

## Research Article

# Prediction of Vegetable Supply in Henan Province Based on PSO-GM (1, N) Model

Xueqiang Guo <sup>1</sup> and Bingjun Li <sup>2</sup>

<sup>1</sup>College of Mechanical and Electrical Engineering, Henan Agricultural University, Zhengzhou 450002, China

<sup>2</sup>College of Information and Management Science, Henan Agricultural University, Zhengzhou 450046, China

Correspondence should be addressed to Bingjun Li; [zylbjun@163.com](mailto:zylbjun@163.com)

Received 14 July 2021; Revised 13 August 2021; Accepted 16 August 2021; Published 25 August 2021

Academic Editor: Shaojian Qu

Copyright © 2021 Xueqiang Guo and Bingjun Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

GM (1, N) model is one of the grey prediction models considering the influence of many factors. This paper improves GM (1, N) model and constructs PSO-GM (1, N) model. Firstly, Lasso method is used to select the influencing factors, then the priority of influencing factors and the value of parameter  $N$  in GM (1, N) model are determined, and finally PSO method is used to optimize GM (1, N) model. Taking the vegetable supply in Henan Province as the research object, this paper makes an empirical test by using PSO-GM (1, N) model. The results show that the key factors affecting the vegetable supply in Henan Province are the number of rural employees, highway mileage, and application of pesticide. The vegetable supply in Henan Province will continue to show a steady growth trend in the next three years.

## 1. Introduction

Vegetables are essential agricultural products in the daily life of urban and rural residents. Effectively ensuring the supply of vegetables has become a great event related to people's livelihood. Vegetable supply is affected by many factors such as natural climate, production capacity, and transportation. Effectively identifying the key factors affecting vegetable supply and accurately predicting the future vegetable supply are of great practical significance for stabilizing vegetable market supply and stabilizing vegetable price fluctuation.

Scholars have studied the influencing factors analysis and prediction model construction of vegetable supply from different angles. Wu and Mu [1] found that the urban population has the greatest impact on the circulation of vegetables; Hu et al. [2] analyzed the impact of agricultural informatization on vegetable yield; Yu [3] found that the main factors affecting vegetable production are planting area, vegetable consumption, and agricultural financial expenditure; Yang and Sun [4] found that the main factors affecting vegetable yield are economic development, technical investment, and land; Li and Mu [5] analyzed the

impact of financial capital on Farmers' willingness to continue planting vegetables; Yin [6] constructed a wavelet neural network prediction model and predicted tomato yield; Qiao and Liu [7] analyzed the impact of the aging of vegetable farmers on vegetable yield. Jin et al. [8] analyzed the role of logistics service providers in detail. From the existing research results, scholars mainly focus on the field of vegetable production, which cannot completely reflect the whole process of vegetable supply. Scholars rarely or even did not consider the grey characteristics of some known information, some unknown information, and poor information of the influencing factors of vegetable supply. In addition, when constructing the prediction model, the multicollinearity and overfitting between the influencing factors are not considered deeply. When considering the influence of multiple factors, Lasso (least absolute shrinkage and selection operator) method can eliminate those unimportant factors from many influencing factors, select a few important factors that have a significant impact on the system, and effectively eliminate the problems of multicollinearity and overfitting. GM (1, N) model is one of the grey prediction models considering the influence of many

factors. In view of the large error of the grey prediction model in practical application, this paper proposes an improved PSO-GM (1,  $N$ ) model, which provides a new way to study the prediction of vegetable supply.

## 2. PSO-GM (1, $N$ ) Model

For a multifactor prediction problem, let  $X_0 = (x_0(1), x_0(2), \dots, x_0(n))$  be the system characteristic data sequence, and  $X_i = (x_i(1), x_i(2), \dots, x_i(n))$  be the influencing factor sequence. In order to eliminate the multicollinearity problem of relevant influencing factors, PSO-GM (1,  $N$ ) first screens the relevant influencing factors by Lasso method, then determines the priority of influencing factors and the value of parameter  $N$  in GM (1,  $N$ ) model, that is, the second selection of relevant influencing factors, and finally optimizes GM (1,  $N$ ) model by PSO method.

