**RESEARCH ARTICLE**

# Research on Data-Driven Fresh Produce Joint Distribution Network Optimization Under Distribution Center Sharing

## MEILIN ZHU AND XIAOYE ZHOU

School of Management, Shenyang University of Technology, Shenyang 110870, China

Corresponding author: Xiaoye Zhou (928911873@qq.com)

**ABSTRACT** To address the problem of low efficiency and ineffective utilization of resources in the distribution of fresh produce at the end of the city, and taking into account the seasonal characteristics of the logistics demand of fresh produce. An innovative rolling adjustment framework model based on data-driven for optimizing fresh produce joint distribution network was proposed, which follows seasonal changes. The cycle of the adjustment framework was divided according to seasonal changes, with each cycle including four steps: fresh produce logistics demand prediction, fresh produce joint distribution network optimization, data collection, and parameters adjustment of the prediction model, further to achieve data-driven optimization of joint distribution network for fresh produce. A catastrophe adaptive genetic algorithm with variable neighborhood search (CAGA-VNS) is developed to solve the fresh produce joint distribution network optimization model. Finally, several numerical experiments are conducted to validate the model and algorithm. The results demonstrate that: 1) the rolling adjustment framework model can provide effective fresh produce distribution network optimization decisions when the fresh produce demand changes according to the season changes. 2) The CAGA-VNS algorithm can be more stable with the lowest difference percentage being 9.16%. 3) distribution center sharing strategy can effectively improve utilization of resources, reduce the total distribution cost of 19.46% and save travel distance of 14.72%.

**INDEX TERMS** Fresh produce, distribution network optimization, data-driven, distribution center sharing, CAGA-VNS, seasonality.

## I. INTRODUCTION

With the development of big data science, a new change has taken place in the traditional management decision paradigm, and a new management decision paradigm has emerged— data-driven management decision paradigm [1]. In the past decades companies have invested heavily in information technologies and digitalization. Increased computing power and the explosion of data have changed the way organisations capture data, analyse information, and make decisions. These changes provide opportunities for the operations management community to develop new models for data driven decision-making, and might become more and more applicable to logistics and supply chain management. However,

The associate editor coordinating the review of this manuscript and approving it for publication was Liandong Zhu.

the research on how to integrate the big data analysis and OR models to drive business performance improvements is still in its infancy [2].

Data driven demand prediction of fresh produce and optimization of distribution based on demand prediction of fresh produce has been proven to of great value to fresh produce distribution [3]. Such optimization problem of fresh produce distribution decisions is often theoretically optimal due to assumptions that demand is deterministic or follows certain known statistical distributions. However, in real-world fresh produce distribution, the distribution process can be divided into multiple stages through corresponding information collection and fresh produce demand prediction can be realized with the help of data science during the fresh produce distribution. Hence, issues concerning how to make fresh produce joint network optimization decisions under restrained

distribution resource and periodically acquired fresh produce demand and its corresponding features information should be addressed.

This paper is specifically motivated by making fresh produce distribution network optimization decisions in real-world fresh produce distribution. The logistics demand for fresh produce is booming sharply, however, fresh produce logistics distribution network in last mile has become one of the bottlenecks in the development of fresh produce industry, which cannot provide corresponding services for the existing demand. There are two obvious reasons existing: on one hand, fresh produce has short shelf lives and the change characteristics of maturity and seasonality, which is different from other produce. Enterprises have had to adapt quickly and respond to the change of demand given their special seasonality; On the other hand, the characteristics of urban last-miles' customers' dispersion and small batch make the logistics distribution of agricultural products more difficult, with repeated transportation, roundabout transportation and other unreasonable transportation situations existing.

Joint distribution is an effective mean of solving this problem. Joint distribution can improve the distribution efficiency of fresh produce by integrating distribution resources. Accordingly, higher requirements are put forward to the optimization of urban last-mile fresh produce distribution network. In order to further improve the distribution efficiency of fresh produce and respond to the constant adjustment of fresh produce demand, how to make rural fresh produce quickly move to the dining table with the help of accurate and fast logistics distribution network is a problem that we urgently need to solve at present [4].

To address the above challenges, the main contribution of our study is that an innovative data-driven urban last-mile fresh produce joint distribution network is designed. Considering the seasonality of fresh produce logistics demand, from the perspective of data-driven, this paper innovatively formulates a rolling adjustment framework model of fresh produce joint distribution network optimization that follows the seasonal change. The rolling adjustment framework model in each decision-making cycle includes the following parts: fresh produce logistics demand prediction, fresh produce joint distribution network optimization, data collection and processing, and parameters adjustment of the prediction model. More specifically for the above rolling adjustment framework model, in each decision-making cycle, the agricultural demand prediction model is fitted. Then a mixed integer programming is applied to solve a fresh produce joint distribution network optimization problem (FPJDOP) based on agricultural demand prediction. While the optimal FPJDOP decisions derived, a rolling-horizon manner is taken and implemented only the optimal solution for the immediately next cycle, the uncertain parameters of the prediction model are also updated to obtain the optimal demand for the next cycle. For the part of constructing the optimization model of the joint distribution network, based on the strategy of distribution center sharing, the sum of the total cost was minimized

by optimizing the minimum number of distribution centers and the shortest delivery distance of vehicles. To solve the optimization model, this paper proposed a catastrophe adaptive genetic algorithm with variable neighborhood search.

The rest of this paper is organized as follows. Section II reviews the literature and fills the research gap. Section III introduces the data-driven fresh produce joint distribution network optimization rolling adjustment framework model and presents the mathematical formulation of fresh produce joint distribution network. Section IV shows the solving approach of the formulation in Section III and a catastrophe adaptive genetic algorithm with variable neighborhood search is developed. Section V provides the computational results and analyses the performance of the solving approach. Finally, Section VI gives a summary of our major contributions and delineates the future research directions.

## II. LITERATURE REVIEW

Given that this work intends to explicitly focus on data-driven fresh produce joint distribution network optimization under distribution center sharing, the relevant literature is categorized into three main streams, which include fresh produce distribution network optimization, joint distribution network optimization and data driven optimization. Afterwards, based on the existing literature and analysing the research gaps, the main contributions of this study are presented.

### A. FRESH PRODUCE DISTRIBUTION NETWORK OPTIMIZATION

Fresh produce distribution logistics network has been proven of great value to fresh produce distribution [5], [6], [7], [8]. Motivated by reducing carbon emission and promoting the development of fresh e-commerce. Chen et al. [9]studied the fresh produce multi-compartment vehicle routing problem with time window (MCVRPTW) and proposed a variable neighborhood search (VNS) approach with the key steps of local search and shaking to solve the problem. Han et al. [10] explored the impact of the agricultural Internet of Things (IOT) and the special requirements of each distribution channel on joint planning of production and distribution of fresh produce. Ma et al. [11]discussed a perishable food location-routing problem with conflict and coordination (PFLRP-CC) between a fresh food seller and a hired transportation company and developed a hybrid algorithm to solve the problem. Zhang et al. [12]established a multi-objective optimization model, including minimizing distribution cost, freshness of fresh produce, carbon emissions. The main target method and fruit fly algorithm were used to solve it. Wei and Wang [13]designed a fuzzy C means clustering-improved simulated annealing (FCM-ISA) algorithm to solve continuous location problem of fresh produce distribution centers. Bortolini et al. [14]extended the problem to multi-modal distribution networks of fresh produce to minimize operating cost, carbon footprint and delivery time goals. A unique tool called Food Distribution

Planner is applied to deal with the distribution of fresh fruits and vegetables based on a set of Italian producers to several European retailers. Jaigirdar et al. [15] investigated a multi-objective optimization model of distribution network for perishable goods. The CPLEX solver and a weighted sum method were considered to solve it. Golestani et al. [16] formulated a mixed-integer linear programming (MINLP) for hub location problem, considering multi-item and multi temperature joint distribution for perishable produce. The GAMS software and $\varepsilon$-Constraint method were adopted to solved this bi-objective model.

It is important to note that the existing publications only address the perishable nature and short shelf lives of fresh produce and do not address importance the seasonality nature on fresh produce distribution. For example, in the north of China, fresh produce is harvested once a year, while fresh produce is harvested twice a year or even more in the south of China. If the demand is predicted not accurately, the fresh produce distribution network configuration will not be designed reasonably. By taking the gaps into consideration, considering the seasonality of fresh produce, this paper establishes innovatively a rolling adjustment framework model which optimize the fresh produce joint distribution network from the perspective of data-driven.

### B. JOINT DISTRIBUTION NETWORK OPTIMIZATION

Sheng et al. [17] focused on rural electronic commerce logistics distribution, introduced a joint distribution strategy sharing customers and vehicles to address the simultaneous pick-up and delivery. Wang et al. [18] formulated a mathematical model of multi-depot green vehicle routing problem based on a vehicle sharing strategy. The Clarke and Wright savings heuristic algorithm combined with the sweep Algorithm and the multi-objective particle swarm optimization algorithm was designed. Li et al. [19] found shared depot resources more beneficial for multi-depot vehicle routing problem compared to unshared depots strategy. Wang et al. [20] investigated a collaborative multiple centers vehicle routing problem with simultaneous delivery and pick-up (CVRPSDP) which shared vehicles and customers. The adoption of k-means and non-dominated sorting genetic algorithm-II (NSGA-II) constituted a hybrid algorithm to solve the problem. In the light of the issues low loading rate and cost of distribution, Wang [21] came up with a joint distribution pattern which is based on cargo resources and remaining vehicle resources. An improved genetic algorithm with dynamic parameters was built to solve the model. In comparison with the published articles in the domain, the essence of joint distribution is the integration of resource such as depots, vehicles, customers or cargos.

