

Research on optimal crop planting strategy based on particle swarm algorithm

Qianyun Zhou[#], Huyi Li[#], Shihan Yang[#]

Savaid Stomatology School, Hangzhou Medical College, Hangzhou, 311399, China

[#]These authors contributed equally.

Abstract: A village in the mountainous region of North China faces the challenges of limited arable land resources and low-temperature environments, and needs to optimise crop planting strategies to enhance production efficiency and reduce risks. To this end, this paper proposes an optimal crop planting strategy model that integrates linear programming, Monte Carlo simulation and particle swarm algorithm. Firstly, the optimal planting strategy was formulated by linear programming and particle swarm algorithm without considering the uncertainty; secondly, the uncertainty of the expected future sales volume, mu yield, planting cost and sales price was further considered, and a dynamic planning model was established by Monte Carlo simulation and particle swarm algorithm, which yielded the average profit of the calendar year to be increased to 62,346,800 Yuan; finally, inter-crop substitutability was introduced to explore the correlation between crops using K-means cluster analysis and a linear regression model of expected sales was developed by least squares. The results showed that the planting strategy after considering inter-crop correlation could further improve the production efficiency. The study in this paper is of great value for future extension and can be applied to areas with similar agro-ecological environments and cropping conditions to provide guidance for local agricultural cropping strategies.

Keywords: Linear Programming, Particle Swarm Algorithm, Monte Carlo Simulation, K-means Clustering

1. Introduction

The research background of this thesis is that a rural village in the mountainous region of North China is facing the challenges of limited arable land resources and low-temperature environments, and needs to optimise crop planting strategies to enhance production benefits and reduce risks. The traditional crop planting strategy is the study of planting structure optimisation through linear programming, BI Chunning et al [1] constructed the E-FOS-MOP model to explore the crop planting structure of the Yellow River estuary area under different river green assurance, and achieved better results. ZHANG P F et al [2] optimised the adjustment of agricultural planting structure based on the water-energy-grain correlation relationship, and obtained the results of economic benefit improvement. Guo Ping et al [3] based on the Pareto solution set of agricultural water and soil resources optimisation allocation planning model, but this kind of linear planning model lacks more generality of the solution procedure, and the use of mean value instead of crop prices and other uncertainty values, has certain limitations. ZHANG H Hao et al. [4] introduced an on-the-fly dynamic planning model to take full account of the uncertainty of water availability in the irrigation area, so that the irrigation area can play a high and stable economic efficiency. In recent years, machine learning algorithms have been widely used in global optimal search problems, DONG Chen-chao [5] constructed a genetic algorithm-generative adversarial neural network model for optimal scheduling of water resources in artesian irrigation districts, and found that the efficiency of water resources utilisation has been significantly improved through validation and application. WANG Zhipeng et al[6] reviewed the methods of determining the objectives of planting structure optimisation and the traditional and intelligent algorithms for model solving, PAN Yue et al[7], solved the multi-objective optimal allocation model by using DSF-GWO algorithm, and got the optimization scheme of agricultural planting structure in irrigation areas under different development scenarios. LI Yanbin [8],ZHANG Qian [9] et al. used improved particle swarm algorithm in agricultural planting structure optimisation, introducing two improvements of relational weight attenuation and particle swarm variation strategy, which led to the improvement of planting structure.

However, through analysis, it was found that these studies were not perfect enough for the

consideration of constraints, and lacked the consideration of uncertainties such as the expected future sales volume and planting costs. In this paper, based on the in-depth study of the optimal planting strategy based on particle swarm algorithm, Monte Carlo method, K-means cluster analysis, least squares and other methods are introduced to study the impact of the uncertain factors and the correlation between them. Firstly, this paper establishes the optimal planting strategy model based on particle swarm algorithm on the basis of the data in 2023, taking the type of arable land, crop rotation, planting area limitation, seasonal limitation and other constraints, and maximising the revenue as the objective function; secondly, the distribution of the fluctuating ranges of the expected future sales volume, mu yield, planting cost and sales price are described through the assumptions and Monte Carlo simulation is used for the simulation to formulate the optimal. Finally, the correlation between price, cost and expected sales volume was introduced into the study, and K-means cluster analysis was used to analyse the differences in the characteristics of crops, and a linear regression model of expected sales volume was established by the least squares method.

