Research on Crop Planting Strategies Based on Monte Carlo Simulation and Genetic Algorithm

Kehan Li *,#, Yao Sun #, Shenyang Li #, Yuxiang Li #, Han Yang #

School of Software, Xinjiang University, Urumqi, China, 830091

* Corresponding Author Email: LIKEHAN041007@outlook.com

#These authors contributed equally.

Abstract. Sustainable agricultural development is challenged by global population growth and resource scarcity. Efficient land use and crop management are crucial for stable yields and maximising benefits from limited land. This paper aims to improve land profitability and marketability through scientific cropping plans that take into account soil nutrient cycling, crop rotation, seasonal adaptability and market dynamics. The study first developed a linear programming model to optimise crop planting in a mountainous area in northern China from 2024 to 2030. By constructing objective functions and constraints and solving them using a genetic algorithm, the maximum annual returns were 6.2 million yuan and 7.2 million yuan under two scenarios, with high model stability. Next, planting costs and selling prices were incorporated and the Monte Carlo algorithm was used to simulate changes in the indicators, further optimising the model to achieve a maximum annual return of 5.5 million yuan. Finally, taking into account crop substitution and complementarity, systematic clustering and multiple linear regression were used to derive the optimal planting configuration, increasing the maximum annual return to 5.6 million yuan. These results demonstrate the effectiveness of integrating advanced techniques to improve agricultural productivity and economic returns. Future research will focus on expanding the scope to more diverse regions and crops, integrating real-time data and advanced analytics to improve model adaptability and accuracy. In addition, incorporating external factors such as climate change and policy changes will improve the model's ability to deal with uncertainties in agricultural production. These efforts aim to support sustainable agriculture globally, ensuring higher productivity and resilience for food security and rural development. The ultimate goal is to provide policy makers and farmers with actionable insights to optimise agricultural practices in a rapidly changing world.

Keywords: Genetic Algorithm, Monte Carlo Algorithm, Systematic Cluster.

1. Introduction

Crop planting strategy research in the field of agriculture has an extremely important position, which is closely related to the improvement of crop yield and quality, and has important theoretical and practical significance for guaranteeing food security and promoting the sustainable development of agriculture. Through scientific planting strategy, it can optimise the allocation of resources and improve the efficiency of agricultural production. In recent years, this field has been developing rapidly, and scholars have used a variety of methods and techniques to conduct research and achieved many results.

In the study of crop planting strategy, Jianxin Li et al [1] constructed a linear programming model with the planting area of crops in different years and different plots in different seasons as the decision variable, and yield maximization as the goal, which provides a way of quantitative analysis for crop planting planning.Yanyang Hu [2] introduced Monte Carlo simulation, constructed a bi-objective containing sales, selling price and cost change robust optimisation model and used the improved NSGA-II algorithm for optimisation solution to make the planting strategy more robust. These studies provide new methods and ideas for crop planting strategies, but they also have limitations.

Yuan et al [3] took the actual situation in North China as an example, used a linear programming model to formulate a regional planting strategy, introduced a robust optimization model to construct an uncertainty set, incorporated an objective function to obtain the optimal planting strategy in the worst case, and also used a K-Means cluster analysis model to improve the objective function and refine the model.Cai et al [4] took a village in the mountainous area of North China as a research object, based on the 2023 crop mu yield, cost and sales price data, established a linear programming model with the objective of maximising total revenue, analysed constraints such as crop selection, continuous cropping, legume crop rotation and plot area allocation to determine the constraints of the objective function, and also computed the objective function for two scenarios, namely, product stagnation and abandonment and price reduction of 50% of the sales.Yuan [5] established an optimisation model with the objective of maximising profit. The expected sales volume, planting area, planting cost and sales price of various crops and their uncertainty risks were considered, and multiple uncertainties such as changes in market dynamics and fluctuations in climatic conditions were taken into account by analysing four scenarios, namely optimal, sub-optimal, intermediate and stochastic, and the substitution and complementary relationships among different crops were explored using K-means clustering method.

Lv Linpu et al [6] developed a robust optimisation model based on 0 - 1 integer programming for the problem of cropping strategy on arable land. They transformed the uncertain planting area problem into a crop and plot selection matching problem by introducing the demand price elasticity coefficient. Hui Honglin et al [7] considered the impacts of climate change and market fluctuations on crop planting strategies in a rural area in the mountainous region of North China as an example. They considered two market scenarios, simulated market fluctuations by Monte Carlo algorithm, and combined the greedy algorithm to find the optimal planting combinations to maximise the net returns under different market conditions.