### C. DATA DRIVEN OPTIMIZATION

In recent years, data-driven optimization approaches are widely used to management decisions or operations research, which can better improve efficiency or reduce costs with the help of data resource and big data analysis. Liu et al. [22]

highlighted the fuzzy boundary and time-varying decision scenarios of an expected epidemic outbreak and proposed an innovative decision framework for optimizing the epidemic-logistics network based on data-driven. Under this new decision framework, the entire emergency response process could be converted to an interactive evolution process of data learning and resource optimization. Du et al. [23] investigated Cholera has clear spatial variation in its transmission pattern, proposed a data-driven optimization approach to determine the optimal strategy of intervention resource allocation at each period and each community in a rolling-horizon manner. At each period, they integrated single-period model parameter fitting and scenario-based on stochastic programming to make decisions under uncertainty with newly acquired system understanding. Du [24] held the opinion of the interplay of epidemic detection and health-care resource allocation requiring joint decision optimization. A multi-stage joint decision-making approach for data-driven outbreak detection and dynamic allocation of healthcare resource was proposed. Xiang [25] extended [22] research study, considering the limited resource of the government and the need for emergency procurement from external suppliers, aiming at the problem of choosing the optimal time point for the start of emergency procurement, the rule for judging the optimal time for stopping (or starting procurement) of epidemic observation was designed, and the boundary characteristics and influencing factors of the optimal stopping time were obtained by combining with theoretical analysis, and a data-driven model was constructed based on the cyclic decision-making idea of "epidemic prediction, emergency effect comparison, stop-time judgment, and parameter update". A data driven methodology which employed the empirical risk minimisation (ERP) principle was introduced in the inventory management, Clausen and Li [26] formulated a big data driven dynamic order-up-to level inventory model which took multiple features into account and did not need classical distributional assumptions and designed a machine learning algorithm to solve the model. In view of the problem of unseasonable order assignments and long distribution distance, taking *Meituan* and *Eleme* as examples, Xiong and Yan [27] analysed the mechanism of dynamic real-time optimization and the mechanism of algorithm optimization mechanism for the intelligent delivery order dispatching on the delivery platform. To tackle the challenge of uncertain delivery time in last-mile food delivery, Chu et al. [28] proposed a data-driven optimization approach that combined machine learning techniques with capacitated vehicle routing problem and used a new smart predict-then-optimize framework. An efficient mini-batching gradient and heuristic algorithms were designed to solve the joint order assignment and routing problem. However, fresh produce distribution network optimization based on data-driven is not common.

### D. GAP ANALYSIS AND CONTRIBUTIONS

To sum up, there have been certain achievements in research on the optimization of fresh produce distribution network

optimization, joint distribution, and data-driven optimization, but there are still some research gap existing: (1) no scholar has combined the optimization of fresh produce joint distribution network under distribution center sharing strategy with data-driven optimization approach, so as to improve the delivery efficiency of fresh produce; (2) no scholar has included the impact of seasonality in fresh produce logistics demand into the optimization of fresh produce distribution network; (3) scholars always use sensitivity analysis to discuss the impact of fresh produce demand on the optimization of fresh produce distribution network. However, fresh produce prediction and fresh produce joint distribution network optimization are multi-stage decision-making optimization problems with mutual influence and coupling, which need decision-making optimization jointly. In response to the gaps of the above research, an innovative rolling adjustment framework model based on data-driven for optimizing fresh produce joint distribution network was proposed, which follows seasonal changes. The cycle of the rolling adjustment framework model was divided according to seasonal changes, with each cycle including four steps: fresh produce logistics demand prediction, fresh produce joint distribution network optimization, data collection, and parameters adjustment of the prediction model. Based on the strategy of sharing distribution centers, a catastrophe adaptive genetic algorithm with variable neighborhood search (CAGA-VNS) is proposed. The crossover operator and mutation operator are adaptively set to avoid premature convergence. Whereas dynamic catastrophe mechanism and neighborhood search structures are introduced to improve the quality of solution, numerical experiments are taken to test the effectiveness and superiority of the algorithm.

## III. FRAMEWORK MODEL FORMULATION

This section explains the investigated fresh produce joint distribution network rolling adjustment framework model based on data-driven. The rolling adjustment framework model involves the following four steps. First, the support vector machine optimized by the grey wolf optimizer and nonlinear principal component analysis (NLPCA) is used to forecast seasonal logistics demand of fresh produce in the future one cycle. Secondly, according to the demand forecasting results, the CAGA-VNS algorithm is designed to further optimize the joint distribution network of fresh produce (FPJDOP). Then, executing the real-world fresh produce distribution network according to the computational results of FPJDOP, collect and record the new actual data of fresh produce logistics demand and its corresponding data in real-world distribution network. Finally, new actual data will be incorporated into the prediction model, and the parameters of the prediction model will be adjusted to improve the accuracy of the prediction model, as shown in Figure 1.

### A. A BRIEF REVIEW OF GREY WOLF OPTIMIZER (GWO)
The GWO is a swarm intelligent algorithm that simulates the group hunting behaviour of grey wolf packs. A social
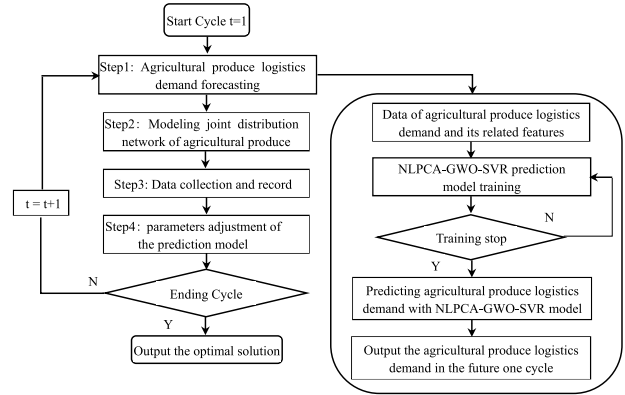


**FIGURE 1.** The rolling adjustment framework model of fresh produce joint distribution network optimization based on data-driven.

hierarchy exists in a wolf pack; to model this when designing the GWO, the algorithm regards the fittest solution in each iteration as the alpha wolf ($\alpha$). Sequentially, the second- and third-best solutions are named beta ($\beta$) and delta ($\delta$) wolves, respectively. The rest of the candidate solutions are assumed to be $\varphi$ wolves. In the GWO algorithm, the hunting behaviour is guided by the $\alpha$, $\beta$ and $\delta$ wolves, which the $\varphi$ wolves follow.

A mathematical model of wolf encircling behaviour is represented by the equations:

$$\vec{D} = \left| \vec{H} \cdot \vec{X}_p(t) - \vec{X}(t) \right| \tag{1}$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D} \tag{2}$$

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \tag{3}$$

$$\vec{H} = 2 \cdot \vec{r}_2 \tag{4}$$

where $\vec{D}$ denotes the distance between the prey and the wolf, $t$ is the current iteration, $\vec{X}_p(t)$ is the position of the prey at iteration $t$, $\vec{X}$ is the position vector of a wolf, $r_1$ and $r_2 \in [0,1]$ are random vectors and vary linearly from 2 to 0, $\vec{A}$ and $\vec{H}$ are the coefficient vectors of wolves $\alpha$, $\beta$ and $\delta$, $\vec{A}$ and $\vec{H}$ are updated by (3) and (4).

The mathematical model of hunting behaviour is represented by the equations:

$$\begin{cases} \vec{D}_\alpha = \left| \vec{H}_1 \cdot \vec{X}_\alpha(t) - \vec{X}(t) \right| \\ \vec{D}_\beta = \left| \vec{H}_2 \cdot \vec{X}_\beta(t) - \vec{X}(t) \right| \\ \vec{D}_\delta = \left| \vec{H}_3 \cdot \vec{X}_\delta(t) - \vec{X}(t) \right| \end{cases} \tag{5}$$

$$\begin{cases} \vec{X}_1 = \vec{X}_\alpha(t) - \vec{A}_1 \cdot \left( \vec{D}_\alpha \right) \\ \vec{X}_2 = \vec{X}_\beta(t) - \vec{A}_2 \cdot \left( \vec{D}_\beta \right) \\ \vec{X}_3 = \vec{X}_\delta(t) - \vec{A}_3 \cdot \left( \vec{D}_\delta \right) \end{cases} \tag{6}$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \tag{7}$$

where Eq. (5) denotes the distance between wolves $\varphi$ and $\alpha$, $\beta$ and $\delta$, the position of prey is calculated by (6)&(7).

## B. FRESH PRODUCE LOGISTICS DEMAND FORECASTING BY NLPCA-GWO-SVR

Aiming at improving the forecasting performance of fresh produce logistics demand. First, the effects of factors related to fresh produce are focused on, customers and logistics distribution enterprises on fresh produce logistics demand prediction. In this fresh produce joint distribution network, for each customer, exacting the following features, such as season, fresh produce price, average wage, and calculate the average delivery time damaged fresh produce ratio as the input of the NLPCA-GWO-SVR prediction model. There may be complex nonlinear correlations between these input features, the correlations between the features logistics demand may cause information overlap. That is, using the SVR model directly may produce a "dimensionality disaster" in the input variables. On this basis, this paper uses NLPCA to reduce the dimensionality of the input features of a traditional SVR and eliminate their correlations. Moreover, the predictive accuracy of SVR has a severe dependence on the penalty factor and kernel parameters, GWO algorithm is used to optimize the penalty factor, and bandwidth parameters of the kernel function of the SVR model. Then, the NLPCA-GWO-SVR prediction model is constructed, because of its superiority in learning and modelling nonlinear and complex relationships, SVR has been deployed in various disciplines [29], [30]. Due to space limitations, the concepts related to nonlinear principal component analysis (NLPCA) and support vector regression (SVR) will not be discussed here. More details on NLPCA, GWO and SVR can refer to [31]. Figure 2 illustrates the flowchart of NLPCA-GWO-SVR prediction model.

## C. FRESH PRODUCE JOINT DISTRIBUTION NETWORK OPTIMIZATION MODEL FORMULATION

### 1) PROBLEM DESCRIPTION

In order to improve the efficiency of fresh produce at urban last mile and reduce the waste caused by the repeated allocation of distribution resource, a distribution center resource sharing strategy is introduced to tackle this problem. Figure 3(a) introduces a typical fresh produce distribution network without sharing centers which deploys four distribution centers and eight vehicles to serve customers. As figure 3(a) shows, unreasonable distribution phenomena such as cross transportation and roundabout transportation may occur, and it is easy to violate the customer's time window, and may even reduce the freshness of fresh produce. Figure 3(b) illustrates the implementation and influence of distribution center resource sharing, where only three distribution centers and six vehicles are deployed. Each customer is served by the closest distribution centers, improving utilization of distribution centers and eliminating long and empty-vehicle trips.

