 2021 Playing Conditions, Page 4, Section 1.2.5," 2021, https://resources.platform.iplt20.com/IPL/document/2021/04/04/35d7aa60-14d1-4260-bcc1-72b8d1e461ba/IPL-2021-Match-Playing-Conditions.pdf.

[2] I P L, "Indian Premier League," 2007, https://en.wikipedia.org/wiki/Indian_Premier_League#Organisation.

[3] R. Greene, R. Devillers, J. E. Luther, and B. G. Eddy, "GIS-based multiple-criteria decision analysis," *Geography Compass*, vol. 5, no. 6, pp. 412–432, 2011.

[4] M. El Amine, JérômePailhes, and N. Perry, "Comparison of different Multiple-criteria decision analysis methods in the context of conceptual design: application to the development of a solar collector structure," in *Proceedings of theJoint Conference on Mechanical, Design Engineering & Advanced Manufacturing*, pp. 1–6, Toulouse, France, June 2014.

[5] E. Roszkowska, "Multi-criteria decision making models by applying the TOPSIS method to crisp and interval data," *Multiple Criteria Decision Making/University of Economics in Katowice*, vol. 6, no. 1, pp. 200–230, 2011.

[6] G. Kabir and M. Hasin, "Comparative analysis of topsis and fuzzy topsis for the evaluation of travel website service quality," *International Journal for Quality Research*, vol. 6, no. 3, 2012.

[7] L. F. A. M. Gomes and M. M. P. P. Lima, "TODIM: basics and application to multicriteria ranking of projects with environmental impacts," *Foundations of Computing and Decision Sciences*, vol. 16, no. 4, pp. 113–127, 1992.

[8] R. A. Krohling and T. T. M. de Souza, "Combining prospect theory and fuzzy numbers to multi-criteria decision making," *Expert Systems with Applications*, vol. 39, no. 13, Article ID 11487, 2012.

[9] J. Wang, G. Wei, and M. Lu, "TODIM method for multiple attribute group decision making under 2-tuple linguistic neutrosophic environment," *Symmetry*, vol. 10, no. 10, p. 486, 2018.

[10] Google, "Google Trends," 2004, https://trends.google.com/trends/explore?date=all&q=TOPSIS,TODIM.

[11] R. Lourenzutti, R. A. Krohling, and M. Z. Reformat, "Choquet based TOPSIS and TODIM for dynamic and heterogeneous decision making with criteria interaction," *Information Sciences*, vol. 408, pp. 41–69, 2017.

[12] R. A. Krohling, A. D. Santos, and A. G. Pacheco, "TODIM and TOPSIS with Z-Numbers," 2016, https://arxiv.org/ftp/arxiv/papers/1609/1609.05705.pdf.

[13] A. ArshadiKhamseh and M. Mahmoodi, "A new fuzzy TOPSIS-TODIM hybrid method for green supplier selection using fuzzy time function," *Advances in Fuzzy Systems*, vol. 2014, Article ID 841405, 10 pages, 2014.

[14] V. D. Baloyi and L. D. Meyer, "The development of a mining method selection model through a detailed assessment of multi-criteria decision methods," *Results in Engineering*, vol. 8, Article ID 100172, 2020.

[15] M. Hanine, O. Boutkhoum, A. Tikniouine, and T. Agouti, "Comparison of fuzzy AHP and fuzzy TODIM methods for landfill location selection," *SpringerPlus*, vol. 5, p. 501, 2016.

[16] S. Mufazzal and S. M. Muzakkir, "A new multi-criterion decision making (MCDM) method based on proximity indexed value for minimizing rank reversals," *Computers & Industrial Engineering*, vol. 119, pp. 427–438, 2018.

[17] Z. Xing, "Study on the application of TOPSIS method to the introduction of foreign players in CBA games," *Physics Procedia*, vol. 33, pp. 2034–2039, 2012.

[18] P. Kumar Dey and D. N. Ghosh, "Abhoy chand mondal,IPL team performance analysis," *A Multi-Criteria Group Decision Approach in Fuzzy Environmentsag*, vol. 7, no. 8, pp. 8–15, 2015.

[19] P. K. Dey, D. N. Ghosh, and A. C. Mondal, "Multi-criteria decision tree approach to classify all-rounder in Indian premier league," *Journal of Emerging Trends in Computing and Information Sciences*, vol. 2, no. 11, pp. 563–573, 2011.

[20] P. K. Dey, D. N. Ghosh, and A. C. Mondal, "Modified group decision algorithm for performance appraisal of Indian premier league cricketers," *WSEAS Transactions on Information Science and Applications*, vol. 10, no. 10, pp. 325–333, 2013.

[21] U. Kashid, S. N. Mehta, and V. Basotia, "A Hybrid AHP-TOPSIS Mathematical Multi Criteria Decision-Making Model for Players Performance Evaluation and Selection in IPL a Case Study," in *Proceedings of the National Conference on New Frontiers of Innovation in Management, Social Science, and Technology and Their Impact on Societal Development AU-FAIT*, vol. 5, no. II, India, October 2019.

[22] P. K. Dey, D. N. Ghosh, and A. C. Mondal, "Statistical based multi-criteria decision making analysis for performance measurement of batsmen in Indian premier league," *International Journal of Advanced Research in Computer Science*, vol. 3, no. 4, 2012.

[23] M. Khandelwal, J. Prakash, and T. Pradhan, "A novel approach for performance analysis and optimal selection of players in Indian premier league auction," in *Emerging Research in Computing, Information, Communication and Applications*, N. Shetty, N. Prasad, and N. Nalini, Eds., Springer, New Delhi, India, 2016.

[24] M. Saqlain, N. Jafar, R. Hamid, and A. Shahzad, "Prediction of cricket world Cup 2019 by TOPSIS technique of MCDM-A mathematical analysis," *International Journal of Scientific Engineering and Research*, vol. 10, no. 2, pp. 789–792, 2019.

[25] J.-J. Yang, Y.-C. Chuang, H.-W. Lo, and T.-I. Lee, "A two-stage MCDM model for exploring the influential relationships of sustainable sports tourism criteria in Taichung city," *International Journal of Environmental Research and Public Health*, vol. 17, no. 7, p. 2319, 2020.

[26] C.-Y. Wang, P.-H. Tsai, and Z. Hu, "Constructing taipei city sports centre performance evaluation model with fuzzy MCDM approach based on views of managers," *Mathematical Problems in Engineering*, vol. 2013, Article ID 138546, 13 pages, 2013.

[27] P. H. Tsai and C. F. Lin, "Creating a management strategy evaluation model for taipei city sports centre by using hybrid MCDM models," *Journal of Testing and Evaluation*, vol. 45, no. 5, pp. 1820–1836, 2017.

[28] D. Kahneman and A. Tversky, "Prospect theory: an analysis of decision under risk," *Econometrica*, vol. 47, pp. 263–292, 1979.

[29] D. Kahneman and A. Tversky, "Chapter 6: prospect theory: an analysis of decision under risk," in *Handbook of the Fundamentals of Financial Decision Making: Part I*, pp. 99–127, World Scientific, Hackensack, NJ, USA, 2013.

[30] B. F. Dehaghia and K. Ali, "Modified TODIM method for water reuse application: a case study from Iran," *Desalination and Water Treatment*, vol. 195, pp. 186–200, 2020.

[31] L. F. A. M. Gomes, M. A. S. Machado, and L. A. D. Rangel, "Behavioral multi-criteria decision analysis: the TODIM method with criteria interactions," *Annals of Operations Research*, vol. 211, no. 1, pp. 531–548, 2013.

[32] L. A. D. Rangel, L. F. A. M. Gomes, and F. P. Cardoso, "An application of the TODIM method to the evaluation of broadband internet plans," *Pesquisa Operacional*, vol. 31, pp. 235–249, 2011.

[33] P. Li, H. Qian, J. Wu, and J. Chen, "Sensitivity analysis of TOPSIS method in water quality assessment: I. Sensitivity to the parameter weights," *Environmental Monitoring and Assessment*, vol. 185, pp. 2453–2461, 2013.

[34] M. Li, C. Wu, L. Zhang, and L. N. You, "An intuitionistic fuzzy-TODIM method to solve distributor evaluation and selection problem," *International Journal of Simulation Modelling*, vol. 14, no. 3, pp. 511–524, 2015.

[35] P. Liu, M. Shen, F. Teng, B. Zhu, L. Rong, and Y. Geng, "Double hierarchy hesitant fuzzy linguistic entropy-based TODIM approach using evidential theory," *Information Sciences*, vol. 547, pp. 223–243, 2021.

[36] M. Zhao, G. Wei, C. Wei, and J. Wu, "Improved TODIM method for intuitionistic fuzzy MAGDM based on cumulative prospect theory and its application on stock investment selection," *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 3, pp. 891–901, 2021.

[37] Z. Lu, Y. Gao, and W. Zhao, "A TODIM-based approach for environmental impact assessment of pumped hydro energy storage plant," *Journal of Cleaner Production*, vol. 248, Article ID 119265, 2020.

[38] M. Zhao, G. Wei, C. Wei, and J. Wu, "TODIM method for interval-valued Pythagorean fuzzy MAGDM based on cumulative prospect theory and its application to green supplier selection," *Arabian Journal for Science and Engineering*, vol. 46, no. 2, pp. 1899–1910, 2021.

[39] S. S. He, Y. T. Wang, J. J. Peng, and J. Q. Wang, "Risk ranking of wind turbine systems through an improved FMEA based on probabilistic linguistic information and the TODIM method," *Journal of the Operational Research Society*, vol. 73, pp. 1–14, 2020.

[40] M. Zhao, G. Wei, C. Wei, and J. Wu, "Pythagorean fuzzy TODIM method based on the cumulative prospect theory for MAGDM and its application on risk assessment of science and technology projects," *International Journal of Fuzzy Systems*, vol. 23, pp. 1027–1041, 2021.

[41] X. Tian, W. Li, L. Liu, and G. Kou, "Development of TODIM with Different Types of Fuzzy Sets: A State-Of The-Art Survey," *Applied Soft Computing*, vol. 111, Article ID 107661, 2021.

[42] V. Belton and T. Gear, "On a short-coming of Saaty's method of analytic hierarchies," *Omega*, vol. 11, no. 3, pp. 228–230, 1983.

[43] B. Kizielewicz, A. Shekhovtsov, and W. Sałabun, "A new approach to eliminate rank reversal in the mcda problems," in *Proceedings of the International Conference on Computational Science*, pp. 338–351, Springer, Krakow, Poland, June 2021.

[44] W. Yang and Y. Wu, "A new improvement method to avoid rank reversal in VIKOR," *IEEE Access*, vol. 8, Article ID 21261, 2020.

[45] R. K. Tiwari and R. Kumar, "G-TOPSIS: a cloud service selection framework using Gaussian TOPSIS for rank reversal problem," *The Journal of Supercomputing*, vol. 77, no. 1, pp. 523–562, 2021.

[46] M. Keshavarz-Ghorabaee, M. Amiri, E. K. Zavadskas, Z. Turskis, and J. Antucheviciene, "A Comparative analysis of the rank reversal phenomenon in the EDAS and TOPSIS methods," *Economic Computation & Economic Cybernetics Studies & Research*, vol. 52, no. 3, 2018.

[47] R. F. D. F. Aires and L. Ferreira, "The rank reversal problem in multi-criteria decision making: a literature review," *Pesquisa Operacional*, vol. 38, pp. 331–362, 2018.

[48] R. Agrawal, "Sustainable material selection for additive manufacturing technologies: a critical analysis of rank reversal approach," *Journal of Cleaner Production*, vol. 296, Article ID 126500, 2021.

[49] A. Piegat and W. Sałabun, "Comparative analysis of MCDM methods for assessing the severity of chronic liver disease," in *Proceedings of the International Conference on Artificial Intelligence and Soft Computing*, pp. 228–238, Springer, Zakopane, Poland, June 2015.

[50] L. F. Gomes, "Eliminating rank reversal in multicriteria analysis of urban transportation system alternatives," *Journal of Advanced Transportation*, vol. 24, no. 2, pp. 181–184, 1990.

[51] R. F. D. F. Aires and L. Ferreira, "A new approach to avoid rank reversal cases in the TOPSIS method," *Computers & Industrial Engineering*, vol. 132, pp. 84–97, 2019.

[52] W. Yang, "Ingenious solution for the rank reversal problem of TOPSIS method," *Mathematical Problems in Engineering*, vol. 2020, Article ID 9676518, 12 pages, 2020.

[53] N. Zoraghi, M. Amiri, G. Talebi, and M. Zowghi, "A fuzzy MCDM model with objective and subjective weights for evaluating service quality in hotel industries," *J IndEngInt*, vol. 9, p. 38, 2013.

[54] E.. Roszkowska, "Rank ordering criteria weighting methods-A comparative overview," *Optimum StudiaEkonomiczne NR*, vol. 5, no. 65, pp. 14–33, 2013.

[55] W. G. Stillwell, D. A. Seaver, and W. Edwards, "A comparison of weight Approximation techniques in multiattribute utility decision-making," *Organizational Behavior & Human Performance*, vol. 28, pp. 62–77, 1981.

[56] C.-L. Hwang and K. Yoon, *Methods for Multiple Attribute Decision making. Multiple Attribute Decision Making*, Springer, Berlin, Germany, pp. 58–191, 1981.

[57] S. C Gupta and V. K Kapoor, *Fundamentals of Mathematical Statistics*, Sultan Chand and Sons, New Delhi, India, 1970.

[58] B. Keçeci, Y. T. Iç, and E. Eraslan, "Development of a spreadsheet DSS for multi-response taguchi parameter optimization problems using the TOPSIS, VIKOR, and GRA methods," *International Journal of Information Technology and Decision Making*, vol. 18, pp. 1501–1531, 2019.

[59] D. Kannan, R. Khodaverdi, L. Olfat, A. Jafarian, and A. Diabat, "Integrated fuzzy multi criteria decision making method and multi-objective programming approach for supplier selection and order allocation in a green supply chain," *Journal of Cleaner Production*, vol. 47, pp. 355–367, 2013.

[60] A Mohammed, M Yazdani, A Oukil, and E. D. R Santibanez Gonzalez, "A hybrid MCDM approach towards resilient sourcing," *Sustainability*, vol. 13, no. 5, p. 2695, 2021.

[61] A. Tversky and D. Kahneman, "Advances in prospect theory: cumulative representation of uncertainty," *Journal of Risk and Uncertainty*, vol. 5, pp. 297–323, 1992.

[62] M. Zhao, G. Wei, Y. Guo, and X. Chen, "CPT-TODIM Method for Interval-Valued Bipolar Fuzzy Multiple Attribute Group Decision Making and Application to Industrial Control Security Service Provider Selection," *Technological and Economic Development of Economy*, vol. 27, 2021.

[63] N. Liao, G. Wei, and X. Chen, "TODIM method based on cumulative prospect theory for multiple attributes group decision making under probabilistic hesitant fuzzy setting," *International Journal of Fuzzy Systems*, vol. 24, pp. 322–339, 2022.

[64] Q. Zhu, J. Aparicio, F. Li, J. Wu, and G. Kou, "Determining closest targets on the extended facet production possibility set in data envelopment analysis: modeling and computational aspects," *European Journal of Operational Research*, vol. 296, 2021.

[65] Q. Zhu, X. Li, F. Li, J. Wu, and D. Zhou, "Energy and environmental efficiency of China's transportation sectors under the constraints of energy consumption and environmental pollution," *Energy Economics*, vol. 89, Article ID 104817, 2020.

[66] Q. Zhu, F. Li, J. Wu, and J. Sun, "Cross-efficiency evaluation in data envelopment analysis based on the perspective of fairness utility," *Computers & Industrial Engineering*, vol. 151, Article ID 106926, 2021.

[67] The hindu, "Ipl-2021-auction-overseas-players-full-list-smith-labuschagne-maxwell," 2021, https://sportstar.thehindu.com/cricket/ipl-2021-auction-overseas-players-full-list-smith-labuschagne-maxwell/article33821497.ece.

Hindawi
Complexity
Volume 2022, Article ID 4069896, 10 pages
https://doi.org/10.1155/2022/4069896

Research Article

A Corpus-Based Study of Public Attitudes towards Coronavirus Vaccines

Ganlin Xia[1] **Yiting Chen**[2] **and Lijing Lu**[3]

[1]*Shanghai University of Political Science and Law, Shanghai 200030, China*
[2]*Shanghai Meiyuan Primary School, Shanghai 200082, China*
[3]*Hebei University of Economics and Business, Shijiazhuang 050067, China*

Correspondence should be addressed to Ganlin Xia; xiaganlin@shupl.edu.cn

Received 22 December 2021; Revised 8 February 2022; Accepted 16 February 2022; Published 16 March 2022

Academic Editor: Feng Li

Since the beginning of 2020, COVID-19 has been sweeping the world on an unprecedented scale. As an important means of fighting the virus, vaccines have provoked heated discussions. Motivated by this practical concern, the present study aims at contributing to the understanding of public opinions about vaccines, which may provide implications for the government in their making and implementation of related policies. This research adopts a corpus-based approach in conjunction with a Critical Discourse Analysis (CDA). Data for this study drawn from the Coronavirus corpus show what people are actually saying in online newspapers and magazines in 20 different English-speaking countries. The collocation and frequency of the word "vaccine" are arranged in concordance contextually. Overall, this study reveals that the collocation of vaccine can be divided into several categories, and people's major concerns about COVID-19 vaccination include global progress, equality, and the latest development.

1. Introduction

Outbreaks of infectious diseases such as Ebola, SARS, and H1N1 flu are common around the world. Since the first half of 2020, Coronavirus has been sweeping the world on an unprecedented scale, affecting every aspect of people's lives. The outbreak is reported in Wuhan, a major city in China's Hubei province. In a very short time, the disease spread so fast that it reached almost every part of China. As a result of mass migration during the Chinese Lunar New Year, the virus began to spread abroad. It was proclaimed by the World Health Organization (WHO) as a pandemic on March 11, 2020. At present, the global Coronavirus is still ongoing, posing a serious threat to the life safety of people all over the world. To some extent, we have come into a new postcorona era.

With no particularly effective drugs to treat Coronavirus, the development of vaccines has been racing against time since the outbreak. As an important means of fighting the virus, the development of vaccines has attracted the attention of people all over the world. However, although laboratories and medical institutions are vigilant in developing a vaccine, their attempts are not effective enough. Besides, because of the lack of knowledge and information about fatalities, diagnosis, medicines, and vaccines of COVID-19, unnecessary fear and nervousness are generated [1]. These factors led to a war of words and an exchange of blame. Governments, public health agencies, and awareness organizations must be prepared to tackle hesitancy and improve vaccine awareness so that, where possible, the population supports immunization. Antivaccination groups are now lobbying against the need for a vaccine in several nations, with others questioning the presence of COVID-19 entirely. Misinformation transmitted across different networks may have a major influence on acceptance.

However, there is only very limited research on public attitude towards COVID-19 vaccines. Therefore, motivated by this practical concern, the present study is aimed at contributing to the understanding of public opinions about vaccines, which may provide implications for governments in their making and implementation of related policies.

This research uses a corpus-based approach in conjunction with a Critical Discourse Analysis (CDA), following the Discourse Analytical Perspective in Dijk [2] to probe public attitudes towards the COVID-19 vaccine. Data for this study drawn from the Coronavirus corpus show what people are actually saying in online newspapers and magazines in 20 different English-speaking countries. In May 2020, the corpus was initially published. Today, there are about 901 million words in the corpus, and every day it keeps rising by 3-4 million words. It is the definitive record of the social, cultural, and economic impact of Coronavirus (COVID-19) in 2020 and beyond. Collocation and frequency of the word "vaccine" can be concordanced contextually. Through word frequency, word collocation, sentiment analysis towards adjectives, and other research methods, the public's attitude towards Coronavirus vaccines can be analyzed and perceived [3]. In addition, since the media is the main source for the public to acquire information about vaccine development, its role in describing the current phenomenon is also worth exploring. Specifically, this paper wants to focus on the following research questions:

(1) What are the most common adjective collocations for "vaccines"?

(2) How are "vaccines" portrayed contextually in online discourses?

(3) What is the general picture of the public attitude towards Coronavirus vaccines?

This study reveals that the collocation of the word "vaccine" can be divided into several categories, indicating that people's major concerns about COVID-19 vaccination include global progress, equality, and the latest development.

The rest of this paper is organized as follows. In Section 2, the theoretical framework and related studies are reviewed, introducing the history of vaccine, research methods of CDA and corpus linguistics, and their applications in studies related to vaccines. In Section 3, instruments used and data collection procedures for this study are described, while in Section 4, results of data analysis are reported. In Section 5, research questions put forward in Section 1 are answered, policy implications are proposed for governments and other organizations, and further research directions are suggested accordingly.

2. Literature Review

2.1. Theoretical Framework

2.1.1. Controversy over Vaccines and Vaccination. Vaccination ranked first among the ten greatest public health achievements of the 20th century [4], since vaccines have saved millions of lives by preventing contagious diseases from spreading [5]. Vaccines are regarded as an important guarantee to maintain social security and improve social productivity. Smallpox vaccination was mandatory in Europe and North America in the 19th century; in the 20th century, certain vaccinations were also included in the admission requirements of public schools [6].

However, vaccination is controversial. A lot of problems we have today can date back to the 1790s, when the first human vaccine was produced by Edward Jenner for curing smallpox ([6]: 612). Since then, the argument of safety, money, and proper immunization schedules has aroused a growing concern from the public, including parents, medical profession, policymakers and the media. Their attitudes vary from "awe of a seeming scientific miracle to skepticism and outright hostility" ([6]: 613). Margaret Chan, the Director-General of the World Health Organization, once described the wide distrust of vaccination as "worrisome" ([7]:1151).

Increasing anxiety and panic leads to some quitting. The antivaccination movement began to gain support in the United States at the end of the 19th century when smallpox outbroke. Parents' anxiety and protest were triggered by the smallpox vaccine itself, which violates their personal rights. Tensions had escalated when a mandatory vaccination program was issued by the government. More than 100 years ago, Jacobson, a citizen of the city of Cambridge in Massachusetts, declined to be vaccinated for smallpox as he argued that the law violated his right to care for his own body. The Court dismissed the challenge raised by Jacobson. In order to protect the public's welfare, the U. S. Supreme Court ruled in 1905 that, in the case of infectious disease, the state of Massachusetts had the authority to protect the public with mandatory laws. That is the first U.S. Case on the power of states in public health law [8]. There is often a dispute between preserving individual liberties and protecting public health, and a major challenge is how to balance individual rights with community needs. Certain religions and value systems also fostered alternate vaccine views. In general, moral opposition to vaccination is based on the ethical dilemmas involved in the use of human tissue cells to manufacture vaccines and the conviction that such drugs or blood or tissues should not be received from animals and that they should be cured by God or by natural means [9].

In the mid-1970s, an international dispute arose in Europe, Asia, Australia, and North America over the immunological protection of DTP [10]. The UK is once again the center of antivaccination protests nearly 25 years after the DTP controversy, this time against the MMR vaccine. As Johnston [11] observed, most opponents of vaccines are not entirely opposed to immunization. What they are concerned about is the safety and efficacy of vaccines. Therefore, they resist only some specific vaccines with potential safety hazards.

2.1.2. CDA and Corpus Linguistics. Following the Discourse Analytical Approach in van Dijk [12], the CDA framework analyzes media texts not just in the text but also from the context. It was described as a sort of study of dialogue that focuses on the examination of social motivations or the philosophy of the speaker behind his or her language choice in a debate. The reason for using CDA to analyze these texts is that CDA is an approach to discourse research aimed at analyzing the use of language, whether spoken or written, in daily conversation. From this perspective, language is seen as a social activity [13] that reflects not only other social practices but also elements such as domination, opposition, control, and ideology [14].

Much of the research that is combined CDA with Corpus Linguistics (henceforth CL) has placed their emphasis on "the implications of lexical choices in the text" ([15]:16) after analyzing ideological ideas in corpora. According to Hunston [16], such studies usually choose high-frequency words as their research emphasis. The use of CL tends to establish integrity and trustworthiness for the research results since the researcher tries to gather a specialized corpus that is representative of the text type [16]. The textual study would analyze the language used in the texts in expressing the view of writers and their semantic and pragmatic meaning [17]. However, the qualitative method still needs to be adopted to analyze the newspaper's fiscal, political, and cultural values. The purpose of CL is not to discover the meaning in isolated words that are out of context. Instead, in a specific sense, the purpose is to identify patterns of word items used in discourse (consistent lines) and words that occur together (collocations) [17].

2.1.3. CDA and Media Discourse. This research focuses on the analysis of the reports in online newspapers, a typical discourse type. As the main function of news reports is to give definitions and labels to people's lives, their expression is the key to portraying real life. From the angle of linguistics, the most integrated study was done in Bednarek [18], who believes that the essence of media discourse is "manufactured." Bednarek and Caple [19] also show the embedded bias of newspapers. The media discourse of this study refers to vaccination in online magazines and newspapers that are deposited in the corpus. Their creators are journalists, and readers are the general public. This research will concentrate on the discourse, adopt the method of Van Dijk, and highlight the theme.

2.2. Related Studies. As Coronavirus started to spread rapidly, the scientific community was motivated to work together in order to gather, organize, and analyze data. At the beginning of the outbreak, most of the scientific research was focused on biology or medicine, and linguistic research was rare.

Katermina and Yachenko [20] and Kim [21] examined the linguistic phenomenon related to COVID-19 in English mass media texts to determine if mass media creates, reproduces, and transmits axiological values about COVID-19. The research studied the word collocation of "COVID-19," "Coronavirus," "virus," and "disease." Their results revealed that COVID-19 is widely discussed in the mass media in many ways. For example, some conceptualize the virus as war. The researchers concluded that the COVID-19 pandemic had affected culture and language to the extent that it can be considered a social calamity.

A contrasting analysis conducted in Awad AlAfnan [22] studied the coverage of COVID-19 in two newspapers, America's Washington Post and China's People's Daily. It is found that mass media expressed biased opinions in their coverage and declaration. In the study, both CDA and CL approaches were used, which revealed that the Chinese virus, Wuhan Virus, and KongFlu were used to refer to the COVID-19. In Washington Post, these labels were focused on creating a derogatory hollow among Asian Americans.

Azizan et al. [23] performed a study based on 15 Facebook posts about COVID-19. They used Constructive Discourse Analysis and Critical Discourse Analysis approaches. A thematic analysis was made to codify and categorize data, which revealed the creation of positive discourse. The posts highlighted the extensive use of collective pronouns such as "we" and "us" that symbolize unity and empowerment in dealing with COVID-19 among Malaysians. Regarding Facebook posts, four positive trends have been identified, which are faith, patriotism, call for heroism, and public awareness. The researchers concluded that individual actions subsequently served as a powerful buffer against negative discourse and as an attempt to make a difference to the emergence of subtle influence.

In most of the studies, Natural Language Processing (NLP), machine learning, sentiment analysis, and text mining were used for analysis by computer software. Many different aspects of COVID-19 and its impacts have been analyzed. There are some studies on the public opinions about the introduction of the COVID-19 vaccines. Some of these studies are based on the text of social media, seeking to understand the general public's acceptance or hesitance of vaccines and the influencing factors (e.g., [24–26]). There are also studies designed to understand the attitude of medical professionals towards vaccines, aiming to provide advice and guidance to nonprofessionals (e.g., [27, 28]). Compared with most of the above studies, the most remarkable characteristic of the current study is as follows. The text style in the corpus selected for this study is news reports, and the corpus includes the news media, journals, and magazines of 20 English-speaking countries in the world, so it can better represent the global public attitude.

3. Methodology

3.1. The Coronavirus Corpus. The discourse analyzed in this study is from English Corpora (https://www.english-corpora.org/). It is the most widely used collection of corpora in the world. Currently, the corpora are visited by more than 130,000 users every month from over 140 countries worldwide. Besides, hundreds of universities worldwide have applied for academic licenses, which allow their students and scholars to use the expanded and additional functions of the corpora.

The Coronavirus Corpus selected for this study is a subset of the NOW Corpus (News on the Web), a 14.3-billion-word corpus based on online newspapers and magazines from 2010 to the present time, with 180–200 million words added each month. The Coronavirus Corpus, with 1,350 million words by January 12, 2022, is also characterized by daily updates. It is selected for this study mainly because it is constructed centering around the news content related to COVID-19 on the Internet. Articles meeting the following two requirements are selected from the news websites: first, each article should have at least two occurrences of one of the three words, i.e., Coronavirus, COVID, or COVID-19; Second, each article should have one of the following words/strings in the title: at-risk, cases, confirmed, contagious, containm*(words start with "containm," the same below), Coronavirus, covid*, curbside, curve, deaths, disinfect*, distanc*, epicenter, epidemic, epidemiol*, flatten*, flu, high-risk, hoard*, hospital*, hydroxychloroquine, infect∗, influenza, isolat*, lockdown, lockdown, mask*, nursing, outbreak, pandemic, panic, patient*, pneumon*, preventative, preventive, quarantin*, reopen*, reopen*, respiratory, sanitiz*, self-isolat*, shelter*, shutdown, spread, spreading, stay-at-home, stay at home, stockpil*, testing, vaccine*, ventilator*, virus.

As claimed in the introduction of the corpus on the website, the Coronavirus Corpus can be regarded as "the definitive record of the social, cultural, and economic impact of the coronavirus (COVID-19) in 2020 and beyond" (https://www.english-corpora.org/corona/).

3.2. Data Collection and Analysis. Data between January 2020 and the end of April 2021 was collected from the abovementioned Coronavirus Corpus. Two functions of this corpus are used to analyze this study, frequency list, and collocation analysis.

The frequency list displays the frequency of words or phrases searched in 10-day increments since Jan 2020, which can not only provide the total frequency of the retrieved words but also indicate the changing characteristics of the frequencies in a certain period of time.

Collocation analysis was conducted to investigate people's attitudes towards COVID-19 vaccines, and the results list words that usually appear together with the searched word in context. Baker emphasizes that collocation has an ideological nature and that the recurrent collocation of lexical items demonstrates the association of two concepts in people's minds (2006: 114). These words can be found on the node word's left (L) side and right (R) side. The span of the word was set to search for the 2L and 2R around the node word "vaccine*" (words starting with vaccine including the word vaccines and other compounds) in this study. The minimum frequency of collocation was controlled to be 40, so the words shown in the list must have appeared around the node word over 40 times in the corpus. These settings are meant to filter out words that have no real meaning or do not occur frequently enough. Besides, in this research, mutual information3 (MI3) is adopted to measure the effect size of associations between a node word and its collocates (see [29] for detailed discussion).

4. Results

4.1. Frequencies. Figure 1 shows that the frequencies of the word "vaccine" in NOW Corpus have increased significantly since 2020. In 2020, the PER MIL (per million) frequency of it was 203.06, 8–11 times that of the previous years, while in 2021, the frequency per million words was 613.87, three times that of 2021. There is no doubt that while the epidemic is raging all over the world, people's attention to vaccines is bound to increase.

From Figure 2, it is clear that the frequency of the word "vaccine" has been increasing since January 2020. Vaccination has been put on the agenda in Asian countries such as China, European and American countries, and people around the world are paying increasing attention to and widely talking about vaccines. In December 2020, many countries, including China and the United States, successively approved the conditional marketing of vaccines [32, 33]. Therefore, since December 2020, the frequency of the word in the corpus has increased sharply and maintained this high frequency so far.

4.2. Collocation Analysis. The result of collocation concordance is shown in Figure 3. Adjectives modifying the word "vaccine" are grouped into three emotional categories, i.e., positive, negative, and neutral.

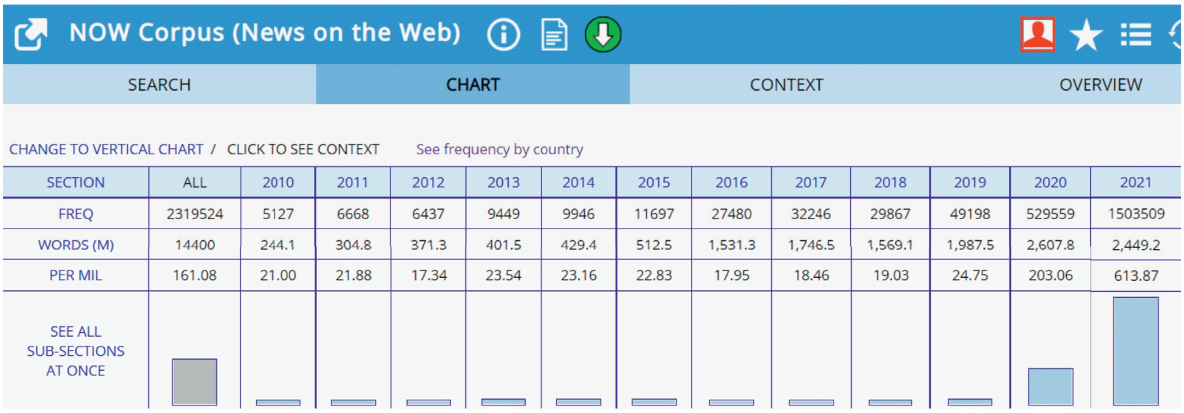
4.2.1. Positive Vaccine Portrayals. Based on the concordance results, the top100 adjectives are taken to illustrate the point. Among the top 100 most frequently used adjectives, positive and optimistic one are: "*available*," "*effective*," "*safe*," "*successful*," "*promising*," "*leading*," "*good*," "*viable*," "*positive*," *and* "*eligible*." Their specific rankings and frequency of use are shown in Table 1.

It can be discovered from Table 1 that the frequencies of the first and the second adjectives "*available*" and "*effective*" are very high, and the word "*safe*" ranked sixth and the word "*successful*" ranked 10, which are also the focus of public attention. It is quite reasonable that vaccine, as a special public security product, its safety and effectiveness are two of its characteristics that cannot be ignored. In order to gather a deeper understanding of how these words are used contextually, we referred to the concordance lines. Following Barnbrook [34], the power of concordance lines lies in the way "of placing each word back in its original context, so that the details of its use and behaviour can be properly examined" (p. 65). The following example shows a random sampling of concordance lines from a total of 2042 occurrences for the word "*successful*:"

(1) Given that at least 200,000 in the U.S. and more than a million globally have already died from COVID-19, we desperately need as many of the vaccines under development to succeed as soon as possible. # Success in dealing with COVID-19 requires far more than *successful* vaccine development, however. We need to eradicate the happy talk that, somehow, the pandemic will end when a vaccine is approved. #

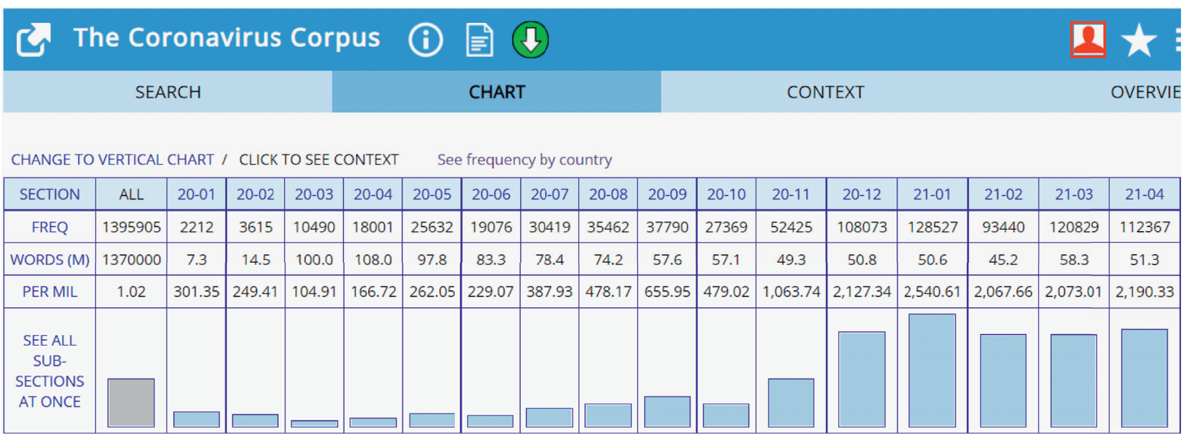
(2) "A *successful* vaccine against Sars-Cov-2 could be used to prevent infection, disease, and death in the whole population, with high-risk populations such as hospital workers and older adults, prioritized to receive the vaccination." #

(3) The global race to find a *successful* coronavirus vaccine continues at full pace. Meanwhile, the virus has claimed over 8 lakh lives across the world. #



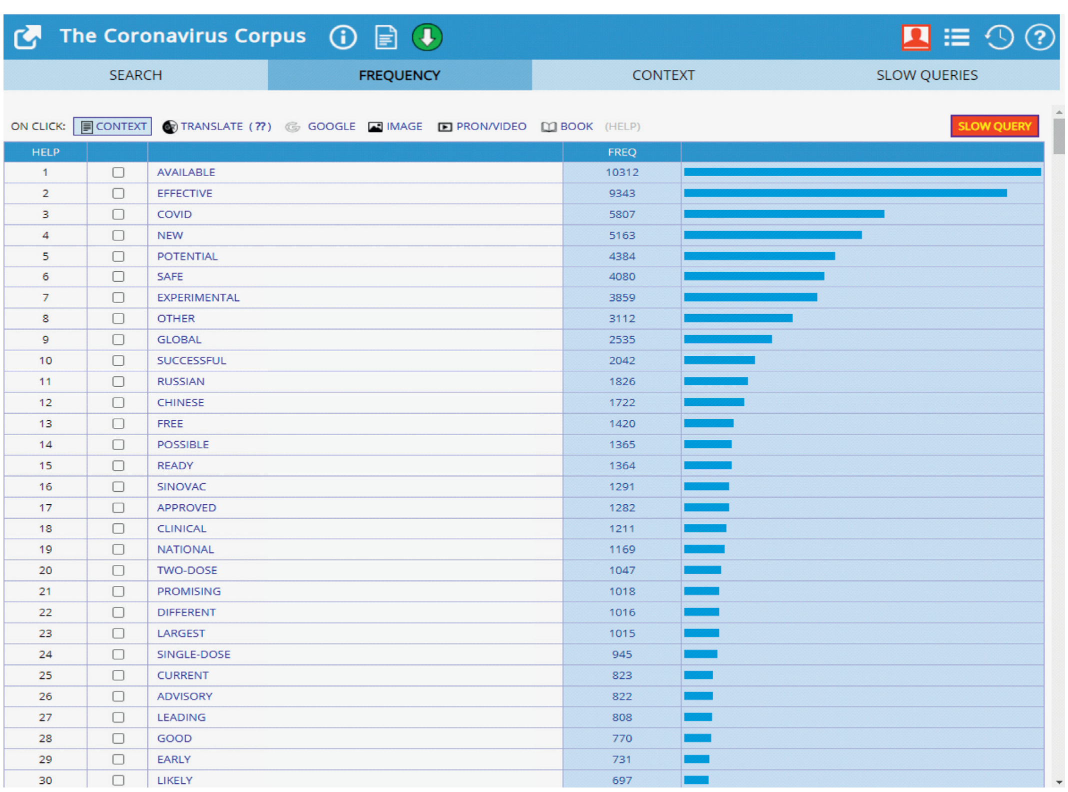
SECTION	ALL	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
FREQ	2319524	5127	6668	6437	9449	9946	11697	27480	32246	29867	49198	529559	1503509
WORDS (M)	14400	244.1	304.8	371.3	401.5	429.4	512.5	1,531.3	1,746.5	1,569.1	1,987.5	2,607.8	2,449.2
PER MIL	161.08	21.00	21.88	17.34	23.54	23.16	22.83	17.95	18.46	19.03	24.75	203.06	613.87

Figure 1: The frequencies of "VACCINE" in NOW Corpus (news on the web) [30].



SECTION	ALL	20-01	20-02	20-03	20-04	20-05	20-06	20-07	20-08	20-09	20-10	20-11	20-12	21-01	21-02	21-03	21-04
FREQ	1395905	2212	3615	10490	18001	25632	19076	30419	35462	37790	27369	52425	108073	128527	93440	120829	112367
WORDS (M)	1370000	7.3	14.5	100.0	108.0	97.8	83.3	78.4	74.2	57.6	57.1	49.3	50.8	50.6	45.2	58.3	51.3
PER MIL	1.02	301.35	249.41	104.91	166.72	262.05	229.07	387.93	478.17	655.95	479.02	1,063.74	2,127.34	2,540.61	2,067.66	2,073.01	2,190.33

Figure 2: The frequencies of "VACCINE" in The Coronavirus Corpus [31].



HELP			FREQ
1	☐	AVAILABLE	10312
2	☐	EFFECTIVE	9343
3	☐	COVID	5807
4	☐	NEW	5163
5	☐	POTENTIAL	4384
6	☐	SAFE	4080
7	☐	EXPERIMENTAL	3859
8	☐	OTHER	3112
9	☐	GLOBAL	2535
10	☐	SUCCESSFUL	2042
11	☐	RUSSIAN	1826
12	☐	CHINESE	1722
13	☐	FREE	1420
14	☐	POSSIBLE	1365
15	☐	READY	1364
16	☐	SINOVAC	1291
17	☐	APPROVED	1282
18	☐	CLINICAL	1211
19	☐	NATIONAL	1169
20	☐	TWO-DOSE	1047
21	☐	PROMISING	1018
22	☐	DIFFERENT	1016
23	☐	LARGEST	1015
24	☐	SINGLE-DOSE	945
25	☐	CURRENT	823
26	☐	ADVISORY	822
27	☐	LEADING	808
28	☐	GOOD	770
29	☐	EARLY	731
30	☐	LIKELY	697

Figure 3: The collocation list in The Coronavirus Corpus [31].

Table 1: Positive collocation of the word *vaccine*.

Word	Rank	Freq
Available	1	10312
Effective	2	9343
Safe	6	4080
Successful	10	2042
Promising	21	1018
Leading	27	808
Good	28	770
Viable	34	653
Positive	46	538
Eligible	73	356

All the three examples above illustrate the public's desire for an effective vaccine, but they are slightly different. In the first example, we can see that the speaker wants to warn people not to rely entirely on the development of vaccines but to be prepared for a long battle.

When analyzing the specific context, we can dig out the deeper inner meaning behind the surface phenomena, which is also an important reason why this study combines the method of CDA with that of corpus linguistics so that both the concordance function of corpus tools and more detailed and critical analysis of CDA can be made full use of.

4.2.2. Negative Vaccine Portrayals. Among the top 100, there are only five adjectives with negative ideas and attitudes (Table 2). However, it is noteworthy that the use of "*potential*," which ranks fifth, indicates that people still pay critical attention to the potential risks of vaccines. In addition, the use of words such as "*so-called*" and "*adverse*" reflects a high degree of public distrust of vaccines, especially in terms of the effect and quality of vaccines. This point is also well confirmed in the following specific context.

(1) # However, one controversial part of this potential "vaccine bubble" is that people will feel like their freedom will still be restricted unless they are part of the "bubble." #

(2) # In recent days, China has vowed to deploy a potential vaccine as a "global public good" that would be accessible and affordable. Toward that end, the World Health Organization also laid out principles to encourage collaboration and information sharing on a COVID-19 treatment. #

In the analysis of adjectives indicating negative emotions, the word *potential* is taken as an example. In one context, the word "*potential" is used together with words such as* risk or threat. In another context, the word "*potential*" suggests that vaccines will be developed in the future, and China will use vaccines for global health. The difference between these two examples lies in the use of the potential for two different meanings of the word. That is to say, the contextual results of the corpus will not distinguish the meanings, and at this time, we need to make semantic division manually to obtain more precise and accurate meanings.

Table 2: Negative collocation of the word "vaccine."

Word	Rank	Freq
Potential	5	4384
Mandatory	53	456
So-called	61	410
Adverse	84	327
Fake	90	315

4.2.3. Neutral Vaccine Portrayals. Most adjectives do not have a clear tendency of emotion; thus, they are classified as neutral (Table 3). These neutral adjectives are divided into four categories (Table 4).

The first category can be labeled as "region, country, and organization." Countries on the top of the list are "*Chinese*" and "*Russian*." In fact, China and Russia have been among the fastest in the world in terms of vaccine research and development, and their progress has attracted the attention of many countries. It is worth noting that SINOVAC, the major vaccine research company in China, is mentioned frequently. Also, Global Alliance for Vaccines and Immunization (GAVI) is the only organization in the top 100. GAVI is a crucial public-private global health alliance established in 1999 that works with governments and nongovernmental organizations to promote global health and immunization. During the period selected by this study, it actively updated the latest vaccine research situation on their website and encouraged global vaccination. GAVI is coleading COVAX raised by WHO, the vaccines pillar of the Access to COVID-19 Tools (ACT) accelerator. It has established a global risk-sharing mechanism for pooled procurement and equitable distribution of COVID-19 vaccines. COVAX has transported 53 million COVID-19 vaccines to 121 participants so far and has played an important role in the transportation and distribution of vaccines around the world.

The second category is labeled as "medical terms," including information about testing and production of vaccines and some of the words that are specific to the medical field of vaccines. This shows that the media consciously used medical terms and principles related to vaccines to inform the public of the basic information about vaccines so that the public can better understand the working mechanism of vaccines.

The third category is labeled as "development process," mainly including adjectives related to the research and development process and time points of the vaccine. This indicates that the media has been paying close attention to the research and development process of vaccines all the time. This also fully reflects the urgent need for effective vaccines and the expectation that vaccine development can achieve a breakthrough.

5. Conclusions and Discussions

5.1. Conclusions. Research questions put forward here are as follows:

Table 3: Neutral collocation of the word "vaccine."

Word	Rank	Freq
Covid	3	5807
New	4	5163
Experimental	7	3859
Other	8	3112
Global	9	2535
Russian	11	1826
Chinese	12	1722
Free	13	1420
Possible	14	1365
Ready	15	1364
SINOVAC	16	1291
Approved	17	1282
Clinical	18	1211
National	19	1169
Two-dose	20	1047
Different	22	1016
Largest	23	1015
Single-dose	24	945
Current	25	823
Advisory	26	822
Early	29	731
Likely	30	697
Additional	31	681
Inactivated	32	680
Single	33	653
Indigenous	35	653
Federal	36	598
Oral	37	591
Slow	39	577
Only	40	575
International	41	570
Future	42	568
Novel	43	551
Local	44	550
One-shot	45	546
One-dose	51	461
Similar	52	459
Major	54	446
Eventual	55	440
Viral	56	435
Universal	57	435
Biggest	58	434
Investigational	59	412
Latest	60	412
Particular	62	407
Developed	63	407
Widespread	64	407
Domestic	65	403
Mass	66	396
Two-shot	67	393
Pandemic	68	392
High	69	390
Multiple	70	387
Live	71	386
Ongoing	72	360
Indian	74	355
Rapid	75	353
Public	76	352
Coronvavac	77	352
Existing	78	348
Therapeutic	79	341
Actual	80	340
Specific	81	340
Recent	82	334
Equitable	83	331
Seasonal	85	327
Top	86	320
Proven	87	319
Developing	88	319
Sure	89	318
Recombinant	91	312
Pneumococcal	92	312
Various	93	309
Working	94	298
GAVI	95	298
Real	96	289
Important	97	288
African	98	280
Late	99	278

(1) What are the most common adjective collocations for "vaccines"?

(2) How are "vaccines" portrayed contextually in online discourses?

(3) What is the general picture of the public attitude towards Coronavirus vaccines?

Next, these questions will be answered based on the analysis results obtained in this study.

5.1.1. Adjective Collocations for "Vaccines". To answer the first two research questions, a total of 447 adjectives were included, and the lowest frequency of use was more than 40 times. The classified percentage of these adjectives is shown in Table 5.

Among the 447 adjectives, the number of positive adjectives was about twice as many as the number of negative adjectives. Therefore, it can be concluded that people are more inclined to express positive and hopeful emotions when expressing their opinions on COVID-19 vaccines. Many positive expressions have appeared in reports describing the latest progress of the Coronavirus vaccines, which is bound to play an optimistic role in the fight against the Coronavirus. At the same time, the study found that many of the positive adjectives found their antonyms in the negative adjectives, indicating that there are different views on the same issue among the public.

5.1.2. Public Attitude towards Vaccine. The third research question, "What is the general picture of the public attitude towards Coronavirus vaccines?" can be answered based on the results of collocation analysis. Public concerns about vaccines fall into three categories.

The first is the globalization of vaccines. As the first category of adjectives mentioned in Section 4, among the 447 adjectives, frequent occurrences of *global, national*

Table 4: The classification of neutral adjectives.

1	Region, country, organization	Global, Russian, Chinese, SINOVAC, national, indigenous, federal, international, local, universal, domestic, Indian, GAVI, African
2	Medical terms	Covid, two-dose, single-dose, inactivated, oral, one-dose, one-shot, two-shot, viral, live, coronaVac, therapeutic, recombinant, pneumococcal
3	Development process	New, experimental, ready, current, early, slow, future, eventual, latest, ongoing, working, rapid, recent, seasonal, developing, late, clinical
4	Descriptions	Other, free, possible, approved, different, largest, advisory, likely, additional, limited, only, similar, major, biggest, investigational widespread, mass, pandemic, mass, high, multiple, public, existing, actual, specific, equitable, proven, top, sure, various, real, important

Table 5: Adjective collocations for "vaccines."

Positive	41	9.1%
Negative	22	4.9%
Neutral	384	85.9%
Total	447	100%

worldwide, nationwide rank high in the list, indicating that people are concerned not only about the development of vaccines in their own country but also about the development of vaccines in the world. In the face of a huge crisis like COVID-19, all human beings should fight together. Here is an example in this sense.

Equally clear is that the drive for *a global COVID-19 vaccine regimen and the global surveillance grid* are moving ahead in conPcert to transform the world as we know it–if we allow it to happen. # Professor Michel Chossudovsky said in his recently republished timeline' COVID-19 Coronavirus "Fake" Pandemic.

The second category is about the equal distribution of vaccines. Such adjectives as *equitable, equal, and fair* frequently appeared in the context mostly reflect people's worry about whether everyone can be immunized. As a special public health measure, the distribution of vaccination can be a test to the governments in different countries and to WHO, and the improper sequence of vaccination may even lead to social disorder. The following example indicates concern of this type.

Climate activist Greta Thunberg says governments, vaccine developers, and the international community must "step up their game" to fight global vaccine inequity. The Swedish teen who inspired the Fridays for Future environmental movement cites estimates that one in four people in high-income countries have received coronavirus vaccines, compared with one in 500 in the middle- and lower-income countries. # Thunberg says it is "*completely unethical* that high-income countries are now vaccinating young and healthy people if that happens at the expense of people in risk groups and on the front lines in low- and middle-income countries." #

Finally, most people pay attention to the effect of vaccines. A group of frequently used adjectives is words like *latest, novel, update,* which express concerns about the latest situation and progress. After studying the context of these adjectives carefully, it is not hard to find that discourses with these words are mainly about whether the vaccine developing now can resist the attack from the variant of the virus. In fact, this is a big problem we are facing. The Coronavirus is mutating at a rapid rate that even vaccinated people may be infected again. The timeliness of vaccines has become one of the concerns of the public. The following statement is an example of this type.

" The problem with a novel virus is it has got so much room to grow and shift and change to optimize itself, " said Harvard School of Public Health's Dr. Michael Mina. " The question is how quickly is it going to keep updating itself " " Whatever mutation happens, the vaccine manufacturers can keep up with it, but given the multiplicity of strains circulating at one time, that may be a challenge, " said Dr. Schaffner. # Dr. Mina, from Harvard, said he is less confident that the existing vaccines can be updated for any and all future variants.

5.2. Discussions. This study mainly studied the public's attitude towards COVID-19 vaccines and their focus of attention. It confirms the positive portrayal of those who choose to be vaccinated. In addition, this study analyzed the neutral adjectives and divided them into four categories. Such classification can help policymakers and healthcare practitioners to understand people's concerns about vaccines and help the media to understand what people most want to know. Based on these findings, it is recommended that governments and public health agencies improve their strategies to better communicate the benefits of vaccination to the public, i.e., to increase vaccine confidence. Such strategies could include methods like increasing transparency of data and introducing data from scientific studies to make vaccines more credible. For example, Chinese media has constantly reported the development of the COVID-19 vaccines in China and around the world at a very early stage, and there are a lot of discussions about this issue on social media platforms like Sina Weibo (a blogging platform where people leave comments and express attitudes) [25]. Moreover, public attitudes are influenced by many factors. Thus, it is suggested that strategies should be adopted to eliminate negative emotions of the public on vaccination and improve positive feelings at the same time [35]. A strategy called "prosocial motivation" proved effective in making people take preventive measures by Jordan et al. [36] and Heffner et al. [37] belong to this category, which emphasizes how vaccination will protect the community and will bring life back to a state where everyone could be intimately connected. This position is supported by the study of Lyu et al. [38], which indicates that improving the public's pandemic experience and increasing their sentiment scores can promote their acceptance of the COVID-19 vaccines.

This study is limited to examining data from January 2020 to April 2021. With the continuous development of the epidemic, human beings have entered the postvaccine era, with various virus variants emerging in an endless stream. It is believed that people's attitudes towards vaccines will also continue to evolve. At the same time, this corpus only contains data from a few countries, so a corpus composed of international newspapers and media will also be a valuable supplement to the study.

Now, the practice of vaccination has entered a new stage, and the governments of all countries are vigorously advocating and encouraging people to be vaccinated against COVID-19. Although China adheres to the principle of voluntary vaccination, people's enthusiasm about vaccination is very high. However, at the same time, there are still people who refuse to be exposed to the COVID-19 vaccine because of concerns about the safety of the vaccine. Lin et al. [40] investigate and summarize the steps taken by the Chinese government to control the spread of COVID-19 and reopen lockdown cities, emphasizing the effect of awareness diffusion in this process. Therefore, further research can be conducted to probe into factors underlining different attitudes and ways to promote people's awareness so that both vaccines and vaccination policies can be improved accordingly.

Data Availability

The data analyzed in this study were taken from Coronavirus Corpus from English Corpora (https://www.english-corpora.org/).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] J. Zhao, Z. Ahmad, S. K. Khosa, M. Yusuf, O. A. Alamri, and M. S. Mohamed, "The Role of Technology in COVID-19 Pandemic Management and its Financial Impact," *Complexity*, vol. 2021, Article ID 4860704, 12 pages, 2021.

[2] T. A. V. Dijk, *News as Discourse*, University of Groningen, Groningen, Netherlands, 1988.

[3] V. Raghupathi, J. Ren, and W. Raghupathi, "Studying public perception about vaccination: a sentiment analysis of tweets," *International Journal of Environmental Research and Public Health*, vol. 17, no. 10, p. 3464, 2020.

[4] Centers for Disease Control and Prevention (Cdc), "Ten great public health achievements--United States, 1900-1999. MMWR," *Morbidity & Mortality Weekly Report*, vol. 48, no. 12, pp. 241–243, 1999.

[5] H. Markel, "Taking Shots: The Modern Miracle of Vaccines. Medscapem," 2004, https://www.medscape.com/viewarticle/481059.

[6] A. M. Stern and H. Markel, "The history of vaccines and immunization: familiar patterns, new challenges," *Health Affairs*, vol. 24, no. 3, pp. 611–621, 2005.

[7] A. Kennedy, K. LaVail, G. Nowak, M. Basket, and S. Landry, "Confidence about vaccines in the United States: understanding parents' perceptions," *Health Affairs*, vol. 30, no. 6, pp. 151–159, 2011.

[8] T. Baroiant, "Combining Critical Discourse Analysis and Corpus Linguistics to Examine the Term "vaccine" in Online News media," Unpublished master's thesis), San Francisco State University, San Francisco, USA, 2015.

[9] R. A. Goodman, R. E. Hoffman, W. Lopez, G. W. Matthews, M. A. Rothstein, and K. L. Foster, Eds., *Law in Public Health Practice*, Oxford University Press, Oxford, UK, 2007.

[10] "The History of Vaccines. (n.d.)," 2020, http://www.historyofvaccines.org/content/articles/history-anti-vaccination-movements.

[11] R. D. Johnston, *The Politics of Healing: Histories of Alternative Medicine in Twentieth-century North America*, Psychology Press, Hove, East Sussex, UK, 2004.

[12] T. A. Van Dijk, *News as Discourse*, University of Groningen, Groningen, Netherlands, 1988.

[13] M. A. K. Halliday, "Modes of meaning and modes of saying: types of grammatical structure and their determination by different semantic functions," in *Context in Linguistic Analysis. Essays Offered to William Haas*, D. J. Allerton, E. Carney, and D. F. Holdcroft, Eds., Cambridge University Press, Cambridge, England, 1979.

[14] A. Atkins, "Critical Discourse Analysis: A Letter to Expatriates from the Rt. Hon. Sir Norman Fowler MP," 2002, http://www.cels.bham.ac.uk/resources/essays/Andrew%20Atkins%20Assignment%205.pdf.

[15] M. A. Kandil, *The Israeli-Palestinian Conflict in American, Arab, and British media:Corpus-based critical discourse analysis*, Georgia State University, Atlanta, Georgia, USA, 2009.

[16] S. Hunston, *Corpora in Applied Linguistics*, Cambridge University Press, Cambridge, UK, 2002.

[17] M. Stubbs, *Text and Corpus Analysis: Computer-Assisted Studies of Language and Culture*, Blackwell, Oxford, UK, 1996.

[18] M. Bednarek, *Evaluation in media Discourse: Analysis of a Newspaper Corpus*, A&C Black, London, UK, 2006.

[19] M. Bednarek and H. Caple, "Why do news values matter? Towards a new methodological framework for analyzing news discourse in Critical Discourse Analysis and beyond," *Discourse &Society*, vol. 25, 2014.

[20] V. Katermina and E. Yachenko, "Axiology of COVID-19 as a linguistic phenomenon in English mass media discourse," *Advances in Journalism and Communication*, vol. 8, no. 02, p. 59, 2020.

[21] Y. Kim, "On media coverage of the COVID-19 outbreak: a corpus-based collocation study," *SNU Working Papers in English Language and Linguistics*, vol. 17, 2020.

[22] M. Awad AlAfnan, "COVID 19-The foreign virus: media bias, ideology and dominance in Chinese and American newspaper articles," *International Journal of Applied Linguistics & English Literature*, vol. 9, no. 1, pp. 56–60, 2020.

[23] M. Azizan, H. H. Ismail, and S. N. Qaiwer, "Power and solidarity in positive Facebook postings amidst covid-19 in Malaysia," *Journal of Nusantara Studies (JONUS)*, vol. 5, no. 2, pp. 329–364, 2020.

[24] S. Das and A. Dutta, "Characterizing public emotions and sentiments in COVID-19 environment: a case study of India," *Journal of Human Behavior in the Social Environment*, pp. 1–14, 2020.

[25] H. Gao, D. Guo, J. Wu, Q. Zhao, and L. Li, "Changes of China's public attitudes to domestic COVID-19 vaccination after the vaccines approved: a semantic network and

sentiment analysis based on Sina Weibo texts," *Frontiers in Public Health*, p. 1677, 2021.

[26] A. Wisniak, H. Baysson, N. Pullen et al., "COVID-19 Vaccination Acceptance in the canton of Geneva: A Cross-Sectional Population-Based Study," *medRxiv*, vol. 151, 2021.

[27] K. H. Nguyen, D. Yankey, P. J. Lu et al., "Report of health care provider recommendation for COVID-19 vaccination among adults, by recipient COVID-19 vaccination status and attitudes—United States, april–september 2021," *Morbidity and Mortality Weekly Report*, vol. 70, no. 50, p. 1723, 2021.

[28] S. Bell, R. M. Clarke, S. A. Ismail et al., "COVID-19 vaccination beliefs, attitudes, and behaviours among health and social care workers in the UK: a mixed-methods study," *PLoS One*, vol. 17, no. 1, Article ID e0260949, 2022.

[29] D. Gablasova, V. Brezina, and T. McEnery, "Collocations in corpus-based language learning research: identifying, comparing, and interpreting the evidence," *Language and Learning*, vol. 67, no. 1, pp. 155–179, 2017.

[30] M. Davies, "Corpus of news on the Web (NOW)," 2016, https://www.english-corpora.org/now/.

[31] M. Davies, "The coronavirus corpus," 2019, https://www.english-corpora.org/corona/.

[32] Xinhuanet, "China approves first self-developed COVID-19 vaccine. Xinhuanet," 2020, https://www.xinhuanet.com/english/2020-12/31/c_139632053.htm/.

[33] FDA, "FDA approves first COVID-19 vaccine," 2021, https://www.fda.gov/news-events/press-announcements/fda-approves-first-covid-19-vaccine.

[34] G. Barnbrook, *Language and Computers*, Edinburgh University Press, Edinburgh, Scotland, 1996.

[35] W. Y. S. Chou and A. Budenz, "Considering emotion in COVID-19 vaccine communication: addressing vaccine hesitancy and fostering vaccine confidence," *Health Communication*, vol. 35, no. 14, pp. 1718–1722, 2020.

[36] J. J. Jordan, E. Yoeli, and D. G. Rand, "Don't get it or don't spread it: comparing self-interested versus prosocial motivations for COVID-19 prevention behaviors," *Scientific Reports*, vol. 11, no. 1, pp. 1–17, 2021.

[37] J. Heffner, M. L. Vives, and O. FeldmanHall, "Emotional responses to prosocial messages increase willingness to self-isolate during the COVID-19 pandemic," *Personality and Individual Differences*, vol. 170, Article ID 110420, 2021.

[38] H. Lyu, J. Wang, W. Wu et al., "Social media study of public opinions on potential COVID-19 vaccines: informing dissent, disparities, and dissemination," *Intelligent medicine*, 2021.

[39] S. F. Chung, "A corpus-based study of SARS in English news reporting in Malaysia and in the United Kingdom," *International Review of Pragmatics*, vol. 3, no. 2, pp. 270–293, 2011.

[40] L. Li, Y. Fan, A. Zeng, and Z. Di, "Understanding the Anticontagion Process and Reopening of China during COVID-19 via Coevolution Network of Epidemic and Awareness," *Complexity*, vol. 2021, Article ID 6623427, 11 pages, 2021.

Hindawi
Complexity
Volume 2022, Article ID 7458978, 14 pages
https://doi.org/10.1155/2022/7458978

Research Article

The Influence Mechanism of Different Cash Flow Availability on R&D Investment: Evidence from China

Xiaobo Wu,[1] Ye Hua,[2] and Hao Lu [3]

[1]*School of Management, Zijingang Campus, Zhejiang University, Hangzhou 310000, Zhejiang, China*
[2]*School of Management, Zhejiang University of Technology, Hangzhou 310014, China*
[3]*Business School, Jiangning Campus, Hohai University, Nanjing 210000, Jiangsu, China*

Correspondence should be addressed to Hao Lu; haolu@hhu.edu.cn

Received 28 October 2021; Revised 3 January 2022; Accepted 10 January 2022; Published 8 March 2022

Academic Editor: Qingyuan Zhu

It has been widely assumed and proven that a firm's research and development (R&D) investment is limited if the availability of cash flow is constrained. The purpose of this study is to verify the opposite proposition: whether firms invest heavily in R&D when more cash flow is available. This paper discusses the heterogeneous relationship between cash flow from different sources and the R&D investment of firms. The study divides the firm's cash flow into three categories according to the business activities that generate firm finance: cash flow from operating activities (CFO), cash flow from investing activities (CFI), and cash flow from financing activities (CFF). On this basis, a dynamic R&D investment model is constructed, and the relevant data for Chinese listed firms are used for the empirical research. The study finds that Chinese listed firms do not necessarily spend more cash on R&D investment, even if the availability of cash flow is not constrained. For young firms, CFO and CFI do not significantly correlate with R&D investment, and CFF significantly negatively correlates with R&D investment. For mature firms, the correlations between cash flow generated from different activities and R&D investment are nonsignificant.

1. Introduction

R&D is a vital input for innovation and thus can increase the competitiveness of firms and promote the economic development of countries and regions [1]. Studies have shown that R&D investment can be constrained by the availability of cash flow [2]. If such financing constraints are sufficiently severe and broad, a country's economic growth will suffer [3]. Based on these studies, an immediate question regarding the promotion of innovation is the following: do firms with sufficient cash flow necessarily have a high level of R&D investment? Little empirical research exists to indicate whether this proposition is valid. Therefore, concerning cash flow from different business activities, it is necessary to study whether firms spend more on R&D investment and what the relationship might be between different sources of cash flow and innovative inputs when adequate capital is available (i.e., the availability of cash flow is less constrained).

This study focuses on cash flow from firms' operating, investing, and financing activities and how changes in cash flow cause changes in firm R&D investment. The previous literature on cash flow and R&D investment has examined various factors affecting firm investments, such as corporate governance level, firm political background, industry, firm nature, financial leverage, financing ability, and investment opportunities [2]. As the development of capital markets steadily improves, sources of cash flow become increasingly abundant, and it is vital to explore the heterogeneous influences of different cash flow sources on firm R&D investment. However, this has rarely been addressed in the previous literature. To address this deficiency, the study explores the sources of cash flow. According to the Chinese Generally Accepted Accounting Principles (GAAP), firm cash flow can be categorised as cash flow from operating activities (CFO), cash flow from investing activities (CFI), and cash flow from financing activities (CFF) [4], which links business activities with R&D investment. This article explores cash flow sources that help potential investors and shareholders of firms improve their understanding of the composition of R&D investment and provides recommendations for choosing a firm's innovation projects and cash flow sources.

According to industrial organisation theory, in the developed market environment of Europe and the U.S., an implicit assumption often underpins the study of firm R&D investment behaviour, namely, the market size remains unchanging [5]. This is markedly different from the situation in emerging economies, as represented by China, where the market has continued to expand significantly over the long term. In emerging economies, where market demand is expanding, R&D and investment behaviour are taking on new characteristics [6]. In the past, industrial organisation approaches that studied the mature oligopolistic markets of Europe and U.S. found that mature firms had advantages in R&D investment competition and usually carried out R&D investment. This can be explained bearing in mind that compared with young firms, mature firms have more money for investments and enjoy cost advantages in R&D [7]. In emerging economies, however, the monopolistic advantage of mature firms has failed, giving young firms room for development and resulting in changes in their R&D and investment strategies. Due to the impact of market demand expansion on firm R&D investment, young firms can largely focus on R&D behaviour and innovation investment to improve their competitive advantage because they do not have economies of scale. In contrast, mature firms use the advantages of scale for short-term profit-making investment [8]. Therefore, it is improper to use existing industrial organisation theory to analyse firm R&D behaviour for the economic development of emerging economies, such as China. The decision-making mechanisms and competition behaviours of firms with different levels of maturity, from the perspective of market demand expansion, are new problems for industrial organisation theory. Hence, this study has great significance for the R&D investment and cash flow decision making of firms with different maturity levels in emerging economies, such as China.

This study is based on the data for Chinese listed firms over the period 2010–2017, covering 18 industries, including manufacturing, construction, mining, and education. CFO, CFI, and CFF are used to measure the financial position of firms, and firms' innovation investment is measured by their R&D investment [3]. We adopt a variant of the dynamic investment model developed by Brown et al. [9] to analyse the impact of the cash flows from the three types of business activities on R&D investment.

This study makes several contributions to the literature on cash flow sources and R&D investment. First, it puts forward a new perspective on financing availability for innovative investment. Previous literature on innovation financing has argued that firms do not invest in innovation if they are underfunded [10]. By analysing the relationship between different cash flow sources and R&D investment, this study finds that even if the availability of cash flow is less constrained, Chinese listed firms do not necessarily spend more money on R&D. The analysis of the financing activities shows that Chinese listed firms rarely support their innovative investment with financing activities.

Second, this paper clarifies the influencing mechanism between different cash flow sources and innovative inputs, which is the first attempt by related research. Since innovation can enhance firm competitiveness, firms are urged to raise funds from multiple channels for R&D investment. It is of great significance to analyse how the cash flow from different financing channels affects R&D investment differently.

Finally, this paper explains the heterogeneous influence of firm cash flow sources with different maturities on R&D investment. Since firms with different maturities differ greatly in their ability to withstand financial risks and bankruptcy, there are obviously significant differences in the composition of cash flow sources and in decision-making attitudes towards R&D investment. The previous research seldom considers this. The conclusion of the study may guide the strategic decision making of firms with different levels of maturity in their cash flow sourcing and R&D investment.

1.1. Literature Review and Hypotheses. Many previous studies have shown that R&D investments can be constrained by the availability of cash flow. For example, Brown et al. [9] documented that R&D investments are constrained by the internal and external financing availability of young high-tech firms that are publicly traded in U.S. Sasidharan et al. [11] reported a positive impact of internal cash flow on R&D and a nonsignificant relationship between external funds and R&D for Indian manufacturing firms. Brown [3] argued that firms with financing constraints tend to have lower levels of R&D investment than firms without financing constraints. Guariglia and Liu [10] illustrated that the difference between the cost of obtaining funding within a firm and the cost of obtaining external funding constrained the innovation activities of the firm. Based on the findings, these studies have identified a necessity to establish policies to reduce financing constraints to improve innovation [3, 10].

The existing studies on financing constraints for innovation have often made the assumption that firms underinvest in innovation because they lack the necessary money and that when firms obtain more money, they will use it for innovation purposes. Some studies have used dynamic investment models based on the Euler equation to analyse the impact of financing availability on innovation [12]. Ahiadorme et al. [13] identified three possible scenarios for such a model: the firm has sufficient funds for investment; the firm does not have sufficient funds for investment and cannot obtain funds for investment through equity issuance; and the firm does not have sufficient funds but can obtain funds for investment through equity issuance. However, will firms with sufficient (funds from different sources of) cash flow really invest more in R&D? This possibility has rarely been considered in the previous literature.

This paper discusses the impact of financing availability on R&D investment from the perspective of CFO, CFI, and CFF. The status and proportion of these three cash flows for a firm is an important reflection of whether the firm can make continuous R&D investment. In China and other emerging economics whose capital markets are imperfect, a large number of firms experience a lack of CFO and CFF and cannot meet the requirement for investing activities (e.g., continuous R&D investment) [14]. However, even if these funds are abundant, corporate managers and investors may still prefer to use cash for projects that are more profitable in the short term rather than to support R&D investment. This is mainly due to the characteristics of R&D investment: high adjustment costs and high uncertainty.

First, the adjustment costs of R&D investment are very high. Bernstein and Mohnen [15] argued that a great amount of adjustment costs exists in the innovation-intensive sectors of developed countries (i.e., U.S. and Japan). Similarly, Guariglia and Liu [10] stated that the adjustment costs of innovation activities in emerging economies, such as China, are also considerable. Unlike ordinary investments, R&D is designed to acquire technological innovation and intellectual property rights, which are intangible assets. In cases where the product has not yet been developed or has not been patented, the "value" of R&D is actually rooted in the human capital of the developers [16]. A significant portion of R&D investment is spent on the hiring, firing, and training costs of R&D personnel (Hall & Lerner, 2010). Once the original R&D employees leave a firm, the training costs spent on these employees by the firm become sunk costs. Moreover, if the employees are rehired by the firm's competitors, the firm faces the risk of its R&D secrets being acquired by competitors [17].

Second, R&D has high uncertainty. Patel and Chrisman [18] showed that no more than 25% of R&D projects in the development phase are successful [18]. According to Guariglia and Liu [10], uncertainty can be driven by irreversibility, long-term returns of R&D activities, and market factors. According to Lee et al. [19], the funds spent on R&D activities are irreversible, and the abrupt halt of R&D activities may lead to technological backwardness and value conversion. In addition, R&D activities are by no means short-term. Not only does the R&D phase take a long time but also the subsequent transformation and product development phases also involve long cycles. In addition, the feasibility of R&D content and the estimation of R&D investment are subject to factors such as market demand and competition [20].

Based on the above inferences, firms have difficulty investing their funds in R&D activities even if their funds are sufficient. The following hypothesis is developed.

H1: when CFO, CFI, and CFF are readily available, it remains difficult for firms to invest in R&D.

Due to the immature capital markets in emerging economies (e.g., China) as well as the existence of information asymmetry, agency problems, and transaction costs, there is a significant cost difference between internal finance and external finance that cannot be completely compensated [21]. According to pecking order theory, because firms are free to control internal funds for R&D investment, internal financing is preferred to external financing [9]. Under actual market conditions, because information asymmetry exists between the firm seeking financing and the external investor, the information-deficit party assumes the valuation risk. External creditors' rights financing and equity financing require the firm seeking funding to bear the high underwriting, audit, legal, regulatory, and other direct or indirect costs, resulting in high external financing costs, which are not conducive to sustainable R&D investment. Firms, on the other hand, can use their own internal reserves of cash for R&D without incurring additional costs, which is a relatively low-cost option. According to financing constraint theory, firms with low availability of cash flow can use internal finance as a buffer for R&D investment to reduce external financing difficulties [22]. Hence, internal cash flow can be more available than external cash flow to finance R&D investment [9].

What is the link between internal and external finance and cash flow from the three different sources? According to the GAAP, CFO mainly includes cash and cash-equivalent inflows and outflows from firm operating activities, such as manufacturing and selling goods or offering services. CFO belongs to internal finance and is the normal source of cash for firms. According to Xu et al. [23], CFO represents capital turnover and improves the ability of the firm to handle risks, which can ensure the smoothing of R&D investment. CFI mainly includes cash inflows and outflows from buying machinery and equipment, plants, advanced technology, intangible assets, etc. The portion of CFI derived from financial instruments constitutes external finance, while the capital investment of CFI can be regarded as internal finance. In addition, CFF mainly includes cash and cash-equivalent inflows and outflows from financing activities, such as issuing or taking debt (e.g., bond or loan) or equity (e.g., stock or dividend). CFF comprises external funds. External creditors and investors are only willing to invest in investments that are considered to be financially sound and, hence, are reluctant to invest in R&D activities, which have the characteristics of high uncertainty [23]. Based on the statements, CFO can be the most accessible source of finance for firm R&D investment, and CFF can be the least accessible source of finance for firm R&D investment. That is, R&D investment can be least constrained by CFO and most constrained by CFF. Therefore, hypothesis 2 is presented as follows.

H2: among the three types of firm financing sources, R&D investment is influenced most by CFF availability and then by CFI availability, and it is least influenced by CFO availability.

According to previous studies, there are significant differences in the sensitivity of cash flow sources for R&D investment for firms with different levels of maturity [9, 24]. Compared with young firms, mature firms with scale advantages have stronger R&D financing capacity based on industrial organisation theory [3, 10]. However, firms with a higher level of maturity are less likely to invest in R&D [25]. In contrast, young firms can be more willing to engage in R&D activities, especially young firms with high technical requirements [9].

Following industrial organisation theory, an implicit assumption often underpins the study of firm R&D investment in developed markets, such as Europe and the U.S.: the market size is unchanging [16]. However, the market size in emerging economies, such as China, has continued to expand significantly over the long term. In the case of the dynamic expansion of market demand, the cash flow characteristics of R&D investment will be different from those of unchanging markets. Therefore, we cannot simply use existing industrial organisation theory to analyse the R&D cash flow strategies of Chinese firms with different levels of maturity.

In the previous literature on industrial organisation theory, the mature firms in European and U.S. mature oligopolistic markets are found to have advantages in R&D investment competition (e.g., sufficient funds and cost advantages) [26]. As the market in emerging economies continues to expand, the monopolistic power of mature firms fails, giving young firms room for development and resulting in changes to their R&D investment strategies [27]. In view of the impact of market expansion on firm R&D investment, young firms have no advantage of scale or brand and need to invest in R&D to develop a unique competitive advantage, while mature firms use the advantages of scale to focus on short-term profit-making investment strategies [28]. This is because when the market expands and the demand for products is strong, mature firms can promote efficiency by making simple investments to expand production. In the context of market expansion, although mature firms also invest in R&D, the intensity of their investment is reduced. The expansion of market size contributes most significantly to the R&D investment of young firms [29]. On the whole, in emerging economies, where markets are expanding, mature firms invest mainly to increase their scale and gain market pricing power. Simple investment to expand production can promote efficiency gains without the need for high-risk R&D investment activities. However, young firms mainly invest in R&D to produce innovative products and develop a unique competitive advantage [30]. Therefore, in emerging economies where market size is expanding, the impact of cash flow sources on R&D investment can vary greatly for firms with different levels of maturity. Hypothesis 3 is presented as follows.

H3: in emerging economies, the R&D investment of young firms is more influenced by CFO, CFI, and CFF than that of mature firms.

1.2. Data Collection

1.2.1. Sample Selection. The sample includes all Chinese firms from the Main Board, Small and Medium Enterprise (SME) Board, and ChiNext Board listed on the Shanghai Stock Exchange and the Shenzhen Stock Exchange. These listed firms cover 18 industries, and the details are presented in Table 1. Since the ChiNext Board was officially launched in the Shenzhen Stock Exchange on 23 October 2009, our data range starts in 2010. Due to the reform of the Chinese accounting system in 2018, the data range ends in 2017 to ensure the consistency of the accounting-related measures (e.g., cash flow and R&D). For the source of the dataset, we obtain the majority of statistics from Wind and acquire the disclosed information on financing activities from China Stock Market & Accounting Research Database (CSMAR).

Following the previous literature [31, 32], we require a firm to have positive assets and sales to be included in the sample. Firms with no R&D observations are excluded. In addition, to reduce potential bias from the outliers, we trim all the variables at the first and ninety-ninth percentiles; thus, the final sample is an unbalanced panel of 2,401 firms and 12,245 observations. In addition, we differentiate between young and mature firms in the analysis. Following Brown et al. [9] and Gkypali et al. [33], firms with less than 15 years of listed experience are categorised as young firms, and the remaining firms are counted as mature firms [9]. All the variable definitions are described in Table 2.

1.2.2. Descriptive Statistics. The descriptive statistics for the variables are shown in Table 3. The first column reports the information for the full sample. The mean rd_t is 0.024, which is smaller than the mean value of the R&D-to-assets ratio for U.S. industrial firms (0.043) reported by He and Wintoki [31]. The mean cfo_t is 0.047, which is equal to the mean cff_t (0.047). This suggests that CFO and CFF can both be important finance sources for firms. The mean cfi_t is −0.081, showing that investing cash outflows are always greater than the inflows. In addition, the standard deviation of rd_t (0.023) is much smaller than the standard deviations for cfo_t (0.079), cfi_t (0.109), and cff_t (0.144), indicating that R&D spending is relatively stable when cash flow from operating, investing, and financing activities fluctuates.

Columns 2 and 3 present the information for the young firm and mature firm subsamples, respectively. In the last column, we show the standard t-test results, comparing the mean difference between the two subsamples. Since the test results are all significant, it is evident that young firms are significantly different from mature firms in all of the dimensions [31].

1.3. Model

1.3. Model. Based on the previously stated hypotheses, the study modifies a dynamic investment model taken from Tori and Onaran [34] to examine the impacts of different types of cash flow on R&D spending:

$$\begin{aligned}\mathrm{rd}_{i,t} = {} & \beta_1 \mathrm{rd}_{i,t-1} + \beta_2 \mathrm{rd}^2_{i,t-1} + \beta_3 \mathrm{sales}_{i,t-1} + \beta_4 \mathrm{cfo}_{i,t-1} \\ & + \beta_5 \mathrm{cfi}_{i,t-1} + \beta_6 \mathrm{cff}_{i,t-1} + \beta_7 Z_{i,t} + u_t + v_i + \varepsilon_{i,t},\end{aligned} \tag{1}$$

where $\mathrm{rd}_{i,t}$ is the dependent variable, denoting the R&D spending for firm i in period t; $\mathrm{cfo}_{i,t-1}$, $\mathrm{cfi}_{i,t-1}$, and $\mathrm{cff}_{i,t-1}$ denote firm i's net CFO in period t−1, net CFI in period t-1, and net CFF in period t-1, respectively; and sales_{t-1} represents the firms' net sales, denoting the firm's output. All these variables are scaled by total assets. Following the previous studies [3], we include prior R&D expenditure and its quadratic terms in the model because these might influence the relationship between cash flow and R&D. Z is a vector of the control variables, which includes growth opportunities (tobinq_t), the use of cash for R&D smoothing ($\Delta\,\mathrm{cashholdings}_t$), and firm size ($\mathrm{size}_t$). In addition, we establish the firm's fixed effects (v_i) to control for unobserved (time-invariant) firm characteristics and time fixed

Table 1: Industry types.

1	Agriculture, hunting, and forestry
2	Mining
3	Manufacturing
4	Electricity, gas, and water supply
5	Construction
6	Wholesale and retail trades
7	Transportation, warehousing, and postal service
8	Accommodation and catering
9	Information transmission, software, and information technology services
10	Financial
11	Real estate
12	Renting and business services
13	Scientific research and technical services
14	Water conservancy, environment, and public facilities management
15	Education
16	Health and social work
17	Culture, sports, and entertainment
18	Comprehensive

Note. The firms listed on the Shanghai Stock Exchange and the Shenzhen Stock Exchange are from 18 industries in all. The classification of industries follows *Industrial Classification for National Economic Activities* listed on National Bureau of Statistics website (http://www.stats.gov.cn/tjsj/tjbz/).

Table 2: Variable definitions.

Variable	Definition
rd_t	R&D expenses in period t divided by the book value of total assets at the beginning of period t.
ocf_t	Net operating cash flow in period t divided by the book value of total assets at the beginning of period t.
icf_t	Net investing cash flow in period t divided by the book value of total assets at the beginning of period t.
fcf_t	Net financing cash flow in period t divided by the book value of total assets at the beginning of period t.
sale_t	Net sales in period t divided by the book value of total assets at the beginning of period t.
tobinq_t	Natural logarithm of Tobin's Q in period t. Tobin's Q is equal to market value over total assets.
Δ cashholdings_t	Difference between cash and cash equivalents in periods t and t-1 divided by the book value of total assets at the beginning of period t.
size_t	Natural logarithm of the total number of a firm's employees in period t. The alternative measure is equal to natural logarithm of total assets.
Alternative size_t	Natural logarithm of the total assets in period t.
Alternative tobinq_t	Natural logarithm of alternative Tobin's Q in period t. Alternative Tobin's Q is equal to market value over (total assets-net intangible assets-good will).
roa_t	Return-on-assets ratio in period t.
tax_t	Total tax divided by the book value of total assets in period t.
Age_a_t	A variable that takes the value of one when a firm is listed less than 15 years and zero otherwise.
Age_b_t	A variable that takes the value of one when a firm is listed less than 10 years and zero otherwise.
Age_c_t	The number of years that a firm is listed.

effects (u_t) to control for the impact of unobserved time events. $\varepsilon_{i,t}$ is the firm-specific idiosyncratic shock [35].

In this study, the dynamic investment model is used to determine whether the costs of cash flow obtained by firms from different business activities differ. Then, by analysing the sensitivity of R&D investment to different cash flows, the study provides the new support for R&D investment by listed firms in emerging economies, such as China.

We estimate the model using the one-step GMM method instead of two-step estimates because the standard errors from two-step GMM are downward-biased for small samples (e.g., [9]). Since equation (1) is a dynamic model with firm fixed effects, we apply the first-difference GMM [36]. Similar approaches have also been employed in many recent studies, such as Brown and Petersen [9]; Sasaki [37]; and Weng and Soderbom [2]. The results are shown in the next section.

2. Results

2.1. Cash Flow Sources and R&D Investment. Table 4 presents the first-difference GMM coefficient estimates of equation (1) for the full sample of Chinese listed firms. The study begins with a modified dynamic investment model containing the three types of cash flow, and then we add more control variables. All explanatory variables in these regression specifications are treated as potentially endogenous. Referring to Brown and Peterson [9]; Sasaki [37]; and Tori and Onaran [34], we choose lagged values dated t−3 to t−5 as instruments. The p values of the AR(1) and AR(2) statistics are reported. The p values for the AR(1) statistics show first-order autocorrelation in the errors, and the p values for the AR(2) statistics do not reject the null hypothesis of no second-order autocorrelation. In addition, we conduct Hansen tests for the overidentification of the

Table 3: Sample descriptive statistics.

Variables		Full (1)	Young (2)	Mature (3)	Difference (4)
rd_t	Mean	0.024	0.027	0.018	-0.009^{***}
	Median	0.019	0.022	0.011	
	SD	0.023	0.023	0.021	
	N	12245	8702	3543	
cfo_t	Mean	0.047	0.049	0.040	-0.009^{***}
	Median	0.044	0.048	0.037	
	SD	0.079	0.079	0.080	
	N	12117	8628	3489	
cfi_t	Mean	−0.081	−0.093	−0.052	0.041^{***}
	Median	−0.056	−0.067	−0.034	
	SD	0.109	0.112	0.093	
	N	12068	8605	3463	
cff_t	Mean	0.047	0.054	0.028	-0.026^{***}
	Median	0.004	0.008	−0.002	
	SD	0.144	0.149	0.129	
	N	12015	8598	3417	
$sales_t$	Mean	0.708	0.691	0.750	0.059^{***}
	Median	0.579	0.575	0.588	
	SD	0.518	0.483	0.595	
	N	12091	8628	3463	
$tobinq_t$	Mean	0.666	0.690	0.605	-0.085^{***}
	Median	0.579	0.610	0.501	
	SD	0.476	0.467	0.494	
	N	11587	8288	3299	
Δ $cashholdings_t$	Mean	0.012	0.010	0.019	0.009^{***}
	Median	0.002	0.001	0.005	
	SD	0.101	0.103	0.094	
	N	12067	8577	3490	
$size_t$	Mean	7.795	7.650	8.150	0.500^{***}
	Median	7.714	7.570	8.143	
	SD	1.151	1.088	1.222	
	N	12134	8606	3528	

Column 1 is for the full sample, column 2 is for the young firm subsample, and column 3 is for the mature firm subsample. In the last column 4, we show the standard t-test results to compare the mean difference between the two subsamples in columns 2 and 3. The p value of the t-test results are all significant with 3 asterisks, that is, $^{***}p$ significant at the 1.0% level. It is evident that young firms are significantly different from matured firms in all of the dimensions. * Significant at the 1.0% level. ** Significant at the 0.5% level. *** Significant at the 0.1% level.

Table 4: Dynamic R&D regressions for full sample.

	(1)	(2)
rd_{t-1}	0.433^{*}	0.907^{***}
	(0.247)	(0.169)
${rd_{t-1}}^2$	0.007	-3.603^{***}
	(1.850)	(1.398)
cfo_{t-1}	0.022	−0.011
	(0.024)	(0.018)
cfi_{t-1}	0.012	0.001
	(0.011)	(0.010)
cff_{t-1}	-0.022^{**}	-0.025^{***}
	(0.009)	(0.010)
$sales_{t-1}$	-0.011^{***}	−0.002
	(0.004)	(0.003)
$tobinq_t$		−0.003
		(0.002)
Δ $cashholdings_t$		0.019^{*}
		(0.011)
$size_t$		−0.003
		(0.003)
Time dummies	Yes	Yes
Firm dummies	Yes	Yes
AR(1)	0.000	0.000
AR(2)	0.955	0.062
Hansen	0.538	0.137
Obs.	7,176	6,261
Firms	1,898	1,826

Note. Robust standard errors are in parentheses; $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$. Estimation is by first-difference GMM. Time and firm fixed effects are included in both regressions. Lagged values dated from t-3 to t-5 are utilised as instruments for all explanatory variables.

instruments [3]. The tests shows that the validity of the instruments cannot be rejected in Table 4.

In Table 4, the coefficients for rd_{t-1} are significantly positive, indicating the persistence of R&D. However, the coefficients for rd_{t-1} are always smaller than 1; the coefficients for rd^2_{t-1} are often negative; and the coefficients for $sales_t$ are negative but not always significant. Hence, these results reject the null hypothesis of the Euler equation, which could be explained by the financial constraints of the firms in imperfect capital markets [9]. The coefficients for cfo_{t-1} and cfi_{t-1} are always nonsignificant. In addition, the coefficients for cff_{t-1} are always negatively significant.

These results suggest that firms with greater CFO and CFI may not have spent more money on R&D, and firms with greater CFF tend to invest less in innovation. The results align exactly with hypothesis 1. Since R&D investment can have long-term risk, unpredictability, and irreversibility, firm managers and external investors may be reluctant to invest largely in R&D, even if CFO, CFI, and CFF are sufficiently available.

Moreover, the results show that R&D spending is more sensitive to financing cash flow than the other two types of cash flow. That is, an increase in CFO and CFI does not necessarily lead to an increase in firm R&D investment, but an increase in CFF can be significantly associated with a decrease in R&D investment. Hence, hypothesis 2 is partially supported. This can be explained by the high adjustment costs of R&D investment [3, 18], which make firms sustain a relatively stable R&D level instead of changing their R&D significantly with the increase/decrease in CFI and CFO. In addition, it is highly possible that firms that need more CFF are generally firms whose CFO and CFI are insufficient, indicating that these firms are in relatively poor operational conditions. Therefore, they tend to invest less in R&D.

For other control variables, the coefficient for Δ $cashholdings_t$ is positive and slightly significant. In addition, the coefficients for $tobinq_t$ and $size_t$ are nonsignificant. This is possibly because the growth opportunity and size of firms have already been selected, as we utilise listed firms to construct our sample. In China, for instance, a firm is required to have sustainable growth ability and a minimum of 30,000,000 RMB capital stock with over 25% publicly issued shares to be listed on the ChiNext Board market (http://www.szse.cn/disclosure/notice/company/t20120420_508712.html).

Since there can be significant variations between young and mature firms [2, 9], we split the full sample into young and mature firm subsamples based on the firms' age to further check the findings.

2.2. Firm Maturity, Cash Flow Sources, and R&D Investment. Table 5 provides the dynamic regressions for the young and mature firm subsamples. All explanatory variables in these regressions are treated as potentially endogenous, and we refer to Brown and Peterson [9]; Sasaki [37]; and Tori and Onaran [34] to use lagged values dated t-3 to t-5 as instruments. The p values for the AR(1) statistics indicate first-order autocorrelation in the errors, and the p values for the AR(2) statistics do not reject the null hypothesis of no second-order autocorrelation. Moreover, the Hansen tests cannot reject the instruments' validity [3].

In Table 5, columns 1 and 3 document the results for the young firm subsample, and columns 2 and 4 report the results for the mature firm subsample. Past R&D is always significantly and positively associated with current R&D. Similar to the results of the full sample, the coefficients for tobinq_t and size_t are nonsignificant. The coefficients for cfo_{t-1} and cfi_{t-1} in the regression specifications are always nonsignificant, which indicates that the lagged CFO and the lagged CFI do not have a significant relationship with contemporaneous R&D for both young and mature firms. The coefficients for cff_{t-1} are only negatively significant in columns 1 and 3, showing that an increase in lagged CFF is significantly related to a decrease in contemporaneous R&D for young, but not mature, firms. Hence, the result confirms hypothesis 3 only for CFF. The R&D investment of both young and mature firms is less likely to be influenced by the availability of CFO and CFI, while an increase in CFF leads to a more significant decrease in R&D investment by young Chinese firms than by mature firms.

The finding on CFF is different between young and mature firms in this study. Mature firms are more capable of relying on their internal funds to finance R&D because they have accumulated more profits over the years [38]. In contrast, young firms always lack internal funds. Hence, young firms are more likely to rely on CFF than mature firms are [39]. Since investors in CFF are normally inclined to pursue short-term and low-risk investments, they are less likely to support R&D investment, which has a long-term risk [40, 41]. In addition, as discussed above, firms tend to sustain a relatively stable R&D level with changes in CFI and CFO and are reluctant to invest more in R&D when CFF increases, which can be explained by the high adjustment costs and high uncertainty of R&D investment [42].

2.3. Intentions of Financing Activities and R&D Investment. Based on the analyses above, it is found that firms with available CFO, CFI, and CFF may not spend more on R&D. Moreover, young firms may spend significantly less on R&D when they have more CFF. Hence, this section verifies the intentions for firms to acquire CFF to further support these findings.

Table 5: Dynamic R&D regressions for young and mature firm subsamples.

	Young (1)	Mature (2)	Young (3)	Mature (4)
rd_{t-1}	0.455**	0.359	0.854***	0.503**
	(0.253)	(0.307)	(0.211)	(0.233)
${\text{rd}_{t-1}}^2$	−0.599	0.119	−3.958**	−1.071
	(1.879)	(2.761)	(1.626)	(2.630)
cfo_{t-1}	0.016	0.030	−0.002	−0.003
	(0.023)	(0.028)	(0.019)	(0.016)
cfi_{t-1}	0.007	−0.004	−0.002	−0.012
	(0.012)	(0.014)	(0.012)	(0.013)
cff_{t-1}	−0.027***	−0.012	−0.027**	−0.012
	(0.009)	(0.010)	(0.011)	(0.010)
sales_{t-1}	−0.008*	−0.005	−0.001	−0.001
	(0.005)	(0.006)	(0.004)	(0.003)
tobinq_t			−0.002	−0.003
			(0.003)	(0.003)
$\Delta\ \text{cashholdings}_t$			0.019	0.001
			(0.014)	(0.011)
size_t			−0.005	0.001
			(0.005)	(0.004)
Time dummies	Yes	Yes	Yes	Yes
Firm dummies	Yes	Yes	Yes	Yes
AR(1)	0.000	0.012	0.000	0.003
AR(2)	0.603	0.627	0.218	0.846
Hansen	0.573	0.208	0.538	0.289
Obs.	4,778	2,398	4,160	2,101
Firms	1,406	713	1,347	674

Note. Robust standard errors are in parentheses; ***$p < 0.01$, **$p < 0.05$, and *$p < 0.1$. Estimation is by first-difference GMM. Time and firm fixed effects are included in all regressions. Lagged values dated from t-3 to t-5 are utilised as instruments for all explanatory variables. Columns 1 and 3 are regressions for the young firm subsample, and columns 2 and 4 are regressions for the mature firm subsample.

Financing activities basically involve issuing or taking debt or equity. To verify the intentions behind financing activities, we examine the disclosed records of Chinese listed firms in the CSMAR database during the period 2010 to 2017. The results are shown in Table 6. It is acknowledged that only few firms disclose the intentions behind their financing activities, but the statistics can provide an overview of possible trends for firms in raising funds. We classify the intentions of financing activities into six groups: innovation, capital investment, operation, funding, debt repayment, and others.