**2.1. Lasso Method.** Lasso is a regularized compression estimation method proposed by statistician Robert Tibshirani in 1996. Regularization is a strategic method that can constrain the characteristics of the model and prevent overfitting. Based on the loss function of the least squares (ordinary least square (OLS)), Lasso uses the sum of the absolute values of the regression coefficients as a penalty function to compress the regression coefficients. When the sum of the absolute values of the regression coefficients is small enough, some regression coefficients can be compressed to zero, and then the variables with zero coefficients can be eliminated, so as to achieve the effect of variable selection [9]. Assumed linear regression model  $X_0 = \beta X + \varepsilon$ , where  $X_0$  is the system behavior characteristic vector,  $X$  is the influence factor variable matrix, and  $\beta$  is the coefficient vector, then the coefficient estimation of Lasso method is as follows:

$$\hat{\beta} = \arg \min \left\{ \|Y - X\beta\|^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}. \quad (1)$$

Here,  $X_0 = (x_0(1), x_0(2), \dots, x_0(n))$ .

$$X = \begin{bmatrix} x_1(1) & x_2(1) & \dots & x_s(1) \\ x_1(2) & x_2(2) & \dots & x_s(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(n) & x_2(n) & \dots & x_s(n) \end{bmatrix}, \quad (2)$$

where  $\|Y - X\beta\|^2$  is the loss function,  $\lambda \sum_{j=1}^p |\beta_j|$  is the regularization function, and  $\lambda$  is the adjustment parameter, also known as regularization parameter; it is mainly used to balance loss function and regularization function. When  $\lambda$  ( $\lambda \geq 0$ ) is increasing gradually from 0, the regularization function can set the estimated values of some coefficients to zero, so as to eliminate the corresponding variables and achieve the purpose of variable selection. The value of  $\lambda$  is determined by cross validation, that is, one part of the influencing factor sample data is selected as the training set and the other part is selected as the validation set, and different models are generated under the same training set. When the error of the verification set in these models is the smallest, at this time,  $\lambda$  is selected. Using Lasso method, if  $m$  main influencing factors are selected from  $s$  influencing factors, i.e., the influencing factor set is screened for the first time, it can be recorded as  $X'_i$  ( $i = 1, 2, \dots, m$ ).

**2.2. Priority of Influencing Factors.** Based on the set of influencing factors selected for the first time, Deng's grey relational analysis model is used to determine the priority of each influencing factor. Due to the different dimensionless processing methods of the original data, it has different effects on the grey correlation order of the influencing factors. Here, this paper uses six processing methods, i.e., ① initial value, ② average value, ③ minimization, ④ maximization, ⑤ centralization, and ⑥ difference.

Because the correlation order is reflected in the difference of the influence factors on the system behavior, and the more obvious the difference is, the better the reasonable correlation order is selected based on the following principles. Suppose six dimensionless processing methods are used to get the correlation degree of influencing factors as  $r_k^i$  ( $i = 1, 2, \dots, m, k = 1, 2, \dots, 6$ ). On the premise of small amount of calculation, two principles are proposed to judge the reasonable correlation order [10]:

$$\Delta_k = \max\{r_i^k\} - \min\{r_i^k\}, \quad (k = 1, 2, \dots, n), \text{ as large as possible,} \quad (3)$$

$$\sigma_k = \frac{1}{n} \sqrt{\left( r_i^k - \frac{1}{n} \sum_{i=1}^n r_i^k \right)^2}, \text{ as large as possible.} \quad (4)$$

For the influencing factor sequence  $X'_i(i = 1, 2, \dots, m)$ , according to the above principles, the reasonable correlation order of influencing factors can be determined. For convenience of expression, it is recorded here as follows:

$$X'_1 > X'_2 > \dots > X'_i \dots > X'_m, \quad (i = 1, 2, \dots, m). \quad (5)$$