2. Materials and methods

2.1 Data Acquisition

The research data for this paper was obtained from the open source website Github.

2.2 Particle Swar Algorithms

In carrying out the actual solution, it is found that the complexity of the model operation is greatly increased due to the addition of new constraints, so we introduce the particle swarm algorithm to solve the objective function so as to reduce the complexity of the operation process. The particle swarm algorithm is actually an intelligent optimisation method that simulates the foraging of a flock of birds to find the optimal solution, which can search for the optimal solution by iterating on the basis of a random solution, and search for the global optimal solution by using the fitness as the evaluation condition.

Suppose that in an M dimensional search space, The initial coordinates of the D particles are $U_i = (u_{i1}, u_{i2}, \dots, u_{im})$, the initial flight speed is $(m = 1, 2, \dots, M, i = 1, 2, \dots, D)$, D , the optimal position of the n particles is $P_i = (p_{i1}, p_{i2}, \dots, p_{im})$.

2.3 Monte Carlo method

This paper focuses on the simulation of random variables in different scenarios using the Monte Carlo method, based on a random sampling computational method, in order to achieve the optimal solution. The general steps are as follows.

- (1) Model the probability distribution according to the problem $\psi(x)$.
- (2) Generate random number x according to the probability distribution and calculate $f(x)$ according to the function model $\psi(x)$.
- (3) Solve the Monte Carlo simulation final result with the mathematical expression:

$$E = \frac{D_s}{N} \sum_i^N f(X_1^n, X_2^n, \dots, X_n^n) \quad (1)$$

- (4) Completion Exit.

2.4 K-means cluster analysis

In order to reduce the differences between different crops due to their own characteristics, this paper introduces K-means cluster analysis to discuss the classification of each crop. K-means cluster analysis is the most basic and commonly used clustering algorithm. The main idea is to randomly select k clustering centres from the sample, and then according to the Euclidean distance to assign each point to the cluster closest to its mean, and then calculate the mean vector of the points assigned to each cluster,

and recursively as a new centre, and ultimately get the clustering results. The specific process is as follows:

(1) Select K objects as initial cluster centres: C_1, \dots, C_k

(2) Calculate the distance d from each point I_i to the centre C_j of each cluster, and assign it to the cluster L_j corresponding to the smallest d .

$$d = |I_i - C_j| (i \in \{1, 2, \dots, m\}, j \in \{1, 2, \dots, k\}) \quad (2)$$

(3) Use the mean value of all objects in each cluster as the new cluster centre:

$$C_j = \sum I_i / |L_j| \quad (3)$$

Where $I_i \in L_j, |L_j|$ is the number of data points in cluster L_j .

(4) Cycle through steps (2) and (3) until each cluster no longer changes. k-means clustering of K clusters with output minimum standard variance.

3. Results and analysis

3.1 Crop planting strategy modelling based on particle swarm algorithm

3.1.1 Identification of decision variables

In this paper, $x_{y,j,i}^m$ is defined as a decision variable indicating the area of the i th crop cultivated on the m th plot of land in the y th quarter of the year. Where y denotes the year, j denotes the parcels of land (in the order of flat dry land, terraced land, hillside land, irrigated land, watered land, irrigated land, ordinary greenhouses, and smart greenhouses), and m denotes the m th quarter, which takes the value of 1 or 2 to denote the first and the second quarter of the season.