Column 1 of Table 6 presents the analysis of all the disclosed intentions relating to stock issuance or reissuance. A total of 2,777 out of 5,410 observations disclosed intentions. Of these observations, approximately 6.80% of funds are raised for innovation. Column 2 of Table 6 provides the analysis of all the disclosed intentions for borrowing bank loans, which is a main source of debt finance. A total of 1,334 out of 2,179 observations disclosed intentions. Of these observations, only 0.29% of loans are borrowed for innovation. These statistics show that firms spend merely a small portion of funds from financing activities on innovation, which supports our findings that firms with more CFF do not necessarily invest more in R&D.

Table 6: Disclosed intents of financing activities.

Intents	Activities	Stock issues (1) (%)	Bank loans (2) (%)
Innovation	Research and development, technical reform, and new product development	6.80	0.29
Operation	Enlarging firm size, merger and acquisition, and purchasing fixed assets	35.26	69.62
Capital investment	Operation maintenance and projects	22.73	9.01
Debt repayment	Adding liquid funds	3.90	4.76
Funding	Paying for debt	29.69	10.90
Others	All other purposes such as advertising, training, and export security	1.62	5.43

Based on the intentions disclosed on the CSMAR, the study classifies them into six groups: innovation, capital investment, operation, funding, debt repayment, and others. Details of the activities for each group are presented. The statistics are reported for the disclosed records of Chinese listed firms on the CSMAR during the period 2010 to 2017. Column 1 provides the percentage of the funds utilised for different intentions, obtaining from stock issues. Column 2 provides the percentage of the funds utilised for different intentions, obtaining from borrowing bank loans.

We further examine the mean values of cash flow and R&D for the disclosed observations and present the results in Table 7. Compared with Table 3, the mean values of cfo_t are quite close, showing that the operational state of firms that obtained finance by issuing stock is similar to the average level for listed firms. The mean R&D spending is 0.029 for young firms and 0.018 for mature firms, which is quite close to the values of 0.027 and 0.018 in Table 3. However, the absolute mean values of cfi_t in Table 7 are much larger than those in Table 3. This suggests that firms in good operating conditions can raise more funds by issuing and reissuing stocks, but they tend to conduct more capital investment than R&D investment. The results can be explained by the conservative attitude of Chinese listed firms due to the high uncertainty of innovation investment. In addition, we find that firms that need bank loans have a much smaller mean cfo_t, indicating that these firms are in a relatively poor operational condition and find it more difficult to sustain their business by operational profits. These firms are less likely to invest in R&D, even if they obtain funding through bank loans.

According to the statistics from the CSMAR database, the mean value of net debt issuance is approximately three times greater than that of the net stock issuance. However, the analyses show that the firms are more likely to support R&D by equity financing than bank loans. One explanation is that the firms that can raise funds from stocks are in better operational condition than the firms that need loans; hence, even if more funds are acquired from debt issuance, the funds are used for purposes such as maintenance and debt repayment. Previous studies on innovation finance have suggested that a large amount of stock or debt issuance may be a main source of R&D spending [3, 9], which is consistent with the previous research conclusions in this paper. It is necessary to verify the intentions of financing activities. The results confirm that firms may not invest in innovation even if they can obtain additional funds from financing activities.

2.4. Robustness. To check the stability of our findings, we conduct the following analyses. The results are shown in Tables 8–13. First, in columns 1–3 of Table 8, we employ the alternative Tobin's *Q* measure and firm size measure [31] and obtain consistent results. Second, in columns 4–6 of Table 8, we add return on assets (roa) to control for profitability [1] and tax-to-asset ratio (tax) to control for the firm tax environment [43]. The results are consistent when additional control variables are added. Third, we adjust the instrument set to include lagged values dated from t-2 to t-5 and lagged values dated from t-3 to t-4. The results are presented in Table 9 and are consistent with the findings in Table 5.

Table 7: Analysis of the disclosed observations.

	Stock issues			Bank loans		
	Full (1)	Young (2)	Mature (3)	Full (4)	Young (5)	Mature (6)
cfo_t	0.049	0.047	0.054	0.026	0.029	0.018
cfi_t	−0.152	−0.171	−0.091	−0.077	−0.075	−0.082
cff_t	0.255	0.280	0.171	0.068	0.059	0.087
rd_t	0.027	0.029	0.018	0.011	0.012	0.010

The mean values of cash flow and R&D for the disclosed observations are analysed in this table.

Fourth, the sample utilised above includes firms with at least one R&D observation. Since some relevant studies construct their samples with at least three R&D observations [37], this paper analyses the regressions using an alternative sample with at least three R&D observations as well in Table 10. The results are invariant with the sample selection criterion for the number of R&D observations.

Fifth, while we drop observations with no R&D in the analyses above, we set missing R&D values to zero as in Pang & Wang [44] to check the robustness. The results are shown in Table 11, which are verified to be robust.

In addition, we employ 10 years as an alternative firm age selection criterion, referring to Brown et al. [9]; Haltiwanger et al. [45]; and Coad et al. [46]. The results, presented in Table 12, show that the findings are robust when firm age selection criterion is different. Further, to verify that the effects of various types of cash flow on R&D investment are significantly different among young and mature firms, we also interact CFO, CFI, and CFF with firm age variables (Age_a_t, Age_b_t, and Age_c_t), respectively, for the full sample, following the research of Howell [47]. The results are reported in Table 13, which corroborate the significant effects of CFF on R&D investment for young firms.

Table 8: Additional controls and alternative controls.

	Alternative control			Additional controls		
	Full (1)	Young (2)	Mature (3)	Full (4)	Young (5)	Mature (6)
rd_{t-1}	0.926***	0.877***	0.503**	0.796***	0.774***	0.556**
	(0.165)	(0.201)	(0.210)	(0.163)	(0.202)	(0.219)
${rd_{t-1}}^2$	−3.829***	−4.052**	−1.095	−3.290**	−3.701**	−1.824
	(1.359)	(1.576)	(2.330)	(1.432)	(1.629)	(2.616)
cfo_{t-1}	−0.009	−0.001	−0.002	−0.019	−0.010	−-0.006
	(0.018)	(0.020)	(0.017)	(0.015)	(0.016)	(0.017)
cfi_{t-1}	−0.001	−0.004	−0.015	0.004	−0.002	−0.015
	(0.010)	(0.012)	(0.014)	(0.011)	(0.011)	(0.013)
cff_{t-1}	−0.021**	−0.028***	−0.008	−0.019**	−0.019*	−0.016
	(0.010)	(0.010)	(0.011)	(0.010)	(0.010)	(0.010)
$sales_{t-1}$	−0.002	−0.000	−0.001	−0.003	−0.003	−0.000
	(0.003)	(0.004)	(0.003)	(0.003)	(0.005)	(0.003)
$tobinq_t$	−0.002	−0.001	−0.002	−0.004	−0.004	−0.003
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
Δ $cashholdings_t$	0.024**	0.015	0.013	0.014	0.022*	-0.011
	(0.011)	(0.012)	(0.012)	(0.010)	(0.013)	(0.009)
$size_t$	−0.003	−0.003	−0.001	−0.001	−0.005	0.005
	(0.003)	(0.004)	(0.005)	(0.003)	(0.004)	(0.004)
roa_t				0.001**	0.001**	0.001***
				(0.000)	(0.000)	(0.000)
tax_t				0.039	0.071	-0.020
				(0.064)	(0.080)	(0.089)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Firm dummies	Yes	Yes	Yes	Yes	Yes	Yes
AR(1)	0.000	0.000	0.001	0.000	0.000	0.003
AR(2)	0.059	0.252	0.476	0.444	0.769	0.465
Hansen	0.114	0.262	0.566	0.239	0.756	0.450
Obs.	6,316	4,190	2,126	6,043	4,044	1,999
Firms	1,835	1,354	676	1,800	1,329	653

Note. Robust standard errors are in parentheses; $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$. Estimation is by first-difference GMM. Time and firm fixed effects are included in all regressions. Lagged values dated from t-3 to t-5 are utilised as instruments for all explanatory variables. Columns 1–3 are regressions with alternative control variables. Columns 4–6 are regressions with additional control variables. Among them, regressions in columns 1 and 4 are for the full sample; regressions in columns 2 and 5 are for the young firm subsample; and regressions in columns 3 and 6 are for the mature firm subsample.

Table 9: Other instruments.

	Full (t-2, t-5)	Young (t-2, t-5)	Mature (t-2, t-5)	Full (t-3, t-4)	Young (t-3, t-4)	Mature (t-3, t-4)
	(1)	(2)	(3)	(4)	(5)	(6)
rd_{t-1}	0.800***	0.751***	0.621***	1.028***	1.112***	0.618***
	(0.133)	(0.169)	(0.180)	(0.205)	(0.257)	(0.236)
${rd_{t-1}}^2$	−3.193***	−2.702**	−2.591	−4.328***	−5.742***	−1.841
	(1.078)	(1.359)	(1.576)	(1.592)	(1.973)	(2.654)
cfo_{t-1}	−0.005	−0.005	0.001	−0.003	−0.002	−0.014
	(0.004)	(0.006)	(0.006)	(0.022)	(0.023)	(0.019)
cfi_{t-1}	−0.001	−0.001	0.001	−0.004	−0.008	−0.015
	(0.003)	(0.003)	(0.006)	(0.013)	(0.013)	(0.015)
cff_{t-1}	−0.010***	−0.012***	−0.003	−0.031***	−0.027**	−0.012
	(0.003)	(0.004)	(0.004)	(0.011)	(0.012)	(0.012)
$sales_{t-1}$	−0.003	−0.004	−0.001	−0.000	−0.000	−0.001
	(0.002)	(0.004)	(0.002)	(0.003)	(0.004)	(0.003)
$tobinq_t$	0.001	−0.002	0.002	−0.002	−0.002	−0.003
	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)
Δ $cashholdings_t$	0.017	0.029	0.006	0.012	0.019	-0.005
	(0.015)	(0.025)	(0.015)	(0.013)	(0.014)	(0.011)
$size_t$	−0.004	−0.009	0.001	−0.001	−0.003	0.003

Table 9: Continued.

	Full (t-2, t-5)	Young (t-2, t-5)	Mature (t-2, t-5)	Full (t-3, t-4)	Young (t-3, t-4)	Mature (t-3, t-4)
	(0.004)	(0.006)	(0.006)	(0.004)	(0.005)	(0.004)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Firm dummies	Yes	Yes	Yes	Yes	Yes	Yes
AR(1)	0.000	0.000	0.000	0.000	0.000	0.005
AR(2)	0.098	0.215	0.924	0.066	0.232	0.892
Hansen	0.743	0.696	0.501	0.073	0.219	0.386
Obs.	6,261	4,160	2,101	6,261	4,160	2,101
Firms	1,826	1,347	674	1,826	1,347	674

Note. Robust standard errors are in parentheses. *Significant at the 1.0% level. **Significant at the 0.5% level. ***Significant at the 0.1% level. Estimation is by first-difference GMM. Time and firm fixed effects are included in all regressions. Columns 1-3 are regressions with instruments lagged twice to five times. Columns 4-6 are regressions with instruments lagged three to four times. Among them, regressions in columns 1 and 4 are for the full sample; regressions in columns 2 and 5 are for the young firm subsample; and regressions in columns 3 and 6 are for the mature firm subsample.

Table 10: Alternative sample selection criterion: at least three R&D observations.

	Full (1)	Young (2)	Mature (3)	Full (4)	Young (5)	Mature (6)
rd_{t-1}	0.433*	0.455*	0.359	0.907***	0.854***	0.503**
	(0.247)	(0.253)	(0.307)	(0.169)	(0.211)	(0.233)
${rd_{t-1}}^2$	0.007	−0.599	0.119	−3.603***	−3.958**	−1.071
	(1.850)	(1.879)	(2.761)	(1.398)	(1.626)	(2.630)
cfo_{t-1}	0.022	0.016	0.030	−0.011	−0.002	−0.003
	(0.024)	(0.023)	(0.028)	(0.018)	(0.019)	(0.016)
cfi_{t-1}	0.012	0.007	−0.004	0.001	−0.002	−0.012
	(0.011)	(0.012)	(0.014)	(0.010)	(0.012)	(0.013)
cff_{t-1}	−0.022**	-0.027***	−0.012	−0.025***	−0.027**	−0.012
	(0.009)	(0.009)	(0.010)	(0.010)	(0.011)	(0.010)
$sales_{t-1}$	-0.011***	−0.008*	−0.005	−0.002	−0.001	−0.001
	(0.004)	(0.005)	(0.006)	(0.003)	(0.004)	(0.003)
$tobinq_t$				−0.003	−0.002	−0.003
				(0.002)	(0.003)	(0.003)
Δ $cashholdings_t$				0.019*	0.019	0.001
				(0.011)	(0.014)	(0.011)
$size_t$				−0.003	−0.005	0.001
				(0.003)	(0.005)	(0.004)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Firm dummies	Yes	Yes	Yes	Yes	Yes	Yes
AR(1)	0.000	0.000	0.012	0.000	0.000	0.003
AR(2)	0.955	0.603	0.627	0.062	0.218	0.846
Hansen	0.538	0.573	0.208	0.137	0.538	0.289
Obs.	7,176	4,778	2,398	6,261	4,160	2,101
Firms	1,898	1,406	713	1,826	1,347	674

Note. Robust standard errors are in parentheses; $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$. Estimation is by first-difference GMM. Time and firm fixed effects are included in all regressions. Lagged values dated from t-3 to t-5 are utilised as instruments for all explanatory variables. An alternative sample with at least three R&D observations is executed. Regressions in columns 1 and 4 are for the full sample; regressions in columns 2 and 5 are for the young firm subsample; and regressions in columns 3 and 6 are for the mature firm subsample.

Table 11: Replacement of all missed R&D values as zero.

	Full (1)	Young (2)	Mature (3)	Full (4)	Young (5)	Mature (6)
rd_{t-1}	1.816***	1.267**	1.891***	1.497***	1.025**	1.343***
	(0.417)	(0.541)	(0.538)	(0.340)	(0.416)	(0.492)
${rd_{t-1}}^2$	−12.199***	−6.483	−15.365***	−9.425***	−5.015	−10.037***
	(3.862)	(5.084)	(4.490)	(2.778)	(3.644)	(3.583)
cfo_{t-1}	−0.027	0.024	0.014	−0.002	−0.004	0.008

Table 11: Continued.

	Full (1)	Young (2)	Mature (3)	Full (4)	Young (5)	Mature (6)
	(0.025)	(0.026)	(0.028)	(0.025)	(0.021)	(0.021)
cfi_{t-1}	0.017	0.031	0.006	0.005	−0.008	0.012
	(0.022)	(0.022)	(0.028)	(0.014)	(0.016)	(0.013)
cff_{t-1}	−0.039***	−0.032**	−0.010	−0.023**	−0.034***	−0.007
	(0.015)	(0.015)	(0.018)	(0.010)	(0.013)	(0.008)
$sales_{t-1}$	−0.012**	−0.013**	0.001	−0.005*	−0.003	0.002
	(0.005)	(0.006)	(0.006)	(0.003)	(0.004)	(0.003)
$tobinq_t$				0.001	0.003	−0.003
				(0.002)	(0.003)	(0.004)
Δ $cashholdings_t$				0.012	0.031**	−0.009
				(0.015)	(0.014)	(0.015)
$size_t$				−0.007*	−0.008*	0.004
				(0.004)	(0.004)	(0.006)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Firm dummies	Yes	Yes	Yes	Yes	Yes	Yes
AR(1)	0.000	0.000	0.005	0.000	0.000	0.003
AR(2)	0.368	0.646	0.210	0.921	0.215	0.065
Hansen	0.553	0.727	0.740	0.529	0.582	0.869
Obs.	9,490	5,553	3,937	8,186	4,822	3,364
Firms	2,347	1,626	1,057	2,255	1,557	995

Note. Robust standard errors are in parentheses; ***$p < 0.01$, **$p < 0.05$, and *$p < 0.1$. Estimation is by first-difference GMM. Time and firm fixed effects are included in all regressions. Lagged values dated from t-3 to t-5 are utilised as instruments for all explanatory variables. All missing R&D values are set to zero. Regressions in columns 1 and 4 are for the full sample; regressions in columns 2 and 5 are for the young firm subsample; and regressions in columns 3 and 6 are for the mature firm subsample.

Table 12: Alternative firm age selection criterion.

	Young (1)	Mature (2)	Young (3)	Mature (4)
rd_{t-1}	0.794***	0.191	0.824***	0.717***
	(0.283)	(0.316)	(0.242)	(0.177)
${rd_{t-1}}^2$	−3.035	1.492	−4.300**	−2.850
	(1.998)	(2.692)	(1.715)	(1.930)
cfo_{t-1}	0.015	0.006	−0.018	−0.012
	(0.025)	(0.031)	(0.023)	(0.018)
cfi_{t-1}	0.010	0.005	0.004	−0.005
	(0.013)	(0.015)	(0.010)	(0.012)
cff_{t-1}	−0.022**	−0.012	−0.025**	−0.014
	(0.009)	(0.010)	(0.010)	(0.010)
$sales_{t-1}$	0.001	−0.015***	0.009**	−0.005*
	(0.004)	(0.005)	(0.004)	(0.003)
$tobinq_t$			0.001	−0.002
			(0.003)	(0.003)
Δ $cashholdings_t$			0.018	0.017
			(0.012)	(0.013)
$size_t$			−0.004	0.003
			(0.004)	(0.004)
Time dummies	Yes	Yes	Yes	Yes
Firm dummies	Yes	Yes	Yes	Yes
AR(1)	0.000	0.016	0.000	0.000
AR(2)	0.140	0.979	0.840	0.122
Hansen	0.697	0.533	0.539	0.366
Obs.	3,533	3,643	3,063	3,198
Firms	1,116	1,005	1,073	954

Note. Robust standard errors are in parentheses; ***$p < 0.01$, **$p < 0.05$, and *$p < 0.1$. Estimation is by first-difference GMM. Time and firm fixed effects are included in all regressions. Lagged values dated from t-3 to t-5 are utilised as instruments for all explanatory variables. An alternative firm age selection criterion is executed. The firms with less than 10 years' listed experience are categorised as young firms, and the remaining firms are counted as mature firms. Regressions in columns 1 and 4 are for the full sample; regressions in columns 2 and 5 are for the young firm subsample; and regressions in columns 3 and 6 are for the mature firm subsample.

Table 13: Interactions.

	(1)	(2)	(3)
rd_{t-1}	0.903***	0.922***	0.893***
	(0.168)	(0.166)	(0.173)
${rd_{t-1}}^2$	−3.595**	−3.953***	−3.463**
	(1.432)	(1.423)	(1.428)
$sales_{t-1}$	−0.002	−0.003	−0.003
	(0.003)	(0.003)	(0.003)
$tobinq_t$	−0.003	−0.002	−0.002
	(0.002)	(0.002)	(0.002)
Δ $cashholdings_t$	0.021**	0.025**	0.019*
	(0.010)	(0.010)	(0.011)
$size_t$	−0.003	−0.001	−0.004
	(0.003)	(0.004)	(0.003)
cfo_{t-1} × Age	−0.006	0.007	−0.005
	(0.020)	(0.027)	(0.007)
cfi_{t-1} × Age	−0.001	−0.002	0.002
	(0.013)	(0.014)	(0.005)
cff_{t-1} × Age	−0.033***	−0.030**	−0.009***
	(0.012)	(0.013)	(0.004)
Time dummies	Yes	Yes	Yes
Firm dummies	Yes	Yes	Yes
AR(1)	0.000	0.000	0.000
AR(2)	0.086	0.125	0.071
Hansen	0.183	0.188	0.133
Obs.	6,266	6,266	6,237
Firms	1,827	1,827	1,823

Note. Robust standard errors are in parentheses; ***$p < 0.01$, **$p < 0.05$, and *$p < 0.1$. Estimation is by first-difference GMM. Time and firm fixed effects are included in all regressions. Regressions in columns 1 and 3 are for the full sample. Here, Age represents the firm age variables, which are Age_a_t, Age_b_t, and Age_c_t in columns 1,2, and 3, respectively.

3. Conclusions

R&D is an important driver of firms' competitiveness and economic growth. Using firm-level dynamic panel data for 2,401 Chinese listed firms from 2010 to 2017, this study investigates whether firms' R&D spending would increase when more cash flows from different business activities became available. In addition, the differential impacts of the cash flow sources of firms with different levels of maturity on R&D investment are discussed.

The findings concern mainly two aspects. First, it is found that Chinese firms will not necessarily spend more cash on R&D even if funds are available. In contrast, they would rather use funds for projects that are more profitable in the short term. This explanation can be supported by the disclosure information on financing activities presented above. Second, the findings suggest that CFO and CFI are nonsignificantly related to R&D spending for both young and mature firms, and CFF has a significantly negative impact on R&D spending for young, but not mature, firms.

With regard to implications, this study provides new insight into financing availability for innovative investment. In China and other emerging economies, financial markets are imperfect, and many firms lack funds [14]. Since continuous R&D investments require a great amount of cash, firm CFO and CFF may not be sufficient to meet such needs. Furthermore, even if cash is abundant, firm managers and external investors may be unwilling to invest in R&D due its high uncertainty. Hence, alleviating financing constraints may not lead to greater innovation investments. Measures should be implemented to increase the willingness of firm managers to invest in R&D.

In addition, this paper clarifies the influencing mechanism between different cash flow sources and R&D investment. Firms are suggested to plan and manage CFO, CFI, and CFF as an organic unified system. According to the characteristics of the cash flow throughout the whole process of R&D, based on the differential management of firm cash flow, the investment management of CFO, CFF, and CFI on R&D should be integrated into an organic and unified management system. The differential management of CFF minimises financial costs while providing start-up funds for R&D investment. The differential management of CFI offers an important guarantee for R&D financing. The differential management of CFO would maximise the benefits of R&D investment by increasing R&D efficiency and reducing fees. In turn, the differential management of CFO would promote the steady progress of R&D financing and thus realise the integration of all aspects and the whole process of R&D activities.

Finally, this study explains the heterogeneous impacts of the cash flow sources of firms with different levels of maturity on R&D investment. Compared with the R&D investment of mature firms, the R&D investment of young firms is more significantly and negatively influenced by CFF. This is because young firms that need more CFF are generally in relatively poor operational condition. In addition, external investors are inclined to pursue short-term and low-risk investments and are less likely to support R&D investment, which has high uncertainty. To ensure the investment in R&D for young firms, policies and measures to stimulate innovation funds from external investors are suggested.

Since this study uses a different categorisation of finance from the perspective of business activities, future research might analyse the impact of funds from specific business activities (e.g., instrument investments, mergers, and acquisitions) on innovation activities. As for the study limitations, since a certain number of listed firms did not disclose their financing intents, we acknowledge that we only analysed the disclosed firms' intents regarding stock issuing, reissuing, and bank loan borrowing in this study. Therefore, it is recommended that these results are used only as support.

Data Availability

The sample of this study includes all Chinese firms from the Main Board, Small and Medium Enterprise (SME) Board, and ChiNext Board listed on the Shanghai Stock Exchange and the Shenzhen Stock Exchange. These listed firms cover 18 industries, and the details are presented in Table 1. The majority of statistics and disclosed information on financing activities used to support the findings of this study were, respectively, supplied by Wind (https://www.wind.com.cn/newsite/about.html) and China Stock Market & Accounting Research Database (CSMAR, https://cn.gtadata.com) under license and so cannot be made freely available without permission. 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 acknowledge financial support from the National Natural Science Foundation of China (Project nos. 71832013 and 71801200), Social Science Foundation of Jiangsu Province (no. 20EYC002), and Fundamental Research Funds for the Central Universities of Hohai University.

References

[1] J. Cornaggia, Y. Mao, X. Tian, and B. Wolfe, "Does banking competition affect innovation?" *Journal of Financial Economics*, vol. 115, no. 1, pp. 189–209, 2015.

[2] Q. Weng and M. Söderbom, "Is R&D cash flow sensitive? Evidence from Chinese industrial firms," *China Economic Review*, vol. 47, pp. 77–95, 2018.

[3] J. R. Brown, G. Martinsson, and B. C. Petersen, "Do financing constraints matter for R&D?" *European Economic Review*, vol. 56, no. 8, pp. 1512–1529, 2012.

[4] F. Moshirian, V. Nanda, A. Vadilyev, and B. Zhang, "What drives investment-cash flow sensitivity around the World? An asset tangibility Perspective," *Journal of Banking & Finance*, vol. 77, pp. 1–17, 2017.

[5] M. Hermosilla and Y. Wu, "Market size and innovation: the intermediary role of technology licensing," *Research Policy*, vol. 47, no. 5, pp. 980–991, 2018.

[6] A. Steinmetz, "Competition, innovation, and the effect of R&D knowledge," *Journal of Economics*, vol. 115, no. 3, pp. 199–230, 2015.

[7] S. Sciascia, M. Nordqvist, P. Mazzola, and A. De Massis, "Family ownership and R&D intensity in small- and medium-sized firms," *Journal of Product Innovation Management*, vol. 32, no. 3, pp. 349–360, 2015.

[8] J. Baumann and A. S. Kritikos, "The link between R&D, innovation and productivity: are micro firms different?" *Research Policy*, vol. 45, no. 6, pp. 1263–1274, 2016.

[9] J. R. Brown, S. M. Fazzari, and B. C. Petersen, "Financing innovation and growth: cash flow, external equity, and the 1990s R&D boom," *The Journal of Finance*, vol. 64, no. 1, pp. 151–185, 2009.

[10] A. Guariglia and P. Liu, "To what extent do financing constraints affect Chinese firms' innovation activities?" *International Review of Financial Analysis*, vol. 36, pp. 223–240, 2014.

[11] S. Sasidharan, P. J. Jijo Lukose, and S. Komera, "Financing constraints and investments in R&D: evidence from Indian manufacturing firms," *The Quarterly Review of Economics and Finance*, vol. 55, pp. 28–39, 2015.

[12] A. Ferrando and A. Ruggieri, "Financial constraints and productivity: evidence from euro area companies," *International Journal of Finance & Economics*, vol. 23, no. 3, pp. 257–282, 2018.

[13] J. W. Ahiadorme, A. Gyeke-Dako, and J. Y. Abor, "Debt holdings and investment cash flow sensitivity of listed firms," *International Journal of Emerging Markets*, vol. 13, no. 5, pp. 943–958, 2018.

[14] S. Malamud and F. Zucchi, "Liquidity, innovation, and endogenous growth," *Journal of Financial Economics*, vol. 132, no. 2, pp. 519–541, 2019.

[15] J. I. Bernstein and P. Mohnen, "International R&D spillovers between U.S. and Japanese R&D intensive sectors," *Journal of International Economics*, vol. 44, no. 2, pp. 315–338, 1998.

[16] L. Gu, "Product market competition, R&D investment, and stock returns," *Journal of Financial Economics*, vol. 119, no. 2, pp. 441–455, 2016.

[17] C. P. Himmelberg and B. C. Petersen, "R & D and internal finance: a panel study of small firms in high-tech industries," *The Review of Economics and Statistics*, vol. 76, no. 1, pp. 38–51, 1994.

[18] P. C. Patel and J. J. Chrisman, "Risk abatement as a strategy for R&D investments in family firms," *Strategic Management Journal*, vol. 35, no. 4, pp. 617–627, 2014.

[19] Y. Lee, A. Mazzei, and J.-N. Kim, "Looking for motivational routes for employee-generated innovation: employees' scouting behavior," *Journal of Business Research*, vol. 91, pp. 286–294, 2018.

[20] R. K. Tyagi, "New product introductions and failures under uncertainty," *International Journal of Research in Marketing*, vol. 23, no. 2, pp. 199–213, 2006.

[21] T. Wang and S. Thornhill, "R&D investment and financing choices: a comprehensive perspective," *Research Policy*, vol. 39, no. 9, pp. 1148–1159, 2010.

[22] M. Meuleman and W. De Maeseneire, "Do R&D subsidies affect SMEs' access to external financing?" *Research Policy*, vol. 41, no. 3, pp. 580–591, 2012.

[23] K. Xu, C. Geng, and X. Wei, "Research ON financing ecology and financing efficiency OF strategic emerging industries IN China," *Journal of Business Economics and Management*, vol. 20, no. 2, pp. 311–329, 2019.

[24] C. J. Hadlock and J. R. Pierce, "New evidence on measuring financial constraints: moving beyond the KZ index," *Review of Financial Studies*, vol. 23, no. 5, pp. 1909–1940, 2010.

[25] D. Coluccia, M. Dabić, M. Del Giudice, S. Fontana, and S. Solimene, "R&D innovation indicator and its effects on the market. An empirical assessment from a financial perspective," *Journal of Business Research*, 2019.

[26] R. Qiu and J. Cantwell, "General Purpose Technologies and local knowledge accumulation - a study on MNC subunits and local innovation centers," *International Business Review*, vol. 27, no. 4, pp. 826–837, 2018.

[27] S. Ren, A. B. Eisingerich, and H.-T. Tsai, "How do marketing, research and development capabilities, and degree of internationalization synergistically affect the innovation performance of small and medium-sized enterprises (SMEs)? A panel data study of Chinese SMEs," *International Business Review*, vol. 24, no. 4, pp. 642–651, 2015.

[28] H. Lu, J. Hong, Q. Chen, and P. Deng, "Impact of overseas M&As on Chinese service enterprises' market value," *Service Industries Journal*, pp. 1–21, 2019.

[29] A. Riding, B. Orser, and T. Chamberlin, "Investing in R&D: small- and medium-sized enterprise financing preferences," *Venture Capital: An International Journal of Entrepreneurial Finance*, vol. 14, no. 2-3, pp. 199–214, 2012.

[30] Q. Zhu, X. Li, F. Li, J. Wu, and D. Zhou, "Energy and environmental efficiency of China's transportation sectors under

the constraints of energy consumption and environmental pollution," *Energy Economics*, vol. 89, 2020.

[31] Z. He and M. B. Wintoki, "The cost of innovation: R&D and high cash holdings in U.S. firms," *Journal of Corporate Finance*, vol. 41, pp. 280–303, 2016.

[32] K. Singh, M. Misra, M. Kumar, and V. Tiwari, "A study on the determinants of financial performance of U.s. Agricultural cooperatives," *Journal of Business Economics and Management*, vol. 20, no. 4, pp. 633–647, 2019.

[33] A. Gkypali, A. Rafailidis, and K. Tsekouras, "Innovation and export performance: do young and mature innovative firms differ?" *Eurasian Business Review*, vol. 5, no. 2, pp. 397–415, 2015.

[34] D. Tori and Ö. Onaran, "The effects of financialization on investment: evidence from firm-level data for the UK," *Cambridge Journal of Economics*, vol. 42, no. 5, pp. 1393–1416, 2018.

[35] Q. Zhu, J. Aparicio, F. Li, J. Wu, and G. Kou, "Determining closest targets on the extended facet production possibility set in data envelopment analysis: modeling and computational aspects," *European Journal of Operational Research*, vol. 296, no. 3, pp. 927–939, 2022.

[36] M. Arellano and S. Bond, "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *The Review of Economic Studies*, vol. 58, no. 2, pp. 277–297, 1991.

[37] T. Sasaki, "Financial cash flows and research and development investment," *Pacific-Basin Finance Journal*, vol. 39, pp. 1–15, 2016.

[38] G. Pellegrino, "Barriers to innovation in young and mature firms," *Journal of Evolutionary Economics*, vol. 28, no. 1, pp. 181–206, 2018.

[39] M. G. Colombo, D. J. Cumming, and S. Vismara, "Governmental venture capital for innovative young firms," *The Journal of Technology Transfer*, vol. 41, no. 1, pp. 10–24, 2016.

[40] I. S. Demirag, "The impact of managers' short-term perceptions on technology management and R&D in UK companies," *Technology Analysis & Strategic Management*, vol. 8, no. 1, pp. 21–32, 1996.

[41] W. B. Elliott, J. L. Hobson, and K. E. Jackson, "Disaggregating management forecasts to reduce investors' susceptibility to earnings fixation," *The Accounting Review*, vol. 86, no. 1, pp. 185–208, 2011.

[42] P. Deng and H. Lu, "Transnational knowledge transfer or indigenous knowledge transfer: which channel has more benefits for China's high-tech enterprises?" *European Journal of Innovation Management*, vol. 24, 2021.

[43] A. Alam, M. Uddin, and H. Yazdifar, "Financing behaviour of R&D investment in the emerging markets: the role of alliance and financial system," *R & D Management*, vol. 49, no. 1, pp. 21–32, 2019.

[44] C. Pang and Y. Wang, "Stock pledge, risk of losing control and corporate innovation," *Journal of Corporate Finance*, vol. 60, Article ID 101534, 2020.

[45] J. Haltiwanger, R. S. Jarmin, and J. Miranda, "Who creates jobs? Small versus large versus young," *The Review of Economics and Statistics*, vol. 95, no. 2, pp. 347–361, 2013.

[46] A. Coad, A. Segarra, and M. Teruel, "Innovation and firm growth: does firm age play a role?" *Research Policy*, vol. 45, no. 2, pp. 387–400, 2016.

[47] A. Howell, "Firm R&D, innovation and easing financial constraints in China: does corporate tax reform matter?" *Research Policy*, vol. 45, no. 10, pp. 1996–2007, 2016.

Hindawi
Complexity
Volume 2022, Article ID 6917527, 18 pages
https://doi.org/10.1155/2022/6917527

Research Article

Chinese Stock Market's Reaction to COVID-19 in the Short and Long Run

Hongxia Wang[1] **and Zongzheng Yu**[2]

[1]*College of Economics and Management, Nanjing Forestry University, Nanjing, China*
[2]*King's Business School, King's College London, UK*

Correspondence should be addressed to Hongxia Wang; whx2016@njfu.edu.cn

Received 11 November 2021; Accepted 4 January 2022; Published 1 February 2022

Academic Editor: Qingyuan Zhu

We study the impact of COVID-19 on Chinese stock market which can be seen as a complex system. We use the event study method to evaluate its performance change in terms of the return rate, turnover rate, etc. We show that the abnormal return of stock market was significantly negative after the outbreak of COVID-19 and did not turn positive until May 2020. Moreover, the five-factor model is used to estimate the ordinary returns of different industries and show that abnormal returns for medical and food industries were significantly positive, while energy and public utility industries had significantly negative abnormal returns which persisted for a long time. COVID-19 had lag effects on clothes industry, finance industry, transportation industry, and IT industry. We also find that energy and finance industries had negative abnormal turnover rates during the sample period, while other industries, such as healthcare and telecommunications service industries, had positive abnormal turnover rates.

1. Introduction

In 2020, COVID-19 swept the world, and the global capital market fluctuated violently. The Shanghai Composite dropped by 7.72% on February 3, 2020; the stock indexes of almost all countries, including USA, Canada, Brazil, South Korea, and the Philippines, also plummeted in March 2020. Over the past two years, the COVID-19 pandemic tremendously influenced the global financial markets.

Literature about the impacts of the COVID-19 on the financial market of different countries and regions is increasing. For example, Del et al. [1] found that the increase of confirmed cases can influence the stability of African markets; however, the effect of the fatality rate is not significant. Zaremba et al. [2] showed that, for the countries which have relatively low unemployment rates and plenty of firms with conservative investment policies and low P/E ratio, stock markets are more likely to be immune to the healthcare crisis. Additional works also explore how the government policy and reports about COVID-19 influence the financial market [3–5], for example, Pandey and Kumari (2021) [6, 7]. Moreover, researchers show that there are different impacts on different specific markets. For example, Ji et al. [8] demonstrated that the return of gold increases during the pandemic and remains robust as a safe asset; meanwhile Ali et al. [9] argued that the return of gold becomes negatively related to the COVID-19 deaths with the spread of COVID-19. Mazur et al. [10] and Sayed and Eledum [11] investigated the impact of the COVID-19 outbreak on the return of different industries in USA and Saudi Arabia, respectively. Corbet et al. [12] discussed the effect of COVID-19 on the Bitcoin market.

The aim of this work is to study the impact of COVID-19 on China's A-stock; specifically, we use event study to investigate the abnormal return and turnover rate for different industries in the long run and short run. Although related works, for example, those by Ali et al. [9], Dai et al. [13], Liu et al. [14], Mezghani et al. [15], and Nguyen et al. [16], explore the impact of COVID-19 on the Chinese stock market, we put emphasis on the abnormal return of the whole market via the OLS model; moreover, compared with existing works, for example, Liu et al. [17], we study the impact of COVID-19 on the return of different industries by using a five-factor model [18]. In addition, since liquidation is an important feature of stock market, we explore the change of turnover rate in different industries.

COVID-19 may be traced back to December 8, 2019, when the first patient was suspected to catch COVID-19 in Wuhan, China. The Chinese government confirmed the existence of COVID-19 and set up the experts group on January 19, 2020. Therefore, we use January 20, 2020 (the first trading day after January 19, 2020), as the event date in event study. We use a long event window (about one year) to clearly examine the impact of COVID-19 on Chinese stock market in the short and long run.

Our results prove that China's A-stock suffered from the negative abnormal return which became zero after about 30 days. We find that the turnover rates of most industries increased during the sample period, while the responses to COVID-19 were different for different industries. Some industries, such as medical care (energy), benefitted (suffered) from COVID-19 very much; others, such as real estate, were affected slightly. Moreover, for some industries, such as food, the impact of COVID-19 lasted a short period; on the contrary, for some industries, such as energy and medical care, it lasted a long period.

The contribution of our work is twofold. First, it extends the literature about the impact of COVID-19 on emerging market; second, it uses a five-factor model to study the abnormal return of different industries and also examine the change of turnover rates. The remainder of this work is organized as follows. Section 2 presents data and methodology. Section 3 explores the impact of COVID-19 on China's A-stock. Section 4 investigates the impact of COVID-19 on the return rate of different industries. Section 5 studies the impact of COVID-19 on the turnover rate of different industries. Section 6 gives the conclusion.

2. Data and Methodology

We choose data of A-stock index, return rate, turnover rate, the one-year deposit interest rate, and five-factor data from China Stock Market & Accounting Research (CSMAR). All Country World Index (ACWI) is from Yahoo Finance. They are all the daily data of trading days from January 2, 2019, to December 31, 2020, roughly covering the years before and after the event day.

2.1. Event. In the event study method, the estimation window, the event date, and the event window need to be determined. We use the period from January 1, 2019, to December 6, 2019, as estimation window. The event date is January 20, 2020, since it is the first trading day after January 19, 2020, when China started to count the number of COVID-19 patients. The event window involves the period from t_{-5} to t_{230}, where t_x denotes the xth day before or after the event day.

2.2. Calculation of Benchmark Return and Abnormal Return

2.2.1. The Case of A-Stock. To analyze the abnormal return (AR), we firstly specify a benchmark. Following the works of Dyckman et al. [19], Pandey and Kumari [20], and others, we use OLS market model to define the benchmark return:

$$ER_t = \alpha + \beta ER_{mt}, \tag{1}$$

where α and β represent the intercept and slope of the market model, respectively; ER_{mt} is the return of ACWI on day t; and α and β are calculated by the data from January 1, 2019, to December 6, 2019.

To get the abnormal return, we compare the benchmark return with the actual return by using the equations of AR and CAR as follows:

$$\begin{aligned} AR_t &= R_t - ER_t, \\ CAR &= \Sigma_{t_1}^{t_2} AR_t, \end{aligned} \tag{2}$$

where AR_t is the abnormal return on day t, R_{t} is the actual return on day t, $\mathrm{ER_t}$ is the benchmark return on day t ((1) above), and CAR is the cumulative abnormal return from $\mathrm{t_1}$ to $\mathrm{t_2}$.

To calculate the daily return of actual index and benchmark, we use log-returns [21] as follows:

$$R_t = LN\left(\frac{P_t}{P_{t-1}}\right) * 100, \tag{3}$$

where LN is the log of nature number, P_{t} is the price on day t, and P_{t-1} is the price on day $t-1$.

2.2.2. The Case of Industries. Huang [22] and others argued that the five-factor model in China may be superior to other traditional models in explaining the returns of different specific industries. Thus, the five-factor model is adopted to estimate the benchmark return of various industries as follows:

$$\begin{aligned} ER_{it} - R_f = \alpha + \beta_m\left(R_m - R_f\right) + \beta_{SMB}SMB \\ + \beta_{HML}HML + \beta_{RMW}RMW + \beta_{CMA}CMA, \end{aligned} \tag{4}$$

where R_{m} is the return rate of the tradable market value weighted index of A-stock; R_{f} is the risk-free return rate; SMB is the difference between the return rate of the low market value stock portfolio and that of the high market value stock portfolio; HML is the difference between the return rate of the high book value stock portfolio and that of the low book value stock portfolio; RMW is the difference between the return of the high-profit stock portfolio and that of the low-profit stock portfolio; and CMA is the difference between the return of the high investment ratio stock portfolio and that of the low investment ratio stock portfolio.

The AR of industries can be obtained by the following equation:

$$AR_{it} = R_{it} - ER_{it}, \quad (5)$$

where AR_{it} is the AR of one industry return on day t; R_{it} is the actual return of one industry on day t; and ER_{it} is the benchmark return on day t.

2.2.3. The Case of Turnover Rate. Following Liang [23], Michaely et al. [24], and others, the normal turnover rate is defined as the average turnover rate in the estimation window. We calculate the abnormal turnover rate and cumulative abnormal turnover rate as follows:

$$\overline{TO} = \frac{\Sigma_1^n TO_t}{n},$$
$$AT_t = \frac{TO_t - \overline{TO}}{\overline{TO}}, \quad (6)$$
$$CAT = \Sigma_{t_1}^{t_2} AT_t,$$

where n is the total number of days in the estimation window; TO_t is the turnover rate on day t; $\overline{TO}$ is the average turnover rate during the estimation window. AT_t is the abnormal turnover rate on day t; σ_t is the variance of the turnover rate during the estimation window; and CAT is the accumulated abnormal turnover from t_1 to t_2.

2.3. Calculation of t-Statistics. To determine the significance of the AR and CAR, we use popular parameter test t-statistics [25, 26]:

$$\sigma_A = \sqrt{\frac{\Sigma_1^n (AR - AAR)^2}{n}},$$
$$t - \text{statistics}_{AR} = \frac{AR_t}{\sigma_A}, \quad (7)$$
$$t - \text{statistics}_{CAR} = \frac{CAR}{\sqrt{(t_2 - t_1 + 1) * \sigma_A^2}},$$

where σ_A is the standard variance on the estimation window, AAR is the abnormal average return, n is the number of estimation days, and $\text{t} - \text{statistics}_{\text{AR}}$ represents the t-statistics of AR. CAR represents the cumulative abnormal return from t_1 to t_2; and $\text{t} - \text{statistics}_{\text{CAR}}$ represents the t-statistics of CAR.

3. The Impact of COVID-19 on China's A-Stock

As Figures 1 and 2 show, roughly speaking, China's A-stock suffered from two shocks of COVID-19 in the sample period: the outbreak in China and the outbreak in many other countries. However, it recovered quickly from the two shocks.

According to Figures 1 and 2 and Table 1, it can be seen that the AR was significantly negative on t_2 and remained negative until t_4 when the CAR reached −13.96%. Starting from t_4, the CAR gradually turned to zero and returned to a positive value on t_{30}. Subsequently, it turned negative again and reached its lowest point on t_{90}. Around t_{100}, the stock index suddenly rose and got close to the predicted return rate on t_{120}.

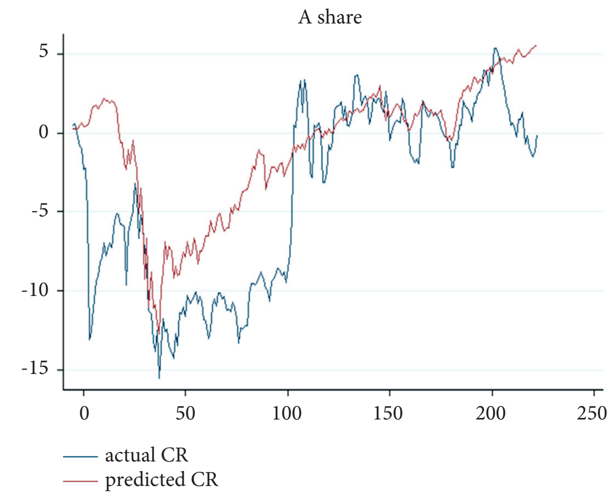


Figure 1: Actual and predicted cumulative return.

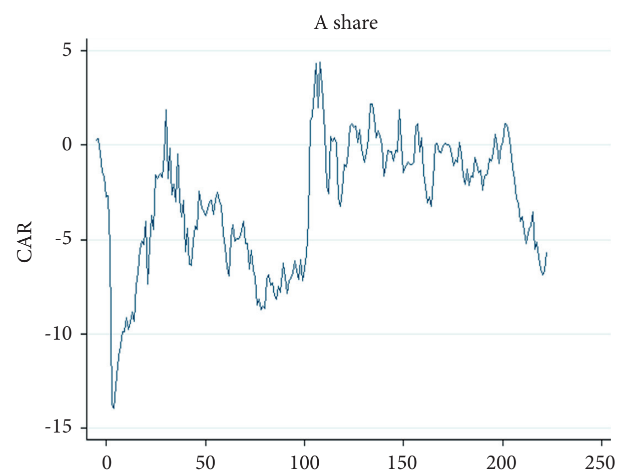


Figure 2: Cumulative abnormal return.

Table 1 presents that, on t_2 and t_3, the market had significantly negative returns. The CAR from day t_3 to day t_{10} was also statistically significant. On t_{30}, the market had significant positive return, but the CAR was no longer significant.

4. The Impact of COVID-19 on the Return Rate of A-Stock in Various Industries

COVID-19 greatly influenced the return of A-stock; however, COVID-19 may have different impacts on various industries. We here investigate this issue by selecting the 12 following industries: healthcare, food and major supplies retail, utilities, energy, durable consumer goods and clothing, media, finance, real estate, materials, transportation, information technology, and telecommunication.

Table 1: AR, CAR, and t-statistics for China's A-stock.

Date t	AR (%)	t-stat	CAR (%)	t-stat	Date t	AR (%)	t-stat	CAR (%)	t-stat
−5	0.24	0.20	0.24	0.20	6	1.59	1.37	−11.20	−2.78
−4	0.05	0.04	0.29	0.18	7	0.57	0.49	−10.64	−2.54
−3	−0.75	−0.65	−0.46	−0.23	8	0.77	0.67	−9.87	−2.27
−2	−0.91	−0.78	−1.37	−0.59	9	0.00	0.00	−9.87	−2.19
−1	−0.26	−0.22	−1.63	−0.63	10	0.70	0.60	−9.17	−1.97
0	−1.13	−0.97	−2.76	−0.97	30	2.38	2.05	1.83	0.26
1	0.05	0.04	−2.71	−0.88	60	−0.64	−0.55	−5.37	−0.57
2	−2.69	−2.32	−5.40	−1.64	90	−0.68	−0.58	−6.97	−0.61
3	−8.38	−7.21	−13.78	−3.95	120	1.34	1.15	−1.06	−0.08
4	−0.17	−0.15	−13.96	−3.80	160	−1.85	−1.59	−1.47	−0.10
5	1.16	1.00	−12.80	−3.32	220	−0.36	−0.31	−6.90	−0.39

4.1. *Industries with Positive Abnormal Return*

4.1.1. Healthcare Industry. From Figures 3 and 4 and Table 2, we know that the actual and predicted cumulative returns of the healthcare industry had a clear upward trend in the sample period. In the short term, the CAR rose; on t_{-1}, t_0, and t_3–t_7, the AR appeared significantly positive. But on t_2, t_8, t_9, and t_{10}, the CAR remained significantly positive. Moreover, there was a a longer-term impact of COVID-19 on the healthcare industry; the positive CAR remained significant until t_{120}.

4.1.2. Food and Major Supplies Retail Industry. Figures 5 and 6 show that the actual cumulative return of food and major supplies retail industry exhibited a trend of rising first and then rebounding. In the short term, COVID-19 had a positive impact. From Table 3, we know that the AR was significantly positive on t_2 , t_4, t_5, t_7, and t_9, and the CAR started to become significantly positive on t_{10}. From a long-term perspective, although the CAR of the food industry fell rapidly after t_{100} when reaching the peak, the CAR of the industry remained significant until t_{160}. This indicates that the impact of COVID-19 was relatively long-term.

4.1.3. Summary. Both the medical and food industries had positive ARs under the impact of COVID-19. In the early stage of the epidemic, the demand for COVID-19 medical testing increased significantly; meanwhile, the demand for medical protective equipment grew substantially. In the middle and late stages of the epidemic, the importance of COVID-19 vaccine was highlighted. Moreover, since the large-scale vaccination by the COVID-19 vaccine is still the only solution to the pandemic, it is obviously beneficial to vaccine-related companies. This may be why healthcare industry can keep a positive AR in the short and long term. For food industry, the traffic blockade led to the insufficient food supply; this boosted the performance of food industry companies. In addition, the food industry is a high-quality defense asset, so it can attract a large amount of capital inflow. However, in the long run, as the epidemic eases, investments are transferred from defensive assets, and the AR of the food industry returns to zero.

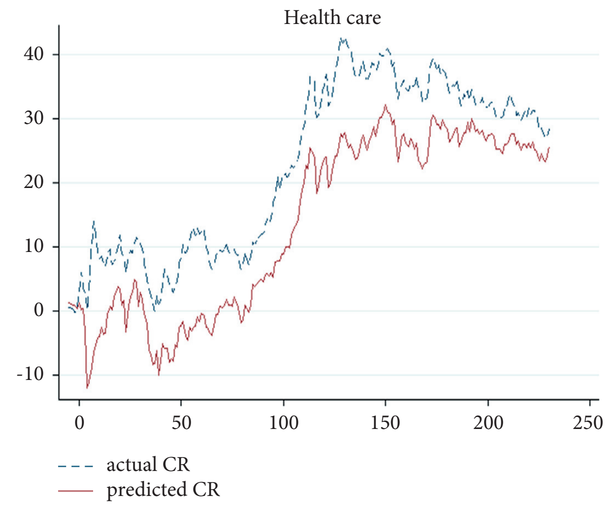


Figure 3: Actual and predicted cumulative returns.

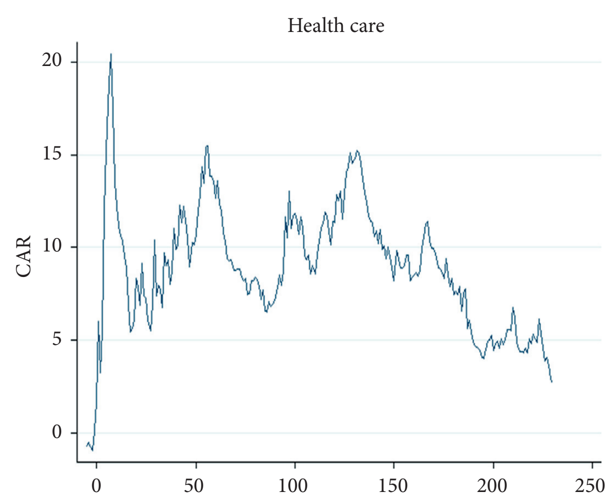


Figure 4: Cumulative abnormal return.

4.2. *Industries with Negative Abnormal Return*

4.2.1. Public Utilities. As Figures 7 and 8 and Table 4 show, from a short-term perspective, the CAR of public utilities has been significantly negative since t_6. On t_9, the CAR reached

Table 2: AR, CAR, and t-statistics for healthcare.

Date	AR (%)	t-stat	CAR (%)	t-stat	Date	AR (%)	t-stat	CAR (%)	t-stat
−5	−0.73	−1.29	−0.73	−1.29	6	3.64	6.46	18.84	9.65
−4	0.22	0.39	−0.51	−0.64	7	1.60	2.85	20.45	10.06
−3	−0.27	−0.49	−0.78	−0.80	8	−3.05	−5.42	17.39	8.25
−2	−0.16	−0.29	−0.94	−0.84	9	−3.87	−6.87	13.52	6.20
−1	1.02	1.82	0.08	0.06	10	−1.30	−2.31	12.22	5.42
0	1.54	2.73	1.62	1.17	30	−3.00	−5.32	7.41	2.19
1	4.36	7.74	5.98	4.01	60	−0.94	−1.67	12.69	2.77
2	−2.75	−4.88	3.23	2.02	90	0.35	0.62	7.33	1.33
3	2.36	4.18	5.58	3.30	120	−0.06	−0.10	11.35	1.79
4	6.47	11.48	12.05	6.76	160	0.10	0.18	8.50	1.17
5	3.15	5.59	15.20	8.13	230	−0.22	−0.39	2.68	0.31

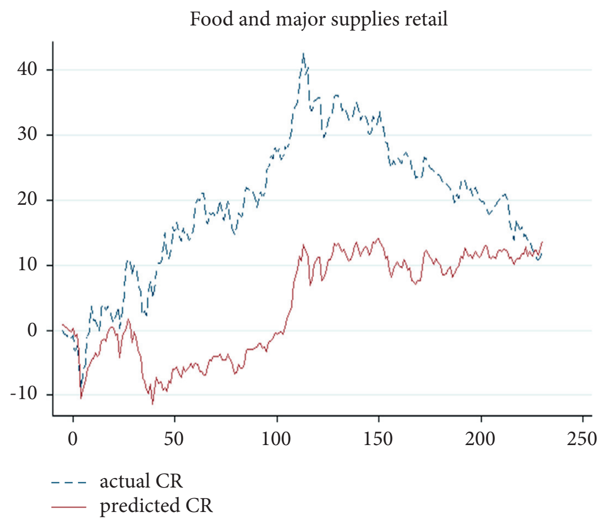


Figure 5: Actual and predicted cumulative returns.

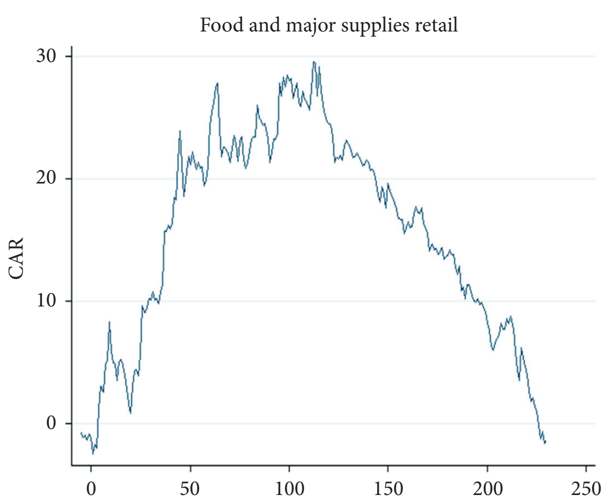


Figure 6: Cumulative abnormal return.

the maximum of −4.34%. In the long run, COVID-19 had a long-term negative impact on public utility companies. Since t_6, the CAR has remained significantly negative for around 100 days. Although the AR turned positive on t_{150}, it quickly fell and turned positive again. Until t_{230}, a significantly negative return in the public utility industry still existed.

4.2.2. Energy Industry. As Figures 9 and 10 and Table 5 show, the energy industry had a bad situation under the impact of COVID-19. From a short-term perspective, significantly negative AR occurred on t_4; and, on the following 5 days, the ARs were all negative. As a result, the CAR became significantly negative on t_9. In the long run, the CAR showed a significantly downward trend and continued until t_{150}. Subsequently, the CAR gradually rebounded, but the CAR in the energy industry was still significantly negative on t_{230}.

4.2.3. Durable Consumer Goods and Clothing Industry. As Figures 11 and 12 and Table 6 show, the CAR of consumer durable goods and clothing industry presented a downward trend. From t_1 to t_{10}, every day presented a significantly negative abnormal return rate. The AR of the first seven days was significant at the 1% significance level, but the t-test statistics on the following days gradually declined. The CAR started to be significantly negative on t_2, showing a trend of long-term slow decline and reaching the lowest point on t_{100}, about −15%. From a long-term perspective, although the CAR tended to be positive, the consumer durable goods and clothing industry generally had a negative AR.

4.2.4. Finance Industry. As Figures 13 and 14 and Table 7show, in the short term, the impact of COVID-19 on the finance industry was not significant. From t_{-5} to t_5, the AR of finance industry did not have a significant decline. However, during the period from t_6 to t_{10}, a significantly negative AR appeared. On the following 100 days, the CAR neared zero; but it was significantly negative on t_{230}.

4.2.5. Transportation Industry. As Figures 15 and 16 and Table 8 show, on t_{-2}, there was a significantly positive gain which was presumably caused by the Spring Festival when many people went hometown to reunite. This led to a significantly positive CAR which remained until t_{10}.

With the implementation of social distancing, the volume of transportation fell sharply. Transportation industry has had a negative AR since t_{120}. On t_{230}, the CAR reached −10.34%. On the whole, the impact of COVID-19 was long-term and negative.

Table 3: AR, CAR, and *t*-statistics for food and major supplies retail.

Date	AR (%)	*t*-stat	CAR (%)	*t*-stat	Date	AR (%)	*t*-stat	CAR (%)	*t*-stat
−5	−0.81	−1.17	−0.81	−1.17	6	−0.44	−0.63	2.58	1.07
−4	−0.41	−0.58	−1.22	−1.24	7	2.26	3.25	4.84	1.93
−3	0.20	0.29	−1.02	−0.84	8	0.26	0.38	5.10	1.96
−2	−0.37	−0.54	−1.39	−1.00	9	3.18	4.58	8.28	3.08
−1	0.48	0.69	−0.91	−0.58	10	−2.38	−3.42	5.91	2.12
0	−0.27	−0.38	−1.17	−0.69	30	−0.11	−0.15	10.09	2.42
1	−1.31	−1.89	−2.49	−1.35	60	3.39	4.88	24.34	4.31
2	0.73	1.05	−1.76	−0.89	90	−1.98	−2.85	21.32	3.13
3	−0.26	−0.38	−2.02	−0.97	120	−0.20	−0.28	24.45	3.13
4	3.19	4.59	1.17	0.53	160	0.50	0.71	16.43	1.83
5	1.85	2.66	3.02	1.31	230	0.28	0.41	−1.32	−0.12

4.2.6. Information Technology Industry. As Figures 17 and 18 and Table 9 show, in the short term, the impact of COVID-19 on the information technology industry was relatively small. Although a significantly negative AR appeared on t_4 and t_7, the CAR was not significantly negative. In the medium and long term, the negative CAR was kept. On t_{120}, the negative CAR was significant.

4.2.7. Media Industry. As Figures 19 and 20 and Table 10 show, in the short term, the media industry was negatively impacted by COVID-19. On t_0 and t_3, the AR was significantly negative. The CAR continued to decline until t_{50}, and the CAR has started to become significantly negative since t_{30}. However the CAR returned to zero on t_{90}.

4.2.8. Summary. Based on the above analysis, the negative impacts of COVID-19 on different industries varied. For energy industry, because of the reduction of global demand for energy, high cost of energy saving, and speculation, the abnormal phenomenon of negative prices in crude oil futures appeared on April 20, 2020. Besides, due to the decrease of demand for some utilities, such as water and electricity, significantly negative ARs in these industries lasted for a long period.

Because of the policy of lockdown, the demand of traveling and cargo transportation dropped significantly. Because the epidemic prevented work and decreased the family income, demand for nonrigid demand products, such as durable consumer goods and clothing, also dropped significantly. Besides, IT industry faced difficulties in the recovery of domestic production lines, as well as domestic demand and foreign demand. Moreover, because companies' short-term debt paying ability declined, the quality of bank assets decreased and bad debt rates increased, and the bank's high-quality lending targets also shrunk, resulting in a credit crunch. Therefore, there were large negative impacts on durable goods and clothing, transportation, information technology, and finance industries.

In terms of media, COVID-19 had different impacts on offline and online media. Offline theaters were shut down and filming stalled, making the media industry face difficulties; however, the online media benefited from the increased number of people moving to home entertainment, driving negative abnormal return of the media industry.

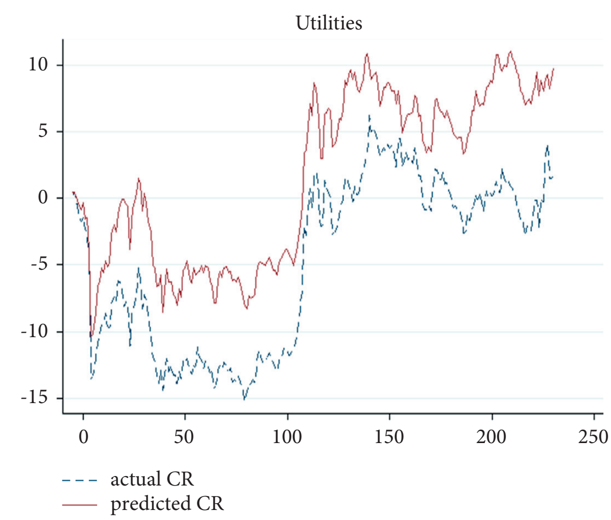


Figure 7: Actual and predicted cumulative returns.

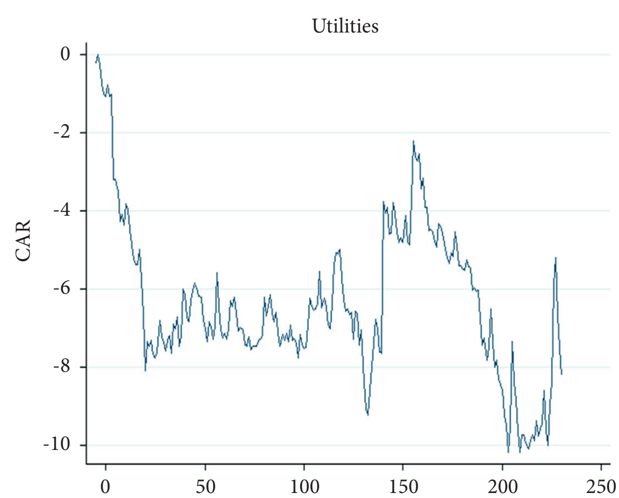


Figure 8: Cumulative abnormal return.

4.3. Industries without Significant Abnormal Return

4.3.1. Real Estate Industry. As Figures 21 and 22 and Table 11 show, real estate was not greatly affected by COVID-19. Within 120 days after the event, neither the AR nor the

Table 4: AR, CAR, and t-statistics for public utilities.

Date	AR (%)	t-stat	CAR (%)	t-stat	Date	AR (%)	t-stat	CAR (%)	t-stat
−5	−0.22	−0.39	−0.22	−0.39	6	−0.30	−0.52	−3.48	−1.74
−4	0.22	0.38	−0.01	−0.01	7	−0.77	−1.33	−4.26	−2.04
−3	−0.25	−0.43	−0.25	−0.25	8	0.18	0.32	−4.07	−1.88
−2	−0.50	−0.87	−0.76	−0.66	9	−0.27	−0.47	−4.34	−1.94
−1	−0.26	−0.44	−1.01	−0.79	10	0.52	0.90	−3.82	−1.65
0	−0.07	−0.11	−1.08	−0.76	30	−0.19	−0.32	−7.56	−2.18
1	0.29	0.51	−0.79	−0.51	60	0.10	0.17	−7.13	−1.52
2	−0.29	−0.50	−1.07	−0.66	90	−0.16	−0.27	−7.32	−1.29
3	0.07	0.12	−1.00	−0.58	120	−0.55	−0.94	−6.28	−0.97
4	−2.19	−3.79	−3.20	−1.75	160	0.26	0.44	−3.16	−0.42
5	0.01	0.02	−3.18	−1.66	230	−0.73	−1.27	−8.20	−0.92

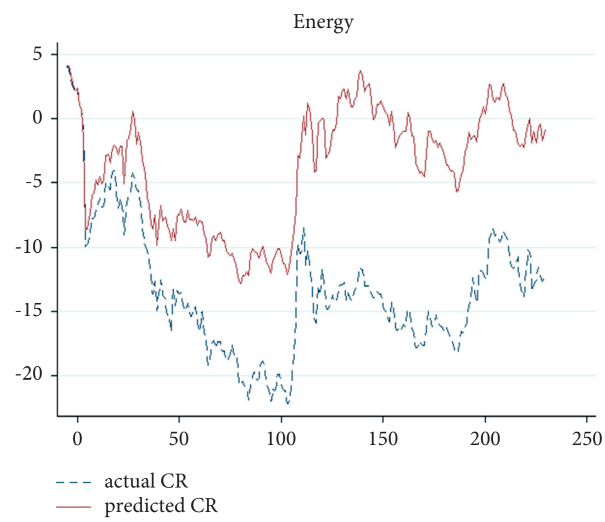


Figure 9: Actual and predicted cumulative returns.

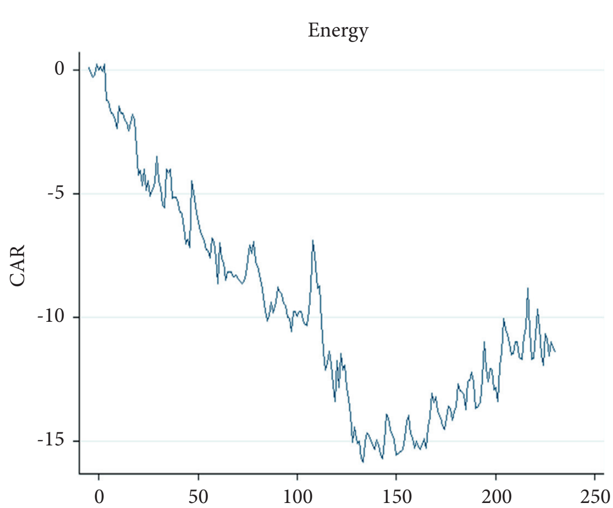


Figure 10: Cumulative abnormal return.

CAR presented significant fluctuation. The large negative AR that started to appear on t_{150} may be related to the government's macro adjustment of housing prices, not directly caused by COVID-19.

4.3.2. Telecommunications Industry. As Figurse 23 and 24 and Table 12 show, COVID-19 did not have an obvious impact on the telecommunications industry. Within 100 days after the event day, neither AR nor CAR was significant; but the CAR dropped largely after t_{100}.

4.3.3. Materials Industry. As Figures 25 and 26 and Table 13 show, the impact of COVID-19 on the materials industry was relatively small. Within 160 days after the incident, the CAR and AR were not significant on most of the days. Only t_{10} had a significantly positive abnormal return of 0.85%, but the CAR at that time was not significant. However, the CAR started to rise quickly around t_{170}.

4.3.4. Summary. Real estate, telecommunications, and materials did not have a significant AR in a short term after the outbreak of COVID-19. Although there were significant ARs after t_{100} for these industries, their occurrences were far from the event date. Therefore, we argue that the impacts of COVID-19 on these industries were limited relatively. As for real estate, even though the lockdown reduced the demand for office space, the real estate stock's return did not drop significantly in the short term after the outbreak. In terms of telecommunications, it is speculated that the increase of the demand for online services compensated the negative impact of COVID-19, making the return of telecommunication industry stable. What is more, although the terminal manufacturing industry was hit hard, the material industry did not suffer a huge impact in the sample period. In the long run, benefiting from the recovery of the world economy, the materials industry gradually recovered and showed a significantly positive AR.

5. The Impact of COVID-19 on the Turnover Rate of A-Shares in Various Industries

Liquidity determines whether one trade can be made shortly at a low cost and therefore influences the value of stocks. We here use turnover rate to represent the liquidity of stock to investigate the impact of COVID-19.

As Figure 27 shows, among the 12 selected industries, the turnover rates of 10 industries, healthcare, real estate, materials, utilities, telecommunications services, durable

Table 5: AR, CAR, and t-statistics for energy industry.

Date	AR (%)	t-stat	CAR (%)	t-stat	Date	AR (%)	t-stat	CAR (%)	t-stat
−5	−0.28	−0.71	−0.28	−0.71	6	−0.48	−1.20	−2.15	−1.56
−4	−0.20	−0.49	−0.48	−0.85	7	−0.05	−0.12	−2.20	−1.53
−3	−0.21	−0.53	−0.69	−1.00	8	−0.16	−0.41	−2.36	−1.58
−2	0.10	0.25	−0.59	−0.74	9	−0.37	−0.92	−2.73	−1.77
−1	0.45	1.12	−0.14	−0.16	10	0.87	2.17	−1.86	−1.17
0	−0.23	−0.58	−0.38	−0.39	30	−0.96	−2.42	−4.85	−2.02
1	0.13	0.32	−0.25	−0.24	60	−1.07	−2.67	−9.04	−2.79
2	−0.19	−0.48	−0.44	−0.39	90	0.64	1.61	−9.14	−2.34
3	0.28	0.71	−0.16	−0.13	120	1.65	4.14	−12.13	−2.71
4	−1.46	−3.65	−1.61	−1.28	160	0.25	0.63	−15.40	−2.99
5	−0.06	−0.16	−1.67	−1.26	230	−0.22	−0.56	−11.79	−1.92

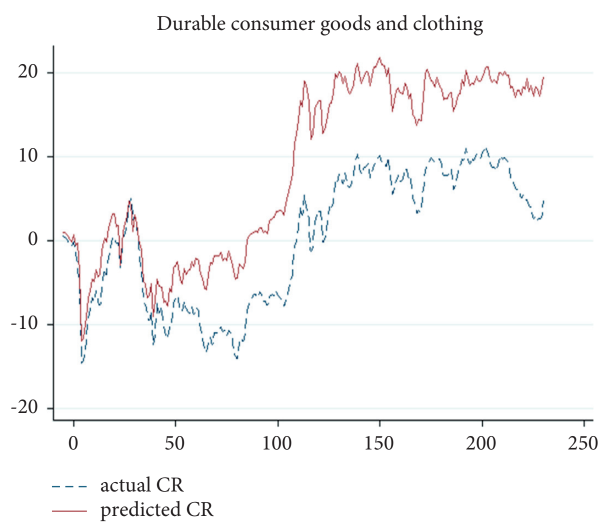


Figure 11: Actual and predicted cumulative returns.

Figure 12: Cumulative abnormal return.

consumer goods and clothing, transportation, media, information technology, and food and major supplies retail, increased significantly after the event date. Among them, the abnormal turnover rates of medical care, materials, utilities, durable consumer goods and clothing, and transportation industries maintained a growth trend during the event window.

For real estate, telecommunications services, media, information technology, and food and major supplies retail industries, the abnormal cumulative turnover rates dropped to some extent on t_{150}. Among them, only the abnormal cumulative turnover rate of the retail sales of food and major supplies dropped to become negative at the end of the event window. In addition, the abnormal cumulative turnover rate of energy and finance declined after event day. However, after the abnormal turnover rate suddenly increased on t_{100}, that of finance directly returned to the normal level, and that of energy also had a significant recovery.

Table 6: AR, CAR, and t-statistics for consumer durable goods and clothing industry.

Date	AR (%)	t-stat	CAR (%)	t-stat	Date	AR (%)	t-stat	CAR (%)	t-stat
−5	−0.47	−1.42	−0.47	−1.42	6	0.31	0.96	−2.54	−2.24
−4	−0.04	−0.11	−0.50	−1.08	7	0.01	0.03	−2.53	−2.14
−3	−0.04	−0.12	−0.54	−0.95	8	0.53	1.62	−2.00	−1.63
−2	−0.05	−0.14	−0.59	−0.89	9	−0.23	−0.72	−2.23	−1.76
−1	−0.10	−0.29	−0.68	−0.93	10	−0.22	−0.66	−2.45	−1.87
0	−0.22	−0.68	−0.91	−1.13	30	−1.41	−4.32	−0.08	−0.04
1	−0.20	−0.61	−1.11	−1.28	60	−0.14	−0.44	−5.97	−2.24
2	−1.38	−4.22	−2.49	−2.69	90	−0.44	−1.33	−7.92	−2.47
3	−0.04	−0.13	−2.53	−2.58	120	−0.10	−0.30	−13.22	−3.59
4	−0.12	−0.36	−2.65	−2.56	160	−0.16	−0.49	−10.58	−2.51
5	−0.20	−0.62	−2.85	−2.63	230	0.67	2.05	−14.89	−2.96

Figure 13: Actual and predicted cumulative returns.

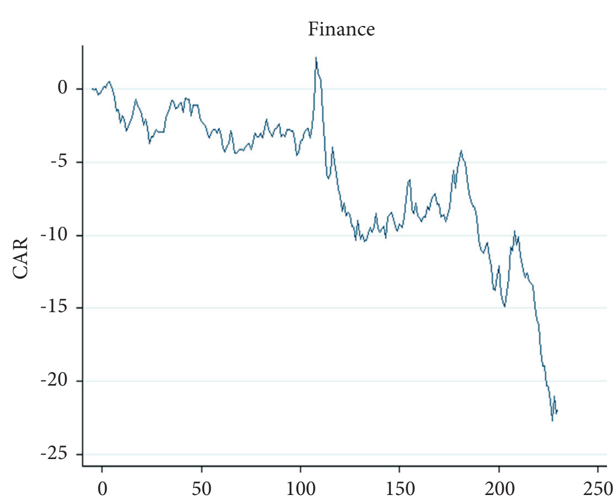


Figure 14: Cumulative abnormal return.

Table 7: AR, CAR, and t-statistics for finance industry.

Date	AR (%)	t-stat	CAR (%)	t-stat	Date	AR (%)	t-stat	CAR (%)	t-stat
−5	0.05	0.10	0.05	0.10	6	−0.79	−1.40	−0.57	−0.29
−4	−0.07	−0.13	−0.02	−0.02	7	−0.91	−1.62	−1.48	−0.73
−3	0.07	0.12	0.05	0.05	8	0.08	0.14	−1.41	−0.67
−2	−0.44	−0.78	−0.39	−0.34	9	−0.87	−1.55	−2.28	−1.04
−1	0.11	0.19	−0.28	−0.22	10	0.44	0.79	−1.83	−0.81
0	0.19	0.35	−0.09	−0.06	30	0.01	0.02	−2.89	−0.85
1	0.24	0.43	0.16	0.11	60	−0.47	−0.84	−3.17	−0.69
2	−0.05	−0.09	0.11	0.07	90	−0.89	−1.57	−3.21	−0.58
3	0.30	0.53	0.41	0.24	120	−0.48	−0.85	−7.34	−1.16
4	0.06	0.11	0.47	0.26	160	−0.22	−0.40	−8.86	−1.22
5	−0.25	−0.45	0.22	0.12	230	0.31	0.56	−21.94	−2.53

Figure 15: Actual and predicted cumulative returns.

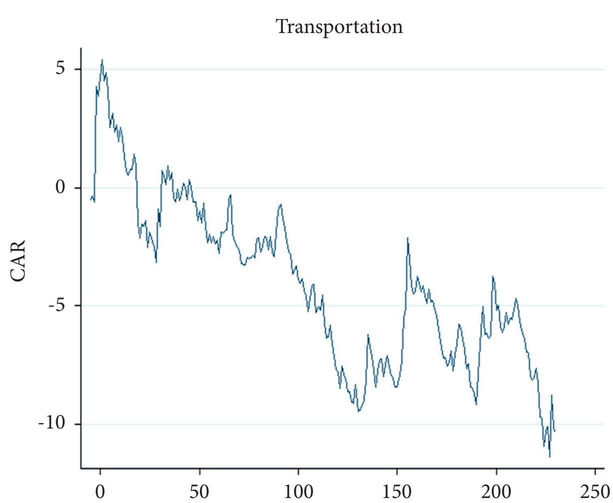


Figure 16: Cumulative abnormal return.

Table 8: AR, CAR, and t-statistics for transportation industry.

Date	AR (%)	t-stat	CAR (%)	t-stat	Date	AR (%)	t-stat	CAR (%)	t-stat
−5	−0.52	−0.94	−0.52	−0.94	6	0.62	1.13	3.14	1.65
−4	0.15	0.27	−0.37	−0.47	7	−0.81	−1.48	2.33	1.18
−3	−0.25	−0.45	−0.61	−0.64	8	0.31	0.56	2.64	1.28
−2	4.88	8.88	4.27	3.88	9	−0.71	−1.30	1.93	0.91
−1	−0.42	−0.76	3.85	3.14	10	0.63	1.14	2.55	1.16
0	0.79	1.44	4.64	3.45	30	−0.77	−1.40	−1.66	−0.50
1	0.75	1.37	5.40	3.72	60	−0.51	−0.92	−2.76	−0.62
2	−0.89	−1.61	4.51	2.90	90	0.58	1.05	−0.83	−0.15
3	0.32	0.58	4.83	2.93	120	−0.10	−0.18	−7.76	−1.26
4	−0.70	−1.28	4.13	2.38	160	0.62	1.13	−3.76	−0.53
5	−1.61	−2.93	2.52	1.38	230	−0.18	−0.32	−10.34	−1.23

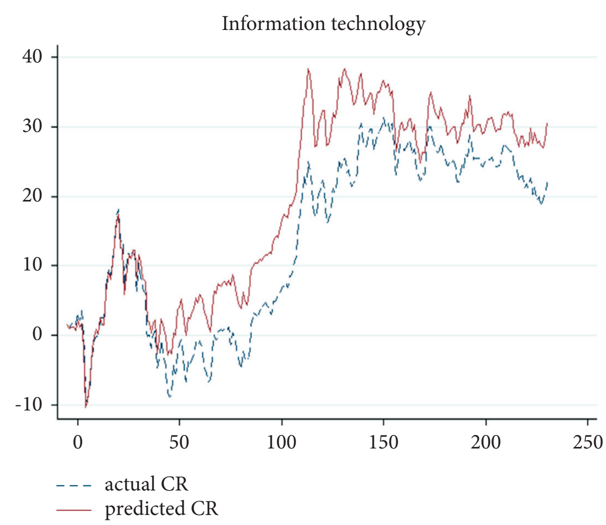


Figure 17: Actual and predicted cumulative returns.

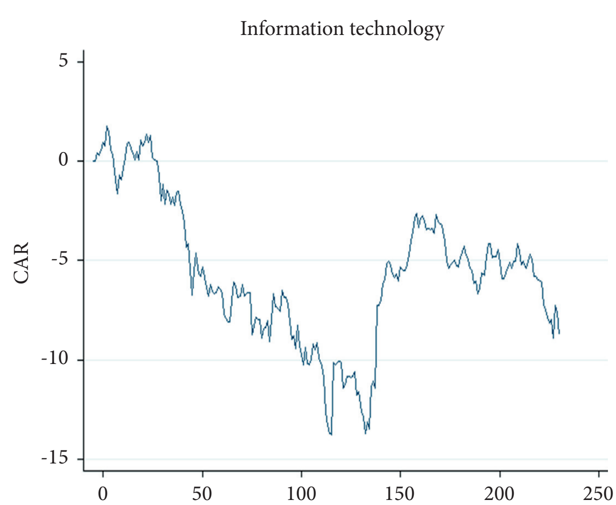


Figure 18: Cumulative abnormal return.

Table 9: AR, CAR, and t-statistics for information technology industry.

Date	AR (%)	t-stat	CAR (%)	t-stat	Date	AR (%)	t-stat	CAR (%)	t-stat
−5	−0.02	−0.05	−0.02	−0.05	6	−1.28	−2.62	−0.91	−0.54
−4	0.03	0.06	0.01	0.01	7	−0.71	−1.44	−1.61	−0.92
−3	0.41	0.83	0.41	0.49	8	0.93	1.91	−0.68	−0.37
−2	−0.10	−0.20	0.31	0.32	9	−0.23	−0.47	−0.91	−0.48
−1	0.22	0.44	0.53	0.48	10	0.59	1.20	−0.33	−0.17
0	0.44	0.90	0.97	0.81	30	0.81	1.66	−1.15	−0.39
1	−0.20	−0.41	0.77	0.59	60	−0.15	−0.30	−6.61	−1.66
2	1.02	2.08	1.78	1.29	90	1.03	2.10	−6.51	−1.36
3	−0.27	−0.55	1.51	1.03	120	−0.03	−0.06	−10.14	−1.85
4	−0.95	−1.94	0.56	0.36	160	0.45	0.91	−2.90	−0.46
5	−0.19	−0.38	0.37	0.23	230	−1.08	−2.20	−8.73	−1.16

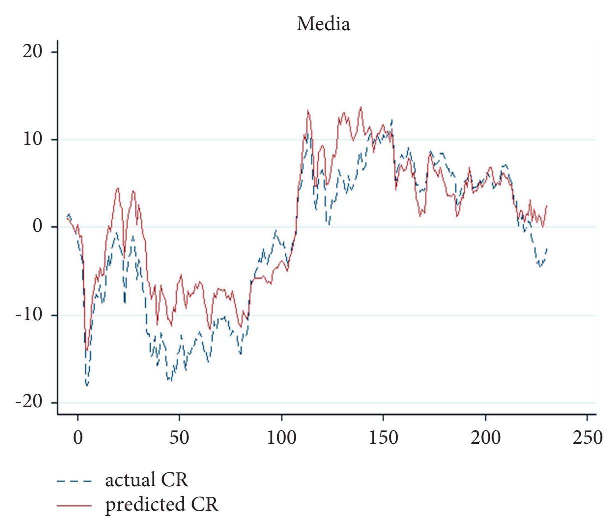


Figure 19: Actual and predicted cumulative returns.

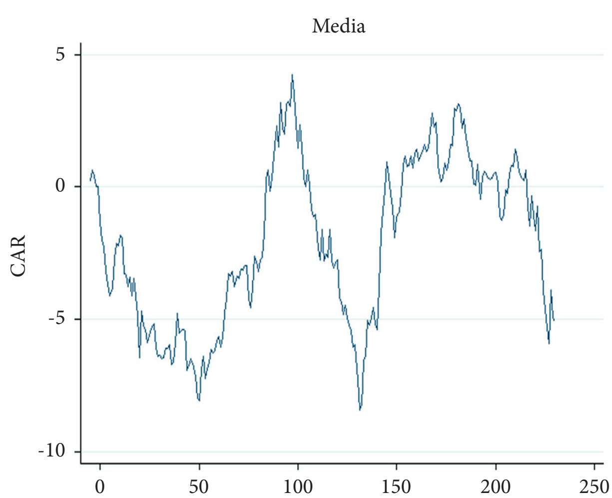


Figure 20: Cumulative abnormal return.

Table 10: AR, CAR, and t-statistics for media industry.

Date	AR (%)	t-stat	CAR (%)	t-stat	Date	AR (%)	t-stat	CAR (%)	t-stat
−5	0.25	0.43	0.25	0.43	6	0.31	0.53	−3.79	−1.86
−4	0.37	0.62	0.62	0.75	7	1.07	1.82	−2.72	−1.28
−3	−0.20	−0.34	0.42	0.41	8	0.56	0.95	−2.16	−0.98
−2	−0.40	−0.69	0.02	0.01	9	−0.07	−0.11	−2.22	−0.97
−1	0.02	0.03	0.04	0.03	10	0.38	0.65	−1.84	−0.78
0	−1.39	−2.35	−1.35	−0.94	30	0.07	0.12	−6.34	−1.79
1	−0.65	−1.10	−2.00	−1.28	60	0.15	0.25	−5.64	−1.18
2	−0.26	−0.44	−2.26	−1.35	90	−0.80	−1.36	1.52	0.26
3	−0.92	−1.56	−3.18	−1.80	120	0.09	0.16	−2.76	−0.42
4	−0.51	−0.87	−3.69	−1.98	160	0.16	0.28	1.44	0.19
5	−0.41	−0.69	−4.10	−2.10	230	−0.19	−0.32	−5.10	−0.56

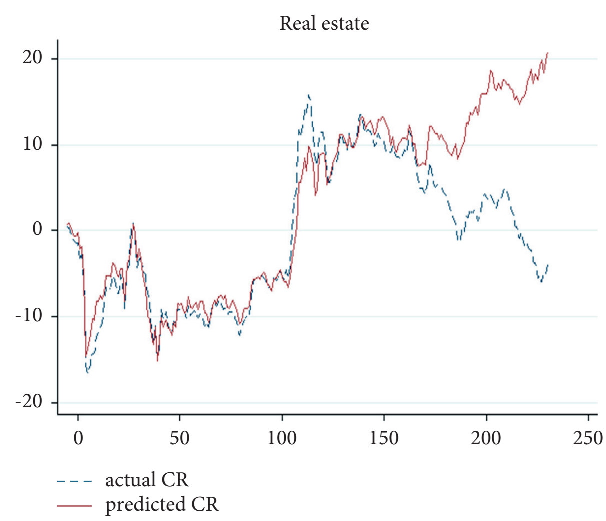


Figure 21: Actual and predicted cumulative returns.

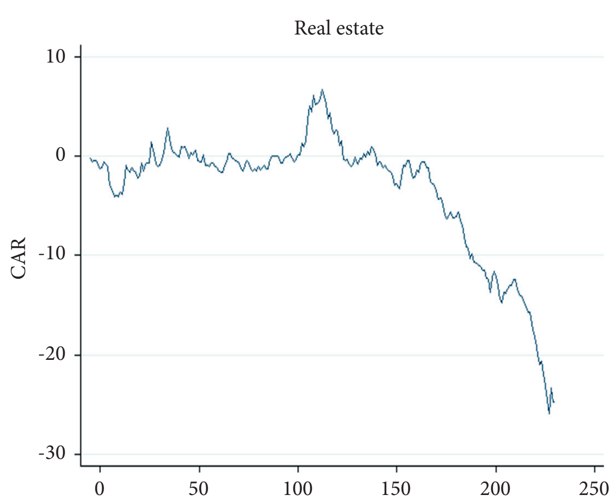


Figure 22: Cumulative abnormal return.

Table 11: AR, CAR, and t-statistics for real estate industry.

Date	AR (%)	t-stat	CAR (%)	t-stat	Date	AR (%)	t-stat	CAR (%)	t-stat
−5	−0.26	−0.54	−0.26	−0.54	6	−0.62	−1.30	−3.52	−2.13
−4	−0.27	−0.57	−0.53	−0.79	7	−0.59	−1.23	−4.11	−2.39
−3	0.01	0.03	−0.52	−0.63	8	0.13	0.27	−3.98	−2.23
−2	0.03	0.06	−0.49	−0.51	9	−0.17	−0.35	−4.15	−2.25
−1	−0.30	−0.62	−0.78	−0.73	10	0.54	1.12	−3.62	−1.89
0	−0.48	−1.00	−1.26	−1.08	30	0.11	0.23	−0.88	−0.31
1	0.11	0.23	−1.15	−0.91	60	−0.35	−0.73	−1.53	−0.40
2	0.52	1.08	−0.64	−0.47	90	−0.12	−0.25	−0.11	−0.02
3	−0.18	−0.38	−0.82	−0.57	120	−0.14	−0.29	2.43	0.45
4	−0.24	−0.50	−1.06	−0.70	160	0.70	1.47	−1.37	−0.22
5	−1.84	−3.87	−2.90	−1.83	230	−0.03	−0.05	−24.80	−3.38

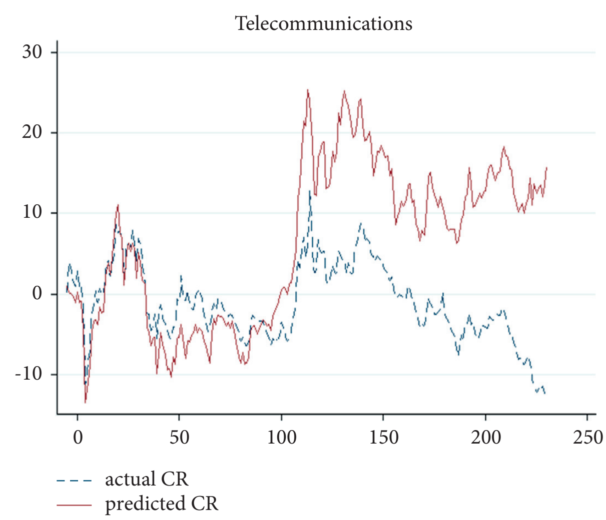


Figure 23: Actual and predicted cumulative returns.

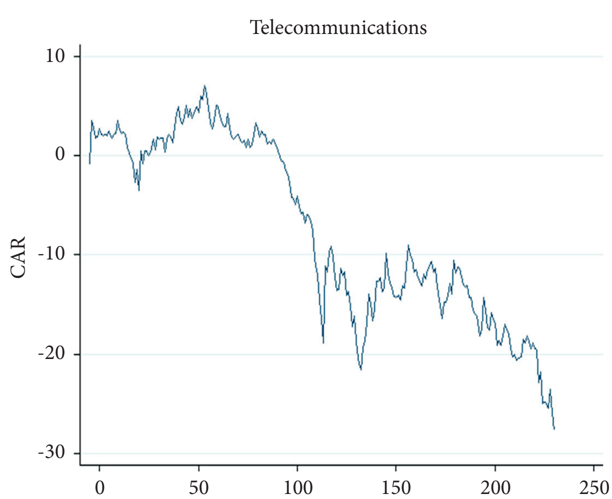


Figure 24: Cumulative abnormal return.

Table 12: AR, CAR, and t-statistics for the telecommunications industry.

Date	AR (%)	t-stat	CAR (%)	t-stat	Date	AR (%)	t-stat	CAR (%)	t-stat
−5	−0.78	−0.54	−0.78	−0.54	6	−0.69	−0.48	1.81	0.36
−4	4.35	3.00	3.57	1.74	7	0.28	0.19	2.08	0.40
−3	−0.73	−0.51	2.84	1.13	8	0.19	0.13	2.27	0.42
−2	−1.03	−0.71	1.81	0.62	9	1.24	0.86	3.51	0.63
−1	0.06	0.04	1.86	0.58	10	−0.92	−0.64	2.59	0.45
0	0.83	0.57	2.69	0.76	30	−0.14	−0.10	1.70	0.20
1	−0.61	−0.42	2.09	0.54	60	−0.19	−0.13	4.90	0.42
2	−0.04	−0.03	2.05	0.50	90	−0.64	−0.44	0.45	0.03
3	0.12	0.08	2.17	0.50	120	−1.30	−0.90	−13.56	−0.83
4	−0.20	−0.14	1.97	0.43	160	0.24	0.16	−11.44	−0.61
5	0.53	0.37	2.50	0.52	230	−1.81	−1.25	−27.72	−1.25

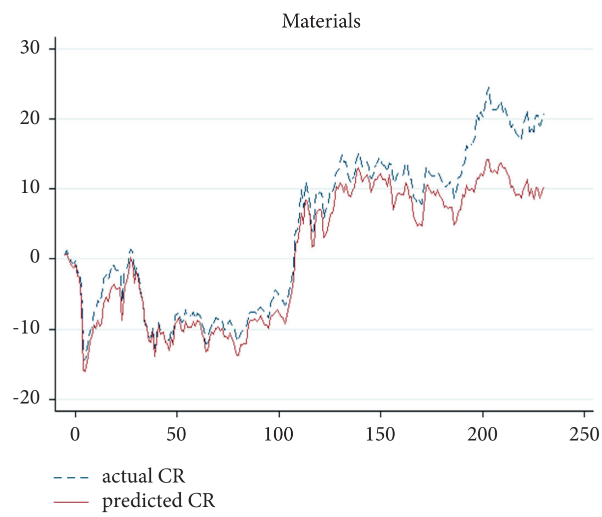


Figure 25: Actual and predicted cumulative returns.

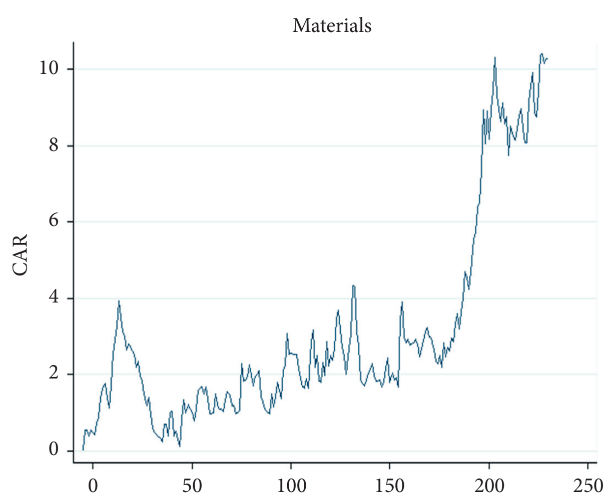


Figure 26: Cumulative abnormal return.

Table 13: AR, CAR, and t-statistics for materials industry.

Date	AR (%)	t-stat	CAR (%)	t-stat	Date	AR (%)	t-stat	CAR (%)	t-stat
−5	0.03	0.08	0.03	0.08	6	0.14	0.31	1.77	1.17
−4	0.50	1.14	0.53	0.86	7	−0.37	−0.84	1.41	0.89
−3	0.01	0.02	0.54	0.71	8	−0.27	−0.62	1.14	0.70
−2	−0.16	−0.36	0.38	0.44	9	0.58	1.32	1.71	1.01
−1	0.15	0.35	0.54	0.55	10	0.85	1.94	2.56	1.47
0	−0.04	−0.09	0.50	0.47	30	−0.35	−0.80	0.59	0.23
1	−0.07	−0.17	0.43	0.37	60	−0.01	−0.03	0.97	0.27
2	0.29	0.66	0.71	0.58	90	0.54	1.24	1.50	0.35
3	0.17	0.40	0.89	0.68	120	0.28	0.63	2.50	0.51
4	0.47	1.07	1.36	0.98	160	−0.15	−0.35	2.76	0.49
5	0.28	0.64	1.64	1.13	230	−0.01	−0.01	10.27	1.53

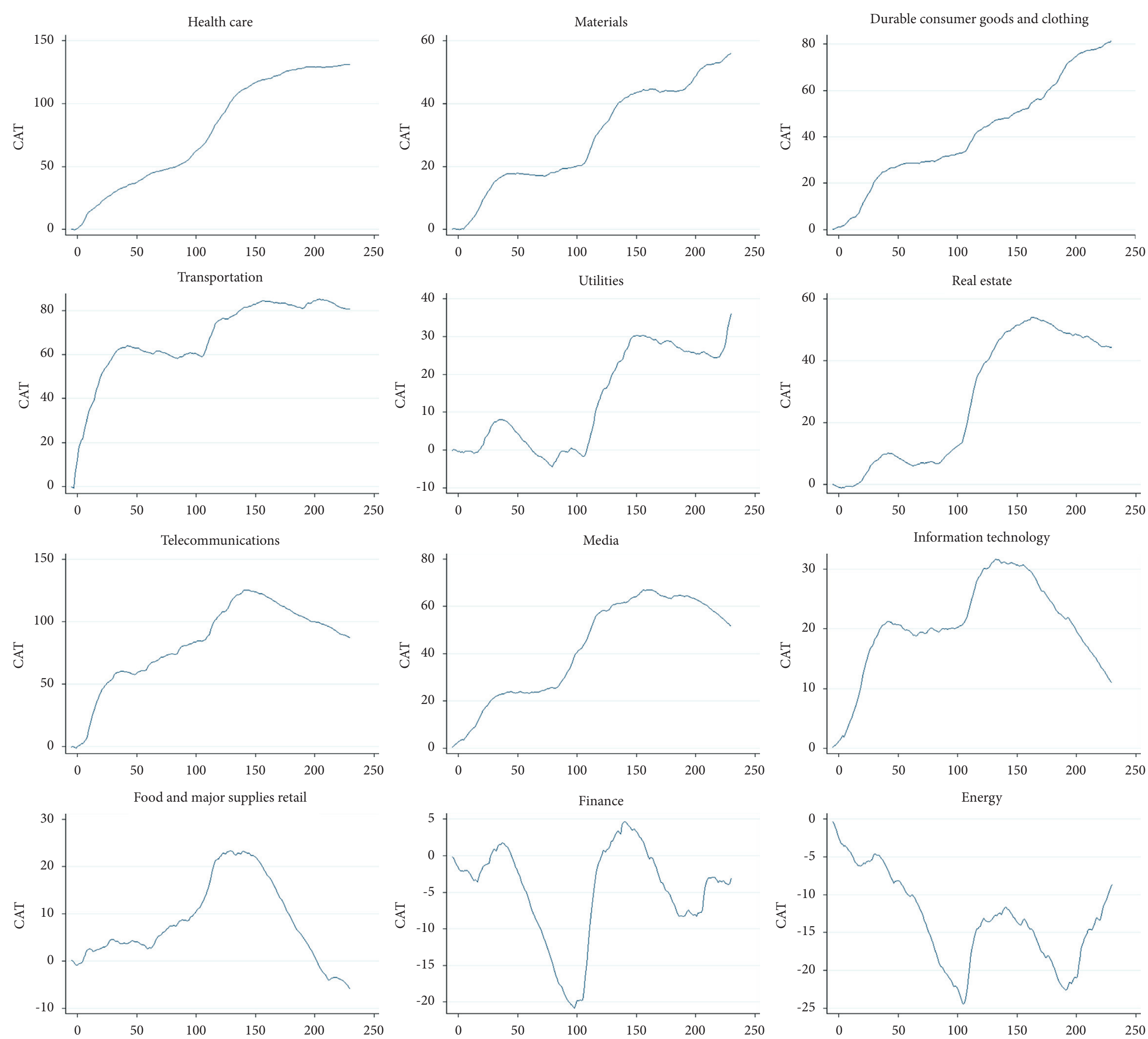


Figure 27: The abnormal turnover rates of industries.