In this FPJDOP, the decision-making issues that need to be considered in this problem include:

(i) Determining the number of joint distribution centers from potential distribution centers;
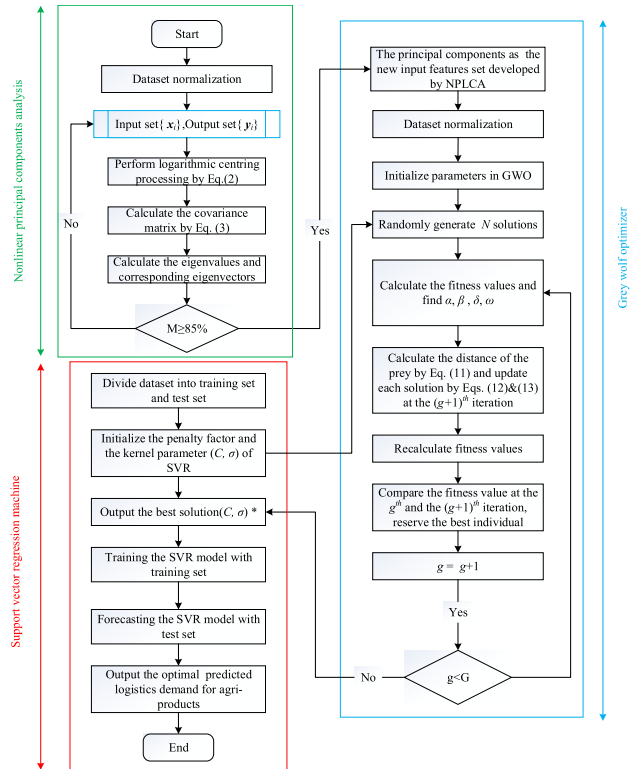


**FIGURE 2.** The flowchart of NLPCA-GWO-SVR prediction model.

(ii) Reassigning all customers of each enterprise;

(iii) Vehicle routine arrangement for each joint distribution center;

### 2) MODEL ASSUMPTIONS

To represent the real-world conditions of the fresh produce joint distribution network as close as possible, model assumptions are designed as follows:

(i) Each distribution center has the same type of vehicles, and the driving speed of the vehicles between each node is constant;

(ii) The geographical location and service time window of each customer node are known;

(iii) Each order of an enterprise at an identical customer node cannot be split;

(iv) Each vehicle can be used for (at most) one route, and it must start and end at the same depot;

(v) The capacity of vehicle is limited and customers' demand cannot exceed the vehicle capacity.

According to the above description and assumption, the parameters are defined in Table 1.

$S = H \cup G(i, j \in S)$, $H = \{i | i = 1, 2, \ldots, N\}$ is the set of customers, $G = \{r | r = 1, 2, \ldots, R\}$ is the set of potential distribution centers, $V = \{k | k = 1, 2, \ldots, \Gamma\}$ is the set of vehicles, $T = \{t | t = 1, 2, \ldots, \Phi\}$ is the set of decision cycles.

The following variables are defined in Table 2.

### 3) OBJECTIVE COST ANALYSIS
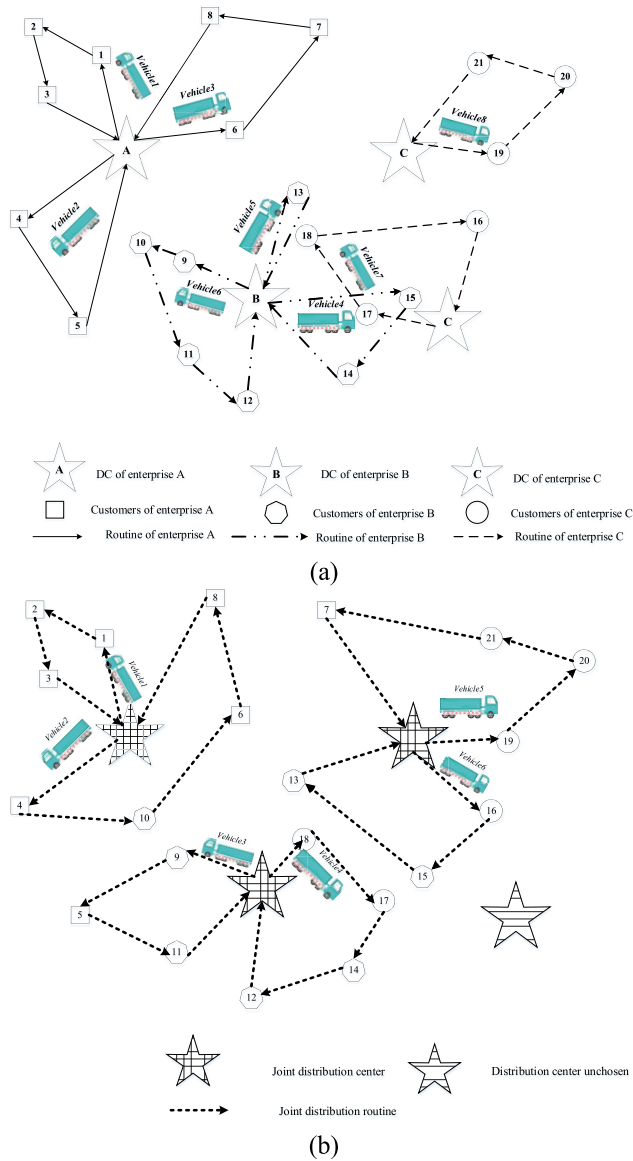
The total objective cost is consisted of five parts:

**FIGURE 3.** Comparison of before and after achieving joint distribution under distribution center sharing.

**TABLE 1.** Parameters.

| Symbol | Definition |
|---|---|
| $c_{ij}$ | average cost of traveling from node $i$ to node $j$ |
| $c_k$ | fixed cost of each used vehicle $k$ |
| $F_r$ | opening cost of distribution center $r$ |
| $d_{ij}$ | distance between node $i$ and node $j$ |
| $q_j(t)$ | demand required by customer $i$ in decision cycle $t$ |
| $Q_k$ | capacity of vehicle $k$ |
| $t_{ijk}$ | travel time of vehicle $k$ from node $i$ to node $j$ |
| $t_i$ | service time needed by customer $i$ |
| $[ET_i, LT_i]$ | service time window of customer $i$ |
| $AT_i$ | arrival time to customer $i$ |
| $c_w$ | waiting penalty cost on early arrival/unit time for customers |
| $c_l$ | lateness penalty cost on late arrival/unit time for customers |
| $f_{ave}$ | maturity change rate of fresh produce |
| $P$ | price of fresh produce |
| $|SR^k|$ | number of customers served by vehicle $k$ |

**TABLE 2.** Variables.

| Symbol | Definition |
|---|---|
| $z_r(t)$ | binary variable: 1, if potential distribution center $r$ is selected in decision cycle $t$; 0, otherwise |
| $x_{ijk}(t)$ | binary variable: 1, if vehicle $k$ travels form node $i$ to node $j$ in decision cycle $t$; 0, otherwise |

4) Time window violation cost, which is bound up to arrival time of vehicles, can be calculated by:

$$f_4(t) = \sum_{i \in H} [\max\{ET_i - T_i, 0\} + \max\{T_i - LT_i, 0\}] \quad (11)$$

5) Damage cost, which is interrelated to the maturity change of fresh produce, if the maturity is not out of a consistent limit, damage cost will not be generated. Damage cost can be obtained by:

$$f_5(t) = P \cdot \left(1 - e^{-f_{ave} \sum_{i \in G} \sum_{j \in G} \sum_{k \in K} x_{ijk} t_{ij}}\right) \quad (12)$$

4) MATHEMATICAL MODELING

For each decision cycle $t$, the mathematical model of FPJDOP is defined as follows.

$$\min f_1(t) + f_2(t) + f_3(t) + f_4(t) + f_5(t) \quad (13)$$

$$\sum_{k \in V} \sum_{i \in S} x_{ijk} = 1, \quad \forall j \in H \quad (14)$$

$$\sum_{i \in S} \sum_{j \in H} q_j x_{ijk} \leq Q_k, \forall k \in V \quad (15)$$

$$\sum_{i \in S} x_{ipk} - \sum_{j \in S} x_{pjk} = 0, \quad \forall k \in V, p \in S \quad (16)$$

$$\sum_{r \in G} \sum_{j \in H} x_{rjk} \leq 1, \quad \forall k \in V \quad (17)$$

$$\sum_{k \in V} \sum_{j \in H} x_{rjk} - z_r \geq 0, \quad \forall r \in G \quad (18)$$
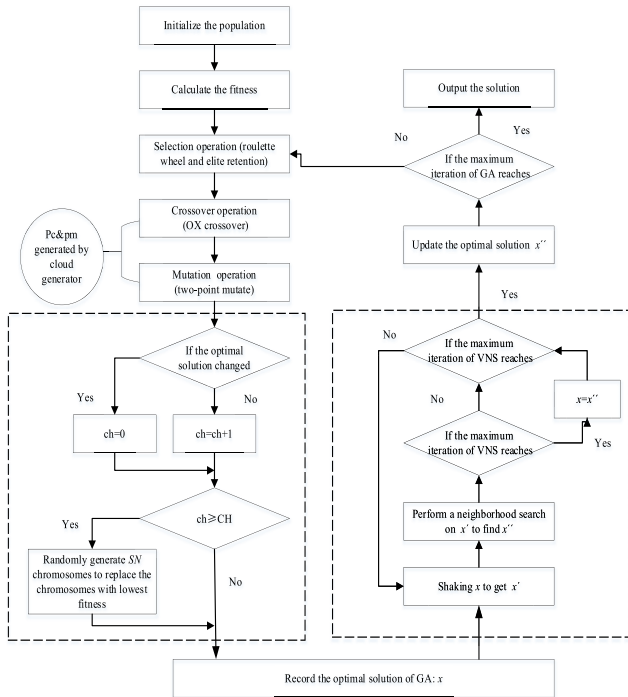
1) Fixed cost of opening common distribution centers, which is related to the number of chosen distribution centers, can be calculated by:

$$f_1(t) = \sum_{r \in G} F_r z_r \quad (8)$$

2) Fixed cost of vehicles, which is corresponding to the number of used vehicles, can be described as:

$$f_2(t) = \sum_{r \in G} \sum_{k \in V} c_k x_{rik} \quad (9)$$

3) Travel cost, which only need to consider the effect of distance, for simplicity. Travel cost can be calculated by:

$$f_3(t) = \sum_{k \in V} \sum_{i \in S} \sum_{j \in S} c_{ij} d_{ij} x_{ijk} \quad (10)$$