3.1.2 Setting the objective function: maximising returns

In this paper, maximising revenue is used as the objective function, so that when there is a situation where the production of a crop exceeds the expected sales, the excess portion is sold at half price. The objective function is obtained as maximising the total revenue from the normal sale portion plus the revenue from the reduced price sale portion (at 50% of the 2023 sale price) minus the cost of cultivation.

$$\text{Maximize} = \sum_{y=2024}^{2030} \sum_{j=1}^{54} \sum_{i=1}^{41} \sum_{m=1}^2 (\min(ES_i^m, S_{y,j,i}^m * x_{y,j,i}^m) * P_{y,j,i}^m - C_{y,j,i}^m * x_{y,j,i}^m + 0.5 * P_{y,j,i}^m * \max(0, S_{y,j,i}^m * x_{y,j,i}^m - ES_i^m)) \quad (4)$$

Where ES_i^m refers to the expected sales volume of the first crop in the m th quarter, $P_{y,j,i}^m$: refers to the unit price of the i th crop on the first plot of land in the m th quarter of the y th year, and $C_{y,j,i}^m$: refers to the cost of growing the i th crop on the first plot of land in the m th quarter of the y th year,

3.1.3 Determining constraints

(1) Limitations on the total area of each plot.

$$\sum_m \sum_{i=1}^{41} x_{y,j,i}^m \leq SQ_j \forall j, y \in \{2024, \dots, 2030\} \quad (5)$$

(2) Types of crops grown on specific plots.

$$x_{y,j,i}^m \leq u_{j,i}^m \cdot SQ_j, \forall i, j, m \in \{1, 2\}, y \in \{2024, \dots, 2030\} \quad (6)$$

where $u_{j,i}^m$ is a newly defined ternary decision variable (a value of 1 indicates that it can be planted,

and a value of 0 indicates that it cannot be planted), T_j represents the type of plot j (e.g., flat dry land, terraced land, hillside land, etc.), and L_i represents the type of crop i (e.g., grain, grain (legume), vegetable, vegetable (legume), and edible mushroom).

(3) Each crop cannot be grown consecutively on a single plot of land.

That is, the crop types cannot be the same in two consecutive years or between seasons:

$$\begin{cases} x_{y,j,i}^1 * x_{y,j,i}^2 = 0(m = 2) \\ x_{y,j,i}^1 * x_{y-1,j,i}^2 = 0(m = 2) \\ x_{y,j,i}^1 * x_{y-1,j,i}^1 = 0(m = 1) \end{cases} \quad (7)$$

(4) Each plot must be planted with a legume crop once in three consecutive years.

Indicates that each plot must be planted with a legume crop once in three consecutive years:

$$x_{y-1,j,i}^1 + x_{y-1,j,i}^2 + x_{y,j,i}^1 + x_{y,j,i}^2 + x_{y+1,j,i}^1 + x_{y+1,j,i}^2 \geq 1 \forall j, i \in \{1, 2, 3, 4, 5, 17, 18, 19\} \forall y \quad (8)$$

(5) the seasonal limitation of crops

In view of the function of the greenhouse, the seasonal limitation of crop cultivation should also be considered. For ‘ordinary greenhouses’ ($T_j = T_5$), only one season of vegetables and one season of mushrooms can be grown each year:

$$\sum_{i \in \{i|N_i=N_s\}} x_{y,j,i}^1 + \sum_{i \in \{i|N_i=N_s\}} x_{y,j,i}^2 \leq SQ_j, \forall j \text{ with } K_j = K_5, y \in \{2024, \dots, 2030\} \quad (9)$$

For ‘smart greenhouses’ ($T_j = T_6$), two seasons of vegetables can be grown each year:

$$\sum_{i \in \{i|N_i=N_f\}} x_{y,j,i}^m \leq SQ_j, \forall j \text{ with } K_j = K_6, m \in \{1, 2\}, y \in \{2024, \dots, 2030\} \quad (10)$$

3.1.4 Range of independent variables

In the process of model building, the code readability of the particle swarm algorithm also needs to be considered, so the range of values of the independent variables should be defined according to the limitations of the actual scene. For example, to ensure that the planting area of the same crop is not too dispersed, the planting area is set here as a multiple of 0.1, i.e:

$$x_{i,j,k,t} \geq 0.1 \cdot b_{i,j,k,t}, \forall i, j, k \in \{1, 2\}, t \in \{2024, \dots, 2030\} \quad (11)$$

Where $b_{i,j,k,t}$ is a binary variable, when $x_{i,j,k,t} > 0$, $b_{i,j,k,t} = 1$; otherwise, $b_{i,j,k,t} = 0$.