6. Conclusion

In the short term, the actual return on A-stock was significantly lower than the expected return after the outbreak of COVID-19. After suffering two waves of shock, the return of A-stock came normal on May 2020. The medical and food industries benefited from the influence of COVID-19. However, energy and utility suffered the negative impact of COVID-19 for a long time; the negative impact on the durable consumer goods and clothing, finance, transportation, and information technology industries did not appear to be significant until several weeks after the event date. The industries of real estate, telecommunications, and materials were not influenced relatively by COVID-19 too much. In order to hedge the various impacts of the epidemic on A-stock and different industries, a more proactive fiscal policy and a flexible monetary policy from government may be helpful.

In terms of turnover rates, the energy and finance industries had significantly negative abnormal turnover rates after the outbreak of the pandemic; moreover, the cumulative abnormal turnover rates of real estate, telecommunications service, clothes, media, information technology, and food and major products started to drop after reaching the highest point on t_{150}; but those of healthcare, materials, durable consumer goods and clothing, transportation, and utilities kept increasing in the event window.

Data Availability

The data are from China Stock Market & Accounting Research (CMSAR), Yahoo Finance.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] L. G. Del, B. Théophile, and S. Babacar, "COVID-19 and the African financial markets: less infection, less economic impact," *Finance Research Letters*, Article ID 1021481544, 6123 pages, 2021, In press.

[2] A. Zaremba, R. Kizys, P. Tzouvanas, Y. D. Aharon, and E. Demir, "The quest for multidimensional financial immunity to the COVID-19 pandemic: evidence from international stock markets," *Journal of International Financial Markets, Institutions and Money*, vol. 71, Article ID 101284, 2021.

[3] B. N. Ashraf, "Economic impact of government interventions during the COVID-19 pandemic: international evidence from financial markets," *Journal of Behavioral and Experimental Finance*, vol. 27, Article ID 100371, 2020.

[4] M. Costola, M. Iacopini, and C. R. M. A. Santagiustina, "Google search volumes and the financial markets during the COVID-19 outbreak," *Finance Research Letters*, vol. 42, Article ID 101884, 2020.

[5] O. Haroon and S. A. R. Rizvi, "COVID-19: media coverage and financial markets behavior—a sectoral inquiry," *Journal of Behavioral and Experimental Finance*, vol. 27, Article ID 100343, 2020.

[6] S. S. Wang, J. R. Goh, D. Sornette, H. Wang, and E. Y. Yang, "Government support for SMEs in response to COVID-19: theoretical model using Wang transform," *China Finance Review International*, vol. 11, no. 3, pp. 406–433, 2021.

[7] X. Wei and L. Han, "The impact of COVID-19 pandemic on transmission of monetary policy to financial markets," *International Review of Financial Analysis*, vol. 74, Article ID 101705, 2021.

[8] Q. Ji, D. Zhang, and Y. Zhao, "Searching for safe-haven assets during the COVID-19 pandemic," *International Review of Financial Analysis*, vol. 71, Article ID 101526, 2020.

[9] M. Ali, N. Alam, and S. A. R. Rizvi, "Coronavirus (COVID-19) — an epidemic or pandemic for financial markets," *Journal of Behavioral and Experimental Finance*, vol. 27, Article ID 100341, 2020.

[10] M. Mazur, M. Dang, and M. Vega, "COVID-19 and the march 2020 stock market crash. Evidence from S&P1500," *Finance Research Letters*, vol. 38, Article ID 101690, 2021.

[11] O. A. . Sayed and H. Eledum, "The short-run response of Saudi Arabia stock market to the outbreak of COVID-19 pandemic: an event-study methodology," *International Journal of Finance & Economics*, 2021, In press.

[12] S. Corbet, Y. Hou, Y. Hu, L. Oxley, and D. Xu, "Pandemic-related financial market volatility spillovers: evidence from the Chinese COVID-19 epicentre," *International Review of Economics & Finance*, vol. 71, pp. 55–81, 2021.

[13] P.-F. Dai, X. Xiong, Z. Liu, and T. L. D. Huynh, "Preventing crash in stock market: the role of economic policy uncertainty during COVID-19," *Financial Innovation*, vol. 7, p. 31, 2021.

[14] Z. Liu, T. L. D. Huynh, and P. F. Dai, "The impact of COVID-19 on the stock market crash risk in China," *Research in International Business and Finance*, vol. 57, Article ID 101419, 2021.

[15] T. Mezghani, M. Boujelbène, and M. Elbayar, "Impact of COVID-19 pandemic on risk transmission between googling investor's sentiment, the Chinese stock and bond markets," *China Finance Review International*, vol. 11, no. 3, pp. 322–348, 2021.

[16] Q. T. T. Nguyen, D. L. T. Anh, and C. Gan, "Epidemics and Chinese firms' stock returns: is COVID-19 different?" *China Finance Review International*, vol. 11, no. 3, pp. 302–321, 2021.

[17] H. Liu, Y. Wang, D. He, and C. Wang, "Short term response of Chinese stock markets to the outbreak of COVID-19," *Applied Economics*, vol. 52, no. 53, pp. 5859–5872, 2020.

[18] E. F. Fama and K. R. French, "A five-factor asset pricing model," *Journal of Financial Economics*, vol. 116, no. 1, pp. 1–22, 2015.

[19] T. Dyckman, D. Philbrick, and J. Stephan, "A comparison of event study methodologies using daily stock returns: a simulation approach," *Journal of Accounting Research*, vol. 22, pp. 1–30, 1984.

[20] D. K. Pandey and V. Kumari, "Event study on the reaction of the developed and emerging stock markets to the 2019-nCoV outbreak," *International Review of Economics & Finance*, vol. 71, pp. 467–483, 2021.

[21] F. L. Elad, "Event study on the reaction of stock returns to acquisition news," *International Finance and Banking*, vol. 4, no. 1, pp. 33–43, 2017.

[22] T. Huang, "Is the Fama and French five-factor model robust in the Chinese stock market?" *Asia Pacific Management Review*, vol. 24, no. 3, pp. 278–289, 2019.

[23] B. Liang, "Price pressure: evidence from the "Dartboard" column," *Journal of Business*, vol. 72, pp. 119–134, 1999.

[24] R. Michaely, R. H. Thaler, and K. L. Womack, "Price reactions to dividend initiations and omissions: overreaction or drift?" *The Journal of Finance*, vol. 50, pp. 573–608, 1995.

[25] B. M. Barber and J. D. Lyon, "Detecting abnormal operating performance: the empirical power and specification of test statistics," *Journal of Financial Economics*, vol. 41, no. 3, pp. 359–399, 1996.

[26] S. J. Brown and J. B. Warner, "Using daily stock returns," *Journal of Financial Economics*, vol. 14, no. 1, pp. 3–31, 1985.

Hindawi
Complexity
Volume 2022, Article ID 3139870, 8 pages
https://doi.org/10.1155/2022/3139870

Research Article

Assessing the Probability of Drought Severity in a Homogeneous Region

Rizwan Niaz [1] **Mohammed M. A. Almazah** [2,3] **Ijaz Hussain** [1] **Joao Dehon Pontes Filho**,[4] **Nadhir Al-Ansari** [5] **and Saad Sh Sammen** [6]

[1]*Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan*
[2]*Department of Mathematics, College of Sciences and Arts (Muhyil), King Khalid University, Abha 61421, Saudi Arabia*
[3]*Department of Mathematics and Computer, College of Sciences, Ibb University, Ibb 70270, Yemen*
[4]*Hydraulic and Environmental Engineering Department (DEHA), Federal University of Ceará, Fortaleza, Ceará 60020-181, Brazil*
[5]*Department of Civil, Environmental and Natural Resources Engineering, Lulea University of Technology, Lulea 971 87, Sweden*
[6]*Department of Civil Engineering, College of Engineering, University of Diyala, Diyala Governorate, Baqubah, Iraq*

Correspondence should be addressed to Saad Sh Sammen; saad123engineer@yahoo.com

Received 16 October 2021; Revised 21 December 2021; Accepted 6 January 2022; Published 29 January 2022

Academic Editor: Qingyuan Zhu

The standardized precipitation index (SPI) is one of the most widely used indices for characterizing and monitoring drought in various regions. SPI's applicability has regional and time-scale constraints when it observes in several homogeneous climatic regions with similar characteristics. It also does not provide sufficient knowledge about precipitation deficits and the spatiotemporal evolution of drought. Therefore, a new method, the regional spatially agglomerative continuous drought probability monitoring system (RSACDPMS), is proposed to obtain spatiotemporal information and monitor drought characteristics more expeditiously. The proposed framework uses spatially agglomerative precipitation (SAP) and copulas' functions to continuously monitor the drought probability in the homogenous region. The RSACDPMS is validated in the region of the Northern area of Pakistan. The outcomes of the current study provide a better quantitative way to obtain appropriate information about precipitation deficits and the spatiotemporal evolution of drought.

1. Introduction

Drought is a creeping phenomenon that gradually spreads in an area over a period and may continue for a long period [1–4]. It is a multifaceted phenomenon that periodically fluctuates in many regions worldwide and becomes a cause of negative impacts [5, 6]. Its impacts directly or indirectly affect humans' activities more than other natural hazards [3, 4, 7, 8]. Furthermore, it negatively affects vast areas in several ways, such as distressing the region's economy, recreation, water resources, waterfowl, hydroelectric energy, forestry, and other environmental locations [9–11]. However, suitable drought monitoring measures can assist to reduce the negative impacts of drought in the affected area regarding the needs such as the region's economy, food, water resources, hydroelectric energy, and social security of the people. Although drought is primarily a water-associated hazard, it has been assessed and defined by several fields from various perspectives [12–14]. To bring some order to measure drought that substantially helps in investigating the various effects, the climatological scientists grouped the drought into four major types [15, 16]. These definitions are associated with a shortage of precipitation over time [12–14].

The standardized procedures are commonly used for drought assessment and its characterization [17]. Numerous studies proposed several standardized drought indices for characterizing and monitoring drought [18, 20]. The indices provide quantitative information about drought monitoring forecasting that helps decision makers with drought characteristics. Over the years, several drought indices have been developed for drought assessment. However, due to the complexity of the drought phenomena, several researchers proposed various drought indices to the specific climatic conditions. Some of the indices are used in specific regions and have some limitations of use under specific climatic conditions. For instance, Palmer [21] provided an index which is known as "Palmer drought severity index (PDSI)" and is extensively used in the United States. Gibbs and Maher [22] developed an index which is known as "Rainfall deciles as drought indicators (RDDIs) and functioning" in Australia. The China *Z* index (CZI) was developed in China and is frequently used in China to identify and monitor drought events. However, among the various standardized indices, the standardized precipitation index (SPI), developed by McKee et al. [23], is frequently used. The SPI uses precipitation data to assess meteorological drought on various time scales. It prevails in developing drought monitoring and mitigating policies [24, 25]. Furthermore, the standardized values of the SPI, therefore, can be used to compare in different climatic zones [13, 19, 26, 27]. The SPI has been frequently considered to identify or envisage drought events; however, standardized numerical values of the index become challenging when associated with the precipitation deficits and the temporal evolution of droughts [28, 29]. Therefore, it is important to develop a new method that provides information about the precipitation deficits, drought probability, and temporal evolution of the droughts.

Usually, drought occurrences are identified by the univariate setting. However, the climatological characteristics of hydrological phenomena consist of the dependence structure. The univariate setting could not be able to consider the dependence structure. Since univariate approaches cannot perform well and become insufficient in hydrological phenomena [30]. Therefore, multivariate techniques were developed to address the dependence structure of the characteristics to enhance the efficiency of the estimates in hydrological studies. These techniques have several shortcomings; for example, these techniques do not explicitly model more than two variables. These techniques are also based on the condition that marginal must have a similar probability distribution that restricts them from illustrating other individual dependence structures. The research is therefore needed to develop techniques that explicitly describe the dependence structure to overcome these issues. In this regard, the concept of copula-based modeling was introduced that has some flexibility. For instance, copula-based modeling extensively assesses a multivariate dependence structure and joint distributions through mainly categorizing the dependence structure of random variables from their marginal distributions [31]. The copulas and their applications in drought have achieved significant importance for joint modeling of drought indices [32]. Furthermore, copulas' functions were used to evaluate varying characteristics of the drought events (e.g., duration, magnitude, intensity, and spatial distribution). In climatic regions such as the Northern area of Pakistan, the major source of drought events is insufficient precipitation during the rainy season. An advanced copula-based methodology is employed to observe the advancement in drought probability by using continuous precipitation information.

The present study proposes the Regional Spatially Agglomerative Continuous Drought Probability Monitoring System (RSACDPMS) to calculate more accurate and comprehensive information about homogeneous stations and monitor regional drought characteristics more expeditiously. The RSACDPMS utilizes spatially agglomerative precipitation (SAP) and copulas' functions to continuously monitor the drought probability in the homogeneous region. The proposed framework is validated on six meteorological stations in the Northern area of Pakistan. The study provides a better quantitative way of analyzing drought at the regional level.

2. Methods

2.1. Copula's Theory. The predictive relationship between two or more than two variables can be analyzed by studying their dependence structure. Mostly, the relationship among the variables can be analyzed using the Pearson correlation coefficient. However, this method cannot consider structural dependence but the degree of dependence. Furthermore, the structural dependence remains unimportant by using this method. Therefore, the nonlinear dependence structure between the variables can be judged by the rank correlation coefficient. Generally, this nonlinear dependence structure is assessed by using Spearman rank correlation and Kendall's Tau. Moreover, the use of Kendall's Tau is frequent because it helps to determine the concordant or discordant pairs' probability directly. The relationship between the correlation coefficient and copula function allows assessing the liner dependencies [33]. For this purpose, Sklar's theorem has gained substantial importance on the various characteristics for the analysis in literature [33, 34]. For instance, random variables, we can say $y_1, \ldots, y_d$ which are following a marginal probability distribution function $F_1(y_1), \ldots, F_d(y_d)$, respectively, there exists a copula, C [35], that can be used to join these functions of marginal distribution in the form of a joint distribution function, as shown in the following equation:

$$\begin{aligned} H(y_1, y_2, y_3 \ldots, y_d) &= (F_1(y_1), F_2(y_2), F_3(y_3), F_d(y_d, F_d(y_d))) \\ &= C(u_1, u_2, u_3 \ldots, u_d), \end{aligned} \tag{1}$$

where $F_k(y_k) = u_k$ for $k = 1, 2, 3, \ldots\ldots, m$, with $u_k \sim u(0, 1)$, and $C(u_1, u_2, \ldots, u_d)$ is envisioned as the copula function. The copulas have been employed as a convenient and useful method in distinct parts of several fields [33, 35]. These various characteristics can be predicted by helping multivariate copulas.

2.2. Regional Spatially Agglomerative Continuous Drought Probability Monitoring (RSACDPMS). The RSACDPMS uses regional-level characteristics to continuously monitor the drought probability in the Northern area of Pakistan. The RSACDPMS provides knowledge for several homogeneous climatic regions and monitors drought characteristics more expeditiously. To complete this work, the following steps are described accordingly.

2.2.1. Seasonal Threshold at the Regional Level Definition. The major concern of the study is to monitor advances regarding precipitation at the regional level. For this purpose, the present study proposed RSACDPMS that calculates the regional thresholds. This calculation is based on the following steps. The precipitation thresholds for four drought severity conditions ("extremely dry, severely dry, median dry, and normal dry") are calculated. These severity conditions are defined in the literature and are classified according to literature [36]. The RSACDPMS can be used for other drought severity conditions accordingly. However, the present study considers them according to current requirements for the research. The rainy season of six months (January to June) is selected for the current analysis. During this selected season, most of the rain falls for this region [37, 38]. The selected rainy season is important for the climatological characteristics. This season accounts for 70% to 80% of the annual rainfall from January to June. The dependency of the other parts of the country is highly linked to this rainy period. The rainy season provides high precipitation, which substantially contributes to the major river system (Indus) in Pakistan. For the rainy period of six months, the SPI is used to drought-triggering precipitation thresholds. Furthermore, the data of the selected stations present a homogeneous environment. Niaz et al. [39] proposed a framework to accumulate information from homogeneous stations. The mentioned study selected the standardized values of the drought category from the varying stations. However, the present study uses the same methodology for selecting precipitation data of the homogeneous stations, which is called "spatially agglomerative precipitation (SAP)."

2.2.2. Converting SPI for the Period. The SPI, which was firstly introduced by McKee et al. [23], has frequently been considered for assessing and characterizing the meteorological drought. The SPI can quantify the standardized discrepancy from selected probability distributions that model the raw data to observe precipitation [19, 26]. Furthermore, SPI reliability found significant distribution in different climatic scenarios (geographical and temporal distribution) and makes it more recognized worldwide [19, 40]. However, the SPI index is insufficient to monitor continuing drought characteristics because it cannot provide expeditious information associated with precipitation deficits, drought probability, and the temporal evaluation of droughts [28, 29]). Furthermore, the deficiency in the precipitation causes the main factor of the drought occurrences. Therefore, the RSACDPMS based on copulas and steady-state probabilities is developed to give the regional level probability of having drought as the rainy season advances by assigning the SPI-6 to drought-triggering precipitation thresholds. The purpose of the proposed RSACDPMS is related to the precipitation advances in the region.

2.2.3. Copula's Fitting. The time series of precipitation data is used for the characterization of the drought. The rainy season of the time-series data is used to envisage the drought occurrences in the selected stations. The thresholds concept proposed by Santos et al. [28] and used by Pontes Filho et al. [29] is also used for the analysis with adaptation. The mentioned studies calculate the precipitation thresholds corresponding to each drought intensity using the generic probabilities proposed by Agnew [41]. However, in the current study, the probabilities for each drought severity condition are obtained by the steady-state probabilities, specifically for the application site. Moreover, the significant difference is that the RSACDPMS was initiated to continuously monitor the drought probability over the Northern area of Pakistan. The precipitation threshold is a significant concern of the analysis. The precipitation thresholds for a region and drought severity are calculated by converting the SPI from January to June, SPI-6. In the studied region, the selected season (January to June) accounts for 70% to 80% of the annual rainfall. The current study is performed to obtain accurate information and timely inform the meteorologists and policymakers to understand the growing risk of drought. For this perspective, the present analysis is accomplished by considering the copula-based method. Numerous studies discuss the families of the copulas in literature [35]. Moreover, the copulas were grouped into four prominent families: extreme value type, meta-elliptical copulas, and Archimedean copulas. Because archimedean copulas have accessible properties in hydrological analyses, it is therefore very prevalent in modeling dependence structures, especially in measuring the dependent tail structures [42]. The Archimedean copulas have limitations for modeling higher-order dependency structures between/among variables. Meta-elliptical copulas [43] can be used to assess such higher-order types in the dependent structure. Furthermore, Archimedean copulas (Gumbel) and meta-elliptical copulas (Gaussian and Student t) were selected as candidates given in the following equations–:

$$\text{Gumbel}\exp\left\{-\left[(-ln u_1)^{\varnothing} + (-ln u_1)^{\varnothing}\right]\frac{1}{\varnothing}\right., \tag{2}$$

$$\text{Gaussian}\theta_{\rho}\left(\theta^{-1}(u_1)\right) + \theta^{-1}(u_2), \tag{3}$$

$$\text{t}-\text{Student}\, T_{\rho,v}\left(T_v^{-1}(u_1)\right), \left(T_v^{-1}(u_2)\right). \quad (4)$$

3. Application

The suitable selection of the region (see Figure 1) for the analysis can improve the capabilities for drought assessment. The selected region has important features regarding the climatological characteristics, including the dependency of the other regions of the country and high altitude, which significantly influences the different parts; it holds a large water frozen reservoir, which substantially contributes to the major river system (Indus) in Pakistan. Therefore, the study selects the most suitable region for the analysis region and develops a widespread basis of permeating evaporative and considerable effect for the irrigation of agriculture sectors of the country [44–46]. The agriculture sectors play a significant role in the development of the country [47–50]). Recently, the agriculture sectors of the country were affected by global warming [51]. Global warming affects widespread places of the world; its impacts are stirring Pakistan's water temperature scarcity. Generally, the drought existences have damaged the economic sector, farming, and agriculture sectors. Particularly, the drought has severely damaged human life and the agriculture sector for the last three decades in Sindh (Province of Pakistan). Therefore, the country needs a comprehensive and substantial method to monitor drought characteristics more quickly by developing inclusive and trustful tools. However, the current outcomes are determined to improve the ability of drought monitoring and mitigation policies significantly.

3.1. Results. The deficiency of precipitation and disturbance to an expected precipitation pattern becomes the main forcing factor for a drought. The drought produces adverse consequences on society and the economy. Therefore, the researchers developed various methods and procedures to reduce the potential negative impacts of drought. However, there are very few studies that focus on drought monitoring in homogeneous regions [39, 52, 53]. The current study used monthly precipitation data to classify the different drought severities in a homogeneous region. The various drought severity conditions are classified according to literature [36]. The steady-state probabilities are used to calculate precipitation thresholds to monitor drought probability and temporal evaluation (Table 1). The steady-state probability is defined as the probability of drought severity in the long run duration. The calculated precipitation thresholds, R_N^* for the rainy season of six months, January to June, at varying drought severities for a region are given in Table 2. The Brier skill score (BSS) values are used to profess the performance of the RSACDPMS. The mathematical description of BSS is available in [29, 54]. In Table 3, the BSS values for normal droughts are given. A value close to 1 represents that model performance is better. Furthermore, n rises the predicting performance of the RSACDPMS rises as well. The average value of all the stations for $n = 1$ is 0.30, while it increases to 0.32 for $n = 2$ and increases to 0.87 for $n = 5$. Furthermore, the probability distributions frequently used for the standardization of SPI, steady-state probabilities, and copula functions (to evaluate the dependence structure of a rainy season's precipitation and its subperiod precipitation) are used in RSACDPMS to calculate the precipitation thresholds in the selected region.

This study tested the most popular Archimedean copulas (Clayton, Gumbel, and Frank) and meta-elliptical copulas (Gaussian and Student t). These copulas' families have been defined in literature [35, 42]. The candidate copulas families are selected based on the Akaike information criterion (AIC). The parameters' estimation of the selected families of the copulas is carried out by the maximum pseudolikelihood (MPL). Furthermore, the bivariate models are used for selecting each subperiod (n), for January ($n = 1$) and May ($n = 5$) (Table 4). The progressive records of the precipitation are observed by Kendall's Tau correlation coefficient (KTCC). KTCC witnesses that the precipitation integrated progressively in the model. The stronger structure of the dependencies R_N and R_n is observed which means that the 30% precipitation of the rainy season is explained by n_1, while n_5 explains 90% of rainfall. The slight disparity between the model and the empirical Kendall's Tau estimates infers that R_N and R_n are suitably modeled by using copulas. The small p-values provide considerable indication against the independence of the structure. When bivariate models are fitted, the next step is to monitor drought probability in chosen stations for every month.

The year 2017 is being selected as an example to continuously show the capabilities of RSACDPMS to monitor drought for that year (Table 5). In January, the observed precipitation is 16 mm, and the climatological precipitation for this month is 22.78 mm. Thus, RSACDPMS has already started to show the risk of a drought in the first semester given the rain in the first month. To occur a normal dry, its probability is 0.72, as the first value is already below the mean. However, it is not so low, and the other drought categories do not present high probabilities as well, and 0.39, 0.34, and 0.009 are for median, severe, and extremely dry conditions. February comes with another month's below mean values, and the two-month accumulated precipitation, which is 34.2, already shows a deficit of 17.9 mm from the climatological accumulated mean of 52.13 mm. This situation kept the normal dry probabilities high but is enough to significantly decrease the other categories' probabilities as they got closer to their precipitation thresholds. In March, the drought severity level (i.e., extremely dry (ED)) is already out of option as the precipitation accumulated by this month, 60.9 mm, is already higher than the ED threshold of 55.29. By April, the observed accumulated precipitation is already greater than all drought severities but one, the closest to normal condition. When June is ended and with it, the rainy season, the accumulated precipitation for 2017 is 102.8 mm, which is below the climatological mean (134.42 mm). Moreover, the mean of the accumulated precipitation (102.8 mm) is not enough to meet the normal dry (ND) threshold (120.99 mm). Therefore, RSACDPMS permitted decision makers to understand the growing risk of drought in that year and what category it would represent.

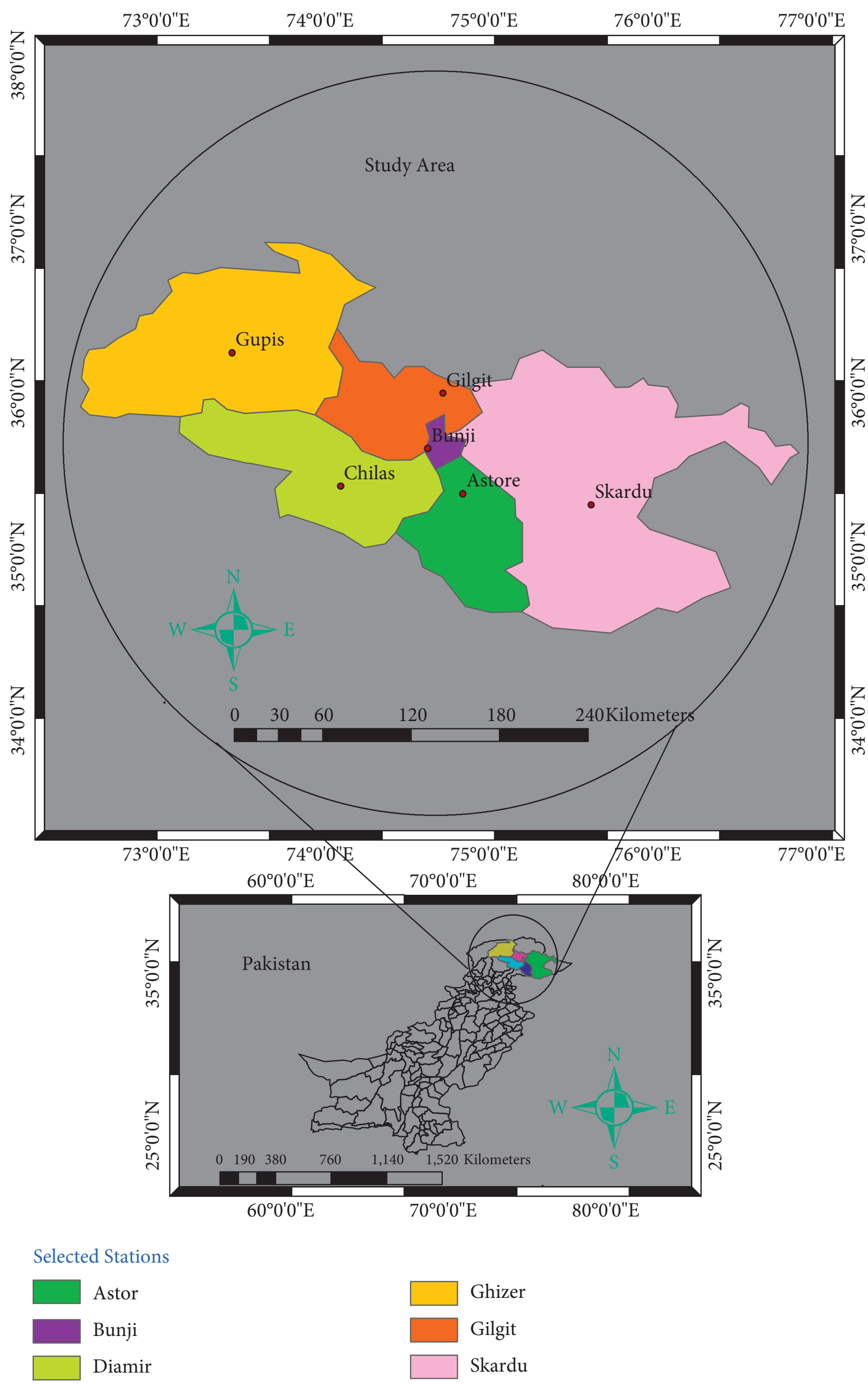


Figure 1: Geographical locations of the six selected stations of the Northern area of Pakistan.

For example, in March, the ND condition has 0.78 drought risk while other categories have much less drought risk. Therefore, anticipated measures that fit the exact proportions of normal drought could be implemented to mitigate its potential negative outcomes.

3.2. Discussion. The present study develops a new method to timely inform the meteorologists and policymakers to understand the growing risk of drought. For this purpose, Pontes Filho et al. [29] proposed the continuous drought probability monitoring system (CDPMS) to monitor drought occurrences

Table 1: The steady-state probabilities observed for various drought conditions.

SPI	SPI > −1 & SPI ≤ 1	SPI > −1.5 & SPI ≤ −1	SPI > −2 & SPI ≤ −1.5	SPI ≤ −2
Drought classes	Normal dry	Median dry	Severely dry	Extremely dry
Steady state probabilities	0.53	0.17	0.14	0.05

Table 2: The precipitation thresholds for various drought conditions (extremely dry, severely dry, median dry, and normal dry) are provided.

Drought intensity	Extremely dry	Severely dry	Median dry	Normal dry
Precipitation thresholds (mm)	55.29	66.05	70.04	120.99

Table 3: BSS values $n = 1$ to $n = 5$ for the normal droughts.

Region	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$	Average
Northern area	0.30	0.32	0.39	0.71	0.87	0.65

Table 4: Each coupled (R_N, R_n) series for bivariate models, their parameters, Kendall's Tau correlation (according to the model and empirical), AIC, and p values at the Northern area.

		Parameters		Kendall's tau			
R_n	Family	∅ or ρ	v	Model	Empirical	AIC	p value
n_1	Gumbel	1.50	—	0.33	0.30	−10.49	<0.05
n_2	Gumbel	1.98	—	0.49	0.45	−27.98	<0.05
n_3	T	0.74	2.0	0.53	0.54	−40.71	<0.05
n_4	Gaussian	0.91	—	0.74	0.72	−81.71	<0.05
n_5	Gumbel	10.23	—	0.90	0.91	−162.6	<0.05

Table 5: Probability of extreme, severe, median, and normal drought events along the rainy season of 2017 (January to June) at scale 6 according to the RSACDPMS framework for the Northern area.

		January	February	March	April	May	June
Mean monthly precipitation (mm)	Monthly	22.78	29.35	22.41	27.55	20.28	12.05
	Accumulated	22.78	52.13	74.54	102.09	122.37	134.42
Observed precipitation (mm)	Monthly	16	18.2	26.7	15.2	16.1	10.6
	Accumulated	16	34.2	60.9	76.1	92.2	102.8
Drought category (threshold)				Drought risk			
Extremely dry (55.29)		0.09	0.07	0.00	0.00	0.00	No drought
Severely dry (66.04)		0.34	0.03	0.01	0.00	0.00	No drought
Median dry (70.05)		0.39	0.04	0.02	0.00	0.00	No drought
Normal dry (120.90)		0.72	0.59	0.78	0.94	0.96	Drought

and translate their probability of occurrence into user-friendly information mathematical transformations. It was applied at the rain gauge level in Portugal, but droughts are usually widespread in large areas. The mentioned study used generic probabilities proposed by Agnew [41] and Gamma distribution to calculate precipitation thresholds in their study. Furthermore, Niaz et al. [54] proposed modified CDPMS to monitor drought occurrences in Punjab and Pakistan. The study used various probability distributions (instead of Gamma distribution), copulas, and steady-state probabilities (instead of generic probabilities) to translate the probability of drought occurrence into user-friendly information mathematical transformations. However, the present study proposes the RSACDPMS to calculate more accurate and comprehensive information about the homogeneous regions and monitor regional drought characteristics more expeditiously. The inclusion of regional spatiotemporal information makes this study innovative. RSACDPMS may receive more importance than CDMPS [29] and MCDPMS [54] at the regional level, specifically, in the homogeneous region. Now, the modification in CDMPS and MCDPMS allows users to calculate precipitation thresholds for the homogeneous region. This modification increases the accuracy and efficiency to determine thresholds for the drought severity.

4. Conclusion

In this study, the dependency between precipitation of the season's total and the observed precipitation at individual months is addressed by using bivariate copula-based models.

The selection of the classes of the copulas families is based on the AIC criterion. The MPL method is employed for estimating the parameters of the copula's families. Furthermore, to continuously monitor the drought probability over the Northern area and monitor advances regarding precipitation at the regional level, the current study proposes RSACDPMS that provides information about the regional advances including the precipitation deficits, drought probability, and spatiotemporal evolution of the drought. The RSACDPMS is validated in the Northern area of Pakistan. Consequently, RSACDPMS enables decision makers to identify the increasing risk of drought severity in the homogeneous region. Furthermore, the RSACDPMS provides information to execute the exact proportions of such a drought's anticipated measures to decrease the possible negative effects of the specific drought condition. For example, in March 2017, the ND condition appears with drought risk (0.78), while other classes have much less drought risk. Therefore, anticipated measures that fit the exact proportions of ND could be implemented to decrease its potential negative effects. The outcomes obtained from the RSACDPMS may improve the monitoring abilities in rainy season advances of the selected region.

Data Availability

The data used for the preparation of the manuscript are available with the corresponding author and can be provided upon request.

Ethical Approval

All procedures followed were under the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000.

Consent

All authors voluntarily agreed to participate in this research study. All authors agreed for publication, and there is no legal constraint in publishing the data used in the manuscript.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work under grant number (RGP.2/4/43), received by Mohammed M. Almazah (http://www.kku.edu.sa).

References

[1] A. C. Chipanshi, K. M. Findlater, T. Hadwen, and E. G. O'Brien, "Analysis of consecutive droughts on the Canadian Prairies," *Climate Research*, vol. 30, no. 3, pp. 175–187, 2006.

[2] J. Maybank, B. Bonsai, K. Jones et al., "Drought as a natural disaster," *Atmosphere-Ocean*, vol. 33, no. 2, pp. 195–222, 1995.

[3] L. Zhang, W. Jiao, H. Zhang, C. Huang, and Q. Tong, "Studying drought phenomena in the Continental United States in 2011 and 2012 using various drought indices," *Remote Sensing of Environment*, vol. 190, pp. 96–106, 2017.

[4] R. Staupe-Delgado, "Progress, traditions and future directions in research on disasters involving slow-onset hazards," *Disaster Prevention and Management: International Journal*, vol. 28, 2019.

[5] R. Stephan, M. Erfurt, S. Terzi et al., "An inventory of Alpine drought impact reports to explore past droughts in a mountain region," *Natural Hazards and Earth System Sciences*, vol. 21, no. 8, pp. 2485–2501, 2021.

[6] Y. Liu, R. Zhou, Z. Wen et al., "Assessing the impacts of drought on net primary productivity of global land biomes in different climate zones," *Ecological Indicators*, vol. 130, Article ID 108146, 2021.

[7] M. Savari and M. Shokati Amghani, "Factors influencing farmers' adaptation strategies in confronting the drought in Iran," *Environment, Development and Sustainability*, vol. 23, pp. 4949–4972, 2021.

[8] O. Rojas, "Next generation agricultural stress index system (ASIS) for agricultural drought monitoring," *Remote Sensing*, vol. 13, no. 5, 2021.

[9] G. Salmoral, B. Ababio, and I. P. Holman, "Drought impacts, coping responses and adaptation in the UK outdoor livestock sector: insights to increase drought resilience," *Land*, vol. 9, no. 6, 2020.

[10] V. Iese, A. S. Kiem, A. Mariner et al., "Historical and future drought impacts in the Pacific islands and atolls," *Climatic Change*, vol. 166, no. 1, pp. 1–24, 2021.

[11] A. Fu, W. Li, Y. Chen et al., "The effects of ecological rehabilitation projects on the resilience of an extremely drought-prone desert riparian forest ecosystem in the Tarim River Basin, Xinjiang, China," *Scientific Reports*, vol. 11, no. 1, pp. 1–11, 2021.

[12] A. Abdelraheem, N. Esmaeili, M. O'Connell, and J. Zhang, "Progress and perspective on drought and salt stress tolerance in cotton," *Industrial Crops and Products*, vol. 130, pp. 118–129, 2019.

[13] M. Fooladi, M. H. Golmohammadi, H. R. Safavi, and V. P. Singh, "Application of meteorological drought for assessing watershed health using fuzzy-based reliability, resilience, and vulnerability," *International Journal of Disaster Risk Reduction*, vol. 66, Article ID 102616, 2021.

[14] T. Caloiero and S. Veltri, "Drought assessment in the sardinia region (Italy) during 1922-2011 using the standardized precipitation index," *Pure and Applied Geophysics*, vol. 176, no. 2, pp. 925–935, 2019.

[15] J. Zhao, Q. Zhang, X. Zhu, Z. Shen, and H. Yu, "Drought risk assessment in China: evaluation framework and influencing factors," *Geography and Sustainability*, vol. 1, no. 3, pp. 220–228, 2020.

[16] A. Y. Kassaye, G. Shao, X. Wang, and S. Wu, "Quantification of Drought Severity Change in Ethiopia during 1952–2017," *Environment, Development and Sustainability*, vol. 23, no. 11, pp. 1–26, 2020.

[17] K. Alsafadi, S. A. Mohammed, B. Ayugi, M. Sharaf, and E. Harsányi, "Spatial–temporal evolution of drought characteristics over Hungary between 1961 and 2010," *Pure and Applied Geophysics*, vol. 177, pp. 1–18, 2020.

[18] T. M. Erhardt and C. Czado, "Standardized drought indices: a novel univariate and multivariate approach," *Journal of the*

Royal Statistical Society: Series C (Applied Statistics), vol. 67, no. 3, pp. 643–664, 2018.

[19] D. H. Mlenga, A. J. Jordaan, and B. Mandebvu, "Monitoring droughts in Eswatini: a spatiotemporal variability analysis using the Standard Precipitation Index," *Jàmbá: Journal of Disaster Risk Studies*, vol. 11, no. 1, pp. 1–11, 2019.

[20] M. Agnoletti, A. Errico, A. Santoro, A. Dani, and F. Preti, "Terraced landscapes and hydrogeological risk. Effects of land abandonment in Cinque Terre (Italy) during severe rainfall events," *Sustainability*, vol. 11, no. 1, p. 235, 2019.

[21] W. C. Palmer, *Meteorological Droughts*, Vol. 45, US Department of Commerce, Weather Bureau Research Paper, Washington, DC, USA, 1965.

[22] W. J. Gibbs and J. V. Maher, "Rainfall Deciles as Drought Indicators," *bureau of meteorology bulletin*, Commonwealth of Australia, vol. 48, no. 29, Melbourne, Australia, 1967.

[23] T. B. McKee, N. J. Doesken, and J. Kleist, "The relationship of drought frequency and duration to time scales," in *Proceedings of the 8th Conference on Applied Climatology*, vol. 17, no. No. 22, pp. 179–183, American Meteorological Society, Boston, MA, January 1993.

[24] B. Yisehak and A. Zenebe, "Modeling multivariate standardized drought index based on the drought information from precipitation and runoff: a case study of Hare watershed of Southern Ethiopian Rift Valley Basin," *Modeling Earth Systems and Environment*, vol. 7, no. 2, pp. 1005–1017, 2021.

[25] A. L. Long, "Drought," *Disturbance and Sustainability in Forests of the Western United States*, p. 992, 2021.

[26] K. R. Dhurmea, R. Boojhawon, and S. D. D. V. Rughooputh, "A drought climatology for Mauritius using the standardized precipitation index," *Hydrological Sciences Journal*, vol. 64, no. 2, pp. 227–240, 2019.

[27] J. P. Bloomfield and B. P. Marchant, "Analysis of groundwater drought building on the standardised precipitation index approach," *Hydrology and Earth System Sciences*, vol. 17, no. 12, pp. 4769–4787, 2013.

[28] J. F. Santos, M. M. Portela, M. Naghettini, J. P. Matos, and A. T. Silva, "Precipitation thresholds for drought recognition: a further use of the standardized precipitation index, SPI," *River Basin Management VII*, vol. 172, 2012.

[29] J. D. Pontes Filho, M. M. Portela, T. Marinho de Carvalho Studart, and F. D. A. Souza Filho, "A continuous drought probability monitoring system, CDPMS, based on copulas," *Water*, vol. 11, no. 9, 2019.

[30] S. Grimaldi and F. Serinaldi, "Asymmetric copula in multivariate flood frequency analysis," *Advances in Water Resources*, vol. 29, no. 8, pp. 1155–1167, 2006.

[31] A. K. Nikoloulopoulos, H. Joe, and H. Li, "Vine copulas with asymmetric tail dependence and applications to financial return data," *Computational Statistics & Data Analysis*, vol. 56, no. 11, pp. 3659–3673, 2012.

[32] H. Van de Vyver and J. Van den Bergh, "The Gaussian copula model for the joint deficit index for droughts," *Journal of Hydrology*, vol. 561, pp. 987–999, 2018.

[33] F. Durante, J. Fernández-Sánchez, and C. Sempi, "A topological proof of Sklar's theorem," *Applied Mathematics Letters*, vol. 26, no. 9, pp. 945–948, 2013.

[34] B. Schweizer, "OPERATIONS ON DISTRIBUTION FUNCTIONS NOT DERIVABLE FROM OPERATIONS ON RANDOM VARIABLES," *Studia Mathematica*, vol. 52, no. 1, 1974.

[35] R. B. Nelsen, *An Introduction to Copulas*, Springer Science & Business Media, Germany, 2007.

[36] X. Li, B. He, X. Quan, Z. Liao, and X. Bai, "Use of the standardized precipitation evapotranspiration index (SPEI) to characterize the drying trend in southwest China from 1982–2012," *Remote Sensing*, vol. 7, no. 8, Article ID 10917, 2015.

[37] Q. Z. Chaudhry and G. Rasul, "Global warming and expected snowline shift along northern mountains of Pakistan," in *Proceedings of the 1st Asiaclic Symposium*, Yokohama, Japan, 2007.

[38] N. Faisal and N. Sadiq, "Climatic zonation of Pakistan through precipitation effectiveness index," *Pak. J. Meteorol*, vol. 6, no. 16, pp. 51–60, 2009.

[39] R. Niaz, I. M. Hussain, Z. Ali et al., "A novel spatially weighted accumulative procedure for regional drought monitoring," *Tellus A: Dynamic Meteorology and Oceanography*, vol. 72, 2020.

[40] I. Nalbantis and G. Tsakiris, "Assessment of hydrological drought revisited," *Water Resources Management*, vol. 23, no. 5, pp. 881–897, 2009.

[41] C. T. Agnew, *Using the SPI to Identify Drought*, 2000.

[42] V. P. Singh and L. Zhang, "IDF curves using the Frank Archimedean copula," *Journal of Hydrologic Engineering*, vol. 12, no. 6, pp. 651–662, 2007.

[43] X. Zhang and K. A. Hoo, "Effective fault detection and isolation using bond graph-based domain decomposition," *Computers & Chemical Engineering*, vol. 35, no. 1, pp. 132–148, 2011.

[44] S. A. Awan, "The climate and flood risk potential of northern areas of Pakistan," *Science Vision*, vol. 7, no. 3-4, pp. 100–109, 2002.

[45] G. Naheed and A. Mahmood, "Water requirement of wheat crop in Pakistan," *Pakistan Journal of Meteorology*, vol. 6, no. 11, pp. 89–97, 2009.

[46] N. U. Ain, M. Latif, K. Ullah et al., "Investigation of seasonal droughts and related large-scale atmospheric dynamics over the Potwar Plateau of Pakistan," *Theoretical and Applied Climatology*, vol. 140, no. 1, pp. 69–89, 2020.

[47] R. Jilani, M. Haq, and A. Naseer, "A study of glaciers in northern Pakistan," in *The First Joint PI Symposium of ALOS Data Nodes for ALOS Science Program in Kyoto Secretariat*, pp. 19–23, Japan Aerospace Exploration Agency (JAEA), Kyoto International Conference Centre, Main Hall, 2007.

[48] A. B. Farooqi, A. H. Khan, and H. Mir, "Climate change perspective in Pakistan," *Pakistan Journal of Meteorology*, vol. 2, no. 3, 2005.

[49] M. Abid, J. Schilling, J. Scheffran, and F. Zulfiqar, "Climate change vulnerability, adaptation and risk perceptions at farm level in Punjab, Pakistan," *The Science of the Total Environment*, vol. 547, pp. 447–460, 2016.

[50] I. Khan, H. Lei, I. A. Shah et al., "Farm households risk perception, attitude and adaptation strategies in dealing with climate change: promise and perils from rural Pakistan," *Land Use Policy*, vol. 91, Article ID 104395, 2020.

[51] W. Malik, H. Shahid, R. Zafar et al., "Role of Pakistan in global climate change through greenhouse gas emissions (GHGs)," *Research Journal of Environmental and Earth Sciences*, vol. 4, no. 11, pp. 996–1001, 2012.

[52] F. Tosunoglu and I. Can, "Application of copulas for regional bivariate frequency analysis of meteorological droughts in Turkey," *Natural Hazards*, vol. 82, no. 3, pp. 1457–1477, 2016.

[53] Z. Ali, I. Hussain, M. Faisal, A. M. Shoukry, S. Gani, and I. Ahmad, "A framework to identify homogeneous drought characterization regions," *Theoretical and Applied Climatology*, vol. 137, no. 3, pp. 3161–3172, 2019.

[54] R. Niaz, M. Almazah, and I. Hussain, "A new framework to substantiate the prevalence of drought intensites," *Theoretical and Applied Climatology*, vol. 147, no. 3, pp. 1079–1090, 2022.

Hindawi
Complexity
Volume 2022, Article ID 9096744, 15 pages
https://doi.org/10.1155/2022/9096744

Research Article

Engaging Stakeholders in Extraction Problems of the Chilean Mining Industry through a Combined Social Network Analysis-Analytic Network Process Approach

Rocio Poveda-Bautista,[1] **Hannia Gonzalez-Urango**,[1] **Edison Ramírez-Olivares**,[2] **and Jose-Antonio Diego-Mas**[3]

[1]*Ingenio (CSIC-UPV), Universitat Politècnica de València, València 46022, Spain*
[2]*Departamento Ingeniería de Minas, Universidad de La Serena, La Serena, Chile*
[3]*i3B, Institute for Research and Innovation in Bioengineering, Universitat Politècnica de València, València 46022, Spain*

Correspondence should be addressed to Hannia Gonzalez-Urango; hkgonzal@upvnet.upv.es

Received 3 November 2021; Accepted 30 December 2021; Published 27 January 2022

Academic Editor: Alireza Amirteimoori

This paper proposes a state-of-the-art methodology for the analysis of stakeholders and their role and performance related to SMEs in the mining industry in the Chilean region of Coquimbo. The relationships between the actors are studied and prioritized according to their support network position. An individual index for each actor based on their influence on solving problems is provided. The social network analysis was used to know the influence of the actors in the sector through the centrality measures. Furthermore, a methodology to measure stakeholders' influence based on the multicriteria method analytic network process approach is proposed. Both methods are used to identify the main stakeholders, study their relationships, and identify the most influential actors involved in executing strategies to boost the sector performance. The results show that the network remains cohesive thanks to certain actors, while the links between private actors must be strengthened; likewise, some public actors should assume a more proactive role in dealing with the problems of the sector.

1. Introduction

Mining is one of the most important economic activities worldwide. Globally, Chile is recognized as one of the largest producers and exporters of minerals, mainly copper. In 2019, Chile produced 5,822 thousand metric tons, equivalent to 28.4% of the world's copper production, which, at the national level, generates a high impact on the gross domestic product (GDP) of the country (9.4% in 2019) [1].

Small and medium-sized companies (SMEs) in the mining industry are significant contributors to this sector and important employment generators [2]. They depend on national development and promotion policies. As a drawback, this generates high inefficiency in productive activities. In this study, we focus on some of the problems related to these companies, given their social and economic impact and the scarce information available in this regard [3].

Mining activities are among the main polemical industries and constitute a real concern worldwide, especially in developing countries like Chile [4]. The modernization of the mineral extraction and treatment processes should be one of the sector's major concerns. However, these activities have been carried out in the same way for more than a century in Chile, where traditional exploitation and treatment practices are difficult to give up. Several other problems are also frequent, e.g., the lack of innovation in their processes and the error in estimating mineral deposits' value [5].

The problems of the sector are intensified in SME mining and are reflected in their production. Additionally, the limited allocation of resources and the absence of efficient plans and programs threaten the implementation of their projects.

The active and coordinated participation of the sector's actors in resolving the problems mentioned above is key for decision making in the mineral extraction processes. Also known as "stakeholders," actors can positively or negatively influence the treatment of a given problem [6]. In the case of SMEs involved in Chilean mining, stakeholders usually apply their ability to act in favor. However, they can also become a threat, developing practices and activities incompatible with the sector's progress, which can turn them into allies or enemies.

As in other previous works, the first step to reach concerted solutions is to identify the related stakeholders and analyze the relationships between them [7]. A solution proposed from the joint work of stakeholders would imply considering participatory methods. They are widely used to address the difficulties of decision-making processes in different scenarios. Nonetheless, one of the challenges of participative processes is selecting proper stakeholders [8].

Within the well-known stakeholder theory, social network analysis (SNA) helps address this challenge. SNA depicts, models, and analyzes a community of agents using a network structure with some nodes and links that represent participating actors and their relationships [8].

The network view has reshaped conventional theory [9], and for that reason, network thinking has been effectively utilized in various areas at different levels. For instance, in economic geography, SNA has been considered a promising tool for empirically investigating the structures and evolution of inter-organization networks and knowledge flows within and across regions [10]. Regarding the mining industry [11], SNA has been applied to examine the strategic role of human resource management in decision-making processes and to determine how the human resources department collaborates with other departments or external organizations.

Considering the network of stakeholders involved in SMEs' problems in the mining industry in Chile, we believe that the action of one or more actors can affect the entire network. Therefore, the structure of this network may be the result of the actions of a few stakeholders. From this point of view, we consider it interesting to use the network approach to understanding the structure of this network.

Describing the existing relationships between the small and medium-sized mining industry stakeholders and their influence in the treatment of the sector's problems can help assess new solutions that support the development of this group. In this way, promoting new ways of cooperation between the different actors could correct the negative impacts caused by the influence of a few actors. The complex network method can be used to study this kind of complex structural relation among these stakeholders, in which there are different individuals with a large number of interactions [12].

The research on social networks has proved that a complex network is more suitable to the characteristics of the real network and that this complex network is composed of several subnetworks [13]. Many SNA-based methods, especially corporate-level and industry-level research, do not fully use SNA constructs to unravel the underlying complexity of the networks [14]. There are many traditional social network analysis methods but few studies from the perspective of complex networks [13].

For this reason and as a novelty of this study, the prioritization logic proposed by the analytic network process (ANP) technique [15] is used to analyze the influence of the actors and compare it with the standard centrality measures of the SNA. Thus, a new SNA-ANP approach to measure the influence among stakeholders exert is proposed. ANP is a well-known multicriteria decision method (MCDM) that models a typical decision problem in the form of a network based on the influence between the problem's elements and prioritizes them based on the influence relationships between them [16,17]. Another advantage of the application of these methodologies is their usefulness in circumstances where there is little precise information, either because it has not been obtained, processed, and disseminated by decision makers; because it is not feasible to generate the type of accurate and complete information that other techniques require; or simply because the one that exists is out of date or not relevant [18,19].

To summarize, the main objective of this study is to analyze the influence relationships between the actors and, based on these influence relationships, prioritize the stakeholders to involve them in future decision-making processes. The aim is to provide different individual indexes for each stakeholder, analyzing the concept of influence from a multirelational point of view based on a combined SNA-ANP methodology. This methodology is used to analyze stakeholders using as a case study the small and medium-sized mining industry of the Chilean region of Coquimbo. In addition, we propose how to take advantage of different actors' positions for the treatment of problems related to the extraction of the mineral.

The rest of the paper is organized as follows. Section 2 presents the context of the case. Section 3 presents the literature on SNA and ANP. Section 4 details the methods proposed and illustrates how the proposed model works through the case study. The results of the case study are discussed in Section 5. Section 6 proposes some implications of the results in the development of strategies for the sector. Finally, Section 7 concludes the study.

2. The Context: SMEs in the Chilean Mining Industry

Small and medium-scale mining comprises all those privately owned mines with an extraction capacity of up to 200 tons of "ore" per day. This group's main characteristic is that the smelting and refining processes are carried out by the National Mining Enterprise ENAMI [5, 20]. ENAMI is a company for promoting and developing small and medium-sized enterprises in the mining industry of the Government of Chile, which provides them with subsidized payment treatment [5, 21].

This segment comprises a large group of small miners who adhere to the government's promotion policy, and most of them sell their products to ENAMI's purchasing powers through a tariff system [5, 21]. Unlike large-scale mining, the demands on small-scale mining in terms of reserves, implementation time, and initial investment are minimal, the capacities and infrastructure needs are moderate, and the employment per unit of production is high. Hence, the specific production pattern makes small-scale mining an attractive option for developing countries [22]. However, several studies indicate that this sector has less equipment, difficult working conditions, and extreme sensitivity to low price cycles [2].

This study focuses on SMEs in the Coquimbo Region (with a total of 33.4% of the mines located in Chile), recognized as one of the most productive areas of the country. The Coquimbo Region contains 11.1% of the total area of exploration mining concessions and 11.5% of the country's exploitation in 2019 [1]. The interest in analyzing SMEs' situation in Coquimbo lies in the productive specialization in the mining activity of this region. This specialization has given rise to a regional development strategy based on creating mining alliances in which SMEs play a leading role.

Chile promotes some policies for the development of small and medium-scale mining through ENAMI [23, 24]. However, these policies also generate some inefficiencies, mainly in terms of incentives. By providing fixed rates and assuring the purchase of their products, small mining producers are not incentivized to reduce costs or to improve the quality of the products they deliver.

Table 1 shows the most critical problems related to the mineral extraction and sale processes faced by small and medium-sized mining companies in the Coquimbo Region [25, 26]. This research will focus on the problems of the mineral extraction process. These processes are the basis for the subsequent sale of the mineral, which must comply only with the buyer's procedures and requirements: ENAMI.

To solve this type of problem, the commitment of mining producers and the organizations that represent them, public and private, is essential. They must understand that this is a common challenge and that the only long-term solution lies in the conviction to share coordinated efforts among all actors [23]. This study intends to provide an individual index for each actor based on his or her influence on the problems and propose some recommendations about how to engage actors in addressing the problems.

3. Theoretical Background

3.1. The Importance of Stakeholder Analysis. Stakeholders can be defined as any group or individual who can affect or is affected by the achievement of an organization's purpose [6]. The stakeholder approach proposed by stakeholder management theory is strongly related to managing the relationship with those groups and individuals.

Interest in analyzing how stakeholders influence management processes, decision making, and conflict resolution has grown significantly, as can be seen in the literature [6, 27]. Therefore, stakeholder management has been adopted in many different areas such as corporate governance, social responsibility, project management [28], environmental management [29], urban development [30], or public management [31], among many others.

However, the identification of the agents, the analysis of their mutual influences and interrelations, the study of the impact or importance in decision making, and the measurement of these influences constitute a real problem that is not fully resolved in the literature. For this reason, so-called "stakeholder analysis" has become an increasingly popular area. The objective of stakeholder analysis is to generate knowledge about the relevant agents so that their behavior, intentions, relationships, and the influences or resources they can contribute to the decision-making processes can be better understood.

Although this analysis is mainly focused on defining methods for identifying and selecting key stakeholders [31–33], many tools have been proposed to improve management and stakeholder engagement. According to Bourne and Weaver [34], there are three basic approaches followed to identifying stakeholders: customer relationship management (CRM), techniques for listing and mapping stakeholders, and social network analysis (SNA).

In general, the analysis of the stakeholders' qualities helps us identify each one of them and thus to know more precisely the expectations they may have. Therefore, selecting the appropriate way of managing stakeholders is crucial. Yang [30] considers that there is no perfect method of identification and prioritization of actors. Therefore, it is convenient to combine different techniques, depending on the case study to which it is applied, that is, according to the number of stakeholders and interest groups, according to the relationships to be studied, and according to the type of information to be extracted.

3.1.1. Social Network Analysis (SNA). SNA was initiated in the early 1920s [35]. It enables relationships to be represented and described systematically and compactly [36]. SNA sees the social world in terms of interactions, rather than as an aggregation of actors who act independently, and thus focuses on patterns of relationships as the unit of analysis [37, 38] and in the implications of these relationships [9, 35, 39].

Since its institutionalization in the 1980–90s, SNA has grown significantly, both in terms of the number of publications and the number of disciplines involved using the SNA approach [40].The method has been applied in many fields, including resource utilization, social communication, disease transmission, strategic planning, science interaction, smart specialization, and project management [35, 41].

Many SNA metrics and concepts have been developed to characterize and compare network structures and positions within them [41]. When selecting each actor's centrality, it is important to keep in mind what we want to consider in the study. This is what a centrality indicates to a specific property. An actor may be visible in terms of one centrality but not another. Therefore, this actor may be more useful in one respect than another may be. A detailed description of centrality measures can be found in [35, 37, 41, 42].

3.1.2. Analytic Network Process (ANP). The analytic network process (ANP) is a theory of relative measurement of intangible features proposed by Saaty [15]. It defines the prioritization model as a network in which the relationships between decision elements represent the influence between them [43–45].

Table 1: Main problems of SME mining in Coquimbo.

Problems of production technology: mineral obtaining process	Commercialization problems: mineral sales process
Lack of professional qualifications	Lack of orientation of small and medium-sized companies in the mining industry for the sale of mineral ore
Lack of strategic business vision	Lack of knowledge in the process of buying and selling minerals
Lack of entrepreneurial capacity	The inefficiency of the state-owned company ENAMI
Technological obsolescence	Low production capacity of the state-owned company ENAMI
Financial weaknesses	Setting of high treatment charges by the state-owned company ENAMI

In a network of actors, the decision elements can be considered the actors of the network, among which it is necessary to identify those with the greatest relevance to the problem, determined by the influence they exert and receive, and what makes them the most relevant actors of the analyzed network.

Like SNA, ANP represents the influence of the elements of a decision problem in a matrix. However, an essential step of ANP is identifying the network elements and their relationships. This is a network design that forces the decision maker to conduct an in-depth analysis of the problem.

In general, the steps to carry out a prioritization process using ANP are

(i) Identifying the components and elements of the network and their relationships (influence matrix).

(ii) Conducting pairwise comparison and calculating relative importance weights.

(iii) Placing the resulting relative importance weights (eigenvectors) within the original influence matrix (unweighted matrix).

(iv) Weighting the unweighted matrix elements by the corresponding priorities of the clusters so that they can be column stochastic (weighted matrix).

(v) Raising the weighted matrix to limit powers until the weights converge and remain stable (superlimit matrix).

The steps and mathematical formulation of the ANP process can be found in [15, 16, 46].

In this paper, we propose the use of ANP logic to quantify the relative influence of stakeholders on the problem of mineral extraction. We conducted the analogy in which the stakeholders identified are the elements in the decision model that should be ranked to prioritize their role in solving the problems for the Chilean sector of mining.

Moreover, to maintain the binary scale of relationships, as in the SNA, no paired comparisons are made between elements. Therefore, the influence matrix is directly transformed into a stochastic column matrix (weighted matrix). This ANP approach allows us to analyze the stakeholders' interdependencies to reflect deeply on the influences of stakeholders on the others (among stakeholders).

4. The SNA-ANP Approach for Stakeholder Analysis

This study proposes the analysis of the network of the small and medium-sized mining industry of the Chilean region of Coquimbo through a methodology that combines SNA, in the stakeholder's identification and the analysis of their influences, and ANP, for the analysis of these influences and the prioritization of the most influential stakeholders in dealing with the problems of this group of companies. Therefore, we have named the proposed methodology SNA-ANP-based approach for stakeholder analysis. The purpose is to associate the actors' positions to decision-making processes and the responses to the mineral extraction problems that this group of companies faces.

The study is exploratory as it examines ANP's approach to stakeholder analysis. It is also descriptive because it combines the SNA-ANP approach to describe the network and its actors' characteristics. The following steps were used to conduct this research (Figure 1).

The combined use of the two techniques has previously been performed in the works of [47–49] for the identification of actors to link in multicriteria decision-making processes. However, as far as the authors know, both techniques have not been combined to study influences between the same elements and compare results.

4.1. Problem Formulation. The correct definition of the problem to be dealt with is the initial stage and is the basis for applying any methodology related to decision making. As mentioned in Section 2, among the possible drawbacks to be addressed in this work, we have focused on problems related to production technology. These problems were identified and classified in previous works [25, 26] through a survey answered by 340 companies belonging to the National Mining Society Database (SONAMI database of small and medium-sized companies in the Coquimbo Region) from which the main problems that are shown in Table 1 of this work were extracted.

4.2. Design: Definition of the Elements of the Stakeholder Analysis. In both techniques (SNA and ANP), the relations are as important as the actors are; therefore, the two fundamental components to identify are the list of stakeholders and the type of interaction to be studied.

4.2.1. Nodes: Identification of Stakeholders. The main interest groups related to mining exploitation in Chile are summarized in Table 2. These actors were selected among associations of small miners, medium-sized companies, or services that usually carry out functions of promotion,

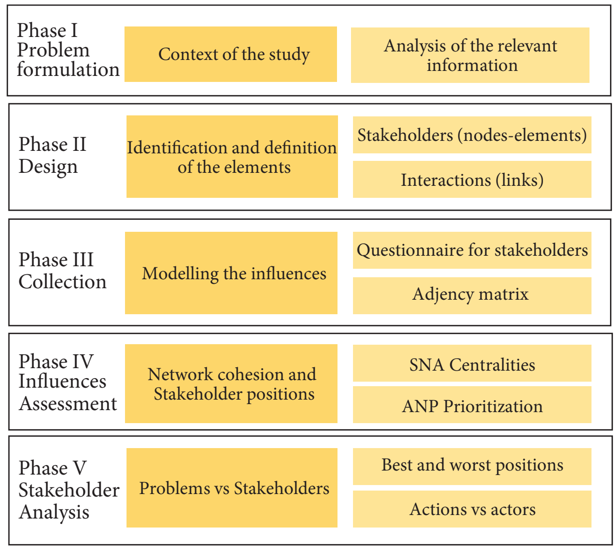


Figure 1: The SNA-ANP proposed approach.

support, technical support, and control of the sector's activities.

Since this study was carried out in the Coquimbo Region, the actors that are directly related to the mining SMEs in this region were identified. For this, a survey of the information was carried out through the Regional Government (GORE) and the Regional Ministerial Secretariat of Mining (SEREMIA). On the other hand, the Ministry of Mining was consulted to define the central government actors that have a direct relationship with SMEs in the different regions of Chile. The list of key stakeholders was compiled, including all these identified actors.

Their public or private nature determines the first characteristic of these actors. The first group corresponds to public institutions, while the second one is conformed of those that do not belong to the Government of Chile and are characterized as being private organizations and trade associations.

The actors represent the nodes in the SNA and the decision elements in the ANP.

4.2.2. Interactions: One-Mode Network. The second key elements are the links between the nodes that represent the relationship between different stakeholders. As this paper analyzes the actor's influence in the problems related to mineral extraction, the relationships (relational content) are those interactions between the actors related to the treatment of the problems.

According to Hanneman and Riddle [36] and many other studies, sharing of information and interactions can be used to establish links between two nodes in a social network. In this work, the relationships between all the actors were studied.

4.3. Collection and Modeling of Influences. A questionnaire was designed as a tool to collect data (Table 3) and sent to the 31 actors listed in Table 2. Responses were received from 17 actors out of the 31 (Table 4). The rest of the actors did not respond despite several attempts. In these cases, the methodology advises against interviewing them because their unwillingness would bias their answers and, therefore, the results. If they were key actors, they would be nominated by the rest of the participants [10].

The network was formed with the 17 actors who answered, taking into account their representativeness in decision-making problems related to mineral extraction.

With respect to the trade associations, of the nine existing ones, the six that participated in the study were included in the network since they are representative in terms of their production size, number of workers, and sales distance from their deposit to ENAMI, while the remaining three are much smaller and not very representative. On the other hand, the eight mining companies existing in the Coquimbo Region, classified as medium-sized mining enterprises, were considered not relevant for this study. These companies sell the final product of their process in a different way than trade associations, and therefore their sales relationship with ENAMI is different. That is, trade associations sell their product extracted from their deposit directly to ENAMI, while the companies, once their product has been exploited, process it in their own plants, obtaining a product and by-product as benefits. Considering these eight companies, with different relationships in the extraction and sales problems with the main promoter of the sector in the Chilean Government´s mining industry (ENAMI), would have altered the relationships of the key network players and their connections with the rest of the network.

Regarding safety entities, in the initial list, we had ACHS (Chilean Association of Security), CCHC (Chilean Chamber of Construction), and IST (Worker Safety Institute). All of them are private non-profit entities, whose missions are very similar and which provide full coverage for work accidents and develop risk prevention programs in Chile. Therefore, CHCH was considered as representative of the three occupational health and safety entities.

Table 2: List of actors.

No	Acronym	Description	Type	Main activity
1	ENAMI	National Mining Enterprise	Public	Promoting, processing, trading
2	SONAMI	National Mining Society	Private/ business	Promotion and advisory services
3	SERNAGEOMÍN	National Geology and Mining Service	Public	Promotion
4	CORFO	Production Development Corporation	Public	Promotion
5	SERCOTEC	Technical Cooperation Service	Public	Promotion
6	SEREMÍA	Regional Ministerial Secretariat of Mining	Public	Technical support
7	GORE	Regional Government	Public	Promotion
8	SENCE	National Training and Employment Service	Public	Training-technical support
9	DDT	Labor Department	Public	Control
10	IST	Work Safety Institute	Private/ business	Training and prevention
11	ACHS	Chilean Security Association	Private/ business	Prevention
12	CCHC	Chilean Chamber of Construction	Private/ business	Promotion
13	ULS	University of La Serena	Public	Training and technical support
14	EB	Banking Entities	Private/ business	Control
15	AGM La Higuera	La Higuera Mining Association	Private/ business	
16	AGM Andacollo	Andacollo Mining Association	Private/ business	
17	AGM Punitaqui	Punitaqui Mining Association	Private/ business	
18	AGM Ovalle	Ovalle Mining Association	Private/ business	
19	AGM El Huacho	El Huacho Mining Association	Private/ business	Small-scale mining: grouping and representing miners
20	AGM Combarbalá	Combarbalá Mining Association	Private/ business	
21	AGM Salamanca	Salamanca Mining Association	Private/ business	
22	AGM Illapel	Illapel Mining Association	Private/ business	
23	AGM La Serena	La Serena Mining Association	Private/ business	
24	CMSG	San Geronimo Mining Company	Private/ business	
25	CMPN	Palo Negro Mining Company	Private/ business	
26	CMLL	Los Linderos Mining Company	Private/ business	
27	CMDP	Punitaqui Mining Company	Private/ business	Medium mining: exploration, extraction, production, and processing: mine + plant
28	CMAC	Minera Andacollo Copper Company	Private/ business	
29	CMLP	Los Pingos Mining Company	Private/ business	
30	CMNE	Nueva Esperanza Mining Company	Private/ business	
31	CMLC	La Cocinera Mining Company	Private/ business	

Table 3: Example of the questionnaire for stakeholders.

Which of the actors do you contact in order to address issues related to mineral extraction problems? Please tick.
Actor:
National Mining Enterprise (ENAMI)
National Mining Society (SONAMI)
National Geology and Mining Service (SERNAGEOMÍN)
. . .
Others

Regarding the participation of the Banking Entity (EB), this was considered unnecessary, since the financing of small mining and trade associations through financial entities is almost nil because the mining sector does not provide benefits for financial entities.

Table 3 shows an example of the questionnaire that was given to the final list of stakeholders.

The adjacency matrix is built based on the actors' responses about the relationships between them to improve the problems related to the mineral extraction process. An actor (A) pointing to another actor (B) indicates that actor B supports actor A. Therefore, this matrix reflects the influence of the elements located in rows on the elements located in columns. A binary measurement scale has been used in which "1" (green) indicates that influence exists between the actors and "0" (red) indicates that there is no influence at all. Once the nodes and lines are identified, a 1-mode adjacency matrix was generated (Table 4).

4.4. Stakeholder Influence Assessment

4.4.1. Analysis through SNA Approach. Three structural or cohesion measures (network density, network centrality, and average path length) and four types of centrality measures (degree, closeness, betweenness, and eigenvector) were used. We have chosen UCINET [50] as the software to carry out the visualization and SNA analysis. The results are presented in Section 5.

4.4.2. Analysis through ANP Approach. Using ANP logic, the prioritization of the actors was obtained according to their influence in solving the mineral extraction process's problems.

According to ANP terms, the adjacency matrix obtained (Table 4) corresponds to the influence matrix representing the influences relationships. Since the calculation of relative importance weights has been avoided, to maintain the same binary scale used in SNA analysis, the influence matrix is transformed into a stochastic column matrix to obtain the weighted matrix (Table 5). Finally, the weighted supermatrix is raised until the weights converge and remain stable to obtain the superlimit matrix. These calculations have been developed using Superdecisions® software [15, 51]. The ANP results are shown in the last column of Table 6.

5. Results and Discussion

5.1. Structural Analysis of the Whole Network. Based on the adjacency matrix, a 1-mode network of stakeholders was generated. In the network, the identified stakeholders are mapped into nodes linked with arrows (Figure 2). The nodes' shape represents the stakeholder's nature, while the nodes' color represents their main activity.

A brief scan of Figure 2 shows that the network is built on several central actors that establish multiple relationships (A1, A2, A6, . . .). Likewise, some peripheral nodes can only communicate through each other, for example, A5, A8, and A9. There is only one isolated node (A10). Therefore, this is a sociocentric type network connected by actors positioned in the center.

Considering only the governmental actors (Figure 3(a)), all of them maintain multiple connections with each other. On the other hand, isolating the non-governmental actors in a subnet (Figure 3(b)), the number of connections diminishes. In the last case, there are two disconnected actors (A10 and A16), while the rest are only held together by actor A2. This subnet also shows a complete disconnection among the associations.

Table 7 shows the main cohesion measures. The social network density is 0.228, indicating that the network is sparse and poorly connected. Such a value is low since the total number of nodes is 17, which indicates that the social network is not very large and it should not be challenging to connect many actors. The number of existing relationships in this network is weak; hence, it is necessary to improve further exchanges and relationships in the sector. This is reflected especially in the group of associations and in some entities that should be recognized as more central actors, such as A5 and A10, given their supporting role.

The social centralization of the network indicates a significant difference among some stakeholders' degree centrality. This is reflected in the network graph in the central position of some actors versus the disconnection of others.

The average degree of each stakeholder is 3.6. This finding indicates that one specific stakeholder can, on average, support more than three stakeholders. Given the size of the network, the average path length is relatively small, with a value of 2. It means that the network connections are efficient, indicating that one actor could easily be exchanged for another. This value facilitates the rapid transfer of information or resources, reducing the loss of information or resources in the process. This is especially important for the small business community that is hard hit by resource constraints.

In general, the results of the cohesion measures show that the relationships between stakeholders are not very close despite the size of the network. However, since the network is small and compact, it is not difficult to improve the connection between the most remote actors.

5.2. Power Quantification Analysis of Chilean Small Miners' Network. Some of the conclusions that emerged from the structure analysis of the whole network carried out in the previous section are reinforced and complemented by the analysis of individual indicators. Table 6 shows the measures calculated for each of the actors in the small mining industry in Chile.

Table 4: Adjacency matrix.

		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17
A1	ENAMI	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0
A2	SONAMI	1	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0
A3	SERNAGEOMÍN	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
A4	CORFO	1	0	1	0	1	1	1	0	0	0	1	0	0	0	0	0	0
A5	SERCOTEC	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
A6	SEREMÍA	1	1	1	1	1	0	1	0	0	0	1	0	0	0	0	0	0
A7	GORE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A8	SENCE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A9	DDT	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
A10	CCHC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A11	ULS	1	0	1	1	0	1	1	1	1	0	0	1	1	1	1	1	1
A12	AGM La Higuera	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
A13	AGM Andacollo	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A14	AGM Punitaqui	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
A15	AGM Ovalle	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0
A16	AGM Combarbalá	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A17	AGM Salamanca	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0

Table 5: Weighted supermatrix.

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17
A1.ENAMI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A2.SONAMI	0	0	0.125	0	0	0.100	0	0	0	0	0.333	0	0	0	0	0	0
A3.SERNAGEOMÍN	0.083	0	0.125	0.200	0	0.100	0.167	0	0	0	0	0	0	0	0	0	0
A4.CORFO	0.083	0.143	0	0	0	0.100	0	0	0	0	0	0	0	0	0	0	0
A5.SERCOTEC	0.083	0	0.125	0	0.500	0.100	0.167	0	0	0	0.333	0	0	0	0	0	0
A6.SEREMÍA	0	0	0	0.200	0	0.100	0	0.500	0	0	0	0	0	0	0	0	0
A7.GORE	0.083	0.143	0.125	0.200	0.500	0	0.167	0	0	0	0.333	0	0	0	0	0	0
A8.SENCE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A9.DDT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A10.CCHC	0.083	0	0.125	0	0	0	0.167	0	0	0	0	0	0	0	0	0	0
A11.ULS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A12.AGM La Higuera	0.083	0	0.125	0.200	0	0.100	0.167	0.500	1.00	0	0	1.00	1.00	1.00	1.00	1.00	1.00
A13.AGM Andacollo	0.083	0.143	0	0.200	0	0.100	0	0	0	0	0	0	0	0	0	0	0
A14.AGM Punitaqui	0.083	0.143	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A15.AGM Ovalle	0.083	0.143	0	0	0	0.100	0	0	0	0	0	0	0	0	0	0	0
A16.AGM Combarbalá	0.083	0.143	0.125	0	0	0.100	0.167	0	0	0	0	0	0	0	0	0	0
A17.AGM Salamanca	0.083	0	0.125	0	0	0	0	0	0	0	0	0	0	0	0	0	0

In Table 6, red circles are used to mark low values for each indicator, amber circles are used for average values, and green circles are used for the most central actors. These measures assess different types of actor relationships and provide insights into how the stakeholders are connected within the mining sector.

There is a significant difference between the most important positions of the OutDegree and InDegree centralities. This is to be expected given the meaning of the relationship studied. Regarding OutDegree, A11ULS has the highest value [13], indicating that 13 out of the 17 stakeholders look for help from this actor to deal with the sector's problems; therefore, it is the most consulted actor. This is consistent given the role of technical support and the typology of problems related to mining production. A11ULS role is critical, especially for small miners' associations, since they only draw on it to solve the problems of the sector. Two hypotheses can be developed to explain this situation. First, the associations do not consider contacting more actors due to lack of knowledge, and second, other actors have not dealt with them in previous experiences. A7GORE, A8SENCE, and A10CCHC are actors that do not provide support to the network.

Regarding InDegre, the values of A1ENAMI [12] and A6SEREMÍA [10] are relatively high, indicating that both institutions are ones that contact more actors to deal with extraction problems. The results of both centralities give us a first idea about the strategy to be implemented considering the contact power of the triad A1, A6, and A11. In general, Associations have the lowest values, which is an evident weakness of the network, as these actors are the ones that require the most support.

Considering OutCloseness, A11ULS remains the closest actor to the rest, along with A6SEREMÍA, A4CORFO, and

Table 6: The SNA-ANP measures.

Actor	OutDeg	Indeg	OutClose	InClose	OutEigen	InEigen	Between	ANP
A1.ENAMI	3	12	31	26	0.47	1.00	50.98	0.107
A2.SONAMI	5	7	37	31	0.49	0.56	9.17	0.056
A3.SERNAGEOMÍN	3	8	39	30	0.37	0.91	6.64	0.033
A4.CORFO	6	5	28	35	0.64	0.58	13.72	0.137
A5.SERCOTEC	3	2	38	39	0.30	0.35	3.13	0.043
A6.SEREMÍA	7	10	27	28	0.73	0.99	47.03	0.157
A7.GORE	0	6	64	31	0.00	0.66	0.00	0.000
A8.SENCE	0	2	64	43	0.00	0.20	0.00	0.000
A9.DDT	3	1	40	46	0.19	0.13	0.00	0.013
A10.CCHC	0	0	64	64	0.00	0.00	0.00	0.000
A11.ULS	13	3	21	35	1.00	0.57	94.62	0.261
A12.AGM La Higuera	4	1	37	46	0.52	0.13	0.14	0.060
A13.AGM Andacollo	2	1	40	46	0.21	0.13	0.14	0.017
A14.AGM Punitaqui	3	1	38	46	0.37	0.13	0.14	0.033
A15. AGM Ovalle	5	1	36	46	0.46	0.13	0.14	0.037
A16.AGM Combarbalá	2	1	42	46	0.19	0.13	0.00	0.013
A17.AGM Salamanca	3	1	38	46	0.37	0.13	0.14	0.033

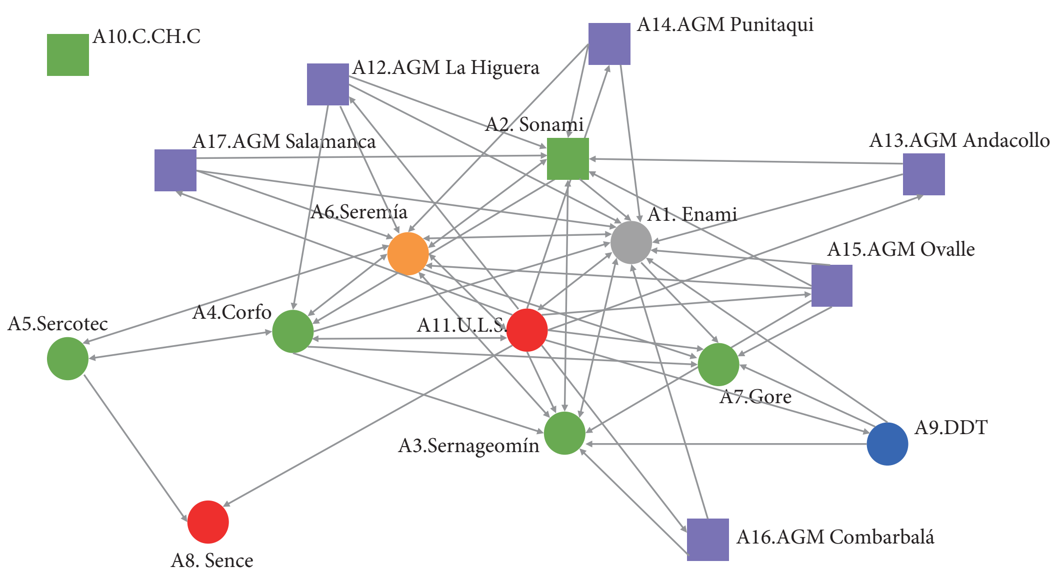


Figure 2: The social network of actors linked to the problem of the mineral extraction process of small mining companies in the Coquimbo Region, Chile (UCINET®) (circles: public actors; squares: private organizations or associations; grey: ENAMI; green: promoting; red: training; blue: control, orange: technical support; purple: association.

A1ENAMI. This suggests that A4CORFO could be an ally for the triad of actors mentioned above. Moreover, given its aim of promoting production, this actor's role must be better used and strengthened. The analysis of InCloseness shows a larger group of actors with the best results. However, given the nature of the network, associations should be able to arrive quickly and get close to the rest of the actors who support them.

OutEigen considers the privileged position of nearby nodes. This centrality shows some different results among the associations, which could be used when designing or applying strategies that seek to impact small miners directly. Especially, A12AGM_La_Higuera and A15AGM_Ovalle, which have the highest values, should make less effort to disseminate or deliver information or resources than the rest of the associations. Regarding the InEigen, A3SERNAGEOMIN seems to play a role in searching for resources for the sector.

As expected, for betweenness, A11ULS has the highest value, indicating that it has the highest potential to connect

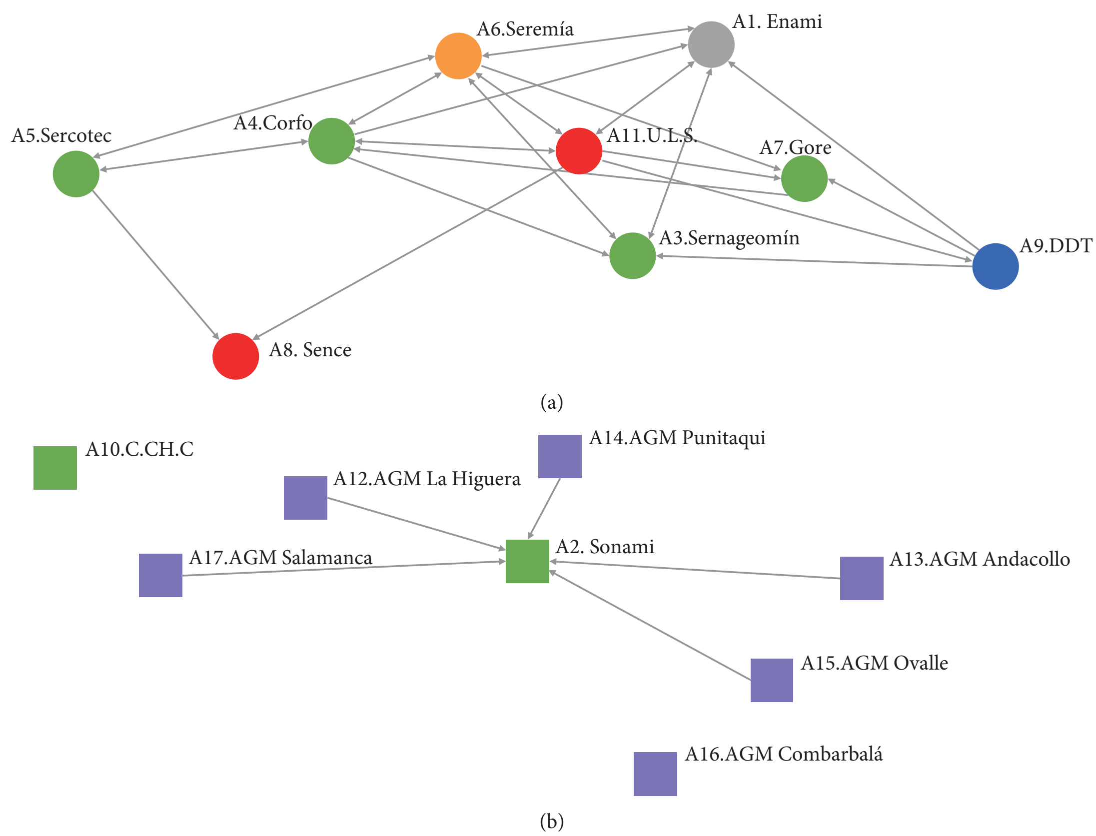


FIGURE 3: Subnets according to the nature of the actors. (a) Public actors. (b) Non-public actors: private organizations or associations.

TABLE 7: Cohesion measures.

Measure	Value
Density	0.228
Deg-centralization	0.663
Out-central	0.621
In-central	0.555
Avg-degree	3.647
Avg-distance	2.076

and control other stakeholders in the network. Stakeholders with higher values of betweenness centrality can act as brokers since they lie on many different paths in the public network and the non-public network. Consequently, they connect many different pairs of actors. Other organizations can also have power as intermediate institutions, such as A1ENAMI and A6SEREMÍA, which can help build up more connections for supporting exchanges, especially to connect the most distant actors. Moreover, these public institutions have adequate drivers and capacities to promote this sector.

Besides, it is not recommended that the network depends so much on a single actor. Actors with the highest betweenness (high and differentiated) have the power to break the link among various actors in the network, weakening the group, and thus must be treated with caution [52]. Centrality analysis also uncovers the betweenness centrality of several actors such as A7, A8, A9, and A10 and all the associations located on the edges of the network, indicating the presence of structural holes. This lack is more evident among non-public actors. They cannot frequently interact with other organizations to discuss mineral extraction problems.

Finally, we analyze the result of the proposed analysis using ANP logic. The result presents the prioritization of the actors that participate in the small companies' mineral extraction process of the mining industry of the region of Coquimbo of Chile. It establishes the dimension of an actor's influence on the problem. This facilitates the comparison between the positions of an actor with respect to others and highlights the differences between the most central and most peripheral actors.

The result reinforces that the A11ULS is the most influential actor. Its priority in this problem is 26%. This value is high, but it is also quite far from the rest. Hence, it is necessary to take actions so that the actors, especially small miners, seek more support in other types of organizations. Moreover, other actors, such as A2SONAMI, could take more relative importance, taking advantage of its proximity to the rest of the actors.

The second group comprises A6SEREMÍA, A4CORFO, and A1ENAMI, with values of 15.7%, 13.7%, and 10.7%. These three public institutions are called upon to lead the process of modernization of the sector that allows small miners to overcome extraction problems. All strategies should be led by them and aimed at those disadvantaged actors of the network, such as small miners' associations, which are the least influential as regards the problem.

The rest of the actors have zero or very low influence values. In this group, we find all the miners' associations

Table 8: Groups of stakeholders

Group	Actors	Actions to be promoted
Group I	A11.ULS	Key actor to lead the actions to be performed
Group II	A6.SEREMÍA A4.CORFO A1.ENAMI	To take a more active role
Group III	**A2.SONAMI** A5.SERCOTEC A3.SERNAGEOMÍN	To improve their connections
Group IV	A9.DDT A7.GORE A8.SENCE A10.CCHC	To increase their presence
Group V	**A12.AGM La Higuera** **A15.AGM Ovalle** A14.AGM Punitaqui A17.AGM Salamanca A13.AGM Andacollo A16.AGM Combarbalá	They are the neediest in the network. Urgent actions are required.

(A12–A17) and A2SONAMI, A3SERNAGEOMIN, A5SERCOTEC, A7GORE, A8SENCE, A9DDT, and A10CCHC. This suggests that 13 out of the 17 stakeholders do not offer any kind of support to the rest of the network. Therefore, the group addresses the same actors (A1ENAMI, A4CORFO, A6SEREMÍA, and A11ULS) regardless of the type of problem, for example, whether it is a problem related to technological obsolescence or financial weaknesses.

An advantage of using ANP logic is that it allows us to complement and summarize the situation depicted from the SNA centralities. The joint analysis using the SNA-ANP approach also allows the classifying of the actors according to the weaknesses and strengths derived from their position in the network. Each measure reveals the location and importance of actors from several different viewpoints. However, to summarize the previous results, following the proposal of Ahmadi et al. [8], we propose a classification of the actors in four groups according to their actions to promote the mining sector (Table 8).

Public actors, especially A11ULS, can easily control the network. The actors in the associations' group (Group V) have less influence because they cannot support other actors and seek support only in A11ULS. For this reason, the intervention of more central actors (Group I and Group II) to establish more support connections is imperative and urgent. Greater availability is required to help these types of actors. The associations A12AGM_La_Higuera and A15AGM_Ovalle may be the most important allies since they are the ones that hold the best position in this group.

Therefore, it is crucial to prepare more policies to encourage non-public stakeholders to ask for more support and involve more public actors supporting small businesses (Group III). A2SONAMI can serve as an intermediate to connect both groups since it is the non-public actor with better linkages among these groups.

Finally, the actors must demand a more significant presence from the actors of Group IV since their role of promoting, training, technical support, and control does not have any influence on the network.

6. Implications of the Results in Handling the Extraction Problems

Two types of actions can be proposed to strengthen the sector. The first aimed at stimulating relations between actors to improve the role and presence of certain actors in the network [42]. The second aimed at solving the problems found by taking advantage of the actors' current position [9, 35]. The objective is to detect opportunities for improvement in management and to support decision-making processes related to the mineral extraction process in the region of Coquimbo in Chile from the perspective of the actors and their positions.

University of La Serena (ULS) is the actor with the most significant influence on the problem of the mineral obtaining process of the SMEs of the mining industry of the region of Coquimbo. This is because this public institution is committed to this region and is aiming at transferring and disseminating knowledge, techniques, and technologies related to mining topics. The Regional Ministerial Secretary of Mining (SEREMÍA) is a relevant public institution since its central role is to execute regional policies, plans, and projects. In a nutshell, these two actors could provide support to companies, helping them increase their strategic vision of the business and improve their entrepreneurial capacity.

In the third and fourth positions appear the Corporation for the Promotion of Production (CORFO) and the National Mining Enterprise (ENAMI). ENAMI is an institution for the promotion of small and medium-sized mining. CORFO is a world-class agency intended to project Chile towards the new knowledge economy of the 21st century. Both institutions play a crucial role in supporting companies' professionals and providing technological training.

Table 9: SNA-ANP analysis vs. key characteristics.

Key characteristics	Measure	Best positions	To be strengthened in
Contacted by other stakeholders Greater support capacity [7,41]	OutDeg	A11.ULS A6.SEREMÍA	A10.CCHC A4.CORFO A2.SONAMI A1.ENAMI A8. SENCE A7.GORE
Contact more actors Request support Greater access to support [7,41]	InDeg	A1.ENAMI A6.SEREMÍA A3.SERNAGEOMÍN A2.SONAMI	A15.AGM Ovalle A12.AGM La Higuer A14.AGM Punitaqui A17.AGM Salamanca A13.AGM Andacollo A16.AGM Combarbalá
Closer to support the rest of the network [37]	OutClose	A11.ULS A6.SEREMÍA A4.CORFO A1.ENAMI	A2.SONAMI A3.SERNAGEOMÍN A5.SERCOTEC A7.GORE A8.SENCE A9.DDT A10.C.CH.C
Get help faster [35]	InClose	A1.ENAMI A6.SEREMÍA A3.SERNAGEOMÍN A2.SONAMI A7.GORE	A15.AGM Ovalle A12.AGM La Higuera A14.AGM Punitaqui A17.AGM Salamanca A13.AGM Andacollo A16.AGM Combarbalá
Reduced delivery effort [29,54]	OutEigen	A11.ULS A6.SEREMÍA	A7.GORE A8.SENCE A10.CCHC A9.DDT A5.SERCOTEC A3.SERNAGEOMÍN
Less effort to seek/demand support or help [29,54]	InEigen	A1.ENAMI A6.SEREMÍA A3.SERNAGEOMÍN	A15.AGM Ovalle A12.AGM La Higuera A14.AGM Punitaqui A17.AGM Salamanca A13.AGM Andacollo A16.AGM Combarbalá A2.SONAMI
Gatekeepers Close structural holes [41,50]	Between	A11.ULS A1.ENAMI A6.SEREMÍA	A4.CORFO A2.SONAMI A3.SERNAGEOMÍN A7.GORE
Influence [15]	ANP	A11.ULS A6.SEREMÍA A4.CORFO	A1.ENAMI A2.SONAMI A3.SERNAGEOMÍN A5.SERCOTEC A7.GORE A8. SENCE A9.DDT A10.CCHC

The active, adequate, constructive, and organized participation of the actors can contribute to better managing the human, technical, and financial resources required for improvements in the mineral extraction process, favoring the access and exchange of crucial information [53], as well as the generation and adoption of commitments by the actors to contribute to management programs and reduce conflicts.

Therefore, we have related SNA-ANP measures to key characteristics of the actors. This makes it easier to identify those actors with specific characteristics required for particular improvement actions. Table 9 shows the best-positioned actors for each measure and those that should be promoted, either because they should offer or because they should ask for more support from the network.

Table 10: Mineral obtaining process problems vs. key stakeholders.

Mineral obtaining process	Required characteristics	Key actors
Professional qualifications and technology obsolescence	Providing technical advice quickly and avoiding the loss of resources.	A1.ENAMI A4.CORFO
Lack of strategic business vision and lack of entrepreneurial capacity	Providing connections between actors and influence	A6.SEREMÍA A11.ULS
Financial weaknesses	Resources and availability	A6.SEREMÍA A11.ULS

Finally, Table 10 shows the relationship between the production technology problems and the stakeholders that could support the decision making and solutions according to their position in the network and their main activity.

7. Conclusions

The application of the methodology proposed in this paper facilitates participatory decision making in the problem of the mineral extraction process. This methodology prioritizes the most influential actors and develops a novel stakeholder analysis model that recommends the participation of the most relevant stakeholders to evaluate opportunities for improvement in the management of mineral extraction processes. In this sense, this methodological approach can be a complementary method in order to go deeper into the relationship between the complexity of real social networks and the decision-making processes of these relevant actors [8].