### 2.3. GM (1, N) Model and Its N Value Determination Method

**2.3.1. GM (1, N) Model.** GM (1, N) model is a first-order differential equation prediction model with N variables, which reflects the influence of N-1 influencing factor variables on the first derivative of one system behavior variable. In order to be consistent with the traditional GM (1, N) model representation method, the corresponding system behavior characteristic sequence and the influencing factor sequence selected and determined for the first time are reexpressed as follows:

$$X_1^{(0)} = (x_1^{(0)}(1), \dots, x_1^{(0)}(i), \dots, x_1^{(0)}(n)). \quad (6)$$

Influencing factor sequence  $X'_i(i = 1, 2, \dots, m)$  can be expressed as

$$\begin{aligned} X_2^{(0)} &= (x_2^{(0)}(1), \dots, x_2^{(0)}(i), \dots, x_2^{(0)}(n)), \\ X_3^{(0)} &= (x_3^{(0)}(1), \dots, x_3^{(0)}(i), \dots, x_3^{(0)}(n)), \\ &\vdots \\ X_N^{(0)} &= (x_N^{(0)}(1), \dots, x_N^{(0)}(i), \dots, x_N^{(0)}(n)). \end{aligned} \quad (7)$$

Let  $X_j^{(1)}$  be the 1-AGO sequence of  $X_j^{(0)}$ , ( $j = 1, 2, \dots, N$ ), and  $Z^{(1)}$  be the nearest neighbor mean generating sequence of  $X_1^{(1)}$ , then

$$x_1^{(0)}(k) + az_1^{(1)}(k) = \sum_{j=2}^N b_j x_j^{(1)}(k), \quad k = 1, 2, \dots, n. \quad (8)$$

is called GM (1, N).

When the amplitude of change of  $X_j^{(1)}$ , ( $j = 1, 2, \dots, N$ ), is very small, the approximate time response of GM (1, N) is as follows:

$$\hat{x}_1(k+1) = \frac{1}{\alpha} \sum_{j=2}^N b_j x_j^{(1)}(k+1) + \left[ x_1^{(1)}(0) - \frac{1}{a} \sum_{j=2}^N b_j x_j^{(1)}(k+1) \right] e^{-ak}, \quad k = 1, 2, \dots, n, \quad (9)$$

where  $x_1^{(1)}(0)$  is taken as  $x_1^{(0)}(1)$ .

The predicted values of system behavior variables are as follows:

$$\hat{x}_1^{(0)}(k+1) = \hat{x}_1^{(1)}(k+1) - \hat{x}_1^{(1)}(k). \quad (10)$$

**2.3.2. Determination of N Value in GM (1, N) Model.** At present, GM (1, N) model has been widely used and innovated. Ma and Liu [11] proposed discrete GM (1, N) model and analyzed the basic law of oil production decline and the influence of related factors; Fan [12] proposed an improved GM (1, N) soybean price forecasting model; Xie et al. [13] proposed the GICM model, screened the key indicators of complex products, and improved the prediction accuracy of GM (1, N) model; Ren [14] used GM (1, N) method to model anaerobic digestion system and predict methane production; Zeng et al. [15] proposed a new multivariable grey prediction GM (1, N) model; Xiong et al. [16] improved GM (1, N) model and proposed AWGM (1, N) model to predict housing demand; Yang and Liu [17] used GM (1, N) model to predict China's grain output; Cheng et al. [18] established grey GM (1, 3) model to simulate and predict the main influencing factors of clean energy consumption; Zhang [19] analyzed the influencing factors of ginger planting area from the aspects of economy, input, and output and established a grey prediction model.

From the existing research results, no scholar has studied the value of parameter N in GM (1, N) model. This paper proposes a method to determine the value of the parameter N in GM (1, N) model: according to the correlation order of

the influencing factors of system behavior, the influencing factor variables are selected in order to establish the model.