Ma Hengtian [8] used a linear programming model combined with Monte Carlo simulation with the aim of maximising net returns under limited land and climate conditions. The study considered two scenarios: excess production is either wasted or sold at low prices. For the planting scheme under stable market conditions, a goal planning model was established to optimise the planting plan for 2024 - 2030 by maximising the annual planting income and combining the constraints of various economic indicators.

This paper focuses on how to rationally allocate 41 crops on 54 different plots in a mountainous region of North China to maximise the overall income. Specifically, the following three problems need to be solved: firstly, based on the data in 2023, how to construct a linear programming model to determine the optimal planting layout strategy to maximise the revenue; secondly, how to use the Monte Carlo algorithm to find the optimal planting plan under uncertain environments, taking into account the variations of planting costs and selling unit prices; and lastly, how to take into account the substitutability and complementarities of crops, as well as the Finally, how to further optimise the planting allocation scheme by considering the correlation between the expected sales volume, sales price and planting cost. The solution of these problems is of great significance for improving the economic efficiency of agricultural production and the efficiency of resource use.

2. Method

2.1. Genetic algorithm

Genetic algorithms are search algorithms that simulate natural selection and genetic variation to solve optimisation problems[9]. In this paper, the genetic algorithm is used to solve the optimal solution of a linear programming model. The steps are as follows:

Step1. Initial population: Randomly generated initial population, with each individual representing a planting scheme.

Step2. Fitness function: The fitness function is used to assess the strengths and weaknesses of each individual, i.e. to calculate the value of the objective function:

$$\max\sum_{i=1}^{34}\sum_{j=1}^{54}\sum\nolimits_{k=1}^{7}\left(W_i\times b_{ih}-C_i\times x_{ijk}a_{ih}\right)+\sum_{i=17}^{41}\sum_{j=55}^{82}\sum\nolimits_{k=1}^{7}\left(W_i\times b_{ih}-C_i\times x_{ijk}a_{ih}\right) \quad (1)$$

Step3. Selection operation: A roulette selection method was used to select individuals into the next generation based on fitness values.

Step4. Crossover operation: A single-point crossover was used, in which two parents were randomly selected and some of their genes were exchanged to generate offspring.

Step5. Mutation operation: An individual is randomly selected and some of its genes are changed to introduce a new solution.

Step6. Reward and punishment mechanisms: Reward or punish individuals for satisfying or violating constraints and adjusting fitness values.

2.2. Monte Carlo simulation

Monte Carlo simulation is a method of estimating uncertainty by random sampling[10]. In this paper, Monte Carlo simulation is used to deal with the uncertainty of indicators such as planting cost and unit selling price in the following steps:

step1. Generate random rates of change for indicators: Randomly generate the rate of change of each crop indicator, such as planting cost, unit selling price, etc.

step2. Generate total random yields etc.: Based on generated rates of change, calculate total yields, costs and revenues for each crop.

step3. Solve the optimisation model: Substituting the generated variables into a linear programming model and solving for the optimal planting scheme using genetic algorithms.

3. Results

3.1. Integrated model construction and constraints

3.1.1. Construction of the objective function

Data source: https://www.mcm.edu.cn/

In the data cleaning session, missing and duplicate values were eliminated and outliers were detected using box-and-line plots. After analysis, the data were more complete and without obvious anomalies. Subsequently, the area and yield information of the same type of crop in the same season were merged to reduce redundancy and improve modeling efficiency.