The active participation of the most relevant actors in these processes contributes to better managing human, technical, and financial resources, relating these actors to specific improvement actions to be applied in the sector. This relationship allows establishing a connection between the sector's production technology problems and the stakeholders, who should make the decisions related to each type of problem, according to their main activity and their position in the network.

The results of the SNA centralities assign a value to the actors' influence on the problem and establish the importance of an actor concerning others, highlighting the differences between the most central and the most peripheral actors.

On the other hand, an advantage of using ANP logic is that it allows complementing the analysis of SNA centralities. The joint analysis, using the SNA-ANP approach, classifies the actors according to the weaknesses and strengths derived from their position in the network. This new approach could be applied in other complex networks with complex structural relations among stakeholders, e.g., at corporate or industrial level, since these levels have not received much attention in network research [14].

According to the results obtained, it can be concluded that government agencies are competent and of great importance for the SMEs of the mining industry of the region of Coquimbo in Chile. It is necessary to re-evaluate the role of the ENAMI, since, as the only purchasing power of minerals of the State of Chile, it results in SMEs of the mining industry of the Coquimbo Region having greater difficulties in complying with the requirements demanded by the state entity, especially with the extraction of the mineral.

Finally, as one of the main limitations of this study, we can mention the low participation of medium-sized companies. However, the information collected with the participating actors allowed us to understand the role of the actors who did not participate [54].

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.

References

[1] "Sernageomin Servicio Nacional de Geología y Minería de Chile. Anuario de la Minería de Chile 2019 [Internet]," [cited 2020 Aug 10]. Available from: http://www.sernageomin.cl, 2020.

[2] C.. Comisión, "Chilena del Cobre [Internet]," Available from: https://www.cochilco.cl/Paginas/Inicio.aspx, 2016.

[3] L. Danielson, "Artisanal and small-scale mining from an NGO perspective," *Journal of Cleaner Production*, vol. 11, no. 2, pp. 97-98, 2003.

[4] A. Reyes, M. Thiombane, A. Panico et al., "Source patterns of potentially toxic elements (PTEs) and mining activity contamination level in soils of Taltal city (northern Chile)," *Environ Geochem Health [Internet*, vol. 42, no. 8, pp. 2573–94, 2020.

[5] P. I. Faúndez, C. Marquardt, J. J. Jara, and J. I. Guzmán, "Valuation and prioritization of early-stage exploration projects: a case study of Cu–Ag and Au-mineralized systems in the tiltil mining district, Chile," *Natural Resources Research*, vol. 29, no. 5, pp. 2989–3014, 2020 Oct 1.

[6] R. E. Freeman, J. S. Harrison, A. C. Wicks, B. L. Parmar, and S. De Colle, *Stakeholder Theory: The State of the Art [Internet]*, http://books.google.com/books?id=xF8-WN1QlIMC&pgis=1, p. 343, Cambridge University Press, Cambridge, Uk, 2010, .

[7] J. Yuan, K. Chen, W. Li, C. Ji, Z. Wang, and M. J. Skibniewski, "Social network analysis for social risks of construction projects in high-density urban areas in China," *Journal of Cleaner Production*, vol. 198, pp. 940–961, 2018.

[8] A. Ahmadi, R. Kerachian, R. Rahimi, and M. J. Emami Skardi, "Comparing and combining social network analysis and stakeholder analysis for natural resource governance,"

Environmental Development, vol. 32, no. September 2018, Article ID 100451, 2019.

[9] Y. Kurt and M. Kurt, "Social network analysis in international business research: an assessment of the current state of play and future research directions," *International Business Review*, vol. 29, no. 2, Article ID 101633, 2020.

[10] A. L. J. Ter Wal and R. A. Boschma, "Applying social network analysis in economic geography: framing some key analytic issues," *Ann Reg Sci [Internet]*, vol. 43, no. 3, pp. 739–56, 2009.

[11] L. Kan and S. Lim, "Strategic human resource management in the Afghanistan Ministry of Mines and Petroleum: a network perspective," *Sustainable Times*, vol. 11, no. 14, 2019.

[12] M. Dong and G. Liu, "Optimization of stakeholder relation network of the qingdao elderly livable community construction project," *Complexity*, vol. 2020, Article ID 8883316, 10 pages, 2020.

[13] N. Li, Q. Huang, X. Ge et al., "A review of the research progress of social network structure," *Complexity*, vol. 2021, Article ID 6692210, 14 pages, 2021.

[14] D. Cao and S. Shao, "Towards complexity and dynamics: a bibliometric-qualitative review of network research in construction," *Complexity*, vol. 2020, pp. 13–17, 2020.

[15] T. L. Saaty, *The Analytic Network Process: Decision Making with Dependence and Feedback*, RWS Publications, Pittsburgh, Pennsylvania, 2001.

[16] T. L. Saaty, *Theory and Applications of the Analytic Network Process: Decision Making with Benefits, Opportunities, Costs, and Risks*, p. 352, RWS Publications P, Pittsburgh, Pennsylvania, 2005.

[17] S. Sipahi and M. Timor, "The analytic hierarchy process and analytic network process: an overview of applications," *Management Decision*, vol. 48, no. 5–6, pp. 775–808, 2010.

[18] N. P. Jesiya and G. Gopinath, "Groundwater suitability zonation with synchronized GIS and MCDM approach for urban and peri-urban phreatic aquifer ensemble of southern India," *Urban Water Journal*, vol. 15, no. 8, pp. 801–811, 2018.

[19] J. Peris, M. García-Melón, T. Gómez-Navarro, and C. Calabuig, "Prioritizing local agenda 21 programmes using analytic network process: a Spanish case study," *Sustainable Development*, vol. 21, no. 5, pp. 338–352, 2013.

[20] P. García, K. Knights, and E. J. Tilton, *Labor Productivity and Comparative Advantage in Mining: The Copper Industry in Chile*, Colorado School of Mines, Division of Economics and Business, Colorado, USA, 2001.

[21] J. M. Sánchez and S. M. Enríquez, "Impacto Ambiental de la Pequeña y Mediana Minería en Chile," 1996, http://biblioteca.unmsm.edu.pe/Redlieds/Recursos/archivos/pequenamineria/Chile/impacto-ambiental.pdf.

[22] M. Ghose, *A Perspective on Community and State Interests in Small-Scale Mining in India Including the Role of Women*, Centre of Mining Environment, India School of Mines University, Q.N° B1, Dhanbad, India, 2007.

[23] ENAMI Empresa Nacional de Minería, http://www.enami.cl/SobreEnami, 2019.

[24] Enami Empresa Nacional de Minería, "Radiografía ilustrada de la minería de menor escala," 2020, https://es.calameo.com/read/00637377508fc441dc9a8.

[25] P. Bautista, P. Ferrando, J. Pascual, R. Olivares, and E. Raúl, "Analysis of the problem of SMEs in the mining industry of Chile," *A study of stakeholders using ANP*, vol. CIDIP 2016, 2016, http://dspace.aeipro.com/xmlui/handle/123456789/776.

[26] E. Ramírez Olivares, *Análisis de la problemática de las PYMES de la industria minera de la IV región de Chile. Estudio de actores participantes mediante ARS y ANP*, Universidad Politécnica de Valencia, Valencia, Spain, 2015.

[27] J. M. Bryson, "What to do when Stakeholders matter," *Public Management Review*, vol. 6, no. 1, pp. 21–53, 2004.

[28] K. O'Toole, M. Keneley, and B. Coffey, "The participatory logic of coastal management under the project state: insights from the estuary entrance management support system (EEMSS) in victoria, Australia," *Environmental Science & Policy*, vol. 27, pp. 206–214, 2013 Mar.

[29] C. Prell, K. Hubacek, and M. Reed, "Stakeholder analysis and social network analysis in natural resource management," *Society & Natural Resources*, vol. 22, no. 6, pp. 501–518, 2009.

[30] R. J. Yang, "An investigation of stakeholder analysis in urban development projects: empirical or rationalistic perspectives," *International Journal of Project Management*, vol. 32, no. 5, pp. 838–849, 2014.

[31] R. Brugha and Z. Varvasovszky, "Stakeholder analysis: a review," *Health Policy and Planning*, vol. 15, no. 3, pp. 239–246, 2000.

[32] E. Mu and H. A. Stern, "A structured stakeholder self-identification approach for the deployment of public information systems: the case of surveillance technology in the city of Pittsburgh," no. 4, pp. 50–66, 2012.

[33] M. C. Achterkamp and J. F. J. Vos, "Critically identifying stakeholders," *Systems Research and Behavioral Science*, vol. 24, no. 1, 2007 Feb 14.

[34] L. Bourne and P. Weaver, "Mapping stakeholders," in *Construction Stakeholder Management [Internet]*Wiley-Blackwell, New Jersey, USA, 2010.

[35] X. Song, Y. Geng, H. Dong, and W. Chen, "Social network analysis on industrial symbiosis: a case of Gujiao eco-industrial park," *Journal of Cleaner Production*, vol. 193, pp. 414–423, 2018.

[36] R. A. Hanneman and M. Riddle, *Introduction to Social Network Methods [Internet]*, Riverside: University of California, Riverside, Losangeles, 2005.

[37] B. K. Wichmann and L. Kaufmann, "Social network analysis in supply chain management research," *International Journal of Physical Distribution & Logistics Management*, vol. 46, no. 8, pp. 740–762, 2016.

[38] A. Marin and B. Wellman, "Social network analysis: an introduction," in *The SAGE Handbook of Social Network Analysis*pp. 11–25, California, USA, 2011, SAGE Publications.

[39] S. Wasserman and K. Faust, *Social Network Analysis*, Cambridge University Press, Cambridge, UK, 15th print edition, 2007.

[40] D. Maltseva and V. Batagelj, "Social network analysis as a field of invasions: bibliographic approach to study SNA development," *Scientometrics*, vol. 121, no. 2, pp. 1085–1128, 2019.

[41] C. Y. Lee, H. Y. Chong, P. C. Liao, and X. Wang, "Critical review of social network analysis applications in complex project management," *Journal of Management in Engineering*, vol. 34, no. 2, pp. 1–15, 2018.

[42] L. G. A. de Souza, M. A. F. D. de Moraes, M. E. S. Dal Poz, and J. M. F. J. da Silveira, "Collaborative networks as a measure of the innovation systems in second-generation ethanol," *Scientometrics*, vol. 103, no. 2, pp. 355–372, 2015.

[43] I. Yüksel and M. Dagdeviren, "Using the analytic network process (ANP) in a SWOT analysis - a case study for a textile firm," *Information Scientist*, vol. 177, no. 16, pp. 3364–3382, 2007.

[44] C.-W. Hsu and A. H. Hu, "Applying hazardous substance management to supplier selection using analytic network

process," *Journal of Cleaner Production*, vol. 17, no. 2, pp. 255–264, 2009 Sep 14.

[45] P. Boateng, Z. Chen, and S. O. Ogunlana, "An Analytical Network Process model for risks prioritisation in megaprojects," *International Journal of Project Management*, vol. 33, no. 8, pp. 1795–1811, 2015.

[46] E. Mu and H. Stern, "A contingent/assimilation framework for public interorganizational systems decisions: should the City of Pittsburgh and Allegheny County consolidate information technology services?" *International Journal of Information Technology and Decision Making*, vol. 17, no. 6, pp. 1611–1658, 2018.

[47] M. M. De Brito, M. Evers, and A. D. S. Almoradie, "Participatory flood vulnerability assessment: a multi-criteria approach," *Hydrology and Earth System Sciences*, vol. 22, no. 1, 2018 Jan.

[48] H. Gonzalez-Urango and M. García-Melón, "Stakeholder engagement to evaluate tourist development plans with a sustainable approach," *Sustainable Development*, vol. 26, no. 6, pp. 800–811, 2018.

[49] H. Gonzalez-Urango, G. Inturri, M. Le Pira, and M. García-Melón, "Planning for pedestrians with a participatory multicriteria approach," *Journal of Urban Planning and Development*, vol. 146, no. 3, 2020.

[50] S. P. Borgatti, M. G. Everett, and L. C. Freeman, "Ucinet for Windows: Software for Social Network Analysis [Internet]," 2002, https://sites.google.com/site/ucinetsoftware/home Harvard, M. 2002 [cited 2018 Dec 9]. Available from:.

[51] T. L. Saaty, *Fundamentals of the Analytic Network Process.*RWS Publications, Pittsburgh, Pennsylvania, 1999.

[52] F. O. Dias, E. Reyes, and F. Saab, "Seeded word-of-mouth marketing strategy: mapping and analysis of a network of political supporters," *Revista Brasileira de Marketing*, vol. 18, no. 4, pp. 177–195, 2019.

[53] J. Pretty, "Social capital and the collective management of resources," *Science*, vol. 302, no. 5652, p. 1912, 2003.

[54] X. Liang, T. Yu, and L. Guo, "Understanding stakeholders' influence on project success with a new SNA method: a case study of the green retrofit in China," *Sustainability*, vol. 9, no. 10, p. 1927, 2017.

Hindawi
Complexity
Volume 2021, Article ID 5766855, 13 pages
https://doi.org/10.1155/2021/5766855

Research Article

An Extended FMEA Model for Exploring the Potential Failure Modes: A Case Study of a Steam Turbine for a Nuclear Power Plant

Huai-Wei Lo[1] **James J. H. Liou**[2] **Jen-Jen Yang,**[1,3] **Chun-Nen Huang**[4] **and Yu-Hsuan Lu**[2]

[1]*Department of Business Administration, Chaoyang University of Technology, Taichung, Taiwan*
[2]*Department of Industrial Engineering and Management, National Taipei University of Technology, Taipei, Taiwan*
[3]*Department of Industrial Education and Technology, National Changhua University of Education, Changhua, Taiwan*
[4]*Department of Fire Science, Central Police University, Taoyuan, Taiwan*

Correspondence should be addressed to James J. H. Liou; jamesjhliou@gmail.com

Received 15 September 2021; Revised 1 December 2021; Accepted 15 December 2021; Published 27 December 2021

Academic Editor: Alireza Amirteimoori

Critical types of infrastructure are provided by the state to maintain the people's livelihood, ensure economic development, and systematic government operations. Given the development of ever more complicated critical infrastructure systems, increasing importance is being attached to the protection of the components of this infrastructure to reduce the risk of failure. Power facilities are one of the most important kinds of critical infrastructure. Developing an effective risk detection system to identify potential failure modes (FMs) of power supply equipment is crucial. This study seeks to improve upon prior approaches for risk assessment by proposing a hybrid risk-assessment model using the concepts of failure mode and effect analysis (FMEA) and multiple-criteria decision-making (MCDM). The proposed model includes a cost-based factor for decision-makers. The subjectivity and uncertainty in FM assessment are adjusted through the rough number method. The original risk priority number (RPN) can be expanded by including the entropy weights in the risk index. Furthermore, to rank the risk priorities in a rational manner, a modified technique for order preference by similarity to ideal solution (modified TOPSIS) is adopted. The applicability and effectiveness of the proposed method were demonstrated by considering an example of a turbine steam engine in a nuclear power plant.

1. Introduction

Critical infrastructure networks, such as technological networks, information and communication technology systems, transport networks, health care systems, and finance and government systems, are vital assets for every country [1, 2]. When a piece of critical infrastructure is destroyed, degraded, or rendered unavailable, lives can be lost, and economic development can be hindered. For example, in mid-August 2017, a large-scale unexpected power outage occurred in Taiwan. The primary reason for this incident was that the supply pipeline for the power supply plant stopped operating, which resulted in a large number of generator sets being shut down. The resultant blackout indirectly caused one death and multiple injuries. The area affected by the incident included a metropolitan area with a high population concentration. The main effects of the blackout included the suspension of business operations in the area and loss of road lighting leading to traffic congestion. To avoid critical infrastructure failure events like this, many countries have begun to focus considerable efforts on protecting critical infrastructure [2]. Interest in the field of risk analysis has grown in recent years, and risk assessment has emerged as a reliable and stable process that supplements and complements many aspects of citizens' lives [3, 4].

Critical infrastructure systems are interdependent, and the infrastructure for producing electrical power is the key system powering the functioning of other facilities [5, 6]. The efficiency of nuclear power generation is higher than that of other power supply systems. One of the main pieces of power generation equipment for nuclear power plants is the turbine steam engine [7]. If the potential for failures can be detected before a failure incident occurs, maintenance measures and improvement strategies can be developed to effectively reduce the probability of failure. This is the purpose of the failure mode and effect analysis (FMEA). FMEA is one of the most popular risk detection tools used to identify, assess, and remove potential or known failure modes (FMs) to improve the safety and robustness of intricate systems. *t* is intended to provide suggestions for risk management decisions [8–10].

Basically, potential FMs in FMEA are assessed and sorted according to the risk priority number (RPN), which is obtained by multiplying three risk elements: severity (S), occurrence (O), and detection (D) [11–13]. Unfortunately, cost, which is often of most concern to the organization or enterprise, is not included as a risk element in the FMEA [8]. In addition, there are only a few FMEA studies that discuss the failure analysis of steam turbines in nuclear power plants.

However, there are problems with the method for determining RPN values when FMEA is applied to real-world problems. The method has been criticized because of the equal weightings of its elements, its high duplication rate, and its failure to address the subjective perceptions from analysts. These shortcomings can significantly affect the accuracy of the analysis results [8, 14]. It is also the case that many current FMEA models use arithmetic averaging to integrate the judgments of multiple experts/decision-makers/analysts. This means that outliers/extreme values are ignored [9].

This study develops an extension of the FMEA model aimed at enhancing the effectiveness of the methodology. Multiple-criteria decision-making (MCDM) techniques that use analysts' experience and judgment to strengthen the risk assessment process are utilized for making critical risk management strategies to enhance the efficacy and empirical validity of risk analysis results. Numerous MCDM models have been proposed to improve the FMEA methodology [15, 16]. The methods for determining the weights of three risk elements include the analytic hierarchy process (AHP) [17], the analytic network process (ANP) [18], data envelopment analysis (DEA) [19], and the best worst method (BWM) [8]. The FM sorting methods include the technique for ordering preference by similarity to ideal solution (TOPSIS) [9], grey relational analysis (GRA) [8], and VIsekriterijumska Optimizacija i Kompromisno Resenje (VIKOR) [20]. FMEA methods using MCDM techniques have been increasingly used for solving real-world cases in recent years. In one study, a hierarchical MCDM approach based on the fuzzy concept and the VIKOR technique was proposed to deal with site evaluation in municipal solid waste management systems [21]. Silva et al. [22] proposed an approach for risk assessment of information security encompassing FMEA and fuzzy set theory. A modified VIKOR method was used to explore the effects of FMs. Researchers have applied the AHP method based on the decision-making trial and evaluation laboratory (DEMATEL) to obtain the influential weights [11]. Instead of calculating the RPN, Safari et al. [23] prioritized risk elements by using fuzzy VIKOR because of the drawbacks of the conventional FMEA method. Mohsen and Fereshteh [14] applied the Z-number technique to reflect the inherent uncertainty in decision-makers' perceptions and the Shannon entropy method to obtain objective weights. A fuzzy VIKOR approach was applied to prioritize the potential risks. Previous studies have significantly advanced risk analysis with fuzzy linguistic information. However, if a risk-assessment model is developed without suitably considering the comprehensive risk elements, the model may produce inaccurate solutions, which could lead to confusion in FM ranking.

In this study, a novel priority model is proposed by applying MCDM methods for FM assessment and ranking in FMEA. In addition, a risk element called the expected cost (E) is added to the process of evaluating the RPN value for financial considerations. The four elements exploited in the FMEA implementations encompass the large range of causal factors leading to an FM, which can reduce the probability of mistakes, uncertainties, and ambiguities in evaluation. The proposed method includes three important steps. In the first step, the concept of rough numbers is employed to handle the uncertainty, subjectivity, and fuzziness arising from the analysts' subjective perceptions and differences in experience during the risk-assessment process. Instead of using the arithmetic mean to obtain crisp values, the rough number is used to effectively integrate analyst information for forming a set of interval values. Next, the entropy technique is employed to generate the objective weights of each risk element. The modified TOPSIS technique is then applied to rank the FMs. In contrast to the traditional TOPSIS, all the alternatives and weight preferences are considered in the modified TOPSIS. Finally, a numerical example of critical infrastructure is shown to illustrate the real application of the proposed model. This study can provide a reference for the industry or organizations to evaluate and prioritize risk in different scenarios. The contributions of this study and the advantages of this methodology are summarized below:

(i) Entropy is used to assign the weight of the risk elements according to the FM assessment data. The proposed model does not require a pairwise comparison questionnaire of risk elements to be issued.

(ii) The increased expected cost is considered a risk element, the inclusion of which enhances the risk-assessment ability of the FMEA.

(iii) The proposed extended FMEA model effectively assesses the potential FMs of nuclear power plants.

(iv) The proposed FMEA model can be applied to other kinds of critical infrastructure. The time and quality of the analysis are not affected by an increase in the criteria and alternatives.

The remainder of the article is organized as follows. A review of the literature on critical infrastructure, risk analysis, and FMEA is presented in Section 2. The research methodology is discussed in Section 3. A real-world case study demonstrating the application of the proposed model is presented in Section 4. Section 5 describes the results and subsequent discussion. Finally, some conclusions and suggestions for future work are presented.

2. Literature Review

This section briefly reviews the topics related to this study, including the interdependence of critical infrastructure systems, type of risk analysis, and FMEA.

2.1. Interdependence of Critical Infrastructure Systems. Critical infrastructure refers not to a single facility but to a collection of numerous facilities, including water, energy, information, telecommunication, financial, transportation, government, and emergency rescue systems [13, 24]. Critical infrastructure is the backbone of the economy in many countries. Energy sustainability, as well as economic and social development, cannot be achieved if the operations of critical infrastructure are at risk of damage or disruption [25]. In recent years, natural disasters and terrorist attacks have been frequently reported, and infrastructure systems have failed, which affects the functioning of all aspects of society [1].

Interdependencies between systems vary widely, and according to their characteristics and effects on infrastructure agents, there are four main categories of interdependencies: cyber, geographic, physical, and logical interdependencies [4]. Geographic interdependency occurs when components of multiple structures are in close spatial proximity. Physical interdependency is related to material flows among infrastructures. Cyber interdependency occurs when a state relies on information transmitted through the internet infrastructure. Logical interdependency includes all types of interdependencies that are not cyber, physical, or geographic connections. Therefore, the reliability of each component of critical infrastructure is the basis of all protective operations.

Huang et al. [5] noted that most types of critical infrastructure are interrelated, but the loss of the electrical power supply is a major factor affecting other infrastructure systems. For example, the water supply, communication, and transportation systems require the provision of electricity to maintain normal operations. Wang et al. [6] proposed a critical infrastructure responsibility framework based on the concept of ethics. In today's high-tech society, electricity is the most basic energy demand, without which all communication systems fail. Große and Olausson [26] believed that when electrical equipment is unavailable, an increasing number of disaster events, such as the interruption of financial transactions, blocked transportation, and failure to monitor security, would occur that would affect society. Therefore, they called for strengthening the protection of energy infrastructure. Developing a comprehensive risk analysis method before an unpredicted disaster occurs can significantly reduce reconstruction costs.

2.2. Type of Risk Analysis. There are two types of risk assessment methods: quantitative and qualitative risk analysis approaches. Quantitative risk analysis methods, such as the proportional risk-assessment technique [27, 28] and the decision matrix risk-assessment technique [10], use the value obtained from mathematical and statistical equations to represent the degree of risk. However, it is difficult to conduct risk evaluation by using mathematical measures because of the complicated structures and widespread usage of information systems. Qualitative risk assessment such as in the FMEA what-if analysis method [12] is based on analyst interviews, and evidence-based results are obtained through soft computing tools rather than mathematical calculations alone.

One major drawback of qualitative risk analysis is that it often does not yield exactly the same results. Furthermore, because qualitative methods do not apply mathematical tools to model the risk, the risk-assessment results are quite dependent on the perceptions of the people who conduct the risk evaluation. The danger of obtaining subjective results when employing qualitative risk analysis approaches exists. Today's information systems have more complicated structures than previous information systems and more widespread usage. Consequently, the intensive mathematical measures used to model risk for complex environments make the risk-assessment process increasingly difficult. Quantitative methods may actually increase the difficulty of the evaluation process. Risk assessment tools based on qualitative measures are more appropriate than other risk-assessment methods for today's intricate risk environment of information systems [10, 29]. A hybrid risk-assessment model is proposed in this study based upon the concepts of qualitative risk analysis, FMEA, and rough numbers. The rough number technique is applied to handle subjectivity and uncertainty, which are significant weaknesses in qualitative risk evaluation.

2.3. FMEA. The purpose of FMEA is to identify all possible risk elements and assess their causes as well as their subsequent effects on the function of the system under consideration [9]. FMEA is a reliable qualitative method applied for accident prevention and risk detection. This technique can be applied to discover and eliminate recognized or potential FMs to enhance the robustness and safety of complex products or systems [8]. In contrast to other risk analysis tools, the major focus of FMEA is to engage in proactive prevention rather than finding a solution after system failure occurs. This proactive prevention can help decision-makers adjust existing strategies, add compensatory provisions, apply the recommended actions to decrease the likelihood of failures, reduce the probability of risk, and minimize accident hazards [11, 12]. Due to its effectiveness, FMEA has been widely applied and proven to be successful in many fields related to the knitting process [30], the aerospace [31], automotive [32], and medical [33] industries, to name a few [34, 35]. The most common way to assess the risk of failure in FMEA involves determining the RPN, which is the product of the S, O, and D. However, when applied in real-world problems, the crisp RPN method also has some shortcomings and limitations. The limitations of the conventional FMEA model are as follows [8, 9, 11, 17, 36]:

(i) The three elements applied in FMEA analysis do not encompass the entire range of causal risk elements

(ii) The measurement of *S* and *D* is relative and subjective, with no holistic characterization of group judgment

(iii) The *S*, *O*, and *D* are difficult to evaluate precisely in numerical form

(iv) The *S*, *O*, and *D* are often given no or equal importance weights

(v) RPN values are not continuous, and there exists no mechanism to interpret the meaning of the differences among different RPNs

(vi) Different combinations of *S*, *O*, and *D* may produce the same RPN, thereby causing some high-risk FMs to be ignored

(vii) Many numbers in the 1–1,000 range cannot be formed from the product of *S*, *O*, and *D*

(viii) Small variations in each rating may lead to considerably different effects on the RPN

The conventional FMEA method has been proven to be one of the most important early preventative initiatives for systems, processes, and services; however, the aforementioned limitations may reduce the reliability of the conventional FMEA model. To deal with the drawbacks of the conventional FMEA model, the entropy-based rough FMEA method is proposed. A detailed introduction of the proposed method is provided in Section 3, and the practical application of the proposed method is described in Section 4.

3. Proposed Extended FMEA Model

In this section, we introduce the mathematical tools used in the proposed method, including rough numbers, entropy, and the modified TOPSIS method. The analysis process of the proposed model is illustrated in Figure 1.

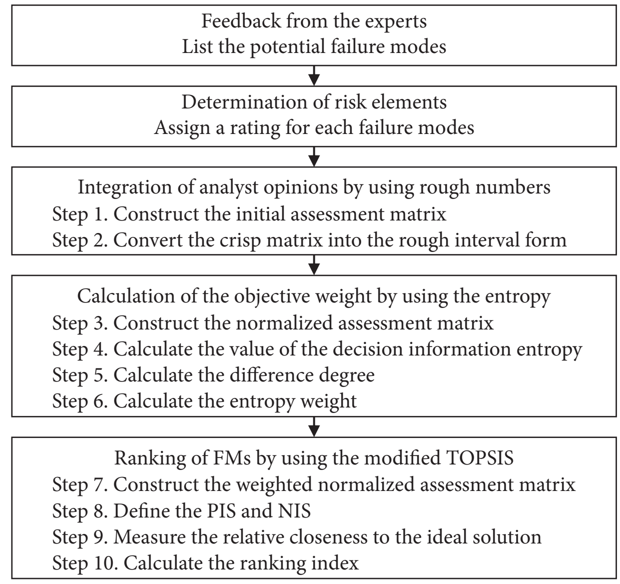


Figure 1: Analysis process of the proposed model.

3.1. Determination of Risk Elements. The *S*, *O*, and *D* used in the FMEA execution process do not encompass the entire range of causative risk elements that need to be considered during decision-making. An increasing number of decision-makers also consider the expected cost when prioritizing FM for prevention. Therefore, for a more comprehensive evaluation of the FMs, the proposed FMEA is not limited to the conventional three risk elements. An additional risk element, namely the expected cost (*E*), is added as a criterion for decision-making. During the risk assessment procedure, analysts selected a linguistic term to describe the degree of *S*, *O*, *D*, and *E* for every FM. The corresponding scores for these terms are presented in Table 1. If an analyst decided that the expected cost for a specific FM was close to the original price, the relative score for the expected cost for that FM would be 10. For example, turbine blade breakage is hazardous (H), and the maintenance cost when the blade fails is extremely high (EH); however, risk detection is very high (VH), and the failure does not occur frequently (very slight, VS). The assessment score would thus be represented as follows: *S*: 10, *O*: 3, *D*: 2, and *E*: 9.

3.2. Integration of Analyst Opinions by Using Rough Numbers. In practice, FM evaluation is an uncertain and subjective group decision-making process because the analysts on the FMEA team judge the importance of risk elements and assess different FMs according to their own knowledge and experience. Thus, practitioners and engineers should find a reliable method to solve the problems of the analysts' subjectivity as well as uncertain or insufficient information. As a new soft computing tool for adjusting the uncertainty and ambiguity of information, rough numbers can provide potential new knowledge without any prior information as well as a relatively objective and reasonable description of the decision issue.

The rough number method is a mathematical extension of the rough set theory proposed by Pawlak [37] and an effective tool for handling vague, imprecise, and uncertain information. There is some overlap with several other theories for dealing with fuzziness and uncertainty, especially with fuzzy set theory. Nevertheless, rough numbers can be viewed as an independent and complementary discipline [38]. The steps involved in the rough number method are described in the following explanation.

3.2.1. Step 1: Construct the Initial Assessment Matrix. After obtaining the corresponding scores of risk elements from the linguistic form (see Table 1), the analyst assesses the risk scores of all FMs. Suppose the FMEA team has k analysts, i FMs, and j risk elements, where $k = 1, 2, \ldots, p$; $i = 1, 2, \ldots, m$; and $j = 1, 2, \ldots, n$. The initial assessment matrix A is represented as follows:

Table 1: Corresponding scores of linguistic terms [9].

Severity	Occurrence	Detection	Expected cost	Corresponding scores
No (N)	Almost never (AN)	Almost certain (AC)	Almost no cost (N)	1
Very slight (VS)	Remote (R)	Very high (VH)	Remote (R)	2
Slight (S)	Very slight (VS)	High (H)	Low (L)	3
Minor (MI)	Slight (S)	Moderately high (MH)	Relatively low (RL)	4
Moderate (MO)	Low (L)	Moderate (M)	Moderate (M)	5
Significant (SI)	Medium (M)	Low (L)	Moderately high (MH)	6
Major (M)	Moderately high (MH)	Very low (VL)	High (H)	7
Extreme (E)	High (H)	Remote (R)	Very high (VH)	8
Serious (SE)	Very high (VH)	Very remote (VR)	Extremely high (EH)	9
Hazardous (H)	Almost certain (AC)	Absolute uncertainty (VU)	Almost or close to original price (O)	10

$$A = \left[a_{ij}^{(k)}\right]_{m\times n} = \begin{bmatrix} a_{11}^{(1)}, a_{11}^{(2)}, \ldots, a_{11}^{(p)} & a_{12}^{(1)}, a_{12}^{(2)}, \ldots, a_{12}^{(p)} & \cdots & a_{1n}^{(1)}, a_{1n}^{(2)}, \ldots, a_{1n}^{(p)} \\ a_{21}^{(1)}, a_{21}^{(2)}, \ldots, a_{21}^{(p)} & a_{22}^{(1)}, a_{22}^{(2)}, \ldots, a_{22}^{(p)} & \cdots & a_{2n}^{(1)}, a_{2n}^{(2)}, \ldots, a_{2n}^{(p)} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}^{(1)}, a_{m1}^{(2)}, \ldots, a_{m1}^{(p)} & a_{m2}^{(1)}, a_{m2}^{(2)}, \ldots, a_{m2}^{(p)} & \cdots & a_{mn}^{(1)}, a_{mn}^{(2)}, \ldots, a_{mn}^{(p)} \end{bmatrix}, \quad (1)$$

where $a_{ij}^{(k)}$ represents the score provided by RN(A) analyst k for FM i under risk element j.

3.2.2. Step 2: Convert the Crisp Matrix into the Rough Interval Form. Next, the rough assessment matrix RN(A) is generated by converting the crisp assessment score into the rough interval form to generate the rough assessment matrix. The rough number calculation is discussed in detail in Lo et al. [9].

$$RN(A) = \left[a_{ij}^{L}, a_{ij}^{U}\right]_{m\times n} = \begin{bmatrix} \left[a_{11}^{L}, a_{11}^{U}\right] & \left[a_{12}^{L}, a_{12}^{U}\right] & \cdots & \left[a_{1n}^{L}, a_{1n}^{U}\right] \\ \left[a_{21}^{L}, a_{21}^{U}\right] & \left[a_{22}^{L}, a_{22}^{U}\right] & \cdots & \left[a_{2n}^{L}, a_{2n}^{U}\right] \\ \vdots & \vdots & \ddots & \vdots \\ \left[a_{m1}^{L}, a_{m1}^{U}\right] & \left[a_{m2}^{L}, a_{m2}^{U}\right] & \cdots & \left[a_{mn}^{L}, a_{mn}^{U}\right] \end{bmatrix}, \quad (2)$$

where a_{ij}^{L} and a_{ij}^{U} represent the rough upper and lower bounds, respectively. The rough matrix is defuzzified for subsequent entropy and modified TOPSIS operations, as given in equation (4). The final assessment matrix Y is represented as follows:

$$Y = \left[y_{ij}\right]_{m\times n} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix}, \quad (3)$$

$$y_{ij} = \frac{a_{ij}^{L} + a_{ij}^{U}}{2}. \quad (4)$$

3.3. Calculation of the Objective Weight by Using the Entropy. In this study, the entropy method was used to eliminate the problem of no or equal relative weights for different risk elements and generate the objective weight of the assessment indicators in FMEA.

The entropy method originated from the thermodynamics field and was initially used to describe the irreversible phenomenon of a motion or process. Shannon [39] then introduced the entropy method into information theory. Entropy is defined as the degree of uncertainty in a random variable. Therefore, the entropy method can be used to determine the degree of disorder and its utility in the system information. If the evaluated parameters have a considerable dissimilarity from each other for a specific risk element, the entropy is small, which indicates that the risk element provides effective information and should be assigned a large weight. By contrast, the smaller the dissimilarity, the larger the value of the entropy weight.

The entropy approach is based on inherent information and is used to obtain the index objective weight. Thus, the method can eliminate the effect of subjective elements and provide a reasonable solution. When adopting the entropy method, only one calculation needs to be performed to obtain a set of weights suited for all FMs.

The entropy method is generally used in the problem of supplier selection; however, in this study, the entropy method was adopted to obtain the weight of the risk-assessment factors in FMEA. The process of calculation is described below.

3.3.1. Step 3: Construct the Normalized Assessment Matrix. The final assessment matrix Y is obtained through rough number calculation. The matrix is used to normalize the rating through equation (5) for eliminating the effect of the rating dimension on incommensurability. Many normalized approaches exist, but in this study, the following equation is selected for normalization:

$$r_{ij} = \frac{y_{ij}}{\sqrt{\sum_{i=1}^{m} y_{ij}^2}}. \tag{5}$$

3.3.2. Step 4: Calculate the Value of the Decision Information Entropy. The evaluation information for each risk element can be represented by

$$f_j = -k \sum_{i=1}^{m} r_{ij} \ln r_{ij}, \tag{6}$$

where $k = (1/\ln n)$.

3.3.3. Step 5: Calculate the Difference Degree. The difference degree is calculated using

$$g_j = 1 - f_j. \tag{7}$$

3.3.4. Step 6: Calculate the Entropy Weight. The entropy weight of the risk element j index is determined using

$$w_j = \frac{g_j}{\sum_{j=1}^{n} g_j}. \tag{8}$$

In this study, the entropy weight represents useful information of the risk-evaluating factors. Consequently, the larger the entropy weight, the more useful the risk element.

3.4. Ranking of FMs by Using the Modified TOPSIS. Due to the drawbacks of the conventional FMEA method mentioned in Section 2, a more comprehensive and flexible FMEA approach for ranking potential risk elements is proposed in this section. The proposed method is based on the traditional TOPSIS, with some modifications made to eliminate flaws. The modified TOPSIS was then used to solve the problem of crisp multiplication in the traditional FMEA.

The TOPSIS technique, which was proposed by Hwang and Yoon [40], has been widely applied in different research areas. The TOPSIS approach is a ranking method based on the concept of a compromise solution. We attempted to determine the solutions farthest from the negative ideal solution (NIS) and nearest to the positive ideal solution (PIS) simultaneously. In the past decades, many studies have attempted to extend the TOPSIS. The modified TOPSIS method used in this study was proposed by Kuo [41]. The ranking index in the traditional TOPSIS does not consider the weights of the separations of an alternative from the PIS and NIS. Therefore, Kuo [41] proposed a new ranking index to deal with this drawback. The proposed solution steps are as follows:

3.4.1. Step 7: Construct the Weighted Normalized Assessment Matrix. The weighted assessment matrix V is determined by multiplying the normalized assessment elements r_{ij} and the corresponding entropy weights w_j, as given in the following equation:

$$v_{ij} = r_{ij} \cdot \left(w_j\right). \tag{9}$$

3.4.2. Step 8: Define the PIS and NIS. The PIS is composed of the maximum of every risk element from the weighted decision matrix, and the NIS is composed of the minimum of every risk element from the weighted decision matrix. The PIS and NIS are determined as follows:

$$\begin{aligned} \text{PIS} &= \left\{v_1^+, v_2^+, \ldots, v_j^+, \ldots, v_n^+\right\}, \\ \text{NIS} &= \left\{v_1^-, v_2^-, \ldots, v_j^-, \ldots, v_n^-\right\}. \end{aligned} \tag{10}$$

3.4.3. Step 9: Measure the Relative Closeness to the Ideal Solution. The distance of every feasible solution from the PIS and NIS is determined using equations (11) and (12), respectively. In this paper, the PIS and NIS are expressed as the highest risk and the lowest risk, respectively. This can facilitate decision-makers to understand the relative risks of these failure modes.

$$d_i^+ = \sqrt{\sum_{j=1}^{n} \left(v_j^+ - v_{ij}\right)^2}, \tag{11}$$

$$d_i^- = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_{ij}^-\right)^2}. \tag{12}$$

3.4.4. Step 10: Calculate the Ranking Index. Compared with the traditional TOPSIS, the modified TOPSIS, which is presented in equation (13), has two advantages in the decision-making process. First, the "relative importance" is considered for two criteria in the modified TOPSIS by adding weights. Thus, if a decision-maker prefers to separate an alternative from the PIS, the modified TOPSIS can provide a different ranking index that caters to the requirements. Second, the ranking index is more comprehensive, thus avoiding the unrecognized ranking of certain alternatives.

In attempting to find a compromise solution, the ranking index in equation (13) is more intelligible than and superior to the original TOPSIS. The calculation process of the ranking index (RC_i) is given as follows:

$$\mathrm{RC}_i = w^+\left(\frac{d_i^-}{\sum_{i=1}^m d_i^-}\right) - w^-\left(\frac{d_i^+}{\sum_{i=1}^m d_i^+}\right), \tag{13}$$

where w^+ and w^- denote the weights that reflect the corresponding importance of the two separation measures obtained from the decision-makers.

4. Empirical Case: Steam Turbine at a Nuclear Power Plant

The proposed methodology was applied to an empirical case for a steam turbine in a nuclear power plant. The smooth functioning of a steam turbine is essential for the reliability and stability of operations in a nuclear power plant. A steam turbine is simply a heat engine that performs mechanical work by using steam as its working fluid. Compared with traditional steam engines of the reciprocating type, steam turbines have considerable improved heat transfer efficiency and are commonly used in thermal and nuclear power plants. To ensure the reliability of the turbine system, the potential FMs of the system should be evaluated, and risk assessment conducted. The effectiveness of the proposed FMEA model was verified by comparison with the conventional FMEA approach.

4.1. Survey Data. This study investigated a nuclear power plant in Taiwan as an example to demonstrate the usefulness and practicability of the proposed FMEA model. The FMEA team consisted of 24 analysts, including government regulators, professors in the relevant fields, and nuclear power plant engineers. Each analyst had at least 10 years of work experience in the nuclear power industry. Currently, two major nuclear power plants are in operation in Taiwan. The occurrence of unplanned downtime at a nuclear power plant has a significant impact on the functioning of society. Nuclear power plants contain numerous and complex pieces of equipment, but we selected the most critical component, namely the "steam turbine," for investigation. Twelve major potential FMs were identified for nuclear power plants by the group of analysts: high temperature of the engine (FM_1), clogged lubricating oil system (FM_2), foreign objects within the system (FM_3), fracture of the vane (FM_4), loose valve (FM_5), bearing damage (FM_6), broken chassis (FM_7), mechanical transmission breakdown (FM_8), rotor breakdown (FM_9), sensor malfunction (FM_{10}), leakage of the pipeline (FM_{11}), and measurement instrument breakdown (FM_{12}).

4.2. Ranking Priority of FMs. After identifying the major potential FMs, the 24 analysts subjectively ranked the importance of these risk elements using a questionnaire format. To save space, we have only presented one criterion, namely the severity scores of analysts for the 12 FMs (Table 2).

The rough assessment matrix comprises the interval values obtained from the analysts' feedback, which includes the uncertainty of their subjective judgments. The rough number method provides a larger amount of implicit information than the arithmetic mean method does. The rough assessment matrix (Table 3) can be obtained using equations (1) and (2). Here, use the fourth FM (FM_4) as a rough number calculation case, as shown in Appendix A.

Using equations (3) and (4), the intervals obtained from Table 3 were transformed into final crisp values. Then, by executing the entropy calculation program described in Section 3.3 (equations (5)–(8)), the weights of the risk elements were obtained as follows: $w_S = 0.4925$, $w_O = 0.1944$, $w_D = 0.0806$, and $w_E = 0.2325$. The severity (S) was rated as the most important risk element, with a weight value of 0.4926. The expected cost (E) was the second most important risk element, which indicated that cost considerations are necessary for critical infrastructure risk-assessment systems. In fact, the government allocates maintenance budgets for critical infrastructure in specific cycles. Finally, the modified TOPSIS was used to calculate the ranking of the FMs, as described in Section 3.4. Table 4 presents the ranking, d_i^+ value, and d_i value of the FMs.

4.3. Results and Management Implications. Steam turbines are a key mechanism for energy conversion in nuclear power plants. According to the information provided by the 24 experts, the results of our analysis indicate that rotor breakdown (FM_9), fracture of the vane (FM_4), foreign objects (FM_3), a clogged lubricating oil system (FM_2), bearing damage (FM_6), and mechanical transmission breakdown (FM_8) are the top 6 FMs leading to the failure of steam turbines. The modified TOPSIS can provide the relative risk level of the FMs through the ranking index (RC_i). Because the sum of RC_i is equal to 0, when RC_i is greater than 0, the FMs have a high-risk level. For example, rotor breakdown (FM_9) is the highest-ranking FM. This mode has the shortest distance from the PIS ($d_9^+ = 0.0396$) and farthest distance from the NIS ($d_9^- = 0.1083$). In addition, RC_9 is positive and represents the largest ranking index (0.0391) of the FMs.

The purpose of FMEA is captured in the phrase "An ounce of prevention is worth a pound of cure." The results of FMEA can provide engineers with guidance as to what precautionary measures should be taken before an accident occurs. We conducted detailed interviews with the analysts to identify the prevention methods for the six aforementioned FMs. Several inspection and maintenance measures were obtained, as presented in Table 5.

4.4. FMEA Model Comparison. To demonstrate the effectiveness of the proposed method, it was compared with the conventional and cost-based RPN methods. In the conventional RPN method, which is based on an engineering perspective, S, O, and D are multiplied. In the cost-based RPN method, the expected cost is considered as a risk element, and then S, O, D, and E are multiplied. Both the cost-based RPN and proposed methods consider the management perspective to reflect the actual budget constraints of risk management. The calculation results obtained with the three approaches are listed in Table 6.

The correlation coefficient between the ranking results of the conventional and cost-based RPN methods was 0.78, which indicates that the ranking results

Table 2: Crisp severity (S) ratings from the analysts for the 12 FMs.

Analysts	FM_1	FM_2	FM_3	FM_4	FM_5	FM_6	FM_7	FM_8	FM_9	FM_{10}	FM_{11}	FM_{12}
1	SI	H	M	H	MI	E	MO	M	H	MO	S	SI
2	SI	M	M	H	M	H	SE	H	M	MI	MI	SE
3	SI	SI	M	H	M	SI	MI	SE	SE	VS	MO	MO
4	SI	SE	SE	H	SI	SI	SI	SI	SE	SI	MO	SI
5	SE	M	M	H	M	E	SE	SE	SE	MO	MO	MO
6	MO	M	H	H	MO	M	H	E	SE	S	MO	MO
7	S	M	SE	H	SI	SI	SE	SE	SE	S	S	MO
8	SI	SE	H	H	MO	H	SI	S	H	S	S	S
9	SI	SE	SE	H	SI	M	SE	SE	H	SI	M	SI
10	MO	SE	SI	H	SI	M	SI	SI	SE	MO	SE	SI
11	M	SE	M	M	M	M	MO	M	M	MO	MO	MO
12	MO	SE	SE	H	M	M	SI	SI	SE	MO	SI	MO
13	VS	M	SE	H	MO	M	MI	S	SE	S	S	N
14	M	E	E	SE	MO	SI	SI	E	E	N	MI	SI
15	SI	SE	M	SE	SI	E	E	M	SE	MO	MO	MI
16	MI	SE	M	E	SI	M	SI	M	H	S	MO	M
17	MO	E	SE	H	M	M	SE	SI	H	S	MO	MI
18	M	M	E	E	E	E	E	E	E	SI	MO	MI
19	SI	M	M	SE	SI	SI	SI	M	SE	S	SI	S
20	MI	M	SE	E	M	M	MI	M	SE	VS	MI	M
21	SI	E	SE	H	M	SE	SI	M	H	N	M	SI
22	M	SE	E	SE	E	SE	M	E	M	MI	SI	MI
23	MO	SE	E	SE	M	SI	M	SI	E	S	M	SI
24	SI	H	SE	SE	MI	E	M	E	E	MI	SI	MO

Table 3: Rough assessment matrix.

FM_i	S	O	D	E
FM_1	[4.635, 6.541]	[2.497, 4.451]	[1.504, 3.444]	[3.613, 6.115]
FM_2	[7.456, 8.863]	[3.114, 5.545]	[2.101, 4.209]	[4.463, 6.392]
FM_3	[7.411, 8.832]	[3.383, 6.571]	[3.706, 6.42]	[3.435, 6.515]
FM_4	[8.854, 9.834]	[1.799, 3.927]	[2.536, 5.71]	[4.936, 7.299]
FM_5	[5.466, 6.906]	[4.018, 6.478]	[2.827, 4.656]	[3.628, 5.584]
FM_6	[6.632, 8.212]	[2.284, 5.524]	[2.327, 4.182]	[3.813, 6.974]
FM_7	[5.526, 8.007]	[1.724, 5.113]	[2.332, 5.076]	[3.96, 6.121]
FM_8	[5.926, 8.214]	[2.706, 5.611]	[2.589, 4.876]	[4.15, 6.638]
FM_9	[8.216, 9.407]	[2.434, 5.007]	[2.201, 3.961]	[5.626, 8.394]
FM_{10}	[2.718, 4.754]	[5.522, 7.931]	[2.843, 5.832]	[3.617, 5.688]
FM_{11}	[4.154, 6.169]	[5.316, 7.489]	[2.745, 4.617]	[3.809, 6.172]
FM_{12}	[4.008, 6.202]	[3.589, 6.674]	[3.33, 5.626]	[3.698, 5.646]

Table 4: Results of the modified TOPSIS calculation.

FM_i	d_i^+	d_i^-	RC_i	Ranking
FM_1	0.0915	0.0386	−0.0329	11
FM_2	0.0441	0.0925	0.0271	4
FM_3	0.0415	0.0943	0.0298	3
FM_4	0.0481	0.1163	0.0386	2
FM_5	0.0738	0.0581	−0.0108	8
FM_6	0.0568	0.0769	0.0104	5
FM_7	0.0709	0.0628	−0.0063	7
FM_8	0.0598	0.0708	0.0050	6
FM_9	0.0396	0.1083	0.0391	1
FM_{10}	0.1182	0.0478	−0.0436	12
FM_{11}	0.0894	0.0522	−0.0237	9
FM_{12}	0.0933	0.0409	−0.0326	10

Table 5: Top six FMs as well as the inspection and maintenance measures for them.

Ranking	FM_i	Failure mode	Inspection and maintenance measures
1	FM_9	Rotor breakdown	Monitor the vibration and temperature of bearings and perform nondestructive inspections on a regular basis to ensure that there is no degradation of the components.
2	FM_4	Fracture of vane	Set standard operating hours and perform regular inspections of turbine vanes.
3	FM_3	Foreign objects	Implement a foreign material exclusion (FME) control system.
4	FM_2	Clogged lubricating oil system	Regular cleaning or replacement of related devices. In addition, the quality of the lubricating oil must be monitored and tested regularly.
5	FM_6	Bearing damage	Regular maintenance, inspection, and replacement of bearings with lubricating oil.
6	FM_8	Mechanical transmission breakdown	Monitor the speed and vibration of the machinery and perform nondestructive tests on a regular basis.

Table 6: Calculation results and rankings obtained with the three methods.

FM_i	Conventional RPN (S, O, and D)		Cost-based RPN (S, O, D, and E)		Proposed FMEA method	
	RPN	Ranking	RPN	Ranking	RC_i	Ranking
FM_1	36	12	180	12	−0.0329	11
FM_2	96	9	480	9	0.0271	4
FM_3	200	1	1,000	1	0.0298	3
FM_4	108	7	648	3	0.0386	2
FM_5	120	3	600	5	−0.0108	8
FM_6	84	10	420	10	0.0104	5
FM_7	84	10	420	10	−0.0063	7
FM_8	112	5	560	7	0.0050	6
FM_9	108	7	756	2	0.0391	1
FM_{10}	112	5	560	7	−0.0436	12
FM_{11}	120	3	600	5	−0.0237	9
FM_{12}	125	2	625	4	−0.0326	10

obtained with the conventional and cost-based RPN methods were somewhat similar. However, certain large differences were observed between them, such as the differences in FM_4 and FM_9. These differences were caused by the inclusion or exclusion of the expected cost as a risk element in the FMEA. In both the cost-based RPN and proposed methods, S, O, D, and E are considered to be risk elements. Nevertheless, the correlation coefficient between the results of the cost-based RPN and proposed methods was 0.4511. The ranking results obtained with these two methods were different, especially for FM_2, FM_6, FM_{10}, FM_{11}, and FM_{12}. The proposed method provides a more rational risk assessment than the other two methods because it addresses the drawbacks in the calculation of the RPN. Rezaee et al. [42] increased the cost element in the conventional FMEA model to optimize the model. They applied the optimized model to data in the marble processing industry. Rezaee et al. [42] verified that the analysis results obtained with the optimized FMEA model were closer to reality than the results obtained using the conventional FMEA model. In the proposed method, linguistic variables and rough numbers are used to capture and express analysts' subjective opinions regarding the importance of each risk element in FMEA. The weights of the risk elements were generated using the concept of entropy, and the ranking index was obtained using the modified TOPSIS.

After analyzing the FMEA calculation results, four main differences were observed among the three methods.

4.4.1. Comprehensively Consider Available and Significant Factors for Decision-Making. Three risk elements, namely severity, occurrence, and detection difficulty, are used in the conventional RPN method. However, the conventional RPN method does not encompass the entire range of causal factors. It does not consider the expected cost, which is an important element. If two FMs receive the same ratings for S, O, and D, the FM with a higher expected cost should have a higher priority. It is not possible for any organization, whether a government agency or private enterprise, to assign unlimited resources to ensure the system or product reliability. In Taiwan, the government allocates a certain budget to the Homeland Security Office every year for maintaining critical infrastructure. Therefore, the cost-based RPN and proposed methods, which consider the expected cost as a risk element, are more suitable for obtaining a sufficient decision-making index.

4.4.2. Evaluate the Ratings and Consider Subjectivity during FM Evaluation. In the conventional RPN method, the final score of every risk element is derived from the arithmetic average, but information loss may occur during the calculation process. In many cases, the obtained score may be unrepresentative because of some uncertain ratings. However, in the proposed method, the subjectivity and uncertainty in FMEA can be adjusted. The rough number mechanism in the proposed FMEA method can provide reasonable and realistic risk element ratings because the

Table 7: Main differences in the three methods.

Method selection	Factors considered			
	Expected cost consideration	Rating	Consideration of parameter weight	Reduce ranking repeatability
Conventional RPN method	Partial	No	No	No
Cost-based RPN method	Yes	No	No	Partial
Mohsen and Fereshteh [14]	No	No	Yes	Yes
Lo and Liou [8]	Yes	No	Yes	Yes
Yucesan et al. [36]	No	No	Yes	Yes
Proposed FMEA method	Yes	Yes	Yes	Yes

flexible and dynamic rough interval represents the subjectivity in the determination of the risk element scores.

4.4.3. Deploy Objective Weights for Every Risk Element. The conventional RPN method assumes that every risk element has equal weight. This method fails to examine the importance of each risk element. In addition, some important elements tend to be ignored. In the proposed technique, the entropy weight method, which is based on inherent information, is used to determine the index weight. Therefore, the proposed method not only considers the weight of each element but also determines the weights in an objective manner. Our determining weights method is different from the studies of Lo et al. [9], Lo and Liou [8], and Yucesan and Gul [35], since they generate weights based on the judgment of experts. This method requires the design of another questionnaire (such as AHP, ANP, or BWM questionnaires). Undoubtedly, these documents all pointed out that risk elements (S, O, D, and E) should have corresponding importance weights.

4.4.4. Lower the Repetition Ratio for Prioritization. The simple multiplication used in the conventional FMEA means that many repeated ranking values are obtained (Table 6). When the conventional RPN method was used, the third, fifth, seventh, and tenth ranks were repeated twice. The priority orders assessed were identical with a high frequency, and the repetition rate was over 66%. In this situation, the decision-maker may find it difficult to decide which FM to prioritize. With the cost-based RPN method, the repetition rate was less than 50%. However, the aforementioned problem still persisted. With the proposed method, the repetition rate among all FMs was 0%, which indicates that the method could clearly distinguish the priorities for all the FMs. The main differences between the conventional RPN, cost-based RPN, and proposed methods are presented in Table 7.

On the other hand, we use sensitivity analysis to explore whether changes in risk element weights will affect the ranking of failure modes. The sensitivity analysis method refers to Lo and Liou's [8] study. The severity (the highest risk element weight) among all the risk elements is assigned from 0.1 to 0.9, with the other weights assigned in proportion. Table 8 presents the results of nine weight configurations. The "severity weight" changes from run 1 to run 9 in units of 0.1. Figure 2 is the ranking result of the failure mode after nine runs of the sensitivity analysis. Under the change of the risk element S, the failure modes will have a significant change, which means that our model is highly sensitive to the change of the risk element weight. Therefore, the relative importance of risk elements must be evaluated. In this case, it is reasonable to use entropy to determine the objective weights of the risk elements, and the analysis results are also unanimously approved by experts.

Table 8: All the risk element weights change according to severity.

	S	O	D	E
Entropy	0.493	0.194	0.081	0.233
Run 1	0.100	0.345	0.143	0.412
Run 2	0.200	0.306	0.127	0.367
Run 3	0.300	0.268	0.111	0.321
Run 4	0.400	0.230	0.095	0.275
Run 5	0.500	0.191	0.079	0.229
Run 6	0.600	0.153	0.063	0.183
Run 7	0.700	0.115	0.048	0.137
Run 8	0.800	0.077	0.032	0.092
Run 9	0.900	0.038	0.016	0.046

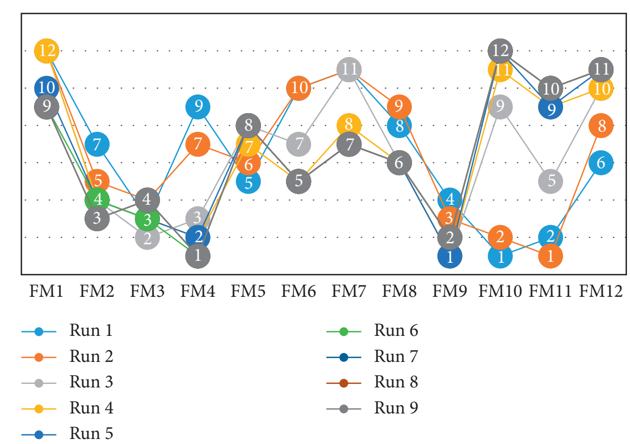


Figure 2: The ranking result of the failure mode after nine runs of the sensitivity analysis.

5. Conclusions and Remarks

Critical infrastructure systems are important for the normal functioning of society and economic development [43]. Natural disasters and accidents are often not predictable in advance. Adopting suitable prevention strategies can significantly reduce the cost and time of reconstruction after a disaster. Many major incidents of critical infrastructure failure have occurred, and governments have actively developed risk management policies to cope with them. The FMEA method, which can reveal the potential causes and problems of failure modes during the process of risk evaluation, is widely used in many fields to control the stability and reliability of any facility [36, 44, 45]. However, some intrinsic problems in the conventional FMEA method exist. The entropy-based rough FMEA method proposed in this study has four advantages, which can alleviate the problems of the conventional FMEA method. The proposed approach considers the uncertainty of information provided by analysts with varied backgrounds. It determines the weight of the relationship among evaluation parameters and reduces the repetition rate for prioritization. Consequently, decision-makers can more effectively identify high-risk FMs and take the appropriate corresponding measures in advance.

To validate the applicability of the proposed approach under a vague and subjective environment, an illustrative example, for a nuclear power plant, was taken into consideration. Via the collection of data from the analysts, the results obtained using three risk-assessment methods, namely the conventional RPN, cost-based RPN, and proposed methods, were compared. The comparison reveals that the proposed method provides a more reasonable and robust ranking system than the other two methods. Sensitivity analysis also confirmed the necessity of risk elements weights evaluation. Decision-makers or risk analysts can use the failure mode risk ranking to determine which part we first maintain and improve. The concept of this article overcomes the traditional methods of qualitative risk assessment. It is feasible and effective to show the degree of risk of failure modes through quantitative calculations. In addition to steam turbine failure mode assessment, the proposed method can also be applied to other nuclear power plant systems or components. Future studies can apply the proposed methodology in other interdisciplinary fields.

Appendix

A. Example of Rough Number Calculation

For the fourth FM (FM_4), the 24 analysts' crisp rankings are 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 7, 10, 10, 9, 9, 8, 10, 8, 9, 8, 10, 9, 9, and 9. By applying the rough number calculation process, the rough interval of the severity importance for the fourth FM is obtained as described in the following text.

Step 1. Obtain the lower and upper limits of the rough numbers

$$\begin{aligned}
&\underline{\text{Lim}}\,(10) = \frac{1}{24}\,(10+10+\cdots+9) = 9.375, \overline{\text{Lim}}\,(10) = \frac{14}{14}\,(10+10+\cdots+10) = 10,\\
&\Rightarrow \text{RN}\,(10) = [9.375,\ 10],\\
&\underline{\text{Lim}}\,(7) = \frac{1}{1}\,(7) = 7, \overline{\text{Lim}}\,(7) = \frac{1}{24}\,(10+10+\cdots+9) = 9.375,\\
&\Rightarrow \text{RN}\,(7) = [7,\ 9.375],\\
&\underline{\text{Lim}}\,(9) = \frac{1}{10}\,(7+9+\cdots+9) = 8.5, \overline{\text{Lim}}\,(9) = \frac{1}{20}\,(10+10+\cdots+9) = 9.7,\\
&\Rightarrow \text{RN}\,(9) = [8.5,\ 9.7],\\
&\underline{\text{Lim}}\,(8) = \frac{1}{4}\,(7+8+8+8) = 7.75, \overline{Lim}\,(8) = \frac{1}{23}\,(10+10+\cdots+9) = 9.478,\\
&\Rightarrow \text{RN}\,(8) = [7.75,\ 9.478].
\end{aligned} \tag{A.1}$$

Step 2. Obtain the interval values of the rough numbers

The set of scores can be obtained by averaging as follows:

$$\text{RN}\,(a_{4S}) = \left[\frac{(9.375+7+8.5+7.75)}{4}, \frac{(10+9.375+9.7+9.478)}{4}\right] = [8.854, 9.834]. \tag{A.2}$$

Data Availability

All data generated or analysed during the study are included in this published article.

Conflicts of Interest

All authors declare that they have no conflicts of interest.

Authors' Contributions

H.-W.L. and Y.-H. L. contributed to the conceptualization; H.-W.L. contributed to the methodology; J.J-H. L. and C.-N. H. contributed to the investigation; J.-J. Y. and C.-N. H. contributed to data curation; H.-W.L. and Y.-H.L. contributed to preparing the original draft; J.J-H. L. performed the review and editing work; and J.-J. Y. and C.-N. H contributed to the project administration. All authors have read and agreed to the final version of the manuscript.

References

[1] L. M. Shakou, J.-L. Wybo, G. Reniers, and G. Boustras, "Developing an innovative framework for enhancing the resilience of critical infrastructure to climate change," *Safety Science*, vol. 118, pp. 364–378, 2019.

[2] F. Aros-Vera, S. Gillian, A. Rehmar, and L. Rehmar, "Increasing the resilience of critical infrastructure networks through the strategic location of microgrids: a case study of Hurricane Maria in Puerto Rico," *International Journal of Disaster Risk Reduction*, vol. 55, Article ID 102055, 2021.

[3] E. G. Quijano, D. Ríos Insua, and J. Cano, "Critical networked infrastructure protection from adversaries," *Reliability Engineering & System Safety*, vol. 179, pp. 27–36, 2018.

[4] J.-S. Chou and C. S. Ongkowijoyo, "Hybrid decision-making method for assessing interdependency and priority of critical infrastructure," *International Journal of Disaster Risk Reduction*, vol. 39, Article ID 101134, 2019.

[5] C.-N. Huang, J. J. H. Liou, and Y.-C. Chuang, "A method for exploring the interdependencies and importance of critical infrastructures," *Knowledge-Based Systems*, vol. 55, pp. 66–74, 2014.

[6] F. Wang, J. J. Magoua, N. Li, and D. Fang, "Assessing the impact of systemic heterogeneity on failure propagation across interdependent critical infrastructure systems," *International Journal of Disaster Risk Reduction*, vol. 50, p. 101818, 2020.

[7] H. H. Xie, F. Zhang, Y. Y. He, J. C. Yang, and W. Q. Ma, "Research on digital prototype technologies of personnel airlock for nuclear power plant," in *Proceedings of the 2018 International Conference on Power System Technology (POWERCON)*, pp. 4684–4691, IEEE, Guangzhou, China, November 2018.

[8] H.-W. Lo and J. J. H. Liou, "A novel multiple-criteria decision-making-based FMEA model for risk assessment," *Applied Soft Computing*, vol. 73, pp. 684–696, 2018.

[9] H.-W. Lo, J. J. H. Liou, C.-N. Huang, and Y.-C. Chuang, "A novel failure mode and effect analysis model for machine tool risk analysis," *Reliability Engineering & System Safety*, vol. 183, pp. 173–183, 2019.

[10] H.-C. Liu, J.-X. You, and C.-Y. Duan, "An integrated approach for failure mode and effect analysis under interval-valued intuitionistic fuzzy environment," *International Journal of Production Economics*, vol. 207, pp. 163–172, 2019.

[11] H.-C. Liu, J.-X. You, X.-F. Ding, and Q. Su, "Improving risk evaluation in FMEA with a hybrid multiple criteria decision making method," *International Journal of Quality & Reliability Management*, vol. 32, no. 7, pp. 763–782, 2015.

[12] W. Wang, X. Liu, X. Chen, and Y. Qin, "Risk assessment based on hybrid FMEA framework by considering decision maker's psychological behavior character," *Computers & Industrial Engineering*, vol. 136, pp. 516–527, 2019.

[13] J. Qin, Y. Xi, and W. Pedrycz, "Failure mode and effects analysis (FMEA) for risk assessment based on interval type-2 fuzzy evidential reasoning method," *Applied Soft Computing*, vol. 89, Article ID 106134, 2020.

[14] O. Mohsen and N. Fereshteh, "An extended VIKOR method based on entropy measure for the failure modes risk assessment - a case study of the geothermal power plant (GPP)," *Safety Science*, vol. 92, pp. 160–172, 2017.

[15] H. Zhang, Y. Dong, I. Palomares-Carrascosa, and H. Zhou, "Failure mode and effect analysis in a linguistic context: a consensus-based multiattribute group decision-making approach," *IEEE Transactions on Reliability*, vol. 68, no. 2, pp. 566–582, 2018.

[16] J. Xiao, X. Wang, and H. Zhang, *"Exploring the ordinal classifications of failure Modes in the reliability management: an optimization-based consensus model with bounded confidences"*, Group Decision and Negotiation, Springer, Berlin, Germany, 2021.

[17] M. Gul, B. Guven, and A. F. Guneri, "A new Fine-Kinney-based risk assessment framework using FAHP-FVIKOR incorporation," *Journal of Loss Prevention in the Process Industries*, vol. 53, pp. 3–16, 2018.

[18] M. Alrifaey, T. Sai Hong, E. Supeni, A. As'arry, and C. Ang, "Identification and prioritization of risk factors in an electrical generator based on the hybrid FMEA framework," *Energies*, vol. 12, no. 4, p. 649, 2019.

[19] S. Yousefi, A. Alizadeh, J. Hayati, and M. Baghery, "HSE risk prioritization using robust DEA-FMEA approach with undesirable outputs: a study of automotive parts industry in Iran," *Safety Science*, vol. 102, pp. 144–158, 2018.

[20] M. Gul, M. F. Ak, and A. F. Guneri, "Pythagorean fuzzy VIKOR-based approach for safety risk assessment in mine industry," *Journal of Safety Research*, vol. 69, pp. 135–153, 2019.

[21] H.-C. Liu, J.-X. You, Y.-Z. Chen, and X.-J. Fan, "Site selection in municipal solid waste management with extended VIKOR method under fuzzy environment," *Environmental Earth Sciences*, vol. 72, no. 10, pp. 4179–4189, 2014.

[22] M. M. Silva, A. P. H. de Gusmão, T. Poleto, L. C. e. Silva, and A. P. C. S. Costa, "A multidimensional approach to information security risk management using FMEA and fuzzy theory," *International Journal of Information Management*, vol. 34, no. 6, pp. 733–740, 2014.

[23] H. Safari, Z. Faraji, and S. Majidian, "Identifying and evaluating enterprise architecture risks using FMEA and fuzzy VIKOR," *Journal of Intelligent Manufacturing*, vol. 27, no. 2, pp. 475–486, 2016.

[24] W. Suo, J. Zhang, and X. Sun, "Risk assessment of critical infrastructures in a complex interdependent scenario: a four-stage hybrid decision support approach," *Safety Science*, vol. 120, pp. 692–705, 2019.

[25] E. Roe and P. R. Schulman, "A reliability & risk framework for the assessment and management of system risks in critical

infrastructures with central control rooms," *Safety Science*, vol. 110, pp. 80–88, 2018.

[26] C. Große and P. M. Olausson, "Blind spots in interaction between actors in Swedish planning for critical infrastructure protection," *Safety Science*, vol. 118, pp. 424–434, 2019.

[27] R. Lin, Y. Man, C. K. M. Lee, P. Ji, and J. Ren, "Sustainability prioritization framework of biorefinery: a novel multi-criteria decision-making model under uncertainty based on an improved interval goal programming method," *Journal of Cleaner Production*, vol. 251, Article ID 119729, 2020.

[28] L. Wang, F. Yan, F. Wang, and Z. Li, "FMEA-CM based quantitative risk assessment for process industries-A case study of coal-to-methanol plant in China," *Process Safety and Environmental Protection*, vol. 149, pp. 299–311, 2021.

[29] M. Mei and Z. Chen, "Evaluation and selection of sustainable hydrogen production technology with hybrid uncertain sustainability indicators based on rough-fuzzy BWM-DEA," *Renewable Energy*, vol. 165, pp. 716–730, 2021.

[30] E. K. Tekez, "Failure modes and effects analysis using fuzzy topsis in knitting process," *Tekstil ve Konfeksiyon*, vol. 28, no. 1, pp. 21–26, 2018.

[31] J. Joseph, K. V. Sriram, A. O. Mathew, and A. Kanoor, "The failure mode effects analysis to prevent delays for development projects in aerospace industry—a case study," *International Journal of Mechanical and Production Engineering Research and Development*, vol. 9, no. 2, pp. 123–136, 2019.

[32] D. M. Popa, "Engineering risk of series production in automotive industry," *Management Systems in Production Engineering*, vol. 27, no. 1, pp. 5–11, 2019.

[33] A. Mora, L. Ayala, R. Bielza, F. Ataúlfo González, and A. Villegas, "Improving safety in blood transfusion using failure mode and effect analysis," *Transfusion*, vol. 59, no. 2, pp. 516–523, 2019.

[34] M. Gul and M. F. Ak, "A modified failure modes and effects analysis using interval-valued spherical fuzzy extension of TOPSIS method: case study in a marble manufacturing facility," *Soft Computing*, vol. 25, no. 8, pp. 6157–6178, 2021.

[35] M. Yucesan and M. Gul, "Failure modes and effects analysis based on neutrosophic analytic hierarchy process: method and application," *Soft Computing*, vol. 25, pp. 1–18, 2021.

[36] M. Yucesan, M. Gul, and E. Celik, "A holistic FMEA approach by fuzzy-based Bayesian network and best–worst method," *Complex & Intelligent Systems*, vol. 7, pp. 1–18, 2021.

[37] Z. a. Pawlak, "Rough sets," *International Journal of Computer & Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.

[38] Z. Pawlak, "Rough sets: theoretical aspects of reasoning about data," *Springer Science & Business Media*, vol. 9, 2012.

[39] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[40] C. L. Hwang and K. S. Yoon, *Multiple Attribute Decision Making: Methods and Applications*, Springer-Verlag, Berlin, 1981.

[41] T. Kuo, "A modified TOPSIS with a different ranking index," *European Journal of Operational Research*, vol. 260, no. 1, pp. 152–160, 2017.

[42] M. J. Rezaee, A. Salimi, and S. Yousefi, "Identifying and managing failures in stone processing industry using cost-based FMEA," *International Journal of Advanced Manufacturing Technology*, vol. 88, no. 9–12, pp. 3329–3342, 2017.

[43] H.-W. Lo, J. J. H. Liou, C.-N. Huang, Y.-C. Chuang, and G.-H. Tzeng, "A new soft computing approach for analyzing the influential relationships of critical infrastructures," *International journal of critical infrastructure protection*, vol. 28, Article ID 100336, 2020.

[44] S. Yousefi, M. Valipour, and M. Gul, "Systems failure analysis using Z-number theory-based combined compromise solution and full consistency method," *Applied Soft Computing*, vol. 113, Article ID 107902, 2021.

[45] H.-W. Lo, C.-C. Hsu, C.-N. Huang, and J. J. H. Liou, "An ITARA-TOPSIS based integrated assessment model to identify potential product and system risks," *Mathematics*, vol. 9, no. 3, p. 239, 2021.

Hindawi
Complexity
Volume 2021, Article ID 7998655, 17 pages
https://doi.org/10.1155/2021/7998655

Research Article

Social Stability Risk Diffusion of Large Complex Engineering Projects Based on an Improved SIR Model: A Simulation Research on Complex Networks

Zhengqi He,[1,2] **Dechun Huang**,[1,2] **and Junmin Fang**[1]

[1]*Business School, Hohai University, Nanjing 211100, China*
[2]*Jiangsu Provincial Collaborative Innovation Center of World Water Valley and Water Ecological Civilization, Nanjing 211100, China*

Correspondence should be addressed to Junmin Fang; fangjunmin@hhu.edu.cn

Received 11 November 2021; Revised 10 December 2021; Accepted 14 December 2021; Published 26 December 2021

Academic Editor: Qingyuan Zhu

The development of China's new urbanization has driven the rapid increase in large complex engineering projects, which have the characteristics of large-scale investment, long-term construction, and wide social influence, easily causing benefit conflicts among relevant stakeholders, and breaking out social stability risks. In the previous research, the risks of large complex engineering projects mainly concentrated on the assessment of economic risk, schedule risk, etc. However, there were few studies on social risks, and they did not consider how the risks spread on the complex networks based on the social connections such as interpersonal relationship. From the subject of social stability risk diffusion of large complex engineering projects, this paper constructs a related risk diffusion model based on the SIR model to analyze risk diffusion mechanism. Through NetLogo simulation platform, the model is placed under a small-world network environment that is closest to the topology structure of real social interpersonal relationship network for simulation research, aiming to find out key factors of social stability risk intervention for large complex engineering projects, which greatly contributes to the social stability risk management of large complex engineering projects.

1. Introduction

With the acceleration of economic growth and modernization, large complex engineering projects have entered a period of rapid growth, making great contributions to social and economic development. Taking a major water conservancy project as an example, the Chinese government plans to construct 150 projects from 2020 to 2022, with a total investment of 1,290 billion yuan, which can drive direct or indirect investment of 6,600 billion yuan. While different from general projects, the large complex engineering projects have a high degree of complexity and uncertainty. With the characteristics of large-scale investment, long period of construction, as well as numerous stakeholders that involves the problem of multi-party interest balance, it can easily lead to social conflicts and form social stability risk events extending over a large area and with a wide range of impacts. Taking the currently controversial PX project in China as an example, some newly constructed PX projects have triggered many social stability risk events due to the adverse impact on the environment. For instance, the social stability risk event triggered by "Ningbo PX incident" caused direct economic loss of 6.4 billion yuan. There are also some other social stability risk events triggered by large complex engineering projects that have a great impact on the normal order of society and economy. For example, the "Qidong Incident" triggered by the sewage disposal of a paper mill led to a demonstration by about 2,000 people, severely affecting the image of the local government, social order, and economic development. The "Suijiang 325 Incident" triggered by the construction of the Xiangjiaba major water conservancy project caused traffic interruption for more than 100 hours, greatly affecting local social stability.

The social stability risks have always attracted the attention of the Chinese government. In the report on the work of the government in 2015, it emphasized that “social stability risks which major policy decisions may pose are fully assessed so as to effectively prevent and resolve social conflicts.” In 2017, the 19th National Congress of the Communist Party of China pointed out that “resolutely fight the battle to prevent and resolve major risks” and “strengthen the construction of mechanisms for preventing and resolving social conflicts.” All have emphasized the importance of preventing and resolving social stability risks. In China’s current context of emphasis on preventing and resolving major risks, since social stability risks of large complex engineering projects belongs to the category of major risks, it has also become the focus of attention from all walks of life and the hot topic in academic research.

The social system is a special complex large system. The conflict of interests among subjects is the root of the social stability risks of large complex engineering projects, and the complex network formed by the interaction among subjects is an important carrier for the spread of social stability risks. Although some regions and departments in China have established local social stability risk assessment systems, and achieved certain outcomes in social stability risk management, social stability risks in large complex engineering projects still occur and have not been fundamentally avoided, causing the failure of the “stability evaluation” safety valve. The reason is that, on the one hand, it is difficult to effectively deal with subject conflicts under the existing social stability risk system of large complex engineering projects. On the other hand, it ignores the influence of complex social network relationships on the diffusion process of social stability risks. Many systems in the real world can be described with complex networks. With the rise of social networks such as Weibo, WeChat, and Twitter, the research on complex networks has expanded from the complex network of natural sciences to that of human society, and has become the focus of attention in the academic and practical circles. On the complex network, the conflicts among the subjects of large complex engineering projects demonstrate some new characteristics. For example, the different network structures formed among the subjects have changed the evolution results of the subject conflicts and significantly affected the process of social stability and risk diffusion. These new features weaken the effective understanding and accurate grasp of traditional engineering project impact and social risk management theory on the social stability risk diffusion mechanism of large complex engineering projects. Therefore, by ignoring the influence of complex networks on the spread of social stability risks of large complex engineering projects, and conducting research only from the social impact of conflicts of interest in large complex engineering projects, neither the diffusion mechanism of social stability risks of large complex engineering projects can be grasped, nor a comprehensive, effective, and realistic social stability risk management mechanism can be put forward. As a result, the “safety valve” of the institutionalized arrangements for social stability risk management of large complex projects will fail. Therefore, under the complex network environment, studying the social stability risk diffusion process of large complex engineering projects and proposing a social stability risk management mechanism that conforms to its scientific laws is an alternative way to solve the social stability risk problem of large complex engineering projects in China, which is also an urgent task in practice.

2. Literature Review

2.1. Research on the Impact of Large Complex Engineering Projects on Social Stability. As the characteristics of large-scale investment have a wide range of impacts, the construction of large complex engineering projects will generally affect the local ecological environment and human social environment [1–4]. Especially, in the process of reconstruction of nature and human society, the evolution of various social elements in time and space will occur. This is an uncontrollable variable, which will have a huge impact on social stability without proper handling, causing social conflicts and social stability risks [5, 6]. Taking the construction of a mega industrial park project as an example, Liu et al. [7] analyzed its impact on social stability, and proposed a practical framework of social risk management. He et al. [8] studied the impact of large-scale engineering projects on social stability from the perspective of stakeholder conflicts.

2.2. Research on Social Stability Risks of Large Complex Engineering Projects. Social stability can be literally explained as a steady, settled, and harmonious state of politics, economy, culture, ecology, etc. It includes not only political stability and economic stability but also normal social order and peace of mind. In a broad sense, social risks can be understood as the social instability and disharmony caused by the political, economic, cultural, ecological, and other risk factors, triggering social conflicts, and endangering social stability and social order with certain probability, which is accumulated to a certain extent on the subject’s interactive relationship network and becomes reality; social risks will evolve into social crises, which will cause a catastrophic impact on social stability and social order. Social stability risks can be defined as the shift of steady status for some uncertain factors, resulting in the risks of social disorder and social environment disharmony in the social system [9]. In recent years, with the Chinese government attaching great importance to social stability risk management, research on social stability risks of large complex engineering projects has gradually become a hot spot. Scholars’ research mainly focuses on two aspects. On the one hand, from the perspective of social stability, risk assessment to construct an assessment index system and model for empirical research is of great importance [10, 11]. Wu et al. [12] established an assessment index system about social stability risks of major water conservancy projects in fragile eco-environment regions (FEER), and by using the set pair analysis (SPA) and the principal component analysis (PCA) methods, they established an assessment model for social stability risks, which was applied to the major water conservancy projects in Shaanxi Province, China. Ma and Cong [13] used cloud model and ordered weighted averaging and matter-element model to establish an assessment model for social stability risks of Not In My Back Yard major projects. On the other hand, many scholars have focused on the social stability risks of specific types of large complex engineering projects, studying their formation and prevention mechanisms. It includes major water conservancy projects [9, 14], power energy projects [15, 16], urban demolition projects [17–19], transportation infrastructure projects [20, 21], environmental engineering Projects [22, 23] and so on.

2.3. Research on Risk Diffusion Based on Infectious Disease Model. Once the social stability risks of a large complex engineering project form, it is prone to spread in the society due to its high-risk, unexpected, and diffused characteristics, leading to a wider range of social instability, and the proliferation of social stability risks gradually attract the attention of scholars [24, 25]. The most common method for the risk diffusion research is the infectious disease model, including those like SIS model and SIR model, which are mainly applied in the medical field to study the spread of various infectious diseases [26]. Based on improving the SIR model, Jiao and Huang [27] constructed the SIHR model containing four types of subjects to study the infectiousness of COVID-19 and put forward a policy to control the spread of COVID-19. With the development of research, infectious disease models have gradually expanded from the medical field to the risk diffusion research in the economic, social, and other fields [28, 29]. Among them, there are many researches on the spread of rumors, which is quite similar to that of social stability risks. Li et al. [30] constructed a SIHR model based on probability generating function to study the spread of rumors, and conducted numerical simulation to examine the differences between rumor transmission recovery mechanism and disease transmission recovery mechanism. Qiu et al. [31] constructed a novel SIR rumor spreading model with the influence mechanism to study the spread of rumors in society. Sun et al. [32] constructed an uncertain SIR rumor spreading model to study the influence of perturbation in the transmission mechanism of rumor spreading.

2.4. Research on Risk Diffusion on Complex Networks. The well-known physicist Hawking said that, "the 21st century will be the century of complexity science." Complexity science has triggered tremendous changes in the natural science. The basis of complex network theory is the random graph theory established by famous mathematicians, Erdos and Renyi [33]. Qian et al. [34] gave a stricter definition of complex networks: networks with some or all of the properties of self-organization, self-similarity, attractor, small world, and scale-free are called complex networks. With the deepening of people's research on complex networks, it has become a highly active interdisciplinary research field, encompassing multiple disciplines such as computer science, statistics, physics, and social sciences. At present, scholars' research on complex networks is mainly concentrated on two aspects: First of all, based on the analysis of the realist world by powerful data analysis tools to study the topological characteristics of realist complex networks [35]. Secondly, use complex network theory to study the normal operation of many realist systems [36, 37]. In reality, the spread of various risks can be regarded as the infectious dynamic behavior on complex networks that follow specific laws, while the infectious dynamic behavior has always received extensive attention from scholars [38, 39]. Research on risk diffusion on complex networks has been widely used in supply chain, finance, project management, and other fields [40–43]. In the late 20th century, the "Collective dynamics of 'small-world' networks" [44] published in Nature magazine and the "Emergence of scaling in random networks" [45] published in Science magazine proposed the small-world network model and the scale-free network model, providing a way for researchers to describe the actual system with networks. It further verified that the real-world network is a network with small-world effects and scale-free characteristics, and provides a realistic and feasible way to solve many important issues in the real world. With the in-depth study on various types of complex networks such as small-world network and scale-free network, infectious models based on small-world network and scale-free network have also been studied in depth [46]. Many scholars have proposed effective risk prevention strategies by exploring the spread of various risks on different and specific complex network structures [47–49].

Therefore, this paper will analyze the relevant characteristics of the social stability risk diffusion of large complex engineering projects, discuss the applicability of the infectious disease model, improve the traditional SIR model, and have a simulation analysis on complex networks to study the diffusion process of social stability risks of large complex engineering projects.

3. Social Stability Risk Diffusion Model of Large Complex Engineering Projects Based on an Improved SIR Model

3.1. Definition of Large Complex Engineering Projects. The project is a very broad concept, ranging from the construction of a basketball court to the construction of a national high-speed railway. A project can span decades, like the construction of the Three Gorges Project, or it can be completed in a few days, like the renovation of a football field lawn. It can stretch thousands of kilometers in space, such as the Beijing-Shanghai high-speed rail, or it can be completed in a small office, such as the development of an enterprise management platform. The investment of a project can reach hundreds of billions, such as the South-to-North Water Diversion Project, or less to tens of thousands, such as the construction of a country road. With the rapid development of China's economy, the scale of engineering projects is increasing, and the projects are getting more and more complicated. Projects with investment of billions or even tens of billions are emerging in an endless stream. Regarding which standard the investment quota reaches can be regarded as large complex engineering projects, people from all circles of life have not reached a consensus. This paper believes that large complex engineering projects cannot be simply defined from the investment quota but should be considered in terms of investment scale, construction period, technical difficulty, and social influence. The large complex engineering projects studied in this paper refer to those characterized by huge construction scale, huge investment scale, many factors involved, and those that have a major and far-reaching impact on the region and the entire country. The engineering projects here have a wide range of connotations, including water conservancy projects, high-speed railways, large-scale petroleum and chemical projects, etc., such as the Three Gorges Project, the South-to-North Water Diversion Project, the Beijing-Shanghai High-speed Railway, the Guangdong-Hong Kong-Macao Bridge, the West-East Gas Pipeline, and strategic Petroleum reserve project, etc.

3.2. The Subjects of Social Stability Risk Diffusion of Large Complex Engineering Projects. After the social stability risks of large complex engineering projects form, due to their diffusion characteristics, they will spread among key subjects such as the government, project legal persons, and local people as well as the public, the media, and other relevant subjects, shaping a complex network of social stability risk diffusion of large complex engineering projects. On this complex network, different types of subjects such as the government, project legal persons, local people, and the public have different understanding of and attitudes towards large complex engineering projects, leading to different attitudes towards risk diffusion, and the different risk perceptions further make their risk spreading capabilities different. Some subjects are not supportive of large complex engineering projects, or for the reason that the benefit compensation received is lower than what was expected, when the social stability risks caused by large complex engineering projects spread out, they will actively participate in the diffusion process and become a new transmission source.

There are also some other subjects who support the large complex engineering projects, or as the benefit compensation meets their psychological expectation, they will not get involved when the social stability risks spread out, but choose to ignore it or even try to persuade others. These subjects are divided based on the role they play in the spread of social stability risks of large complex engineering projects. For example, some of the local people will directly participate in the spread of social stability risks, while some will not and become the new type of subjects in the risk diffusion. Similarly, for other subjects such as the government and the public, a certain proportion of subjects in each type will participate or not participate in the spread of social stability risks. By combining with the attitudes of the subjects involved in the spread of social stability risks and the division of the subjects of the classic infectious disease model, this paper divides the subjects of social stability risk diffusion of large complex engineering projects into four categories: the ignorant, malcontents, staggerers, and the rational.

(1) *The Ignorant*: in the process of spreading the social stability risks of large complex engineering projects, the ignorant refer to the kind of subjects who are not familiar with the large complex engineering projects and the related information about social stability risks, and they will not spread the social stability risks without being exposed to social stability risk information.

(2) *Malcontents*: the malcontents refer to those subjects who are resistant to the large complex engineering projects and dissatisfied with the relevant benefit compensation, so that they appeal for benefits by violent protests and other means and actively spread the social stability risks.

(3) *Staggerers*: the staggerers can be understood as those subjects who have known the social stability risk information related to the large complex engineering projects. However, for certain reasons such as their own lack of active participation and that benefit compensation has reached expectations, they maintain a wait-and-see attitude at this stage and will not actively participate in the spread of social stability risks temporarily.

(4) *The Rational*: the rational are the subjects who get access to a large amount of information about social stability risks of the large complex engineering projects. However, due to certain reasons such as support for large complex engineering projects and social stability maintenance, they do not participate in the spread of social stability risks (or are immune to social stability risks). The difference between the rational and staggerers is that the latter do not participate in the spread of social stability risks for the time being, but when they get access to a certain amount of social stability risk information, they may become a transmission source. While the rational are already immune to risks, they will not participate in the spread of any social stability risks.

3.3. Interactive Model of Social Stability Risk Diffusion of Large Complex Engineering Projects. In the process of spreading the social stability risks of large complex engineering projects, the four types of subjects, namely, ignorant, malcontents, staggerers, and the rational, interact and influence each other, forming a diffusion network about social stability risks. In this process, on the basis of certain social connections, one subject can pass the social stability risk information to other subjects after receiving it. Therefore, the diffusion network can be regarded as a directional and powerless network with people as nodes and relationships among people as edges, as shown in Figure 1. When perceiving the social stability risks, subject A will spread it to those subjects in contact with it such as B, C, and D. B and C may not be interested in the risks or may doubt their

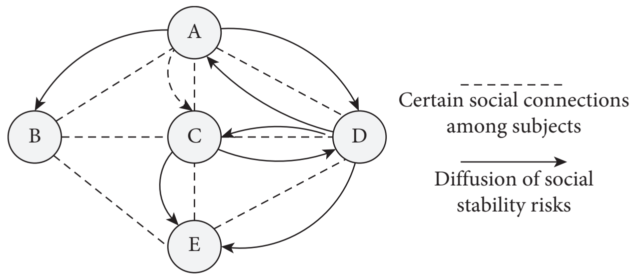


Figure 1: Network diagram of social stability risk diffusion of large complex engineering projects.

existence after receiving the risk information, so they will not spread them again. While affected by the risk information and perceiving the social stability risks, subject *D* will panic and further spread the risks to those subjects in contact with it such as *A*, *C*, and *E*. Subject *C* is not used to participating in the spread of social stability risks, but due to the increased exposure to risk information, it gradually turns into a communicator, further spreading social stability risks to related subjects with connections.

The interactive model of social stability risk diffusion of large complex engineering projects is shown in Figure 2, with four subjects of the ignorant, malcontents, staggerers, and the rational. When the ignorant perceive the social stability risks, some become malcontents, some become staggerers, and the others become the rational. Those who become malcontents will spread the risks again, causing the rise in the number of malcontents. Those who become staggerers will not spread risks temporarily but may turn to malcontents with increased connections with social stability risk information. In the process of social stability risk diffusion, some malcontents will turn to staggerers under appropriate social handling or for some personal reasons they will not spread risks for the moment, while the others will not participate in the spread of risks and become the rational.

3.4. Infectious Disease Characteristics of Social Stability Risk Diffusion of Large Complex Engineering Projects. The spread of social stability risks of large complex engineering projects is essentially the process wherein various subjects perceive the social stability risks caused by the construction of large complex engineering projects and spread them to other subjects with connections through various transmission channels. Infectious diseases generally have the characteristics such as pathogens, infectious agents, infectiousness, and immunity, to which the spread of social stability risks of large complex engineering projects is similar.

(1) Pathogen of Social Stability Risk Diffusion of large complex Engineering Projects: In the process of social stability risk diffusion of large complex engineering projects, the pathogen is the source of risks, namely, the various conflict events triggered during the construction of large complex engineering projects. The source of risks is the prerequisite for the spread of risks, or risk diffusion will not exist. When interest conflicts among subjects triggered by large complex engineering projects occur, group incidents break out, forming a source of social stability risks, which will be further diffused through infectious media to other relevant subjects with connections.

(2) Infectious media of Social Stability Risk Diffusion of large complex Engineering Projects: The infectious medium is the carrier through which the risk source transmits. The infectious medium of social stability risks of large complex engineering projects is the relationship network among the subjects. It includes not only the Internet, newspapers, and other new media and traditional media but also the various social relationships formed in the regions. Once the source of social stability risks form, risk information spreads to a larger area through these infectious media, which have a huge impact on the entire society.

(3) The Infectiousness of Social Stability Risk Diffusion of Large complex Engineering Projects: Like infectious diseases, the spread of social stability risks of large complex engineering projects is also infectious. When perceiving the social stability risks, a subject will spread it to the external environment through the infectious media. Especially, in today's era, when the Internet is so developed, risk information will spread rapidly. Under the great impact of large complex engineering projects, other subjects will also accept the risk information and form a certain degree of panic. These subjects will become staggerers and malcontents. With the increased degree of impact, the staggerers may further turn into malcontents. This in turn affects the surrounding subjects, thus forming the infectiousness of the social stability risk diffusion of large complex engineering projects.

(4) Immunity of Social Stability Risk Diffusion of large complex Engineering Projects: In the process of infectious disease transmission, some people will build immunity due to their antibodies and will no longer be infected by infectious diseases. Similarly, in the spread of social stability risks of large complex engineering projects, some subjects will not be affected by social stability risk information due to their relatively complete knowledge structure, a good understanding of large complex engineering projects, and strong psychological endurance. They will not participate in the spread and become rational people of the social stability risks of large complex engineering projects, that is, they have a certain degree of immunity to the social stability risks.

3.5. Applicability of the Infectious Disease Model. The transmission model of infectious diseases originated from the analysis of smallpox infection in the 18th century. With the deepening of research, scholars have successively constructed some infectious disease models that quantitatively study the spread of infectious diseases, gradually expanding

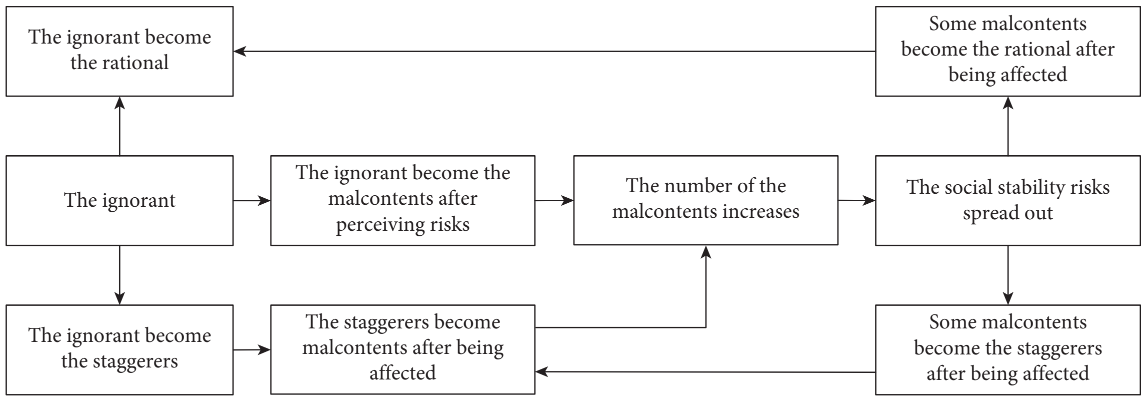


Figure 2: The interactive model of social stability risk diffusion of large complex engineering projects.

from the field of medical research to general transmission mechanisms. Infection models such as SI model, SIS model, and SIR model were formed. Based on the SIR model, the follow-up research content of this paper will improve it, so the SIR model will be mainly introduced in the following part. There are three types of people in the SIR model. In addition to healthy people and infected patients, there is also an immune group. Immune group means that the infected patient has immune function after being cured or disappears in the group, and will no longer have any influence on the infection process. In the SIR model, S represents the population in a healthy state, I represents those who have been infected, and R represents the immune population. The SIR model has the following assumptions (Figure 3):

(1) $S(t)$ represents the proportion of people in a healthy state at time t, $I(t)$ represents the proportion of people in an infected state at time t, and $R(t)$ represents the proportion of people in an immune state after being cured at time t.
(2) People in a healthy state will be infected by infected patients with a certain probability to become patients, and we will record the infection rate as λ.
(3) Patients who have been infected will be cured with a certain probability and become healthy people. We record the cure rate of patients as μ. People in the cured immune group will not be infected again, nor will they participate in any infection process (Figure 4).