In order to test the accuracy of the prediction results of the model, the last three samples are selected from the n sample of the system as the comparison value of the prediction results of the model, and the model is established by using n-3 sample data. Through the test of the prediction effect, the optimal N value is determined. The specific steps are as follows:

In Step 1, based on the reexpression of influencing factor sequence in relation (5) and 2.3.1, select  $X_2^{(0)}$ , establish GM (1, 2) model, and predict the three periods of n-2, n-1, and n. Through comparing with the actual value and calculating the average relative error, the result is recorded as  $\varepsilon_2$ .

In Step 2, according to the above method and the correlation order of influencing factors determined in relation (5), the GM (1, 3) model, GM (1, 4) model, ..., GM (1, N) model are established, respectively, and the average relative errors are calculated, which are recorded as  $\varepsilon_3, \varepsilon_4, \dots, \varepsilon_N$ .

In Step 3, when  $\varepsilon_i (i = 2, 3, \dots, N)$  is the minimum, the prediction accuracy of the model is the highest. In this case, the value of i is the optimal solution, that is, the GM (1, i) model is the optimal prediction model.

This process can be called the second selection influencing factor set.

**2.3.3. PSO-GM (1, N) Prediction Model.** Particle swarm optimization (PSO) initializes a group of random particles and then finds the optimal value through multiple iterations.

In each iteration, each particle updates its speed and position through the individual extreme  $pbest$  and global extreme  $gbest$ . The calculation formula is as follows:

$$\begin{aligned} v_{k+1} &= \omega v_k + c_1 r_1 (pbest_k - x_k) + c_2 r_2 (gbest_k - x_k), \\ x_{k+1} &= x_k + v_{k+1}, \end{aligned} \quad (11)$$

where  $v_k$  is the current velocity of the particle,  $x_k$  is the current position of the particle,  $pbest_k$  is the position of the optimal value currently found by the particle, and  $gbest_k$  is the position of the optimal value currently found by the whole population.  $\omega$  is the inertia weight,  $c_1, c_2$  is the acceleration coefficient, usually  $c_1 = c_2 = 2$ .  $r_1, r_2$  is a random number between (0, 1). Through continuous updating, particles finally reach the position of the global optimal value.

According to the basic principle of GM (1, N) model, let  $u = (a, b_1, b_2, \dots, b_{n-1})^T$ ; the calculation formula is  $u = (B^T B)^{-1} B Y$ , where

$$\begin{aligned} B &= \begin{bmatrix} -z_1^{(1)}(2) & x_2^{(1)}(2) & \dots & x_n^{(1)}(2) \\ -z_1^{(1)}(3) & x_2^{(1)}(3) & \dots & x_n^{(1)}(3) \\ \vdots & \vdots & \ddots & \vdots \\ -z_1^{(1)}(n) & x_2^{(1)}(n) & \dots & x_n^{(1)}(n) \end{bmatrix}, \\ Y &= \begin{bmatrix} x_1^{(0)}(2) \\ x_1^{(0)}(3) \\ \vdots \\ x_1^{(0)}(n) \end{bmatrix}. \end{aligned} \quad (12)$$

It can be seen that the representation of GM (1, N) model depends on matrix  $B$ . The estimated value of the parameter  $u$  is obtained according to the above method, but it is not necessarily optimal. In this paper, PSO-GM (1, N) model is proposed: add other parameters to matrix  $B$ , and then use particle swarm optimization algorithm to obtain the optimal parameter  $u$ . Matrix  $B$  with parameters is reexpressed as follows:

$$B = \begin{bmatrix} -k_1 * z_1^{(1)}(2) & k_2 * x_2^{(1)}(2) & \dots & k_N * x_N^{(1)}(2) \\ -k_1 * z_1^{(1)}(3) & k_2 * x_2^{(1)}(3) & \dots & k_N * x_N^{(1)}(3) \\ \vdots & \vdots & \ddots & \vdots \\ -k_1 * z_1^{(1)}(n) & k_2 * x_2^{(1)}(n) & \dots & k_N * x_N^{(1)}(n) \end{bmatrix}. \quad (13)$$