**FIGURE 4.** The flowchart of CAGA-VNS algorithm.

$$\sum_{r \in G} \sum_{i \in H} x_{rik} + \sum_{j \in H} \sum_{r \in G} x_{jrk} = 0, \quad \forall r \in G, k \in V \quad (19)$$

$$\sum_{i \in S} \sum_{j \in S} x_{ijk} \leq \left| SR^k \right| - 1, \quad \forall k \in V \quad (20)$$

In the above model, the objective function (13) minimizes the total cost in each decision cycle, where the first component is the fixed cost of opening common distribution centers, the second component is the fixed cost of vehicles, the third component is the travel cost, the fourth component is the time window violation cost, and the last component is the damage cost. Constraint (14) forces each customer can only be served by one vehicle. Constraint (15) ensures that the vehicle capacity cannot be exceeded. Constraint (16) means that the vehicle must leave the node after arrival. Constraint (17) states each vehicle can only depart from at most one distribution center. Constraint (18) vehicles can be dispatched to serve customers only if the distribution center is selected. Constraint (19) forces each vehicle to depart from and return to the same distribution center, if employed. Constraint (20) eliminates the sub-tour.

### D. DATA COLLECTION

According to the above FPJDOP model and corresponding solution methodology, enterprises will adjust the actual fresh produce joint distribution network according to the best solution of this problem, that is, determine which joint distribution center to use, how to allocate all customers and arrange the vehicle routing. After the implementation of the actual fresh produce joint distribution network, new features are generated such as delivery time and other features such as seasonality will be known early, which are fed

| substring1 | 2 | 3 | 1 | 3 | 3 | 2 | 1 | 2 | 1 |
| substring2 | 7 | 3 | 1 | 2 | 8 | 2 | 9 | 6 | 5 |
| substring3 | 3 | 3 | 1 | | | | | | |

**FIGURE 5.** Chromosome coding.

into the NLPCA-GWO-SVR model to reforecast the fresh produce logistics demand. Aiming at the problems of errors between the actual logistics demand and the predicted logistics demand, the parameters of the prediction model need to be adjusted combining data-driven approaches.

### E. PARAMETERS ADJUSTMENT OF THE PREDICTION MODEL

Under the rolling adjustment framework model established in this paper, for each decision cycle, when the fresh produce joint distribution network optimization (Section B) result execution is over (Section C), the indicator data related to agricultural product logistics demand within the decision cycle has been obtained, and parameters ($c$, $\sigma$) of the NLPCA-GWO-SVR prediction model need to be adjusted. A prediction model with rolling time domain framework is proposed [32], the actual logistics demand in decision cycle t will be considered as a new index put into the NLPCA-GWO-SVR model when predicting the logistics demand in (t+1) decision cycle.

## IV. SOLUTION APPROACH OF THE ROLLING ADJUSTMENT FRAMEWORK MODEL
### A. SOLUTION APPROACH OF THE MATHEMATICAL MODEL OF FPJDOP

Genetic algorithm (GA) is a highly parallel algorithm with strong global search ability, which is developed from the natural selection and evolution ideas of the biological world and has a wide range of applications in dealing with NP-hard problems [33], [34]. However, GA has the disadvantage of easily falling into local optima. To improve this disadvantage, a catastrophe adaptive genetic algorithm with variable neighborhood search (CAGA-VNS), which is more suitable for solving the model of fresh produce joint distribution network established in this paper. Figure 4 illustrates the flowchart of CAGA-VNS algorithm.

### 1) ENCODING AND THE DECODING

The natural number coding method is used to encode chromosomes in CAGA-VNS population. Each chromosome consists of three substrings, for example, there are 9 customers, 3 vehicles and 3 potential distribution centers. Figure 5 illustrates the process of chromosome coding.

1) Substring1 shows which vehicle is responsible for each customer. There are N locations in substring1, each of which is randomly generated from 1 to K (K is the maximum number of vehicles that can be selected). If the value at the nth location is k, it means that the customer at the corresponding

locus is served by the vehicle k, and the customer at the corresponding locus is determined by the value above the corresponding gene in substring2.

2) Substring1 and substring2 correspond one-to-one to determine the delivery routine of each vehicle. There are N locations in substring2, each of which is randomly generated from 1 to N (N is the number of customers).

3) Substring3 shows which distribution center each vehicle belongs to. As seen in substring1, there are K vehicles served, substring3 is related to substring1, so there are K locations in substring3, and the value of each location in substring 3 is up to the number of potential distribution centers, each of which is randomly generated from 1 to R (R is the total number of potential distribution centers). If the value at the kth locus is r, it means that the r distribution center is selected and the vehicle k belongs to distribution center r, meanwhile the vehicle k starts from the distribution center r and returns to the same. If distribution center r is not showing up in substring3, the potential distribution center r is not selected.

The value of second gene in *substring3* is 3, while the value of the second gene in corresponding *substring2* is 3, which means that customer 3 is delivered by vehicle 3. Similarly, customer 4 & 8 are also delivered by vehicle3. The distribution route of vehicle 3 is {Customer3→ Customer4→ Customer8}. The value of the third gene in *substring3* is 1, indicating that vehicle 3 belongs to distribution center 1 (DC1). Therefore, the complete distribution route of vehicle 1 is {DC1→ Customer3→ Customer4→ Customer8→ DC1}. Similarly, the chromosome can be decoding as follows:

Vehicle1: DC3→ Customer1→ Customer9→ Customer5→ DC3

Vehicle1: DC3→ Customer7→ Customer2→ Customer6→ DC3

Vehicle1: DC1→ Customer3→ Customer4→ Customer8→ DC1

As shown in *substring3*, DC2 is not selected.

### 2) INITIALIZING THE POPULATION AND FITNESS
Chromosomes are randomly generated. Each chromosome corresponds to a feasible solution and a corresponding fitness value can be calculated, as expressed by (21), which indicates that the objective of the mathematical model is to achieve the minimum total cost.

$$fit = 1/F_t \qquad (21)$$

where $F_t$ represents the total cost, *fit* is the fitness value of a chromosome. The smaller the total cost, the higher the fitness value, the better the genes of the chromosome.

### 3) GENETIC OPERATION
#### a: CROSSOVER OPERATION
Due to the chromosome designed in this paper consisting of three substrings, and each substring has a different range. Crossover operation is operated, respectively. OX crossover is adopted to *substring1~substring3*. Taking *substring2* as
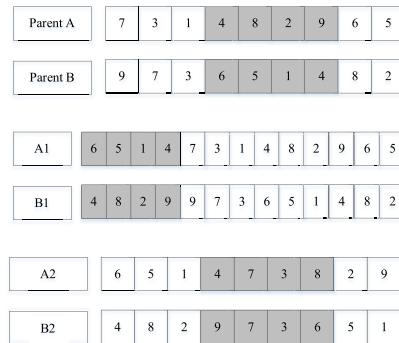


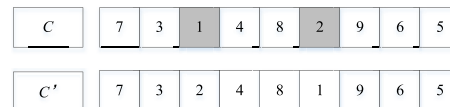**FIGURE 6.** Chromosome coding.



**FIGURE 7.** The process of mutation operation.

an example to demonstrate the crossover process, the crossing area selected by the two parent individuals is between two vertical lines, Parent A = 731|4829|65, Parent B = 973|6514|82, place the cross-section in B at the front of A and the cross-section in A at the front of B, to get A1 and B1: A1 = 6514|731482965, B1 = 4829|973651482. By sequentially deleting the same natural number as the self-crossing fragments of A1 and B1, the final legal chromosomes are obtained: A2 = 651473829, B2 = 482973651. OX crossover has advantage of maintaining the diversity of population, the process of OX crossover is shown in Figure 6.

#### b: MUTATION OPERATION
Randomly choose a chromosome from the population to perform two-points mutation. Mutation rate $p_m$ is defined by cloud generator according the whole level of the population and the chosen population, and generate a random number $rand \in [0,1]$, if $rand < p_m$, then perform mutation to the chosen chromosome, randomly choose two gene to get the new individual $C'$. Figure 7 shows the process of mutation.

#### c: DYNAMIC CATASTROPHIC MECHANISM
When the genetic algorithm iterates multiple times and its optimal solution does not change, catastrophic mechanism needs to be called. Set the initial value of the catastrophic counter $ch = 0$, and the critical value of the catastrophic mechanism $CH = 0$. If the optimal solution changes, $ch = 0$; otherwise $ch = ch+1$. When $ch \geq CH$, catastrophic mechanism works. *SN* chromosomes with low fitness in the population will be eliminated, and *SN* chromosomes will be randomly generated and added to the current population, to improve the diversity of the population. *SN* is not constant, but changes with the number of iterations, the number of
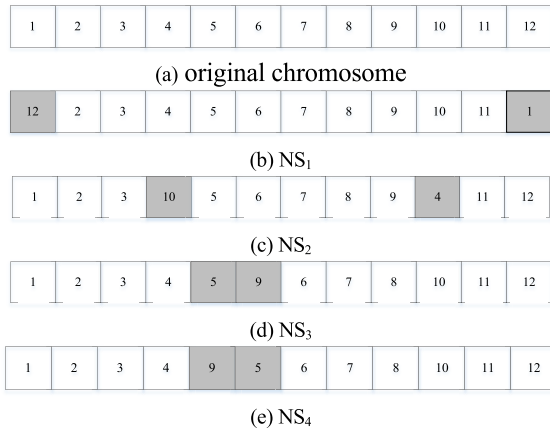
**FIGURE 8.** Neighborhood structures.

eliminated chromosomes *SN* can be calculated by:

$$SN = \left[ SN_0 * \exp\left\{ -\frac{\lambda * iter\_GA}{iter\_GA\_\max} \right\} \right] \quad (22)$$

where *SN* is the number of chromosomes that will be eliminated, $SN_0$ is the preset catastrophic scale, $\lambda$ is a controlling parameter, *iter_GA* is the number of current iteration of GA, *iter_GA_max* is the number of maximum iteration of GA.