Figure 4 shows the infection process of the SIR model. According to the above assumptions, there are $\lambda S(t)I(t)$ people with a healthy state who are infected and become patients, and $\mu I(t)$ patients are cured and become immune populations at every moment. Therefore, the SIR model can be obtained as:

$$\begin{cases} \dfrac{\mathrm{d}S(t)}{\mathrm{d}t} = -\lambda S(t)I(t), \\ \dfrac{\mathrm{d}I(t)}{\mathrm{d}t} = \lambda S(t)I(t) - \mu I(t), \\ \dfrac{\mathrm{d}R(t)}{\mathrm{d}t} = \mu I(t). \end{cases} \quad (1)$$

From this differential equation, it can be seen that the number of infected people will gradually increase. When the time reached a certain point, the number of infected people will start to decrease due to the decline in the number of healthy people until the number of infected patients is reduced to zero.

The spread of social stability risks of large complex engineering projects has the characteristics of infectious diseases, and the spread on social network is similar in terms of spreading environment, spreading mechanism, spreading direction, and spreading stage, etc.

(1) *Diffusion environment*: Infectious diseases spread on social network based on the connections among people. Human individuals are equivalent to the nodes of social network, and communication among people is the way of infection, which is equivalent to the edges of a social network. While the social stability risks of large complex engineering projects also spread among stakeholders, namely, the nodes, connections, and communication among related subjects are the edges. Therefore, the spreading environment of infectious diseases is similar to that of social stability risks of large complex engineering projects, and both spread on a complex network.
(2) *Diffusion mechanism*: When an infectious disease spreads, the virus carrier infects others through connections, and then the infected people further infect other people with connections. The transmission is based on the direct connections among people, and the virus is transmitted outward by depending on people's connections; this is the mechanism of virus transmission. The spread of social stability risks of large complex engineering projects is similar. After perceiving the social stability risks, the relevant subjects will pass risk information to subjects who have connections with them, and further spread it to other subjects and through them to the entire society. The position and capability of the individual on the network determines the infectiousness of the social stability risks. Therefore, the spreading mechanism of infectious diseases is similar to that of social stability risks of large complex engineering projects.

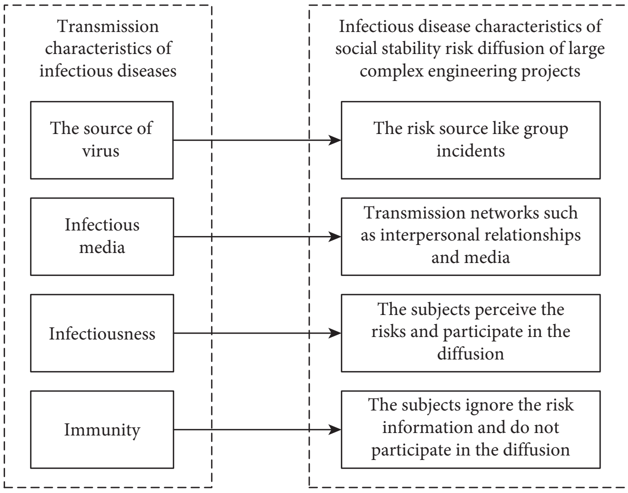


Figure 3: The infectiousness of social stability risk diffusion of large complex engineering projects.

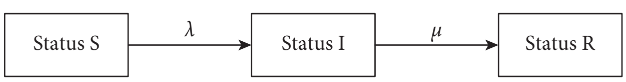


Figure 4: Schematic diagram of the infection process of the SIR model.

(3) *Diffusion Direction*: The spread of infectious diseases is radial and directionless. In the process of spreading social stability risks of large complex engineering projects, although the relationship among the subjects has certain sequential characteristics due to different status, the spread of risk information is also radial and can be affected by each other. There is no one-way influence, that is, the spread of social stability risks is also directionless. Therefore, the diffusion direction of infectious diseases is similar to that of social stability risks of large complex engineering projects.

(4) *Diffusion Stage*: There are the incubation period, the outbreak period, and the recovery period in the spread of infectious diseases. There are also certain periods in the social stability risks of large complex engineering projects. When the risk information spreads out, not all the subjects will become malcontents and actively participate in the spread of social stability risks, as some will take a wait-and-see attitude. At this time, the social stability risks are not very serious, which is equivalent to the incubation period. With the group incidents becoming critical, social stability risks will erupt on a large scale, and eventually die out because of the disposal of all parties. Therefore, the diffusion stage of infectious diseases is similar to that of social stability risks in the large complex engineering projects.

To sum up, the spread of social stability risks of large complex engineering projects on the social network is similar to that of infectious diseases, so it is very appropriate to introduce infectious disease models into the research on the social stability risks of large complex engineering projects.

3.6. IMSR Model Construction. Based on the traditional SIR model, this paper constructs the IMSR model for social stability risk diffusion of large complex engineering projects and improves it from the following four aspects. First of all, compared with the three types of subjects in the traditional SIR model, this paper increases the model subjects into four types. According to the previous subject analysis of social stability risk diffusion of large complex engineering projects, the subjects in the hypothetical model include four categories: the Ignorant (abbreviated as I), Malcontents (abbreviated as M), Staggerers (abbreviated as S), and The Rational (abbreviated as R). Secondly, assuming that the subjects have memory and forgetting functions as time passed, the staggerers and malcontents will spontaneously change their state with a certain probability if not contacting other types of subjects. This is called memory mechanism and forgetting mechanism [50]. Thirdly, after the subject is exposed to social stability risks, it is assumed that there is an acceptance probability, which is not fixed since it will be affected by external influences such as the amount of risk information and the social effect of the risks. Lastly, assuming that the social stability risks of large complex engineering projects spread under the complex network topology, by learning from the average field theory in statistical physics and considering the influence of the network node degree in the model, the average degree of the complex network is introduced into the model.

Based on the previous analysis of the four types of subjects, this paper gives the diffusion rules of social stability risks:

(1) When contacting the malcontents, the ignorant will be affected by the social stability risks diffused. During the spread of social stability risks, there is a probability $p(x)$ for the ignorant to become a malcontent, and we called it risk diffusion probability.

(2) When contacting the malcontents, the ignorant temporarily neither spread the social stability risks nor explicitly reject them. The probability of becoming a staggerer is α, since the staggerers just temporarily do not spread social stability risks, we call α the risk dormancy rate.

(3) When contacting the malcontents, the ignorant will refuse to spread the risks on the ground of not being affected by social stability risks. The probability of becoming the rational is β, which is called the risk rejection rate.

(4) It is believed that the staggerers have memory function. With the increase in exposure to risk information or for other influencing factors, they will be awakened spontaneously with a certain probability θ and become the malcontents. We call θ the risk recovery rate. At the same time, after contacting

the malcontents, there is a probability $p(x)$ for the staggerers to turn into the malcontents.

(5) It is believed that the malcontents have forgetting function. As time goes by, the malcontents tend to calm down and they will spontaneously turn into staggerers with a certain probability σ; as they will not participate in the spread of social stability risks for the time being, we call σ the risk hibernation rate.

(6) When the malcontents contact other malcontents, staggerers, or the rational, there is a probability μ that the malcontents will turn into the rational under the persuasion of other subjects, and will no longer spread the social stability risks, and we call μ the risk extinction rate.

The risk diffusion framework of IMSR model is shown as Figure 5.

It is believed that the risk diffusion probability is not a fixed value, but it is related to how many times the ignorant contact the malcontents. Drawing on the assumption of the rumor acceptance probability in the literature [51], this paper considers the risk diffusion rate as:

$$p(x) = \left|(M-\rho)e^{(-|\mathrm{SI}|\times(x-1))} - M\right|. \tag{2}$$

In formula (2), x represents the cumulative number of times that the ignorant have contacted the malcontents. M is a function of SI, and $|\mathrm{SI}|$ represents the degree of risk amplification or restraint in society. When $\mathrm{SI} > 0$, it means that risks have been amplified in society, and $M = 1$ at this time. When $\mathrm{SI} < 0$, it means risks have been restrained in society, and $M = 0$. $\rho = p(1)$ represents the risk diffusion probability when the ignorant first contact the malcontents. It can be obtained that when there is a social amplification effect of risks, the risk diffusion probability can be expressed as:

$$p(x) = 1 - (1-\rho)e^{(-|\mathrm{SI}|\times(x-1))}. \tag{3}$$

When there is a social restraint effect of risks, the risk diffusion probability can be expressed as:

$$p(x) = \rho e^{(-|\mathrm{SI}|\times(x-1))}. \tag{4}$$

It respectively shows the changing curves of risk diffusion probability when there is a social amplification effect of risks and a social restraint effect of risks, as shown in Figure 6.

As shown in Figure 6(a), when there is a social amplification effect of risks, the risk diffusion probability increases with the number of times the people contact the malcontents; the greater the degree of social amplification, the faster the growth of risk diffusion probability. At this time, because of the amplification effect of risks, every time the ignorant contact the malcontents, the perceived risks become greater and greater, leading to the greater risk diffusion probability. As shown in Figure 6(b), when there is a social restraint effect of risks, the risk diffusion probability decreases with the number of times the people contact the malcontents, and the greater the degree of risk restraint, the slower the decrease of the risk diffusion probability. At this time, because of the amplification restraint of risks, every time the ignorant contact the malcontents, the ignorant will dislike the risk information transmitted by the malcontents, and become less willing to participate in the risk diffusion as the number of contacts increases; thus, the risk diffusion probability decreases.

On a complex network, $I(t)$, $M(t)$, $S(t)$, and $R(t)$, respectively, represent the density of the ignorant, malcontents, staggerers, and the rational at time t. According to the given risk diffusion rules of large complex engineering projects, dynamic equation of IMSR model is got as below:

$$\begin{cases} \dfrac{\mathrm{d}I(t)}{\mathrm{d}t} = -[p(x)+\alpha+\beta]\langle k\rangle I(t)M(t), \\[2ex] \dfrac{\mathrm{d}M(t)}{\mathrm{d}t} = p(x)\langle k\rangle I(t)M(t) - \mu\langle k\rangle M(t)[M(t)+S(t)+R(t)] - \sigma M(t) + \theta S(t) + p(x)\langle k\rangle S(t)M(t), \\[2ex] \dfrac{\mathrm{d}S(t)}{\mathrm{d}t} = \alpha\langle k\rangle I(t)M(t) + \sigma M(t) - \theta S(t) - p(x)\langle k\rangle S(t)M(t), \\[2ex] \dfrac{\mathrm{d}R(t)}{\mathrm{d}t} = \beta\langle k\rangle I(t)M(t) + \mu\langle k\rangle M(t)[M(t)+S(t)+R(t)]. \end{cases} \tag{5}$$

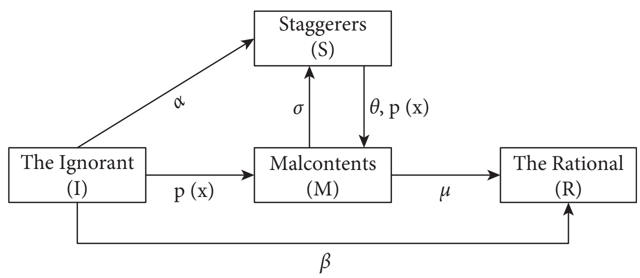


FIGURE 5: The risk diffusion framework of IMSR model.

Among them, $I(t)+M(t)+S(t)+R(t)=1$, and $\langle k\rangle$ represents the average degree of the complex network.

From the above formula, it gets:

$$\begin{aligned}\frac{dR(t)}{dI(t)} &= \frac{\beta\langle k\rangle I(t)M(t)+\mu\langle k\rangle M(t)[M(t)+S(t)+R(t)]}{-[p(x)+\alpha+\beta]\langle k\rangle I(t)M(t)}\\ &= \frac{\beta\langle k\rangle I(t)M(t)+\mu\langle k\rangle M(t)[1-I(t)]}{-[p(x)+\alpha+\beta]\langle k\rangle I(t)M(t)}\\ &= \frac{(\mu-\beta)I(t)-\mu}{[p(x)+\alpha+\beta]I(t)}.\end{aligned} \tag{6}$$

It further gets:

$$dR(t) = \frac{\mu-\beta}{p(x)+\alpha+\beta}dI(t) - \frac{\mu dI(t)}{[p(x)+\alpha+\beta]I(t)}. \tag{7}$$

Assuming that there are a total of N subjects on the network, when the social stability risks break out at the beginning, there is only one malcontent, and the rest are the ignorant. Then, there are:

$$\begin{aligned}I(0) &= \frac{N-1}{N},\\ M(0) &= \frac{1}{N},\\ S(0) &= 0,\\ R(0) &= 0.\end{aligned} \tag{8}$$

In particular, when the number of subjects is large enough, $I(0)=\lim_{N\longrightarrow\infty}N-1/N=1$. In the process of spreading social stability risks of large complex engineering projects, as time goes by, the number of malcontents increases, and then gradually reduces to zero. At this time, the network reaches a steady state, and there are only the ignorant and the rational. Let $R(\infty)=R$, then $I(\infty)=1-R(\infty)=1-R$, and integrating formula (7), it gets:

$$R = -\frac{\mu-\beta}{p(x)+\alpha+\beta}R - \frac{\mu}{p(x)+\alpha+\beta}\ln(1-\beta). \tag{9}$$

Then, it gets:

$$R = 1-e^{-(p(x)+\alpha+\mu/\mu)R}. \tag{10}$$

Let $\tau = p(x)+\alpha+\mu/\mu$, then formula (10) can be written as:

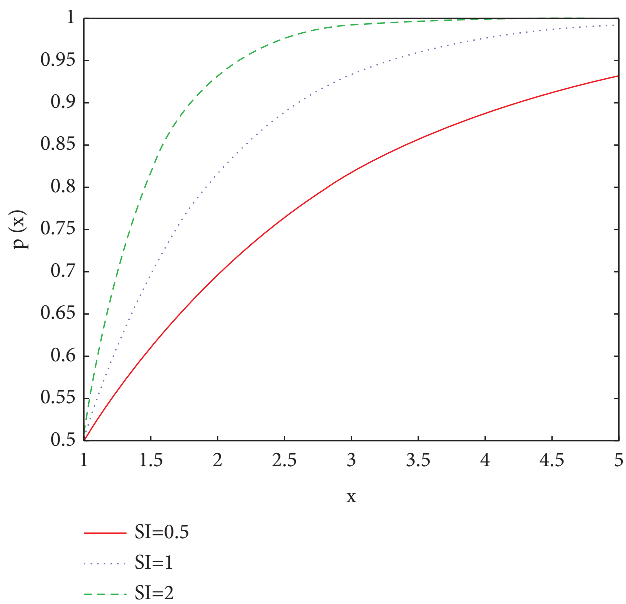
$$R = 1-e^{-\tau R}. \tag{11}$$

For $\tau = p(x)+\alpha+\mu/\mu$, $p(x)>0, \alpha>0$, then $\tau>1$. Let $f(R)=R-1+e^{-\tau R}$ and taking its derivative, we can get $f'(R)=1-\tau e^{-\tau R}$, $f''(R)=\tau^2e^{-\tau R}$. Since $f(0)=0$, $f(1)=e^{-\tau}>0$, and $f'(0)=1-\tau<0$, $f''(R)$ is always greater than 0, there are two solutions of 0 and R_s in formula (11), and $0<R_s<1$. This shows that for the parameters *p(x)*, α, and μ, no matter what the value is, when the network reaches a steady state, there are two possible situations for the proportion of rational people, that is, there is no diffusion threshold.

The above analysis shows that the proportion of rational people R in the final steady state is a function of the risk diffusion rate $p(x)$, the risk dormancy rate α, and the risk extinction rate μ. Next, we use simulation research methods to analyze the relationship between the proportion of rational people R and the risk diffusion rate $p(x)$, the risk dormancy rate α, and the risk extinction rate μ in the final steady state.

4. Simulation Analysis of Social Stability Risk Diffusion of Large Complex Engineering Projects on Complex Networks

4.1. Simulation Rules on Complex Networks. On the basis of constructing the IMSR model for the social stability risk diffusion of large complex engineering projects, for the purpose of more clearly analyzing how the subjects vary in the spread of social stability risks and explaining the process of social stability risk diffusion, further simulation research is needed to have a more vivid analysis of its diffusion process. As it is easily affected by the complexity of large complex engineering projects and inner dynamic changes of diffusion networks, the diffusion process will be accompanied by the mutual impacts and constant accumulation of risks, despite certain regular patterns, which further complicates the process of social stability risk diffusion of large complex engineering projects. The network of social stability risk diffusion of large complex engineering projects is essentially a complex network based on the multi-subject interaction on real social networks, and it is affected by external systems like the socioeconomic environment. It has been verified by a large number of scholars that the real social networks have small-world characteristics, Watt and Strogatz [44] studied social network, thinking its topology structure lied somewhere between completely regular and completely random, and characterized by small average length and large clustering coefficients, which belongs to a typical small-world network. By using "Tomocom" data, Tomochi et al. [52] analyzed the structure of social network and the result showed the social network belonged to the small-world network. Lu et al. [53] studied a community



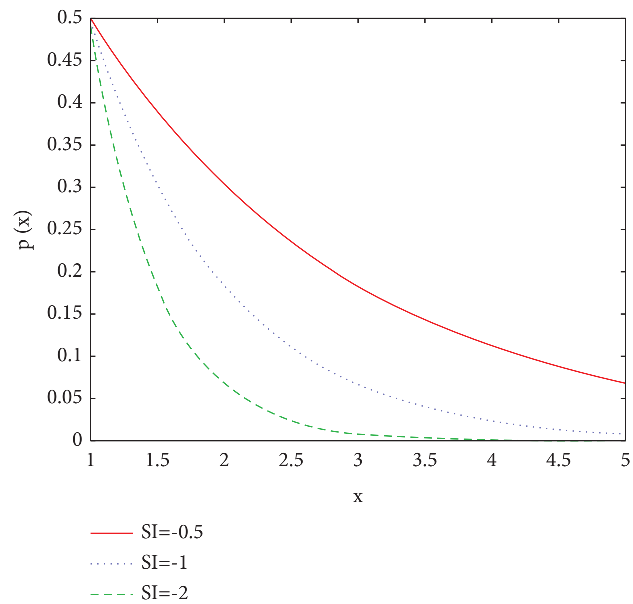


Figure 6: Changing graph of risk diffusion probability $p(x)$ in different situations: (a) there is a social amplification effect of risks; (b) there is a social restraint effect of risks.

model for social networks based on social mobility and analyzed the degree distributions, clustering coefficients, average distances, and diameters of networks. Experimental results demonstrated that the proposed model possessed the small-world property and could reproduce social networks effectively and efficiently. What is more, scholars also verified that social networks such as Facebook and Weibo were also featured with the small-world network, Su and Yen [54] analyzed 226 Facebook users, verifying that Facebook had obvious small-world characteristics. Dong et al. [55] studied the information dissemination on Sina Weibo during natural disasters, believing that the topology structure of Sina Weibo interactive network had the attribute of small world. In view of the fact that small-world networks can provide help to explain the problems related to complex socioeconomic systems and that the actual communication networks are similar to small-world networks, the type of complex network structure constructed in this paper is the small-world network. In the simulation study of social stability risk diffusion of large complex engineering projects on the small-world networks, this paper uses the NetLogo simulation platform, on which the simulation rules of IMSR model are as follows:

(1) Generate a WS small-world network with a certain number of nodes and an average degree of $\langle k \rangle$. The number of nodes indicates the number of subjects on the network of social stability risk diffusion of the entire large complex engineering projects, and the average degree indicates that each subject on the diffusion network is directly related to individuals with the number of $\langle k \rangle$ on average. The types of nodes on the diffusion network are divided into four categories, namely, the ignorant I, malcontents M, staggerers S, and the rational R, and the colors of the nodes correspond to yellow, red, green, and gray, respectively.

(2) In the initial state, it is assumed that there is only one malcontent, who is extremely dissatisfied with the current state, and actively spread the social stability risks in various ways, and we set the node to be red. All other nodes are the ignorant, indicating that they have not yet got the relevant social stability risk information, and we set them to be yellow.

(3) The malcontents transmit information about social stability risks to neighboring nodes of the ignorant; some subjects will believe in the risk information and participate in the spread of social stability risks, and they will become malcontents with the probability $p(x)$ (the color of the nodes turns red). Some subjects cannot judge the risk information well and are temporarily in a wait-and-see state, and they will become the malcontents with the probability α (the color of the nodes turns green). Other subjects who do not believe in the risk information at all and refuse to spread the social stability risks will turn into the rational with probability β (the color of the nodes turns gray).

(4) As time goes by, some staggerers who are exposed to more risk information or begin to believe in risk information for other factors will join in the spread of the social stability risks, and then spontaneously transform into the malcontents with probability θ (the color of the nodes turns red).

(5) When contacting the malcontents, some of the staggerers will be further affected and turn into the malcontents with probability $p(x)$ (the color of the nodes turns red).

(6) Over time, due to factors such as excessive exposure to risk information and under persuasion, some malcontents will temporarily take a wait-and-see attitude and will not participate in the spread of social stability risks, that is, spontaneously transform into the staggerers with probability σ (the color of the nodes turns green).

(7) When surrounded by staggerers, the rational, or other malcontents, some malcontents will be intervened by outside factors and will gradually turn into the ideal, namely, they will become the rational with probability μ (the color of the nodes turn gray).

(8) When there are not any malcontents on the entire network of social stability risk diffusion (no nodes with red color exist), the process of social stability risk diffusion is over.

4.2. Basic Variable Setting of NetLogo Simulation Platform. According to the simulation rules above, it primarily generates the WS small-world network on NetLogo platform and the initial network status will be set, as shown in Figure 7.

In Figure 7, the relevant initial parameters of the model are on the left. "Num-nodes" represent the total number of nodes on the network, namely, the total number of subjects on the diffusion network of the social stability risks of large complex engineering projects. "Rewiring-probability" represents the random reconnection probability p of generating the WS small-world network. "Initial-malcontent" represents the number of malcontents on the entire diffusion network in the initial state, and SI represents the social amplification degree of risks on the diffusion network (in the simulation, we assume that the risks are amplified in the society, setting $\mathrm{SI} > 0$), then the risk diffusion rate $p(x)$ can be obtained from formula (3). "Dormancy-probability" represents risk dormancy rate α, "rejection-probability" represents risk rejection rate β, "recovery-probability" represents risk recovery rate θ, "hibernation-probability" represents risk hibernation rate σ, "extrusion-probability" represents risk extinction rate μ, "clustering-coefficient" represents aggregation coefficient of the generated WS small-world network, and "average-path-length" represents the average distance of the generated WS small-world network.

This paper conducts an investigation on a major water conservancy project in Guangdong Province, and conducts interviews and investigations on the government, project developers, experts, residents, and other related subjects involved in the project to obtain the initial values of the relevant parameters. Due to the sensitivity of social stability risks, this paper has carried out "desensitization" treatment, and the names of large complex engineering projects have been given, but only briefly introduced. The research object is a comprehensive major water conservancy project that focuses on flood control and water supply, and takes into account power generation and shipping. Together with upstream and downstream dikes, it forms a flood control system called "combined with dykes and reservoirs." The

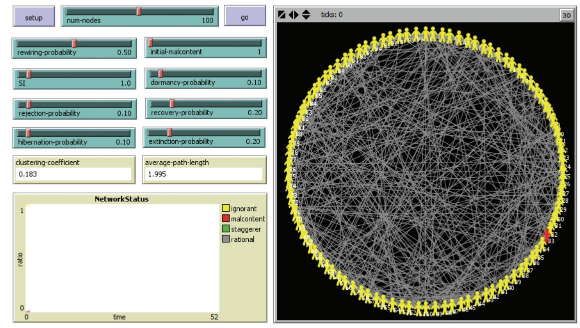


Figure 7: The initial status of IMSR model simulation on WS small-world network.

major water conservancy project plays an important role in ensuring water safety and promoting economic development; thus, it has received extensive attention from the local government and people.

In the initial state, it assumes that the reconnection probability p on the small-world network is 0.5, the number of subjects on the entire network is 100, with the initial number of malcontents being 1, and the rest are all the ignorant. It also assumes that the social amplification degree of risks SI is 1, the risk dormancy rate is 0.1, the risk rejection rate is 0.1, the risk recovery rate is 0.2, the risk hibernation rate is 0.1, and the risk extinction rate is 0.2. The model is run under this state, generating the simulation results as shown in the lower left and right of Figure 7. At the bottom left of Figure 8, it can be seen from the changing trend of subjects in different states that, at the beginning, the number of malcontents on the diffusion network increases rapidly, while the number of ignorant decreases sharply. There are the ignorant, malcontents, staggerers, and the rational on the entire diffusion network. With the spread of social stability risks on the network, the number of malcontents gradually decreases after reaching the peak, and the number of rational people gradually increases, until the malcontents spreading social stability risk factors on the entire network disappear, with only ignorant and rational people left to reach a stable equilibrium state. At this time, the number of rational people on the entire network is 88.

From the theoretical derivation of the previous model, it can be seen that the network equilibrium state is related to the risk diffusion rate, the risk dormancy rate, and the risk extinction rate, among which the risk diffusion rate is determined by the social amplification degree of risks SI. Therefore, the parameters SI, dormancy-probability, and extinction-probability are adjusted to simulate and analyze the influence of changes in risk diffusion rate, risk dormancy rate, and risk extinction rate in the final steady state.

4.3. Simulation Results and Analysis

4.3.1. The Relation between Final Steady State and Risk Diffusion Rate. It can be seen from formula (3) that the risk diffusion rate $p(x)$ is determined by the social amplification degree of risks SI on the diffusion network, and the number of times that the ignorant contact the malcontents x. On the

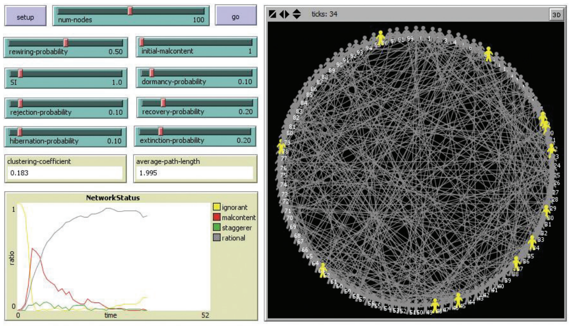


Figure 8: The simulation results of initial parameter values of SWN-IMSR model.

simulation model, x can be calculated by the number of times that the malcontents contact the ignorant, so the influence of SI variation in the final steady state is mainly considered. From formula (3), it can be further seen that when SI increases, $p(x)$ also increases. When the values of SI are 2, 3, and 4, the simulation results are shown in Figure 9.

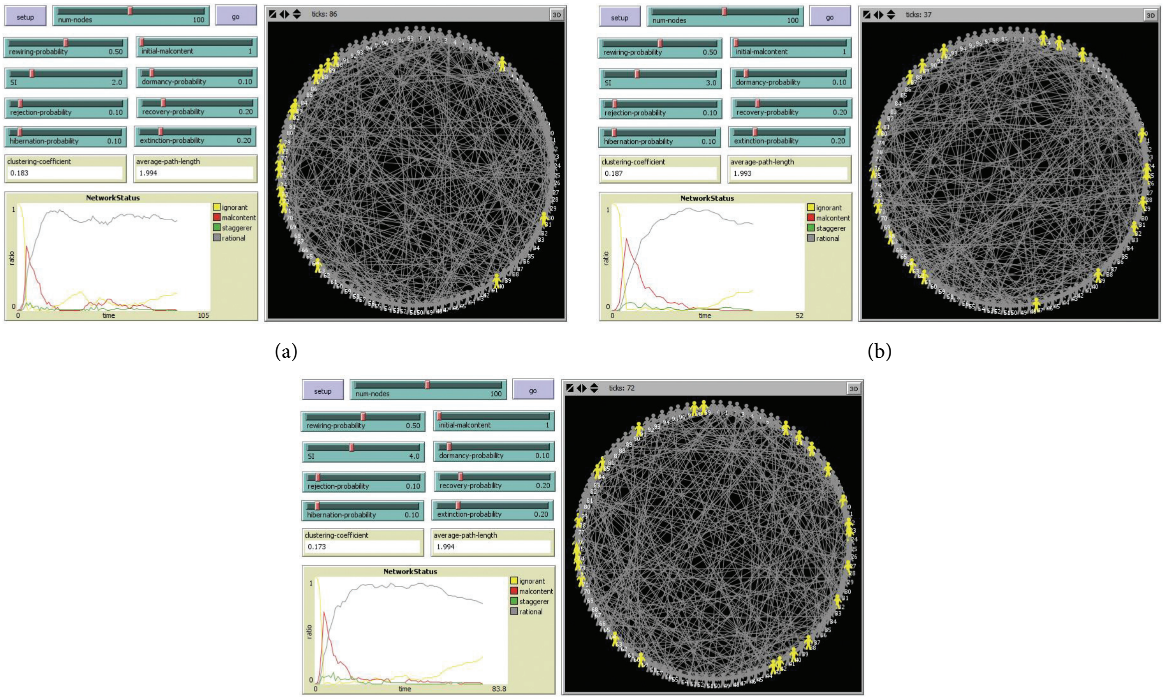
As shown in Figure 9, after a period of spreading of the social stability risks on the diffusion network, it finally reaches a steady state, ending up with only the ignorant and the rational, that is to say, the diffusion process of social stability risks is over. In the final steady state, the final number of rational people are 84 (when $SI = 2$), 81 (when $SI = 3$), and 76 (when $SI = 4$). When the social amplification degree of risks SI increases, the risk diffusion rate $p(x)$ increases; finally, the proportion of rational people on the diffusion network gradually decreases, thus the spread of social stability risks of large complex engineering projects becomes more violent. From the simulation results, it can be seen that in order to effectively control the spread of social stability risks of large complex engineering projects, the value of the social amplification degree of risk SI should be reduced. In the process of social stability risk diffusion of large complex engineering projects, the reduction of information asymmetry can significantly affect the social amplification degree of risk SI. When the information asymmetry decreases, the government's credibility can be improved, the subjects have relatively more trust in the government, and the social amplification degree of risk SI will decrease. Therefore, as the main owner of information resources, the government has the responsibility and obligation to perform the information disclosure function when the social stability risks of large complex engineering projects break out, build a sound information disclosure mechanism, and protect the public's right to know, so as to reduce the risk of the social amplify degree to control the spread of social stability risks.

4.3.2. The Relation between the Final Steady State and Risk Dormancy Rate. It can be seen from the previous analysis that, in the steady state, the proportion of rational people R is inversely proportional to the risk hibernation rate α on the IMSR model. As the risk dormancy rate α increases, the proportion R decreases. When the value of the risk dormancy rate in the initial state is 0.1, the number of rational people in the final steady state is 88. When the values of risk dormancy rate α are 0.2, 0.3, and 0.4 (namely, the dormancy-probability values in the simulation interface are 0.2, 0.3, and 0.4), respectively, the simulation results are as shown in Figure 10.

As shown in Figure 10, after a period of spreading of the social stability risks on the diffusion network, it finally reaches a steady state, ending up with only the ignorant and the rational, that is to say, the diffusion process of social stability risks is over. In the final steady state, the final numbers of rational people are 85 (when $\alpha = 0.2$), 82 (when $\alpha = 0.3$), and 75 (when $\alpha = 0.4$). In the final steady state, when the risk dormancy rate α increases, the proportion of rational people in the entire diffusion network gradually decreases, leading to more violent risk diffusion as well as serious social stability risks of large complex engineering projects. From the simulation results, it can be seen that in order to effectively control the spread of social stability risks of large complex engineering projects, the value of the risk dormancy rate α should be reduced. Similar to the social amplification degree SI of risk, in the process of social stability risk diffusion of large complex engineering projects, the reduction of information asymmetry can effectively reduce the risk dormancy rate α, thereby effectively controlling the social stability risk diffusion of large complex engineering projects.

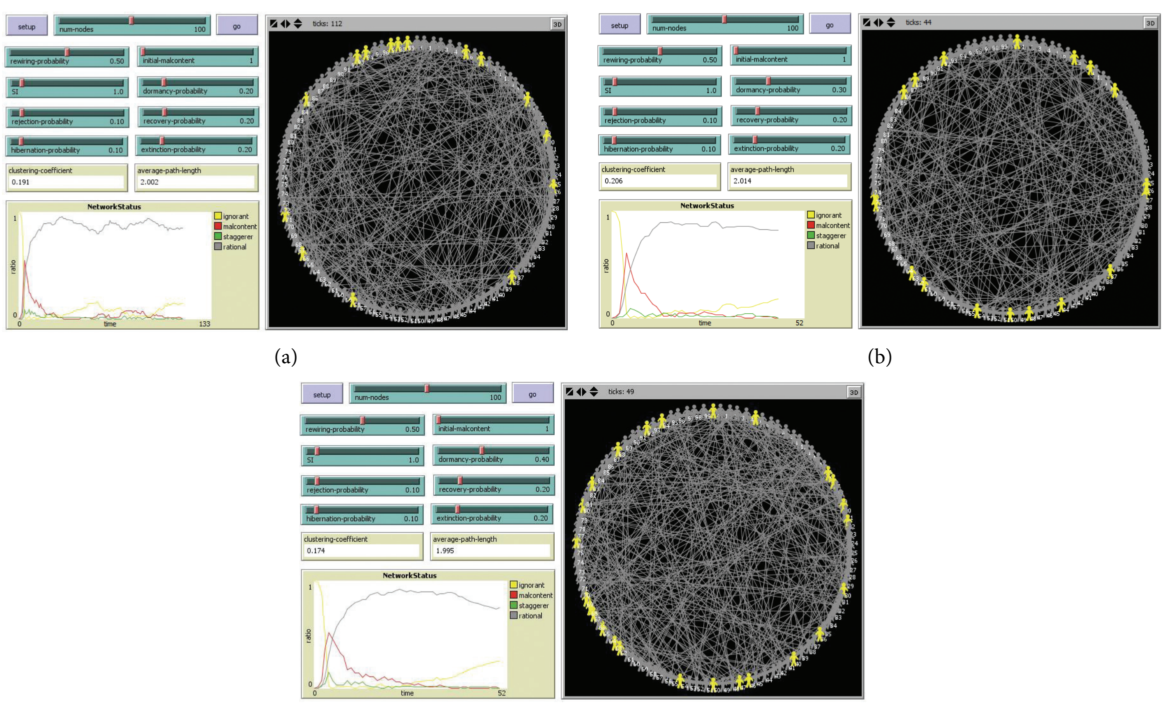
4.3.3. The Relation between the Final Steady State and Risk Extinction Rate. It can be seen from the previous analysis that, in the steady state, the proportion of rational people R is proportional to the risk extinction rate μ on the IMSR model. When the risk extinction rate μ increases, the proportion R increases. When the value of the risk extinction rate μ in the initial state is 0.2, the number of rational people in the final steady state is 88. When the values of risk extinction rate μ are 0.3, 0.4, and 0.5 (namely, the extinction-probability values in the simulation interface are 0.3, 0.4, and 0.5) respectively, the simulation results are as shown in Figure 11.

As shown in Figure 11, after a period of spreading of the social stability risks on the diffusion network, it finally reaches a steady state, ending up with only the ignorant and the rational, that is to say, the diffusion process of social stability risks is over. In the final steady state, the final numbers of rational people are 89 (when $\mu = 0.3$), 93 (when $\mu = 0.4$), and 97 (when $\mu = 0.5$), respectively. In the final steady state, when the risk extinction rate μ increases, the proportion of rational people on the entire diffusion network gradually increases, making it more difficult for risk diffusion on the network, and the social stability risks of large complex engineering projects can be alleviated. Through the simulation results, it can be seen that in order to effectively control the spread of social stability risks in large complex engineering projects, the value of the risk extinction rate μ can be increased. In the process of social stability risk diffusion of large complex engineering projects, the ability to deal with risk events can significantly affect the risk extinction rate μ. When the ability is strong, it can quickly act on relevant subjects (namely dissatisfied) that have been



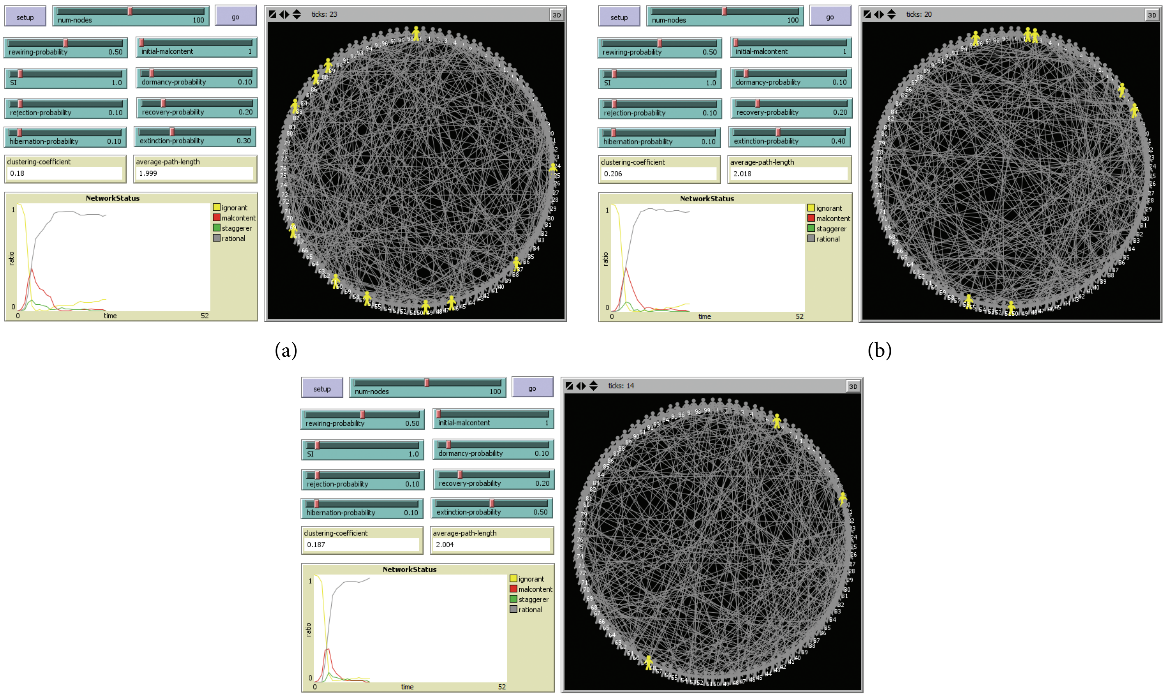
(a) (b) (c)

Figure 9: The simulation results of the IMSR model when the social amplification degree of risks SI takes different values: (a) the simulation result when SI = 2; (b) the simulation result when SI = 3; (c) the simulation result when SI = 4.



(a) (b) (c)

Figure 10: The simulation results of the IMSR model when the risk dormancy rate α takes different values: (a) the simulation result when $\alpha = 0.2$; (b) the simulation result when $\alpha = 0.3$; (c) the simulation result when $\alpha = 0.4$.



(a) (b) (c)

Figure 11: The simulation results of the IMSR model when the risk extinction rate μ takes different values: (a) the simulation result when $\mu = 0.3$; (b) the simulation result when $\mu = 0.4$; (c) the simulation result when $\mu = 0.5$.

affected by risk information, accelerate their transition to staggerers and the rational, and increase their risk extinction rate. Therefore, the government should build a sound social stability risk response mechanism, such as strengthening regional social security and economic compensation at the macro level, and strengthening risk rumors control and psychological counseling at the micro level, so as to control social stability diffusion by increasing the risk extinction rate.

5. Conclusion

Based on the analysis of subjects, interactive model, and infectious disease characteristics of the social stability risk diffusion of large complex engineering projects, this paper discusses the adaptability of the infectious disease model. The traditional SIR model is improved from the aspects of diffusion subjects and diffusion rules, and by regarding the small-world network as the complex network topology, the diffusion process of social stability risks of large complex engineering projects is studied through the NetLogo simulation platform. The results are shown as below:

First of all, we found that the changes in risk diffusion rate $p(x)$, risk dormancy rate α, and risk extinction rate μ can affect the final steady state of social stability risk diffusion of large complex engineering projects, which are basically consistent with the theoretical analysis results. Under the impact of malcontents, the risk diffusion rate $p(x)$ referring to the ignorant participates in the spread of social stability risks and is directly affected by SI, namely, the social amplification degree of risks. The larger the SI is, the greater the risk diffusion rate $p(x)$ is, and finally the proportion of rational people is lower when the state becomes steady. Similarly, the greater the risk dormancy rate α is, the lower the proportion of rational people in the final steady state is. Furthermore, for the purpose of effectively intervening in the spread of social stability risks of the large complex engineering projects, effective measures should be taken to reduce the social amplification degree SI and risk dormancy rate α of risks, which exist on the network of social stability risk diffusion of large complex engineering projects. Then, it affects the diffusion process to effectively defuse the social stability risks.

What is more, the malcontents no longer participate in the spread of social stability risks since they are affected by other subjects or their interest demands are met; thus, the increase in the risk extinction rate μ will increase the proportion of rational people in the final steady state. Therefore, at the same time of reducing social amplification degree SI and the risk dormancy rate α, the risk extinction rate μ on the network of social stability risk diffusion of large complex engineering projects should also be increased as much as possible, aiming to better control the spread of social stability risks of the large complex engineering projects.

There are still three deficiencies in the study. Firstly, from the aspect of diffusion subjects and diffusion rules, the IMSR model constructed in this paper is improved based on the traditional SIR model, but the division of subjects and the analysis of diffusion rules still remain at the theoretical level; whether new subjects and more complex diffusion paths can be added is worthy of further exploration. Secondly, although the small-world network is indeed the network structure most in line with the real social network, this paper only considers the small-world network as a network structure for simulation research, and does not choose other network structures such as scale-free networks for comparative analysis. Thirdly, real data are not used in the simulation study of social stability risk diffusion of large complex engineering projects on the small-world network; the relevant parameters such as the social amplification degree of risks SI, the risk diffusion rate $p(x)$, the risk dormancy rate α, and the risk extinction rate μ are assigned with hypothetical values, which still deviate from the actual situation, although these assignments are determined based on a large number of literatures and interviews with relevant experts. The further research of this paper should focus on the following three aspects. Firstly, further combine the construction practice of large complex engineering projects in China, analyze more types of subjects participating in the spread of social stability risks, consider the diffusion paths among more subjects, and improve the IMSR model. Secondly, improve the model by constructing other network structure types, such as scale-free networks, and conduct further comparative analysis to study the process of social stability risk diffusion under different network structure types. Thirdly, increase the relevant case collection and social survey, and try to use real data from some typical cases, including social networks such as Facebook and Sina Weibo, enabling the selection of relevant parameters in simulation research to be more scientific and reasonable. At the same time, it is possible to propose more appropriate intervention strategies for the spread of social stability risks of the large complex engineering projects based on more accurate parameter selections.

Data Availability

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

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 71573072), the China Postdoctoral Science Foundation (no. 2019M661719), and the Fundamental Research Funds for the Central Universities (no. B200207010).

References

[1] B. Tilt, Y. Braun, and D. He, “Social impacts of large dam projects: a comparison of international case studies and implications for best practice,” *Journal of Environmental Management*, vol. 90, pp. S249–S257, 2009.

[2] X. Xu, Z. Yang, L. Jin, and L. Hao, “Research on risk assessment of port project investment and construction,” *Journal of Coastal Research*, vol. 115, no. sp1, pp. 187–189, 2020.

[3] X. Wei, L. Zhou, G Yang, Y. Wang, and Y. Chen, “Assessing the effects of desertification control projects from the farmers’ perspective: a case study of Yanchi county, northern China,” *International Journal of Environmental Research and Public Health*, vol. 17, no. 22, p. 15, 2020.

[4] C. Zhang, X. Nong, D. Shao, H. Zhong, Y. Shang, and J. Liang, “Multivariate water environmental risk analysis in long-distance water supply project: a case study in China,” *Ecological Indicators*, vol. 125, pp. 70–82, 2021.

[5] C. Folke, “Resilience: the emergence of a perspective for social-ecological systems analyses,” *Global Environmental Change*, vol. 16, no. 3, pp. 253–267, 2006.

[6] J. Miao, D. Huang, and Z. He, “Social risk assessment and management for major construction projects in China based on fuzzy integrated analysis,” *Complexity*, vol. 2019, Article ID 2452895, 17 pages, 2019.

[7] Z.-z. Liu, Z.-w. Zhu, H.-j. Wang, and J. Huang, “Handling social risks in government-driven mega project: an empirical case study from West China,” *International Journal of Project Management*, vol. 34, no. 2, pp. 202–218, 2016.

[8] Z. He, D. Huang, J. Fang, and B. Wang, “Stakeholder conflict amplification of large-scale engineering projects in China: an evolutionary game model on complex networks,” *Complexity*, vol. 2020, Article ID 9243427, 17 pages, 2020.

[9] Z. He, D. Huang, C. Zhang, and J. Fang, “Toward a stakeholder perspective on social stability risk of large hydraulic engineering projects in China: a social network analysis,” *Sustainability*, vol. 10, no. 4, p. 15, 2018.

[10] S. Peng, G. Shi, and R. Zhang, “Social stability risk assessment: status, trends and prospects -a case of land acquisition and resettlement in the hydropower sector,” *Impact Assessment and Project Appraisal*, vol. 39, pp. 1–17, 2019.

[11] H. An, W. Yang, J. Huang, A. Huang, Z. Wan, and M. An, “Identify and assess hydropower project’s multidimensional social impacts with rough set and projection pursuit model,” *Complexity*, vol. 2020, Article ID 9394639, 16 pages, 2020.

[12] Q. Wu, L. Wang, N. Jin, N. Lin, and X. Hu, “Social stability risk evaluation of major water conservancy projects in fragile eco-environment regions,” *Applied Ecology and Environmental Research*, vol. 17, no. 4, pp. 9097–9111, 2019.

[13] L. Ma and X. Cong, “Social stability risk assessment of NIMBY major projects by OWA, matter-element, and cloud model,” *Journal of Intelligent and Fuzzy Systems*, vol. 36, no. 3, pp. 2545–2556, 2019.

[14] L. Wang, L. Jin, Q. Wu, F. Yang, and J. Sun, “Risk assessment of social stability caused by large hydro-juncture project,” *Journal of Yangtze River Scientific Research Institute*, vol. 37, no. 4, pp. 152–159, 2020, in Chinese.

[15] C. Li, Y. Liu, and S. Li, “Risk evaluation of qinghai-tibet power grid interconnection project for sustainability,” *Sustainability*, vol. 8, no. 1, p. 19, 2016.

[16] R. Zhu, F. Zhao, and H. Wang, “Study on social stability risk analysis for nuclear power projects,” in *Proceedings of the 25th International Conference on Nuclear Engineering*, vol. 4, Shanghai, China, July 2017.

[17] Q. Li, “Identification and governance of social stability risk in megacities: based on the relocation event of “Z city village” in shanghai,” *Journal of Tongji University*, vol. 29, no. 6, pp. 89–96, 2018, in Chinese.

[18] H. Wang, “Single factor risk analysis of social stability risk in engineering project-risk in land and house expropriation,” in *Proceedings of the 2nd International Conference on Advanced*

Education and Management Science (AEMS), Shanghai, China, December 2018.

[19] C. Li and Z. Xi, "Social stability risk assessment of land expropriation: lessons from the Chinese case," *International Journal of Environmental Research and Public Health*, vol. 16, no. 20, p. 20, 2019.

[20] C. Jiang, E. Chai, and H. Wang, "Indicator system and assessment of social stability risk in road construction," in *Proceedings of the 4th International Conference on Management Science and Management Innovation (MSMI)*, vol. 31, pp. 207–210, Chengdu, China, June 2017.

[21] Z. Dong, H. Wang, X. Dong, and S. Liu, "Evaluation method for highway project social stability risk based on entropy weight extension matter-element model," *China Journal of Highway and Transport*, vol. 31, no. 9, pp. 191–198, 2018, in Chinese.

[22] G. Zhang, "Study on the causes of social stability risk of environmental engineering project-based on qualitative comparative analysis method," in *Proceedings of the International Conference on Management Science and Engineering-Annual Conference Proceedings*, pp. 684–690, Nomi, Japan, August 2017.

[23] P. Gu and H. Cong, "Research on collaborative governance of engineering social stability risk: taking JiuFeng waste incineration power generation project as an example," *Journal of Dialectics of Nature*, vol. 42, no. 1, pp. 108–114, 2020, in Chinese.

[24] S. Tan, "The formation and prevention to the social risk of NIMBY project from anxiety psychological perspective," *Journal of Beijing University of Aeronautics and Astronautics (Social Sciences Edition)*, vol. 26, no. 3, pp. 25–30, 2013, in Chinese.

[25] D. Huang, Z. He, and C. Zhang, "A study of system dynamics simulation for the social stability risk diffusion of large engineering," *Journal of Hohai University (Philosophy and Social Sciences)*, vol. 21, no. 3, pp. 60–67, 2019, in Chinese.

[26] B. Sartorius, A. Lawson, and R. Pullan, "Modelling and predicting the spatio-temporal spread of COVID-19, associated deaths and impact of key risk factors in England," *Scientific Reports*, vol. 11, no. 1, p. 11, 2021.

[27] S. Jiao, M. Huang, and M. Huang, "An SIHR epidemic model of the COVID-19 with general population-size dependent contact rate," *Aims Mathematics*, vol. 5, no. 6, pp. 6714–6725, 2020.

[28] T. Chen, B. Ma, and J. Wang, "SIRS contagion model of food safety risk," *Journal of Food Safety*, vol. 38, no. 1, Article ID e12410, 2018.

[29] Q. Qian, Y. Yang, J. Gu, and H. Feng, "Information authenticity, spreading willingness and credit risk contagion - a dual-layer network perspective," *Physica A: Statistical Mechanics and its Applications*, vol. 536, Article ID 122519, 2019.

[30] J. Li, Y. Hu, and Z. Jin, "Rumor spreading of an SIHR model in heterogeneous networks based on probability generating function," *Complexity*, vol. 2019, Article ID 4268393, 15 pages, 2019.

[31] L. Qiu, W. Jia, W. Niu, M. Zhang, and S. Liu, "SIR-IM: SIR rumor spreading model with influence mechanism in social networks," *Soft Computing*, vol. 25, no. 22, pp. 13949–13958, 2020.

[32] H. Sun, Y. Sheng, and Q. Cui, "An uncertain SIR rumor spreading model," *Advances in Difference Equations*, vol. 2021, no. 1, p. 22, 2021.

[33] P. Erdos and A. Renyi, "On the evolution of random graphs," *Publications of the Mathematical Institute of the Hungarian Academic Sciences*, vol. 5, pp. 17–61, 1960.

[34] X. Qian, J. Yu, and R. Dai, "A new scientific field: open complex giant system and its methodology," *Chinese Journal of Nature*, vol. 1, pp. 3–10, 1990.

[35] R. Pastor-Satorra and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach*, Cambridge University Press, Cambridge, UK, 2007.

[36] J. Feng, X. Li, B. Mao, Q. Xu, and Y. Bai, "Weighted complex network analysis of the Beijing subway system: train and passenger flows," *Physica A: Statistical Mechanics and Its Applications*, vol. 474, pp. 213–223, 2017.

[37] D. Petrone and V. Latora, "A dynamic approach merging network theory and credit risk techniques to assess systemic risk in financial networks," *Scientific Reports*, vol. 8, no. 1, p. 5561, 2018.

[38] T. Petermann and P. De Los Rios, "Role of clustering and gridlike ordering in epidemic spreading," *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 69, no. 6, Article ID 066116, 2004.

[39] J. Zhang and J. Sun, "Stability analysis of an SIS epidemic model with feedback mechanism on networks," *Physica A: Statistical Mechanics and its Applications*, vol. 394, pp. 24–32, 2014.

[40] C. Ellinas, N. Allan, C. Durugbo, and A. Johansson, "How robust is your project? From local failures to global catastrophes: a complex networks approach to project systemic risk," *PLoS One*, vol. 10, no. 11, Article ID e0142469, 2015.

[41] Y. Li, Z.-p. Du, and L. Zhang, "Agri-Food supply chain network risk propagation research based on complex network," in *Proceedings of the 6th International Asia Conference on Industrial Engineering and Management Innovation.*, pp. 985–994, Atlantis Press, Paris, May 2016.

[42] A. Zhu, W. Chen, J. Zhang, X. Zong, W. Zhao, and Y. Xie, "Investor immunization to Ponzi scheme diffusion in social networks and financial risk analysis," *International Journal of Modern Physics B*, vol. 33, no. 11, Article ID 1950104, 2019.

[43] Z. Su and F. Xu, "Dynamic identification of systemically important financial markets in the spread of contagion: a ripple network based collective spillover effect approach," *Journal of Multinational Financial Management*, vol. 60, Article ID 100681, 2021.

[44] D. J. Watts and S. H. Strogatz, "Collective dynamics of "small-world" networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[45] A. L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science (New York, N.Y.)*, vol. 286, no. 5439, pp. 509–512, 1999.

[46] L. He and S. Li, "Network entropy and systemic risk in dynamic banking systems," *Complexity*, vol. 2017, Article ID 1852897, 7 pages, 2017.

[47] Y. Chen, T. Shu, S. Chen, S. Wang, K. K. Lai, and L. Gan, "Strong-weak collaborative management in coping supply chain disruption risk transmission based on scale-free networks," *Applied Economics*, vol. 49, no. 39, pp. 3943–3958, 2017.

[48] K. Weeden and B. Cornwell, "The small-world network of college classes: implications for epidemic spread on a university campus," *Sociological Science*, vol. 7, pp. 222–241, 2020.

[49] S. Wu, M. Tong, Z. Yang, and T. Zhang, "Interconnectedness, systemic risk, and the influencing factors: some evidence from

China's financial institutions," *Physica A: Statistical Mechanics and Its Applications*, vol. 569, Article ID 125765, 2021.

[50] L. Zhao, J. Wang, Y. Chen, Q. Wang, J. Cheng, and H. Cui, "SIHR rumor spreading model in social networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 391, no. 7, pp. 2444–2453, 2012.

[51] H. Wang, L. Deng, F. Xie, H. Xu, and J. Han, "A new rumor propagation model on SNS structure," in *Proceedings of the 2012 IEEE International Conference onGranular Computing (GrC)*, pp. 499–503, IEEE, Hangzhou, China, June 2012.

[52] M. Tomochi, A. Tanaka, and T. Shichijo, "Stratification and nested structure of small world in a friendship network-data analysis, modeling, and simulation on s social networking service a.k.a Tomocom," *Sociological Theory and Methods*, vol. 26, no. 1, pp. 83–97, 2011.

[53] Z. Lu, Z. Wu, H. Luo, and H. Wang, "Improved community model for social networks based on social mobility," *International Journal of Modern Physics C*, vol. 26, no. 2, Article ID 1550012, 2015.

[54] B. C. Su and T. S. Yen, "Small-world phenomenon and strategies for making friends on social networking sites in mobile environment: random and non-random," *International Journal of Mobile Communications*, vol. 15, no. 4, pp. 355–371, 2017.

[55] R. Dong, L. Li, Q. Zhang, and G. Cai, "Information diffusion on social media during natural disasters," *IEEE Transactions on Computational Social Systems*, vol. 5, no. 1, pp. 265–276, 2018.

Hindawi
Complexity
Volume 2021, Article ID 3191609, 12 pages
https://doi.org/10.1155/2021/3191609

Research Article

Public Transport in Rural Roads: Measures to Increase Its Modal Share in Iran

Alireza Afkham,[1] **Shahriar Afandizadeh**,[2] **and Ali Naderan**[1]

[1]*Department of Civil Engineering Sciences and Research Branch, Islamic Azad University, Tehran, Iran*
[2]*Department of Transportation School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran*

Correspondence should be addressed to Shahriar Afandizadeh; zargari@iust.ac.ir

Received 7 November 2021; Accepted 8 December 2021; Published 24 December 2021

Academic Editor: Qingyuan Zhu

The supply and demand management of rural public transport has gained an important place in Iran, given its decreasing trend over the last decade. Accordingly, an effective approach is to analyze stakeholders' opinions in this field to identify the effective local solutions for increasing the share of this transport method. The used methodology was to design a mixed questionnaire, a part of which included fuzzy pairwise comparison, while, in its first layer, the variables were classified using rotary analysis. In the other two sections of the questionnaire, the respondents were asked to express their opinions regarding the questions in the form of qualitative and words cloud. According to the inquiries made from the users in 20 selected terminals of Iran providing services to more than 85% of rural passengers, the indicators were weighted and divided into two groups based on the rotary analysis. The weighted results obtained from the users' opinions revealed that the safety, dynamics of incentive policies, and traffic system performance in the rural transport had the highest effects on the micro- and macro-level indicators. In the second section, based on the qualitative questions, a multivariate linear estimation model of the number of rural passengers was constructed. Moreover, in the third section, the users' suggested keywords focused on policy-making, travel time optimization, quality of services, and safety. Both the second and third sections had an acceptable agreement with the pairwise comparisons. Given the vast area of Iran and the distance between the population centers in the country, the obtained solutions to increase the share of public rural road transport included reducing the desirability of travel with private cars in short rural distances through interaction with industrial towns around metropolises, along with providing such areas with special services to attract passengers to rural transport with occupational goals.

1. Introduction

This study aims to seek solutions for mitigating the declining share of rural public transport in I.R. Iran since 2008 and to suggest measures to attract passengers to this mode of transport. According to the latest data published by the Iranian Roads and Transportation Organization for 2019, the average number of passengers was 18 per bus, 14 per minibus, and 4 per passenger car. Table 1 presents the number of intrastate and interstate passengers in terms of vehicle type from 2008 to 2019.

According to Table 1, the following remarks can be presented:

Road passengers are more likely to choose a minibus for intrastate trips, while interstate passengers are more likely to choose a bus. In recent years, the popularity of buses and minibuses has declined, while passenger cars have maintained their popularity.

From 2008 to 2019, using buses and minibuses for interstate and intrastate trips has decreased on a yearly basis. It seems that appropriate policies had been adopted for road trips by 2008; however, the market share of this mode has been later absorbed by other modes.

The number of road passengers is decreasing, and it can be seen that, in recent years, the attractiveness of road transport has decreased.

Finally, the question arises as how to increase the utility of road public transport systems. It is noteworthy that the decrease in the total number of public transport passengers

Table 1: Number of public Rural Road Passengers in Terms of Vehicle Type from 2008 to 2019 (Statistical yearbooks of road transport from 2008 to 2019) [1].

Year	Inter-state passenger (thousand)			Intra-state passenger (thousand)			Total (thousand)		
	Bus	Minibus	Passenger car	Bus	Minibus	Passenger car	Bus	Minibus	Passenger car
2008	74075	12004	7107	43901	67923	18945	117976	79927	26052
2009	80981	12757	7489	52324	73755	20305	133305	86512	27794
2010	84252	13670	8593	56290	76455	24382	140542	90125	32975
2011	84268	13366	9568	57675	75020	27057	141943	88386	36625
2012	79006	11211	8962	51593	68171	26026	130599	79382	34988
2013	76725	10640	8488	49026	65139	25770	125751	75779	34258
2014	77239	10136	9226	43600	59173	26176	120839	69309	35402
2015	75772	8588	9120	39209	52259	26330	114981	60847	35450
2016	67547	6790	8769	32599	47227	27207	100146	54017	35976
2017	64240	6334	8507	30137	45407	24323	94377	51741	32830
2018	62923	5816	7724	29100	39684	21742	92023	45500	29466
2019	57514	6642	7776	25137	35354	21552	82651	41996	29328

may occur for two reasons: first, a reduction in the number of intercity passengers due to reduced financial capacity and, consequently, a reduction in the general willingness to travel; secondly, a shift in the tendency of passengers from using public toward private modes of transport. The development of remedial strategies to increase the modal share of public transport requires identifying the main variables affecting the utility of the public transport system. This is to be carried out through extensive interviews, stakeholder surveys, information analysis, and the classification of variables using a rotary analysis and a fuzzy AHP approach. Passengers in all major rural bus terminals in the country (i.e., 20 terminals), accounting for 85% of total passengers, were randomly selected for the interviews.

The fuzzy pairwise comparison was applied due to its good adaptability with the data of the study and the way they were collected. The rotary analysis was also performed to complete the study using a method based on a programming language that readily makes complex calculations and can be suggested as an experience for future studies on rural road transport. It is worth noting that the users' opinions were collected, compared, and verified in three different sections, including rotary pairwise comparison, qualitative assessment, and word cloud.

The study was completed by fuzzy rotary classifying the opinions, words cloud analyzing the users' behavior, constructing a multivariate linear estimation model of the number of rural passengers and identifying an effective solution that can remarkably impact the share of rural public transport.

2. Literature Review

Analyzing the factors influencing the choice of transport and determining its utility leads to identifying the people who are attracted to each mode of transport and meeting the needs of each individual based on the utility of those transport modes. The choice of transport in intercity travels depends on many factors, and this diversity is due to the commercial differences between different modes of transportation, the heterogeneity of their passengers, and using them for modeling and evaluation.

When expressing their opinions, public transport users generally consider their immediate needs rather than the total utility of the transport system. Using the previous empirical criteria, Hoque et al. (2021) [2] indicated that traffic forecasters used their set of values in expressing their opinions. A goal of inquiries from users is to identify the issues of different parts of a transportation system. In this regard, Jafino (2021) [3] sought to identify parts of a network based on utilitarianism. In this framework, two main ideas were investigated: calculation of the transport demand with balanced weights and balance of users in the network. Jafino et al. (2020) [4] stated that the importance of using the opinions of transportation users to support decision-making is increasing, while the principle of balanced distribution resulting from ethical analyses is still applied minimally.

The way passengers value the waiting time and the access quality are among the crucial factors determining the desirability of a public transport system. Instead of conducting a face-to-face survey of passengers, Yap and Cats (2021) [5] investigated the behavior of passengers who denied boarding and their alternative decisions. The competitive evaluation of public transport operators from the users' perspective is of great importance to identify the supply improvement methods. Mo (2021) [6] assessed the competitive approach and a system's performance from four stakeholders' perspectives, including automatic vehicle operators, public transportation operators, passengers, and the transport authority. Introducing the required data and the processing stages, Hörl and Balac (2021) [7] studied a sample of repeatable travel demand for Paris and the suburbs to develop a general design. There are several contradictory approaches, such as the users' needs against the system's performance and goals (e.g., increasing the spatial coverage and increasing the frequency) for the redesign of public transport networks. Weckström (2021) [8] provided a multidimensional navigability approach from the users' perspective.

With the development of social media, we should investigate whether we can access reliable data through them to meet the information needs of public transport. According to studies, this approach has had good results in some routes, but on a large scale like within a country, comprehensive data cannot be accessed given the governmental restrictions. Sala et al. (2021) [9] questioned whether it was possible to extract reliable data on passenger travels from the content of social media (Twitter) to design bus routes in areas around cities. The information was tested for a great musical event in Barcelona, and an acceptable image of supply distribution was suggested. In a similar way, Yao (2021) [10] analyzed the information of twits regarding the sleep/wake status, local events, and planned traffic incidents of each day to describe the traffic of the morning in next day in 53 road segments of Pittsburgh.

Despite the unprecedented volume and variety of displacement data, origin-destination (OD) matrices are still the widest tools for organizing and expressing travel demand. Ballis and Dimitriou (2020) [11] stated that standard ODs could not fully demonstrate the factors affecting travel behavior, such as trip interdependency and trip chaining. They suggested considering home-based trip chains, and consequently, activity schedules in studying personal mobility. Khan et al. (2021) [12] reported increasing the attractiveness of public transport as a necessity for sustainable systems and mentioned a lack of empirical research to analyze how to do so. Santos and *Lima* (2021) [13] evaluated the public transport quality indicators and their performance concerning the users' opinions based on a multicriteria approach.

Fournier et al. (2021) [14] performed a survey of passengers regarding heterogeneous values, which resulted in a computational model for policy-making in determining the transportation price. Transportation based on managed demand can potentially raise the desirability of public travel. Kucharski and Cats (2020) [15] identified a combination of compatible trips, which remain attractive by replacing travel with private cars. They achieved a formulation that obtains optimum time value among prices, delays, and dissatisfactions. Developing and verifying a questionnaire, Deb et al. (2017) [16] analyzed the acceptance level of fully self-driving vehicles. The outcomes revealed that men and the youth residing in cities were more open to changes. Jokubauskaite et al. (2019) [17] investigated different trips and the weights allocated to them and found a framework to develop numerous trips.

Popović et al. (2018) [18] presented a satisfactory road transport system in the European Transport Committee based on the domestic market, fair competition, the realization of workers' rights, the implementation of decarbonization policies, using digital technologies in passenger transport in accordance with the principles of transport, sustainable transfers in cities, and fare and payment modes. Augustin et al. (2017) [19] investigated intercity buses in Germany and Italy based on a number of performance indicators, such as the travel company's development policies, the ticket prices, the frequency of service per day, the type of vehicles, and competitive indicators. Solak (2016) [20] evaluated the effects of the rules and policies of suburban road transport in Turkey as a platform for attracting more passengers, and they provided a structure to expand the contribution of public road transport. Afandizadeh and Safari (2020) [21] presented a set of models developed using clustering methods based on departure time.

The requirements for the intercity bus system are presented in the TCRP-79 (2019) [22] Report and used as a reference, the original version of which was published in 2002. This report addresses the issue of intercity buses in three sections. The first section deals with the requirements for developing intercity bus systems, related industries, and investment issues in this area. The second section presents a number of system improvement strategies in the form of planning, the establishment of affiliated branches, marketing, and the development of service facilities, smart equipment, and related items. The final section involves field studies carried out in different states to improve the system. The Interregional Travel Guide (2016) [23] discusses the topics related to passengers in suburban trips; however, these topics are not limited to the road system. This guide is important for providing a unified structure for developing benchmarks, gathering information, and providing an analytical tool.

Brzeziński et al. (2018) [24] analyzed the possibility of utilizing high-efficiency means of transportation in transportation companies as one of the most important ways to achieve a competitive advantage. This point is especially important in the passenger transport services market in large areas, because, in addition to the economic dimension, this also involves a social aspect. It should be noted that their study focused on groups of passenger transport vehicles. Wang et al. (2020) [25] argued that the rapid construction and development of existing high-speed rail (HSR) lines in China had increased the convenience of this mode of transportation and attracted passengers from other modes of transportation. Yan et al. (2021) [26] stated that the rapid growth in demand for intercity travel had put significant pressure on passenger transport networks in various countries. Wang et al. (2020) [27] evaluated the inevitable transfer in a multimode public transport network, and while confirming it, they explored the planning and management for an efficient transmission network and identified the needs that would lead to understanding the factors of these transfers. Paudel (2021) [28] considered the relationship between congestion and the quality of services in transportation systems as one of the concerns of citizens, planners, transportation managers, and passengers.

3. Methodology

In order for an inquiry, the first step is to determine the size of the needed sample for meaningful analyses. According to the annual statistics recorded by the Road Maintenance and Transportation Organization of Iran (2019) [29], more than 85% of intercity passengers travel through 20 main terminals. Equation (1) was used to determine the number of samples required in the mentioned terminals (Transit Capacity and Quality of Service Manual, 2013) [30].

$$N = \left| \frac{z_{1-\alpha/2}^{2} p(p-1)}{d^{2}} \right|, \tag{1}$$

where Nis the adequate number of passengers, $z^2_{1-(\alpha/2)}$ is the normal distribution at the confidence level of α% ($z^2_{1-(\alpha/2)} = 1.96$ based on the confidence level of 95%), P is the distribution percentage ($p = 0.5$), and d is the acceptable error ($d = 0.5$).

In order to analyze the opinions of stakeholders of road transport, a questionnaire was designed, a part of which represented pairwise comparisons, while the variables required for rotary analysis were included in its hidden layer. The questionnaire was given to technical experts, office managers, representatives of the union, drivers, and passengers of carriers throughout the country, and they were asked to fill it in a given time.

In the first section of the questionnaire, aimed at the pairwise comparison, the respondents were asked to give scores to each of the determined indicators based on the developed questions. In the second section of the questionnaire, entitled quality assessment, the respondents were requested to express their qualitative opinions on road transport infrastructures and services to indicate their satisfaction with the system in the intended sample, according to the variables expressed and the separated provincial socioeconomic data of the National Statistics Center of Iran [31] and the Ministry of Roads and Urban Development of Iran [1, 29]; a suitable multivariate linear estimation model of the number of suburban passengers can be constructed. In the third section, the participants were asked to give their solutions to increase the share of public road transport in passenger transportation in the form of general and brief opinions. A statistical analysis program adaptable with the analytic features of Excel (XLSTAT) was employed to analyze the keywords, and the results were displayed using the word cloud method. This method was effective in identifying the frequency and similarity of opinions from the participants' responses.

Regarding the first section of the questionnaire, since the studied indicators were returned from micro-level to macro-level and vice versa, their direct pairwise comparison indicates a linear curve with frequent breaks while paving the path for a back-analysis. Therefore, in order to classify the indicators, the questions were asked from the participants in the form of pairwise comparison to identify the bilinear and hidden effective level and, accordingly, perform the ultimate classification through rotary analysis (using the R statistical programming language) and introduce the output of the classification as the input to the pairwise comparison weighting through fuzzy AHP. It is worth noting that rotary analysis is a statistical method that uses the studied data in the analysis of the same data and paves the path for the evaluation of other data from the same class concerning their nature. That is why it is called rotary analysis. In the first phase of classification, the concentration density of respondents on the performance levels of indicators is investigated. As can be seen in (2), by using the power cosine distribution function, the classification of the concentration of respondents on the performance levels of indicators can be shown in Figure 1.

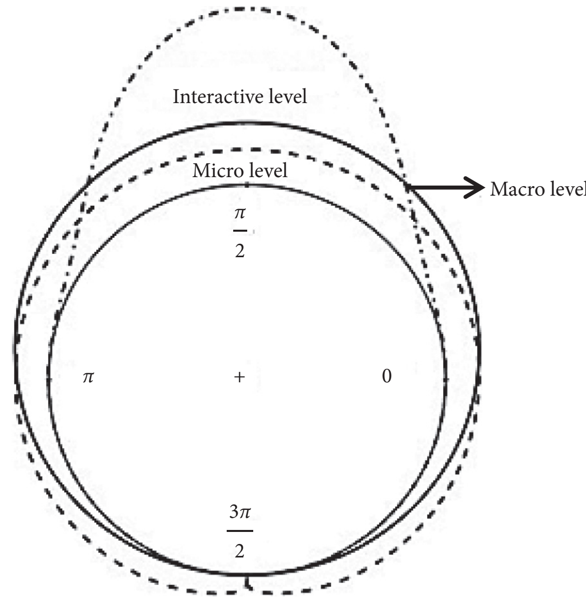


Figure 1: The classification of indicator levels using cosine power distribution.

$$f(\theta) = \frac{2^{-1+1/\zeta}\Gamma^2(1+1/\zeta)}{\pi\Gamma(1+2/\zeta)}(1+\cos(\theta-\mu))^{1/\zeta}. \quad (2)$$

After determining the density of concentration on levels, the indicators are classified to be weighted regardless of their performance levels. According to equation (3), using the von Mises density distribution function and based on the concentration density, the classification levels of indicators can also be determined and displayed. Circular statistics in R (2014) [32]:

$$f(\theta) = \frac{1}{2\pi I_0(\kappa)}e^{\kappa\cos(\theta-\mu)}, I_p(\kappa) = \frac{1}{2\pi}\int_0^{2\pi}\cos p\theta e^{\kappa\cos(\theta)}d\theta. \quad (3)$$

In the next step, by performing the fuzzy hierarchical analysis, the weights of indicators in each class can be determined. Given the advantages of the fuzzy (AHP) hierarchical analysis in reflecting the experts' opinions, this method was employed to weight and analyze the participants' answers in the form of linguistic preferences from equal importance to completely more important and within three fuzzy spectra of the lower, middle, and upper limits. The participants had to score their opinions from 1 to 9, and from 1/2 to 1/9 (reverse structure) in the form of pairwise comparison.

Regarding the numerical representation of fuzzy expressions, it is suggested that a fuzzy number may be expressed as a triangle or a trapezoid. In the triangular case, the corresponding number is displayed as $M = (a, b, c)$, where a, b, and c represent the lowest possible value, the most probable value, and the maximum possible value for the desired number, respectively, and the desired number can be between a and c. The fuzzy hierarchical analysis was conducted in the study according to the method suggested by Wang (2008) [33] as follows:

The first step: plotting the hierarchy curve.

In any multicriteria analysis, plotting the hierarchy curve (decision tree) is one of the first and of course important steps. Because it is the goal after plotting this curve, the structure of the hierarchy of indicators and subindicators and the options are clearly defined. In principle, even before designing a fuzzy AHP questionnaire, the decision hierarchy plan must first be determined.

The second step: defining the fuzzy numbers to make pairwise comparisons.

In this stage, it is necessary to define the fuzzy numbers that are needed to make pairwise comparisons, so that experts can provide their answers accordingly. These numbers are already listed in Table 2.

The third step: creating the pairwise comparison matrix using fuzzy numbers.

At this step, the questionnaires have been provided to the experts, and they have answered them. Therefore, in this step, the matrix of pairwise comparisons containing fuzzy numbers is prepared.

The fourth step: calculating matrix S for each row of the pairwise comparison matrix.

The S are triangular fuzzy numbers calculated from the following equation:

$$\widetilde{S}_i = \sum_{j=1}^{m} \widetilde{M}_{g_i}^{j} \times \left(\sum_{i=1}^{n} \sum_{j=1}^{m} \widetilde{M}_{g_i}^{j} \right)^{-1}, \tag{4}$$

in which M denotes the triangular fuzzy numbers inside the pairwise comparison matrix. In fact, when calculating the matrix S, components of fuzzy numbers are added peer to peer and multiplied by the total inverse fuzzy sum. This step is similar to calculating the normalized weights in the ordinary AHP but with fuzzy numbers.

The fifth step: calculating the relative magnitude of Sis.

In this step, Sis are compared with respect to their magnitude based on the equation (5) and Figure 2:

$$V\left(M_2 \geq M_1\right) = hgt\left(M_1 \cap M_2\right) = \mu_{M_2}(d) = \begin{cases} 1, & \text{if } m_2 \geq m_1, \\ 0, & \text{if } l_1 \geq u_2, \\ \dfrac{l_1 - u_2}{\left(m_2 - u_2\right) - \left(m_1 - l_1\right)}, & \text{otherwise}, \end{cases} \tag{5}$$

From the equation above,

$$M_2 = \left(l_2, m_2, u_2\right), M_1 = \left(l_1, m_1, u_1\right). \tag{6}$$

The sixth step: calculating the weights of criteria and options in pairwise comparison matrices.

In this step, the unnormalized weight vector is obtained by finding the lowest value of the calculated Vs in the previous step.

The seventh step: calculating the ultimate weight vector.

In the last step, the weight vector obtained in the previous step is normalized to achieve the ultimate weight vector as the goal of fuzzy computations.

4. Data Description

The studied data were defined in three sections. In the first section, with the goal of pairwise comparison of indicators, the respondents were asked to give scores to each of the determined indicators based on the designed questions.

(i) Dynamics of incentive policies: This variable indicates whether the policies considered at the macro-level to develop a public road transport system are up-to-date or not.

(ii) Traffic performance of the system: This variable denotes the performance of the road transport system, including its infrastructures and executive tools. Therefore, it involves the performance and function of some cases, such as the qualitative level of roads, terminals, buses, minibuses, and road taxis.

(iii) System coverage: This variable represents the coverage level of the public road transport fleet by the whole origin-destination pairs of the country.

(iv) System integrity: It denotes the function and interaction of different modes of public road transport and the performance of the related executive organizations with each other.

(v) Safety: It indicates the safety level of the public road transport system.

(vi) Environment: This variable represents the indicators associated with the environment, especially air pollution, regarding the road transport system.

(vii) Trip time: It indicates the condition and quality of choosing the public road transport mode concerning trip time.

(viii) Resting services: It indicates the quality of resting services regarding the public road transport infrastructures, such as terminals and roadhouses.

(ix) Smart systems: This variable indicates the effect of using smart systems in attracting passengers and the way they are employed in the current condition of the public road transport system.

Table 2: Fuzzy spectrum corresponding to linguistic preferences.

Linguistic preferences	Importance	Low limit (L)	Midline (M)	High limit (U)
Equal preference	1	1	1	1
Preference lower than medium	2	1	2	3
Medium preference	3	2	3	4
Medium preference to high	4	3	4	5
High preference	5	4	5	6
High preference to very high	6	5	6	7
Very high preference	7	6	7	8
Very high preference to completely high	8	7	8	9
Completely high preference	9	8	9	10

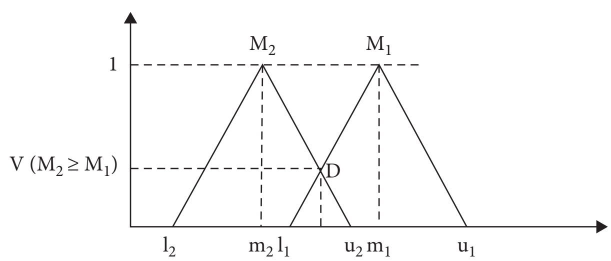


Figure 2: Pairwise comparison of triangular fuzzy numbers.

In the second section of the questionnaire, the qualitative opinions of the participants on the road transport infrastructures and services were received to identify the level of satisfaction with the system.

In the third section of the questionnaire, the respondents' opinions were obtained in the form of solutions to increase the share of public road transport in transferring the passengers.

Given the existing limitations, the inquiries were made from the passengers in four successive days from 2019 to 02-12 to 2019-02-15. The surveys in the West, East, South, and Beyhaghi Terminals of Tehran (the commercial and administrative center of Iran) were made face-to-face. Meanwhile, the coordinated representatives performed the surveys in the other 20 terminals selected (Tabriz Central Bus Terminal, Kaveh Bus Terminal in Isfahan, Shahid Kalantari Bus Terminal in Karaj, Bushehr Bus Terminal, Imam Reza Bus Terminal in Mashhad, Siahat Bus Terminal in Ahvaz, Enqelab Bus Terminal in Zahedan, Shahid Karandish Bus Terminal in Shiraz, Azadegan Bus Terminal in Qazvin, Qom Bus Terminal, Shahid Kaviani Bus Terminal in Kermanshah, Eastern Borujerd Bus Terminal, Sari State Bus Terminal, Persian Gulf Bus Terminal in Bandarabbas, Hamedan Bus Terminal, and Yazd Bus Terminal) during the same period. A total of 330 samples, including 114 passengers, 63 drivers, 32 members of passenger unions, and 21 traffic experts, were selected and surveyed. Of the participants, 45% had B.Sc. Degrees, and 29% had M.Sc. or higher degrees. The experts, separated by expertise and management levels, were selected, such that all levels (from associate to Ph.D. degrees) were covered. The passengers were chosen randomly, and the drivers were selected according to their vehicles and working hours.

5. Results and Discussion

This study aims to replace the traditional methods of attracting passengers to the public road transport system with scientific techniques and identify effective solutions to increase the share of this transportation mode with the aid of the users' opinions. It should be noted that the efficiency of social networking services, especially on regional and local scales, has been recognized based on the studies by Sala et al. (2021) [9] and Yao (2021) [10]. However, given the filtering of such tools as Twitter and Telegram in Iran, and the national scale of this study, using questionnaires seems to be still more efficient unless the conditions change. The study by Ballis and Dimitriou (2020) [11] had a similar suggestion.

5.1. Evaluation of the Information regarding the First Section of the Questionnaire: Pairwise Comparison of the Indicators. Table 3 lists the classification results of the indicators used in this study for policy-making in road transport to increase its share in micro- and macro-levels using the indicators effective in the desirability of public road transport system and the power cosine distribution function.

The classification of the concentration of respondents on the performance levels of indicators revealed that they had equal comparisons in the micro- and macro-levels. However, regarding the indicators with conflicting levels, they expressed different opinions, proving the performance power of these indicators (e.g., safety, environment, and system coverage). It also revealed that the respondents had properly separated the micro- and macro-levels and had not positioned them in one analytical field.

After determining the concentration density on levels, the indicators should be classified to be weighted regardless of their performance levels. According to equation (3), by using the von Mises density distribution function and based on the concentration density, the levels of indicators were divided into two groups, one with five indicators and the other with four indicators, as shown in Table 4. As can be seen in Figure 3, four indicators are grouped and shown by dashes, and four other indicators are in another group shown by dots. Another indicator (shown by a solid line), positioned between the two groups, is classified in the group shown by dots given its shorter polar distance with them.

Table 3: Indicators affecting the utility of the public rural road transport system.

Effective indicators	Micro-level	Macro-level
The dynamics of incentive policies		✓
System traffic performance	✓	
System coverage	✓	✓
System integrity	✓	
Safety	✓	✓
The environment	✓	✓
Time of trip	✓	
Resting services	✓	
Smart systems	✓	

Table 4: Classification of indicators based on rotational analysis.

First category	Second category
The dynamics of incentive policies	System traffic performance
System coverage	Safety
System integrity	Time of trip
The environment	Smart systems
Resting services	—

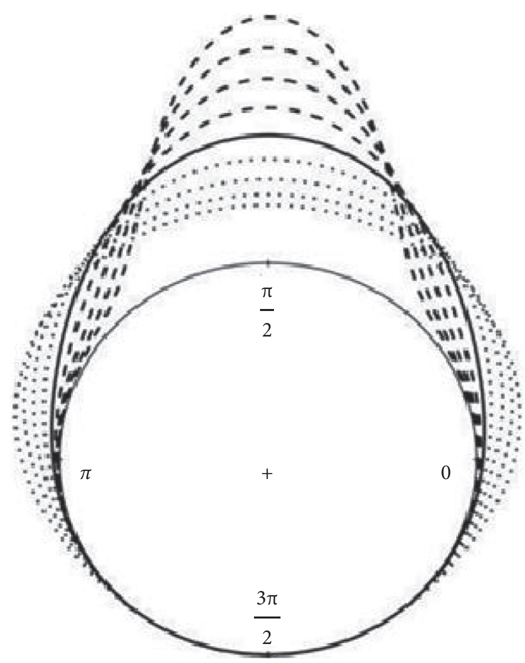


Figure 3: Functional classification of indicators for weighting.

In the next step, regarding the defined classification, the weights of indicators in each class were determined by performing the fuzzy hierarchical analysis. Table 5 provides the ultimate results of the stakeholders' opinions.

According to the statistics published in the World Bank Report on Iran (2016) [34] and National Statistics Center (2019) [31], the main competitor of public road trips is trips by private cars, and rail and air transport methods have no great potential due to their improper infrastructures in Iran. This issue, along with the high importance of safety in trips reported by the users given the high rate of road accidents in Iran, indicates the necessity to improve the safety of public trips. Moreover, according to the Office of Transportation Planning and Economics (2017) [35], public trips are safer than private ones. The delays, dissatisfactions, and optimum time value formulation introduced by Kucharski and Cats (2020) [15] are different from the obtained results. The reason is that the users in Iran pay attention to the high rate of road accidents and underestimate the importance of the environment, smart systems, and trip time, which can be expected from the users in a developing country like Iran. It seems that the concern of Iranian users over safety influences important indicators like trip time mentioned in other studies.

Table 5: Results of weighting the indicators.

The first category	Weight 0.43	The second category	Weight 0.57
Indicator	Weight	Indicator	Weight
The dynamics of incentive policies	0.27	Safety	0.39
Resting services	0.21	System traffic performance	0.26
System coverage	0.19	Time of travel	0.19
System integrity	0.18	Smart systems	0.16
The environment	0.15	—	—

5.2. Evaluation of the Information regarding the Second Section of the Questionnaire. In the second section, six questions were asked to evaluate the quality of the current road transport infrastructures and services. The following provides each question, along with the answers given to it:

(i) The first question: What is your evaluation of the status of the country's roads regarding infrastructures?

Of the participants, 51% described the status of the country's roads as weak in terms of infrastructures. Moreover, 43% gave a good score, and 6% gave an excellent score to the road infrastructures of the country. Given the undeveloped infrastructures of Iran according to the Plan and Budget Organization (2015) [36], the similarity of the users' opinions with the multidimensional attitude of Weckström (2021) [8] and Jafino et al. (2020) [4] indicates the low expectations of the Iranian users.

(ii) The second question: What is your evaluation of the status of the country's roads regarding safety?

Of the respondents, 69% stated that the country's roads had weak safety. Moreover, 26% gave a good score, and 5% gave an excellent score to the safety of the country's roads. Hörl and Balac (2021) [7] mentioned the role of safety in repeatable trips, which can help identify a solution in the current study.

(iii) The third question: What is your evaluation of the services provided by the carriers?

Fifty-one percent of the participants described the services provided by carriers as good. In addition, 40% gave a weak score, and 9% gave an excellent score to the state of their services. The responses to this question have a good agreement with the condition of the carriers that provide services to the users.

(iv) The fourth question: What is your evaluation of the status of passenger terminals regarding their provided services and construction quality?

Of the whole participants, 61% stated that the passenger terminals had good performance regarding their provided services and construction quality. Furthermore, 37% gave a weak score, and 2% gave an excellent score to the state of the services provided by them. The relative satisfaction of the users can indicate the potential to attract more passengers to the public transport system. Mo (2021) [6] mentioned this competitive attitude.

(v) The fifth question: What is your evaluation of the access to passenger terminals?

Sixty-nine percent of the respondents described the condition of access to passenger terminals as good. Meanwhile, 28% gave a weak score, and 3% gave an excellent score to this variable. According to Fournier et al. (2021) [14], it can be realized that the users are relatively satisfied with the access costs.

(vi) The sixth question: In your idea, to what extent do carriers have delays?

Of the respondents, 52% stated that the carriers have a good performance regarding the probable delays. However, 47% gave a weak score, and only 1% gave an excellent score to this variable of the carriers. Although the valuation of the passengers to the delay time has no conflict with the results of Yap and Cats (2021) [5], the Iranian users do not give the delay in trips the importance that they should, due to the lack of fast access systems like subways to most terminals of the country. This issue is more analyzed in the fourth part of this section.

Based on the questions asked and statistics information, a multivariate linear estimation model of the number of rural passengers can be constructed. This model helps measure the impact of changes in different sectors on passenger volume in order to apply appropriate strategies accordingly.

Regarding the bus passenger estimation model, according to the variables expressed and the separated provincial socio-economic data of the National Statistics

Table 6: Results of constructing the bus passenger estimator model in rural trips.

Variables name	Coefficient	t-test	Variables	Coefficient	t-test
1Age	3.13e−2	2.64	11Number of transportation companies	2.34e−5	2.76
2Employment rate	0.231	−2.32	12Number of buses	1.73e−2	3.45
3Family size	0.103	2.23	13Bus age	4.33e−3	−3.23
4Gender composition	2.234	−2.63	14Number of resting complexes	5.97e−5	2.43
5Per capita car ownership	4.25e−2	−3.65	15Freeway and highway length	4.61e−2	3.44
6Level of education	2.67e−5	−2.28	16The length of the arterial pathways	2.27e−4	2.23
7Economic strength of the household	2.43e−2	−3.29	17Waiting time (delay)	4.83e−2	−3.88
8Travel expenses	2.11e−3	−4.12	18Access	5.69e−3	4.45
9The purpose of the trip	6.25e−2	3.75	19Fixed model	0.37	2.23
10Number of accidents (killed)	1.78e−4	−2.89			

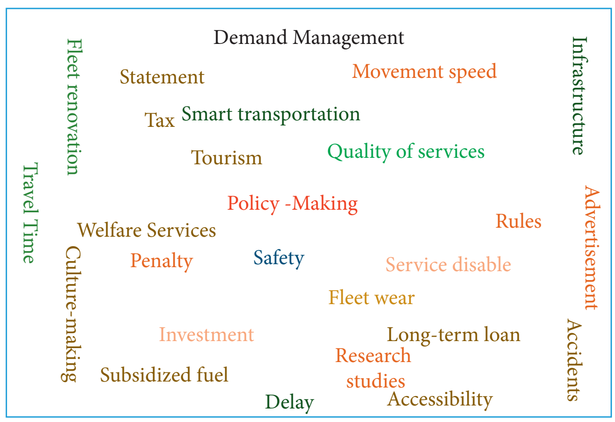


Figure 4: Word cloud map for the answers of the respondents to the questionnaire.

Center of Iran [31] and the Ministry of Roads and Urban Development of Iran [1] [29], the estimated volume of rural bus passengers is as follows:

$$\begin{aligned} U_{Bus} = {} & \overline{\beta_1} \times v1 - \overline{\beta_2} \times v2 + \overline{\beta_3} \times v3 - \overline{\beta_4} \times v4 - \overline{\beta_5} \times v5 - \overline{\beta_6} \times v6 - \overline{\beta_7} \times v7 - \overline{\beta_8} \times v8 \\ & + \overline{\beta_9} \times v9 - \overline{\beta_{10}} \times v10 + \overline{\beta_{11}} \times v11 - \overline{\beta_{12}} \times v12 + \overline{\beta_{13}} \times v13 + \overline{\beta_{14}} \times v14 + \overline{\beta_{15}} \times v15 \\ & + \overline{\beta_{16}} \times v16 - \overline{\beta_{17}} \times v17 + \overline{\beta_{18}} \times v18 + \overline{\beta_{19}} \times v19 + C_i + \varepsilon_i. \end{aligned} \tag{7}$$

This model is made using data from 2008 to 2019. The results of the model construction and names are presented in Table 6.

5.3. Evaluation of the Stakeholders' Responses in the Words Cloud Section. In the third section of the questionnaire, the users were requested to express their opinions about the short-term and long-term solutions and programs to increase the share of public road transport in the transportation of passengers in the form of separate items. Figure 4 demonstrates the obtained word cloud, which indicates the concentration of the suggestions.

The suggested solutions focus on policy-making, trip time optimization, safety level improvement, and quality of services. The other keywords extracted from the suggestions can also be seen in the figure. The opinions are depicted either horizontally or vertically, given the small space of the figure. The larger thickness of the text for an indicator represents its more repetitions.

As can be seen, the results of this section agree with those of the two previous sections. The users preferred their personal benefits to the system, as stated by Hoque et al. (2021) [2], and the utilitarianism principle of Jafino (2021) [3] is also observed. In this section, the users paid slightly lower attention to safety and gave more importance to management issues, such as trip time and quality of services. The difference between the results of this section indicates the importance of the surveying method and the fact that the type of questions can lead to the answers. This can be one of the reasons to consider the data of social networks as a supporting factor.

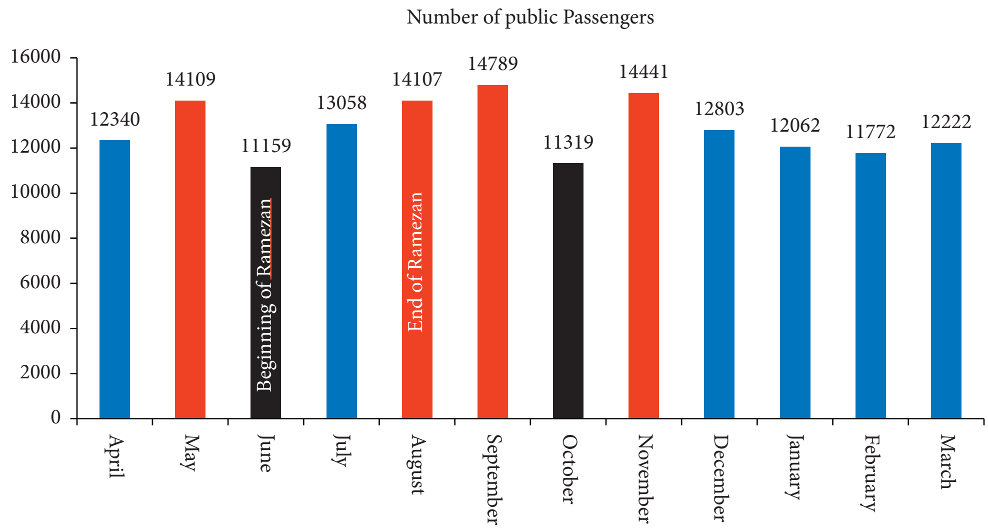


Figure 5: The bar graph of the number of public passengers (thousand) in different months of 2019.

5.4. Integrated Analysis of the Stakeholders' Answers. In order to draw a conclusion from the users' responses, the travel status of the passengers in the rural public transport system during different months of a year is investigated. Figure 5 demonstrates the travels of passengers in 2019. May, August, September, and November have the highest number of public travels.

The least travels have occurred in June due to the beginning of Ramadan and October due to the reopening of the schools and universities. The relative agreement between Figures 5 and 6 indicates that the most travels occur in summer and the months with more holidays, but the goals of trips in these months should also be studied.

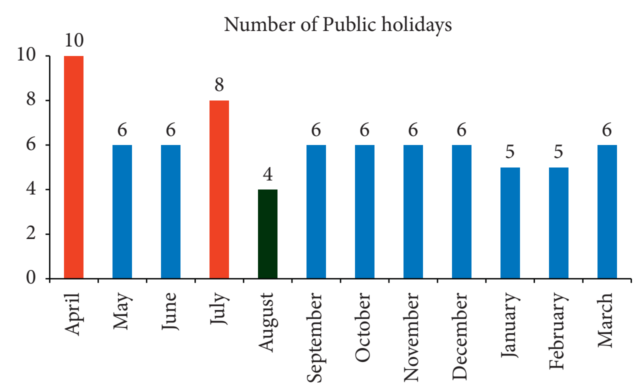


Figure 6: Number of public holidays in each month of 2019.

According to Figure 5, in each month, on average, 12831 thousand passengers are transferred by the public transport system of the country. Meanwhile, according to the annual report of the Road Maintenance and Transportation Organization (2019), a large portion of travels, especially on holidays, take place with private cars (4% with buses and minibuses, and 2% with automobiles). Based on the origin-destination report of the country (2016), most of the travels by bus have occupational (42%) and educational (19%) objectives. Accordingly, of the total monthly travels, 5389 thousand have occupational objectives. A glance at the answers given to the question concerning the delay in travels and performance of terminals reveals that while a large portion of travels happen with occupational objectives, the users have no considerable sensitivity to delays and performance of passenger terminals.