Then, the parameter  $(\hat{a}, \hat{b}_1, \hat{b}_2, \dots, \hat{b}_{n-1})^T$  is reestimated according to formula  $u = (B^T B)^{-1} B Y$ , the parameter in  $(\hat{a}, \hat{b}_1, \hat{b}_2, \dots, \hat{b}_{n-1})^T$  is a function of  $k_i (i = 1, 2, \dots, N)$ , and GM (1, N) is also a function of  $k_i (i = 1, 2, \dots, N)$ .

According to the optimal prediction model GM (1, N) determined in 2.3.2, the relative errors of the model in three periods of  $n-2$ ,  $n-1$ , and  $n$  are calculated, respectively:  $\varepsilon(n-2)$ ,  $\varepsilon(n-1)$ , and  $\varepsilon(n)$ . The minimum average relative error in these three periods is taken as the objective function of particle swarm optimization algorithm, which is specifically expressed as follows:

$$\min \text{avg} = \frac{1}{3} (\varepsilon(n-2) + \varepsilon(n-1) + \varepsilon(n)). \quad (14)$$

### 3. Prediction of Vegetable Supply in Henan Province

**3.1. Data Sources and Influencing Factors.** Due to the lack of a perfect statistical index system for agricultural products logistics, this paper selects the product of the number of urban residents and the per capita fresh vegetable purchase of urban residents in Henan Province from 2007 to 2019 as the vegetable supply ( $X_0$ ) and uses the predicted value of vegetable supply ( $X_0$ ) to reflect the development and change of vegetable supply in Henan Province in the future. According to the existing research results and the principles of comprehensiveness, representativeness, and operability of index selection, 9 influencing factors are selected: vegetable planting area ( $X_1$ ) (thousand hectares), the application amount of pesticide ( $X_2$ ) (10 thousand tons), the usage of plastic film ( $X_3$ ) (10 thousand tons), plastic film coverage area ( $X_4$ ) (thousand hectares), the number of rural employees ( $X_5$ ) (10 thousand people), the investment in transportation, storage, and postal industry ( $X_6$ ) (100 million RMB), highway mileage ( $X_7$ ) (10 thousand km), the tonnage of road truck ( $X_8$ ) (10 thousand tons), and the number of employees in highway transportation industry ( $X_9$ ) (people). The data in this paper are all from the statistical yearbook of Henan Province.

#### 3.2. Data Processing and Result Analysis

**3.2.1. Determination of the First Influencing Factor Set.** Based on the influence factor set selected in 3.1, this paper takes the sample data of influence factors from 2007 to 2016 as the training set and uses Lasso method to determine the influence factors with the help of MATLAB software. The minimum error value can be obtained by iterating the program for 8 times. The factors corresponding to the nonzero term in the regression coefficient are the selection results of Lasso method. The specific results are shown in Table 1.

**3.2.2. Priority of Influencing Factors Selected for the First Time.** According to the influence factor set selected for the first time, the method established in 2.2 is used for processing, and the specific calculation results are shown in Table 2. According to the judgment criteria (3) and criteria (4), the ideal correlation order of influencing factors is obtained after the initial value processing method:

$$X_4' > X_6' > X_1' > X_3' > X_2' > X_5' > X_7'. \quad (15)$$

**3.2.3. Determination of N Value in GM (1, N) Model.** For expression (13), the method determined in 2.3.2 is used for processing, and the results are shown in Table 3. Because  $\varepsilon_4$  (4.93%) is the minimum value of  $\varepsilon_i (i = 2, 3, \dots, N)$ , so  $N = 4$ , that is, the influencing factors of the second selection

TABLE 1: Set of influencing factors for the first selection.