3.1.5 Model aggregation

In summary, the optimal planting scheme model initially established based on the expected future sales volume, planting cost, mu yield and sales price of various crops remain unchanged is shown below:

$$\text{Maximize} = \sum_{y=2024}^{2030} \sum_{j=1}^{54} \sum_{i=1}^{41} \sum_{m=1}^2 (\min(ES_i^m, S_{y,j,i}^m * x_{y,j,i}^m) * P_{y,j,i}^m - C_{y,j,i}^m * x_{y,j,i}^m + 0.5 * P_{y,j,i}^m * \max(0, S_{y,j,i}^m * x_{y,j,i}^m - ES_i^m)) \quad (12)$$

$$\left. \begin{aligned}
 & \sum_m \sum_{i=1}^{41} x_{y,j,i}^m \leq SQ_j \forall j, y \in \{2024, \dots, 2030\} \\
 & x_{y,j,i}^m \leq u_{j,i}^m \cdot SQ_j, \forall i, j, m \in \{1, 2\}, y \in \{2024, \dots, 2030\} \\
 & x_{y,j,i}^1 * x_{y,j,i}^2 = 0 (m = 2) \\
 & x_{y,j,i}^1 * x_{y-1,j,i}^2 = 0 (m = 2) \\
 & x_{y,j,i}^1 * x_{y-1,j,i}^1 = 0 (m = 1) \\
 & x_{y-1,j,i}^1 + x_{y-1,j,i}^2 + x_{y,j,i}^1 + x_{y,j,i}^2 + x_{y+1,j,i}^1 + x_{y+1,j,i}^2 \geq 1 \forall y \\
 & \sum_{i \in \{i | N_i = N_s\}} x_{y,j,i}^1 + \sum_{i \in \{i | N_i = N_s\}} x_{y,j,i}^2 \leq SQ_j, \forall j \text{ with } K_j = K_5 \\
 & \sum_{i \in \{i | N_i = N_j\}} x_{y,j,i}^m \leq SQ_j, \forall j \text{ with } K_j = K_6, m \in \{1, 2\}
 \end{aligned} \right\} \text{s.t.} \tag{13}$$

3.2 Optimal planting strategy modelling based on Monte Carlo method

3.2.1 Data analysis and testing

Specific variables fluctuated as follows: expected sales of wheat and maize increased between 5 and 10 per cent per year; expected sales of other crops varied ± 5 per cent per year relative to the first year; acres of each crop varied ± 10 per cent per year; and the cost of cultivation increased by an average of about 5 per cent per year. Of the sales prices, grain crops are stable; vegetable crops increase by an average of about 5 per cent per year, edible mushrooms decline by 1 to 5 per cent per year, and morel mushrooms decline by 5 per cent per year. Relevant data are then collected and analysed, as shown in Figure 1, Figure 2, Figure 3 and Figure 4.