Figure 7: Location of industrial towns.

Moreover, according to the data of the geographic information system of industrial towns and areas of Iran (2018) [31], the average distance between industrial towns and cities is 30 km, as shown in Figure 7. On the other hand, almost 41700 industrial and manufacturing units are active in the industrial towns of the country, in which 729 thousand people work. With an average of 25 working days per month, there is a potential for 36450 thousand travels, 15% of which have been attracted to public transport. The low share of public business travels from the total potential working travels in the country can be attributed to the indifference of existing users to the coverage and integrity of the system and the quality of the rural road infrastructures.

6. Conclusions

The dynamics of incentive policies in the first class and safety in the second class have greater weights, as an indicator of the macro-level and an indicator of the interactive level, respectively, which are along each other according to the rotary analysis. Thus, the weighting results indicate that the dynamics of incentive policies and policy-making with a focus on safety can highly influence the micro- and macro-level indicators to increase the share of public road transport.

In the first class, following the dynamics of incentive policies, the coverage and integrity of the system have, respectively, greater weights. Regarding the comparison with the opinions of the users in the second section and the 50% satisfaction with the infrastructures, in this class, the users prioritize development and incentive policies to complete the existing network rather than constructing new roads. Almost 70% of the respondents have described safety as weak, indicating the correct weight of this indicator in the second class of the first section of the questionnaire. The satisfaction with passenger terminals (61%) and with access to them (69%), along with the relative satisfaction with the infrastructures, can indicate the potential to attract more passengers in the existing conditions. The only risk leading to the undesirability of this method is the strong belief of the users in the lack of safety in the system. This shows the importance of focusing on making society aware of the higher safety of public travels compared to private ones, as well as trips with short distances to attract passengers in the short term.

In the word cloud section, safety was mentioned after trip time and services, but there is an acceptable agreement with the weights obtained in the first section overall. The access and incentive policies were again emphasized, while the users still did not pay attention to smart transportation. Regarding the delay in trips and satisfaction with carriers, some differences can be observed between the classifications and opinions of users, which is due to the large portion of business travels and can be more analyzed in future studies. The passengers with occupational objectives, who travel for work, have a small share among the total intercity passengers, including public and private transport. Meanwhile, this small share makes up 40% of the public travels of the country. This is the reason for the lower attention of Iranian users to delays and the role of passenger terminals in increasing the desirability of rural public road travels.

By increasing the costs of business travels with private cars in short rural distances through interactions with industrial towns and providing these areas with special services, the finances of public transport can be improved, and the potential to make the number of rural public travels six times larger can be realized. Establishing carriers in industrial towns and encouraging their workers to use public transport, creating smart mechanisms to provide special facilities in this field (e.g., timing and ticket applications), offering special discounts, providing VIP services for managers of workshops and factories to promote using this system, coordination with managers of towns in macro-levels to impose restrictions on the passage of private cars, i.e., increasing the performance costs of this system, are all appropriate solutions that further activate the public rural transport system in this field and effectively increase public travels.

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.

References

[1] R. Maintenance and Transportation Organization, "Road maintenance and transportation organization," *Statistical Yearbooks of Road Transport from 2008 to 2019*, Ministry of Road and Urban Development, I. R. Of Iran, 2019, http://www.rmto.ir.

[2] J. M. Hoque, G. D. Erhardt, D. Schmitt, M. Chen, and M. Wachs, "Estimating the uncertainty of traffic forecasts from their historical accuracy," *Transportation Research Part A: Policy and Practice*, vol. 147, pp. 339–349, 2021.

[3] B. A. Jafino, "An equity-based transport network criticality analysis," *Transportation Research Part A: Policy and Practice*, vol. 144, pp. 204–221, 2021.

[4] B. A. Jafino, J. Kwakkel, and A. Verbraeck, "Transport network criticality metrics: a comparative analysis and a guideline for selection," *Transport Reviews*, vol. 40, no. 2, pp. 241–264, 2020.

[5] M. Yap and O. Cats, "Taking the path less travelled: valuation of denied boarding in crowded public transport systems," *Transportation Research Part A: Policy and Practice*, vol. 147, pp. 1–13, 2021.

[6] B. Mo, Z. Cao, H. Zhang, Y. Shen, and J. Zhao, "Competition between shared autonomous vehicles and public transit: a case study in Singapore," *Transportation Research Part C: Emerging Technologies*, vol. 127, Article ID 103058, 2021.

[7] S. Hörl and M. Balac, "Synthetic population and travel demand for Paris and Île-de-France based on open and publicly available data," *Transportation Research Part C*, vol. 130, Article ID 103291, 2021.

[8] C. Weckström, M. N. Mladenović, R. Kujala, and J. Saramäki, "Navigability assessment of large-scale redesigns in nine public transport networks: open timetable data approach," *Transportation Research Part A: Policy and Practice*, vol. 147, pp. 212–229, 2021.

[9] L. Sala, S. Wright, C. Cottrill, and E. Flores-Sola, "Generating demand responsive bus routes from social network data analysis," *Transportation Research Part C: Emerging Technologies*, vol. 128, Article ID 103194, 2021.

[10] W. Yao and S. Qian, "From Twitter to traffic predictor: next-day morning traffic prediction using social media data," *Transportation Research Part C: Emerging Technologies*, vol. 124, Article ID 102938, 2021.

[11] H. Ballis and L. Dimitriou, "Revealing personal activities schedules from synthesizing multi-period origin-destination matrices," *Transportation Research Part B: Methodological*, vol. 139, pp. 224–258, 2020.
[12] J. Khan, R. Hrelja, and F. Pettersson-Löfstedt, "Increasing public transport patronage - an analysis of planning principles and public transport governance in Swedish regions with the highest growth in ridership," *Case Studies on Transport Policy*, vol. 9, no. 1, pp. 260–270, 2021.
[13] J. B. d. Santos and J. P. Lima, "Quality of public transportation based on the multi-criteria approach and from the perspective of user's satisfaction level: a case study in a Brazilian city," *Case Studies on Transport Policy*, vol. 9, no. 3, pp. 1233–1244, 2021.
[14] N. Fournier, E. Christofa, and E. J. Gonzales, "A continuous model for coordinated pricing of mixed access modes to transit," *Transportation Research Part C: Emerging Technologies*, vol. 128, Article ID 103208, 2021.
[15] R. Kucharski and O. Cats, "Exact matching of attractive shared rides (ExMAS) for system-wide strategic evaluations," *Transportation Research Part B: Methodological*, vol. 139, pp. 285–310, 2020.
[16] S. Deb, L. Strawderman, D. W. Carruth, J. DuBien, B. Smith, and T. M. Garrison, "Development and validation of a questionnaire to assess pedestrian receptivity toward fully autonomous vehicles," *Transportation Research Part C: Emerging Technologies*, vol. 84, pp. 178–195, 2017.
[17] S. Jokubauskaitė, R. Hössinger, F. Aschauer et al., "Advanced continuous-discrete model for joint time-use expenditure and mode choice estimation," *Transportation Research Part B: Methodological*, vol. 129, pp. 397–421, 2019.
[18] V. D. Popović, P. Gladović, M. Miličić, and M. Stanković, "Methodology of selecting optimal fare system for public transport of passengers," *Promet - Traffic & Transportation*, vol. 30, no. 5, pp. 539–547, 2018.
[19] R. Grimaldi, K. Augustin, and P. Beria, "Intercity coach liberalisation. The cases of Germany and Italy," *Transportation Research Procedia*, vol. 25, no. 3, pp. 474–490, 2017.
[20] A. O. Solak, "Regulations in scheduled intercity coach transport sector in Turkey," *International Journal of Economics and Finance*, vol. 8, no. 8, p. 33, 2016.
[21] S. Afandizadeh Zargari and F. Safari, "Using clustering methods in multinomial logit model for departure time choice," *Journal of Advanced Transportation*, vol. 2020, pp. 1–12, 2020.
[22] Transportation research board and National research council, *Effective Approaches to Meeting Rural Intercity Bus Transportation Needs. TCRP REPORT 79, Transportation Research Board*, National Research Council, Italy, 2019.
[23] Transportation research board, *Interregional Travel*, TRB Publications, Washington, DC, 2016.
[24] M. Brzeziński, M. Kijek, P. Owczarek, K. Głodowska, J. Zelkowski, and P. Bartosiak, "Aspects of improvement in exploitation process of passenger means of transport," *Journal of Advanced Transportation*, vol. 2018, pp. 1–13, 2018.
[25] B. Wang, S. Ni, F. Jin, and Z. Huang, "An optimization method of multiclass price railway passenger transport ticket allocation under high passenger demand," *Journal of Advanced Transportation*, vol. 2020, Article ID 8860115, 15 pages, 2020.
[26] H. Yan, X. Zhang, and X. Wang, "Hierarchical passenger hub location problem in a megaregion area considering service availability," *Promet - Traffic & Transportation*, vol. 33, no. 2, pp. 247–258, 2021.
[27] W. Wang, Y. Wang, G. H. d. A. Correia, and Y. Chen, "A network-based model of passenger transfer flow between bus and metro: an application to the public transport system of beijing," *Journal of Advanced Transportation*, vol. 2020, pp. 1–12, 2020.
[28] J. Paudel, "Bus ridership and service reliability: the case of public transportation in Western Massachusetts," *Transport Policy*, vol. 100, pp. 98–107, 2021.
[29] R. Maintenance and Transportation Organization, *Road Transport Traffic Census Yearbook 2019, Management Reports Section*, Road Maintenance and Transportation Organization, Ministry of Road and Urban Development, I.R. Of Iran, 2019, http://www.rmto.ir.
[30] National Academies of Sciences Engineering Medicine, "Transit capacity and quality of service manual, third edition," *Transit Capacity and Quality of Service Manual*, The National Academies Press, Washington, DC, Third Edition, 2013.
[31] National Statistics Center, *Results of the General Population and Housing Census*, National Statistics Center of Iran, Plan and Budget Organization, Presidency Islamic Republic of Iran, Iran, 2018, http://www.amar.org.ir.
[32] A. Pewsey, M. Neuhäuser, and G. D. Ruxton, *Circular Statistics in R*, Oxford University Press, Oxford, United Kingdom, 2014.
[33] Y.-M. Wang, Y. Luo, and Z. Hua, "On the extent analysis method for fuzzy AHP and its applications," *European Journal of Operational Research*, vol. 186, no. 2, pp. 735–747, 2008.
[34] R. Maintenance and Transportation Organization, *World Bank Report on Iran*, Road Maintenance and Transportation Organization Ministry of Road and Urban Development, I.R. Of Iran, 2016, http://www.rmto.ir.
[35] Office of Transportation Planning and Economics, *Comprehensive Plan of the Country's Transportation, Management Report*, Deputy Minister of Transportation, Ministry of Road and Urban Development, I.R. Of Iran, 2016, https://www.mrud.ir.
[36] *Detailed Document of the Sixth Economic, Social and Cultural Development Plan of the Islamic Republic of Iran. Plan and Budget Organization, Presidency Islamic Republic of Iran*, Publications of the Management and Planning Organization, Vienna, Austria, 2015, https://www.mporg.ir.

Hindawi
Complexity
Volume 2021, Article ID 6860754, 16 pages
https://doi.org/10.1155/2021/6860754

Research Article

Analysis on Decision-Making Changes of Multilevel Governments and Influencing Factors in Watershed Ecological Compensation

Yuansheng Wang,[1] Xiaojing Wu,[1] Juqin Shen,[2] Chen Chi,[3] and Xin Gao[4]

[1]*Jiangsu Taihu Planning and Design Institute of Water Resources Co., Ltd., Suzhou 215103, China*
[2]*College of Agricultural Science and Engineering, Hohai University, Nanjing 211100, China*
[3]*Business School, Hohai University, Nanjing 211100, China*
[4]*Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai 200030, China*

Correspondence should be addressed to Xin Gao; gxtz1987@sjtu.edu.cn

Received 19 August 2021; Accepted 17 November 2021; Published 9 December 2021

Academic Editor: Feng Li

Watershed ecological compensation has been widely accepted as a system to promote the cooperation of various stakeholders to solve the problem of transboundary water pollution, but the existing research does not fully consider the impact of compensation fee paid by different governments on stakeholders' decision-making. Therefore, this paper constructs a tripartite game model between upstream governments, downstream governments, and the central government by using evolutionary game theory and determines the influence of different factors on the decision-making process of each player through simulation. The results show the following: (1) the initial probability significantly affects the decision-making behavior of each player; (2) daily supervision of the central government and the reduction of the environmental protection cost can promote the implementation of watershed ecological compensation; (3) the fine to downstream governments makes the decisions of the central government and downstream governments change periodically; and (4) the increase of ecological compensation fee urges downstream governments to choose noncompensation, and compensation fee paid by the central government has a critical value.

1. Introduction

As the birthplace of human social civilization, watershed is one of the main sources of human available freshwater resources and an important part of natural ecosystem [1]. However, with the rapid development of social economy, the destruction of watershed ecosystem and water pollution are becoming more and more serious [2]. Due to the mobility of water resources, the pollution generated by upstream is easily transferred to downstream, resulting in the destruction of the ecological environment of the whole basin [3, 4]. This process forms transboundary water pollution involving multiple administrative regions in the basin [5]. Therefore, how to effectively solve the problem of transboundary water pollution has become the focus of governments and scholars.

Because different regions in the basin have different needs, in order to maximize their own interests, upstream and downstream areas often have interest conflicts around the development, distribution, and utilization of water resources, resulting in the "tragedy of the commons" [6]. For example, if upstream areas protect water resources, the local government needs to pay a high cost; meanwhile downstream areas enjoy the good water; they do not pay compensation to upstream areas. This will eventually lead to the alienation of regional economic development and living standards and cause the conflicts [7]. Since transboundary water pollution usually involves many different regions, it is difficult for them to take unified action to solve this problem [8]. In recent years, many scholars have conducted in-depth research on how to solve transboundary water pollution problem and proposed many solutions [9]. Among them, payment for watershed ecosystem services (PWES), as an effective system that can promote the cooperation of upstream and downstream areas to protect watershed ecological environment, has attracted extensive attention all over the world [10, 11]. Therefore, PWES provides a practical way to effectively solve the transboundary water pollution problem. PWES is the application of payments for ecosystem service (PES) in the basin, which is usually called watershed ecological compensation (WEC) in some countries. Because WEC has the advantages of promoting cooperation between upstream and downstream areas, it is widely used in major watersheds all over the world [12–14], such as Yangtze River and Yellow River in China [15–17], the Elbe River in Germany, and the Murray-Darling Basin in Australia.

For WEC, scholars have done much research on compensation identification [18, 19], standard [20–22], and mode [23, 24], while the key influencing factors and the decision-making changes of compensation subject and object have been attached much attention [25, 26]. For the research of influencing factors, scholars took the compensation subject as research object and combined econometrics and conditional value method to identify them. The studies found that demographic characteristics, water consumption patterns, the close dependence between upstream and downstream areas, the awareness of improving drinking water services, and so forth were the key influencing factors [27, 28], but these studies mostly focused on the compensation subject and paid less attention to compensation object. In addition, scholars only explained the influence results but could not show the influence process of various factors. For the research of decision-making, scholars took multilevel governments (multilevel governments refer to governments with different levels in the government management hierarchy. This paper mainly refers to the central government and local governments, in which local governments mainly refer to the provincial governments along the river) as the research object and used game theory for analysis. Some scholars analyzed the decision-making behaviors by constructing the static game model between the upstream and downstream governments and found that it was difficult to construct WEC between them spontaneously [29, 30]. However, because the static game assumes that the players are completely rational, it is difficult to realize in the real world. Therefore, some scholars have introduced evolutionary game theory [5, 31], but scholars mainly focused on how to promote upstream and downstream governments to establish horizontal compensation between governments at the same level, while ignoring the role of vertical compensation between the superior and subordinate governments. Moreover, the existing research has not discussed the matching relationship between horizontal compensation and vertical compensation.

Based on these research gaps, this paper constructs a tripartite evolutionary game model including upstream governments, downstream governments, and the central government in WEC, analyzes their decision-making behaviors, and identifies the key influencing factors. The innovations and contributions are as follows: (1) This paper explores the influencing factors from the perspective of multiple agents and uses simulation to show the change process of decision-making. (2) This paper also analyzes vertical compensation, horizontal compensation, and their ratio, which provides a new idea for WEC. The results can not only optimize the existing compensation mechanism but also provide reference for countries or regions that have not established WEC.

The remainder of this paper is arranged as follows. Section 2 introduces the theoretical background and research framework; Section 3 constructs the model; Section 4 carries out the simulation experiment and gets the results; Section 5 discusses the results; Section 6 obtains the conclusions and proposes suggestions.

2. Theoretical Background and Research Framework

2.1. Theoretical Background. Evolutionary game theory is an important branch of game theory [32]. It breaks through the hypothesis of complete rationality of players in classical game theory, integrates the evolutionary theory in biology, and uses the hypothesis of limited rationality to study the decision-making behavior of players [33, 34]. This theory has a great influence on the study of the interaction between multiple agents. A complete research paradigm of evolutionary game theory usually includes the following parts:

(1) Determination of strategy set: in the multiagent game, different players have different alternative strategies. Then, determining the selectable strategies of each player and forming the corresponding strategy set play an important role.
(2) Construction of game matrix: game matrix is a matrix composed of payoffs of different players under different strategic combinations. The construction of the matrix can clearly reflect the specific payoffs of each player under different conditions.
(3) Establishment of replication dynamic system: replication dynamic system is a set of equations composed of a set of differential equations. It reflects the proportion of players who choose different strategies in the group and reflects the decision-making change of the whole group through the change of the proportion.
(4) Determination of evolutionary stability strategy (ESS): ESS is the combination of strategies when each player reaches a stable state. Under this strategy combination, each player can obtain its own optimal utility and will not change its own strategy. Looking for ESS is the core of evolutionary game theory [35, 36].

In the WEC, the needs of various stakeholders are different, and it is difficult to maintain complete rationality in the game process, so it is impossible to determine the optimal strategy in a game. Therefore, evolutionary game theory is suitable for the study of WEC [18, 37], and the application of this theory can identify the changes of decision-making behavior and the key influencing factors.

2.2. Research Framework. The effective establishment and implementation of WEC need the full cooperation of multilevel governments, but they have different interest needs [10]. Therefore, when constructing WEC mechanism, upstream governments, downstream governments, and the central government constitute a game. The game process is shown in Figure 1.

According to Figure 1, the central government has two collective strategies: daily supervision and random inspection. When it chooses daily supervision, it hopes that the ecological, economic, and social benefits of the basin will be improved simultaneously, so as to realize the high-quality development of the basin. However, daily supervision will make the central government pay a huge cost, which will make it bear huge capital pressure. Therefore, it will choose random inspection to reduce the cost and supervise the establishment of WEC. After that, the upstream governments also have two optional strategies, namely, protecting water environment and not protecting water environment. When upstream governments choose protecting water environment, this can not only improve the ecological benefits of the whole basin but also provide high-quality water resources to downstream governments, but it will cost upstream governments. When they choose the other strategy, these costs can be used to develop the economy of the region, so as to improve the production and living standards of local residents. Finally, the optional strategy of the downstream governments is to compensate the upstream ones or not. When downstream governments pay compensation fee, this can promote upstream governments to continuously protect water environment and make downstream obtain high-quality water resources continuously. In contrast, downstream governments think that using high-quality water resources is their own right and protecting water environment is the responsibility of upstream governments, so they choose not to compensate them.

Based on the game involving multilevel governments, this paper deeply analyzes their decision-making behaviors and identifies the key influencing factors, so as to provide reference for the design of appropriate WEC mechanism. The research framework is shown in Figure 2.

3. Model

3.1. Hypothesis. In order to fully explain the changes of decision-making behavior of different governments in the WEC and identify the key influencing factors, this paper needs to set the necessary hypotheses before building the model. The hypotheses are as follows:

(1) Each player is bounded rationality, not complete rationality. This means that it is difficult for each player to find its own optimal strategy in one game, but it can find the optimal strategy in multiple games by learning.

(2) The goal of each player is to maximize its own interests.

(3) When the upstream governments protect water environment, governments at all levels can obtain ecological benefits.

3.2. Evolutionary Game Model

3.2.1. Variables. The game model contains many variables, as shown in Table 1.

3.2.2. Payoff Matrices. Because different players have two optional strategies, the payoff of each player is different under different strategy combinations (Tables 2 and 3).

(1) When the central government chooses daily supervision, upstream governments choose protecting water environment and downstream governments choose compensation.

The payoff of upstream governments is

$$V_1 - C_1 + H_1 + H_2. \quad (1)$$

The payoff of downstream governments is

$$V_2 - H_1. \quad (2)$$

The payoff of the central government is

$$V_3 - C_2 - H_2. \quad (3)$$

(2) When the central government chooses daily supervision, upstream governments choose protecting water environment and downstream governments choose no compensation.

The payoff of upstream governments is

$$V_1 - C_1 + H_2. \quad (4)$$

The payoff of downstream governments is

$$V_2 - F_2. \quad (5)$$

The payoff of the central government is

$$V_3 - C_2 - H_2 + F_2. \quad (6)$$

(3) When the central government chooses daily supervision, upstream governments choose not to protect water environment and downstream governments choose compensation.

The payoff of upstream governments is

$$H_1 - F_1. \quad (7)$$

The payoff of downstream governments is

$$-H_1. \quad (8)$$

The payoff of the central government is

$$F_1 - C_2. \quad (9)$$

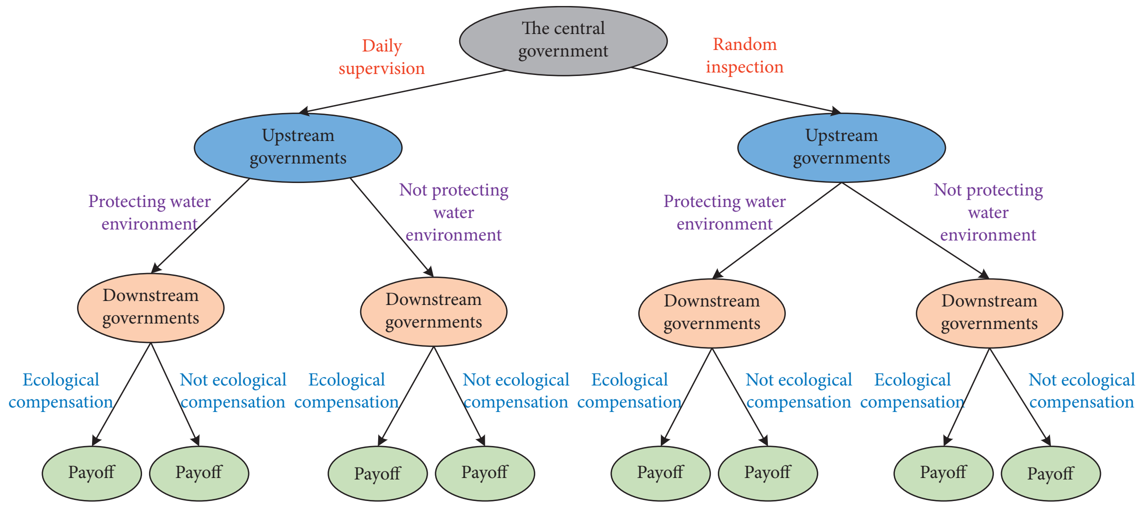


Figure 1: Game process.

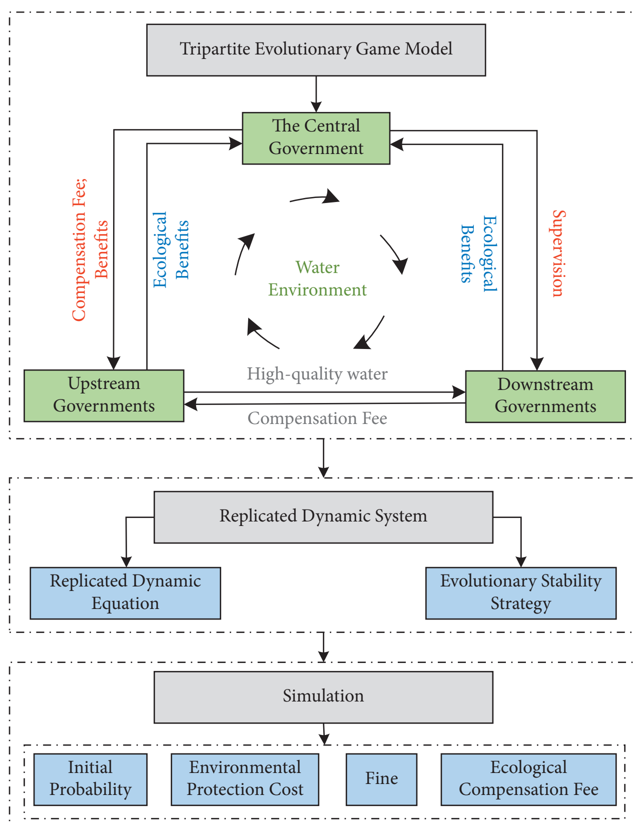


Figure 2: The research framework.

(4) When the central government chooses daily supervision, upstream governments choose not to protect water environment and downstream governments choose no compensation.

The payoff of upstream governments is

$$-F_1. \tag{10}$$

The payoff of downstream governments is

$$-F_2. \tag{11}$$

Table 1: The variables in the game model.

Variable	Definition
C_1	Total costs of upstream governments to protect water environment
C_2	Additional costs paid by the central government when choosing daily supervision
F_1	The fine imposed by the central government on upstream governments when upstream governments do not protect water environment
F_2	The fine imposed by the central government on downstream governments when downstream governments do not compensate
H_1	The ecological compensation fee paid by downstream governments
H_2	The ecological compensation fee paid by the central government
V_1	The ecological benefits obtained by upstream governments in protecting water environment
V_2	The ecological benefits obtained by downstream governments when upstream governments protect water environment
V_3	The ecological benefits obtained by the central government when upstream governments protect water environment

Notes: all variables listed in Table 1 are positive.

Table 2: The payoff matrix of each player with daily supervision.

When the central government chooses daily supervision strategy							
		Downstream governments					
		Ecological compensation			Nonecological compensation		
		Upstream governments	Downstream governments	The central government	Upstream governments	Downstream governments	The central government
Upstream governments	Protecting water environment	$V_1 - C_1 + H_1 + H_2$	$V_2 - H_1$	$V_3 - C_2 - H_2$	$V_1 - C_1 + H_2$	$V_2 - F_2$	$V_3 - C_2 - H_2 + F_2$
	Not protecting water environment	$H_1 - F_1$	$-H_1$	$F_1 - C_2$	$-F_1$	$-F_2$	$F_1 + F_2 - C_2$

The payoff of the central government is

$$F_1 + F_2 - C_2. \tag{12}$$

The payoff matrix with daily supervision of governments is shown in Table 2.

(5) When the central government chooses random inspection, upstream governments choose protecting water environment and downstream governments choose compensation.

The payoff of upstream governments is

$$V_1 + H_1 - C_1. \tag{13}$$

The payoff of downstream governments is

$$V_2 - H_1. \tag{14}$$

The payoff of the central government is

$$V_3. \tag{15}$$

(6) When the central government chooses random inspection, upstream governments choose protecting water environment and downstream governments choose no compensation.

The payoff of upstream governments is

$$V_1 - C_1. \tag{16}$$

The payoff of downstream governments is

$$V_2. \tag{17}$$

The payoff of the central government is

$$V_3. \tag{18}$$

(7) When the central government chooses random inspection, upstream governments choose not to protect water environment and downstream governments choose compensation.

The payoff of upstream governments is

$$H_1. \tag{19}$$

The payoff of downstream governments is

$$-H_1. \tag{20}$$

The payoff of the central government is

$$0. \tag{21}$$

(8) When governments choose random inspection, upstream governments choose not to protect water environment and downstream governments choose no compensation. The payoffs of them are 0, respectively.

The payoff matrix with random inspection of the central government is shown in Table 3.

Table 3: The payoff matrix of each player with random inspection.

		When the central government chooses random inspection strategy					
		Downstream governments					
		Ecological compensation			Nonecological compensation		
		Upstream governments	Downstream governments	The central government	Upstream governments	Downstream governments	The central government
Upstream governments	Protecting water environment	$V_1 + H_1 - C_1$	$V_2 - H_1$	V_3	$V_1 - C_1$	V_2	V_3
	Not protecting water environment	H_1	$-H_1$	0	0	0	0

3.3. Replicated Dynamic Equations. According to hypotheses, each player maximizes its own interests through learning in multiple games. Therefore, this paper uses x, y, and z (x, y, and $z \in\in [0,1]$) to represent the initial cooperation probability of upstream governments, downstream governments, and the central government choosing the strategies of protecting water environment, ecological compensation, and daily supervision, respectively.

In addition, it is assumed that π_{mn} represents the expected payoffs of each player choosing different strategies, and $\overline{\pi_m}$ refers to the average payoffs of each player. m represents each player, where $m = 1$, 2, and 3 denote upstream governments, downstream governments, and the central government, respectively; n refers to the different strategies for each player, where $n = 1$ and 2 denote the first strategy and the second strategy, respectively. The description of each symbol is listed in Table 4.

According to the above analysis of the payoff matrices, we can get the payoff of each player under different strategies, as shown below.

For upstream governments, the payoffs are

$$\begin{aligned}\pi_{11} &= y \times z \times (V_1 - C_1 + H_1 + H_2) + y \times (1 - z) \times (V_1 - C_1 + H_2) \\ &\quad + y \times (1 - z) \times (V_1 + H_1 - C_1) + (1 - y) \times (1 - z) \times (V_1 - C_1), \\ \pi_{12} &= y \times z \times (H_1 - F_1) + y \times (1 - z) \times (-F_1) + y \times (1 - z) \times H_1, \\ \overline{\pi_1} &= x \times \pi_{11} + (1 - x) \times \pi_{12}.\end{aligned} \tag{22}$$

Based on π_{11}, π_{12}, and $\overline{\pi_1}$, the replicated dynamic equation of upstream governments is as follows:

$$\begin{aligned}F_1(x) &= \frac{\mathrm{d}x}{\mathrm{d}t} \\ &= -x \times (x - 1) \times (V_1 - C_1 + F_1 \times z + H_2 \times z).\end{aligned} \tag{23}$$

For downstream governments, the payoffs are

$$\begin{aligned}\pi_{21} &= x \times z \times (V_2 - H_1) + (1 - x) \times z \times (-H_1) + x \times (1 - z) \\ &\quad \times (V_2 - H_1) + (1 - x) \times (1 - z) \times (-H_1), \\ \pi_{22} &= x \times z \times (V_2 - F_2) + (1 - x) \times z \times (-F_2) + x \times (1 - z) \times V_2, \\ \overline{\pi_2} &= y \times \pi_{21} + (1 - y) \times \pi_{22}.\end{aligned} \tag{24}$$

Table 4: The description of different symbols.

Symbol	Description
x	The probabilities that upstream governments choose to protect water environment
$1-x$	The probabilities that upstream governments choose not to protect water environment
y	The probabilities that downstream governments choose ecological compensation
$1-y$	The probabilities that downstream governments choose nonecological compensation
z	The probabilities that the central government chooses daily supervision
$1-z$	The probabilities that the central government chooses random inspection
π_{11}	The expected payoffs when upstream governments choose to protect water environment
π_{12}	The expected payoffs when upstream governments choose not to protect water environment
π_{21}	The expected payoffs when downstream governments choose ecological compensation
π_{22}	The expected payoffs when downstream governments choose nonecological compensation
π_{31}	The expected payoffs when the central government chooses daily supervision
π_{32}	The expected payoffs when the central government chooses random inspection
$\overline{\pi_1}$	The average payoffs of upstream governments
$\overline{\pi_2}$	The average payoffs of downstream governments
$\overline{\pi_3}$	The average payoffs of the central government

Similarly, based on π_{21}, π_{22}, and $\overline{\pi_2}$, the replicated dynamic equation of downstream governments is as follows:

$$\begin{aligned} F_2(y) &= \frac{\mathrm{d}y}{\mathrm{d}t} \\ &= y \times (y-1) \times (H_1 - F_2 \times z). \end{aligned} \tag{25}$$

For the central government, the payoffs are

$$\begin{aligned} \pi_{31} &= x \times y \times (V_3 - C_2 - H_2) + (1-x) \times y \times (F_1 - C_2) + x \times (1-y) \\ &\quad \times (V_3 - C_2 - H_2 + F_2) + (1-x) \times (1-y) \times (F_1 + F_2 - C_2), \\ \pi_{32} &= x \times y \times H_2 + (1-x) \times z \times 0 + x \times (1-y) \times H_2 + (1-x) \times (1-y) \times 0, \\ \overline{\pi_3} &= z \times \pi_{31} + (1-z) \times \pi_{32}. \end{aligned} \tag{26}$$

Similarly, based on π_{31}, π_{32}, and $\overline{\pi_3}$, the replicated dynamic equation of the central government is as follows:

$$\begin{aligned} F_3(z) &= \frac{\mathrm{d}z}{\mathrm{d}t} \\ &= z \times (z-1) \\ &\quad \times (C_2 - F_1 - F_2 + F_1 \times x + F_2 \times y + H_2 \times x). \end{aligned} \tag{27}$$

The replication dynamic system can be obtained by simultaneous equations (23), (25), and (27), as shown in the following equation:

$$\begin{cases} F_1(x) = \dfrac{\mathrm{d}x}{\mathrm{d}t} = -x \times (x-1) \times (V_1 - C_1 + F_1 \times z + H_2 \times z), \\ F_2(y) = \dfrac{\mathrm{d}y}{\mathrm{d}t} = y \times (y-1) \times (H_1 - F_2 \times z), \\ F_3(z) = \dfrac{\mathrm{d}z}{\mathrm{d}t} = z \times (z-1) \times (C_2 - F_1 - F_2 + F_1 \times x + F_2 \times y + H_2 \times x). \end{cases} \tag{28}$$

3.4. Evolutionary Stability Strategies. Because each player needs to play games for many times in order to achieve ESS, the initial probabilities change with time; that is, x, y, and z are functions of time, so they are represented by $x(t)$, $y(t)$, and $z(t)$. In order to obtain ESS of the replicated dynamic system, we make the differential equation (28) equal to 0, respectively. Because $x(t)$, $y(t)$, and $z(t) \in [0,1]$, the space for the solution of the replication dynamic system is a cube with side length of 1. By solving the above differential equations, nine equilibrium points ($E_1(0,0,0)$, $E_2(1,0,0)$, $E_3(0,1,0)$, $E_4(0,0,1)$, $E_5(1,1,0)$, $E_6(1,0,1)$, $E_7(0,1,1)$, $E_8(1,1,1)$, and $E_9(x^*, y^*)$) can be obtained. $E_9(x^*, y^*)$ can be obtained by solving

$$\begin{cases} V_1 - C_1 + F_1 \times z + H_2 \times z = 0, \\ H_1 - F_2 \times z = 0, \\ C_2 - F_1 - F_2 + F_1 \times x + F_2 \times y + H_2 \times x = 0. \end{cases} \tag{29}$$

Although nine equilibrium points are obtained in this paper, it is uncertain whether they are ESS. Since ESS must be a pure strategy Nash equilibrium, E_9 must not be ESS, because E_9 represents a mixed strategy equilibrium. Then, we only need to judge whether the other eight points are ESS. According to Lyapunov's System Stability Theory [38, 39], the eigenvalues of the coefficient matrix of dynamic system can help to judge the stability of the system. When all eigenvalues are negative, the system reaches a stable state; when the eigenvalue is nonnegative, the system is in an unstable state. Therefore, firstly, this paper calculates the Jacobian matrix of the replicated dynamic system; secondly, the coordinates from E_1 to E_8 are substituted into the Jacobian matrix to obtain the corresponding eigenvalues. Finally, ESS is determined according to Lyapunov's stability theory. The Jacobian matrix with each equilibrium point can be obtained (J_1 to J_8), as shown in the Appendix.

According to J_1 to J_8, a_{22} of J_3 and J_5, and a_{33} of J_8 are positive (a_{lr} refers to the eigenvalue shown in the lth row and the rth column of the matrix), so E_3, E_5, and E_8 are not ESS. Other equilibrium points cannot directly judge whether they are ESS but need to introduce constraints for further judgment. In contrast, only E_6 is most in line with the actual situation, because the WEC fee is mainly paid by the central government, and most of the downstream governments do not actively participate in the WEC. Therefore, this paper takes E_6 as the object for in-depth research. In order to make the system reach this stable state, the constraints (equation (30)) are added.

$$\begin{cases} C_1 - H_2 - V_1 < 0, \\ F_2 - H_1 < 0, \\ C_2 - F_2 + H_2 < 0. \end{cases} \tag{30}$$

4. Results of Simulation

According to the above analysis, the final stability of replication dynamic system is $E_6(1,0,1)$. Under equation (30), this paper uses numerical simulation technology to study decision-making of each player and the influencing factors. The initial value of each parameter is shown in Table 5.

4.1. Initial Probabilities. Since the initial probability of each player is the function of time, in order to analyze its impact on the decision-making process of each player, this paper divides the initial probability into high probability group and low probability group for comparative analysis. The results are shown in Figure 3.

In Figure 3, we can find that, whether in the high probability group or in the low probability group, the system is finally stable at E_6 (1, 0, 1); that is, upstream governments choose to protect water environment, downstream governments choose not to carry out ecological compensation, and the central government chooses daily supervision. In addition, it can be observed that the time for upstream governments and downstream governments to reach a stable state is shorter, while the time for the central government to reach a stable state is longer. In the high probability group, the initial probability of the central government has experienced a change process of first decreasing and then increasing, which indicates that the central government initially tends to choose random inspection and then turns to daily supervision.

4.2. Environmental Protection Cost. In order to reflect the influence of environmental production cost (C_1) on each player, this section regards it as a variable and takes values of 1, 3, and 5 respectively, while other parameters are constants (as shown in Table 6). Under this setting, this paper sets both high probability group and low probability group to simulate the system, and the results are shown in Figure 4.

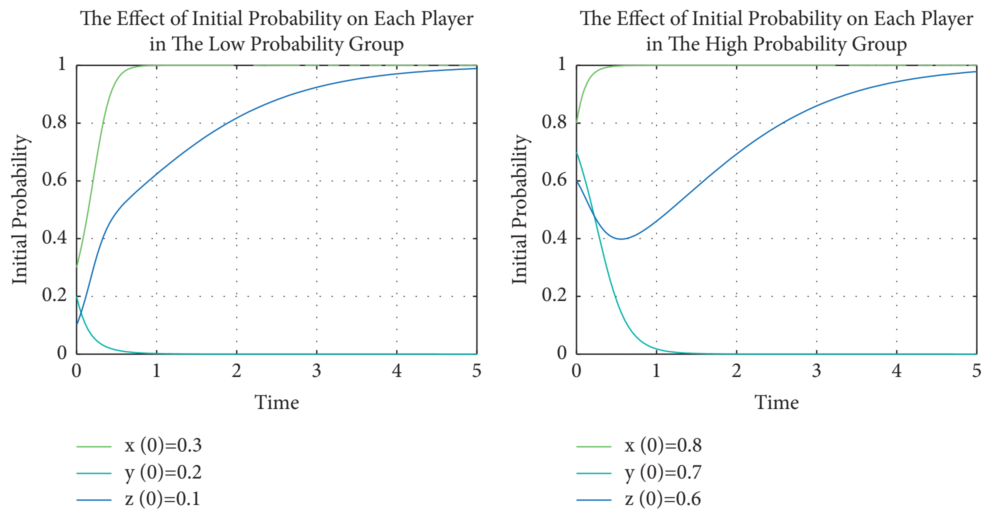
According to Figure 4, it can be found that the environmental protection cost has an impact on upstream governments and the central government in the low probability group (the initial probability is 0.4) but has little impact on downstream governments. In Figures 4(a) and 4(c), the impact of environmental protection cost on upstream governments and the central government is opposite. For upstream governments, the increase of environmental protection cost will prolong the time to reach a stable state. For the central government, the increase of environmental protection cost shortens the time to reach a stable state in the low probability group.

4.3. Fine. In order to illustrate the impact of fine, this paper takes fine (F_1 and F_2) as variables, respectively, with values of 5, 7, and 9, and sets other parameters as constants (as shown in Table 7). On this basis, this paper still carries out system simulation for high initial probability group and low initial probability group, and the results are shown in Figure 5.

According to Figure 5(a), the fine on upstream governments (F_1) has a certain effect on themselves. Whether the initial probability is high or low, when the fine increases, the time for upstream governments to reach stable state 1 is shortened. In Figure 5(c), F_1 only affects the decision-

Table 5: The initial values of parameters.

Parameter	C_1	C_2	F_1	F_2	H_1	H_2	V_1	V_2	V_3
Value	3	2	7	7	8	4	7	10	13

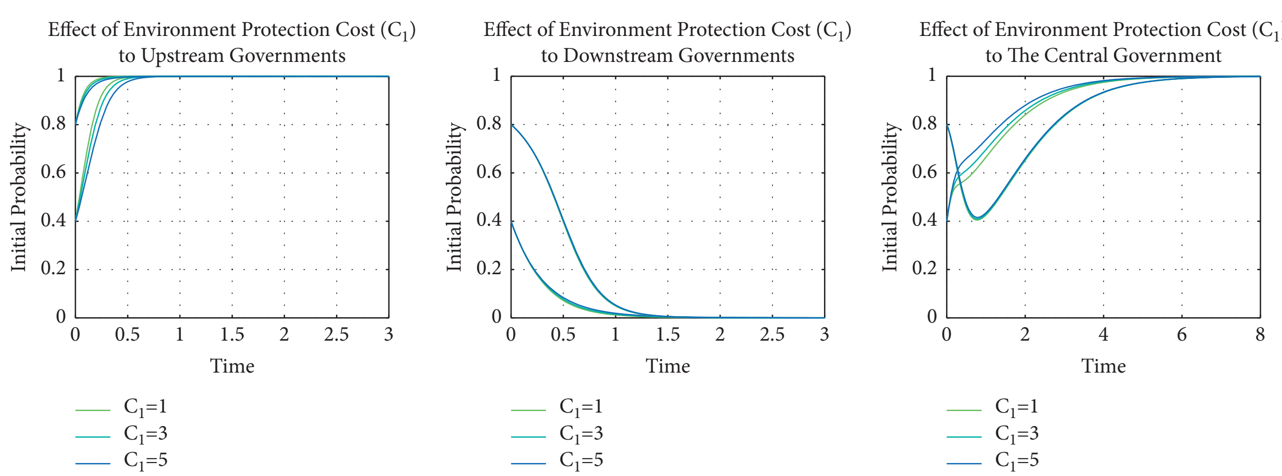


(a) (b)

Figure 3: The effect of initial probability on each player in two groups. (a) The effect of initial probability on each player in the low probability group. (b) The effect of initial probability on each player in the high probability group.

Table 6: The values of parameters in the simulation of environmental protection cost (C_1).

Parameter	C_2	F_1	F_2	H_1	H_2	V_1	V_2	V_3
Value	2	7	7	8	4	7	10	13



(a) (b) (c)

Figure 4: The effect of environmental protection cost (C_1) on each player. (a) Effect of environment protection cost (C_1) on upstream governments. (b) Effect of environment protection cost (C_1) on downstream governments. (c) Effect of environment protection cost (C_1) on the central government.

making of the central government in the low initial probability group; that is, when the fine increases, the time for the central government to reach stable state 1 is shortened to a certain extent. In contrast, F_1 has almost no impact on downstream governments. Similarly, according to Figures 5(d)–5(f), F_2 has little impact on upstream

Table 7: The values of parameters in the simulation of fine (F_1 and F_2).

Parameter	C_1	C_2	H_1	H_2	V_1	V_2	V_3
Value	3	2	8	4	7	10	13

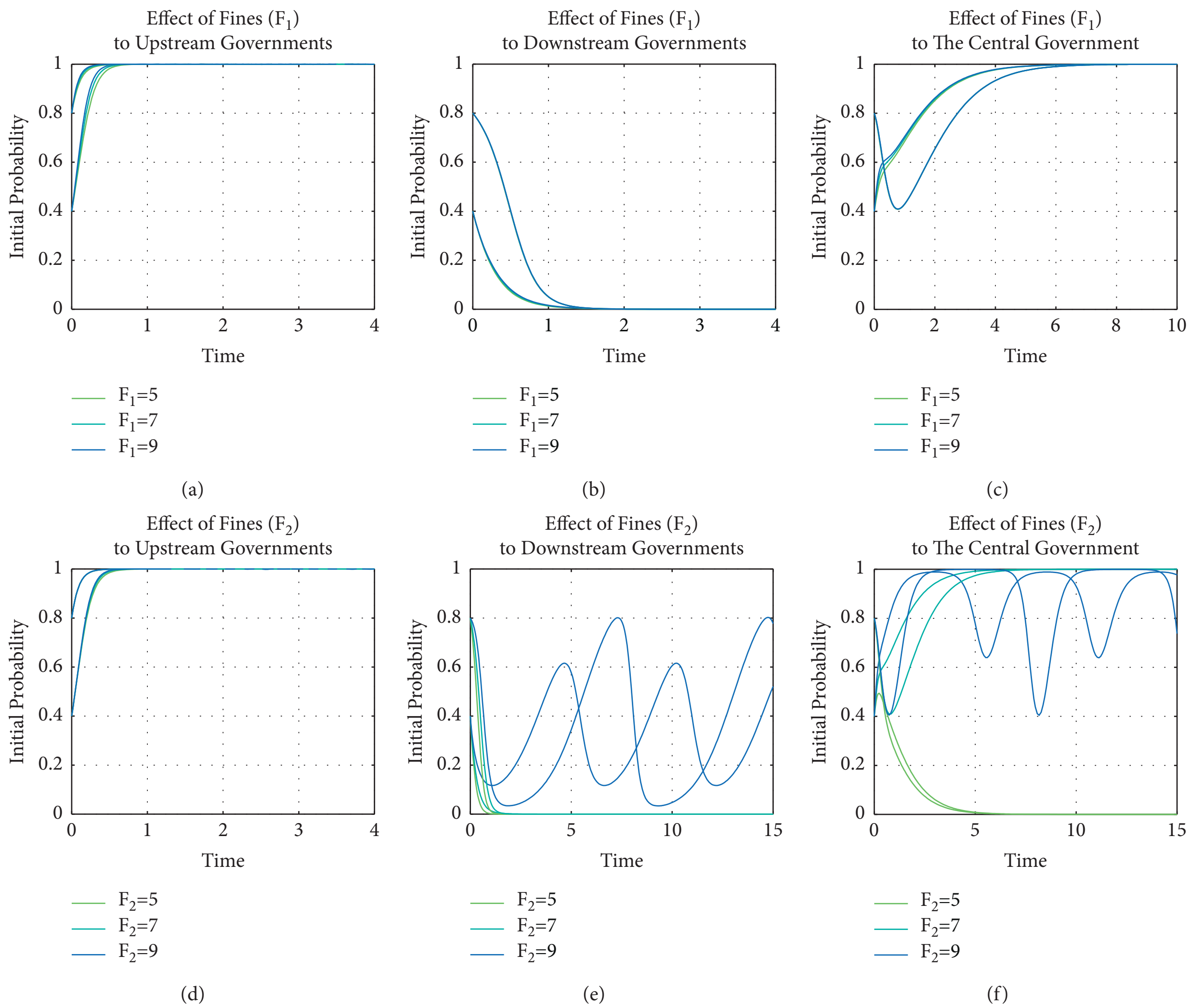


Figure 5: The effect of fines (F_1 and F_2) on each player. (a) Effect of fines (F_1) on upstream governments. (b) Effect of fines (F_1) on downstream governments. (c) Effect of fines (F_1) on the central government. (d) Effect of fines (F_2) on upstream governments. (e) Effect of fines (F_2) on downstream governments. (f) Effect of fines (F_2) on the central government.

governments and has a great impact on downstream governments and the central government. When F_2 changes from 5 to 7, the time for downstream governments to reach stable state 0 increases. When F_2 equals 9, the decision-making of downstream governments fluctuates periodically and finds it difficult to reach a stable state. In Figure 5(f), when F_2 changes from 5 to 7, the central government changes from state 0 to state 1. When F_2 increases to 9, the decision-making of the central government also shows periodic changes.

4.4. Ecological Compensation Paid by Downstream Governments. In this section, ecological compensation paid by downstream governments is regarded as a variable and taken as 6, 8, and 10, respectively, while other parameters are taken as constants (as shown in Table 8). In addition, the system is simulated in high initial probability and low initial probability groups, respectively, and the results are shown in Figure 6.

As can be seen from Figure 6(a), the ecological compensation paid by downstream governments (H_1) has little

Table 8: The values of parameters in the simulation of ecological compensation paid by downstream governments (H_1).

Parameter	C_1	C_2	F_1	F_2	H_2	V_1	V_2	V_3
Value	3	2	7	7	4	7	10	13

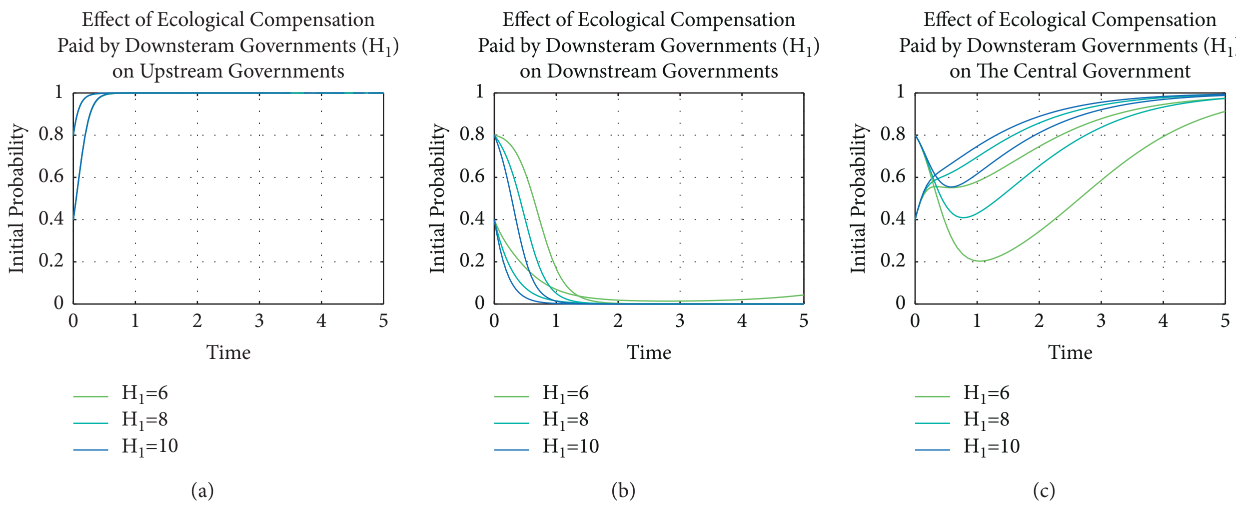


Figure 6: The effect of ecological compensation paid by downstream governments (H_1) on each player. (a) Effect of ecological compensation paid by downstream governments (H_1) on upstream governments. (b) Effect of ecological compensation paid by downstream governments (H_1) on downstream governments. (c) Effect of ecological compensation paid by downstream governments (H_1) on the central government.

impact on upstream governments. In contrast, in Figures 6(b) and 6(c), when ecological compensation paid by downstream governments increases, the speed of downstream governments and the central government reaching a stable state is significantly shortened. At this time, downstream governments will choose no compensation, while the central government will choose daily supervision.

4.5. Ecological Compensation Paid by the Central Government. In this section, ecological compensation paid by the central government (H_2) is regarded as a variable and taken as 2, 4, and 6 respectively, while other parameters are taken as constants (as shown in Table 9). In addition, the system is simulated in high initial probability and low initial probability groups, respectively, and the results are shown in Figure 7.

From Figure 7, ecological compensation paid by the central government has little impact on upstream governments but has a greater impact on downstream governments and the central government. In Figure 7(b), when H_2 increases, the time for downstream governments to reach stable state 0 is shortened. In Figure 7(c), when H_2 changes from 2 to 4, the time for the central government to reach stable state 1 is extended. However, when H_2 increases to 6, the stable state of the central government changes from state 1 to state 0 and stabilizes at state 0. Because the stable state of the central government has changed greatly, this paper further simulates H_2, taking values from 4 to 5.4, and the step value is 0.2. In Figure 7(d), when H_2 is equal to 5, the decision-making curve of the central government becomes a

Table 9: The values of parameters in the simulation of the ecological compensation paid by the central government (H_2).

Parameter	C_1	C_2	F_1	F_2	H_1	V_1	V_2	V_3
Value	3	2	7	7	4	7	10	13

horizontal straight line, which shows that 5 is the critical value of the change of the central government's strategy.

5. Discussion

5.1. Initial Probabilities. From Section 3.4, it can be seen that the initial probability is the key factor for determining ESS. Therefore, the simulation of the initial probability will help to understand the sensitivity of each player's decision. According to the results in Section 4.1, in different initial probability groups, the final stable state of each player is the same; that is, the upstream governments choose to protect water environment, the downstream governments choose not to compensate, and the central government chooses daily supervision. However, there are some differences in the trend of each player reaching a stable state. For upstream governments, the time to reach a stable state in the high probability group is significantly shorter than that in the low probability group. For downstream governments, the time to reach the steady state in the low probability group is shorter than that in the high probability group. For the central government, the times to reach a stable state are close in different groups, but, in the high probability group, the curve shows a downward trend first and then an upward

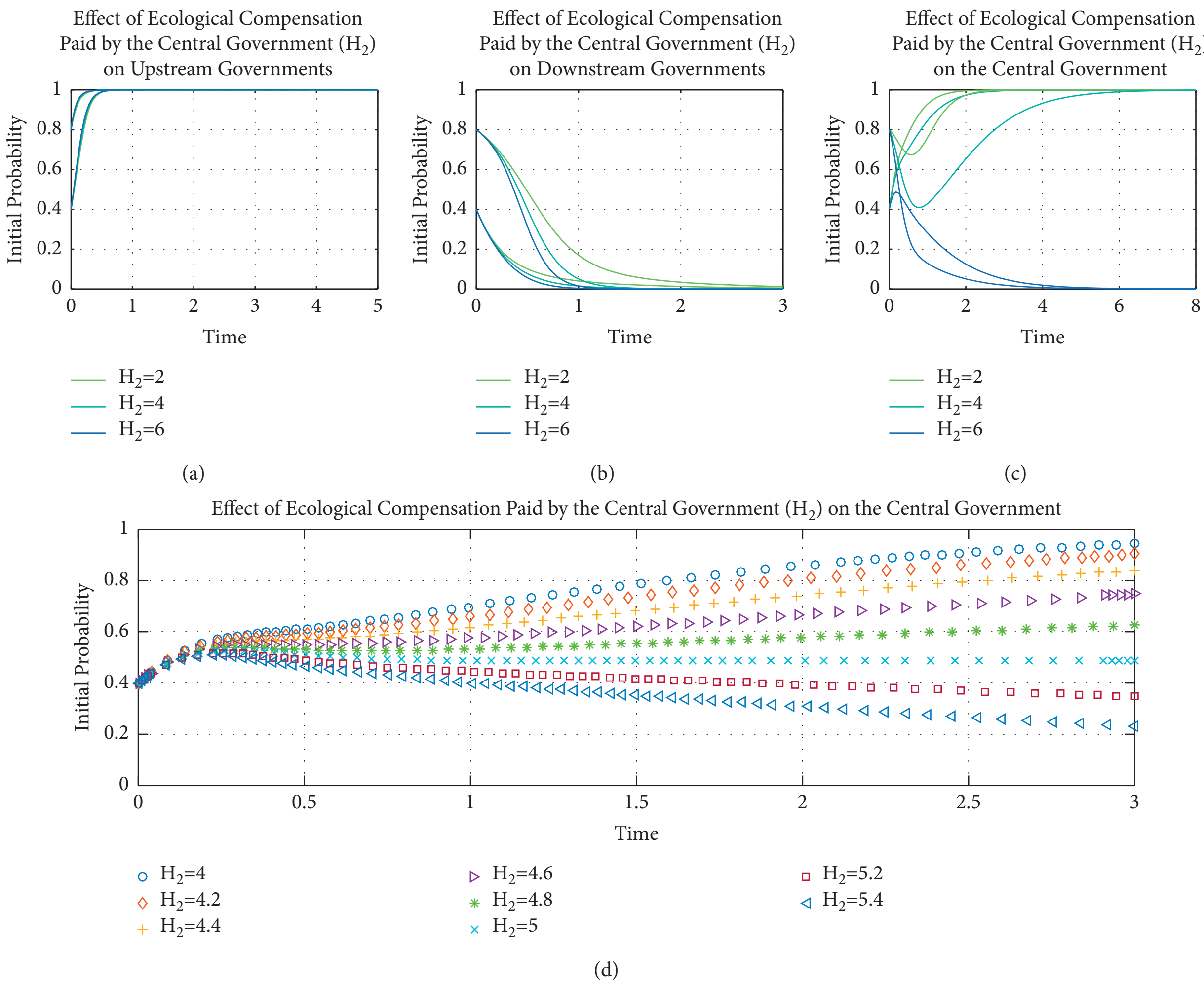


Figure 7: The effect of the ecological compensation paid by the central government (H_2) on each player. (a) Effect of the ecological compensation paid by the central government (H_2) on upstream governments. (b) Effect of the ecological compensation paid by the central government (H_2) on downstream governments. (c) Effect of the ecological compensation paid by the central government (H_2) on the central government. (d) Effect of the ecological compensation paid by the central government (H_2) on the central government.

trend. It can be seen that upstream governments have strong willingness to participate in WEC, downstream governments have weak willingness to participate, and the central government chooses daily supervision. At present, the water resources protection funds of upstream governments include the transfer payment of the central government and the self-raised funds of upstream governments. Due to the slow economic development and weak financing capacity of upstream governments, the water resources protection funds are mainly borne by the central government. This reflects that the vertical compensation is the mainstream of WEC in China. At this stage, the central government hopes to build horizontal compensation between upstream and downstream governments to supplement vertical compensation, but, from the simulation results, the downstream governments are the key to establish horizontal compensation, and their willingness to participate is not high. Therefore, downstream governments need to fully understand the contribution of upstream governments protecting water resources and improve their willingness to participate in WEC, which will change the current situation of the lack of horizontal compensation.

5.2. Environmental Protection Cost. The establishment and implementation of WEC need the cooperation of governments at all levels, and upstream governments need to invest much direct cost and opportunity cost in protecting water environment. Therefore, this cost is the key factor to WEC. According to the analysis in Section 4.2, the higher the environmental protection cost is, the longer upstream governments take to reach the stable state. Downstream governments are hardly affected. In the low probability group, the central government reaches a stable state faster with the increase of environmental protection cost. The reasons for the above changes are as follows: first, when protecting the water environment, upstream governments not only have to pay a lot of costs for the construction of environmental protection infrastructure but also need to implement other measures to maintain the protection effect of water environment, such as limiting the expansion of polluting enterprises and changing the planting structure of local agriculture. Therefore, the higher the cost paid by upstream governments is, the more difficult it is to implement sustainable water environmental protection. Second, due to the continuous increase of environmental costs, upstream governments have the motivation to choose not to protect water environment. In order to ensure the health of water environment in the basin, the central government has to request upstream governments to protect water environment by daily supervision and paying certain compensation. Therefore, effectively reducing the environmental protection cost of upstream governments is the key to promote the sustainable protection of water resources. At present, the Chinese government is reducing the cost of water environmental protection through the following ways: (1) Introduce new green technologies to help upstream governments improve the efficiency of sewage treatment. For example, the central government selects sewage treatment companies with technology and experience through investment to assist upstream governments in building sewage treatment plants. (2) Establish enclaves in the downstream areas and relocate high polluting enterprises in the upstream areas to enclaves and give them corresponding tax relief. On the one hand, it can ensure the normal operation of these polluting enterprises, and, on the other hand, it can reduce the amount of local sewage.

5.3. Fine. The central government's fine to upstream and downstream governments is one of the environmental regulation ways to promote them participating WEC. According to the simulation results in Section 4.3, the fines imposed on upstream governments have no impact on players, while the fines imposed on downstream governments have a great impact on the downstream governments and the central government. When the fine increases to a certain degree, the decision-makings of downstream governments and the central government change periodically. The reasons for the above phenomena are as follows: firstly, the fine will increase the additional cost of downstream governments and increase the income of the central government. Therefore, the decision-making changes between downstream governments and the central government have strong relevance. Secondly, when upstream governments choose to protect water environment, downstream governments try to "free ride" without paying compensation to upstream governments, because it can not only obtain high-quality water resources but also pay no price. However, to ensure the continuous protection of water environment by upstream governments, the central government will adopt daily supervision to urge downstream governments to pay certain compensation fee to upstream governments. At this time, in order to deal with daily supervision of the central government, downstream governments will have the tendency to pay compensation. Therefore, the downstream governments will weigh the impact of "free riding" and "daily supervision," so their decision-making curve fluctuates periodically. Finally, the central government will make corresponding adjustments to the strategic changes of downstream governments and finally form periodic changes. However, the central government is more sensitive to the fine of downstream governments, because when F_2 increases to 9, the central government's strategy curve changes periodically, which can be seen from Figure 5(f). According to the above analysis, the fine of the central government is indispensable for the establishment of WEC mechanism. If downstream governments are sensitive to fine, they can choose compensation by establishing a strict punishment mechanism. However, at present, there is no legislation on ecological compensation in China, so the horizontal compensation between upstream and downstream governments is mainly established through consultation, while the central government only guides. This makes it impossible to punish downstream governments for their no compensation. On September 12, 2021, General Office of the State Council of the PRC issued the opinions on deepening the reform of ecological protection compensation system, which clearly emphasized the use of legal means to regulate ecological protection compensation. Therefore, WEC law will be conducive to the establishment of punishment mechanism, which will promote downstream governments to participate in the construction of WEC.

5.4. Ecological Compensation Fee. Ecological compensation fee is the core to ensure the long-term implementation of WEC. Due to the implementation of WEC, both the downstream governments and the central government will benefit, so how to determine the proportion of compensation funds is the key problem.

According to the simulation results in Sections 4.4 and 4.5, the compensation paid has little impact on upstream governments but has a greater impact on downstream governments and the central government. First, when compensation fee increases, upstream governments are almost unaffected, which is different from common sense because compensation fee is an additional benefit to upstream governments; when it increases, upstream governments should reach a stable state faster, but this is not reflected in the simulation. This shows that upstream governments need to pay huge costs for protecting water environment, which cannot be compensated by compensation fee. However, due to the policy pressure of the central government, upstream governments have to protect water environment. Therefore, upstream governments are not sensitive to limited compensation fee. Second, compensation fee is an additional expenditure for downstream governments. When compensation fee increases, downstream governments will choose not to compensate, which is the same as the simulation results. Finally, when compensation fee paid by downstream governments (H_1) increases, the central government will reach a stable state faster (Figure 6(c)), which is also different from common sense, because if downstream governments pay compensation fee in time, this can form a virtuous cycle of protecting water resources between upstream and downstream areas, so the central government no longer needs daily supervision. However, in the practice of WEC, downstream governments believe that using high-quality water resources is their own right, so they are unwilling to pay compensation fee and continue to choose no-compensation strategy (Figure 6(b)). Then, the central government can only continue to adopt daily supervision to request upstream governments protecting water environment. In addition, when compensation paid by the central government (H_2) increases from 2 to 4, the time to reach stable state 1 will be extended. When H_2 further increases to 6, the central government will change from daily supervision to random inspection. This shows that the central government has an upper limit on compensation fee (Figure 7(d)). Once the compensation fee exceeds the upper limit, the central government will change the strategy. This is consistent with the reality of WEC. At present, the compensation fee paid by the central government is the main source of funds to maintain WEC, which will bring huge financial pressure to itself. Therefore, the central government is also stepping up its research on how to promote the spontaneous construction of ecological compensation mechanism upstream and downstream of the basin, so as to free it from the huge pressure of compensation fee. At present, the Chinese government has carried out horizontal compensation pilot projects in Xin'an River, Chishui River, and other river basins and has gained a lot of practical experience. In addition, it has successively issued guidance on supporting the establishment of horizontal compensation in large river basins such as Yangtze River and Yellow River. Attracting social capital (such as enterprises and the public) to join horizontal compensation is an effective way to increase compensation funds. Improving the trading of water rights and emission rights between upstream and downstream areas is also an important way to effectively promote horizontal compensation.

6. Conclusion and Suggestion

6.1. Conclusion. This paper takes upstream governments, downstream governments, and the central government as the research object and analyzes the changes in the decision-making process and key influencing factors. This is of great significance to the rational design and implementation of WEC. Our results show that the initial probability significantly affects the decision-making behavior of each player. Daily supervision of the central government and the cost of water environmental protection play an important role in the establishment of WEC. The fine imposed on the downstream governments will lead to periodic changes in the decision-making of themselves and the central government. There is a critical value for the ecological compensation paid by the central government, which means that reasonably determining the matching relationship between vertical compensation and horizontal compensation can effectively promote the establishment and long-term implementation of WEC. Although this paper has obtained some interesting results, there are still the following limitations: (1) this paper does not include polluting enterprises and the public into the analysis framework; (2) this paper uses simulation technology to study the decision-making process of each subject in WEC but does not introduce actual cases for comparative analysis. In the future, we will use multiagent simulation to include stakeholders such as polluting enterprises and residents in the research framework. Meanwhile, we will obtain the data of actual WEC case through questionnaire survey combined with econometrics and statistical analysis to further verify the theoretical model and conclusions in this paper.

6.2. Suggestion. According to this paper, the following suggestions are put forward: (1) at this stage, the central government should continue to implement daily supervision to ensure the implementation of WEC; (2) improving the initial willingness of governments at all levels to participate in WEC significantly improves its success rate; (3) upstream governments should study new methods and paths to reduce the cost of protecting water environment; (4) the central government should strengthen the supervision and fine on downstream governments; (5) the central government should build a new scheme to promote upstream and downstream governments to spontaneously establish a horizontal compensation mechanism. The implementation of the above suggestions will effectively promote the establishment and implementation of WEC.

Appendix

The Jacobian matrices of eight equilibrium points are as follows:

$$J_1 = \begin{bmatrix} V_1 - C_1 & 0 & 0 \\ 0 & -H_1 & 0 \\ 0 & 0 & F_1 - C_2 + F_2 \end{bmatrix},$$
$$J_2 = \begin{bmatrix} C_1 - V_1 & 0 & 0 \\ 0 & -H_1 & 0 \\ 0 & 0 & F_2 - C_2 - H_2 \end{bmatrix},$$
$$J_3 = \begin{bmatrix} V_1 - C_1 & 0 & 0 \\ 0 & H_1 & 0 \\ 0 & 0 & F_1 - C_2 \end{bmatrix},$$
$$J_4 = \begin{bmatrix} F_1 - C_1 + H_2 + V_1 & 0 & 0 \\ 0 & F_2 - H_1 & 0 \\ 0 & 0 & C_2 - F_1 - F_2 \end{bmatrix},$$
$$J_5 = \begin{bmatrix} C_1 - V_1 & 0 & 0 \\ 0 & H_1 & 0 \\ 0 & 0 & -C_2 - H_2 \end{bmatrix},$$
$$J_6 = \begin{bmatrix} C_1 - F_1 - H_2 - V_1 & 0 & 0 \\ 0 & F_2 - H_1 & 0 \\ 0 & 0 & C_2 - F_2 + H_2 \end{bmatrix},$$
$$J_7 = \begin{bmatrix} F_1 - C_1 + H_2 + V_1 & 0 & 0 \\ 0 & -F_2 + H_1 & 0 \\ 0 & 0 & C_2 - F_1 \end{bmatrix},$$
$$J_8 = \begin{bmatrix} C_1 - F_1 - H_2 - V_1 & 0 & 0 \\ 0 & -F_2 + H_1 & 0 \\ 0 & 0 & C_2 + H_2 \end{bmatrix}. \tag{A.1}$$

Data Availability

All simulation data used in this paper are listed in the manuscript.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was funded by the Water Conservancy Science and Technology Projects in Jiangsu Province (no. 2018034).

References

[1] N. Li, G. Tang, P. Zhao, Y. Hong, Y. Gou, and K. Yang, "Statistical assessment and hydrological utility of the latest multi-satellite precipitation analysis IMERG in Ganjiang river basin," *Atmospheric Research*, vol. 183, pp. 212–223, 2017.

[2] M. Chen, Y. Lu, L. Ling, Y. Wan, Z. Luo, and H. Huang, "Drivers of changes in ecosystem service values in Ganjiang upstream watershed," *Land Use Policy*, vol. 47, pp. 247–252, 2015.

[3] S. Jichuan and W. Michael, "Incentive coordination for transboundary water pollution control: the case of the middle route of China's South-North water transfer project - sciencedirect," *Journal of Hydrology*, vol. 598, Article ID 125705, 2020.

[4] J. Osán, S. Török, B. Alföldy, and G. Falkenberg, "Characterization of anthropogenic sediment particles after a transboundary water pollution of river Tisza using synchrotron radiation," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 59, no. 5, pp. 701–708, 2004.

[5] X. Gao, J. Shen, W. He et al., "Multilevel governments' decision-making process and its influencing factors in watershed ecological compensation," *Sustainability*, vol. 11, pp. 1–28, 2019.

[6] H. Li, J. Lu, and P. Research, "Can regional integration control transboundary water pollution? A test from the Yangtze river economic belt," *Environmental Science and Pollution Research*, vol. 27, no. 22, pp. 28288–28305, 2020.

[7] Q. Wang, Q. Fu, Z. Shi, and X. Yang, "Transboundary water pollution and promotion incentives in China," *Journal of Cleaner Production*, vol. 261, Article ID 121120, 2020.

[8] J. Lu and P. Research, "Turnover of environmental protection officials and transboundary water pollution," *Environmental Science and Pollution Research*, vol. 28, pp. 1–17, 2020.

[9] L. Zhao, C. Li, R. Huang et al., "Harmonizing model with transfer tax on water pollution across regional boundaries in a China's lake basin," *European Journal of Operational Research*, vol. 225, no. 2, pp. 377–382, 2013.

[10] J. Shen, X. Gao, W. He et al., "Prospect theory in an evolutionary game: construction of watershed ecological compensation system in Taihu lake basin," *Journal of Cleaner Production*, vol. 291, Article ID 125929, 2021.

[11] X. Guan, W. Liu, M. Chen, J. Ecolind, T. Control, and E. C. Standard, "Study on the ecological compensation standard for river basin water environment based on total pollutants control," *Ecological Indicators*, vol. 69, pp. 446–452, 2016.

[12] N. Grima, S. J. Singh, B. Smetschka, and L. Ringhofer, "Payment for ecosystem services (PES) in Latin America: analysing the performance of 40 case studies," *Ecosystem Services*, vol. 17, pp. 24–32, 2016.

[13] X. Kai, K. Fanbin, Z. Ning, N. Lei, and C. Sun, "Analysis of the factors influencing willingness to pay and payout level for ecological environment improvement of the ganjiang river basin," *Sustainability*, vol. 10, p. 2149, 2018.

[14] X. Pan, L. Xu, Z. Yang, and Y. Bing, "Payments for ecosystem services in China: policy, practice, and progress," *Journal of Cleaner Production*, vol. 21, pp. 109–119, 2017.

[15] Mo.Fo.t.Ps.Ro China, "Guiding opinions on establishing and perfecting long-term mechanism of ecological compensation and protection in the Yangtze river economic zone," Availabe online: http://ln.mof.gov.cn/lanmudaohang/zhengcefagui/201802/t20180228_2822651.html (accessed on October, 6), 2018.

[16] Z. Zhang, J. Gao, X. Fan, Y. Lan, and M. Zhao, "Response of ecosystem services to socioeconomic development in the Yangtze river basin, China," *Ecological Indicators*, vol. 72, pp. 481–493, 2017.

[17] China, "Mo.Fo.t.Ps.Ro. guiding opinions on establishing and perfecting the long-term mechanism of ecological compensation and protection in the Yangtze river economic zone," Availabe online: http://www.gov.cn/xinwen/2018-02/24/content_5268509.htm (accessed on October 6), 2018.

[18] X. Gao, J. Shen, W. He et al., "An evolutionary game analysis of governments' decision-making behaviors and factors

influencing watershed ecological compensation in China," *Journal of Environmental Management*, vol. 251, Article ID 109592, 2019.

[19] D. Feng, W. Wu, L. Liang, L. Li, and G. Zhao, "Payments for watershed ecosystem services: mechanism, progress and challenges," *Ecosystem Health and Sustainability*, vol. 4, no. 1, pp. 13–28, 2018.

[20] Y. J. Wang, G. X. Bing, Q. L. Xiao, Y. L. Hong, and Y. Jin, "Ecological compensation standards and compensation methods of public welfare forest protected area," *Chinese Journal of Applied Ecology*, vol. 27, pp. 1893–1900, 2016.

[21] W. Sheng, L. Zhen, G. Xie, and Y. Xiao, "Determining eco-compensation standards based on the ecosystem services value of the mountain ecological forests in Beijing, China," *Ecosystem Services*, vol. 26, pp. 422–430, 2017.

[22] X. Gao, J. Shen, W. He et al., "Changes in ecosystem services value and establishment of watershed ecological compensation standards," *International Journal of Environmental Research and Public Health*, vol. 16, no. 16, p. 2951, 2019.

[23] B. Yu, L. Xu, and S. E. Reviews, "Review of ecological compensation in hydropower development," *Renewable and Sustainable Energy Reviews*, vol. 55, pp. 729–738, 2016.

[24] K. Johst, M. Drechsler, and F. Wätzold, "An ecological-economic modelling procedure to design compensation payments for the efficient spatio-temporal allocation of species protection measures," *Ecological Economics*, vol. 41, no. 1, pp. 37–49, 2002.

[25] Y. M. Jie, G. E. Yan-Xiang, X. U. J. O. R. Guang-Li, and M. Science, "Cooperation evolution of ecological compensation between waterhead region and downstream based on evolutionary game," *Operations Research and Management Science*, vol. 21, pp. 137–143, 2012.

[26] H. Xie, W. Wang, and X. Zhang, "Evolutionary game and simulation of management strategies of fallow cultivated land: a case study in Hunan province, China," *Land Use Policy*, vol. 71, pp. 86–97, 2018.

[27] R. Moreno-Sanchez, J. H. Maldonado, S. Wunder, and C. Borda-Almanza, "Heterogeneous users and willingness to pay in an ongoing payment for watershed protection initiative in the Colombian Andes," *Ecological Economics*, vol. 75, pp. 126–134, 2012.

[28] G. Van Hecken, J. Bastiaensen, and W. F. Vásquez, "The viability of local payments for watershed services: e," *Ecological Economics*, vol. 74, pp. 169–176, 2012.

[29] X. Zhang, X. Gao, and I. Amp, "Research on intergovernmental watershed ecological compensation mechanism based on game theory," *Value Engineering*, vol. 36, pp. 110–112, 2018.

[30] Y. Zhao and S. Dong, "Research on ecological compensation in the Yellow River basin based on the perspective of game theory," *E3S Web of Conferences*, vol. 275, Article ID 02002, 2021.

[31] H. U. Dongbin, H. Liu, X. Chen, and Y. Chen, "Research on the ecological compensation standard of the basin pollution control project based on evolutionary game theory and by taking Xiangjiang river as a case," *Frontiers of Engineering Management*, vol. 6, pp. 575–583, 2019.

[32] D. Friedman, "On economic applications of evolutionary game theory," *Journal of Evolutionary Economics*, vol. 8, no. 1, pp. 15–43, 1998.

[33] K. Ritzberger and J. W. Weibull, "Evolutionary selection in normal-form games," *Econometrica*, vol. 63, no. 6, pp. 1371–1399, 1995.

[34] C. Schmidt, "Are evolutionary games another way of thinking about game theory?" *Journal of Evolutionary Economics*, vol. 14, no. 2, pp. 249–262, 2004.

[35] R. Cressman and J. Apaloo, *Evolutionary Game Theory*, MIT Press, Cambridge, MA, USA, 2016.

[36] A. J. Robson, "Individual strategy and social structure: an evolutionary theory of institutions," *The Canadian Journal of Economics/Revue Canadienne D`economique*, vol. 3, pp. 329–331, 1999.

[37] H. Zhang, Y. Ge, and Y. Jie, "Analysis of replicated dynamics and evolutionary stability strategies of watershed ecological compensation," *Statistics & Decisions*, vol. 34, pp. 50–53, 2018.

[38] Bomze, M. Immanuel, Weibull, and W. Jorgen, "Does neutral stability imply Lyapunov stability?" *Games and Economic Behavior*, vol. 11, pp. 173–192, 1995.

[39] S. R. Vadali and E. S. Kim, "Feedback control of tethered satellites using Lyapunov stability theory," *Journal of Guidance, Control, and Dynamics*, vol. 14, pp. 729–735, 2015.

Hindawi
Complexity
Volume 2021, Article ID 7145168, 11 pages
https://doi.org/10.1155/2021/7145168

Research Article

Prediction for Various Drought Classes Using Spatiotemporal Categorical Sequences

Rizwan Niaz,[1] **Mohammed M. A. Almazah**,[2,3] **Xiang Zhang**,[4,5] **Ijaz Hussain**,[1] **and Muhammad Faisal**[6]

[1]*Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan*
[2]*Department of Mathematics, College of Sciences and Arts (Muhyil), King Khalid University, Muhyil 61421, Saudi Arabia*
[3]*Department of Mathematics and Computer, College of Sciences, Ibb University, Ibb 70270, Yemen*
[4]*National Engineering Research Center of Geographic Information System, School of Geography and Information Engineering, China University of Geosciences (Wuhan), Wuhan 430074, China*
[5]*State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing (LIESMARS), Wuhan University, Wuhan 430079, China*
[6]*Faculty of Health Studies, University of Bradford, BD7 1DP, Bradford, UK*

Correspondence should be addressed to Rizwan Niaz; razwanniaz11@gmail.com and Xiang Zhang; zhangxiangsw@whu.edu.cn

Received 26 September 2021; Accepted 12 November 2021; Published 28 November 2021

Academic Editor: Alireza Amirteimoori

Drought frequently spreads across large spatial and time scales and is more complicated than other natural disasters that can damage economic and other natural resources worldwide. However, improved drought monitoring and forecasting techniques can help to minimize the vulnerability of society to drought and its consequent influences. This emphasizes the need for improved drought monitoring tools and assessment techniques that provide information more precisely about drought occurrences. Therefore, this study developed a new method, Model-Based Clustering for Spatio-Temporal Categorical Sequences (MBCSTCS), that uses state selection procedures through finite mixture modeling and model-based clustering. The MBCSTCS uses the functional structure of first-order Markov model components for modeling each data group. In MBCSTCS, the suitable order K of the components is selected by Bayesian information criterion (BIC). In MBCSTCS, the estimated mixing proportions and the posterior probabilities are used to compute probability distribution associated with the future steps of transitions. Furthermore, MBCSTCS predicts drought occurrences in future time using spatiotemporal categorical sequences of various drought classes. The MBCSTCS is applied to the six meteorological stations in the northern area of Pakistan. Moreover, it is found that MBCSTCS provides expeditious information for the long-term spatiotemporal categorical sequences. These findings may be helpful to make plans for early warning systems, water resource management, and drought mitigation policies to decrease the severe effects of drought.

1. Introduction

Drought is relatively more volatile than other natural disasters, and traditional valuations or forecast procedures are failed to predict it. Its relatively unperceptive onset and the multifaceted impacts cause the new assessment methodologies [1–5]. Since last decades, it has become more prominent to distress the environment and economic sectors worldwide than other natural hazards [6–8]. Moreover, determining the onset and end times of the drought is still challenging for drought management. Structurally, the effects of droughts slowly add over a period, and they may linger for a long period [8–10]. However, it can be characterized by a precipitation deficiency, which has substantial impacts on the agriculture, hydrological systems, and on living standards of the people [11, 12]. Despite perceptible effects of drought, these effects acclimatize severity without appropriate measures and are sustained for the long term even after termination [9].

The advancements in drought assessing and monitoring procedures can lead to better drought preparation and decrease the susceptibility of society to drought and its forgoing influences [8, 10, 13]. Therefore, it is essential to find more suitable techniques and procedures to predict drought occurrences more instantaneously. The improved method can be helpful to make plans for the early warning system, drought mitigation policies, and water resource management and decrease the severe effects of drought. Furthermore, the occurrences and characteristics of drought trigger the discussion about the various methodologies and techniques. Generally based on the occurrences and characteristics of the drought, authors have been categorizing the drought into various groups, including "meteorological, hydrological, agricultural, and socioeconomic" [14]. Chang [15] and Eltahir [16] defined that meteorological drought can be occurred due to the shortage of precipitation over a region for some time. Several studies have considered precipitation data to analyze meteorological droughts [17, 18]. The streamflow data have been frequently used for analyzing hydrological drought [19–21]. Furthermore, the reduction in soil moisture usually causes agricultural drought. The reduction in soil moisture can be affected by meteorological and hydrological droughts. Socioeconomic drought is linked to the shortfall in water resource systems, and in this case, the water supply is unable to meet water demands.

In the past few decades, numerous drought indices have been proposed to assess the drought occurrences [22–26]. The drought indices are frequently used to characterize the drought. The indices are based on various parameters that describe the spatial and temporal extents. Obtaining accurate and precise information about drought occurrences using several drought indices is crucial for an early warning policy; however, consistent and eminent drought information plays a crucial part in preparing drought monitoring and mitigating policies. Numerous drought indices with their strengths and weaknesses exist in the literature and are used by decision-makers who build action plans for drought early warning systems and mitigation policies. For example, Palmer [27] developed a drought index named the Palmer Drought Severity Index (PDSI). The PDSI worked well especially for subhumid and semiarid regions. The PDSI provided weekly information related to abnormal evapotranspiration deficit for the various regions. Information obtained from PDSI can be helpful for the crops in the region. The moisture condition of the regions can be assessed. Gommes and Petrassi [28] have proposed the national rainfall index (NRI). The NRI was used to provide synthetic discussion in sub-Saharan countries in Africa. They used NRI to determine the pattern recognition of rainfall in various regions. The Surface-Water Supply Index (SWSI) was introduced by Shafer and Dezman [24]. The computation of the SWSI is based on two major sources of irrigation water supply, namely, spring-summer streamflow runoff and reservoir carryover. Both sources are accumulatively analyzed to determine the total availability of surface water supply in season. Van Rooy [29] developed the Rainfall Anomaly Index (RAI). The RAI helped to find geographical anomalies of the rainfall pattern in varying regions. Weghorst [26] has introduced the Reclamation Drought Index. Palmer [22] has introduced crop moisture index (CMI). Bhalme and Mooley [23] has developed Bhalme and Mooly drought index (BMDI). The BMDI used precipitation data and provided both negative and positive values to measure drought intensities. McKee et al. [25] developed the Standardized Precipitation Index (SPI). The SPI considered the time series of a long-term record of precipitation in the climatic areas. The dynamic characteristic of SPI is that it can be studied for different time scales and used to compare varying climatic areas. Therefore, SPI is being used extensively for evaluating and recording drought characteristics [30–35]. Furthermore, the drought indices that are mentioned above have been used frequently for drought monitoring in the different studies, although having discrepancies among the indices, to gain consistent interpretation across several regimes and spatial climates. This study utilized SPI, which is often employed to assess and monitor meteorological drought and is recommended by the World Meteorological Organization [36].

Furthermore, many clustering techniques are considered in the literature [37–41]. The clustering techniques focus on grouping the data so that the data group with similar characteristics would be selected within the cluster, while distinct information can exist among other clusters. Various clustering techniques have been frequently considered in machine learning approaches, especially in statistics and computer science, due to the variety of their applications [41–45]. Among the various techniques, model-based clustering groups data and presumes that each data cluster can be perceived as a part of any probability distribution [46, 47]. In various data groups, numerous distributions are preferred, and finite mixture models are desired [48]. The performances of the model-based clustering are outstanding in spectrometry data, text classification, social networks, and distinct grouping objects. Model-based clustering is used for time series [49] and regression time series analyses [50]. Several studies related to model-based clustering are available in the literature; however, it has not yet received greater attention in drought analysis. Therefore, this study developed a new technique known as Model-Based Clustering for Spatio-Temporal Categorical Sequences (MBCSTCS) to precisely predict drought occurrences for spatiotemporal categorical sequences. The performance of the proposed technique is assessed by using six meteorological stations in the northern area of Pakistan.

2. Methods

2.1. Standardized Precipitation Index (SPI). The long-term record of precipitation in the climatic area observed in the time sequence can be used to compute SPI. The vital feature of SPI is that it can be considered for various time scales and is being widely used to calculate and record drought occurrences [34, 35, 51, 52]. The analysis with various time scales can provide varying information. For example, the moisture conditions in different seasons can be assessed using SPI at a three-month time scale. The SPI can assess information related to the water deficiency at a twelve-month time scale. Furthermore, the use of SPI describes the best characteristics in forecasting and analyzing risks as a probabilistic approach [31, 35, 53]. The SPI has been frequently used for drought monitoring in several aspects, for example, spatiotemporal analysis, forecasting, frequency analysis, and climatic studies [33, 35, 51, 52]. As precipitation is only used to determine the climatic condition for a particular area, it offers spatially reliable interpretations across various climates [32, 34, 35]. Therefore, it can be advantageous for the areas where other parameters are available that are required to calculate other indices and of significantly great concern to the various environmental and temporal circumstances [54]. This study focuses on the new methodology developed for monitoring drought more precisely and comprehensively in a specific area. The SPI at various time scales (1, 3, 6, 9, 12, and 24) is used for the current analysis.

2.2. Model-Based Clustering for Spatio-Temporal Categorical Sequences (MBCSTCS). Model-based clustering has been used for time series [49] and regression time series analyses [50]. Various studies associated with model-based clustering are available in the literature; the technique has significant importance for many applications; however, it has not yet received greater attention in drought analysis. Furthermore, in drought classification, categorical sequences are required for obtaining reliable results for the drought characterization. In this perspective, this study proposed MBCSTCS to analyze the categorical drought sequences for various time scales and stations. The MBCSTCS provides more significant results by using a categorical grouping of sequences than traditional approaches that have been used for the prediction. The MBCSTCS reflects the steering behavior of drought classes on various time scales and stations. Moreover, the selected drought classes (states) ("(Extremely Dry (ED), Severely Dry (SD), Normal Dry (ND), Median dry (MD), Median Wet (MW), Severely Wet (SW), and extremely Wet (EW)") are considered for the region [55].

Moreover, the first-order Markov model has a rationale in statistical modeling. The MBCSTCS considers the functional shape of first-order Markov model components for each data group. Furthermore, in the MBCSTCS the data groups consist of various sequences of drought states. For example, we let observation $X = (x_1, \ldots, x_m)^T$ that specifies for an ordered sequence, where each of its elements x_J consists of a categorical value that is specified for varying drought states and coded by natural integers. Furthermore, it is assumed that the number of unique drought states equals p, i.e., $x_J \in \{1, 2, \ldots p\}$ for $j = 1, 2, \ldots m$. Moreover, using a joint probability expression it can be written as $P(X = x) = P(X_1 = x_1, \ldots, X_m = x_m)$. In this format, the first-order Markov model provides an interesting method to describe the transitions between varying states. The probability of transitions of drought states in the next step depends only on the present state and has no connection to the drought states that are observed in the past. The joint probability using the first-order Markov model is given in the following equation:

$$P(X = x) = P(X_1 = x_1, \ldots, X_m = x_m) = P(X_1 = x_1) \prod_{J=2}^{m} P\left(X_j = x_j | X_{j-1} = x_{j-1}\right). \tag{1}$$

Furthermore, to simplify the notations, we use β to denote initial state probability and γ to represent the transition probability. For example, β_{x_1} shows the probability that the initial state is x_1 and transition probability of x_{j-1} to x_j is represented by $\gamma_{x_{j-1} x_j}$. So, utilizing the given notations, we can write as there are p states in the Markov model, and in this case, the initial state probabilities can be represented as $\beta = (\beta_1, \ldots \beta_p)^T$ and the matrix of the transitions as $\Gamma = (\gamma_{jr})_{p \times p}$. $P(X = x) = \beta_{x_1} \prod_{j=1}^{m} \gamma_{x_{j-1} x_j}$. Moreover, for the specific component based on finite mixture modeling the β_{x_1} and $\gamma_{x_{j-1} x_j}$ are replaced by the β_{kx_1} and $\gamma_{kx_{j-1} x_j}$ and the model can be written as follows:

$$f(x;\, \theta) = \sum_{k=1}^{K} \pi_k \beta_{kx_1} \prod_{j=1}^{m} \gamma_{\, kx_{j-1} x_j}. \tag{2}$$

The log-likelihood of equation (2) can be expressed as follows:

$$\log L(\theta; \{x_i\}_{i=1}^{n}) = \sum_{i=1}^{n} \log \left\{ \sum_{k=1}^{K} \pi_k \prod_{l=1}^{p} (\beta_{kl})^{I(=x_{i1}=l)} \prod_{j=1}^{m_i} \gamma_{\, kx_{i(j-1)x_{ij}}} \right\}. \tag{3}$$

In equation (3), the $I(.)$ is indicator function and m_i indicates the length of i^{th} categorical sequence. Expectation-maximization (EM) algorithm is employed to estimate the parameters [56].

2.3. Prediction of Future Drought Occurrences for Spatial-Temporal Categorical Sequences. The setting of transition probability matrices can be represented by $\Gamma_1, \Gamma_1, \ldots, \Gamma_k$ and a probability distribution $\pi_1, \pi_2, \ldots, \pi_K$ connected with mixture components, and the M-step transition probability matrix can be created by

$$\Gamma^M = \sum_{k=1}^{K} \alpha_k \Gamma_k^M, \tag{4}$$

where Γ_k^M indicates the matrix Γ^M raised to the power M. The choice of the appropriate distribution $\pi_1, \pi_2, \ldots, \pi_K$ is linked with the application. However, the $(\widehat{\alpha_1}, \widehat{\alpha_2}, \ldots, \widehat{\alpha_K})$ and (i.e., $\widehat{z_{i1}}, \widehat{z_{i2}}, \ldots, \widehat{z_{iK}}$), which are the mixing proportion estimated vector and the posterior probability estimated vector, respectively, associated with a particular sequence, can hold significant influence for the computation of probability distribution for future drought occurrences.

3. Application

The choice of the region is based on its structural impacts and other climatic characteristics that affect the other parts of the country. The outcomes of the study are obtained from the six selected stations with time-series data from January 1971 to December 2017 of the northern area of Pakistan (Figure 1) using SPI at various time scales. The selected stations have significant importance for the selected region and other regions of the country. For example, the reservoir system and agriculture sector are highly associated with the selected region; therefore, the climatic discrepancy of the region is significant for the other parts of the country [57, 58]. Furthermore, the fluctuation of the weather pattern in other regions within the country also contributed to their impacts on socio-economic and environmental sectors. Most of the parts of the country have been facing the highest temperature, and these parts are being highly influenced by global warming [58, 59]. Undoubtedly, extreme climate events, including high temperatures, rainstorms, and droughts, are frequently associated with global climate warming. Climate warming significantly affects the universe, which usually causes a high temperature and water deficiency. These issues are associated with drought occurrences that damage the environment, natural resources, and lives of the people distinctly more than any other natural hazard. Furthermore, it produces convoluted consequences for society and the economic sectors of the country. Therefore, it is vital to recognize the drought occurrences more instantly by emerging comprehensive and efficient frameworks and techniques. In this regard, a new technique is applied to the selected stations that will influentially expand the capability of detecting drought occurrences and improve the competencies for drought evaluation and its assessment.

3.1. Results. The findings of this study are obtained by using long time series data collected from six climatological stations in the northern area of Pakistan. The selected stations are observed to show homogenous results for the specific indices when calculated for varying stations with a single time scale [55]. However, on the varying time scales, the observations of the indices may vary. Furthermore, the inconsistency in their observations and varying generating processes of the drought states causes for developing a new method (i.e., MBCSTCS). The MBCSTCS considers the various time scales for a particular station as sequences with inconsistency in their sizes and varying data generating processes to analyze the spatiotemporal behavior of the drought states. It means that the observations of the SPI at scale-1 (SPI-1) for Astore station are considered as sequence-1, sequence-2 takes all observations of Astore station in SPI at scale-3 (i.e., SPI-3), and these sequences are continued to the last scale (SPI-24). Accordingly, these sequences can be assigned for other stations and time scales. Furthermore, the observations of each sequence assume that they come from the specific components that are selected appropriately for the data. The selected states are observed corresponding to every calculated value of SPI. These selected states are further distributed categorically for the computation of this study.

Moreover, Niaz et al. [55] proposed a new technique for monthly forecasting drought intensities using model-based clustering of categorical drought state sequences. The mentioned study is performed on various stations based on a single time scale. However, in this study, the various time scales are accumulatively considered for the monthly prediction of drought severity in a region. The outcomes of the current analysis are more appropriate, especially for the selected stations, and help the policymakers to make better policies related to various kinds of droughts including meteorological, hydrological, and socioeconomic. Furthermore, the current analysis is performed by using the R package ClickClust [45] that handles the case of coming observations from various probability distributions (K-components). The package is based on finite mixtures with Markov model components and is used to find the specific outcomes related to the specific sequence. The appropriate order K of the components (i.e., the mixture model) is identified by minimizing the Bayesian information criterion (BIC) [60]. Moreover, for a specific sequence, the mixing proportions estimated vector and the associated posterior probability estimated vector were used to calculate probability distribution associated with future steps of transitions from the last state of these sequences. Furthermore, climatological statistics on the given data of various stations are provided in Figure 2. To accomplish the analysis, the R package named *propagate* is used to provide appropriate findings and permit the specific analysis. In the mentioned package, various distributions are considered; among the given distributions, the appropriate choice of the distribution is based on the BIC values. This selection criterion is helpful to find the best fitting for the time scale and stations specified for the analysis.