Factors influencing the first choice	Unit
Application amount of pesticide $X_1'$	10 thousand tons
Usage of plastic film $X_2'$	10 thousand tons
Plastic film coverage area $X_3'$	1 thousand hectares
Number of rural employees $X_4'$	10 thousand people
Investment in transportation, storage, and postal industry $X_5'$	100 million RMB
Highway mileage $X_6'$	10 thousand km
Tonnage of road truck $X_7'$	10 thousand tons

TABLE 2: Correlation order of various processing methods and its judging criteria.

Processing method	Correlation order of influencing factors	$\Delta_k$	$\sigma_k$
① Initial value	$X_4' > X_6' > X_1' > X_3' > X_2' > X_5' > X_7'$	0.4192	0.0677
② Average value	$X_4' > X_6' > X_1' > X_3' > X_2' > X_7' > X_5'$	0.3573	0.0612
③ Minimization	$X_6' > X_1' > X_3' > X_1' > X_2' > X_7' > X_5'$	0.3367	0.0529
④ Maximization	$X_1' > X_3' > X_6' > X_4' > X_2' > X_5' > X_6'$	0.3885	0.0660
⑤ Centralization	$X_5' > X_6' > X_4' > X_7' > X_2' > X_1' > X_3'$	0.0534	0.0077
⑥ Difference	$X_6' > X_3' > X_6' > X_2' > X_7' > X_1' > X_5'$	0.0801	0.0100

TABLE 3: Variable selection results of GM (1, N) model.

Model	Average relative error of prediction	Error value (%)
GM (1, 2)	$\varepsilon_2$	14.74
GM (1, 3)	$\varepsilon_3$	22.75
GM (1, 4)	$\varepsilon_4$	4.93
GM (1, 5)	$\varepsilon_5$	8.30
GM (1, 6)	$\varepsilon_6$	9.65
GM (1, 7)	$\varepsilon_7$	8.67
GM (1, 8)	$\varepsilon_8$	8.06

are the number of rural employees, highway mileage, and application amount of pesticide.

3.2.4. *Prediction Results of PSO-GM (1, N) Model and Analysis of Influencing Factors.* The PSO-GM (1, N) model established in 2.3.3 is used to predict the vegetable supply in Henan Province from 2017 to 2019, in which the number of particles is 100, the inertia weight is 1, the acceleration coefficient is 2, and the number of iterations is 1000. After many experiments and based on the sample data from 2010 to 2016, the average relative error of vegetable supply in Henan Province from 2017 to 2019 is the smallest, reaching 2.38%.

In order to compare the rationality of the prediction results, PSO-GM (1, 4) model, GM (1, 4) model, GM (1, 1) model, and multiple linear regression method are used to predict the vegetable supply in Henan Province. The actual value of vegetable supply in Henan Province from 2017 to 2019 is taken as the standard. The results are shown in Table 4. It can be seen from Table 4 that the average relative error from high to low is PSO-GM (1, 4) model, GM (1, 4) model, GM (1, 1) model, and multiple linear regression. It can be concluded that compared with the traditional GM (1, N) model, PSO-GM (1, N) model greatly improves the prediction accuracy and is much higher than other

prediction models. It has strong rationality and can effectively predict the future vegetable supply.

According to the result of 3.2.3, PSO-GM (1, 4) model is used to forecast the vegetable supply in Henan Province in the next three years. The results are shown in Table 5. The data of influencing factors in the next three years are obtained by GM (1, 1) model, and the results are shown in Table 6.

It can be concluded from Table 5 that the vegetable supply in Henan Province will still show a steady growth trend in the next three years, and it is expected to reach 6.35 million tons in 2023. The key factors affecting the vegetable supply in Henan Province are the number of rural employees, highway mileage, and pesticide application. Among them, the number of rural employees and road mileage has a positive impact on vegetable supply, while the application amount of pesticides has a negative impact on vegetable supply.

Vegetable production is a labor-intensive industry; the number of labor has a direct impact on vegetable production. With the development of social economy, agricultural machinery has greatly improved labor productivity, but in vegetable production, due to the limitations of vegetable types and terrain, agricultural machinery is difficult to get effective application and still needs a lot of labor for manual operation. Therefore, the impact of rural employment on vegetable supply is positive.