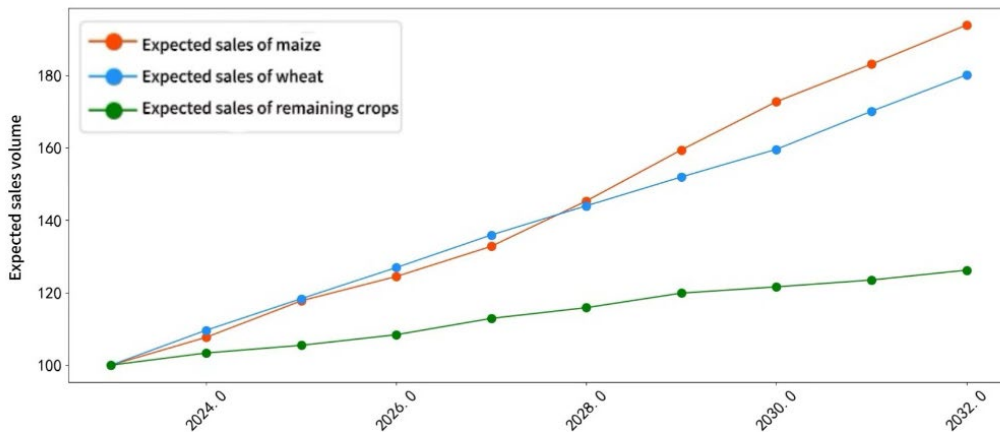


Fig.1: Expected sales various crops

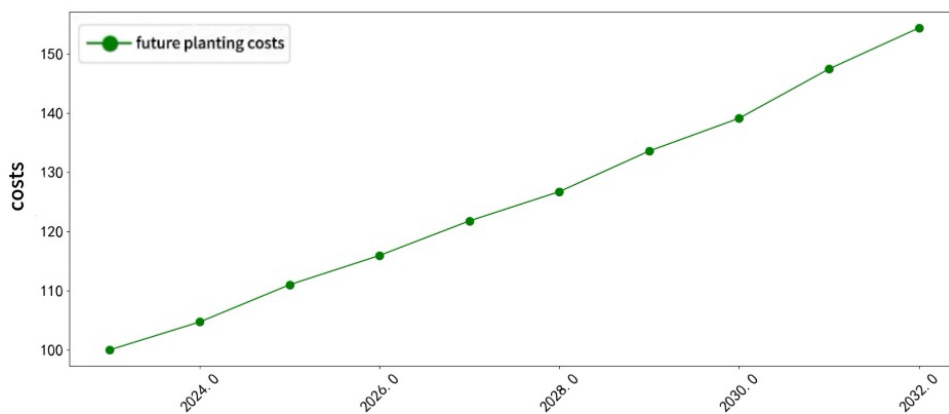


Fig.2: Cost of cultivation of various crops

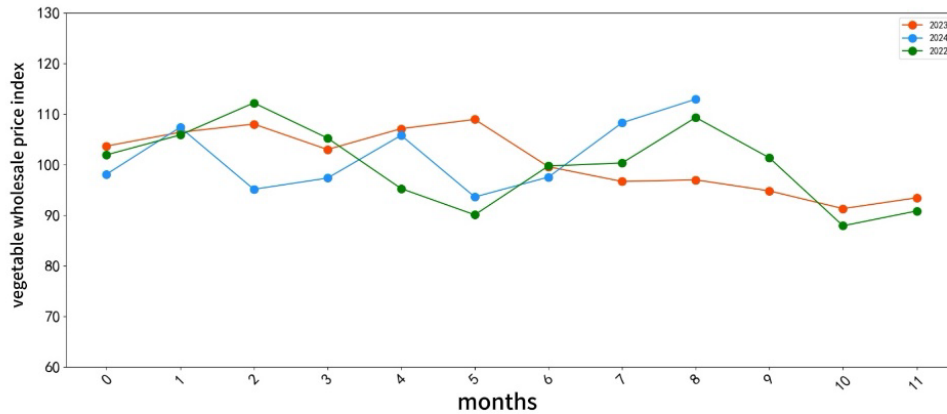


Fig.3: Vegetable wholesale price index

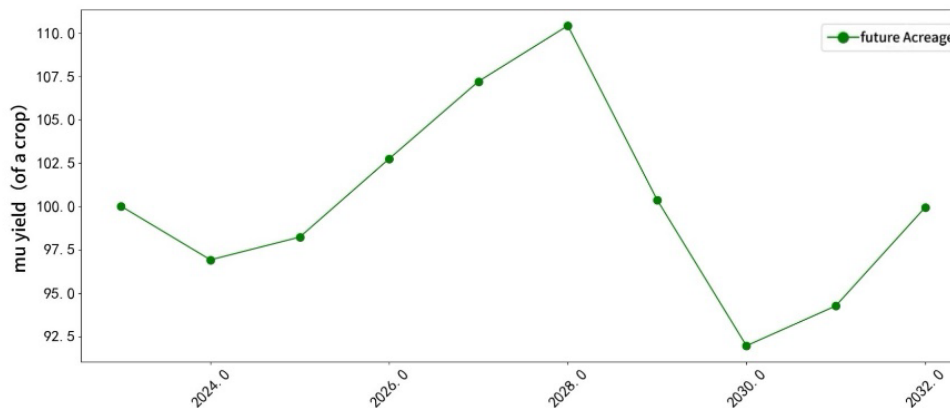


Fig.4: Future acreage

The normality test for each data is obtained as shown in the table:

Table.1. Table of normality test results

variable name	median	average	skewness	kurtosis	S-W test
Expected sales of maize	143.643	146.023	0.212	-1.248	0.749
Expected sales of wheat	136.698	140.984	0.358	-1.033	0.787
Expected sales of remaining crops	111.606	112.304	0.407	-0.551	0.897
Future acreage	10	99.965	0.494	0.084	0.494
Future planting costs	124.234	125.451	0.196	-1.065	0.895
Vegetable wholesale price index	99.58	101.32	0.152	-1.304	0.167

Observing the results in Table 1, it was found that all of them appeared $P > 0.05$, therefore, it was considered that the pattern of change of the above variables obeyed the normal distribution.

3.2.2 Introduction of random variables and ranges of values

Definitions of various types of random variables are introduced based on the assumption of normal distribution.

(1) Changes in sales volume

Sales of wheat and maize are expected to grow at 5-10 per cent per annum:

$$ES_i(y) = ES_i(2023) * (1 + es_{i,y}), es_{i,y} \in [0.05, 0.10], \forall i \in \{L_6, L_7\} \quad (14)$$

For the remaining crops, the expected sales volume varies within ± 5 per cent.

$$ES_i(y) = ES_i(2023) * (1 + es_{i,y}), es_{i,y} \in [-0.05, 0.05], \forall i \notin \{L_6, L_7\} \quad (15)$$