#### d: SELECTION OPERATION

A combination of elite retention and roulette wheel is adopted to select individuals to enter the next generation, which not only preserve the diversity of the population whereas inheriting the optimal solution, but also improve the convergence speed of the population and enhance the search ability of the algorithm. The specific selection process is as follows: the individuals in the parent population are arranged in the order of fitness value from high to low, and the first 10% of the individuals are directly put into the offspring population, and the rest are selected by roulette wheel. The probability of selection can be obtained by:

$$P(x_i) = fit(x_i) \bigg/ \sum_{i \in popsize} fit(x_i) \quad (23)$$

where $fit(x_i)$ is the fitness value of chromosome $x_i$, $P(x_i)$ is the selected probability of chromosome $x_i$.

#### e: VARIABLE NEIGHBORHOOD SEARCHING

Neighborhood structures are the key to improve the solving efficiency and quality of CAGA-VNS algorithm. Figure 8(a), (b) ∼ (e) show the original routine and routines with different neighborhood structures.

Neighborhood Structure 1 (NS1): selects gene with maximum and minimum-value and then exchanges their positions.

Neighborhood Structure 2 (NS2): elects two genes randomly, then swaps their positions.

Neighborhood Structure 3 (NS3): randomly selects two genes from left to right, shifts the second gene to the first gene's neighbor and the units between two gene to the right.

Neighborhood Structure 4 (NS4): based on NS$_3$, reverse two chosen gene.

### B. SOLVING PROCESS OF THE ROLLING ADJUSTMENT FRAMEWORK MODEL

Based on the CAGA-VNS hybrid algorithm designed and the rolling time domain algorithm, sequentially solves the optimal solution of the static problem in each decision cycles, when solving the static problem in the next decision cycle, the relevant parameters in the previous decision cycle will be used as input parameters for the prediction model in the next decision cycle. The detailed steps of the rolling adjustment framework model proposed are given as follows:

*Step1:* Initialize the parameters of CAGA-VNS algorithm, initial decision cycle $t = 1$.

*Step2:* Divide randomly the original real-world instance set into three subsets: a training set, a validation set, and a test set. The training subset has 70 percent of the instances [3] and is used to optimize the SVR parameters. The second subset has 15 percents of the instances that are utilized to monitor the validation error in each training epoch. Then, the rest of the instances are used as the testing data for the NLPCA-GWO-SVR prediction model. Finally, the predicted output logistics demand results are used as input for the optimization of the agricultural product joint distribution network.

*Step3:* Setting the initial iteration number of CAGA-VNS algorithm *gen1* = 1, Generate chromosomes of CAGA algorithm randomly. Calculate the fitness value of each chromosome, and record the current optimal solution.

*Step4:* Conduct genetic operations on chromosomes, and subsequently incorporate parent and offspring chromosomes together, *gen1* = *gen1*+1.

*Step5:* If the optimal solution changes,*ch* = 0; otherwise, *ch* = *ch*+1.

*Step6:* If *ch* ≥ *CH*, use equation (22) to calculate the number of eliminated chromosomes with the lowest fitness; otherwise, record the optimal solution *x*, and *gen2* = 0.

*Step7:* *gen2* = *gen2*+1, shaking to get *x'*, perform the neighborhood search on *x'* to find *x''*.

*Step8:* If *x''* is superior to *x'*, *x* = *x''*; otherwise, turn to **Step9**.

*Step9:* If gen2 ≥ *iter_VNS_max*, update the optimal solution and turn to **Step10**; otherwise, turn to **Step7**.

*Step10:* Implement the above optimal solution in current decision cycle and the actual data on fresh produce logistics demand will be collected. If $t \leq T$, turn to **Step2**, adjust the parameters of NLPCA-GWO-SVR prediction model according to the newly collecting data; otherwise, stop the rolling adjustment framework model.

## V. NUMERICAL EXPERIMENTS

To verify the effectiveness of the CAGA-VNS algorithm, experiment 1 is designed based on a standard testing benchmark (source: http://www.vrp-rep.org/), and then experiment 2 is carried out to validate the correctness of the rolling

**TABLE 3.** $p_c$ & $p_m$.

| $p_c = X\_Cloud\,(Ex, En, He)$ | $p_m = X\_Cloud\,(Ex, En, He)$ |
|---|---|
| $Ex = \bar{f}$ | $Ex = \bar{f}$ |
| $En = (f_{max} - Ex)/C_1$ | $En = (f_{max} - Ex)/C_3$ |
| $He = En/C_2$ | $He = En/C_4$ |
| $En' = randn(En, He)$ | $En' = randn(En, He)$ |
| $p_c = \begin{cases} k_1 e^{\frac{(f'-Ex)^2}{2(En')^2}}, f' \geq \bar{f} \\ k_3, f' \leq \bar{f} \end{cases}$ | $p_m = \begin{cases} k_2 e^{\frac{(f-Ex)^2}{2(En')^2}}, f \geq \bar{f} \\ k_4, f \leq \bar{f} \end{cases}$ |

**TABLE 4.** Comparison of the three mentioned algorithms.

| Number of runs | algorithm | | |
|---|---|---|---|
| | GA | CAGA | **CAGA-VNS** |
| 1 | 5892.05 | 6275.83 | 6073.72 |
| 2 | 6596.49 | 6433.43 | 6282.46 |
| 3 | 6586.12 | 6083.42 | 5977.26 |
| 4 | 6681.26 | 6091.60 | 6031.19 |
| 5 | 6311.55 | 6319.98 | 5997.43 |
| 6 | 6894.36 | 5622.75 | **5543.08** |
| 7 | 6697.32 | 6495.68 | 6393.52 |
| 8 | 6626.33 | 5777.21 | 6154.65 |
| 9 | 6108.79 | 6168.24 | 6093.20 |
| 10 | 6931.21 | 5989.82 | 5963.39 |
| Percentage of divergence (%) | 10.87 | 10.51 | 9.16 |

adjustment framework model proposed in this paper, which is originated form an actual enterprise.

## A. EXPERIMENT 1: THE EFFECTIVENESS OF CAGA-VNS

### 1) EXPERIMENTAL RESULTS

After multiple tests, the specific parameter settings of the CAGA-VNS algorithm are as follows: $popsize = 300$, $iter\_GA\_max = 500$, $iter\_VNS\_max = 10$, $SN_0 = 30$, $\lambda = 0.75$. The probability of crossover and mutation is settled by the cloud generator adaptively, as Table 3 shows.

where $(Ex, En, He)$ is the controlling parameters of $p_c$ & $p_m$, $\bar{f}$ is the average value of all the chromosomes in the current generation, $f_{max}$ is the optimal fitness value of the current generation, $f'$ is the larger fitness of the two individuals selected for crossover operation, $f$ is the individual selected for mutation operation, $k_1 = k_3 = 0.9$, $k_2 = k_4 = 0.4$, $C_1 \sim C_4$ can be calculated by:

$$C_1 = C_3 = 6 \cdot popsize \cdot (iter + 1) \quad (24)$$

$$C_2 = C_4 = 15 - (iter - popsize/2)^2 \quad (25)$$

To validate the performance of the algorithm, the CAGA-VNS proposed in this paper is compared with the classical GA, catastrophic adaptive genetic algorithm (CAGA). Run the above three algorithms 10 times each, and the running results are shown in Table 4. The last row in Table 4 represents the percentage of divergence between each run and the optimal value, the percentage of divergence can be calculated by (26):

$$\frac{1}{10}\left(\sum_{rt}^{10} \frac{cr_{rt} - \min(cr_{rt})}{\min(cr_{rt})}\right) \times 100\% \quad (26)$$

where $cr_{rt}$ is the running result of the $rt$ time.

### 2) COMPARATIVE ANALYSIS

As shown in Table 4, Comparing GA and CAGA, dynamic catastrophic mechanism, the adaptive changes in crossover probability, and mutation probability can enable GA to obtain better solutions within the same runtime. CAGA-VNS has significant advantages, and the optimal solution obtained by the algorithm is only 5543.08. Moreover, the smaller the difference percentage is, the more stable the

algorithm is, CAGA-VNS is the most stable algorithm with the lowest difference percentage being 9.16% among the three algorithms, followed by the CAGA algorithm with a difference percentage of 10.51%, and the GA algorithm with the highest difference percentage being 10.87%. Therefore, the algorithm designed in this paper has more advantages.

For solving the problem of optimizing the joint distribution network of fresh produce, CAGA-VNS is more competitive.

## B. EXPERIMENT 2: THE CORRECTNESS OF THE ROLLING ADJUSTMENT FRAMEWORK MODEL

### 1) REAL-WORLD CASE

For the real-world case, we adopt a representative fresh produce sales company (*Enterprise A*) in Shenyang China, which is highly trusted enterprise with a wide range of business scope. However, *Enterprise A* adopts independent delivery to serve customers, resulting in high costs and low distribution efficiency and customer satisfaction. Therefore, *Enterprise B* is jointed together to perform joint distribution. Table 5 summarizes data with regard to customer nodes in the fresh produce distribution network in part. Table 6 shows the information of potential distribution center. D1~D4 are potential distribution centers belonging to *Enterprise A*, D5~D10 are potential distribution centers belonging to *Enterprise B*.

### 2) EXPERIMENTAL RESULTS

#### a: PREDICTION RESULT ANALYSIS

We select the fresh produce demand data of two mentioned enterprises from 2012 to 2021 (2012Q1~2021Q4) as the training test set, with each cycle length of one quarter and a total of 40 decision cycles. Table 7 shows partly the real fresh produce demand data for all customers served by the two companies during the period of Q1 2020.