TABLE 4: Prediction results of PSO-GM (1, 4)/GM (1, 4)/GM (1, 1)/multiple linear regression.

Year	Actual value	PSO-GM (1, 4)	GM (1, 4)	GM (1, 1)	Multiple linear regression
2017	504.11	518.49	546.50	543.01	535.48
2018	538.41	536.02	571.50	587.93	480.18
2019	570.21	548.23	568.90	636.57	490.72
Average relative error (%)		2.38	4.93	9.52	10.32

TABLE 5: Forecast value of vegetable supply in Henan Province.

Vegetable supply	Year		
	2021	2022	2023
Estimate	578.27	592.07	635.71

TABLE 6: Prediction of influencing factors of vegetable supply.

Influence factors	Year		
	2021	2022	2023
Number of rural employees	4083	3915	4440
Highway mileage	27.17	27.28	27.38
Application amount of pesticide	9.51	8.97	8.45

The modes of freight transportation include road transportation, railway transportation, waterway transportation, and air transportation. Due to the characteristics of short distance transportation of vegetables, highway transportation is widely used in vegetable transportation. Transportation infrastructure construction plays an important role in promoting vegetable transportation. Perfect traffic network can speed up the circulation of vegetables and expand the scale of vegetable market. Highway mileage is an important indicator of regional traffic infrastructure. Therefore, the impact of highway mileage on vegetable supply is positive.

The excessive pesticide residue in vegetables is an important task in the supervision of vegetable quality and safety. In the process of vegetable production, due to the frequent occurrence of diseases and insect pests, farmers have overdose in vegetable production in order to ensure vegetable yield. Due to the strict control of government departments, the supply of vegetable market has decreased significantly. In order to stabilize the normal supply of vegetable market, it is urgent to reduce the use of pesticides. Therefore, the effect of pesticide application on vegetable supply is negative.

#### 4. Conclusions

Vegetable supply is affected by multiple factors, and the relationship between influencing factors is complex. In this paper, when using PSO-GM (1,  $N$ ) model to predict vegetable supply in Henan Province, firstly, Lasso method is used to select many factors affecting vegetable supply for the first time, which eliminates the problems of multicollinearity and overfitting between factors. Then, based on the grey correlation analysis, the priority of the influencing factors is sorted, the determination method of parameter  $N$  in GM (1,  $N$ ) is proposed, and the influencing factors are selected for

the second time. Finally, the PSO method is used to optimize the GM (1,  $N$ ) model, which greatly improves the prediction accuracy. Based on PSO-GM (1,  $N$ ) model, the vegetable supply in Henan Province in the next three years is predicted. The results show that the vegetable supply in Henan Province continues to show an increasing trend. The key factors affecting vegetable supply in Henan Province are the number of rural employees, highway mileage, and application of pesticide. The PSO-GM (1,  $N$ ) model proposed in this paper provides a new method for multifactor grey prediction. However, this study does not consider the impact of adjacent provinces on vegetable supply in Henan Province, especially spatial autocorrelation. Next, the temporal and spatial evolution law of vegetable supply in Henan Province will be explored from the two dimensions of time and space.

#### Data Availability

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

#### Conflicts of Interest

The authors have no conflicts of interest.

#### Authors' Contributions

Bingjun Li was responsible for proposing the overall idea and framework of the manuscript. Xueqiang Guo was responsible for data processing and writing of the first draft of the manuscript.

#### Acknowledgments

This work was supported by the Key Project of Soft Science Research in Henan Province (202400410051).