(2) Change in yield per acre

A variation of ± 10 per cent per year in the acreage yield of a crop can be defined as:

$$S_i(y) = S_i(2023) * (1 + s_{i,y}), s_{i,y} \in [-0.1, 0.1] \quad (16)$$

(3) Changes in planting costs

The cost of growing crops increases by an average of 5 per cent per year, as defined by market conditions:

$$C_i(y) = C_i(2023) * (1 + 0.05)^{y-2023}, \forall i \quad (17)$$

(4) Sales price

According to the changes in the sales prices of crops, a total of four categories were analysed, divided into grain, vegetables, edible mushrooms and morel mushrooms, respectively, with the following range of changes:

The sales price of food crops is basically stable and unchanged:

$$P_i(y) = P_i(2023), \forall i \text{ with } N_i = N_1 \cup N_2 \quad (18)$$

Sales prices of vegetable crops increased by 5 per cent per year:

$$P_i(y) = P_i(2023) \cdot (1 + 0.05)^{y-2023}, \forall i \text{ with } N_i = N_3 \cup N_4 \quad (19)$$

Sales prices of edible mushroom crops have fallen by 1 per cent to 5 per cent annually.

$$P_i(y) = P_i(2023) \cdot (1 - r_{i,y}), r_{i,y} \in [0.01, 0.05], \forall i \text{ with } N_i = N_5 \quad (20)$$

The annual decline in the price of morel mushrooms is set at 5 per cent.

$$P_{41}(y) = P_{41}(2023) * (1 - 0.05)^{y-2023} \quad (21)$$

3.2.3 Model results

Using python programming, the optimal crop planting strategy model is solved, where the number of particle swarm iterations is 1000 and the number of Monte Carlo simulations is 100, and the economic returns for the next seven years are solved as in Table 2.