Compared with the original demand, the results of forecasting counterpart forecasted by NLPCA-GWO-SVR are drawn in Figure 9.

Taken NO.1 customer as an example, the decision-making cycle range from 2012Q1~2021Q4, the NLPCA-GWO-SVR model is used to predict. As seen in Figure 9, the blue line with solid square represents the collected actual logistics

**TABLE 5.** Coordinates and time window of customers.

| No. | Coor | TW | No. | Coor | TW |
|-----|------|----|-----|------|----|
| C1 | (41,49) | [161,171] | C2 | (35,21) | [50,60] |
| C3 | (55,45) | [116,126] | C4 | (55,20) | [149,159] |
| C5 | (15,30) | [34,44] | C6 | (25,30) | [99,109] |
| C7 | (20,50) | [81,91] | C8 | (10,43) | [95,105] |
| C9 | (55,78) | [97,107] | C10 | (30,60) | [124,134] |
| C11 | (20,65) | [67,77] | C12 | (50,35) | [63,73] |
| C13 | (30,25) | [159,169] | C14 | (15,10) | [32,42] |
| C15 | (30,5) | [61,71] | C16 | (10,20) | [75,85] |
| C17 | (5,30) | [157,167] | C18 | (20,40) | [87,97] |
| C19 | (15,60) | [76,86] | C20 | (45,65) | [126,136] |
| C21 | (45,20) | [62,72] | C22 | (45,10) | [97,107] |
| C23 | (55,5) | [68,78] | C24 | (65,35) | [153,163] |
| C25 | (65,20) | [172,182] | C26 | (45,30) | 132,142] |
| C27 | (35,40) | [37,47] | C28 | (41,37) | [39,49] |
| C29 | (64,42) | [63,73] | C30 | (40,60) | [71,81] |
| C31 | (31,67) | [50,60] | C32 | (35,69) | [141,151] |
| C33 | (53,52) | [37,47] | C34 | (65,55) | [117,127] |
| C35 | (63,65) | [143,153] | C36 | (2,60) | 41,51] |
| C37 | (20,20) | [134,144] | C38 | (5,5) | [83,93] |
| C39 | (60,12) | [44,54] | C40 | (40,25) | [85,95] |

[1]Note: Coor is the coordinates of customers, TW means the time window of each customer.

**TABLE 6.** Coordinates and opening cost of potential distribution centers.

| Distribution Center | Coor | $F_r$ |
|---------------------|------|-------|
| 1 | (35,17) | 1032 |
| 2 | (55,60) | 1045 |
| 3 | (18,10) | 1030 |
| 4 | (55,65) | 1040 |
| 5 | (31,52) | 1024 |
| 6 | (70,73) | 1045 |
| 7 | (15,20) | 1020 |
| 8 | (4,50) | 1030 |
| 9 | (40,31) | 1010 |
| 10 | (23,5) | 1025 |

demand of fresh produce. The solid line with a multiplication sign in orange represents the prediction result. By comparing the prediction logistics demand of fresh produce with the actual changing logistics demand of fresh produce. As seen, the NLPCA-GWO-SVR model established can basically meet the actual logistics demand of fresh produce better after adjusting the parameters of proposed model in each decision cycle. At the same time, a certain buffer space exists, which can cope with the logistics demand of fresh produce in emergencies.

To investigate the superiority of the NLPCA-GWO-SVR forecasting model, two widely used evaluation measures for performance comparison are employed, namely, root mean squared error (RMSE) and mean average percentage error

**TABLE 7.** Logistics demand data of fresh produce.

| No. | D | No. | D | No. | D |
|-----|------|-----|------|-----|------|
| C1 | 23.8200 | C2 | 26.9775 | C3 | 26.6476 |
| C4 | 31.2692 | C5 | 28.3648 | C6 | 37.8705 |
| C7 | 38.3747 | C8 | 30.0566 | C9 | 24.4798 |
| C10 | 45.1366 | C11 | 32.0530 | C12 | 36.0610 |
| C13 | 51.5548 | C14 | 39.8069 | C15 | 38.0423 |
| C16 | 57.6293 | C17 | 39.2423 | C18 | 34.2057 |
| C19 | 63.3601 | C20 | 38.0906 | C21 | 43.8255 |
| C22 | 68.7472 | C23 | 36.6434 | C24 | 22.8170 |
| C25 | 79.7863 | C26 | 34.3538 | C27 | 40.8425 |
| C28 | 82.0878 | C29 | 36.9591 | C30 | 39.6308 |
| C31 | 81.6475 | C32 | 39.8689 | C33 | 37.4638 |
| C34 | 78.4652 | C35 | 39.5394 | C36 | 29.9853 |
| C37 | 65.1875 | C38 | 39.9060 | C39 | 32.3942 |
| C40 | 59.4628 | | | | |
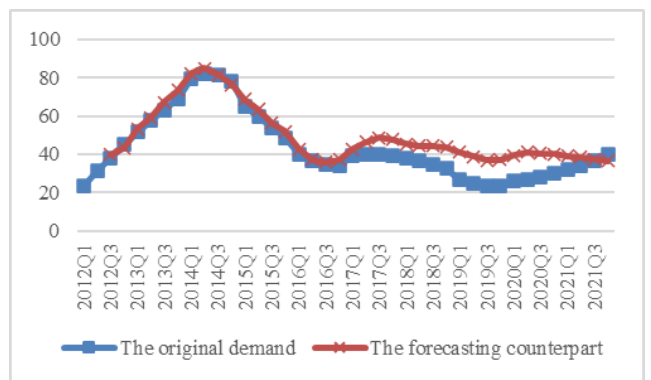


**FIGURE 9.** Comparison between the original demand and the forecasting counterpart.

(MAPE), which are formally defined as follows:

$$RMSE = \sqrt{\frac{1}{J} \sum_{j=1}^{J} \left( y_j(t) - \hat{y}_j(t) \right)^2} \quad (27)$$

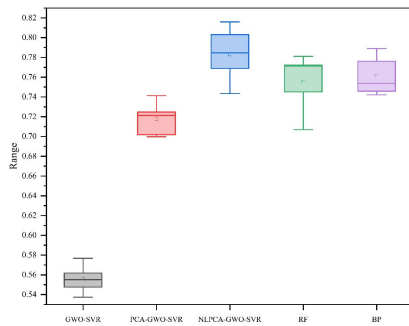$$MAPE = \frac{1}{J} \sum_{j=1}^{J} \left| \frac{y_j(t) - \hat{y}_j(t)}{y_j(t)} \right| \quad (28)$$

$$R^2 = 1 - \frac{\sum_{j=1}^{J} \left( y_j(t) - \hat{y}_j(t) \right)^2}{\sum_{j=1}^{J} \left( y_j(t) - \bar{y}_j(t) \right)^2} \quad (29)$$

where $y_j(t)$ is the ground truth value for the logistics demand of customer $j$ in decision cycle $t$, whereas $\hat{y}_j(t)$ is the predicted counterpart, $\bar{y}_j(t)$ is the mean value of the original demand.

Table 8 demonstrates the performance comparison in terms of RMSE, MAPE and R$^2$ evaluation metrics defined above, comparing with the baseline methods using the same training-test set.

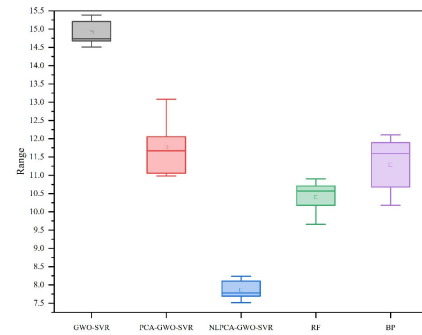**TABLE 8.** Performance comparison of different forecasting models.

| | RMSE | MAPE | $R^2$ |
|---|---|---|---|
| GWO-SVR | 15.3842 | 0.3508 | 0.5374 |
| NLPCA-GWO-SVR | **7.5161** | **0.1849** | **0.8032** |
| PCA-GWO-SVR | 12.0576 | 0.2367 | 0.7213 |
| RF(random forest) | 9.6572 | 0.2012 | 0.7724 |
| BP(back propagation) | 10.1765 | 0.2169 | 0.7538 |



**FIGURE 11.** RMSE of the five prediction models.
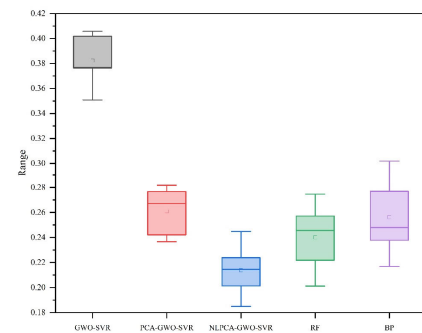


**FIGURE 10.** $R^2$ of the five prediction models.



**FIGURE 12.** MAPE of the five prediction models.

As seen from Table 8, the proposed NLPCA-GWO-SVR model achieves an $R^2$ (R-squared, the coefficient of the determination) of 0.8032, indicating that the forecasting counterpart is close to the ground truth values. Table 8 also summarizes the performance comparison in terms of RMSE and MAPE evaluation metrics. As seen, the proposed NLPCA-GWO-SVR prediction model achieves the lowest RMSE of 7.5161 and MAPE of 0.1849, which outperforms baseline models, such as BP and RF, and shows its ability to implicitly discover complex nonlinear relationships between demand and the influential factors by detecting all possible interactions between predictor variables. The well trained NLPCA-GWO-SVR model with a high prediction accuracy can be implemented for inputting demand parameter, which will be further used in FPJDOP.

Furthermore, Figure 10, 11, 12 present the R-squared, RMSE, MAPE after making a 5-time prediction. There are no exceptional points in these figures. It is clear that the NLPCA-GWO-SVR prediction model has the highest R-squared, compared with other four prediction models in figure 10. However, observing the height of the box, it can be seen that its volatility is relatively high. Figure 11 draws the results of five prediction models for RMSE, NLPCA-GWO-SVR prediction model also achieves the lowest RMSE, and the box of this model is relatively short, its RMSE distribution is very stable, as well as Figure 12. In a nutshell, the prediction model established in this paper achieves a good prediction accuracy, the model fits very well, and basically captures the change trend of the original sample.