In order to reduce the complexity of the model, improve the computational efficiency, ensure that the parameters are set accurately, maximize the return of crops as much as possible, and solve the optimal planting scheme, it can be seen:

Revenue = Expected sales volume x unit sales price - Yield x cost of planting

$$\text{Cereals: } \sum_{i=1}^{15}\sum_{j=1}^{82}\sum_{k=1}^{7}\left(W_i \times b_{ih} - C_i \times x_{ijk}a_{ih}\right) \tag{2}$$

$$\text{Remaining crops: } \sum_{i=16}^{41}\sum_{j=1}^{82}\sum_{k=1}^{7}\left(W_i \times b_{ih} - C_i \times x_{ijk}a_{ih}\right) \tag{3}$$

3.1.2. Linear programming constraints

Point1.Lot size restrictions

The planting area should not be too small resulting in less utilization of each piece of land.

$$\sum_{i=1}^{41} x_{ijk} = b_j, \qquad k = 1,2,3 \cdots 7, \qquad j = 1,2 \cdots 54,55 \cdots 82 \tag{4}$$

Point 2.Requirements for growing legumes

Each plot is required to grow a legume crop at least once in three years, i.e., starting from 2024 to 2030, from which any three years each plot will necessarily have at least one season of growing a legume crop to be used as a fertilizer booster.

$$\sum_{i=1}^{5}\sum_{k=t}^{t+2} y_{ijk} + \sum_{i=17}^{19}\sum_{k=t}^{t+2} y_{ijk} \geq 1 \tag{5}$$

$$\mathrm{t} = 1,2,3,4,5 \ \ \mathrm{j} = 1,2 \cdots 82$$

Point 3.Heavy cropping restrictions

Since successive plantings of the same crop on the same plot would result in reduced yields, none of these crops could be planted the following year.

$$\begin{cases} y_{ijk} + y_{ijk-1} = 1 \\ \quad i = 1,2 \cdots 41 \\ \quad j = 1,2 \cdots 82 \end{cases} \tag{6}$$

Point 4.Planting Requirements

Planting sites for the same species should not be too dispersed.

$$\sum_{j=1}^{82} y_{ijk} \leq M_i \ \ i = 1,2 \cdots 34 \ \ k = 1,2,3 \cdots 7 \tag{7}$$

$$\text{Dispersion} \ \ \gamma_{ijk} = \frac{\sum_j |x_{ijk} - \bar{x}_{jk}|}{\sum_j x_{ijk}} \tag{8}$$

Of these, the molecular$\sum_j |x_{ijk} - \bar{x}_{jk}|$ Indicates the sum of the deviations of the planted area of each crop from the mean, emphasizing the degree of deviation from the distribution of planting; denominator of a fraction$\sum_j x_{ijk}$ Indicates the total planted area, which is used for normalization.

Point 5.Rice Single and Dual Season Specificity.

Due to the particularity that rice can only be grown in the first season of the watered land, it is known that the total area of the watered land is equal to the total number of areas of rice + the number of areas in the second season of the watered land.

$$\sum_{j=1}^{54} x_{16jk} + \sum_{i=1}^{41}\sum_{j=55}^{62} x_{ijk} = S \tag{9}$$

3.1.3. Risk-based constraints

point 1. Five percent annual increase in planting costs

$$= 105\% C_i^t \ \ i = 1,2,\cdots,41 \ \ t = 2023,2024,\cdots,2029 \tag{10}$$

point 2. Vegetable sales unit price increased by 5 percent annually

$$b_{ih}^{t+1} = 105\% b_{ih}^{t} \tag{11}$$

$$\mathrm{i} = 17,18,\cdots,37 \ \ \mathrm{h} = 1,2,\cdots,6$$

point 3. The unit price of morel mushroom sales fell by five percent annually

$$b_{41}^{t+1} = 95\% b_{41}^{t} \tag{12}$$

point 4. Cultivation risk quantification

Indicator I: The coefficient of production volatility, which measures the relative degree of change in production.

$$\text{Coefficient of fluctuation of production} \ \ CV_Y = \frac{\sqrt{\frac{\sum_{k-1}^{n}(P_k - \bar{P})^2}{n-1}}}{\frac{\sum_{k-1}^{n} P_k}{n}} \tag{13}$$

Indicator II: The coefficient of variation of prices, which measures the relative degree of price variability.

$$\text{Price coefficient of variation } CV_P = \frac{\sqrt{\frac{\sum_{k-1}^{n}(E_k-\bar{E})^2}{n-1}}}{\frac{\sum_{k-1}^{n}E_k}{n}} \tag{14}$$

In order to ensure that the planting risk is low, this paper requires that these two coefficients of price variability and product volatility, which can reflect the risk, be within the permitted range.

$$CV_Y \le \beta$$
$$CV_P \le \theta$$

β is the maximum yield volatility coefficient within the allowed range;

γ is the maximum price variability coefficient within the allowed range.