The BIC values are given in Table 1 for the selected probability distributions fitting appropriately to the several time scales and stations. For example, at Astore station for scasle−1, the BIC value (−1036.5) of three-parameter (3P) Weibull distribution is found minimum among other distributions. Therefore, the (3P) Weibull distribution is

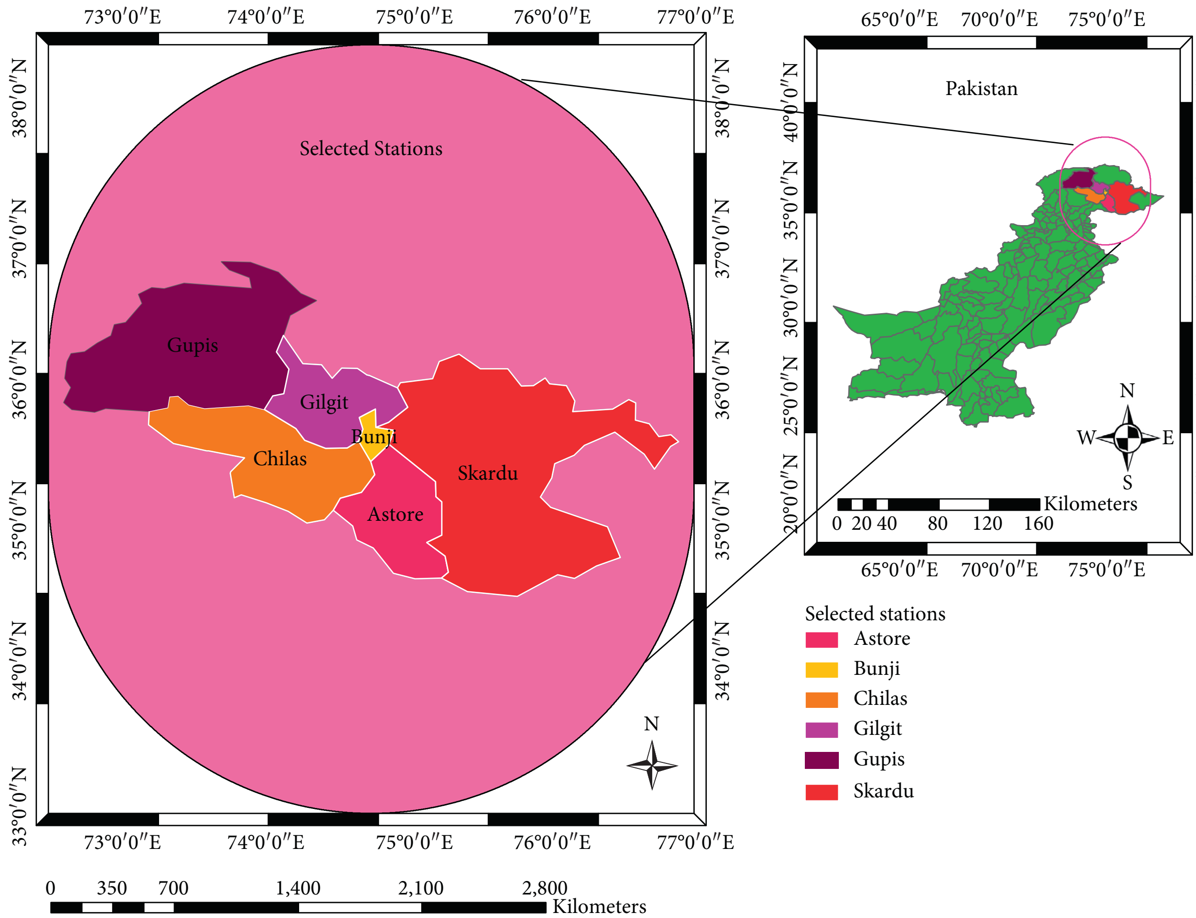


Figure 1: The study area consisting of six required stations in the northern area of Pakistan is presented.

considered as best fitted distribution for the Astore station at a scale−1. Furthermore, in Astore station for scale−3, the Gamma distribution is selected with the minimum BIC value (−1279.1). Moreover, in scale−6 and scale−9, it is also found that the Gamma distribution is best fitted in Astore station with minimum BIC values −892.8 and −896.1, respectively. Furthermore, Cosine and Skewed- normal distributions are considered for scale−12 and scale−24, respectively, in the Astore station. In Bunji station at scale−1, the (3P) Weibull is showing the minimum value of BIC (i.e., 1,031.0) and specified for the computation. At scale−3, scale−6, scale−9, scale−12, and scale−24 at Bunji station, the Gamma, Skewed-normal, Normal, Laplace, and Laplace are selected with BIC values −824.9, −1162.2, −649.1, −688.1, and −843.7, respectively. In Gupis station 4p beta has a minimum value of BIC (−788.7) for scale−1. In other scales (3, 6, 9, 12, and 24), the gamma, Gumbel, Johnson SU, and scaled/shifted t have minimum values of BIC −1264.9, −1305.4, 1519.0, −937.6, and 1408.0, respectively.

Accordingly, various distributions are selected for various time scales for the other three stations (Chilas, Gilgit, and Skardu) (1, 3, 6, 9, 12, and 24). After standardization with a selected probability distribution, the next step is the classification of the SPI for various drought states (Table 2). In Figure 3, the temporal behavior of the SPI at scale-1 is presented for various stations. However, the behavior of SPI for other selected time scales can be presented accordingly. After calculating values to quantify SPI at various time scales, we first categorized SPI for its magnitude. The behavior of several drought classes for SPI at a one-month time scale in selected stations is provided in Table 3, where the observations are taken as an example for various months of the year 2017. Accordingly, the behavior of several drought classes for other years for different time scales is calculated. These observed drought classes are further used to find the probability distribution associated with the three-step transition from the last state in the various sequences. The posterior vector related to these sequences specifies the parameter values (briefly described in Section 2.2). The obtained results show that the most likely state to visit in three steps is ND, which means the probability associated with ND is higher than the other selected states in varying sequences (Table 4). For example, for the Astore station, in sequence-1, the value indicates that the probability of ND occurrence is 0.6668, which is higher than other selected states. This probability of occurrence for ND can be observed from other sequences. Further in sequence-2, the probability of ND occurrence after three months is 0.6729. Moreover, the probabilities of ND in sequence-3, sequence-4, sequence-5, and sequence-6 are 0.6611, 0.6221, 0.6450, and 0.6729, respectively. It means that the policymakers should make their plans accordingly for ND. Other information can

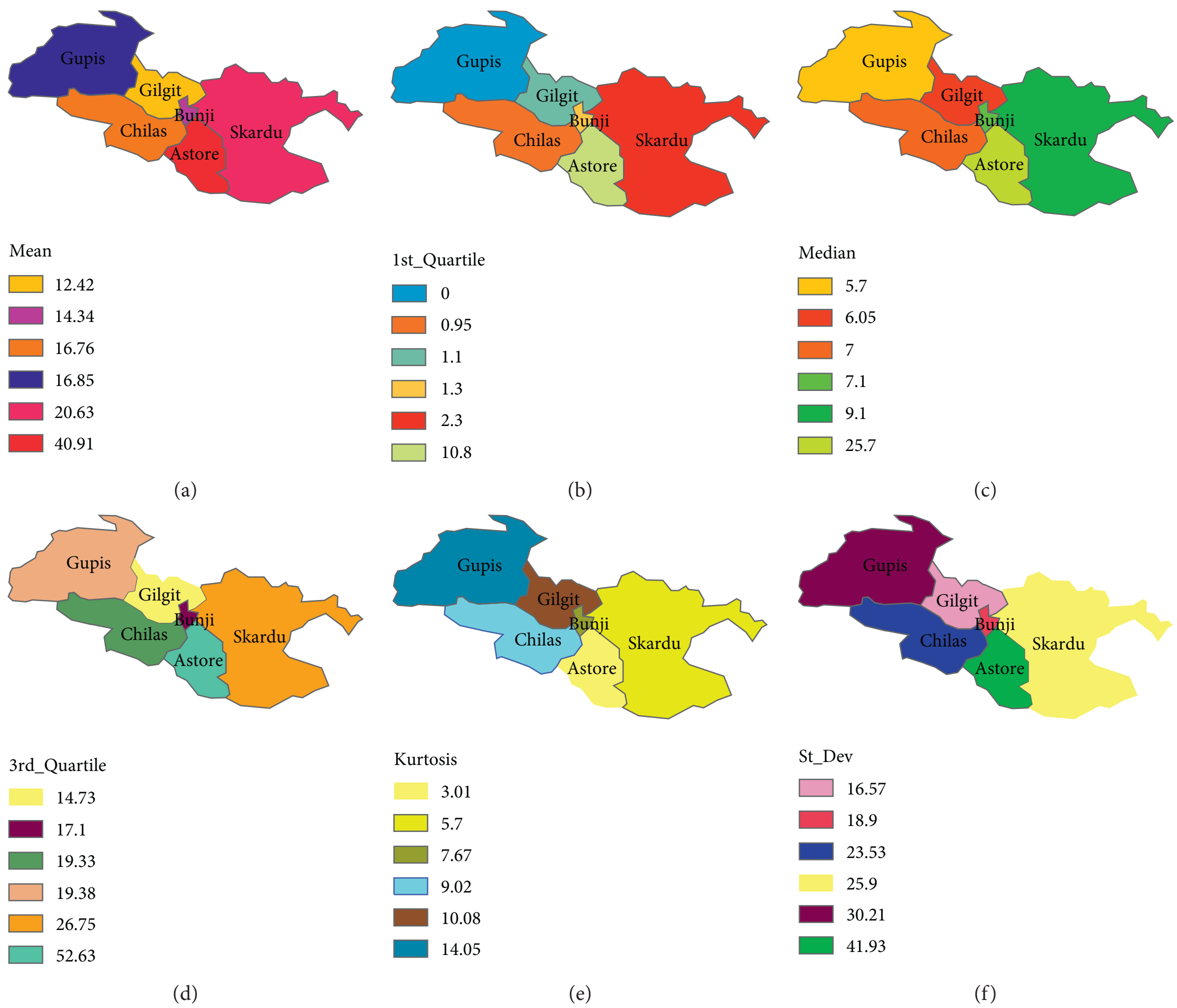


FIGURE 2: The climatological characteristics of the precipitation to describe the period from January 1971 to December 2017 of specified stations. The (a) mean, (b) 1st-quartile, (c) median, (d) 3^{rd}_quartile, (e) kurtosis, and (f) st_dev.

be observed from the various sequences for different time scales. However, the ND is prevailing in all time scales in the selected region. So, the policymakers should work to mitigate negative impacts for this specific drought state (ND).

3.2. Discussion. The time series data were collected from six meteorological stations in the northern area of Pakistan. The drought index SPI is used for the analysis with various time scales for selected stations. The reliable and efficient outcomes of the analysis provide strong indications about the drought occurrences that can significantly help for an early warning system [31, 53, 58, 59, 61]. Therefore, a new MBCSTCS method is developed for the drought monitoring and mitigation policies that explicitly envisage spatiotemporal information. The proposed technique uses the long-run behavior of drought states (categorical sequences) from various time scales and stations in the selected region. If a time scale changes, then the categorical sequence sizes are affected. Therefore, in past studies, researchers have not been studied various time scales accumulatively due to inconsistency in their sizes and the phenomenon that has generated the observations for varying stations. However, these issues are being resolved effectively by the current technique. Furthermore, the outcomes associated with the present technique help to accomplish the current objective and provide more substantial outcomes for the selected drought states based on varying time scales and stations. MBCSTCS uses state selection procedures through finite mixture modeling and model-based clustering. Niaz et al. [55] developed a new model-based clustering technique that predicts probabilities for various drought classes. They computed categorical drought state (classes) sequences for selected drought classes and predicted their probabilities for the future. The mentioned study used a single time scale on various stations. However, in this study, the varying time scales are accumulatively measured for the monthly prediction of drought severity in selected stations. Therefore, it is a novel method for predicting drought severity using spatiotemporal categorical sequences. MBCSTCS is applied to six meteorological stations in the northern area of Pakistan. It is found that MBCSTCS provides expeditious information for the long-term spatiotemporal categorical sequences. The present analysis results are more suitable,

Table 1: The BIC of selected probability distributions for SPI at various timescales (1, 3, 6, 9, 12, and 24) for the selected stations

		Astore		Bunji		Gupis		Chilas		Gilgit		Skardu	
Index	Scale	Distribution	BIC	Distribution	BIC	Distribution	BIC	Distribution	BIC	Distribution	BIC	Distribution	BIC
SPI	1	3P Weibull	−1036.5	3p Weibull	−1031.0	4p beta	−788.7	4P beta	−805.6	3P Weibull	−1097.4	3P Weibull	−735.1
	3	Gamma	−1279.1	Gamma	−824.9	Gamma	−1264.9	3P Weibull	−810.2	Exponential	−824.2	Generalized extreme value	−653.9
	6	Gamma	−892.8	Skewed-normal	−1162.2	Gumbel	−1305.4	Triangular	−1190.6	Gamma	−1204.8	Log−normal	−1018.5
	9	Gamma	−896.1	Normal	−649.1	Johnson SU	−1519.0	Normal	−1266.7	Laplace	−1206.3	Log−normal	−1109.6
	12	Cosine	−913.3	Laplace	−688.1	Johnson SU	−937.6	Laplace	−1325.8	Inverse gamma	−1266.4	Gamma	−1025.7
	24	Skewed-normal	−1294.4	Laplace	−843.7	Scaled/shifted t	−1408.0	Laplace	−826.8	Trapezoidal	−681.7	Inverse Gaussian	−926.8

Table 2: The classification of the SPI for various drought states

SPI	Observed drought states
SPI ≤ −2	Extremely dry (ED)
SPI > −2 and SPI ≤ −1.5	Severely dry (SD)
SPI > −1.5 and SPI ≤ −1	Median dry (MD)
SPI > −1 and SPI ≤ 1	Normal dry (ND)
SPI > 1 and SPI ≤ 1.5	Median wet (MW)
SPI > 1.5 and SPI ≤ 2	Severely wet (SW)
SPI > 2	Extremely wet (EW)

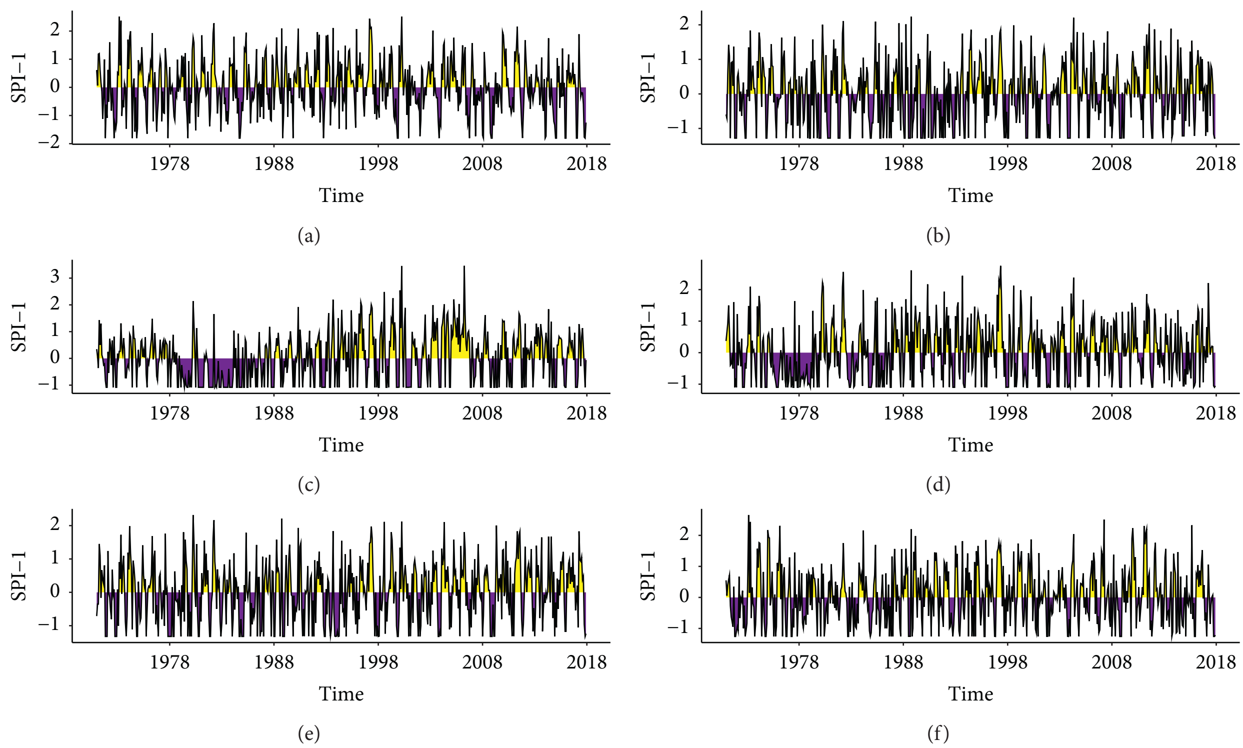


Figure 3: The temporal presentation of SPI at scale-1 for varying stations. (a) Astore, (b) Bunji, (c) Gupis, (d) Chilas, (e) Gilgit, and (f) Skardu.

Table 3: The behavior of various drought classes for SPI at scale-1 in selected stations.

Month	Astore		Bunji		Gupis		Chilas		Gilgit		Skardu	
	SPI	Class	SPI	Class	SPI	Class	SPI	Class	SPI	Class	SPI	Class
Jan	−1.8044	SD	−1.2152	MD	−1.0805	MD	−0.3254	ND	−0.230	ND	−0.4827	ND
Feb	−1.8044	SD	0.2629	ND	−1.0805	MD	0.8075	ND	0.201	ND	−1.2581	ND
Mar	−0.4745	ND	−1.1059	MD	−0.6655	ND	−0.1535	ND	−1.052	MD	−0.3096	ND
Apr	1.8879	SW	0.9377	ND	1.2811	MW	2.2094	EW	1.828	SW	1.0615	EW
May	0.3314	ND	0.6975	ND	0.9813	ND	0.8297	ND	0.972	ND	0.0796	ND
Jun	−0.3302	ND	−0.6380	ND	−0.2883	ND	−0.0637	ND	0.931	ND	−0.8574	ND
Jul	−0.1857	ND	0.8192	ND	0.7872	ND	0.1366	ND	1.057	MW	−0.2401	ND
Aug	−0.1967	ND	0.7181	ND	1.1317	MW	0.2058	ND	0.386	ND	0.0086	ND
Sep	−0.5938	ND	0.2824	ND	0.4029	ND	0.1937	ND	0.476	ND	0.0739	ND
Oct	−1.7685	SD	−1.1576	MD	−0.2883	ND	−1.0539	MD	−0.957	ND	−1.2581	MD
Nov	−1.8044	SD	−1.2806	MD	−1.0805	MD	−1.0995	MD	−1.323	MD	−1.2581	MD
Dec	−1.2499	MD	−1.2806	MD	−0.0621	ND	−1.0539	MD	−1.240	MD	−1.2168	MD

Note. The observations are considered for varying months of the year 2017. The specific year is selected as an example. Accordingly, the behavior of various drought classes for other years for various time scales is calculated.

Table 4: Predicting future three-step transitions from the last state for various drought sequences at Astore station

Three-month prediction in Astore							
Sequence	ED	SD	MD	ND	MW	SW	EW
1	0.0014	0.0766	0.0965	0.6668	0.0950	0.0409	0.0228
2	0.0010	0.0691	0.0927	0.6729	0.0962	0.0436	0.0241
3	0.0017	0.0829	0.0999	0.6611	0.0941	0.0388	0.0215
4	0.0065	0.0127	0.2198	0.6221	0.0815	0.0437	0.0137
5	0.0015	0.0014	0.1987	0.6450	0.0878	0.0506	0.0150
6	0.0014	0.0691	0.0927	0.6729	0.0962	0.0436	0.0241

especially for the selected region, and help the policymakers make better policies related to various kinds of drought, including meteorological, hydrological, and socioeconomic. The MBCSTCS may help to make plans for early warning systems, water resource management, and drought mitigation policies to reduce the severe effects of drought.

4. Conclusions

The outcomes of MBCSTCS provide the future probabilities corresponding to each of the drought states in varying stations and time scales. The obtained outcomes show that the most likely state to visit is ND, which means the probability associated with ND is higher than the other selected states in varying sequences. For instance, in sequence-1, the value shows that the probability of ND is 0.6668, which is higher than other selected states. Further in sequence-2, the probability of ND after three months is 0.6729. This probability of ND also prevails in other sequences. Furthermore, in sequence-6, the ND has a higher probability (0.6729) of occurrence in the future. Therefore, policymakers should work to reduce the negative impacts of this drought state (ND). In conclusion, this study suggests a more appropriate technique that emphasizes evaluating drought occurrences more instantaneously. The MBCSTCS helps the policymakers to make better policies related to various kinds of drought, including meteorological, hydrological, and socioeconomic. Furthermore, the analysis provides the basis to bring more attention to early warning systems. Moreover, the outcomes of the current analysis are only capable of transmitting in the present circumstances of the application site, as the circumstance of the selected stations will change the influence of the outcome for the extrapolations. Furthermore, the study can find some propagations and compute several thresholds for different drought severities for the region. Moreover, other drought indices can be incorporated to envisage the drought occurrences effectively.

Data Availability

The data used for the preparation of the manuscript are available from the corresponding author and can be provided upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China program (41801339). The authors are also thankful to the Deanship of Scientific Research at King Saud University for the support, through research group no. RG-1439-015. Finally, the author also extends his appreciation to the Deanship of Scientific Research at King Khalid University for funding this work under grant number (RGP.1/26/42), received by Mohammed M. Almazah (https://www.kku.edu.sa).

References

[1] E. Gidey, O. Dikinya, R. Sebego, E. Segosebe, and A. Zenebe, "Predictions of future meteorological drought hazard (~2070) under the representative concentration path (RCP) 4.5 climate change scenarios in Raya, Northern Ethiopia," *Modeling Earth Systems and Environment*, vol. 4, no. 2, pp. 475–488, 2018.

[2] M. Trnka, P. Hlavinka, M. Možný et al., "Czech drought monitor system for monitoring and forecasting agricultural drought and drought impacts," *International Journal of Climatology*, vol. 40, no. 14, pp. 5941–5958, 2020.

[3] M. Jehanzaib, J. Yoo, H.-H. Kwon, and T.-W. Kim, "Reassessing the frequency and severity of meteorological drought considering non-stationarity and copula-based bivariate probability," *Journal of Hydrology*, vol. 603, Article ID 126948, 2021.

[4] R. Salam, A. R. M. Towfiqul Islam, B. K. Shill et al., "Nexus between vulnerability and adaptive capacity of drought-prone rural households in northern Bangladesh," *Natural Hazards*, vol. 106, no. 1, pp. 509–527, 2021.

[5] P. Kalura, A. Pandey, V. M. Chowdary, and P. V. Raju, "Assessment of hydrological drought vulnerability using geospatial techniques in the tons river basin, India," *Journal of the Indian Society of Remote Sensing*, vol. 49, pp. 1–15, 2021.

[6] A. H. Payab and U. Türker, "Comparison of standardized meteorological indices for drought monitoring at northern part of Cyprus," *Environmental Earth Sciences*, vol. 78, no. 10, pp. 1–19, 2019.

[7] O. O. Ayantobo and J. Wei, "Appraising regional multi-category and multi-scalar drought monitoring using standardized moisture anomaly index (SZI): a water-energy balance approach," *Journal of Hydrology*, vol. 579, Article ID 124139, 2019.

[8] H. Salimi, E. Asadi, and S. Darbandi, "Meteorological and hydrological drought monitoring using several drought indices," *Applied Water Science*, vol. 11, no. 2, pp. 1–10, 2021.

[9] O. V. Wilhelmi, K. G. Hubbard, and D. A. Wilhite, "Spatial representation of agroclimatology in a study of agricultural drought," *International Journal of Climatology*, vol. 22, no. 11, pp. 1399–1414, 2002.

[10] S. Jiang, L. Wei, L. Ren et al., "Utility of integrated IMERG precipitation and GLEAM potential evapotranspiration products for drought monitoring over mainland China," *Atmospheric Research*, vol. 247, Article ID 105141, 2021.

[11] F. Hui-Mean, Z. Yusop, and F. Yusof, "Drought analysis and water resource availability using standardised precipitation

evapotranspiration index," *Atmospheric Research*, vol. 201, pp. 102–115, 2018.

[12] Y. Lin, F. Liu, and P. Xu, "Effects of drought on infant mortality in China," *Health Economics*, vol. 30, no. 2, pp. 248–269, 2021.

[13] N. Puletti, W. Mattioli, F. Bussotti, and M. Pollastrini, "Monitoring the effects of extreme drought events on forest health by Sentinel-2 imagery," *Journal of Applied Remote Sensing*, vol. 13, no. 2, Article ID 020501, 2019.

[14] D. A. Wilhite, *Drought as a Natural hazard: Concepts and Definitions*, Springer, Berlin, Germany, 2000.

[15] T. J. Chang, "Investigation of precipitation droughts by use of kriging method," *Journal of Irrigation and Drainage Engineering*, vol. 117, no. 6, pp. 935–943, 1991.

[16] E. A. B. Eltahir, "Drought frequency analysis of annual rainfall series in central and western Sudan," *Hydrological Sciences Journal*, vol. 37, no. 3, pp. 185–199, 1992.

[17] T. J. Chang and X. A. Kleopa, "A proposed method for drought monitoring," *Journal of the American Water Resources Association*, vol. 27, no. 2, pp. 275–281, 1991.

[18] T. Estrela and E. Vargas, "Drought management plans in the European Union. The case of Spain," *Water Resources Management*, vol. 26, no. 6, pp. 1537–1553, 2012.

[19] Y. B. Zecharias and W. Brutsaert, "The influence of basin morphology on groundwater outflow," *Water Resources Research*, vol. 24, no. 10, pp. 1645–1650, 1988.

[20] S. Mohan and N. C. V. Rangacharya, "A modified method for drought identification," *Hydrological Sciences Journal*, vol. 36, no. 1, pp. 11–21, 1991.

[21] B. Clausen and C. P. Pearson, "Regional frequency analysis of annual maximum streamflow drought," *Journal of Hydrology*, vol. 173, no. 1-4, pp. 111–130, 1995.

[22] W. C. Palmer, "Keeping track of crop moisture conditions, nationwide: the new Crop Moisture Index," *Weatherwise*, vol. 21, no. 4, pp. 156–161, 1968.

[23] H. N. Bhalme and D. A. Mooley, "Large-scale droughts/floods and monsoon circulation," *Monthly Weather Review*, vol. 108, no. 8, pp. 1197–1211, 1980.

[24] B. A. Shafer and L. E. Dezman, "Development of Surface Water Supply Index (SWSI) to assess the severity of drought condition in snowpack runoff areas," in *Proceedings of the Western Snow Conference*, Reno, Nevada, January 1982.

[25] T. B. McKee, N. J. Doesken, and J. Kleist, "The relationship of drought frequency and duration to time scales," in *Proceedings of the 8th Conference on Applied Climatology*, pp. 179–183, Boston, MA, USA, 1993 January.

[26] K. M. Weghorst, *The Reclamation Drought Index: Guidelines and Practical Applications*, Bureau of Reclamation, Denver, CO, USA, 1996.

[27] W. C. Palmer, *Meteorological Drought*, US Department of Commerce, Washington, DC, USA, 1965.

[28] R. Gommes and F. Petrassi, *Rainfall Variability and Drought in Sub-saharan Africa. SD Dimensions*, FAO, Rome, Italy, 1996.

[29] M. P. Van Rooy, *A Rainfall Anomally Index Independent Of Time And Space*, NOTOS, Kifisia, Greece, 1965.

[30] J. González and J. B. Valdés, "New drought frequency index: definition and comparative performance analysis," *Water Resources Research*, vol. 42, no. 11, 2006.

[31] A. K. Mishra and V. P. Singh, "A review of drought concepts," *Journal of Hydrology*, vol. 391, no. 1-2, pp. 202–216, 2010.

[32] A. Zargar, R. Sadiq, B. Naser, and F. I. Khan, "A review of drought indices," *Environmental Reviews*, vol. 19, pp. 333–349, 2011.

[33] A. K. Mishra and V. P. Singh, "Drought modeling–A review," *Journal of Hydrology*, vol. 403, no. 1-2, pp. 157–175, 2011.

[34] B. Amirataee and M. Montaseri, "The performance of SPI and PNPI in analyzing the spatial and temporal trend of dry and wet periods over Iran," *Natural Hazards*, vol. 86, no. 1, pp. 89–106, 2017.

[35] M. M. Moghimi and A. R. Zarei, "Evaluating performance and applicability of several drought indices in arid regions," *Asia-Pacific Journal of Atmospheric Sciences*, vol. 57, no. 3, pp. 645–661, 2021.

[36] W.-H. Nam, M. J. Hayes, M. D. Svoboda, T. Tadesse, and D. A. Wilhite, "Drought hazard assessment in the context of climate change for South Korea," *Agricultural Water Management*, vol. 160, pp. 106–117, 2015.

[37] K. Nigam, A. K. McCallum, S. Thrun, and T. Mitchell, "Text classification from labeled and unlabeled documents using EM," *Machine Learning*, vol. 39, no. 2, pp. 103–134, 2000.

[38] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645–678, 2005.

[39] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical dirichlet processes," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1566–1581, 2006.

[40] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007.

[41] S. B. Dalirsefat, A. da Silva Meyer, and S. Z. Mirhoseini, "Comparison of similarity coefficients used for cluster analysis with amplified fragment length polymorphism markers in the silkworm, Bombyx mori," *Journal of Insect Science*, vol. 9, no. 1, pp. 1–8, 2009.

[42] J. H. Ward, "Hierarchical grouping to optimize an objective function," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236–244, 1963.

[43] C. Fraley and A. E. Raftery, "Model-based clustering, discriminant analysis, and density estimation," *Journal of the American Statistical Association*, vol. 97, no. 458, pp. 611–631, 2002.

[44] V. Melnykov, "Challenges in model-based clustering," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 5, no. 2, pp. 135–148, 2013.

[45] V. Melnykov, "ClickClust: an R package for model-based clustering of categorical sequences," *Journal of Statistical Software*, vol. 74, no. 9, pp. 1–34, 2016.

[46] M. S. Handcock, A. E. Raftery, and J. M. Tantrum, "Model-based clustering for social networks," *Journal of the Royal Statistical Society: Series A*, vol. 170, no. 2, pp. 301–354, 2007.

[47] J. D. Banfield and A. E. Raftery, "Model-based Gaussian and non-Gaussian clustering," *Biometrics*, vol. 49, no. 3, pp. 803–821, 1993.

[48] G. McLachlan and D. Peel, "Mixtures of factor analyzers," in *Proceedings of the 7th International Conference on Machine Learning*, Bellevue, WA, USA, June 2000.

[49] T. Warren Liao, "Clustering of time series data-a survey," *Pattern Recognition*, vol. 38, no. 11, pp. 1857–1874, 2005.

[50] W.-C. Chen and R. Maitra, "Model-based clustering of regression time series data via APECM-an AECM algorithm sung to an even faster beat," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 4, no. 6, pp. 567–578, 2011.

[51] A. K. Mishra, V. R. Desai, and V. P. Singh, "Drought forecasting using a hybrid stochastic and neural network model," *Journal of Hydrologic Engineering*, vol. 12, no. 6, pp. 626–638, 2007.

[52] V. Mishra, A. D. Tiwari, S. Aadhar et al., "Drought and f," *Geophysical Research Letters*, vol. 46, no. 4, pp. 2075–2083, 2019.
[53] R. R. Heim, "A review of twentieth-century drought indices used in the United States," *Bulletin of the American Meteorological Society*, vol. 83, no. 8, pp. 1149–1166, 2002.
[54] B. Lloyd-Hughes and M. A. Saunders, "A drought climatology for Europe," *International Journal of Climatology: A Journal of the Royal Meteorological Society*, vol. 22, no. 13, pp. 1571–1592, 2002.
[55] R. Niaz, I. M. Hussain, Z. Ali et al., "A novel spatially weighted accumulative procedure for regional drought monotoring," *Tellus A: Dynamic Meteorology and Oceanography*, vol. 72, 2020.
[56] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via theEMAlgorithm," *Journal of the Royal Statistical Society: Series B*, vol. 39, no. 1, pp. 1–22, 1977.
[57] S. A. Awan, "The climate and flood risk potential of Northern areas of Pakistan," *Science Vision*, vol. 7, no. 3-4, pp. 100–109, 2002.
[58] S. Adnan and K. Ullah, "Development of drought hazard index for vulnerability assessment in Pakistan," *Natural Hazards*, vol. 103, no. 3, pp. 2989–3010, 2020.
[59] K. Ahmed, S. Shahid, and N. Nawaz, "Impacts of climate variability and change on seasonal drought characteristics of Pakistan," *Atmospheric Research*, vol. 214, pp. 364–374, 2018.
[60] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.
[61] S. M. Vicente-Serrano, S. Beguería, and J. I. López-Moreno, "A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index," *Journal of Climate*, vol. 23, no. 7, pp. 1696–1718, 2010.

Hindawi
Complexity
Volume 2021, Article ID 5774789, 10 pages
https://doi.org/10.1155/2021/5774789

Research Article

Integrative Analytics for Technological Pedagogical Content Knowledge

Jin-E. Zhang

College of Mathematics and Statistics, Hubei Normal University, Huangshi 435002, China

Correspondence should be addressed to Jin-E. Zhang; zhang86021205@163.com

Received 10 October 2021; Revised 27 October 2021; Accepted 3 November 2021; Published 27 November 2021

Academic Editor: Feng Li

In this paper, we deeply analyze the general framework of technological pedagogical content knowledge (TPACK): it mainly involves technological knowledge (TK), pedagogical knowledge (PK), content knowledge (CK), technological content knowledge (TCK), technological pedagogical knowledge (TPK), and pedagogical content knowledge (PCK). Meanwhile, we also show that the conceptual framework of TPACK can help teachers to integrate a lot of complexity and tension in teaching and learning. Also, this study suggests that the framework of TPACK may be able to change the application mode of technology in education and the training mode of teachers and then provide a panoramic analysis for building a learning community.

1. Introduction

Teachers are an important factor for schools to achieve connotative development and improve education quality. Teachers' development is a necessary guarantee for teaching reform and educational innovation. Teachers' knowledge is the basis for teachers to carry out teaching activities. The composition and structure of teachers' knowledge will directly affect teachers' teaching behavior and students' learning effect. Since entering the new era, the continuous development of informatization technology has put forward higher requirements for teachers' knowledge structure and teaching ability. Only by actively adapting to technological changes such as informatization, networking, and artificial intelligence (AI), then teachers can effectively improve the quality of talent training and promote the vigorous development of education. As "Ten-Year Development Plan for Educational Informatization (2011–2020)" in China points out, we should promote the integration of informatization technology and teaching and then promote the professional development of teachers. The opinions issued by the Ministry of Education of China on the "Implementation of Excellent Teacher Training Plan 2.0" clearly propose to promote the all-round integration of new technologies such as AI and intelligent learning environment with teacher education courses. The "Excellent Teacher Training Plan" proposed by China also aims to cultivate a large number of high-quality professional teachers with noble ethics, solid professional foundation, and outstanding educational and teaching ability. So we have to think about such a question: What knowledge does a teacher need to become a qualified teacher [1]? Since the mid-1980s, scholars have conducted in-depth research on pedagogical content knowledge (PCK) [1–3]. In recent years, for mathematics education, mathematics pedagogical content knowledge (MPCK) is one of the hot issues for educational studies in mathematics.

Knowledge learning is the core value pursuit of teachers' development. The rapid, continuous innovation and progress of informatization technology has reformed the education. With it, teachers' knowledge and ability need to be redefined and designed. Teachers' technical literacy has a direct impact on the behavioral intention of information-based teaching. The main obstacle to the use of technology in classroom is the lack of technical knowledge and skills. If teachers do not know how to operate technology, they will not use technology. We should pay attention to the cultivation of teachers' technical knowledge and help teachers master the use of basic technology. At the same time, teachers should be encouraged and supervised to apply the learned technical knowledge and skills to classroom teaching. Therefore, for the promotion of educational informatization, one should pay particular attention to improving teachers' beliefs and attitudes toward the application of technology teaching, so as to improve teachers' intention of information-based teaching behavior. Today, informatization technology not only affects human thinking and behavior but also plays a great role in economic, cultural, and social activities. At present, informatization technology has been widely used in the field of education at all levels. With the more and more frequent application of digital technology for teaching in the Information Age, teachers generally lack the experience of using digital technology in teaching, which makes the professional knowledge of teachers difficult to meet the needs of students growing up in the Information Age. Teachers' development must pay attention to teachers' existing experience, environment, and surrounding colleagues, which is a circular process. Under the background of informatization, teachers in professional development must think about and pay attention to the effective integration of technology, PCK, and teaching method [4, 5]. That is, they ought to possess a professional knowledge: technological pedagogical content knowledge (TPACK). The hierarchical structure of traditional education (see Figure 1) is a dull and inappropriate structure. Learning under traditional education is always passive, and it is difficult to support today's learners to have rich knowledge, complex thinking ability, and cooperation ability. Contrary, TPACK is an important framework for reshaping teachers' professional development experience [6–8]. Restructuring teachers' knowledge development in pre-service and in-service teacher education to adapt to TPACK is a logical step in teacher education reform. Actually, the framework of TPACK is based on the concept of PCK. TPACK is the basis for teachers to use technology for effective teaching. It has become a more and more important mission for today's mathematics teachers to exert their imagination on the connection between technology and teaching, judge the advantages of relevant teaching strategies, and integrate all factors to achieve an effective mathematics class [9–12].

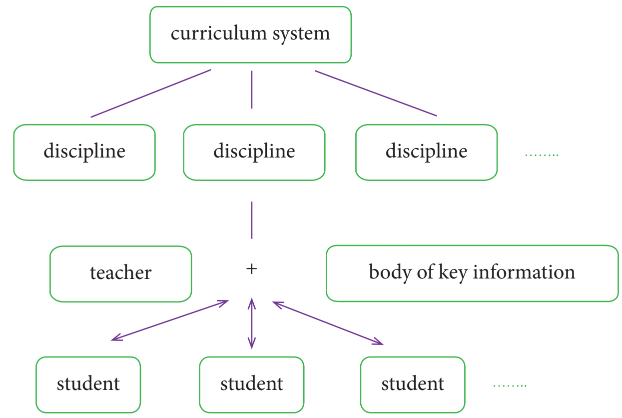


Figure 1: The hierarchical structure of traditional education.

2. Integration of TPACK in Teachers' Professional Quality

TPACK is an important part of teachers' professional development. In normal education and in-service teacher training, teachers should have knowledge and skills, cultivate teachers' tendency to try new technologies and learn from their own teaching, foresee possible problems, and insist on using technology in a way that is conducive to students' learning. Next, we discuss the teachers' information technology literacy from three dimensions, i.e., informatization technology sentiment, informatization technology skills and informatization teaching ability, and analyzing the relationship between the three.

2.1. Informatization Technology Sentiment. For normal students, with the growth of grade, the informatization technology sentiment of mathematics normal students is rising. Because educational practice is different from tutoring and other relatively single guidance, strictly speaking, educational practice is the first time that normal students contact the formal classroom as teachers and carry out formal lesson preparation. It can be said that educational practice means that normal students enter another environment and role. What they see, hear, and experience accumulation during educational practice will certainly affect their views on teaching, including their views on information-based teaching. On the whole, educational practice can promote normal students' understanding and reflection on informatization technology. Therefore, the internship experience may make normal students more interested in informatization technology. Another important factor affecting normal students' informatization technology sentiment is whether they have made courseware or instructional design. After making courseware or instructional design, normal students will find that informatization technology is very helpful to present knowledge and promote students' understanding of knowledge. Using informatization technology to prepare lessons will not only improve the efficiency of lesson preparation but also diversify the presentation of knowledge and support novel teaching methods. Normal students who have made courseware or instructional design generally have higher informatization technology sentiment. For in-service teachers, one of the main factors affecting their informatization technology sentiment is teaching age. On the one hand, teachers with short teaching age are relatively young, and they are willing to accept new things, have higher enthusiasm for learning technology, and therefore have higher intentions for informatization technology. On the other hand, traditional teaching methods are still deeply rooted in the hearts of older teachers, and they are familiar with these methods, which may be one of the reasons for their low intentions for informatization technology.

2.2. Informatization Technology Skills. For any TPACK framework, technology is one of the main factors in its content. Therefore, for teachers, no matter how excellent application of teaching methods are, they should be closely linked with technology in order to integrate, and the basis of all this is to be able to use technology. A survey of normal students and in-service teachers on the use of informatization technology shows that the score of software commonly used in daily office is relatively high. For example, network application, word processing software, office software, and PPT are the most familiar technologies and software for teachers. In addition, benefiting by the promotion of educational informatization carried out by government departments at all levels. By updating equipment, training teachers, and carrying out various teaching research activities, the majority of teachers are naturally familiar with $Z+Z$ education platform, LATEX, SPSS, etc., which are professional mathematical software or statistical software. Although teachers seldom use these software, they can also help teachers' teaching. For instance, as statistical software, we can use SPSS to analyze the data, so as to obtain the development curve of students or make data-based evaluation on students.

2.3. Informatization Teaching Ability. Both normal students and in-service teachers are more optimistic about the presentation of knowledge through technology and the combination of technology and teaching method, while the integration of technology, knowledge, and teaching method and the evaluation of students through informatization technology are not optimistic. This shows that normal students and in-service teachers doubt whether technology can participate in the complete teaching process. The possible reasons may lie in: (1) the technical literacy and professional level of normal students and in-service teachers need to be improved; and (2) some teachers cannot accept the modern educational concept. This requires a long process and hard work to achieve the deep integration of technology, knowledge, and teaching. For the integration of technology and knowledge, normal students believe that technology is more suitable for presenting geometry, image, or dynamic knowledge. In-service teachers are also aware of the convenience of Geometer's Sketchpad in mathematics teaching and are willing to use it to draw geometric graphics or present the changes of function images. The application of informatization technology is diverse, which has incomparable advantages over traditional teaching methods in the teaching of algebra and probability. This requires normal students and in-service teachers to master more knowledge to support their teaching in more knowledge fields.

It is found that the informatization technology literacy of normal students and in-service teachers is significantly related to informatization technology sentiment, informatization technology skills, and informatization teaching ability. Whether for normal students or in-service teachers, the dimension of informatization teaching ability has the greatest impact on informatization technology literacy.

3. Application of TPACK in Teachers' Teaching Practice

TPACK is the basis for teachers to use technology for effective teaching. The National Council of Teachers of Mathematics (NCTM) has pointed out that technology is an essential tool, and teachers must be prepared to ask more why in teaching practice. For example, why is this technology used in the class? How can this technology support the expansion of students' mathematical thinking to the greatest extent? To successfully make this technology as a tool for learning content knowledge, what should the students do to get started?

3.1. Overall Level of TPACK for Mathematics Teachers. Numerous studies have now found that the approach about learning and mastering informatization technology for teachers is mainly self-study. Due to the limited learning resources and learning channels, some teachers have fewer opportunities to receive training than listening to reports and lectures. From the six dimensions of TPACK, i.e., technological knowledge (TK), pedagogical knowledge (PK), content knowledge (CK), technological content knowledge (TCK), technological pedagogical knowledge (TPK), and PCK, as illustrated in Figure 2, the recent survey found that teachers' PK, CK, and PCK are relatively good, that is, PK, CK, and PCK are above the overall level. Among them, the PCK score is the highest. The mastery of TK and TCK is relatively low, among which the level of TK is the lowest, which indicates that teachers are relatively lacking technical knowledge. To a certain extent, it also reflects that teachers have not formed correct attitudes and habits in the aspect of application technique. Due to the relative lack of technical knowledge, the level of relevant technological knowledge is also relatively low. It can be seen that the lack of technical knowledge and technological barriers are the main factors affecting teachers' TPACK level.

3.2. Difference and Correlation of TPACK for Mathematics Teachers. From the perspective of gender, the average score of male teachers is higher than that of female teachers in the dimensions of TK, PK, CK, TCK, TPK, and PCK. As well, there are no significant differences in TK, PK, and CK, but there are significant differences in TCK, TPK, and PCK.

In terms of teaching age, for PK, CK, and PCK, the average score of teachers with teaching age of 11–15 years and more than 16 years is higher than that of teachers with teaching age of less than 3 years and 4–10 years, and there is a significant difference in PK and CK, but there is no significant difference in PCK. For TK and TCK, the average score of teachers with teaching age of 11–15 years is the highest. As far as the teaching age is concerned, there is no significant difference among TK, TCK, TPK, and PCK. In addition, there is a positive correlation between teaching age and PK, CK, and PCK.

In terms of professional titles, teachers' scores on PK, CK, and PCK are higher with the increase of professional titles, and there are significant differences. In TK, TCK, TPK, and PCK, the teachers with the lowest professional titles have the lowest average score. There is no significant difference in professional titles among TK, TCK, TPK, and PCK. In addition, professional titles are positively correlated with PK, CK, and PCK.

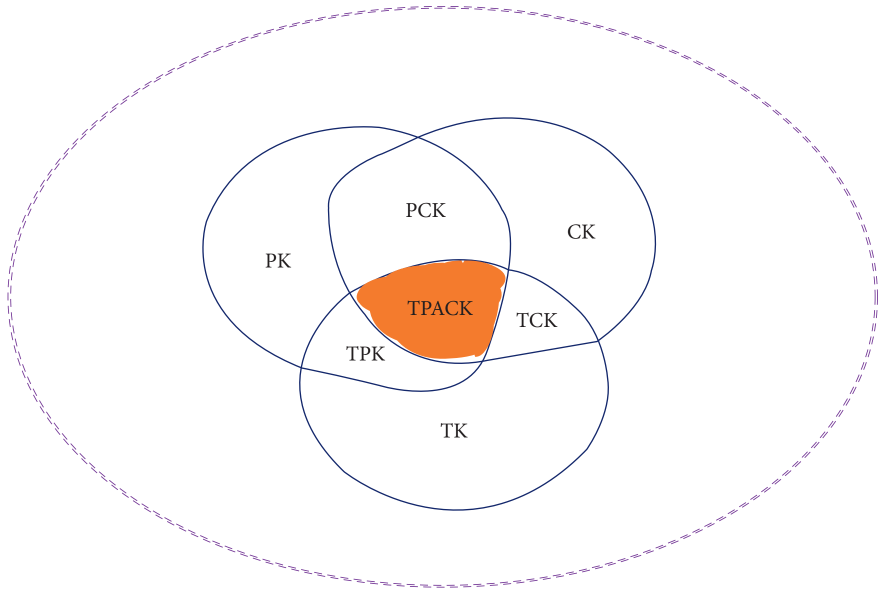


Figure 2: TPACK framework and its constituent components.

From the perspective of school types, teachers' scores on PK, CK, and PCK are related to the types of teaching school (key middle school and ordinary middle school). The better the school level is, the higher the average score is, but there is no significant difference. In terms of school types, there is no significant difference between TK and TCK, but there is significant difference between TPK and PCK.

In terms of the highest educational background, the average score of teachers with bachelor's degree in TK, PK, CK, TCK, and PCK is higher than that of teachers with master's degree, while the average score of teachers with master's degree on TPK is higher than that of teachers with bachelor's degree.

3.3. Application and Value of Informatization Technology. Mathematics teachers generally believe that informatization technology is useful for teaching, but their awareness of the use of informatization technology is not strong, and they are not proficient in using all kinds of informatization technology software. The roles of informatization technology in mathematics teaching include enriching classroom teaching methods, obtaining curriculum resources, enriching students' learning methods, improving teachers' professional quality, evaluating students' learning, and others. Teachers also believe that the main advantages of integrating informatization technology into mathematics teaching are (1) informatization technology can replace traditional blackboard writing, save blackboard writing time, and increase classroom capacity; (2) informatization technology can provide more situational materials and learning resources; (3) we can use informatization technology to analyze and evaluate students' learning situation and carry out dynamic tracking evaluation; (4) informatization technology can present knowledge more vividly, improve students' visual expression and visual control ability, make students more willing to accept and explore, enrich the depth and breadth of teaching topics, and promote students' exploration ability and in-depth understanding of knowledge; and (5) informatization technology can not only enhance the interaction between teachers and students but also improve students' interest in learning.

4. Systematic Characteristics of TPACK Knowledge Framework

The TPACK knowledge framework possesses systematic characteristics, which can be summarized as follows:

4.1. Integrality. The system is composed of more than two elements according to a certain structure, but not the random stacking of these elements. The change of one element in the system will cause the change of other elements, and the change of any element also depends on the change of other elements. The three core elements in TPACK are dynamic and interactive. If any of these elements changes, the other two elements will change to compensate.

4.2. Openness. Knowledge exists and arises in a certain semantic environment. The environment and condition on which this process depends is the knowledge field. Knowledge realizes the acquisition, aggregation, and transformation of information through the interaction between knowledge and knowledge field (environmental conditions). TPACK is an open system, and it is very important to build a benign information environment, which directly affects the input, internal transformation, and output of the system.

4.3. Dynamics. Knowledge is a dynamic mixture of experience, values, background information, expert views, and basic intuition, and its spatial structure changes with time. Teachers' knowledge category, knowledge level, and knowledge relevance always develop and change continuously and affect the structure of knowledge. In different contexts, TK, PK, and CK play different roles, which makes the integration of education and technology widely changeable.

4.4. Hierarchy. The hierarchical structure is the fundamental nature of the system. The level of knowledge is like a pyramid. From bottom to top, the knowledge space is gradually reduced and the complexity level of knowledge is gradually improved. The lower layer is the foundation of the upper layer, and the upper layer is the promotion of the lower layer. TK, PK, and CK are hierarchical knowledge. For example, TK contains two meanings: (1) technology of materialized form and (2) methods and skills of intelligent form. When TK is integrated with PK and CK, it also has hierarchy. The lowest level of integration of informatization technology and teaching is the renewal of teaching media. As a tool of materialization, technology has produced new teaching media during integration.

4.5. Self-Organization. The self-organizing process refers to the spontaneous transformation from disorder to order. This process of transition is from an ordered state to a higher ordered state. Order and disorder are some kinds of structure of the system. The essence of knowledge organization is a series of orderly organizational activities aimed at the disordered state of objective knowledge. The integration process of knowledge is also a spiral process of chaos-order-chaos-order. We expect that teachers' knowledge structure is a self-organizing system with a higher degree of order. When the internal elements of the system reach certain threshold conditions (such as the accumulation of knowledge, the improvement of knowledge level), the internal elements of the system can spontaneously transform to order or more order.

Based on the above, the teacher knowledge system model with systematic structure can be visually represented as the three-dimensional space and hierarchical structure of knowledge, as shown in Figure 3. TPACK comes into being in an open semantic environment, which realizes the acquisition, aggregation, and transformation of information into knowledge. TPACK comes from the multiple interactive integration of three kinds of knowledge. Knowledge is hierarchical. From bottom to top, the knowledge space is gradually reduced and the level of knowledge complexity is gradually improved. When the elements and external environment of TPACK reach certain threshold conditions, it will cause self-organization and promote the transformation of knowledge to order or more order. In addition, when one element of the system changes, in order to maintain the dynamic balance of the whole system, the other two elements will also change to compensate. The lack of elements leads to the weak connection between the elements in the system. This relationship structure is conducive to obtaining information because the weak connection is related to two different types of knowledge. Such two different types of knowledge have different information sources, which is conducive to transmitting fresh or heterogeneous information and knowledge. Figure 4 shows the absence of elements in the TPACK system. Figure 4(a) is missing the content material of TK, so TK and PK, TK and CK constitute a weak connection. Teachers need to obtain new information to supplement the missing parts and re-establish and strengthen the connection. In this system, of the original three knowledge variables, PK and CK become known and independent variables and TK becomes a dependent variable (changing with the combination of PK and CK). Teachers need to design the missing part according to the known content and reset the function of technology according to the purpose of teaching, so as to improve the level of technology application. In Figure 4(b), CK is missing. When PK and TK are independent variables, CK can change multiple results according to different combinations of PK and TK. Even for the same teaching content, when technology is used as tools for presentation, expression and communication, information processing, cognition, and cooperation, different teaching methods can be used accordingly. In Figure 4(c), when there are two missing elements, there are more variable factors, more complex integration methods, and more difficult integration, which can be used as high-level training materials for teachers. By adding the missing part, teachers constantly excavate knowledge and establish the connection of knowledge in the process of self-construction, so as to improve the ability of knowledge integration.

5. AI-TPACK

In the context of AI, teachers should actively use new technologies such as big data to understand students' cognitive state, analyze students' learning characteristics, evaluate students' advantages, potential, and best learning methods, and design personalized learning recommended schemes. We should study how students learn and grow and understand students' noncognitive state, including motivation, emotion, quality, values, etc. Only in this way can we truly respect students and meet their personalized learning and development needs. In the era of AI, does TPACK need to inject new connotation? What new changes will TPACK bring to the teaching form and learning environment after integrating AI technology? In short, re-exploring the relationship among technology, teaching method, and subject content and constructing a new TPACK framework based on AI technology has become an urgent problem to be solved. Isolated AI-TK, AI-PK, and AI-CK are all not enough to support teachers to effectively apply AI technology to teaching. Only the AI-TPACK knowledge formed by the combination of the three is the most effective knowledge basis for teachers to integrate AI technology with classroom teaching. Therefore, there are three main development paths of teachers' AI-TPACK: from PCK to AI-TPACK, from AI-TPK to AI-TPACK, and the synchronous development of AI-TPACK and PCK.

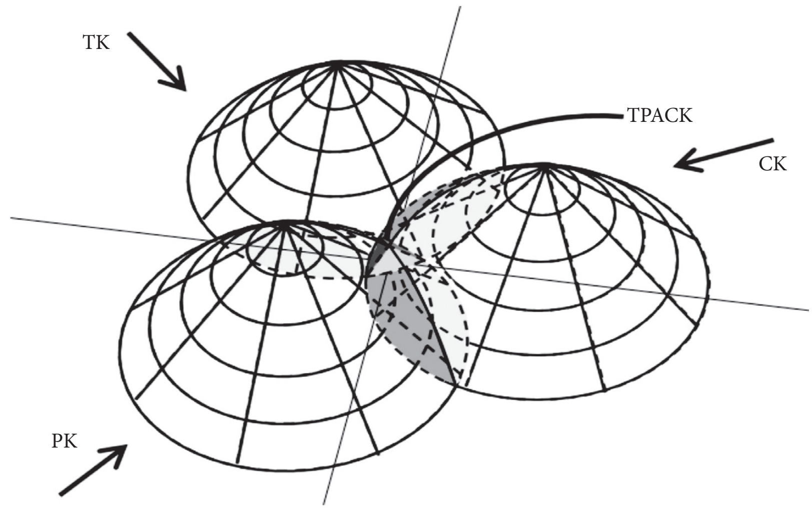


Figure 3: Knowledge system model.

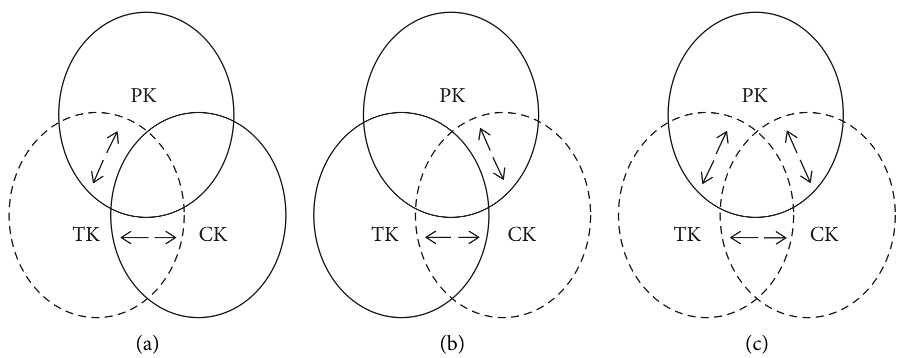


Figure 4: The absence of elements in TPACK system.

5.1. From PCK to AI-TPACK. On the basis of teachers' original PCK, by introducing AI technology, the integration of PCK and AI technology is one of the development ways of teachers' AI-TPACK. For teachers with teaching experience, they not only have a deep understanding of the subject content but also have a lot of accumulation about teaching and learning methods used in the classroom. More importantly, they can quickly establish a relationship between specific teaching contents and teaching and learning methods, that is, they already have rich PCK knowledge. For such teachers, they need to understand the teaching functions of various AI technologies and try to integrate AI technologies into teaching activities related to specific subject contents, so as to realize the development of AI-TPACK. It can be seen that the development from PCK to AI-TPACK is gradual. We can divide this process into five stages: cognition, acceptance, adaptation, exploration, and advancement.

In the cognition stage, teachers begin to understand and use some common AI teaching tools in education and can realize that these tools are suitable for the subject content and teaching activities. In the acceptance stage, teachers will analyze the impact of the application of AI technology on subject teaching, so as to decide whether to adopt AI technology in class. In the adaptation stage, teachers try to select appropriate AI technology to support subject teaching according to the teaching activities. In the exploration stage, teachers will actively integrate AI technology into subject teaching, design the teaching and learning process supported by AI technology in detail, and creatively give full play to the role of AI technology in promoting teaching and learning. In the advancement stage, teachers comprehensively evaluate the teaching and learning supported by AI technology and judge the effectiveness of the integration of AI technology and classroom teaching by analyzing the data information and student feedback in the process of teaching and learning. AI technology has gradually become an integral part of daily teaching and learning activities.

5.2. From AI-TPK to AI-TPACK. Based on the existing AI-TPK, developing teachers' AI-TPACK through the so-called technology mapping method is another feasible path. With the continuous penetration of AI technology in the field of education, teachers have more and more opportunities to learn courses as regards educational application of AI technology in the pre-service education stage. Therefore, we need to pay attention that how to use AI technology to support students' inquiry-based learning and to diagnose students' learning, etc. In this kind of course learning, teachers mainly obtain AI-TPK knowledge, and the technology mapping method can help teachers realize the development from AI-TPK knowledge to AI-TPACK knowledge in subsequent courses.

5.3. Synchronous Development of AI-TPACK and PCK. In order to promote the synchronous development of AI-TPACK and PCK, the following strategies can be adopted: (1) to design artificial products integrating AI technology, such as teaching plans and learning resources supported by AI technology, so as to help learners understand the complexity of the integration of AI technology and classroom teaching in the design process; (2) to critically analyze the design cases of the integration of AI technology and classroom teaching, which promotes learners to deeply think about the relationship among AI technology, teaching methods, and curriculum content; (3) to reflect on the design experience, which can help learners further refine the experience gained in the design process, determine the difficulties encountered, and conduct the self-assessment of AI-TPACK; (4) to apply the design results into the real environment, so as to help teachers further sublimate their understanding about how AI-TPACK to be implemented.

6. Suggestions on Improving Teachers' Informatization Technology Literacy and Application Ability

6.1. Building Teachers' TPACK Belief and Enhancing Their Awareness of the Use of Informatization Technology. Belief is an important factor that determines individual behavior. Teacher knowledge and teacher belief are intertwined. Under the condition of informatization, TPACK is a necessary knowledge framework for teachers. In the classroom, integration technology is a complex and poorly structured problem, including a variety of factors and hovering interactions. There are few fixed and rapid planning methods, which can be applied to various situations and cases. To fully understand the complexity of the integration of teaching and technology, teaching under informatization technology needs teachers' active participation and exploration. New teachers have newer informatization technology knowledge. So the older teachers can learn newer knowledge and technology from new teachers, and new teachers can learn teaching method knowledge and classroom management knowledge from older teachers, so as to promote the integration of informatization technology and classroom teaching and promote the improvement of teachers' TPACK belief. For the development of teachers' TPACK, it is not enough to just train them to master specific operational knowledge. More importantly, we should guide them to form a correct attitude of using technology and the awareness of actively integrating technology.

6.2. Optimizing the Information Resource Environment and Enhancing the Application Ability of Informatization Technology. High-quality informatization technology hardware is the external factor of teachers' TPACK development. Hardware facilities and information resource environment are the premise of technological teaching. Hardware is one of the important reasons for regional differences. In order to speed up the informatization construction, rural schools must provide a good environmental guarantee for the development of teachers' TPACK. Meanwhile, we should provide teachers with a variety of learning resources and constantly update informatization technology resources. Moreover, we can guide teachers to cobuild some technical resource databases for obtaining the required technical support and create an environment for using informatization technology to assist teaching. To effectively improve teachers' understanding of technology and ability of using technology in mathematics teaching, we also construct some teaching strategies, which is helpful for enhancing teachers' confidence about using applied sciences in the actual teaching process, and then will improve teachers' TPACK level.

6.3. Paying Attention to the Selection of Technical Training Mode and Strategy to Enhance the Training Effect. Both pre-service teachers and in-service teachers should make full use of the TPACK framework to formulate corresponding training modes and strategies. The purpose of training is not only to let teachers master technology but also to let teachers learn to use technology for teaching and learn how to integrate technology into teaching, including the integration of technology and teaching content, as well as the integration of technology and teaching method. The choice of training form also has a profound impact on the training effect. Teachers should realize the importance of technical training and understand the focus of training content. Then, the training of hardware equipment and teaching software is carried out, which promotes teachers to achieve a more skilled application level of educational technology. Viewed from the different needs of teachers with different teaching ages, the dimension training of TPACK is performed according to their teaching ages. Besides, typical integration cases of TPACK are selected for in-depth analysis, so as to inspire teachers to master the methods about when to use technology, what technology to use, and how to use technology. Teachers should not only use informatization technology to improve teaching means and methods but also use informatization technology to change the traditional classroom teaching structure. It is important that teachers ought to give full play to the leading role of teachers and reflect the dominant position of students.

6.4. Strengthening the Cooperation between Universities and Primary and Secondary Schools to Promote Mutual Development. Primary and secondary schools should cooperate more with local universities and carry out more exchanges about educational informatization in order to promote the mutual development of both sides. Both sides between universities and primary and secondary schools carry out specific and effective seminars and constantly discuss technology, teaching methods, and teaching knowledge. Through this mutual-assistance mode, the knowledge level of normal students and in-service teachers can be improved, and it can also lay a foundation for normal students to become excellent teachers as soon as possible. With the help of in-service teachers, normal students can deepen their understanding of the subject and learn how to think about the subject content from the perspective of students.

6.5. Strengthening the Integration of Informatization Technology and Mathematics Teaching. In the actual teaching, some teachers blindly use modern informatization technology in pursuit of innovation, which weakens the knowledge penetration in the process of mathematics teaching and cripples the students' understanding of mathematics knowledge. If the process of students' learning mathematical knowledge is completely replaced by computers, it will not be conducive to the cultivation of students' thinking ability. The application of modern informatization technology in mathematics teaching is only an auxiliary means of classroom teaching in order to promote students' mathematical understanding and improve teaching efficiency. We select appropriate technical means for different mathematical contents and find the best bonding point of mathematics teaching and informatization technology in teaching. The integration of informatization technology and curriculum should focus on the improvement of teachers' professional quality, in order to effectively integrate the essence of traditional teaching with the application of informatization technology and handle the relationship between subjectivity and leading. There is a very close relationship between informatization technology and mathematics curriculum. Teachers should try to use STEM courses to integrate the teaching characteristics of informatization technology and mathematics curriculum, formulate scientific and reasonable teaching plans according to teaching contents and teaching requirements, display mathematical knowledge with the help of informatization technology, and then carry out informatization technology education around mathematical concepts to realize the in-depth integration between disciplines.

With knowledge reformation and technological innovation, a new era of education based on big data and Internet+ has come. In addition, how to delimitate the influencing factors of TPACK and the collaborative relationship between the influencing factors and the constituent elements of TPACK, as shown in Figure 5, will become an important breakthrough in the field of TPACK in the future. For MPCK, informatization technology has become an indispensable tool, which is very important to mathematics teaching. The deep integration of informatization technology and mathematics teaching has become a beautiful vision for the development of mathematics education. Under the background of the close combination of informatization technology and mathematics teaching, how to improve teachers' professional quality, how to better serve mathematics teaching with informatization technology, and how to promote students' mathematical understanding are topics that we should explore deeply.

7. Discussion

The research hotspots in the field of TPACK mainly include informatization technology, teacher professional development, teacher education, teacher knowledge, and pre-service teachers. From the perspective of centrality, in recent years, the research in the field of TPACK has mainly focused on teacher education and curriculum construction under the background of informatization technology. As the designer of integrating informatization technology, teachers should find the entry point between informatization technology and classroom teaching, so as to make informatization technology effectively integrate into teaching. In teaching practice, through the integration of technical means, teaching methods, and subject content, teachers create a multimodal learning environment, which can not only enhance their ability of technology learning and technology integration in the real situation but also design teaching activities that help improve learners' ability development. All of these contribute to promote the change of teachers' teaching behavior and the optimization of teaching process.

In addition, formally, the TPACK framework composed of three knowledge rings and a context is relatively easy to understand. From a theoretical point of view, TPACK has a certain complexity: It originates from the concept of PCK containing uncertainty and is essentially a fuzzy and complex concept; as a top-level design model, TPACK only provides teachers with a way of thinking and does not tell them what to do and how to do. A framework is required to satisfy the simplicity of the structure of basic model and reflect the complexity of the essence, and then the theoretical tension of a perfect framework will be unavoidable. Developing brand-new TPACK through design-based learning, design thinking, and knowledge creation is a new research perspective.

AI-TPACK goes beyond TPACK, which looks at technology from the tool level. The connotation of AI-TPACK is the interactive relationship among AI technology, subject content, and teaching methods under human-machine cooperation. The AI-TPACK theoretical framework has certain reference significance for the development of teachers' application ability of AI technology, the integration of AI technology and curriculum teaching practice, and the development of intelligent learning environment. Teachers' development is a long-term process. Teachers should have the consciousness of practice and reflection and actively develop independently. Academic groups should help each other and make mutual improvement. Schools should provide institutional and informatization technological support. Only by doing so can teachers change their role as soon as possible and play the role of multidirectional collaborator.

8. Future Works

In future works, the following topics may become irreversible trends:

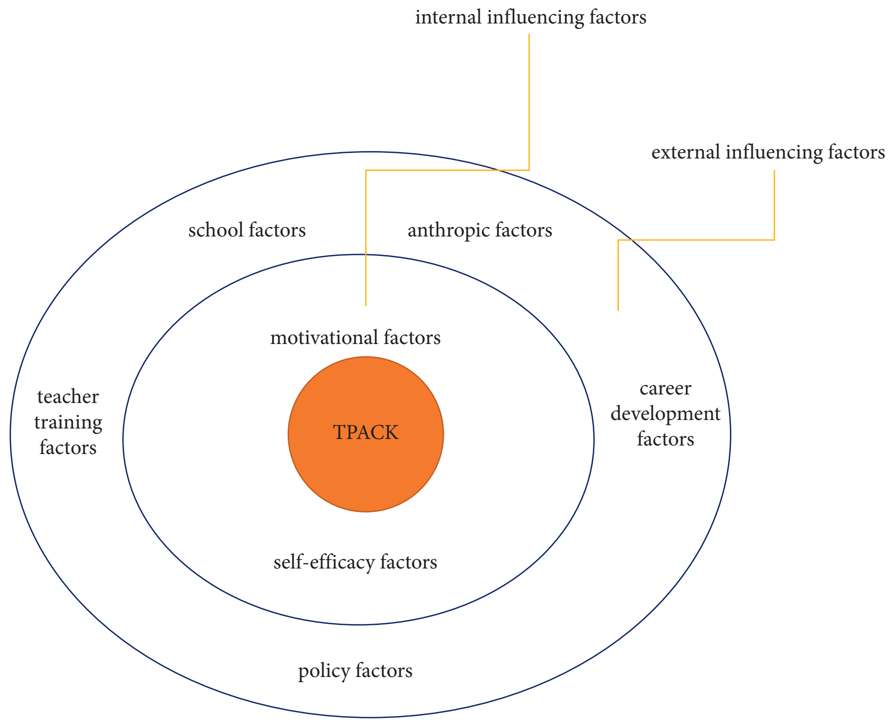


Figure 5: Influencing factors of teachers' TPACK.

(1) Strengthening the subjectivity research on students' closely related to teachers' TPACK. Using educational informatization technology to optimize classroom teaching is the purpose of teachers' TPACK internalization. Therefore, in the future, we can measure teachers' TPACK level from the perspective of students and think about the relationship between TPACK and students' knowledge representation, knowledge system construction, academic achievement, and problem-solving ability. Combined with students' physiological characteristics and cognitive laws, we will explore the impact of teachers' TPACK level on students with different learning stages and characteristics in order to further promote the process of educational informatization.

(2) With the development of informatization technology, great changes have taken place in the way of teaching activities. From the aspect of teaching by educators, teachers' mastery and flexible application of TK in TPACK have become necessary knowledge and skills. The rapid development of informatization technology will cause revolution in the field of education and improve the timeliness of TPACK essence, which is also an important direction of TPACK integrating technology. In addition, one should also pay attention to the practice of informatization technology and curriculum integration based on teaching. The integration strategy of informatization technology and curriculum can start from the entry of knowledge points, the construction of teaching environment, the way of cooperative exploration, and the construction of teaching resources. The specific integration teaching practice can be carried out through group cooperative learning, expanding learning resources, and carrying out diversified evaluation.

(3) For different types of teachers, AI-TPACK promotion strategies should also be different. We should not only continue to pay attention to the development path of AI-TPACK for teachers with different knowledge and experience but also explore the promotion strategies of AI-TPACK for teachers with different cognitive structures, different subject backgrounds, different teaching beliefs, and different thinking ways.

(4) The improvement and expansion of TPACK theoretical framework not only provides a more scientific reference framework for teachers' educational beliefs, knowledge, and skills but also brings important enlightenment to the resource allocation, teaching design, and strategy selection of teaching practice. Therefore, we should take some effective steps for the measurement and evaluation of TPACK framework. The measurement and evaluation of TPACK framework has become an important link in the teaching process. Simultaneously, combined with the application of various measurement tools, we ought to promote the improvement of teachers' knowledge and skills and the teaching activities. In addition, the application research of teacher educational training and instructional design based on TPACK theory is an important direction. The guiding role of theoretical results should be played to realize the development of teaching practice.

9. Concluding Remarks

With the continuous improvement of China's educational informatization level, there are more expectations and challenges for future teachers' knowledge structure and teaching literacy. Under the environment of modern informatization technology, how to better cultivate the knowledge, skills, and professional quality of pedagogics has a direct impact on the future of teacher education reform in China. This paper aims to explore and understand the knowledge required for teachers to effectively integrate technology in a specific pedagogical content field, including all levels from teachers' professional quality to teachers' teaching practice. In addition, this paper enriches and deepens the research of TPACK in the field of teacher education at the two levels of theory and practice.

Data Availability

No data were used to support this study.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

References

[1] J. M. Santos and R. D. R. Castro, "Technological Pedagogical content knowledge (TPACK) in action: application of learning in the classroom by pre-service teachers (PST)," *Social Sciences & Humanities Open*, vol. 3, no. 1, Article ID 100110, 2021.

[2] H. Ozgur, "Relationships between teachers' technostress, technological pedagogical content knowledge (TPACK), school support and demographic variables: a structural equation modeling," *Computers in Human Behavior*, vol. 112, Article ID 106468, 2020.

[3] H. E. Alharbi, "An Arabic assessment tool to measure technological pedagogical and content knowledge," *Computers & Education*, vol. 142, Article ID 103650, 2019.

[4] D. Akyuz, "Measuring technological pedagogical content knowledge (TPACK) through performance assessment," *Computers & Education*, vol. 125, pp. 212–225, 2018.

[5] M. Schmid, E. Brianza, and D. Petko, "Self-reported technological pedagogical content knowledge (TPACK) of pre-service teachers in relation to digital technology use in lesson plans," *Computers in Human Behavior*, vol. 115, Article ID 106586, 2021.

[6] L. M. Archambault and J. H. Barnett, "Revisiting technological pedagogical content knowledge: exploring the TPACK framework," *Computers & Education*, vol. 55, no. 4, pp. 1656–1662, 2010.

[7] C. R. Graham, "Theoretical considerations for understanding technological pedagogical content knowledge (TPACK)," *Computers & Education*, vol. 57, no. 3, pp. 1953–1960, 2011.

[8] B. Kramarski and T. Michalsky, "Preparing preservice teachers for self-regulated learning in the context of technological pedagogical content knowledge," *Learning and Instruction*, vol. 20, no. 5, pp. 434–447, 2010.

[9] J. H. L. Koh and C. S. Chai, "Seven design frames that teachers use when considering technological pedagogical content knowledge (TPACK)," *Computers & Education*, vol. 102, pp. 244–257, 2016.

[10] Q. Zhu, J. Aparicio, F. Li, J. Wu, and G. Kou, "Determining closest targets on the extended facet production possibility set in data envelopment analysis: modeling and computational aspects," *European Journal of Operational Research*, vol. 296, no. 3, pp. 927–939, 2022.

[11] F. Baier and M. Kunter, "Construction and validation of a test to assess (pre-service) teachers' technological pedagogical knowledge (TPK)," *Studies In Educational Evaluation*, vol. 67, Article ID 100936, 2020.

[12] C. H. Tosunoglu and N. G. Lederman, "Developing an instrument to assess pedagogical content knowledge for biological socioscientific issues," *Teaching and Teacher Education*, vol. 97, Article ID 103217, 2021.

Hindawi
Complexity
Volume 2021, Article ID 3724919, 13 pages
https://doi.org/10.1155/2021/3724919

Research Article

Logistic Regression Analysis for Spatial Patterns of Drought Persistence

Rizwan Niaz,[1] **Xiang Zhang**,[2,3] **Nouman Iqbal**,[1] **Mohammed M.A. Almazah**,[4,5] **Tajammal Hussain**,[6] **and Ijaz Hussain**[1]

[1]*Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan*
[2]*National Engineering Research Center of Geographic Information System, School of Geography and Information Engineering, China University of Geosciences (Wuhan), Wuhan 430074, China*
[3]*State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing (LIESMARS), Wuhan University, Wuhan 430079, China*
[4]*Department of Mathematics, College of Sciences and Arts (Muhyil), King Khalid University, Muhyil 61421, Saudi Arabia*
[5]*Department of Mathematics and Computer, College of Sciences, Ibb University, Ibb 70270, Yemen*
[6]*Department of Statistics, COMSATS University Islamabad, Lahore Campus, Lahore 54000, Pakistan*

Correspondence should be addressed to Xiang Zhang; zhangxiangsw@whu.edu.cn

Received 27 August 2021; Revised 26 September 2021; Accepted 26 October 2021; Published 22 November 2021

Academic Editor: Feng Li

Drought is one of the natural hazards with potentially significant impacts on society, economy, and other natural resources over the globe. However, the understanding of drought characteristics and its persistence can significantly help to reduce the potential impacts of drought. Moreover, the knowledge about the spatiotemporal pattern of seasonal drought frequency and drought persistence is important for water resource management, agricultural development, energy consumption, and crop yields. Therefore, the present study is employed to examine the seasonal drought frequency and drought persistence in the region. In this regard, the standardized precipitation index (SPI) at the three-month time scale was used to determine meteorological drought. Furthermore, the logistic regression model is used to calculate the odds and probability of drought persistence from one season to the next for the selected stations by identifying the spatial pattern of seasonal drought frequency and persistence. The potential of the current analysis is validated on six selected stations of the northern area of Pakistan. The outcomes related to the current analysis provide the basis for taking more considerations on early warning systems and help to make the valuable decision for water resource management and agriculture sectors in Pakistan.

1. Introduction

Drought is a complex phenomenon and one of the greatest recurring natural disasters. It causes substantial losses in agriculture sectors [1–6], natural ecosystems [7–9], and forestry [10–12]. It is also called a creeping phenomenon with its influences gradually take in an area over a period and may persist for a long period. In severe cases, drought can last for many years and disturb many social activities, agricultural sectors socioeconomic sectors, and environmental settings [6, 13–16]. The arrival of drought occurrences is complicated; therefore, it is challenging to substantiate the onset of the drought occurrences. The complication in drought occurrence becomes a cause for new methodologies for the assessment of drought [17–19]. However, the effects of droughts have been described in several studies, and various extensive measures are proposed to hold them. However, without proper actions, drought consequences can be severe for the natural environment and sustained for a long period even after termination [20–22].

The speedy improvement in the drought monitoring policies brought several advantages to society [17, 18, 23]. To understand the complexity and improve methodologies of drought monitoring, the researchers have been grouping the drought in several aspects including meteorological, agricultural hydrological, and socioeconomic aspects [24, 25]. Rhee and Im [26] described the meteorological drought as a lack of precipitation in a prolonged period [27, 28]. The agricultural and hydrological droughts can be defined as periods with insufficient soil moisture and river flow or groundwater, respectively [29]. The drought-related socioeconomic occurs due to a shortage in water resource systems. In this case, the water demands are not fully attained by the water supply. The water supply is unable to provide sufficient water at a particular time for society. Since last recent years, droughts cause major factors that distress the society and its economic environment sectors around the world more than other natural hazards [30]. However, the disaster can be minimized by adopting suitable approaches to assess drought occurrences. The appropriate approaches can help to mitigate drought impacts and provide significant outcomes for further drought monitoring and mitigation policies.

In the literature, researchers added several indices for the improvement of drought mentoring policy and forecasting strategies. For instance, for identifying various drought classes in the selected area, Van Rooy, [31] proposed a rainfall anomaly index (RAI) which is designed particularly for the classification of drought classes. Furthermore, a study measured the cumulative deviation of moisture supply and added new research Palmer drought severity index (PDSI) in the literature for monitoring drought more accurately [32]. Many other drought indices were added in the literature for increasing the efficiency and accuracy to determine drought occurrences more accurately and precisely. The drought indices including Crop Moisture Index (CMI) proposed in [33], Bhalme and Mooly Drought Index (BMDI) in [34], Reclamation Drought Index (RDI) introduced in [35], National Rainfall Index (NRI) in [36], Surface-Water Supply Index (SWSI) developed in [37, 38] have proposed Standardized Precipitation Index (SPI). The SPI is prevalent among other indices and frequently used for drought monitoring and forecasting [39, 40]. The obtained standardized values of SPI can be used to compare droughts in distinct climatic zones across regions. Therefore, the SPI is commonly used for assessing drought occurrences and developing new methodologies for drought monitoring and forecasting [41–43].

The current study aimed to examine the drought persistence in the selected stations. The study used logistic regression to compute odds and probability of drought persistence from one season to the next in the selected region by identifying the spatial pattern of seasonal drought frequency and persistence. The standardized precipitation index (SPI) at the three-month time scale was employed to characterize meteorological drought. Furthermore, the potential of the present study is validated on six selected stations of the northern area of Pakistan. The results associated with the present study provide the basis for taking more attention to the early warning systems and making improvements for water resource management and agriculture sectors to observe and substantiate the severity of drought in Pakistan.

2. Methods

2.1. Standardized Precipitation Index (SPI). The various studies consider varying drought indices to categorize and monitor drought occurrences [39, 41, 43]. Since the last decade, the use of SPI is common for drought monitoring [41, 44]. The present study considers the SPI developed by [38] for the estimation of drought occurrences. The time-series data ranging from January 1971 to December 2017 is used for drought characterization. The consistency of SPI is significant in climatic circumstances relating to the geographical and temporal distribution [39, 40]. Furthermore, the simplicity in calculation makes SPI more prominent and recognized worldwide for characterization. The SPI can be computed for various times based on monthly data to observe the meteorological drought [39].

2.2. Logistic Regression Model. The drought is assessed based on SPI for the selected stations. The seasonal drought is analyzed using SPI at a three-month time scale for each selected station. The seasonal drought frequency is calculated as the total number of years in which drought SPI ≤ 1 appears in each climate division of the selected station. The climate divisions are included as follows: (1) "December, January, February" (Winter); (2) "March, April, May" (Spring); (3) "June, July, August" (Summer); (4) "September, October, November" (Autumn). For instance, if the climate brings 10 years in which drought occurs in winter during the total number of 47 years, the drought frequency probability is 10/47 or 21%. Furthermore, the persistence of seasonal drought can be defined as the drought conditions staying from one season to the next. In the current analysis, drought persistence is defined as winter-spring, spring-summer, summer-autumn, and autumn-winter. For each station, the seasonal drought persistence and season combination can be expressed as the probability of persistence. The probability of persistence can be calculated as the total number of years in which drought persists from one season to season combination by dividing the total number of years in which drought appears during the first season in that climate division. For example, if the climate shows 12 years in which drought occurs in winter and drought conditions continue into spring during the 10 of those years, the drought persistence probability is 10/12 or 83%. Furthermore, in the present study, SPI for scale three is considered as a binary variable: drought occurs = 1 and drought does not occur = 0. The drought persistence can be calculated to define the impact of the influence of one season's moisture conditions on moisture conditions in the forthcoming season (Ford and Labosier, 2013). Moreover, in the current study, the logistic regression model is employed to examine the persistence of drought from one season to the coming season [45]. The logistic regression model is frequently used to examine the relationship between the binary dependent variable and one or more independent variables (Ford and Labosier, 2013). Kutner et al. [46]; stated that using an ordinary least squares regression with a binary dependent variable can lead to several issues including nonnormally distributed errors and predicted values in which range does not occur from 0 to 1. Therefore, logistic regression is the more suitable method for analyzing a binary dependent variable. Furthermore, various studies used a logistic regression model to predict the logit transformation of the dependent variable based on the independent variables [46–49]. The logistic regression model can be considered in the following form:

$$F(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}, \quad (1)$$

where t is a linear function of an explanatory variable x, and the logistic function can be written as

$$\pi(x) = \frac{e^{\alpha+\beta x}}{e^{\alpha+\beta x} + 1} = \frac{1}{1 + e^{-\alpha+\beta x}}. \quad (2)$$

The inverse of the logistic function, the logit model, is

$$g(x) = \frac{\rho}{1-\rho} = \alpha + \beta x, \quad (3)$$

where the odds ratio of the drought occurrence is denoted by $(\rho/(1-\rho))$, α shows the intercept, and β denotes the regression coefficient. Furthermore, the normally distributed residuals are not required in the logistic regression model. However, the error distribution between actual and predicted dependent variables follows binomial (Ford and Labosier, 2013). In the current scenario, for SPI at scale 3, the observations were perceived independently, and the binominal assumption meets [45, 47]. The use of the logistic function is beneficial as it can be considered an input with any value from negative infinity to positive infinity, whereas the output ρ is remained to values between 0 and 1 and thus is explainable as a probability. The current study employed a logistic regression model for the specific season based on a binary dependent variable which is drought occurrence (1 = drought, 0 = no drought) and the explanatory variable is the SPI value from the previous season. The outcome of the model can be described for the given season as the log odds ratio of drought occurrence, and the relationship between the dependent and explanatory variables is represented by the slope (β) of the logistic regression. For instance, when β is observed negative, for the specific season, the odds of drought occurrence are negatively related to the SPI value in the previous season. In a more precise way, it can be described that the rise in the preceding season's SPI causes to decline in the odds of drought occurring in the subsequent season.

3. Application

The structural importance of the Northern region is important for the other regions of the country. The northern region is a geographic area that has a group of three mountain ranges, the Himalayas, Karakoram, and the Hindu Kush, which cover most of the region [50]. Many of the world's tallest peaks are found in this region, including *K*-2, Nanga Parbat, and Rakaposhi. The average altitude of Karakorum is 6,100 m, Hindu Kush is 7,690 m, and the Himalayas is 8,848 m [51]. These high altitudes of mountains frequently deliver a significant portion of precipitation [50–52]. The reservoirs of the other regions are also associated with the rainfall of this region. Due to the significant importance of the region for the other parts of the country, the region is selected for analysis. Furthermore, the climatic situations of the chosen region placed vast and persistent effects on climatic environments and weather conditions of several parts of the country. The current analysis is accomplished with time-series data from January 1971 to December 2017 (see [53] for the selected region of the northern area of Pakistan (Figure 1)). The weather pattern fluctuates over the varying periods within the country which brings great socioeconomic influences. These changes in the atmosphere cause disturbance for the agriculture sectors and reservoir system of the country [54, 55]. There are various climatic events related to global climate warming. These climatic events including rainstorms, high temperatures, and droughts are commonly considered for influencing global climate warming. However, the fluctuation of global warming becomes the source of wide-ranging impacts on the atmosphere and people in Pakistan [56–58]. Therefore, it is required greater considerations to recognize the drought incidents more instantly by adding improvements in the drought monitoring processes and methods. In this regard, the current study used logistic regression to evaluate the odds and probability of drought persistence from one season to the next in the selected region by recognizing the spatial pattern of seasonal drought frequency and persistence. The SPI at the three-month time scale is applied to characterize meteorological drought. The current study efficiently improves the capability of identifying the spatial pattern of seasonal drought frequency and persistence. The outcomes of the current analysis can be helpful for drought monitoring and mitigation policies in the selected region.

3.1. Results. The six stations were selected for the current analysis. These stations were chosen from the northern area of Pakistan. The climate of the stations is important for the other parts of the regions. Therefore, these stations were selected in the current analysis. The brief climatology of the selected stations for precipitation is given in Table 1. The precipitation of the selected stations is used for the calculation of the SPI index. The importance of SPI is significantly high from the drought monitoring perspective. The SPI values were standardized by using varying probability distributions. Among the various probability distributions, the distributions for the varying stations were selected based on minimum Bayesian information criterion (BIC) values. The selected distribution and their BIC values are provided in Table 2. For instance, at the three-month time scale, the gamma distribution shows suitable candidacy for the Astore station. The BIC of gamma of Astore is −1279.0507 which is minimum among other distributions. Therefore, gamma distribution is used for standardization in the Astore station. The gamma distribution is also suitable for two other stations. For Bunji station, the BIC value of gamma distribution is −824.8686 and for Gupis, it has a BIC value is −1264.9476. For the Chilas station, the BIC of 3p Weibull is minimum

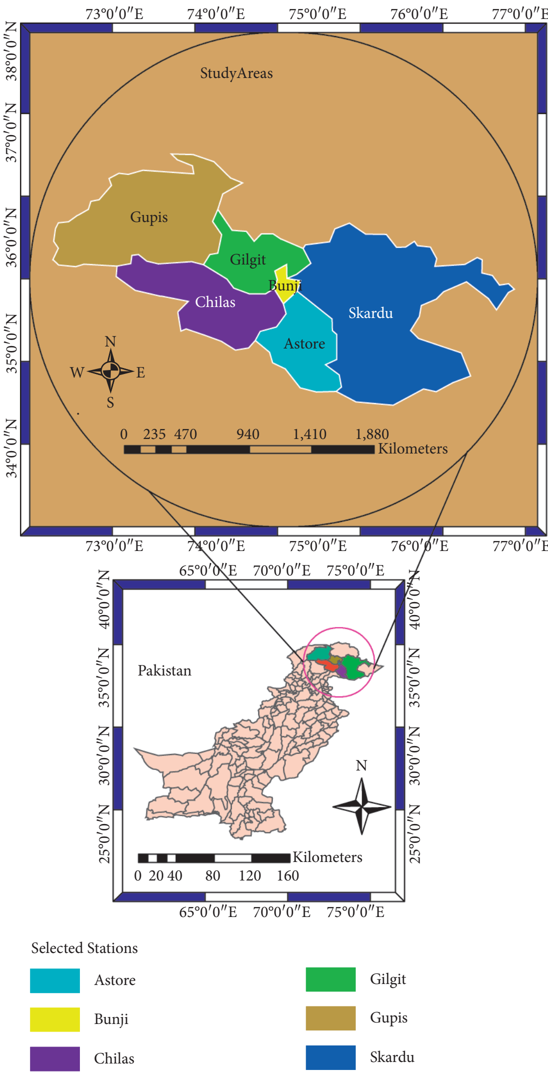


Figure 1: Selected geographical locations from the northern areas of Pakistan.

Table 1: The precipitation characteristics during the period 1971–2017 of six selected stations.

Variable	Station	Mean	1st quartile	Median	3rd quartile	Kurtosis	Std. dev
Precipitation	Astore	39.34	10.80	25.70	52.62	3.01	41.93
	Bunji	13.56	1.30	7.10	17.10	7.55	18.90
	Gupis	15.94	0.00	5.70	19.38	51.38	30.21
	Chilas	15.85	0.95	7.00	19.32	8.88	23.53
	Gilgit	11.75	1.10	6.05	14.72	9.93	16.57
	Skardu	19.51	2.30	9.10	26.75	5.60	25.90

among other distributions and selected for the standardization. The exponential distribution is suitable for the Gilgit station with minimum BIC (−810.2190). In Skardu station, the precipitation data is used for the standardization based on the generalized extreme value (GEV) distribution. The GEV has minimum BIC among other distributions, and therefore, selected for the standardization. Moreover, after standardization, the next step is the drought characterization which is done by the classified values of SPI.

Drought frequency is calculated by the total number of months with SPI ≤1 divided by the total number of months from January 1971 to December 2017. In this way for the current analysis, the seasonal drought frequency is calculated. To calculate seasonal drought frequency, the percentage of seasons in which drought SPI ≤1 occurs over the whole studied period for winter, spring, summer, and autumn, respectively. Furthermore, drought is a complicated phenomenon, and its impacts affect the agriculture,

Table 2: The selected probability distributions of SPI-3 and their BIC values for selected stations.

	Astore		Bunji		Gupis		Chilas		Gilgit		Skardu	
Index	Distribution	BIC	Distribution	BIC	Distribution	BIC	Distribution	BIC	Distribution	BIC	Distribution	BIC
SPI	Gamma	−1279.0507	Gamma	−824.8686	Gamma	−1264.9476	3p Weibull	−810.2973	Exponential	−810.2190	Generalized extreme value	−653.9913

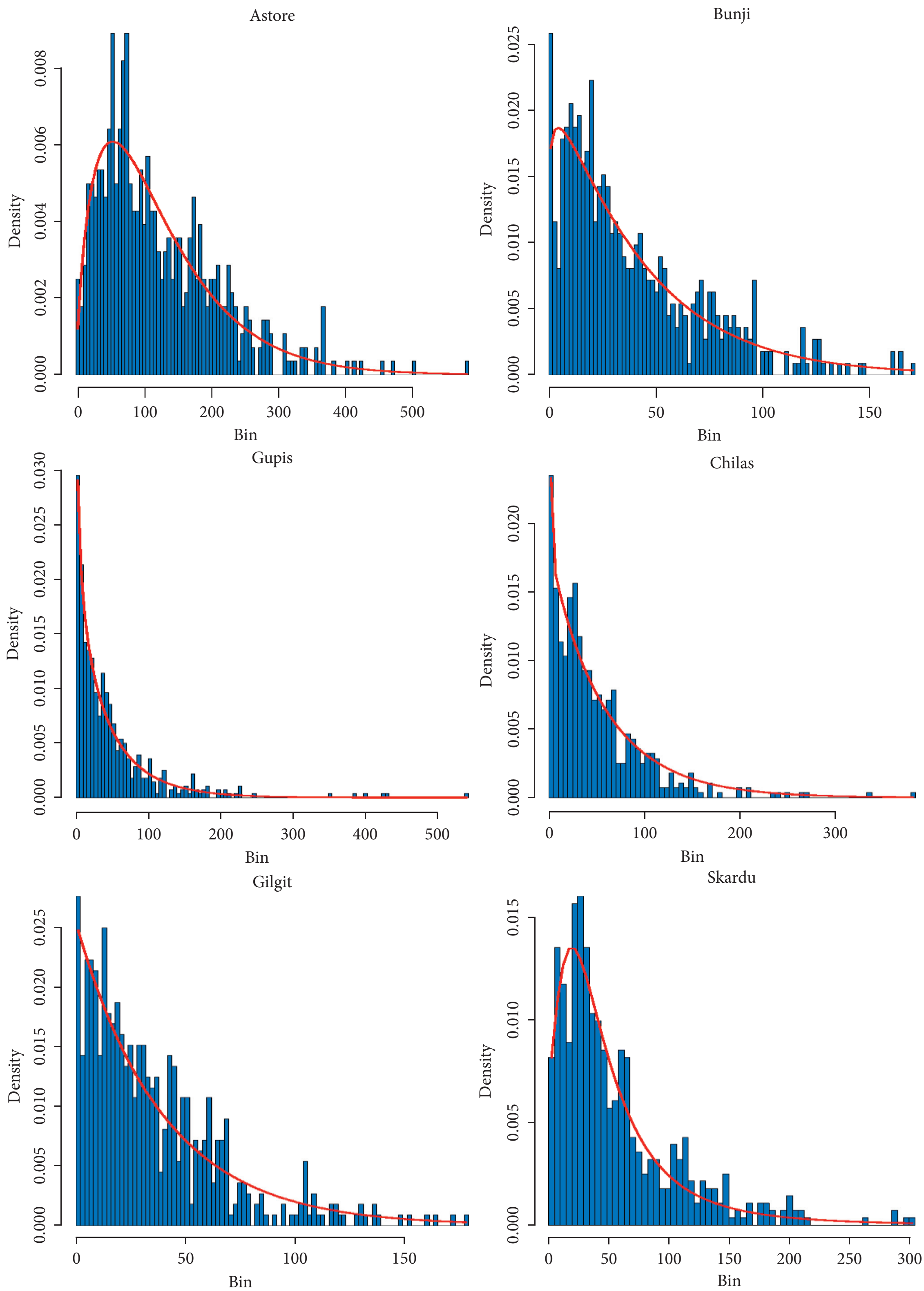


Figure 2: The varying distributions are selected for the six selected stations. The histograms (theoretical vs. empirical) of the selected distributions are presented for SPI-3.

economy, and other social activities. If meteorological drought sustains a prolonged time, it can be hazardous for soil moisture and evapotranspiration and lead to agriculture and hydrological drought. Furthermore, information about seasonal drought frequency and drought persistence are essential for water resource management,

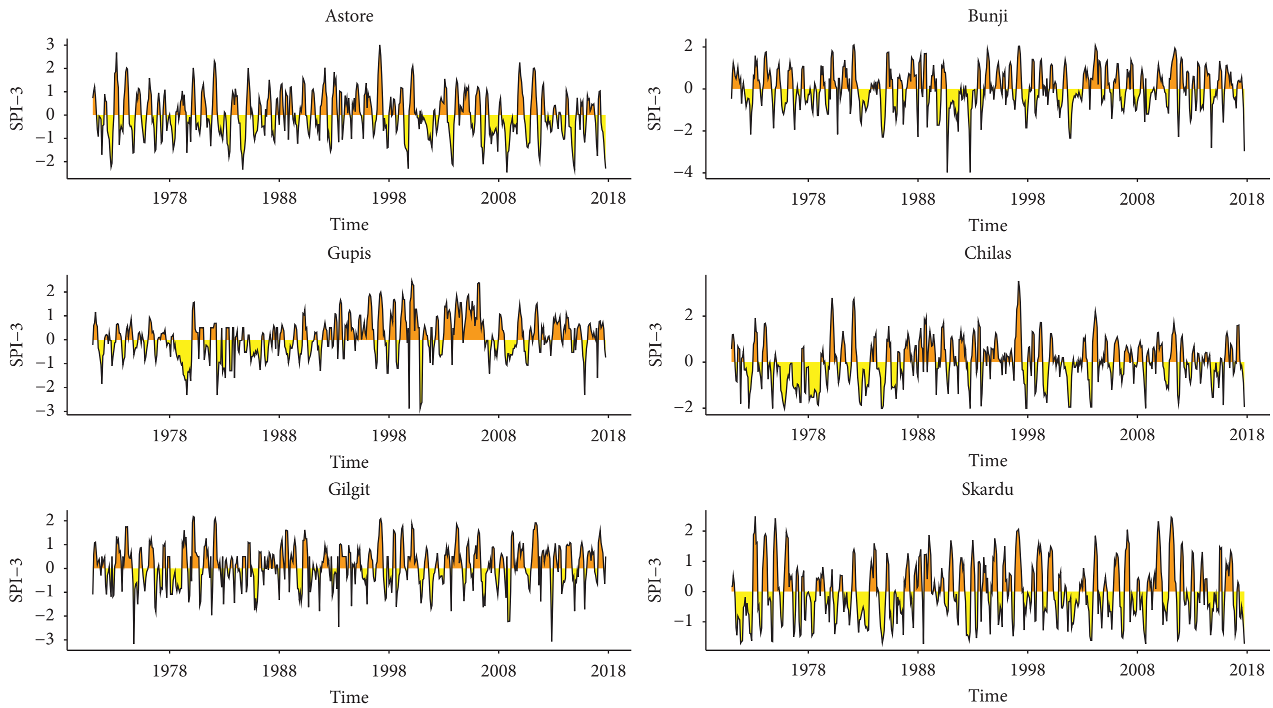


Figure 3: The temporal plots for SPI-3 at selected stations.

agricultural development, and energy consumption. Therefore, the study is applied to determine the frequency of meteorological drought and its persistence in various seasons in the selected stations. The seasonal drought persistence can be expressed by the probability of the persistence and computed as the total number of years in which drought remains from one season to another divided by the overall years in which drought occurs during the first season. Furthermore, the logistic regression is used to compute odds and probability of drought persistence from one season to the next in the selected stations by finding the spatial pattern of seasonal drought frequency and persistence. For example, the logistic regression model is used for winter-spring in the selected stations separately to observe odds for drought occurrences in the current season based on the SPI values of the previous season. The logistic regression model for selected stations based on spring-summer is used to observe drought occurrences. Similarly, the outcomes can be obtained from the logistic regression model for summer-autumn and autumn-winter. Moreover, the results obtained from the various seasons using logistic regression and their spatial characteristics are described by the maps, and the temporal aspects are demonstrated by the graphs.