Meanwhile, to verify the robustness of the proposed NLPCA-GWO-SVR prediction model with high changing fresh produce logistics demand, a new experiment with high changing signals is carried out. 300 samples of fresh produce demand are generated randomly. Figure 13 shows the result of ground truth value of fresh produce logistics demand and the forecasting counter counterpart. The NLPCA-GWO-SVR prediction model achieve a RMSE of 2.1045 in the testing period, a MAPE of 0.1175 in the testing period, which proves the proposed prediction model still has good fitting ability.

*b: JOINT DISTRIBUTION CENTER SETTINGS FOR DIFFERENT DECISION CYCLES*

Figure 14 shows the number of distribution centers selected in 10 decision-making cycles from *2018Q1* to *2020Q2*. For example, in the cycle *2018Q1*, five distribution centers (D2, D4, D6, D7, D9) are selected according to the optimization results, which covered the logistics demand of fresh produce of 35%, 22%, 18%, 15% and 10%, respectively. In the cycle *2018Q4*, only four distribution centers (D5, D7, D9, D10) need to be selected to cover the demand for agricultural products of 47%, 34%, 11% and 8%, respectively. That is, through
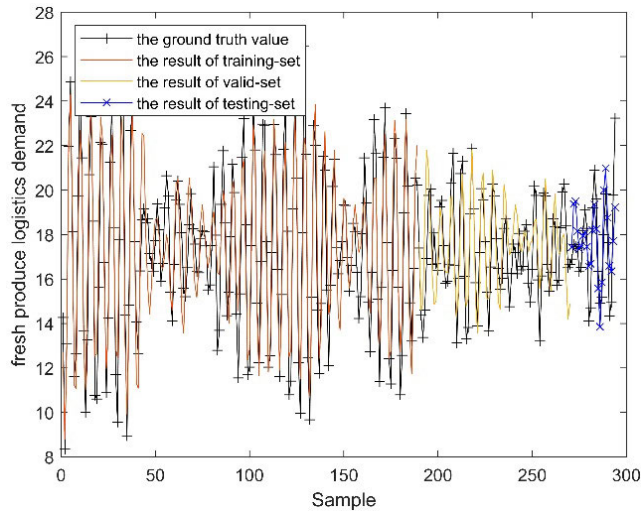
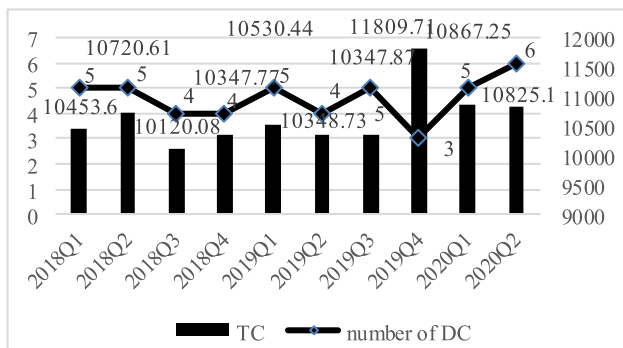**FIGURE 13.** The forecasting result with high changing signals.



**FIGURE 14.** The number of joint distribution center setting in different decision cycles.

the distribution center sharing strategy proposed in this paper, it can be seen that the number of distribution centers has been reduced and the utilization rate of distribution centers has been improved. In the joint distribution network of fresh produce at the end of the city, the setting up of joint distribution centers not only can better cover the logistics demands of customers, but also can make full use of joint distribution to form scale benefits and improve the utilization rate of distribution center. Therefore, with the seasonal changes in the logistics demand of fresh produce, it is necessary for the distribution network to make corresponding adjustments to help save enterprise costs while quickly delivering the fresh produce to customers.

*c: DISTRIBUTION ROUTING ARRANGEMENTS FOR DIFFERENT DECISION CYCLES*

Table 9 displays the optimal distribution routes and costs for the cycles *2018Q4* and *2019Q4*. In two different optimization cycles, different optimal number of distribution centers are selected, where number 0 is used to separate different vehicles within the same distribution center. In the cycle *2018Q4*, four distribution centers were selected, including D1, D5, D7

**TABLE 9.** The optimal distribution routines and cost in different cycles.

| Cycle | Distribution routines | Total cost |
|---|---|---|
| 2018Q4 | **D1:** 33-16-8-46-37-0-17-23-34-0-5-29-2-15-0-64-50-25-77-0-39-38-48-55-43-0-32-57-72-0-11-98-3-26-0-30-44-87-22-68<br>**D5:** 9-76-99-10-20-60-0-62-47-66-88-27<br>**D7:** 33-16-8-46-37-0-83-42-19-61-0-86-81-35-91-89-100-093-14-67-84-0-59-45-0-82-95-69-85-0-36-92<br>**D9:** 65-73-4-56-24-0-28-90-49-53-0-70-52-78-54-96-0-79-31-21-94-6-41-58-0-63-97-18-74-13-0-1-71-40-80-0-7-12 | 10347.77 |
| 2019Q4 | **D2:** 4-12-20-79-0-71-76-34-65-0-84-9-68-81-35-24<br>**D7:** 62-18-41-37-0-63-44-94-0-64-53-3-89-58-0-40-42-22-32-96-0-15-86-57-67-48<br>**D9:** 39-75-11-0-38-61-6-0-92-14-83-0-82-31-78-87-0-45-29-0-19-5-69-54-0-74-21-30-0-36-100-1-55-0-80-16-51-8-91-70-0-73-27-7-25-17-26-93-0-60-59-95-46-97-0-49-10-66-13-0-47-98-85-43-0-2-28-23-0-33-99-88-56-77-0-52-72-90-50 | 11809.71 |

**TABLE 10.** Comparison on computational results of two strategies.

| Strategy | Strategy A | Strategy B |
|---|---|---|
| TD | 1115.30 | 1279.48 |
| TC | 10120.08 | 12089.96 |
| VN | 20 | 24 |

and D9, respectively. In the cycle *2019Q4*, three distribution centers were selected, including D2, D7 and D9, respectively. The selection of distribution centers in the two phases is very different, except D7 and D9. Therefore, reselecting the distribution centers and optimizing the routines in different cycles is necessary.

*d: COMPARISON BETWEEN DISTRIBUTION CENTER SHARING STRATEGY AND NON-SHARED STRATEGY*

With the above algorithm parameters unchanged, using data from the cycle *2018Q3*, this paper compares the distribution network optimization under distribution center sharing strategy (named *Strategy A*) with the distribution network optimization under the independent distribution mode (named *Strategy B*). Run the two programs 10 times, and take the average value as the final result, as shown in Table 10. TD represents the total delivery distance (unit: kilometers), TC is the total cost (unit: Yuan), VN is the number of delivery vehicle.

As can be seen from Table 10, from the perspective of total travel distance TD, the joint distribution strategy designed in this paper (Strategy A) is superior to traditional independent distribution (Strategy B), with the former saving travel distance of 14.72%, compared to the latter; In terms of TC, the former saves up to the cost of 19.46%, compared to the latter, while the former's solution reduces the use of distribution
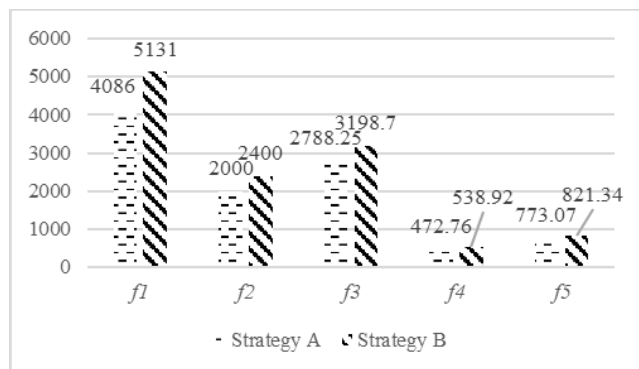
**FIGURE 15.** Comparison of costs under sharing and unsharing strategies.

centers and improves the utilization rate of distribution centers compared to the latter; Regarding the number of delivery vehicles, the former uses fewer vehicles of 20% than the latter. The distribution scheme of the joint distribution strategy designed in this paper is far superior to the traditional independent distribution strategy, which can simultaneously reduce the use of distribution centers and vehicles, shorten the distribution distance, and reduce the cost.

As shown in Figure 15, comparing the optimization results of a joint distribution network based on distribution center sharing with the traditional independent distribution network optimization results in the various costs, each itemized cost is lower than the cost generated under the independent distribution mode. In terms of the opening cost of the distribution centers, the former saves 20% compared to the latter, reducing the use of the distribution center, and improving the utilization rate of the distribution centers. With respect to the time window penalty cost, it can be seen that the distribution center sharing strategy reduces the violation of customer time windows to a certain extent; Regarding the cost of fresh produce damage, the distribution center sharing strategy can help enterprises to quickly deliver agricultural products to end customers while reducing the loss of fresh produce.

## VI. CONCLUSION

The research on fresh produce joint distribution network optimization has become an essential field in agricultural logistics management. The seasonality changes and maturity changes in the demand for fresh produce themselves make the distribution of urban fresh produce significantly different from industrial products. With that in mind, an innovative rolling adjustment framework model based on data-driven is presented for optimizing the distribution network of fresh produce. Each decision cycle under this framework model encompasses four progressive links, i.e., the forecast of fresh produce logistics demand, joint distribution network optimization of fresh produce, data collection and processing, and parameters adjustment of the prediction model. The main theoretical contributions of this paper can be summarized.

First, from the data-driven perspective, a rolling adjustment framework model for the optimization of agricultural joint distribution network is constructed, which is not one-off distribution network optimization in the traditional sense, and has strong applicability; In addition, the implementation of joint distribution network based on the distribution center sharing strategy enables the distribution resources of fresh produce at the end of the city to be rationally allocated according to the logistics demands of fresh produce, and improves the utilization rate of distribution centers. It provides a new way of thinking for the optimization of the urban last-mile logistics distribution network of fresh produce in China, and also provides some instructive decision reference for related enterprises.