The following mathematical model is obtained:

$$\max \sum_{i=1}^{34}\sum_{j=1}^{54}\sum\nolimits_{k=1}^{7}\left(W_i \times b_{ih} - C_i \times x_{ijk}a_{ih}\right) + \sum_{i=17}^{48}\sum_{j=55}^{82}\sum\nolimits_{k=1}^{7}\left(W_i \times b_{ih} - C_i \times x_{ijk}a_{ih}\right) \tag{15}$$

3.2. model solution

Linear programming: Initial allocation of plots and crop types in an uncertain environment with the objective of maximizing returns.

Genetic algorithm iterative optimization: coding “plot-season-crop allocation” as an individual chromosome; setting selection, crossover and mutation operations, and integrating reward and punishment mechanisms for heavy cropping limitations, legume rotation requirements, and patch planting requirements in the fitness function.

Monte Carlo simulation: The key parameters such as “planting cost, unit price of vegetable sales, unit price of morel mushroom sales, planting risk” are randomly sampled several times, and the genetic algorithm is run under each sample to obtain a robust planting plan. The results of the model are shown in Figure 1:

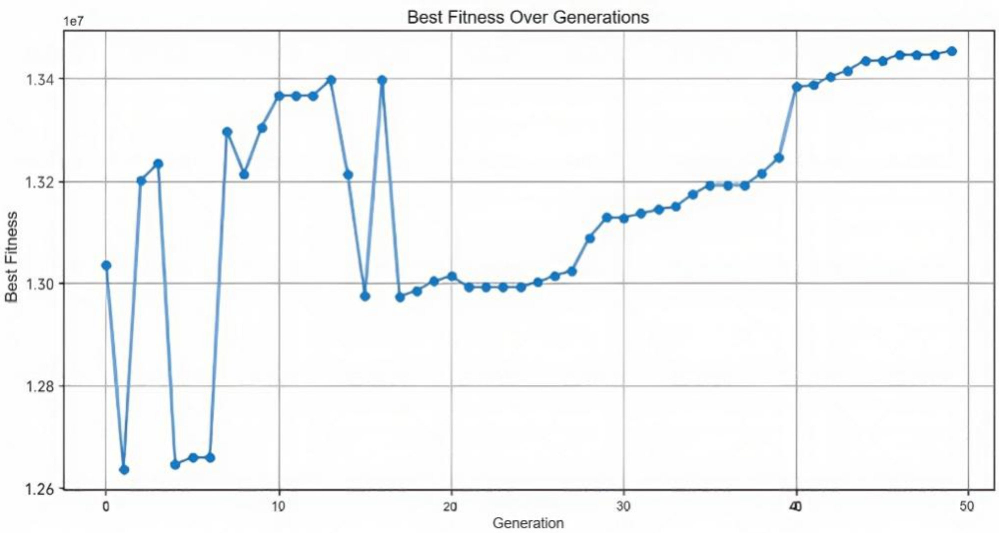

Figure 1. iterative process

The maximum return of the model basically stabilizes when the number of iterations exceeds 40. At this point, it can be seen that, in this case, the average value of the annual crop return is 6195378.94 yuan, compared with 2023, an increase of 330,000 yuan, an increase of 5%. Observing the distribution of crop planting in 2023, it can be seen that only one crop is planted on each piece of land, while the optimal planting strategy of this problem takes into account that a piece of land can be divided into dfferent crops, which leads to an increase in the average value of annual crop income.

In this paper, the rate of change is first simulated by Monte Carlo through python, and the simulated rate of change is imported into the optimization model, which is solved by genetic algorithm to obtain the maximum value of the average annual crop revenue of 5509396.57 yuan.

Next, the model was tested, and through Monte Carlo simulation, the total seven-year average return was 46,159,754.17 yuan across all simulated scenarios. This is an expected value that provides the overall level of returns.