Moreover, in Figure 2, theoretical and empirical distributions for SPI at a three-month time scale (SPI-3) are presented, and the temporal behavior of selected stations for SPI-3 can be observed in Figure 3. Figure 4 shows the temporal variation for the numbers of months/year with SPI≤1. Figure 5 represents the counts of drought occurrence in the selected stations in various seasons. It can be observed that the ND category prevails among other categories in various seasons of the selected stations. Furthermore, SPI ≤1 specifies the drought in the selected stations and greater than 1 shows that there is no drought. The selected stations their latitude and longitude and average months of SPI≤1 is presented in Figure 6. These average months show that the overall, number of months with SPI≤1 in each year varies 6 to 12 months with an average of 10 in the Astore station from the time-series data ranging from January 1971 to December 2017. However, in Bunji station, the number of months with SPI ≤1 in each year varies from 6 to 12 months with an average of 9.8085. Furthermore, average months' SPI ≤1 for the other selected stations is also calculated based on the same rationale. Seasonal drought frequency is calculated as percentage of seasons in which drought (SPI) occurs over the whole study period for winter, spring, summer, and fall, respectively, and are presented in Figure 7. The probability maps in Figure 8 demonstrate the stations that are more vulnerable to prolonged meteorological drought. For example, 90 to 100% summer drought persistence probability in summer-autumn means that 90 to 100% of summer meteorological drought in selected stations of the northern areas of Pakistan can persist to the autumn. It can be observed that there are large spatial variations in seasonal drought persistence. A strong difference is shown in various seasons with stations. The Skardu station has a lower probability of autumn-winter drought persistence than other stations. Drought persistence probabilities in spring-summer and summer-autumn show a more similar pattern among the selected stations. This indicates that, in these seasons, the precipitation occurrences are not sufficient. Furthermore, the drought persistence probabilities in northern areas during summer to autumn are higher than in

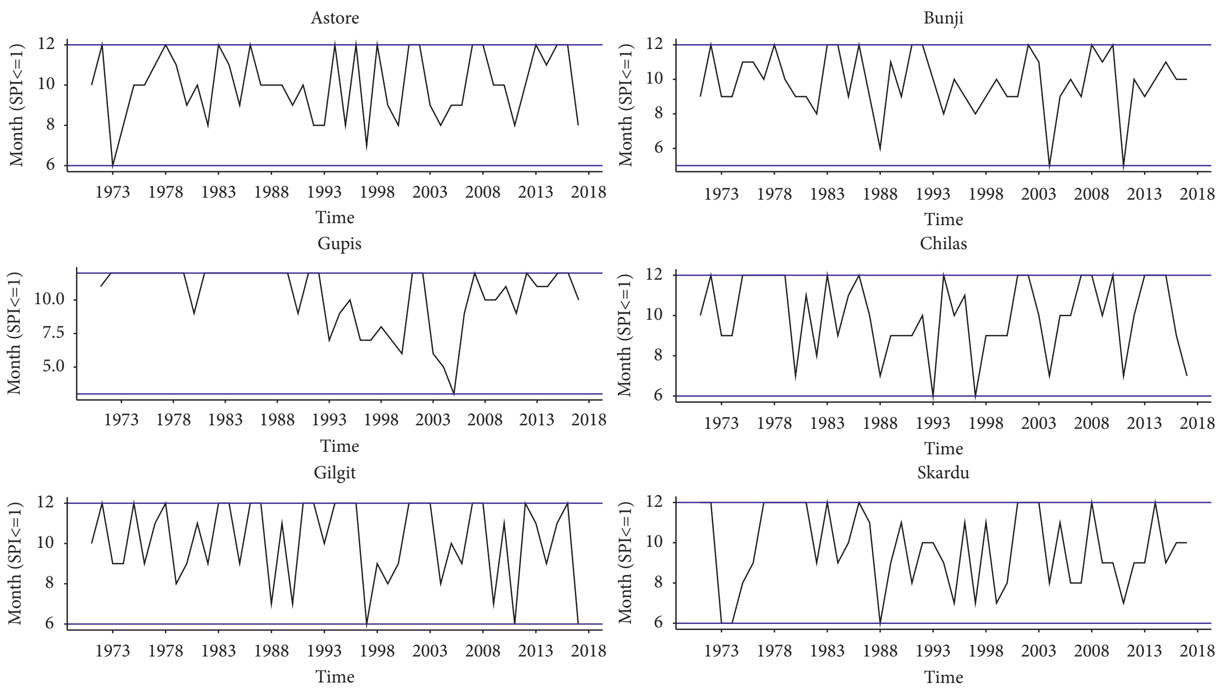


Figure 4: Time series plot: temporal behavior of SPI ≤1 in selected stations.

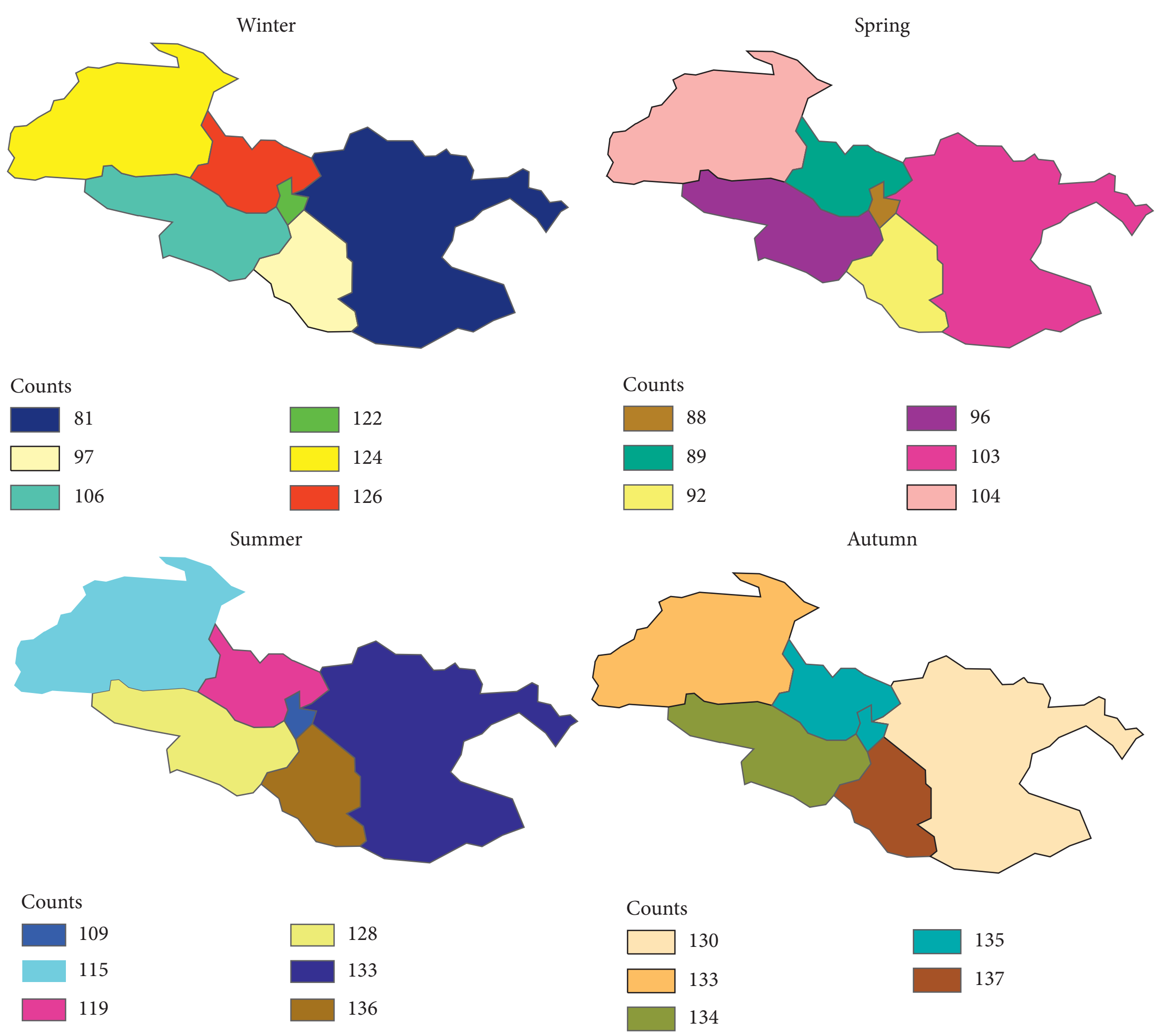


Figure 5: The total number of counts with SPI ≤1 in the selected stations for SPI-3 in various seasons.

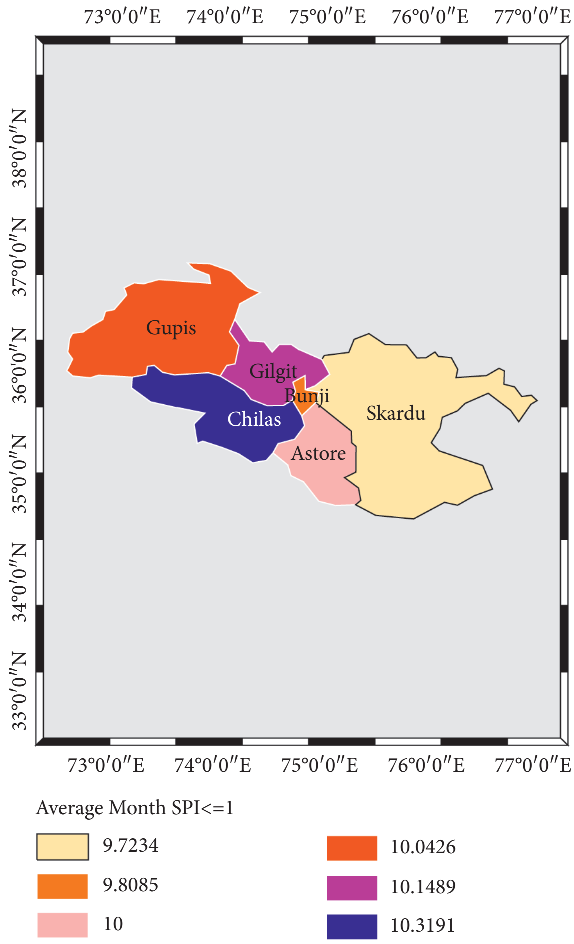


Figure 6: The latitude and longitude of the selected stations and their monthly averages for SPI ≤1.

other seasons. Seasonal drought persistence odds ratios output from the logistic regression model presented in various maps of Figure 9. This shows drought occurrence odds ratio changes as a function of a 1 unit increase in the previous season's SPI. The stations with an odds ratio <1 indicate a decreased odds ratio of drought occurrence in the given season when the SPI of the previous increases while those stations with odds ratio >1 showing that a positive increase in the odds ratio of drought occurrence in the given season when previous season's SPI increases. Furthermore, the significant odds ratio can be observed in the maps. The stations outline bolded are those at which the logistic regression fit is statistically significant at $\alpha < 0.05$. For instance, on Gupis station in spring to summer, the odds ratio is significant and shows that as increases 1 unit in the SPI of spring the odds of drought occurrence are decreasing in summer by 0.20 to <1%. Furthermore, the relationship between various seasons and SPI is negative in selected stations. This means for increases 1 unit in the SPI, the odds of drought occurrence are negatively influenced. Similarly, the odds ratio of drought occurrence can be observed for other stations and seasons. Based on the obtained results drought early warning policies should be prepared in the selected stations accordingly. Moreover, the obtained results from the recent evaluation about seasonal drought frequency and drought persistence can be helpful to water resource management, agricultural development, energy consumption, and crop yields.

3.2. Discussion. The data of precipitation is used for the classification of drought occurrences. These drought occurrences are used to classify the drought categories in the selected region. The classified values of the drought are calculated using SPI-3. The SPI provides the standardized values for the given precipitation in the selected station for monthly data. The varying probability distributions were used for the standardization of the precipitation values for SPI. The selection of the distributions is based on BIC values. Information related to seasonal drought frequency and drought persistence play a vital role in water resource management, agricultural development, energy consumption, and crop yields. Despite their importance to the water resource management and agriculture sectors, there are very few studies available in the literature that focus on seasonal drought frequency and drought persistence [45]. Therefore, dynamic procedures are required in drought monitoring for providing relief to affected areas. Suitable drought monitoring is a key factor in preparation for water resources and disaster mitigation and making a climate-resilient society

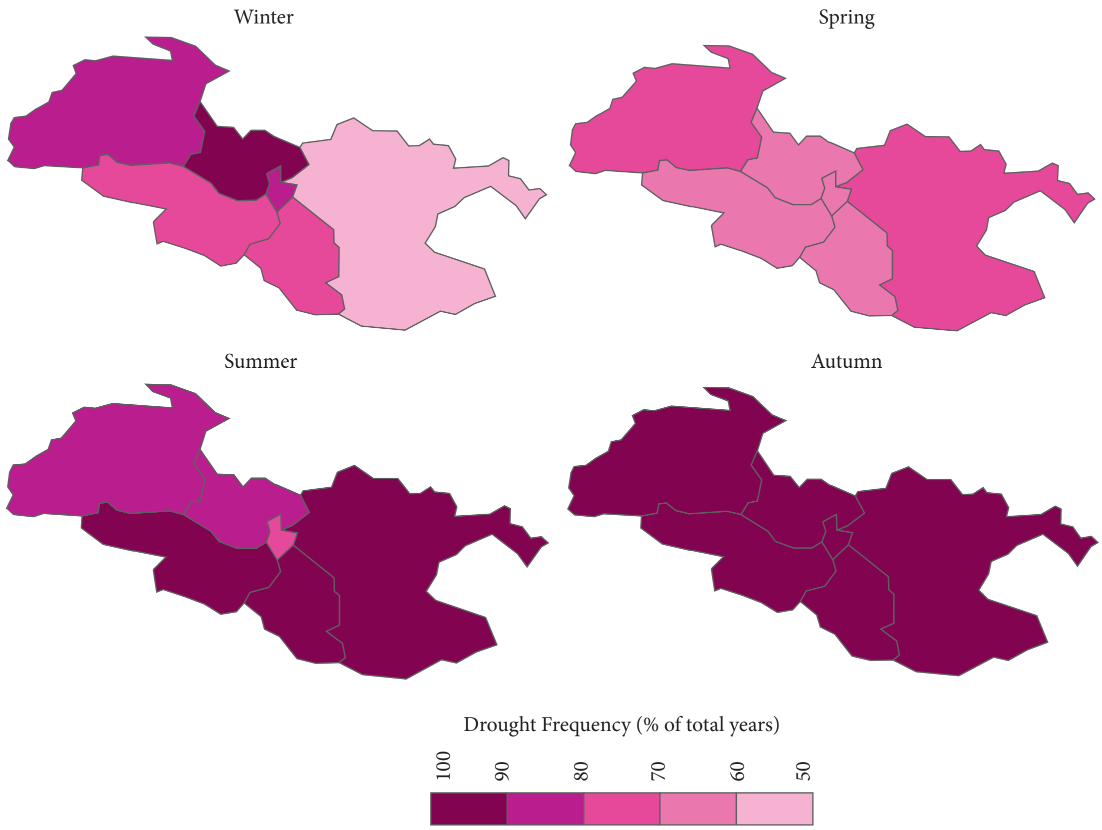


Figure 7: Seasonal drought frequency for each selected station as a percent of total (1971–2017) years.

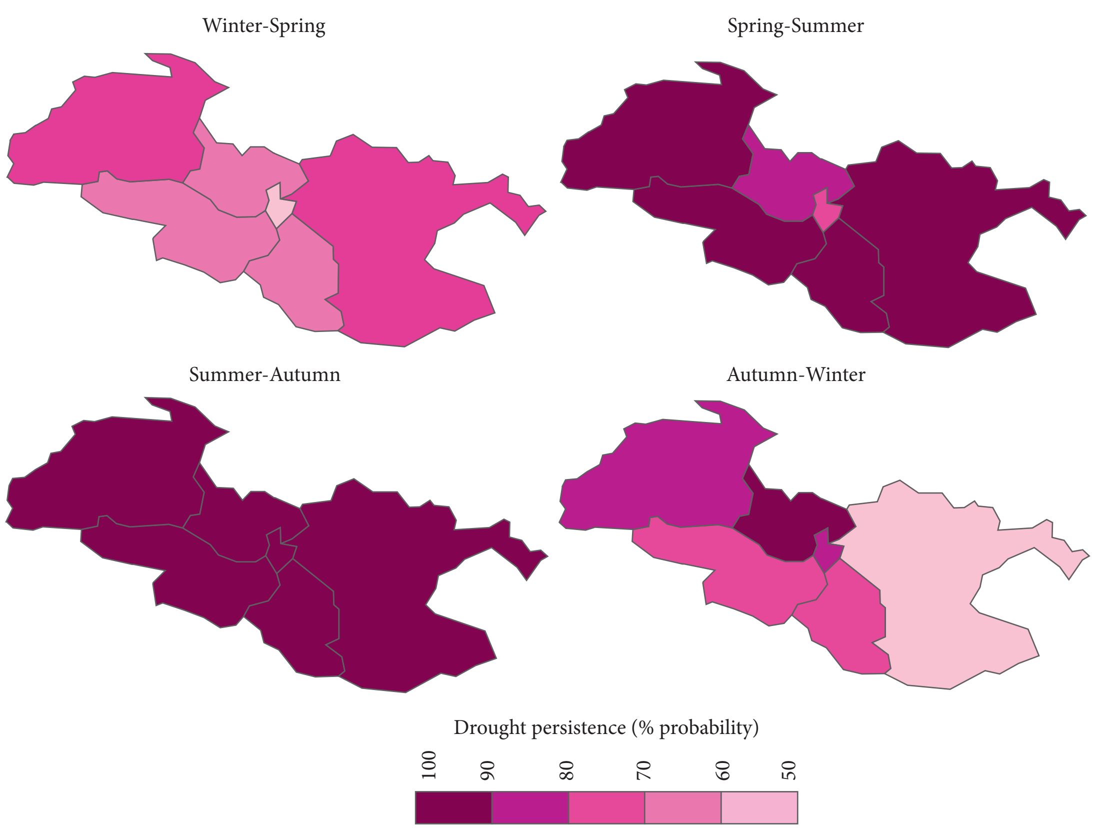


Figure 8: Drought persistence percent probability for all seasons in the selected stations.

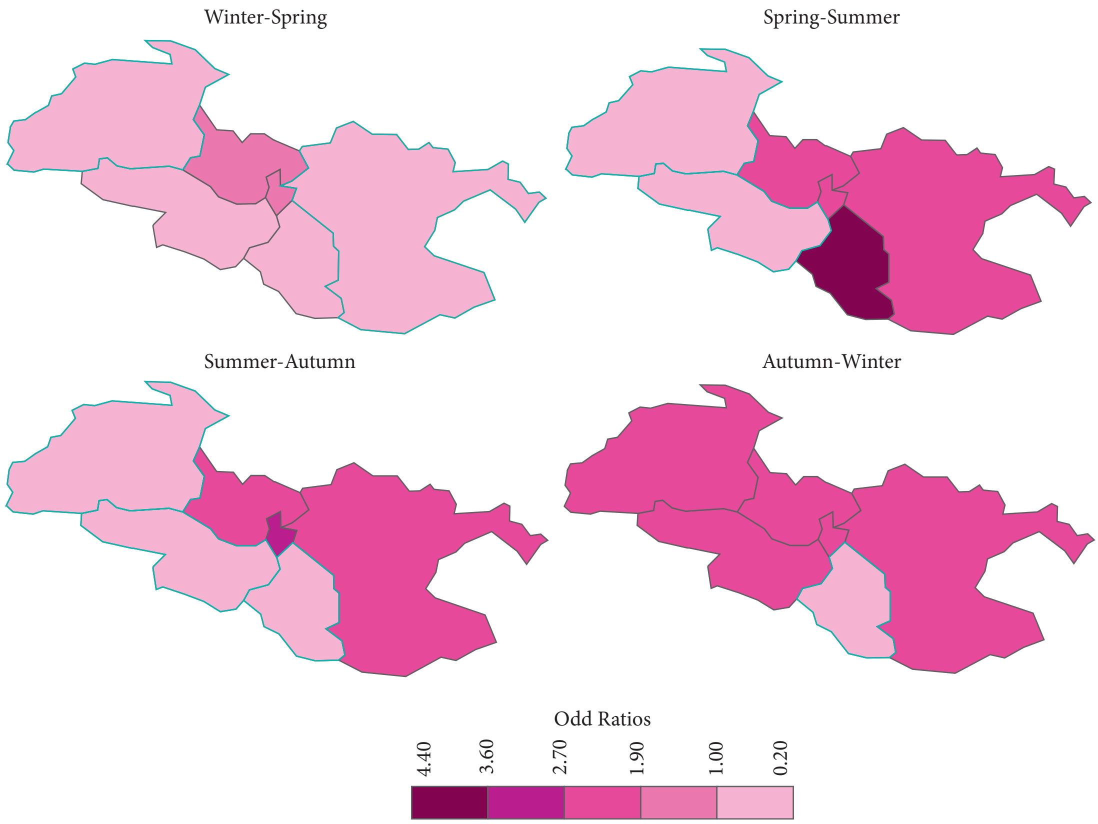


Figure 9: Odds ratio for various seasons in the selected stations. The relationship in the bolded stations is significant at $\alpha < 0.05$.

[59–61]. The present study is employed to identify the seasonal drought frequency and drought persistence in the region. However, none of the authors have worked to identify the seasonal drought frequency and drought persistence for Pakistan in previous research. Therefore, it is a new approach in terms of spatial patterns of drought persistence in northern areas, Pakistan. The outcome of the current analysis effectively works for early warning and mitigation policies and enhances the competencies in forecasting procedures, making better management, and planning to decrease the vulnerability of society to drought and its forgoing impacts.

4. Conclusion

Drought is a kind of complex recurrent climatic phenomenon, least understood to monitor and challenging for the past several decades. Because of the significant influences, various techniques have been developed to overcome its impacts. Due to the development of new methodologies, the drought characteristics have been monitored more accurately. Information about the drought characteristics and its persistence provides the direction to the water resource management and agriculture sectors to minimize the negative impacts of drought. Therefore, the current study used logistic regression to analyze the odds and probability of drought persistence from one season to the next in the selected stations by identifying the spatial pattern of seasonal drought frequency and persistence. From the results, we observed that the ND category is prevailing among other categories in various seasons of the selected stations. Furthermore, the seasonal drought frequency is found significantly high in summer and autumn. The spring meteorological drought in selected stations of the northern areas of Pakistan can persist to the summer with the probability of 90 to 100%. The drought persistence percent probability for summer to autumn is highest among other climate divisions. Furthermore, it was observed that there are large spatial variations in seasonal drought persistence in various seasons and stations. However, the probabilities of drought persistence in spring-summer and summer-autumn show a more similar pattern among the selected stations. However, the Skardu station has a lower probability of autumn-winter drought persistence. Moreover, the odds ratios from the logistic regression model for the seasonal drought persistence showed that the significant relationship between various seasons and SPI is negative in selected stations. This means increases 1 unit in the SPI the odds of drought occurrence are negatively influenced. The Astore, Gupis, Chilas, and Gilgit found significant at $\alpha < 0.05$ in various seasons. Moreover, the current analysis is applied for the present situation and application site; however, for other climatic conditions results cannot be generalized. As climatology conditions of the regions vary, the results can be affected for the assessment of drought occurrence. Furthermore, the use of other indices can significantly increase the capabilities for drought monitoring and its persistence in the selected region. The outcomes associated with the present assessment provide the basis for getting more considerations on early warning systems and can be used for decision-making.

Data Availability

The data used for the preparation of the manuscript are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work under grant number (RGP.1/26/42), received by Mohammed M. Almazah (https://www.kku.edu.sa). This work was supported by grants from the National Natural Science Foundation of China program (41801339) and Hubei Provincial Natural Science Foundation of China (2020CFB615).

References

[1] J. Lorenzo-Lacruz, S. M. Vicente-Serrano, J. I. López-Moreno, S. Beguería, J. M. García-Ruiz, and J. M. Cuadrat, "The impact of droughts and water management on various hydrological systems in the headwaters of the Tagus River (central Spain)," *Journal of Hydrology*, vol. 386, no. 1-4, pp. 13–26, 2010.

[2] S. Beguería, S. M. Vicente-Serrano, F. Reig, and B. Latorre, "Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring," *International Journal of Climatology*, vol. 34, no. 10, pp. 3001–3023, 2014.

[3] B. Dobson, G. Coxon, J. Freer, H. Gavin, M. Mortazavi-Naeini, and J. W. Hall, "The spatial dynamics of droughts and water scarcity in England and Wales," *Water Resources Research*, vol. 56, no. 9, Article ID e2020WR027187, 2020.

[4] M. Alamgir, N. Khan, S. Shahid et al., "Evaluating severity-area-frequency (SAF) of seasonal droughts in Bangladesh under climate change scenarios," *Stochastic Environmental Research and Risk Assessment*, vol. 34, no. 2, pp. 447–464, 2020.

[5] D. Rodziewicz and J. Dice, "Drought risk to the agriculture sector," *Economic Review*, vol. 105, no. 2, 2020.

[6] P. Kalura, A. Pandey, V. M. Chowdary, and P. V. Raju, "Assessment of hydrological drought vulnerability using geospatial techniques in the tons river basin, India," *Journal of the Indian Society of Remote Sensing*, pp. 1–15, 2021.

[7] I. Raev and B. Rosnev, "The impact of drought on natural forest ecosystems," in *Drought in Bulgaria*, pp. 117–136, Routledge, England, UK, 2018.

[8] L. Deng, C. Peng, D. G. Kim et al., "Drought effects on soil carbon and nitrogen dynamics in global natural ecosystems: a meta-analysis," *Earth-Science Reviews*, vol. 214, Article ID 103501, 2020.

[9] I. R. Orimoloye, J. A. Belle, and O. O. Ololade, "Drought disaster monitoring using MODIS derived index for drought years: a space-based information for ecosystems and environmental conservation," *Journal of Environmental Management*, vol. 284, Article ID 112028, 2021.

[10] C. Grossiord, "Having the right neighbors: how tree species diversity modulates drought impacts on forests," *New Phytologist*, vol. 228, no. 1, pp. 42–49, 2020.

[11] B. Schuldt, A. Buras, M. Arend et al., "A first assessment of the impact of the extreme 2018 summer drought on Central European forests," *Basic and Applied Ecology*, vol. 45, pp. 86–103, 2020.

[12] A. L. Long, "Disturbance and sustainability in forests of the western United States," *Drought*, vol. 992, 2021.

[13] S. Huang, P. Li, Q. Huang, G. Leng, B. Hou, and L. Ma, "The propagation from meteorological to hydrological drought and its potential influence factors," *Journal of Hydrology*, vol. 547, pp. 184–195, 2017.

[14] M. Masoudi and M. Elhaeesahar, "GIS analysis for vulnerability assessment of drought in Khuzestan province in Iran using standardized precipitation index (SPI)," *Iran Agricultural Research*, vol. 38, no. 2, pp. 9–16, 2019.

[15] V. Marchionni, E. Daly, G. Manoli, N. J. Tapper, J. P. Walker, and S. Fatichi, "Groundwater buffers drought effects and climate variability in urban reserves," *Water Resources Research*, vol. 56, no. 5, 2020.

[16] S. E. Abdel-Ghany, F. Ullah, A. Ben-Hur, and A. S. N. Reddy, "Transcriptome analysis of drought-resistant and drought-sensitive sorghum (Sorghum bicolor) genotypes in response to peg-induced drought stress," *International Journal of Molecular Sciences*, vol. 21, no. 3, p. 772, 2020.

[17] A. D. Cameron, "Mitigating the risk of drought-induced stem cracks in conifers in a changing climate," *Scandinavian Journal of Forest Research*, vol. 34, no. 8, pp. 667–672, 2019.

[18] A. Gazol, J. J. Camarero, R. Sánchez-Salguero et al., "Drought legacies are short, prevail in dry conifer forests and depend on growth variability," *Journal of Ecology*, vol. 108, no. 6, pp. 2473–2484, 2020.

[19] A. AghaKouchak, A. Mirchi, K. Madani et al., *Anthropogenic Drought: Definition, Challenges, and Opportunities*, AGU, Washington, DC, USA, 2021.

[20] A. Hamdy and G. Trisorio-Liuzzi, *Drought Planning and Drought Mitigation Measures in the Mediterranean Region*, AGU, Washington, DC, USA, 2001.

[21] R. Ragab and A. Hamdy, "Water management strategies to combat drought in the semiarid regions," *Water Management for Drought Mitigation in the Mediterranean*, pp. 47–112, Centre International de Hautes Etudes Agronomiques Méditerranéennes, Tecnomack–Bari, Italy, 2004.

[22] E. Gidey, O. Dikinya, R. Sebego, E. Segosebe, and A. Zenebe, "Predictions of future meteorological drought hazard (~2070) under the representative concentration path (RCP) 4.5 climate change scenarios in Raya, Northern Ethiopia," *Modeling Earth Systems and Environment*, vol. 4, no. 2, pp. 475–488, 2018.

[23] M. A. Nayak and W. U. Hassan, "A synthesis of drought prediction research over India," *Water Security*, vol. 13, Article ID 100092, 2021.

[24] D. A. Wilhite, *Drought as a Natural Hazard: Concepts and Definitions*, AGU, Washington, DC, USA, 2000.

[25] O. V. Wilhelmi, K. G. Hubbard, and D. A. Wilhite, "Spatial representation of agroclimatology in a study of agricultural drought," *International Journal of Climatology*, vol. 22, no. 11, pp. 1399–1414, 2002.

[26] J. Rhee and J. Im, "Meteorological drought forecasting for ungauged areas based on machine learning: using long-range climate forecast and remote sensing data," *Agricultural and Forest Meteorology*, vol. 237-238, pp. 105–122, 2017.

[27] S. Jiang, L. Wei, L. Ren et al., "Utility of integrated IMERG precipitation and GLEAM potential evapotranspiration

products for drought monitoring over mainland China," *Atmospheric Research*, vol. 247, Article ID 105141, 2021.

[28] H. Salimi, E. Asadi, and S. Darbandi, "Meteorological and hydrological drought monitoring using several drought indices," *Applied Water Science*, vol. 11, no. 2, pp. 1–10, 2021.

[29] B. Clausen and C. P. Pearson, "Regional frequency analysis of annual maximum streamflow drought," *Journal of Hydrology*, vol. 173, no. 1-4, pp. 111–130, 1995.

[30] A. Dai, "Drought under global warming: a review," *WIREs Climate Change*, vol. 2, no. 1, pp. 45–65, 2011.

[31] M. P. Van Rooy, "A rainfall anomally index independent of time and space," *NOTOS*, vol. 14, 1965.

[32] W. C. Palmer, "Meteorological drought," *US Department of Commerc*, vol. 30, 1965.

[33] W. C. Palmer, "Keeping track of crop moisture conditions, nationwide: the new Crop Moisture Index," *Weatherwise*, vol. 21, no. 4, pp. 156–161, 1968.

[34] H. N. Bhalme and D. A. Mooley, "Large-scale droughts/floods and monsoon circulation," *Monthly Weather Review*, vol. 108, no. 8, pp. 1197–1211, 1980.

[35] K. M. Weghorst, *The Reclamation Drought Index: Guidelines and Practical Applications*, vol. 6, Bureau of Reclamation, Denver, CO, USA, 1996.

[36] R. Gommes and F. Petrassi, "Rainfall variability and drought in sub-Saharan Africa," vol. 630, 1996.

[37] B. A. Shafer and L. E. Dezman, "Development OF surface water supply index (SWSI) to assess the severity OF drought condition IN snowpack runoff areas," in *Proceedings of the Western Snow Conference*, Fort Collins, CO, USA, 1982, January.

[38] T. B. McKee, N. J. Doesken, and J. Kleist, "The relationship of drought frequency and duration to time scales," in *Proceedings of the 8th Conference on Applied Climatology*, Boston, MA, USA, 1993, January.

[39] I. Nalbantis and G. Tsakiris, "Assessment of hydrological drought revisited," *Water Resources Management*, vol. 23, no. 5, pp. 881–897, 2009.

[40] D. H. Mlenga, A. J. Jordaan, and B. Mandebvu, "Integrating standard precipitation index and normalised difference vegetation index for near-real-time drought monitoring in Eswatini," *Jamba (Potchefstroom, South Africa)*, vol. 11, no. 1, pp. 917–919, 2019.

[41] T. Caloiero and S. Veltri, "Drought assessment in the sardinia region (Italy) during 1922-2011 using the standardized precipitation index," *Pure and Applied Geophysics*, vol. 176, no. 2, pp. 925–935, 2019.

[42] T. Javed, Y. Li, S. Rashid et al., "Performance and relationship of four different agricultural drought indices for drought monitoring in China's mainland using remote sensing data," *The Science of the Total Environment*, vol. 759, Article ID 143530, 2021.

[43] O. Singh, D. Saini, and P. Bhardwaj, "Characterization of meteorological drought over a dryland ecosystem in northwestern India," *Natural Hazards*, vol. 42, 2021.

[44] J. P. Bloomfield and B. P. Marchant, "Analysis of groundwater drought building on the standardised precipitation index approach," *Hydrology and Earth System Sciences*, vol. 17, no. 12, pp. 4769–4787, 2013.

[45] C.-Y. J. Peng, K. L. Lee, and G. M. Ingersoll, "An introduction to logistic regression analysis and reporting," *The Journal of Educational Research*, vol. 96, no. 1, pp. 3–14, 2002.

[46] M. H. Kutner, C. J. Nachtsheim, J. Neter, and W. Li, *Applied Linear Statistical Models*, University of Minnesota, Minneapolis, Minnesota, 2005.

[47] T. Ford and C. F. Labosier, "Spatial patterns of drought persistence in the Southeastern United States," *International Journal of Climatology*, vol. 34, no. 7, pp. 2229–2240, 2014.

[48] S. Bachmair, C. Svensson, I. Prosdocimi, J. Hannaford, and K. Stahl, "Developing drought impact functions for drought risk management," *Natural Hazards and Earth System Sciences*, vol. 17, no. 11, pp. 1947–1960, 2017.

[49] L. Meng, T. Ford, and Y. Guo, "Logistic regression analysis of drought persistence in East China," *International Journal of Climatology*, vol. 37, no. 3, pp. 1444–1455, 2017.

[50] G. Rasul, Q. Z. Chaudhry, A. Mahmood, K. W. Hyder, and Q. Dahe, "Glaciers and glacial lakes under changing climate in Pakistan," *Pakistan Journal of Meteorology*, vol. 8, no. 15, 2011.

[51] Y. Latif, M. Yaoming, M. Yaseen, S. Muhammad, and M. A. Wazir, "Spatial analysis of temperature time series over the Upper Indus Basin (UIB) Pakistan," *Theoretical and Applied Climatology*, vol. 139, no. 1, pp. 741–758, 2020.

[52] M. Adnan, G. Nabi, M. Saleem Poomee, and A. Ashraf, "Snowmelt runoff prediction under changing climate in the Himalayan cryosphere: a case of Gilgit River Basin," *Geoscience Frontiers*, vol. 8, no. 5, pp. 941–949, 2017.

[53] R. Niaz, I. Hussain, Z. Ali et al., "A novel spatially weighted accumulative procedure for regional drought monitoring," *Tellus A: Dynamic Meteorology and Oceanography*, vol. 72, no. 1, pp. 1–13, 2020.

[54] S. A. Anjum, L. Wang, J. Salhab, I. Khan, and M. F. Saleem, "An assessment of drought extent and impacts in agriculture sector in Pakistan," *Journal of Food Agriculture and Environment*, vol. 8, no. 2, pp. 1359–1363, 2010.

[55] S. Hina and F. Saleem, "Historical analysis (1981-2017) of drought severity and magnitude over a predominantly arid region of Pakistan," *Climate Research*, vol. 78, no. 3, pp. 189–204, 2019.

[56] K. Ahmed, S. Shahid, and N. Nawaz, "Impacts of climate variability and change on seasonal drought characteristics of Pakistan," *Atmospheric Research*, vol. 214, pp. 364–374, 2018.

[57] N. Khan, D. A. Sachindra, S. Shahid, K. Ahmed, M. S. Shiru, and N. Nawaz, "Prediction of droughts over Pakistan using machine learning algorithms," *Advances in Water Resources*, vol. 139, Article ID 103562, 2020.

[58] N. Mazhar, M. Nawaz, A. I. Mirza, and K. Khan, "Socio-political impacts of meteorological droughts and their spatial patterns in Pakistan," *South Asian Studies*, vol. 30, no. 1, 2020.

[59] A. H. Payab and U. Türker, "Comparison of standardized meteorological indices for drought monitoring at northern part of Cyprus," *Environmental Earth Sciences*, vol. 78, no. 10, pp. 1–19, 2019.

[60] O. O. Ayantobo and J. Wei, "Appraising regional multi-category and multi-scalar drought monitoring using standardized moisture anomaly index (SZI): a water-energy balance approach," *Journal of Hydrology*, vol. 579, p. 124139, 2019.

[61] N. Puletti, W. Mattioli, F. Bussotti, and M. Pollastrini, "Monitoring the effects of extreme drought events on forest health by Sentinel-2 imagery," *Journal of Applied Remote Sensing*, vol. 13, no. 2, Article ID 020501, 2019.

Hindawi
Complexity
Volume 2021, Article ID 6735210, 11 pages
https://doi.org/10.1155/2021/6735210

Research Article

A Two-Sided Stable Matching Model of Cloud Manufacturing Tasks and Services considering the Nonlinear Relationship between Satisfaction and Expectations

Yujie Zheng, Meiyan Li, and Jiakun Liu

Shandong University of Science and Technology, Qingdao, China

Correspondence should be addressed to Jiakun Liu; lu0shi@163.com

Received 6 July 2021; Revised 17 August 2021; Accepted 6 September 2021; Published 7 October 2021

Academic Editor: Feng Li

To improve the satisfaction of both service demanders (SDs) and service providers (SPs) in the matching of cloud manufacturing (CMfg) tasks and services, a two-sided stable matching model of CMfg tasks and service considering the nonlinear relationship between satisfaction and expectations is proposed. As the expectations of SDs and SPs are difficult to be quantified directly, an evaluation method based on interval-valued hesitant fuzzy linguistic sets (IVHFLSs) is first presented. Next, a nonlinear model of satisfaction and expectations is built to quantify the satisfaction, which achieves accurate quantification of satisfaction. Then, a two-sided stable matching model of CMfg tasks and service is built, which takes the satisfaction of SDs and SPs as the optimization goals and considers the individual rationalities and blocking pairs. Finally, an adaptive genetic algorithm (AGA) is designed to solve the proposed two-sided matching model. A practical application and comparison analysis is used to verify the effectiveness and superiority of the research.

1. Introduction

Cloud manufacturing (CMfg) is a new manufacturing mode that organizes the online service released by service providers (SPs) and allocates them to service demanders (SDs) on demand [1, 2]. SDs expect to cooperate with SPs of low service cost and high service quality. SPs also expect to serve SDs with fast payment speed and high credibility [3, 4]. Hence, both SDs and SPs have expectations on the matching scheme of CMfg tasks and service. The comparison between the expectations of SDs for service and the actual situations of the service forms the satisfaction of SDs [5]. Similarly, the comparison between the expectations of SPs for tasks and the actual situations of the tasks forms the satisfaction of SPs. The higher the satisfaction of SDs and SPs, the stronger the competitiveness of CMfg platforms [6, 7].

Many studies concerning the matching of CMfg tasks and service have been performed. The research can be divided into two categories: single-sided matching and two-sided matching. Single-sided matching research has provided an important reference for two-sided matching research. For example, Lartigau et al. optimized the manufacturing cost, time, reliability, maintainability, and availability in research on CMfg service composition [8]. Zhang et al. developed an optimization model for service configuration in CMfg that took manufacturing cost, time, and quality as the optimization objectives [9]. Hsieh and Lin proposed a dynamic scheme of scheduling complex collaborative tasks for minimizing time and cost [10]. Xiong et al. maximized manufacturing efficiency and balanced service load during the service scheduling process [11]. Joglekar and Ford proposed a service allocation matrix to shorten the project duration [12]. Wang et al. optimized the cost, time, quality, and risk with a mathematical model [13]. Because single-sided matching research has mainly considered the expectations of SDs but ignored the expectations of SPs, and its effectiveness and feasibility need to be improved. To compensate, Zhao and Wang proposed a two-sided matching model of CMfg tasks and service [14]. First, SPs and SDs assessed each other based on linguistic information, and the assessment results were transformed into numerical values. Then, the satisfaction was calculated by using the variable fuzzy recognition method. Finally, a multiobjective optimization model was established, and the multiobjective functions were converted into a single objective function. Li et al. built a two-sided matching model with hesitant fuzzy preference information for configuring CMfg tasks and service [15]. First, SPs and SDs gave their expectations using the hesitant fuzzy element. Then, a two-sided matching model was constructed, and the optimal configuration results were obtained by solving the model. Ren and Ren put forward a one-to-many two-sided matching model, and the expectation utility theory was applied to calculate the satisfaction of SDs and SPs [16]. Zhao and Ding researched the two-sided resource matching mechanism and stability of a CMfg platform [17].

In summary, the two-sided matching research of CMfg service and tasks is in the exploratory stage at present. Relevant research has supposed that the relationship between satisfaction and expectations is linear. The more the matching scheme exceeds the expectations of SDs and SPs, the higher the satisfaction. They took the values of satisfaction indexes or the difference between the values and the expectations as the optimization objectives. However, satisfaction is a psychological feeling that forms after expectation is compared with the actual situation. The functions of expectations and satisfaction are nonlinear and segmented [5]. In some segments, the satisfaction is 0 or a fixed number, whereas in other segments, satisfaction increases or decreases between 0 and the fixed number.

When assessing satisfaction, some information such as expectations, the lower threshold, and the upper threshold needs to be determined in advance. The complexity of satisfaction indexes and the limitation of human cognition, hesitation, and fuzziness always exist in the expression process of information. Under such a situation, qualitative linguistic terms instead of precise quantitative numbers are more suitable for SDs and SPs to express information of satisfaction indexes [18]. Consequently, a number of linguistic decision-making methods emerged, such as the probabilistic linguistic term set (PLT) [18], hesitant fuzzy linguistic sets (HFLSs) [19, 20], interval-valued hesitant fuzzy linguistic sets (IVHFLSs) [21, 22], fuzzy preference relation with self-confidence (FPR-SC) [23], etc. These methods have been applied widely to various fields. In this article, the interval-valued hesitant fuzzy linguistic sets (IVHFLSs) are introduced, which are the extension of the linguistic term sets and interval-valued hesitant fuzzy sets (IVHFSs) [22]. Compared to PLT and HFLSs, IVHFLSs make the membership degrees no longer just relative to some crisp number, thus depicting the hesitancy of SDs and SPs, comprehensively. Compared to IVHFSs, IVHFLSs express the preference of SDs and SPs with linguistic terms, thus qualitatively representing the uncertain information. In a word, IVHFLSs are more convenient for SDs and SPs to express the information related to satisfaction and can effectively describe uncertainty, hesitancy, and inconsistency inherent in the decision-making process.

To bridge the gap, a two-sided matching model of CMfg tasks and service based on the quantification of satisfaction is proposed. Compared with the traditional matching model (as shown in Figure 1(a)), the main contributions of the article are as follows:

(1) A nonlinear model of satisfaction and expectations is built (as shown in Figure 1(b)), which achieves accurate quantification of satisfaction and lays the foundation for the improvement of satisfaction in the matching of CMfg tasks and service.

(2) A two-sided stable matching model of CMfg tasks and service is put forward (as shown in Figure 1(b)), which directly takes the satisfaction of SDs and SPs as the optimization goals and considers the individual rationalities and blocking pairs, thus making the optimization of satisfaction more effective.

The remaining sections of this article are organized as follows: Section 2 establishes a nonlinear model of satisfaction and expectations. Section 3 details a two-sided stable matching model of CMfg tasks and service on the basis of Section 2. Section 4 presents an example application and a comparison analysis of this article. Section 5 provides conclusions.

2. Quantification of Satisfaction Based on the Nonlinear Relationship

2.1. Expression of Expectations Based on IVHFLSs. To make the problem specific, suppose that SDs published I manufacturing tasks on a CMfg platform, and SPs issued J matching service. D_i and P_j indicate the $i-$th CMfg task and the $j-$th CMfg service, respectively.

Let z_m be the m-th satisfaction index of SDs that is difficult to quantify directly, $L = \{l_t | t = 0, 1, \ldots, 2T, T \in N+\}$ be a linguistic set, $l_t(D_{i,m}) \in L$ be the linguistic evaluation result of CMfg task D_i for index z_m, and $\Gamma(D_{i,m}) \in [0,1]$ be the membership interval of $l_t(D_{i,m})$, then the expectation of the service demander published task D_i on index z_m based on IVHFLSs can be expressed as $h_{i,m} = \langle l_t(D_{i,m}), \Gamma(D_{i,m})\rangle$. Moreover, $\Gamma(D_{i,m}) = \{[a_k^-, a_k^+], k = 1, 2, \ldots, \#\Gamma(D_{i,m})\}$, where $\#\Gamma(D_{i,m})$ is the number of intervals. Similarly, if z_n' is a satisfaction index of SPs that is difficult to be quantified directly, the expectation of the service provider published P_j on index z_n' based on IVHFLSs can be expressed as $h_{j,n} = \langle l_t(P_{j,n}), \Gamma(P_{j,n})\rangle$.

2.2. Quantification of Expectations Based on Score Function. According to literature [21], the expectation function $E(\Gamma(D_{i,m}))$ of $\Gamma(D_{i,m})$ can be denoted as follows:

$$E\left(\Gamma\left(D_{i,m}\right)\right) = \frac{\sum_{k=1}^{\#\Gamma(D_{i,m})} \left(a_k^- + a_k^+\right)}{2\#\Gamma\left(D_{i,m}\right)}. \quad (1)$$

Thus, the score function $s(h_{i,m})$ of $h_{i,m}$ can be represented as follows:

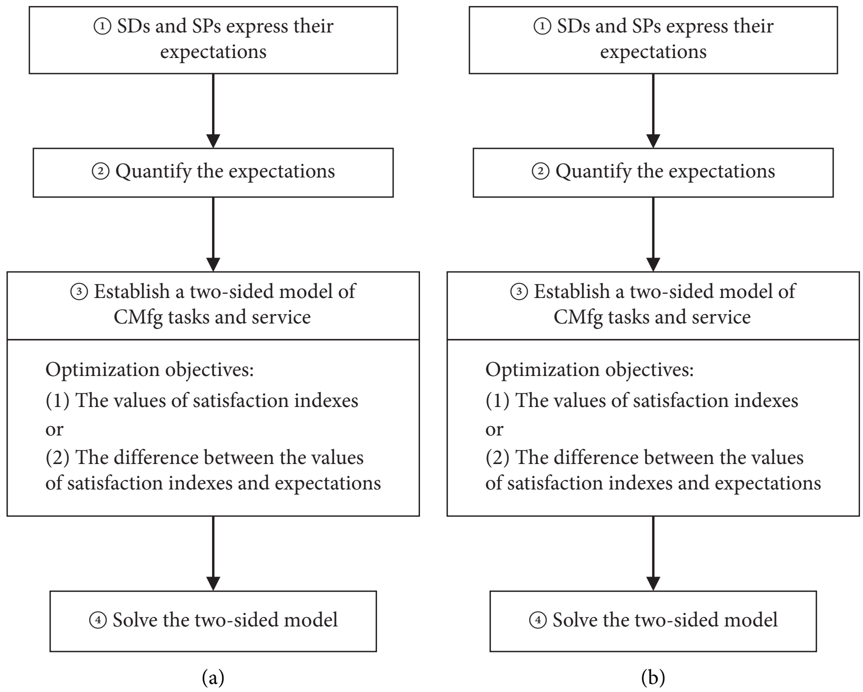


FIGURE 1: Framework of two-sided matching methods of CMfg tasks and service. (a) Traditional matching methods. (b) The proposed two-sided matching methods.

$$s\left(h_{i,m}\right) = f\left(l_t\left(D_{i,m}\right)\right) \times E\left(\Gamma\left(D_{i,m}\right)\right). \quad (2)$$

Similarly, the score function $s(h_{j,n})$ of $h_{j,n}$ can be represented as follows:

$$s\left(h_{j,n}\right) = f\left(l_t\left(D_{j,n}\right)\right) \frac{\sum_{k=1}^{\#\Gamma\left(P_{j,n}\right)} \left(a_k^- + a_k^+\right)}{2\#\Gamma\left(P_{j,n}\right)}. \quad (3)$$

2.3. A Nonlinear Model of Satisfaction and Expectations. By fitting data, Liu et al. proposed that the piecewise exponential function is the subject function to the expression of customer satisfaction [5]. Especially, as the values of satisfaction indexes increase or decrease, the change in customer satisfaction can be divided into four areas: no-response area, defect area, overflow area, and saturation area. In the no-response area, the customer satisfaction is 0; in the defect area, the customer satisfaction increases or decreases rapidly and exponentially; in the overflow area, the customer satisfaction increases or decreases slowly and exponentially; and in the saturation area, satisfaction reaches its maximum and will not change. Again by fitting data, Zhou et al. found that the customer satisfaction has exponent relation to a satisfaction index [24]. Based on the above studies, a nonlinear model of satisfaction and expectations is built.

If z_m is a positive correlation index, the satisfaction degree $\phi_{ij}^{D\longrightarrow P}(m)$ of the service demander published task D_i with service P_j for index z_m can be expressed as in the following equation (the corresponding satisfaction curve is shown in Figure 2(a)):

$$\phi_{ij}^{D\longrightarrow P}(m) = \begin{cases} 0, & g_{j,m}^P \le h_{i,m}^T, \\ e^{\left(g_{j,m}^P - h_{i,m}^D / g_{j,m}^P - h_{i,m}^T\right)}, & h_{i,m}^T < g_{j,m}^P \le h_{i,m}^D, \\ 1 + (K-1)e^{\left(g_{j,m}^P - h_{i,m}^\Delta / g_{j,m}^P - h_{i,m}^D\right)}, & h_{i,m}^D < g_{j,m}^P \le h_{i,m}^\Delta, \\ K, & g_{j,m}^P > h_{i,m}^\Delta. \end{cases} \quad (4)$$

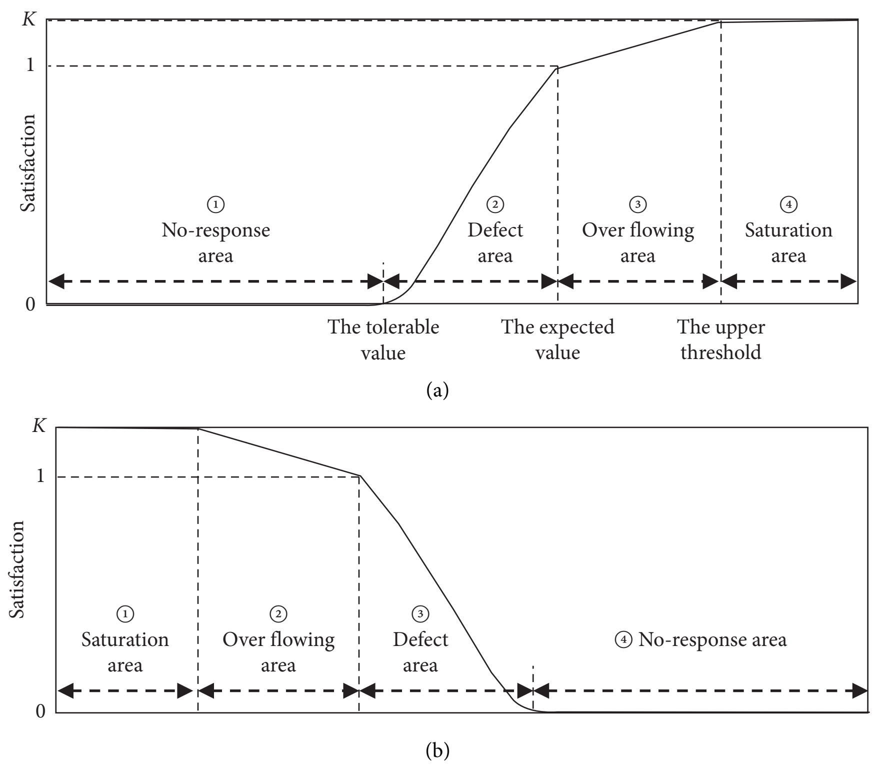


FIGURE 2: The satisfaction curve. (a) The satisfaction curve of a positive correlation index. (b) The satisfaction curve of a negative correlation index.

If z_m is a negative correlation index, the satisfaction degree $\phi_{ij}^{D\longrightarrow P}(m)$ of the service demander published task D_i with service P_j for index z_m can be expressed as in the following equation (the corresponding satisfaction curve is shown in Figure 2(b)):

$$\phi_{ij}^{D\longrightarrow P}(m)=\begin{cases} K, & g_{j,m}^{P}\le h_{i,m}^{\nabla}, \\ K-(K-1)e^{\left(g_{j,m}^{P}-h_{i,m}^{D}/g_{j,m}^{P}-h_{i,m}^{\nabla}\right)}, & h_{i,m}^{\nabla}<g_{j,m}^{P}\le h_{i,m}^{D}, \\ e^{\left(g_{j,m}^{P}-h_{i,m}^{D}/g_{j,m}^{P}-h_{i,m}^{T}\right)}, & h_{i,m}^{D}<g_{j,m}^{P}\le h_{i,m}^{T}, \\ 0, & g_{j,m}^{P}>h_{i,m}^{T}. \end{cases} \quad (5)$$

In equations (4) and (5), $h_{i,m}^{D}$, $h_{i,m}^{T}$, $h_{i,m}^{\Delta}$, and $h_{i,m}^{\nabla}$ represent the expected values, the tolerable values, the upper threshold, and the lower threshold of task D_i for index z_m, respectively. $g_{j,m}^{P}$ means the actual values of service P_j for index z_m.

Similarly, if the satisfaction index of SPs z_n' is a positive correlation index, the satisfaction degree $\phi_{ji}^{P\longrightarrow D}(n)$ of the service provider published service P_j with task D_i for index z_n' can be expressed as follows:

$$\phi_{ji}^{P\longrightarrow D}(n)=\begin{cases} 0, & g_{i,n}^{D}\le h_{j,n}^{T}, \\ e^{\left(g_{i,n}^{D}-h_{j,n}^{P}/g_{i,n}^{D}-h_{j,n}^{\nabla}\right)}, & h_{j,n}^{T}<g_{i,n}^{D}\le h_{j,n}^{P}, \\ 1+(K-1)e^{\left(g_{i,n}^{D}-h_{j,n}^{\Delta}/g_{i,n}^{D}-h_{j,n}^{P}\right)}, & h_{j,n}^{P}<g_{i,n}^{D}\le h_{j,n}^{\Delta}, \\ K, & g_{i,n}^{D}>h_{j,n}^{\Delta}. \end{cases} \quad (6)$$

If z_n' is a negative correlation index, the satisfaction degree $\phi_{ji}^{P\longrightarrow D}(n)$ of the service provider published service P_j with task D_i can be expressed as follows:

$$\phi_{ji}^{P\longrightarrow D}(n)=\begin{cases} K, & g_{i,n}^{D}\le h_{j,n}^{\nabla}, \\ K-(K-1)e^{\left(g_{i,n}^{D}-h_{j,n}^{P}/g_{i,n}^{D}-h_{j,n}^{\nabla}\right)}, & h_{j,n}^{\nabla}<g_{i,n}^{D}\le h_{j,n}^{P}, \\ e^{\left(g_{i,n}^{D}-h_{j,n}^{P}/g_{i,n}^{D}-h_{j,n}^{\Delta}\right)}, & h_{j,n}^{P}<g_{i,n}^{D}\le h_{j,n}^{T}, \\ 0, & g_{i,n}^{D}>h_{j,n}^{T}. \end{cases} \tag{7}$$

In equations (6) and (7), $h_{j,n}^{P}$, $h_{j,n}^{T}$, $h_{j,n}^{\Delta}$, and $h_{j,n}^{\nabla}$ denote the expected values, the tolerant values, the upper threshold, and the lower threshold of service P_j for index z_n', respectively. $g_{i,n}^{D}$ means the actual values of task D_i for index z_n'.

In equations (4)–(7), K is the overflowing coefficient of satisfaction. According to literature [5], the value of K can be set by experts, and its range is usually between 1 to 1.2. Based on equations (4)–(7), the satisfaction degree $\phi_{ij}^{D\longrightarrow P}$ of D_i with P_j and the satisfaction degree $\phi_{ji}^{P\longrightarrow D}$ of P_j with D_i can be expressed as in the following first and second equations, respectively:

$$\phi_{ij}^{D\longrightarrow P}=w(m)\left[\phi_{ij}^{D\longrightarrow P}(m)\right]^{T}, \tag{8}$$

$$\phi_{ji}^{P\longrightarrow D}=w'(n)\left[\phi_{ji}^{P\longrightarrow D}(n)\right]^{T}. \tag{9}$$

In equations (8) and (9), $w(m)$ and $w'(n)$ are the weights of satisfaction indexes z_m and z_n', respectively.

3. A Two-Sided Stable Matching Model of CMfg Tasks and Service

3.1. Model Construction. CMfg tasks can be divided into different types, such as machining, heat treatment, welding, and metal forming. Let q denote the q-th type of task ($q = 1, 2, \ldots, Q$) and the 0-1 variables c_i^q, $c_j'^q$, and y_{ij} denote the type of CMfg tasks, the type of cloud service, and the matching variable of tasks and service, respectively. If task D_i belongs to the q-th task type, then $c_i^q = 1$; otherwise, $c_i^q = 0$. If service P_j can operate the q-th task type, $c_j'^q = 1$; otherwise, $c_j'^q = 0$. If task D_i is matched with service P_j, then $y_{ij} = 1$; otherwise, $y_{ij} = 0$.

Definition 1. The two-sided matching scheme between CMfg tasks and service is defined as mapping λ: $D\cup P \longrightarrow 2^{D\cup P}$. If $\forall D_i \in D$, $\forall P_j \in P$ satisfies the following conditions: (1) $\lambda(D_i) \in \left\{P_j | c_i^q c_j'^q = 1, P_j \in P\right\}$; (2) $\lambda(P_j) \subseteq \left\{D_j | c_i^q c_j'^q = 1, D_i \in D\right\}$; (3) $D_i \in \lambda(P_j)$, then (D_i, P_j) is a matching pair of CMfg tasks and service.

If a service provider is incapable of the q-th type of task, the service provider would rather not provide the service to the q-th type of task. This phenomenon is called the individual rationality of SPs. Its specific definition is given below.

Definition 2. If service $^{j}_{P}$ satisfies one of the following conditions: (1) $c_i^q c_j'^q y_{ij} = 1$; (2) $c_i^q c_j'^q = 0$, $y_{ij} = 0$, then it is called the individual rationality of SPs.

Definition 3. If all the SPs in the matching scheme λ are rational, then λ is an individual rational matching scheme of SPs.

If the SDs would rather not match service than match the unacceptable service, then this phenomenon is called the individual rationality of SDs. Its specific definition is given below.

Definition 4. If task D_i satisfies one of the following conditions: (1) $c_i^q c_j'^q y_{ij} = 1$; (2) $c_i^q = 1$, $c_j'^q = 0$, and $y_{ij} = 0$; (3) $\phi_{ij}^{D\longrightarrow P} = -M$ (M is a large number), $y_{ij} = 0$, then it is the individual rationality of SDs.

Definition 5. If all the SDs in matching scheme λ are rational, then λ is an individual rational matching scheme of SDs.

Definition 6. If λ is a matching scheme of both SPs' individual rationality and SDs' individual rationality, then λ is an individual rational matching scheme and satisfies the following equations:

$$y_{ij}-\sum_{q} c_i^q c_j'^q \le 0, \tag{10}$$

$$y_{ij}\phi_{ij}^{D\longrightarrow P} > -M. \tag{11}$$

If there are blocking pairs in a matching scheme, the matching two sides may match privately, thus affecting the stability and effectiveness of the matching scheme [25]. The blocking pairs in a matching scheme of CMfg tasks and service are defined as follows.

Definition 7. For the matching schemes of CMfg tasks and service λ: $D\cup P \longrightarrow 2^{D\cup P}$, $\exists D_i, D_b \in D$, $\exists P_j, P_f \in P$, $i \ne b$, $j \ne f$, if D_i, D_b, P_j and P_f satisfy one of the following conditions: (1) $|\lambda(P_j)| = 0$, $\lambda(D_i) = P_f$, and $c_i^q c_j'^q \phi_{ij}^{D\longrightarrow P} > \phi_{if}^{D\longrightarrow P}$; (2) $\lambda(D_i) = P_f$, $\lambda(D_b) = P_j$, $c_i^q c_j'^q \phi_{ij}^{D\longrightarrow P} > \phi_{if}^{D\longrightarrow P}$, and $c_i^q c_j'^q \phi_{ji}^{P\longrightarrow D} > \phi_{jb}^{P\longrightarrow D}$, then (D_i, P_j) is a blocking pair in the matching scheme λ.

Definition 8. If the matching scheme of CMfg tasks and service λ is individually rational and contains no blocking pairs, then λ is a stable matching scheme and satisfies the following equation:

$$\left(\sum_{i:\phi_{ij}^{D\longrightarrow P}>\phi_{if}^{D\longrightarrow P}} y_{ij}+y_{if}\right)\left(1+\left(\sum_{j:\phi_{jb}^{P\longrightarrow D}<\phi_{ji}^{P\longrightarrow D}} y_{ij}+y_{bj}\right)\right)>\sum_{q} c_i^q c_j^{\prime q}. \tag{12}$$

Based on the above analysis, the two-sided matching model of CMfg tasks and service can be established as in equation (13). The first and second optimization objectives are to maximize the satisfaction of SDs and satisfaction of SPs, respectively. The third optimization objective is to maximize the matching number of CMfg tasks and service. The meaning of constraints refers to Definitions 6 and 8,

$$\begin{aligned}
&\max Z_1=\sum_{i=1}^{I}\sum_{j=1}^{J}\phi_{ij}^{D\longrightarrow P} y_{ij} y_{ij}-\sum_q c_i^q c_j^{\prime q}\le 0,\ y_{ij}\phi_{ij}^{D\longrightarrow P}>-M\\
&\max Z_2=\sum_{i=1}^{I}\sum_{j=1}^{J}\phi_{ji}^{P\longrightarrow D} y_{ij}\quad \left(\sum_{i:\phi_{ij}^{D\longrightarrow P}>\phi_{if}^{D\longrightarrow P}} y_{ij}+y_{if}\right)\left(1+\left(\sum_{j:\phi_{jb}^{P\longrightarrow D}<\phi_{ji}^{P\longrightarrow D}} y_{ij}+y_{bj}\right)\right)>\sum_q c_i^q c_j^{\prime q}\\
&\max Z_3=\sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij}\quad \sum_{i=1}^{I} y_{ij}=1,\sum_{j=1}^{J} y_{ij}\le 1\\
&\qquad\text{S.T.}\quad y_{ij}\in\{0,1\},\quad i\in I,\ j\in J.
\end{aligned} \tag{13}$$

3.2. Model Solution. To solve the proposed two-sided model, an adaptive genetic algorithm (AGA) was designed.

3.2.1. Coding. Real coding is adopted in this algorithm. $D_i = j$ ($j = 0, 1, 2, \ldots, J$) denotes the task in which D_i and service P_j form a matching pair.

3.2.2. Fitness Function. The objectives Z_1 and Z_2 in equation (13) are mutual restrictions. The optimization solution satisfying the two objectives is difficult to obtain, but the ideal optimal solution and the ideal worst solution of each objective are easy to get. To this end, the technique for order preference by similarity to an ideal solution (TOPSIS) is adopted to convert the optimization of Z_1 and Z_2 into the following equations:

$$\max f_1=\frac{\sqrt{\sum_{t=1}^{2}\left(Z_t^{-}-Z_{tr}\right)^2}}{\left(\sqrt{\sum_{t=1}^{2}\left(Z_t^{+}-Z_{tr}\right)^2}+\sqrt{\sum_{t=1}^{2}\left(Z_t^{-}-Z_{tr}\right)^2}\right)}, \tag{14}$$

$$\max f_2=\sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij}. \tag{15}$$

In equation (14), Z_{tr} represents the t-th objective function value of the r-th individual in population. Z_t^+ and Z_t^- indicate the ideal optimal value and the ideal worst value of the t-th objective function, respectively.

3.2.3. Selection Operation. Given that tournament selection has better convergence than other selection operations [26, 27], the tournament selection strategy is adopted. First, select N_s individuals from population. Next, compare the fitness values between these selected individuals. Finally, put the individuals with the highest fitness values into a crossover pool. Cycle the above process until the crossover pool is full.

3.2.4. Adaptive Crossover. A two-point crossover is adopted here [28]. First, select individuals to be crossed as parents according to the crossover possibility pc. Second, randomly generate two integral points, rnd_1 and rnd_2, within the length of individuals as two crossover points. Finally, exchange the genes between the two integral points, rnd_1 and rnd_2, of two parents. Hereby, two children individuals are obtained. To avoid prematurity and local optimality, the adaptive crossover probability instead of a fixed one is adopted. The calculation method of the adaptive crossover probability is expressed as follows:

$$pc = \begin{cases} pc_{\min} + \dfrac{(pc_{\max} - pc_{\min})(f' - f_{\min})}{f_{\text{avg}} - f_{\min}}, & f' \le f_{\text{avg}}, \\ pc_{\max}, & f' > f_{\text{avg}}. \end{cases} \quad (16)$$

In equation (16), pc denotes an adaptive crossover probability, $pc_{\min}$ indicates the minimum crossover probability, $pc_{\max}$ denotes the maximum crossover probability, $f_{\min}$ indicates the minimum fitness values in contemporary populations, f_{avg} indicates the average fitness values in contemporary populations, and f' denotes the larger fitness values of the two individuals involved in the crossover.

3.2.5. Adaptive Mutation. Likewise, an adaptive mutation strategy is adopted here. The calculation method of the adaptive crossover probability is expressed as follows:

$$pm = \begin{cases} pm_{\min} + \dfrac{(pm_{\max} - pm_{\min})(f - f_{\min})}{f_{\text{avg}} - f_{\min}}, & f' \le f_{\text{avg}}, \\ pm_{\max}, & f' > f_{\text{avg}}. \end{cases} \quad (17)$$

In equation (17), pm denotes an adaptive mutation probability, $pm_{\min}$ indicates the minimum mutation probability, $pm_{\max}$ means the maximum mutation probability, $f_{\min}$ denotes the minimum fitness values in contemporary populations, f_{avg} indicates the average fitness values in contemporary populations, and f' denotes the larger fitness values of the two mutated individuals.

4. Example Application and Comparison Analysis

4.1. Example Application. There are 6 metal forming tasks on a CMfg platform. After preliminary screening, a total of 9 services can operate these welding tasks. The sets of welding tasks and service are expressed as $D = \{D_1, D_2, D_3, D_4, D_5, D_6\}$ and $P = \{P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9\}$, respectively. The satisfaction indexes of SDs include cost (z_1), delivery time (z_2), and credibility (z_3); the satisfaction indexes of SPs consist of payment time (z_1') and credibility (z_2'). The delivery time represents the days required for processing and transportation; payment time is how many days SPs would get 95% of the full amount after a welding task is accepted (the other 5% is usually a warranty deposit).

4.1.1. Expression of Expectations Based on IVHFLSs. First, the SDs and SPs evaluate their expectations and self-conditions. For cost, delivery time, and payment time, which are easy to quantify directly, SDs and SPs can evaluate them with numbers. For credibility, which is difficult to quantify directly, SDs and SPs evaluate it with seven-level linguistic variables: L = {poorest, poorer, poor, general, good, better, best}. Then, the SDs and SPs use intervals to express the membership of linguistic levels. The evaluation results are shown in Tables 1 and 2, respectively. After the discussion of experts, the lower thresholds of cost, delivery time, and payment time are set to 200, 10, and 3, respectively. The overflowing coefficient of satisfaction $K = 1.2$.

4.1.2. Quantification of Expectations Based on Score Function. According to the linguistic scale function $f(l_t) = t/2T$, $(t = 0, 1, 2, \ldots, 2T)$ and equations (2)–(7), the satisfaction of SDs D_1–D_6 with service provider P_1 can be obtained, as shown in Table 3.

Similarly, the satisfaction of SDs D_2–D_6 with the SPs P_1–P_9 and the satisfaction of SPs P_1–P_9 with SDs D_1–D_6 can be obtained. If the weights of cost, delivery time, and credibility are set to $w_1^D = 0.3$, $w_2^D = 0.3$, and $w_3^D = 0.4$, respectively. The weights of payment time and credibility are set to $w_1^P = 0.5$, and $w_2^P = 0.5$, respectively. Then, the satisfaction matrices $H^{D\longrightarrow P}$ and $H^{P\longrightarrow D}$ can be obtained, as shown in Tables 4 and 5.

4.1.3. Obtaining the Soundest Alternatives. Based on the stable two-sided matching model shown in equation (13) and the proposed AGA, the soundest alternative was obtained.

The proposed AGA was programmed in Matlab 2014a. The initial population was set to 100, the maximum generation to 200, the minimum crossover probability to 0.7, the maximum crossover probability to 0.9, the minimum mutation probability to 0.08, and the maximum mutation probability to 0.1. After solution, the ideal optimal values of Z_1 and Z_2 were 6.517 and 6.670, respectively. The ideal worst values of Z_1 and Z_2 were 2.761 and 1.990, respectively. After solution, the best fitness values of every generation are shown in Figure 3.

When the algorithm converges, the optimal value of fitness function f_1 was 0.837, the optimal value of fitness function f_2 was 6, the corresponding optimal solution was $\lambda^* = \{(D_1, P_2), (D_2, P_1), (D_3, P_4), (D_4, P_7), (D_5, P_6), (D_6, P_8)\}$, and the corresponding optimization objectives were $Z_1 = 5.924$, $Z_2 = 5.889$, and $Z_3 = 6$.

4.2. Comparisons and Discussion. To further verify the superiority, the proposed two-sided matching model was compared with the traditional matching model proposed by Li et al. [15], which takes the values of satisfaction indexes as optimization objectives. Let $h_{i,j,m}$ represent the satisfaction expression of the service demander published task D_i on service P_j for index z_m based on IVHFLSs, $h_{j,i,n}$ indicate the satisfaction expression of the service provider published service P_j on task D_i for index z_n' based on IVHFLSs, $s(h_{i,j,m})$ and $s(h_{j,i,n})$ mean the score of $h_{i,j,m}$ and $h_{j,i,n}$, respectively. Hence, the satisfaction optimization goal of SDs based on the thought proposed by Li et al. [15] can be expressed as $\max Z_1 = \sum_{i=1}^{I}\sum_{j=1}^{J} w_m s(h_{i,j,m}) y_{ij}$. In a similar way, the satisfaction optimization goal of SPs can be expressed as $\max Z_2 = \sum_{i=1}^{I}\sum_{j=1}^{J} w_n s(h_{j,i,n}) y_{ij}$. The function of optimal matching numbers can be expressed as

Table 1: Evaluation results of satisfaction indexes of SDs.

		Expected values ($h_{i,m}^{D}$)	Tolerant values ($h_{i,m}^{T}$)	Payment time ($g_{i,1}^{D}$)	Credibility ($g_{i,2}^{D}$)
	z_1	400	500		
D_1	z_2	30	35	5	<better, {[0.5, 0.7], [0.8, 0.9]}>
	z_3	<good, {[0.7, 0.8]}>	<general, {[0.5, 0.6]}>		
	z_1	300	500		
D_2	z_2	30	35	8	<better, {[0.7, 0.9]}>
	z_3	<good, {[0.7, 0.9]}>	<general, {[0.5, 0.6]}>		
	z_1	500	600		
D_3	z_2	30	35	8	<better, {[0.7, 0.8]}>
	z_3	<better, {[0.6, 0.7]}>	<general, {[0.5, 0.6]}>		
	z_1	500	600		
D_4	z_2	20	30	10	<good, {[0.6, 0.8]}>
	z_3	<good, {[0.7, 0.8] }>	<general, {[0.5, 0.6]}>		
	z_1	400	500		
D_5	z_2	30	40	7	<better, {[0.5, 0.6], [0.7, 0.8]}>
	z_3	<good, {[0.4, 0.5]}>	<general, {[0.5, 0.6]}>		
	z_1	300	500		
D_6	z_2	15	20	5	<good, {[0.3, 0.6], [0.7, 0.9]}>
	z_3	<better, {[0.4, 0.6]}>	<general, {[0.5, 0.7], [0.8, 0.9]}>		

Table 2: Evaluation results of satisfaction indexes of SPs.

		Expected values ($h_{j,n}^{P}$)	Tolerant values ($h_{j,n}^{T}$)	Income ($g_{j,1}^{P}$)	Delivery time ($g_{j,2}^{P}$)	Credibility ($g_{j,3}^{P}$)
P_1	z_1'	7	10	300	20	<good, {[0.7, 0.9]}>
	z_2'	<good, {[0.7, 0.9]}>	<general, {[0.6, 0.8]}>			
P_2	z_1'	6	10	300	30	<good, {[0.7, 0.8], [0.8, 0.9]}>
	z_2'	<better, {[0.4, 0.6], [0.7, 0.8]}>	<general, {[0.5, 0.6]}>			
P_3	z_1'	6	10	500	30	<general, {[0.6, 0.8], [0.8, 0.9]}>
	z_2'	<better, {[0.4, 0.5], [0.5, 0.7]}>	<general, {[0.5, 0.6]}>			
P_4	z_1'	7	10	500	20	<better, {[0.6, 0.7], [0.7, 0.8}>
	z_2'	<good, {[0.8, 0.9]}>	<general, {[0.5, 0.6]}>			
P_5	z_1'	7	10	600	20	<best, {[0.5, 0.6], [0.6, 0.7]}>
	z_2'	<good, {[0.7, 0.8], [0.8, 0.9]}>	<general, {[0.6, 0.8]}>			
P_6	z_1'	10	15	300	30	<better, {[0.4, 0.5]}>
	z_2'	<general, {[0.7, 0.8]}>	<general, {[0.2, 0.4]}>			
P_7	z_1'	10	15	300	15	<good, {[0.5, 0.6]}>
	z_2'	<general, {[0.6, 0.7], [0.7, 0.8]}>	<general, {[0.3, 0.4]}>			
P_8	z_1'	5	10	300	15	<best, {[0, 4, 0.6], [0.6, 0.7]}>
	z_2'	<general, {[0.5, 0.7], [0.8, 0.9]}>	<general, {[0.3, 0.4]}>			
P_9	z_1'	5	10	400	20	<good, {[0.5, 0.6]}>
	z_2'	<better {[0.6, 0.7], [0.7, 0.8]}>	<good, {[0.4, 0.5]}			

Table 3: Satisfaction degree of D_1–D_6 with P_1.

Satisfaction degree	Cost	Delivery time	Credibility	Satisfaction degree	Cost	Delivery time	Credibility
$\phi_{11}^{D\longrightarrow P}$	1.126	1.126	1.001	$\phi_{41}^{D\longrightarrow P}$	1	1	1.001
$\phi_{21}^{D\longrightarrow P}$	1	1.126	0.998	$\phi_{51}^{D\longrightarrow P}$	1.126	1.126	1.098
$\phi_{31}^{D\longrightarrow P}$	1.173	1.126	0.966	$\phi_{61}^{D\longrightarrow P}$	1	0	1.047

Table 4: Satisfaction matrix $H^{D\longrightarrow P}$ of SDs D_1–D_6 with SPs P_1–P_9.

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9
D_1	1.076	1.038	0.447	0.758	0.767	0.753	0.788	1.110	0.731
D_2	1.037	0.999	0.409	0.746	0.756	0.682	0.722	1.061	0.513
D_3	1.076	1.038	0.702	1.043	0.752	0.727	0.768	1.111	0.721
D_4	1.000	0.754	0.390	0.865	0.729	0.415	0.731	1.053	0.575
D_5	1.115	1.077	0.702	0.791	0.795	1.039	1.095	1.146	1.038
D_6	0.719	0.719	0.125	0.440	0.446	0.314	0.600	1.036	0.110

Table 5: Satisfaction matrix $H^{P\longrightarrow D}$ of SPs P_1–P_9 with SDs D_1–D_6.

	D_1	D_2	D_3	D_4	D_5	D_6
P_1	1.089	0.881	0.847	0.282	1.022	0.650
P_2	1.071	0.764	0.733	0.377	0.858	0.779
P_3	1.102	0.772	0.756	0.504	0.892	1.039
P_4	1.071	0.875	0.831	0.297	0.955	0.737
P_5	1.089	0.881	0.847	0.282	1.000	0.650
P_6	1.158	1.122	1.107	1.008	1.091	1.092
P_7	1.160	1.123	1.109	1.014	1.097	1.093
P_8	1.067	0.701	0.687	0.511	0.798	1.001
P_9	1.001	0.679	0.628	0.248	0.678	0.620

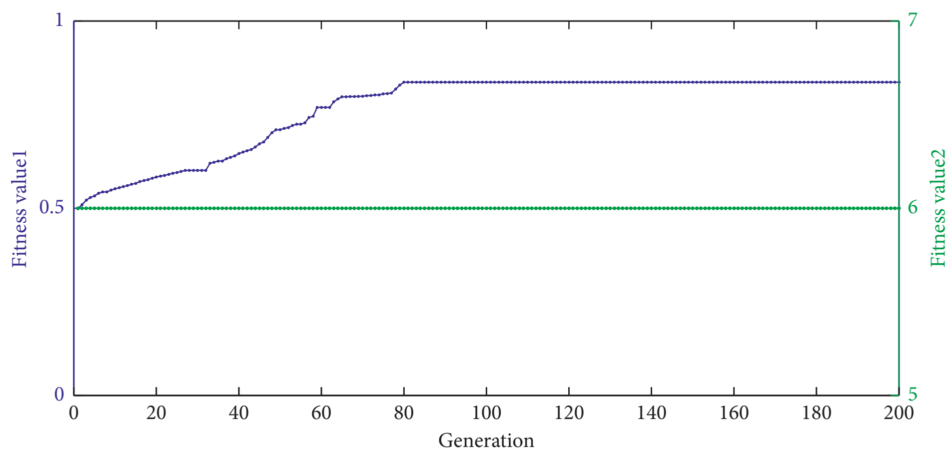


Figure 3: Best fitness value of every generation.

$\max Z_3 = \sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij}$. Combining the constraints in Li et al. [15], the compared matching model is described as follows:

$$\begin{aligned}
\max Z_1 &= \sum_{i=1}^{I}\sum_{j=1}^{J} w_m s\left(h_{i,j,m}\right) y_{ij},\\
\max Z_2 &= \sum_{i=1}^{I}\sum_{j=1}^{J} w_n s\left(h_{j,i,n}\right) y_{ij},\\
\max Z_3 &= \sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij},\\
&\sum_{i=1}^{m} y_{ij} = 1, \sum_{j=1}^{n} y_{ij} \le 1,\\
&y_{ij} + \sum_{f: w_m s\left(h_{i,f,m}\right) < w_m s\left(h_{i,j,m}\right)} y_{if}\\
&+ \sum_{b: w_n s\left(h_{j,b,n}\right) < w_n s\left(h_{j,i,n}\right)} y_{bj} \ge 1,\\
&y_{ij} \in \{0,1\}, \quad i \in I, j \in J.
\end{aligned} \tag{18}$$

Then, the proposed AGA was programmed to solve the matching model proposed by Li et al. [15]. The parameters here are set same as the parameters used when solving the matching model proposed by this article. After solving, the optimal solution was $\lambda^{*'} = \{(D_1, P_1), (D_2, P_2), (D_3, P_4), (D_4, P_7), (D_5, P_6), (D_6, P_8)\}$. The corresponding optimization results were $Z_1 = 5.924, Z_2 = 5.790, Z_3 = 6$. Specifically, the satisfaction of SDs on SPs in the matching scheme is $\phi_{11}^{D\longrightarrow P} = 1.076$, $\phi_{22}^{D\longrightarrow P} = 0.999$, $\phi_{34}^{D\longrightarrow P} = 1.043$, $\phi_{47}^{D\longrightarrow P} = 0.731$, $\phi_{56}^{D\longrightarrow P} = 1.039$, $\phi_{68}^{D\longrightarrow P} = 1.036$ and the satisfaction of SPs on SDs in the matching scheme is $\phi_{11}^{P\longrightarrow D} = 1.089$, $\phi_{22}^{P\longrightarrow D} = 0.764$, $\phi_{34}^{P\longrightarrow D} = 0.831$, $\phi_{47}^{P\longrightarrow D} = 1.014$, $\phi_{56}^{P\longrightarrow D} = 1.091$, $\phi_{68}^{P\longrightarrow D} = 1.001$.

As shown in Figure 4, the satisfaction of SDs optimized by this article is equal to the satisfaction of SDs optimized by Li et al. [15]. However, the satisfaction of SPs optimized by this article is increased 0.099 than the satisfaction of SPs optimized by Li et al. [15]. This indicates that the matching model proposed by this article is superior to the traditional model proposed by Li et al. [15].

Overall, satisfaction of SDs and SPs is an important property for improving the competitiveness of CMfg platforms. Although the two-sided matching problem of CMfg tasks and service has been studied in some articles, they took satisfaction indexes or expectations as the optimization goals of the matching model. Whereas, satisfaction is not linear with expectations. In fact, there is a limit to satisfaction, the bigger the satisfaction index, the higher the satisfaction is not necessarily in a range. In this article, the nonlinear model of satisfaction and expectations is built first. The satisfaction of

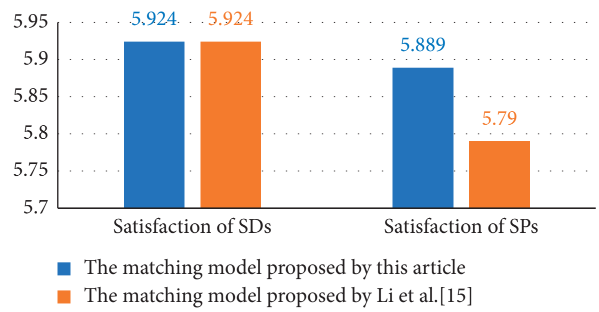


Figure 4: Comparison of different methods.

SDs and SPs is taken as the optimization objectives of the two-sided matching model of CMfg tasks and service, thus making the optimization of satisfaction more effective.

Second, stability is necessary to consider when building the two-sided matching model, or else SDs and SPs may give up the current matching results [15]. In this article, stable matching conditions are considered when establishing the matching model, which not only can maximize satisfaction degrees of SDs and SPs but also can derive the stable matching results.

Third, IVHFLSs are introduced to SDs and SPs to express their expectations, tolerant values, and so on, which effectively describe the uncertainty, hesitancy, and inconsistency inherent in the decision-making process.

Therefore, the proposed matching model enriches the matching studies of CMfg tasks and service. However, the article also has some limitations. First, it cannot be used to deal with the conditions when some complex linguistic expressions are needed, such as multigranular unbalanced linguistic expression [29]. Second, it is assumed that the weight of satisfaction indexes is given by crisp numbers in advance. However, sometimes SDs and SPs may tend to provide criteria weight information with linguistic expressions, like "satisfaction index A is more important than in index B."

5. Conclusions

The matching of CMfg tasks and service is an essential issue in the field of CMfg. Improving the satisfaction of SDs and SPs is significant to enhance the competitiveness of CMfg platforms. A nonlinear model of satisfaction and expectations and a stable two-sided matching model of CMfg tasks and service are proposed in this article. The proposed nonlinear model achieves the accurate quantification of satisfaction. The proposed two-sided matching model not only takes the satisfaction of SDs and SPs as the optimization objectives, but also considers the rationality of SDs and SPs and blocking pairs in the matching schemes, thereby improving the effectiveness.

In the future research, the following studies can be further conducted:

(1) For the matching of CMfg tasks and service, due to different cultural and knowledge backgrounds of SDs and SPs, complex linguistic expressions like multigranular unbalanced linguistic expression may be needed. Therefore, it is necessary to develop new models with complex linguistic expressions.

(2) In the proposed matching model, the weight of satisfaction indexes are given by SDs and SPs with crisp numbers. However, sometimes SDs and SPs may tend to provide criteria weight information with linguistic expressions in the real-world applications. Hence, how to deal with the weights of satisfaction indexes with linguistic expressions needs further study.

Data Availability

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

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 71971130).

References

[1] F. Tao, Y. Cheng, L. Zhang, and A. Y. C. Nee, "Advanced manufacturing systems: socialization characteristics and trends," *Journal of Intelligent Manufacturing*, vol. 28, no. 5, pp. 1079–1094, 2017.
[2] Y. Yin, J. Zhou, C. Zhang, and D. Chen, "Adaptive SLA mechanism for service sharing in virtual environments," *Kybernetes*, vol. 45, no. 7, pp. 1036–1051, 2016.
[3] Y. Xiao, C. Li, L. Song, J. Yang, and J. Su, "A multidimensional information fusion-based matching decision method for manufacturing service resource," *IEEE Access*, vol. 9, pp. 39839–39851, 2021.
[4] X. L. Xie, L. Liu, and Y. Z. Cao, "Trust model based on feedback evaluation in cloud manufacturing environment," *Advanced Materials Research*, vol. 308–310, pp. 1740–1745, 2011.
[5] J. Liu, H. Zhao, Z. Xu, and Z. Wu, "Research on demand-overflowing customer satisfaction model," *China Mechanical Engineering*, vol. 24, no. 21, pp. 2880–2884, 2013.
[6] Y. K. Lin and C. S. Chong, "Fast GA-based project scheduling for computing resources allocation in a cloud manufacturing

system," *Journal of Intelligent Manufacturing*, vol. 28, pp. 1–13, 2017.

[7] Z. Zhao, "The research on cloud manufacturing service discovery strategy based on mapping reuse," *Advanced Materials Research*, vol. 860–863, pp. 2898–2901, 2013.

[8] J. Lartigau, X. F. Xu, L. S. Nie, and D. C. Zhan, "Cloud manufacturing service composition based on QoS with geo-perspective transportation using an improved artificial bee colony optimisation algorithm," *International Journal of Production Research*, vol. 53, no. 14, pp. 1–25, 2015.

[9] Y. F. Zhang, D. Xi, R. Li, and S. D. Sun, "Task-driven manufacturing cloud service proactive discovery and optimal configuration method," *The International Journal of Advanced Manufacturing Technology*, vol. 84, no. 1, pp. 29–45, 2016.

[10] F.-S. Hsieh and J.-B. Lin, "A dynamic scheme for scheduling complex tasks in manufacturing systems based on collaboration of agents," *Applied Intelligence*, vol. 41, no. 2, pp. 366–382, 2014.

[11] Y. H. Xiong, J. Wang, M. Wu, and J. She, "Virtual resource scheduling method of cloud manufacturing oriented to multi-objective optimization," *Computer Integrated Manufacturing Systems*, vol. 21, no. 11, pp. 3079–3087, 2015.

[12] N. R. Joglekar and D. N. Ford, "Product development resource allocation with foresight," *European Journal of Operational Research*, vol. 160, no. 1, pp. 72–87, 2005.

[13] S. L. Wang, Z. Q. Zhu, and L. Kang, "Resource allocation model in cloud manufacturing," *ARCHIVE Proceedings of the Institution of Mechanical Engineers Part C Journal of Mechanical Engineering Science*, vol. 230, no. 10, pp. 203–210, 2016.

[14] J. H. Zhao and X. H. Wang, "Two-sided matching model of cloud service based on QoS in cloud manufacturing environment," *Computer Integrated Manufacturing Systems*, vol. 22, no. 1, pp. 104–112, 2016.

[15] B. Li, Y. Yang, J. Su, N. Zhang, and S. Wang, "Two-sided matching model for complex product manufacturing tasks based on dual hesitant fuzzy preference information," *Knowledge-Based Systems*, vol. 186, Article ID 104989, 2019.

[16] L. Ren and M. Ren, "One-to-many two-sided matching method of wisdom manufacturing task based on competition and synergy effect," *Computer Integrated Manufacturing Systems*, vol. 24, no. 5, pp. 50–63, 2018.

[17] D. Zhao and L. Ding, "Two-sided re-source matching mechanism and stability of cloud manufacturing platform," *Systems Engineering*, vol. 35, no. 2, pp. 109–115, 2017.

[18] Y. Li, X. K. Wang, J. Q. Wang, J. B. Li, and L. Li, "Probability distribution-based processing model of probabilistic linguistic term set and its application in automatic environment evaluation," *International Journal of Fuzzy Systems*, 2021.

[19] J.-Y. Dong, F.-F. Yuan, and S.-P. Wan, "Extended VIKOR method for multiple criteria decision-making with linguistic hesitant fuzzy information," *Computers & Industrial Engineering*, vol. 112, pp. 305–319, 2017.

[20] Z. Zhang, J. Gao, Y. Gao, and W. Yu, "Two-sided matching decision making with multi-granular hesitant fuzzy linguistic term sets and incomplete criteria weight information," *Expert Systems with Applications*, vol. 168, Article ID 114311, 2021.

[21] X.-B. Mao, S.-S. Hu, J.-Y. Dong, S.-P. Wan, and G.-L. Xu, "Multi-attribute group decision making based on cloud aggregation operators under interval-valued hesitant fuzzy linguistic environment," *International Journal of Fuzzy Systems*, vol. 20, no. 7, pp. 2273–2300, 2018.

[22] J.-Q. Wang, J.-T. Wu, J. Wang, H.-Y. Zhang, and X.-H. Chen, "Interval-valued hesitant fuzzy linguistic sets and their applications in multi-criteria decision-making problems," *Information Sciences*, vol. 288, pp. 55–72, 2014.

[23] Z. Zhang, X. Kou, W. Yu, and Y. Gao, "Consistency improvement for fuzzy preference relations with self-confidence: an application in two-sided matching decision making," *Journal of the Operational Research Society*, vol. 72, no. 8, pp. 1–14, 2020.

[24] K. Zhou, X. Han, X. Zhu, and F. Liu, "Research on customization model based on customer satisfaction," *Computer Integrated Manufacturing Systems*, vol. 10, no. 11, pp. 1338–1342, 2004.

[25] D. Gale and L. S. Shapley, "College admissions and the stability of marriage," *The American Mathematical Monthly*, vol. 69, no. 1, pp. 9–15, 1962.

[26] D. E. Goldberg and K. Deb, "A comparative analysis of selection schemes used in genetic algorithms," *Foundations of Genetic Algorithms*, vol. 42, pp. 69–93, 1991.

[27] K.-M. Lee, J.-T. Tsai, T.-K. Liu, and J.-H. Chou, "Improved genetic algorithm for mixed-discrete-continuous design optimization problems," *Engineering Optimization*, vol. 42, no. 10, pp. 927–941, 2010.

[28] J. F. Su, F. T. Zhang, S. Chen, N. Zhang, H. Wang, and J. Jian, "Member selection for the collaborative new product innovation teams integrating individual and collaborative attributions," *Complexity*, vol. 2021, Article ID 8897784, 14 pages, 2021.

[29] Z. Zhang, Z. Li, and Y. Gao, "Consensus reaching for group decision making with multi-granular unbalanced linguistic information: a bounded confidence and minimum adjustment-based approach," *Information Fusion*, vol. 74, pp. 96–110, 2021.