Nevertheless, in this paper some directions are suggested for further research. One exciting research path would investigate the optimization of fresh produce distribution network under low-carbon background, due to global warming which is mainly caused by carbon dioxide emission and poses growing threat to environment and human beings. Another promising future research path would introduce disruption management to fresh produce distribution network to cope with the disruption.

## REFERENCES

[1] Q. Chen, G. Wu, Y. D. Gu, B. J. Lu, and Q. Wei, "The challenges for big data driven research and applications in the context of managerial decision-making: Paradigm shift and research directions," *J. Manage. Sci. Chin.*, vol. 21, no. 7, pp. 1–10, Jul. 2018.

[2] Q. Chen, D. J. Zeng, Q. Wei, M. Y. Zhang, and X. H. Guo, "Transitions of decision-making paradigms and enabled innovations in the context of big data," *Manage. World*, vol. 36, no. 2, pp. 95–220, Feb. 2020.

[3] J. Liu, W. Chen, J. Yang, H. Xiong, and C. Chen, "Iterative prediction-and-optimization for e-logistics distribution network design," *INFORMS J. Comput.*, vol. 34, no. 2, pp. 769–789, Mar. 2022.

[4] M. T. Li, Q. L. Ding, and K. J. Liu, "An optimization approach of collaborative distribution considering common customers," *Oper. Res. Manag. Sci.*, vol. 32, no. 3, pp. 104–110, Mar. 2023.

[5] F. Goodarzian, D. Shishebori, F. Bahrami, A. Abraham, and A. Appolloni, "Hybrid meta-heuristic algorithms for optimising a sustainable agricultural supply chain network considering CO₂ emissions and water consumption," *Int. J. Syst. Sci., Oper. Logistics*, vol. 10, no. 1, Dec. 2023, Art. no. 2009932.

[6] A. Cheraghalipour and E. Roghanian, "A bi-level model for a closed-loop agricultural supply chain considering biogas and compost," *Environ., Develop. Sustainability*, pp. 1–47, May 2022.

[7] T. Perdana, B. Tjahjono, K. Kusnandar, S. Sanjaya, D. Wardhana, and F. R. Hermiatin, "Fresh agricultural product logistics network governance: Insights from small-holder farms in a developing country," *Int. J. Logistics Res. Appl.*, pp. 1–24, Aug. 2022.

[8] Y. Zhao, S. Cheng, and F. Lu, "Seasonal characteristics of agricultural product circulation network: A case study in Beijing, China," *Agronomy*, vol. 12, no. 11, pp. 1–16, Nov. 2022.

[9] J. Chen, B. Dan, and J. Shi, "A variable neighborhood search approach for the multi-compartment vehicle routing problem with time windows considering carbon emission," *J. Cleaner Prod.*, vol. 277, Dec. 2020, Art. no. 123932.

[10] J. Han, N. Lin, J. Ruan, X. Wang, W. Wei, and H. Lu, "A model for joint planning of production and distribution of fresh produce in agricultural Internet of Things," *IEEE Internet Things J.*, vol. 8, no. 12, pp. 9683–9696, Jun. 2021.

[11] Y. F. Ma, B. Ying, X. Y. Zhou, and P. Wang, "Multi-agent optimization model and algorithm for perishable food location-routing problem with conflict and coordination," *Syst. Eng. Theor. Pract.*, vol. 40, no. 12, pp. 3194–3209, Dec. 2020.

[12] Q. Zhang, Y. Xiong, M. K. He, and H. Zhang, "Multi-objective model of distribution route problem for fresh electricity commerce under uncertain demand," *J. Syst. Simul.*, vol. 31, no. 8, pp. 1582–1590, Aug. 2019.

[13] J. Wei and J. X. Wang, "FCM-ISA algorithm for continuous location model of fresh produce multi-distribution centers and application," *Oper. Res. Manage. Sci.*, vol. 28, no. 11, pp. 85–90, Nov. 2019.

[14] M. Bortolini, M. Faccio, E. Ferrari, M. Gamberi, and F. Pilati, "Fresh food sustainable distribution: Cost, delivery time and carbon footprint three-objective optimization," *J. Food Eng.*, vol. 174, pp. 56–67, Apr. 2016.

[15] S. M. Jaigirdar, S. Das, A. R. Chowdhury, S. Ahmed, and R. K. Chakrabortty, "Multi-objective multi-echelon distribution planning for perishable goods supply chain: A case study," *Int. J. Syst. Sci., Oper. Logistics*, vol. 10, no. 1, Dec. 2023, Art. no. 2020367.

[16] M. Golestani, S. H. Moosavirad, Y. Asadi, and S. Biglari, "A multi-objective green hub location problem with multi item-multi temperature joint distribution for perishable products in cold supply chain," *Sustain. Prod. Consump.*, vol. 27, pp. 1183–1194, Jul. 2021.

[17] Y. Sheng, C. S. Liu, and R. Y. Lu, "Vehicle routing problem with simultaneous pick-up and delivery in rural electronic commerce based on joint distribution strategy," *Syst. Eng.*, vol. 37, no. 3, pp. 98–104, May 2019.

[18] Y. Wang, K. Assogba, J. Fan, M. Xu, Y. Liu, and H. Wang, "Multi-depot green vehicle routing problem with shared transportation resource: Integration of time-dependent speed and piecewise penalty cost," *J. Cleaner Prod.*, vol. 232, pp. 12–29, Sep. 2019.

[19] J. Li, R. Wang, T. T. Li, Z. X. Lu, and P. M. Pardalos, "Benefit analysis of shared depot resources for multi-depot vehicle routing problem with fuel consumption," *Transport. Res. D, Transport. Environ.*, vol. 59, pp. 417–432, Mar. 2018.

[20] Y. Wang, J. Zhang, K. Assogba, Y. Liu, M. Xu, and Y. Wang, "Collaboration and transportation resource sharing in multiple centers vehicle routing optimization with delivery and pickup," *Knowl.-Based Syst.*, vol. 160, pp. 296–310, Nov. 2018.

[21] Z. P. Wang, "Study of sharing distribution model based on dynamic genetic algorithm," *Oper. Res. Manag. Sci.*, vol. 30, no. 12, pp. 115–122, Dec. 2021.

[22] M. Liu, J. Cao, and D. Zhang, "Dynamic adjustment method for optimizing epidemic-logistics network based on data-driven," *Syst. Eng. Theory Pract.*, vol. 40, no. 2, pp. 437–448, Feb. 2020.

[23] M. Du, A. Sai, and N. Kong, "A data-driven optimization approach for multi-period resource allocation in cholera outbreak control," *Eur. J. Oper. Res.*, vol. 291, no. 3, pp. 1106–1116, Jun. 2021.

[24] M. Du, "Data-driven joint decision method for epidemic detection and dynamic allocation of medical resources," *Chin. J. Manage. Sci.*, vol. 13, p. 11, Mar. 2023, doi: 10.16381/j.cnki.issn1003-207x.2021.2214.

[25] Y. Xiang, "Selection of optimal start-up time point for emergency procurement at the beginning of an outbreak," *Chin. J. Manage. Sci.*, vol. 20, p. 13, Apr. 2023, doi: 10.16381/j.cnki.issn1003-207x.2021.2631.

[26] J. B. B. Clausen and H. Li, "Big data driven order-up-to level model: Application of machine learning," *Comput. Oper. Res.*, vol. 139, Mar. 2022, Art. no. 105641.

[27] H. Xiong and H. L. Yan, "Research on the data-driven mechanism of intelligent order dispatching of takeaway platform," *Nankai. Bus. Rev. Int.*, vol. 25, no. 2, pp. 15–29, Feb. 2022.

[28] R. H. Chu, W. S. Zhang, P. F. Bai, and Y. H. Chen, "Data-driven optimization for last-mile delivery," *Complex Intell. Syst.*, vol. 2021, pp. 1–14, Feb. 2021.

[29] X. Kang, "Combining rough set theory and support vector regression to the sustainable form design of hybrid electric vehicle," *J. Cleaner Prod.*, vol. 304, Jul. 2021, Art. no. 127137.

[30] M. Liu, K. Luo, J. Zhang, and S. Chen, "A stock selection algorithm hybridizing grey wolf optimizer and support vector regression," *Exp. Syst. Appl.*, vol. 179, Oct. 2021, Art. no. 115078.

[31] X. Y. Zhou, M. L. Zhu, X. Y. Ma, and Y. Zhong, "Forecasting agricultural product logistics demand by nonlinear principal component analysis and a support vector machine optimised by the grey wolf optimizer," *Int. J. Internet Manuf. Serv.*, vol. 8, no. 2, pp. 150–171, Apr. 2021.

[32] Z. F. Qiu, M. Zhang, H. L. Wu, and B. Q. Li, "Dynamic forecasting of emergency materials based on rolling horizon framework," *Aeronaut. Comput. Tech.*, vol. 49, no. 4, pp. 62–66, Jul. 2019.

[33] X. L. Ge, X. Wang, and L. B. Xing, "Multi-vehicle scheduling problems and cloud GA based on the dynamic needs," *J. Syst. Eng.*, vol. 27, no. 6, pp. 823–832, Dec. 2012.

[34] H. Z. Zhang, Y. Liu, and J. Ni, "A partheno-genetic hybrid ant colony algorithm for solving the MOVRPFTW based on customer satisfaction," *J. Syst. Manage.*, vol. 28, no. 5, pp. 927–933, Sep. 2019.

**MEILIN ZHU** received the B.Eng. degree in logistics management from the School of Management, Shenyang University of Technology, Shenyang, China, where she is currently pursuing the Ph.D. degree in management science and engineering. Her current research interests include fresh produce logistics and data-driven optimization.

**XIAOYE ZHOU** received the Ph.D. degree from the Shenyang Institute of Automation, Chinese Academy of Sciences. She is currently a Professor and a Doctoral Supervisor with the School of Management, Shenyang University of Technology, Shenyang, China. She has run and taken charge of many funded research projects as a principal investigator. She has published three monographs and more than 20 articles in international journals, such as IEEE Access and *International Journal of Production Economics*. Her current research interests include fresh produce logistics, smart logistics, and data-driven optimization.

● ● ●