Table.2. The economic returns for the next seven years

Year	2024	2025	2026	2027	2028	2029	2030
Economic return/ ¥10,000	5263.19	5921.81	6132.33	6212.02	6626.25	6542.24	6944.92

3.3 Optimisation models based on relevance and substitutability considerations

3.3.1 K-means cluster analysis

Table.3. Clustering results table

categories	frequency	Percentage/%	cluster naming
1	13	31.7	moderate demand
2	4	9.8	high demand
3	24	58.5	low demand

Observing Table 3, based on the main criteria for judging the clusters, the three types of crops can be classified into three categories according to the clusters: high demand, moderate demand, and low demand. Subsequently, data can be extracted from each of these three types of clusters for correlation modelling.

3.3.2 Least squares based sales volume model forecasting

We selected representative crops for linear regression analyses from three categories of moderate, high, and low demand: pumpkin, wheat, and white mushroom.

Table.4. Linear regression results table

Crop	Pumpkin (moderate demand)	Wheat (high demand)	White Ling Mushroom (low demand)
P	<0.001	<0.001	<0.001
R ²	0.995	0.998	0.997

Observing Table 4, the P-values of the three crops are less than 0.05, observing R², it is found that

they are all more than 0.9, indicating that the model works well. The final expression of the function on the expected sales volume was obtained as follows:

$$ES_{12} = 6.008 - 6.123C_{12} + 6.562P_{12} \quad (22)$$

Where, ES is the expected sales volume, C is the planting cost, P is the selling unit price, and 12, 6, and 40 represent pumpkin, wheat, and white mushroom, respectively.

3.3.3 Model results

In order to analyse the practical implications of correlation and substitutability factors, this paper compares the expected sales volumes with and without considering correlation. The comparison in Figure 5 reveals that the planting strategy after considering the correlation between crops can further enhance the production efficiency.

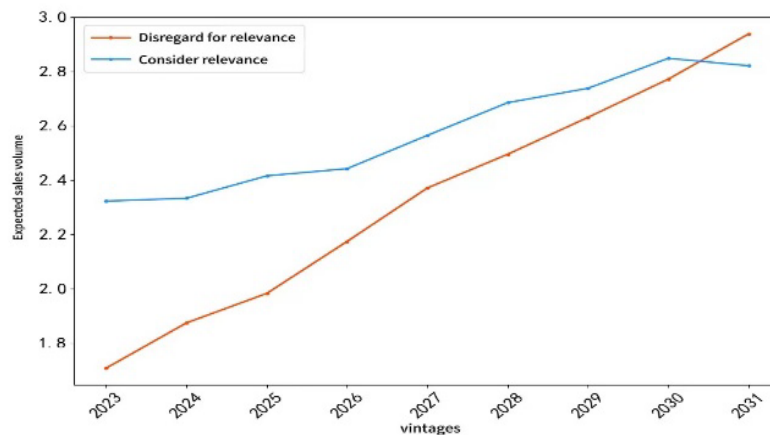


Figure.5: Comparison of expected sales volume

3.4 Analysis and discussion of results

This research has far-reaching practical significance. The proposed optimal planting strategy based on particle swarm algorithm, Monte Carlo simulation, K-mean cluster analysis and least squares method provides a comprehensive framework for agricultural planning. Firstly, the Monte Carlo simulation results shown in Fig. 1 and Table 2 show the fluctuations in expected sales and planting costs over the next seven years. These fluctuations are critical for farmers to plan their crop cultivation. By incorporating uncertainty into the model, the study provides a more realistic representation of the agricultural market and allows farmers to anticipate potential challenges and opportunities. As shown in Table 2, the upward trend in economic returns indicates that the proposed cropping strategy is sustainable and profitable in the long run.