#### References

- [1] S. Wu and Y. Y. Mu, "Analysis of vegetable circulation and its influencing factors in China based on temporal and spatial characteristics," *Journal of Business Economics*, no. 2, pp. 18–25, 2016.
- [2] S. X. Hu, L. Chong Guang, L. Wei Juan, Z. Chun Yang, and S. Xiang Cheng, "Empirical Study on the influence of agricultural informatization on the development of vegetable industry in Hubei Province," *Journal of Henan Agricultural Sciences*, vol. 46, no. 7, pp. 154–160, 2017.
- [3] Y. S. Yu, "Establishment and correlation analysis of grey prediction model GM (1, 1) for vegetable yield in Guangxi," *Journal of Guangdong Agricultural Sciences*, vol. 45, no. 7, pp. 157–164, 2018.

- [4] X. L. Yang and Z. L. Sun, "An empirical study on the influencing factors of vegetable industry production in Heilongjiang Province," *Northern Horticulture*, no. 1, pp. 177–182, 2019.
- [5] H. R. Li and Y. Y. Mu, "Study on the influence of financial capital on Farmers' willingness to continue planting vegetables," *Chinese Vegetables*, vol. 11, pp. 63–69, 2019.
- [6] Y. Z. Yin, "Prediction of tomato yield in greenhouse based on wavelet neural network," *Zhong Guo Gua Cai*, vol. 33, no. 8, pp. 53–59, 2020.
- [7] Z. X. Qiao and S. Liu, "Influence of aging vegetable farmers on vegetable production -- a case study of Yuzhong County, Gansu Province," *Journal of Chinese Agricultural Mechanization*, vol. 41, no. 6, pp. 231–236, 2020.
- [8] F. F. Jin, M. Cao, J. Liu, L. Martínez, and H. Chen, "Consistency and trust relationship-driven social network group decision-making method with probabilistic linguistic information," *Applied Soft Computing*, vol. 103, 2021.
- [9] L. J. Zhang, "Lasso regression: from interpretation to prediction," *Advances in Psychological Science*, vol. 28, no. 10, pp. 1777–1788, 2020.
- [10] B. J. Li, "Influence of dimensionless processing of original data on grey relational order," *Journal of Henan Agricultural University*, vol. 36, no. 2, pp. 199–202, 2002.
- [11] X. Ma and Z. B. Liu, "Predicting the oil field production using the novel discrete GM (1, N) model," *Journal of Grey System*, vol. 27, no. 4, pp. 63–73, 2015.
- [12] Z. Fan, "Analysis and prediction of soybean price in China based on Improved GM (1, N) model," *Soybean Science*, vol. 35, no. 5, pp. 847–852, 2016.
- [13] N. Xie, C. Hu, and S. Yin, "Key indexes choosing approach of complex equipment's development cost based on grey incidence cluster model," *Grey Systems: Theory and Application*, vol. 6, no. 1, pp. 110–123, 2016.
- [14] J. Ren, "GM (1, N) method for the prediction of anaerobic digestion system and sensitivity analysis of influential factors," *Bioresource Technology*, vol. 247, pp. 1258–1261, 2018.
- [15] B. Zeng, H. Duan, and Y. Zhou, "A new multivariable grey prediction model with structure compatibility," *Applied Mathematical Modelling*, vol. 75, pp. 385–397, 2019.
- [16] X. Xiong, H. Guo, and X. Hu, "The housing demand analysis and prediction of the real estate based on the AWGM (1, N) model," *Grey Systems: Theory and Application*, vol. 11, no. 2, pp. 222–240, 2020.
- [17] W. Yang and B. Li, "Prediction of grain supply and demand structural balance in China based on grey models," *Grey Systems: Theory and Application*, vol. 11, no. 2, pp. 253–264, 2020.
- [18] M. L. Cheng, J. N. Li, Y. Liu, and B. Liu, "Forecasting clean energy consumption in China by 2025: using improved grey model GM (1, N)," *Sustainability*, vol. 12, no. 2, p. 698, 2020.
- [19] K. Z. Zhang, "Prediction of ginger planting area based on GM (1, N) model," *Journal of Chinese Agricultural Mechanization*, vol. 41, no. 10, pp. 139–143, 2020.