3.3. Testing of the model

The calculated sensitivity analysis plot is shown in Fig. 2:

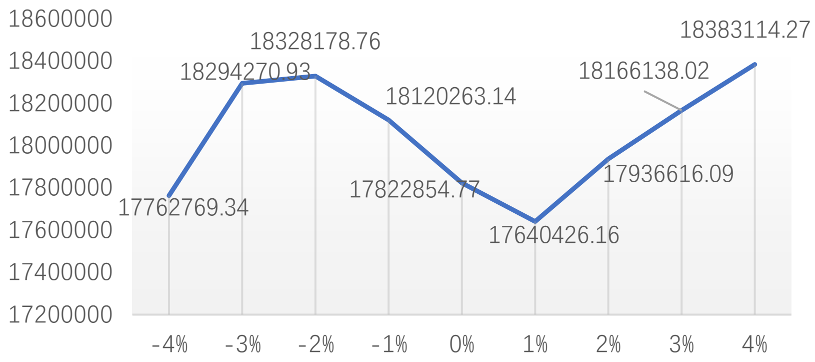

Figure 2. sensitivity analysis

The minimum area fluctuates up and down by 4% and the return fluctuates by no more than 4%, which can show that the model is more stable and passes the sensitivity test.

At this time, the standard deviation of 4,697,165.16 yuan and the confidence interval of [43833310.46 yuan, 49486197.87 yuan] indicate that we have a 95% confidence level that the true return will fall within this interval.

3.4. Substitutability and complementarity analysis

Through systematic clustering to generalize the economic benefits of similar crops into a class, this paper uses the Euclidean distance to quantify the economic benefits, for the correlation between the three indicators, the use of multiple regression analysis, to find out the relationship between the indicators, and to add the above factors into the optimization model, to solve the optimal planting plan.

Heat mapping using Euclidean distances between different crops as shown in Fig. 3

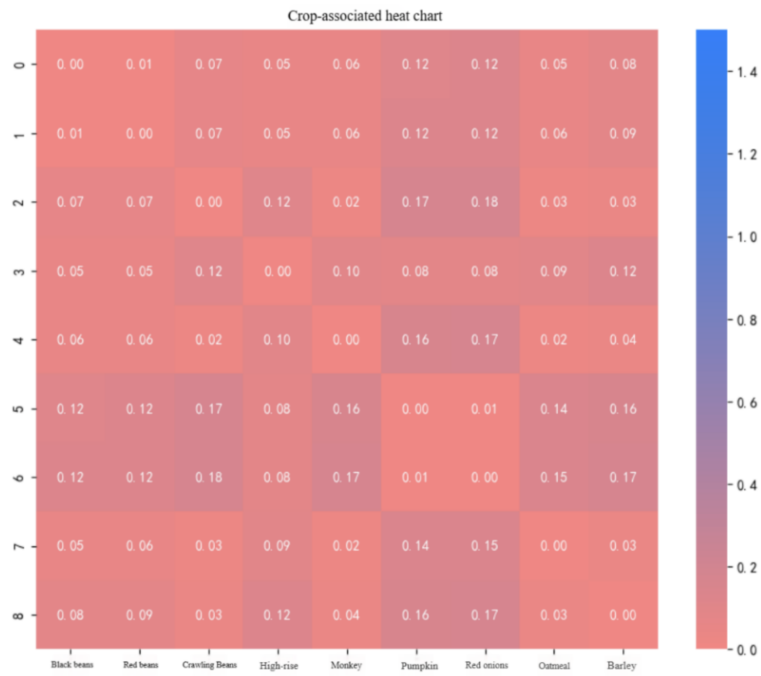

Figure 3. Heat map of European distances between crops of the second category

The Euclidean distance between different crops in this figure is small, the correlation is large, and the economic benefits are similar. And this paper clusters it into one class by systematic clustering method, which also proves that their economic benefits are similar, so it can be seen that the clustering effect of this paper is better.

The adjusted R^2 is 0.434, which can show that the independent variable can explain 43.4% of the variance of the standardized expected sales, and the F-value is 11.243, with P less than 0.01. In this paper, there exists at least one independent variable that can explain the change of the dependent variable, which proves that the model building is relatively successful. As shown in Figure 4.