Secondly, the results of the cluster analysis shown in Table 3 classified the crops into high, medium and low demand categories. By identifying crops with similar demand patterns, farmers can allocate their resources more efficiently and concentrate on high-demand crops to maximise profits. The linear regression models developed for each demand category further refine this strategy by allowing farmers to forecast expected sales based on growing costs and selling prices. This forecasting ability is invaluable for agricultural planning and decision making.

In addition, the flexibility of the cropping strategy was improved by introducing substitutability between crops through K-means cluster analysis and linear regression analysis. By considering correlations between crops, farmers can adjust their cropping plans to accommodate market changes and reduce risk. For example, if the demand for a particular crop declines, farmers can switch to a related crop with higher demand to ensure income stability.

4. Conclusions

In this paper, the optimisation of crop structure is studied taking into account the uncertainties of future expected sales volume, mu yield, planting cost and selling price, utilising particle swarm algorithm, Monte Carlo method, K-means clustering and least squares method. Firstly, the optimal planting strategy was

formulated by particle swarm algorithm without considering the uncertainty; secondly, the uncertainty of future expected sales volume, mu yield, planting cost and sales price was further considered, and a dynamic planning model was established by Monte Carlo simulation and particle swarm algorithm to obtain the average profit of all years to be raised to 62,346,800 Yuan; finally, the effect of substitutability between crops was introduced, and the effect of substitution was analysed using K-means cluster analysis to explore the correlation between crops, and a linear regression model of expected sales volume was established by the least squares method. Compared with traditional linear programming models, our proposed method can significantly improve planting efficiency and profitability. However, our study still has limitations, such as Monte Carlo simulations may have local optima. For future research, we suggest exploring more advanced optimisation algorithms to further improve the robustness and accuracy of the model.

References

- [1] BI Chunning, SUN Bin, XUE Jiangchun, LI Pengming, WU Shuang. *Optimization of Crop Planting Structure in Estuary Area of Lower Yellow River Basin under Water Resources Constraint [J]*. *Yellow River*, 2024, 46(12): 117-122.
- [2] ZHANG P F, XIAO M L, ZHANG Z R, et al. *Optimization and adjustment of agricultural planting structure in Shanxi Province based on water—energy—food nexus [J]*. *Water Saving Irrigation*, 2023(1): 55-63.
- [3] GUO Ping, SHAN Baoying, GUO Shanshan. *A Multi-Objective programming model for irrigation water and planting structure optimization based on pareto Set [J]*. *Journal of Tianjin University (Science and Technology)*, 2019, 52(10): 1008-1016.
- [4] ZHANG H, ZHAO S W, QIAN J, et al. *Optimization of irrigation farming structure in tropical savanna climatic based on stochastic dynamic programming [J]*. *Water Saving Irrigation*, 2024(10): 15-21.
- [5] DONG Chen-chao, *Optimal dispatching of water resources in artesian irrigation district of Ningxia based on genetic algorithm-generative adversarial neural network mode [J]*. *Hubei Agricultural Sciences*, 2021, 60(15): 174-180, 187.
- [6] WANG Zhipeng, LIU Qunchang, DAI Wei, et al. *A review of the optimization method of planting structure based on water constraints [J]*. *Journal of Water Resources and Architectural Engineering*, 2019, 17(3): 231-235, 246.
- [7] PAN Yue, YANG Guang, XUE Liangying, et al. *Application of DSF-GWO algorithm in optimization of planting structure in Manas River irrigation areal [J]*. *Journal of drainage and irrigation machinery engineering (JDlyE)*, 2023, 41(9): 943-951.
- [8] LI Yanbin, MA Jiatong, LI Daoxi, et al. *An Improved Particle Swarming Optimization Method to Optimize Cropping Systems [J]*. *Journal of Irrigation and Drainage*, 2022, 41(1): 62-71.
- [9] ZHANG Qian, ZHANG Jianfeng, LI Tao, et al. *Improvement of particle swarm optimization and its application in agricultural water resources allocation [J]*. *Journal of drainage and irrigation machinery engineering (JDIME)*, 2020, 38(6): 637-642.