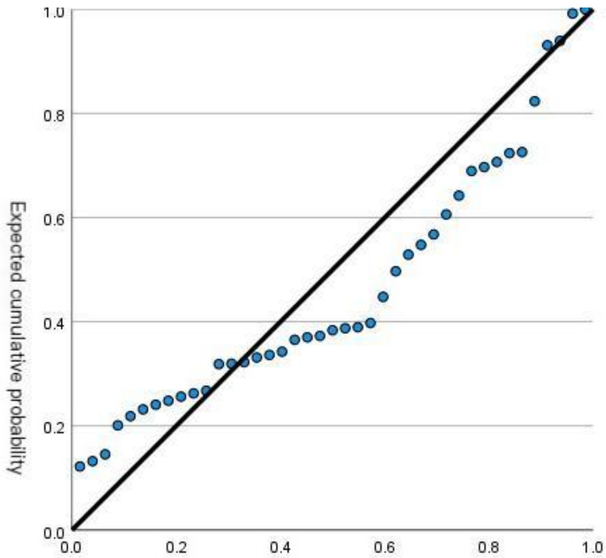

Figure 4. Normal P-P plot of standardized residuals

The histogram of the residuals shows that it satisfies the normal distribution. From the normal P-P plot, it can be seen that the points are uniformly distributed around this diagonal line, which also indicates that it satisfies the normal distribution condition. Based on the above regression equation parameter table it can be seen that:

$$y = 0.211x_1 - 0.281x_2 + 0.686x_3 \tag{16}$$

Considering the correlation between the three indicators, this paper analyzes the multiple linear regression to find the corresponding regression equation and add it to the constraint conditions.

$$\max \sum_{i=1}^{34} \sum_{j=1}^{54} \sum_{k=1}^{7} \left(W_i \times b_{ih} - C_i \times x_{ijk} a_{ih}\right) + \sum_{i=17}^{48} \sum_{j=55}^{82} \sum_{k=1}^{7} \left(W_i \times b_{ih} - C_i \times x_{ijk} a_{ih}\right) \tag{17}$$

4. Conclusion

In this paper, a planting strategy optimization scheme is designed through linear programming and uncertainty modeling, and tested with genetic algorithm and Monte Carlo simulation. With the introduction of uncertainty, the robust returns combined with Monte Carlo simulation have more practical value and can provide a decision basis for operators with different risk preferences. The substitutability between crops is associated with an average crop return value of 6,195,378.94 yuan per year, which is an increase of 330,000 yuan compared to 2023, an increase of 5%. Through Monte Carlo simulation, the total seven-year average return is 46,159,754.17 yuan in all simulated scenarios.

The complementary relationship has a significant impact on the overall efficiency and needs to be considered comprehensively in planting planning. The limitations of the study are that the data source is single and the genetic algorithm has the problem of early convergence. In the future, dynamic factors can be included, combined with deep reinforcement learning and other techniques to improve model adaptability and decision-making efficiency. The genetic algorithm sometimes suffers from premature convergence in multi-objective and multi-constraint large-scale problems, which can be further optimized by combining with other meta-heuristic algorithms. In the future, dynamic factors such as climate change, overlapping growth cycles, and multi-season crop rotation can be incorporated into the model to improve the adaptability to complex agricultural systems.

References

[1] Li J, Gao D, Qiu Z, et al. Research on Planting Planning Strategies Based on Optimization Genetic Algorithms under Multiple Constraints[C]//2024 IEEE 2nd International Conference on Electrical, Automation and Computer Engineering (ICEACE). IEEE, 2024: 1460-1465.

[2] Hu Y. Research on Planting Strategies Based on Improved Multi-Objective Genetic Algorithm and Monte Carlo Simulation[C]//2024 IEEE 2nd International Conference on Electrical, Automation and Computer Engineering (ICEACE). IEEE, 2024: 689-692.

[3] Yuan W, Yin Z, Xia L. Research on Crop Planting Strategies Based on K-means Cluster Analysis and Linear Programming[C]//2024 IEEE 2nd International Conference on Electrical, Automation and Computer Engineering (ICEACE). IEEE, 2024: 838-842.

[4] Zhidan CAI, Wenjie CHEN, Xinyi CHEN. Research on Optimization Model of Crop Planting Strategy Based on Linear Programming[J]. Transactions on Economics, Business and Management Research, 2024. DOI:10.62051/mm41rk73.

[5] Yuan W, Yin Z, Xia L. Research on Crop Planting Strategies Based on K-means Cluster Analysis and Linear Programming[C]//2024 IEEE 2nd International Conference on Electrical, Automation and Computer Engineering (ICEACE). IEEE, 2024: 838-842.

[6] Lv L, Yang H, Yang W, et al. Research on Cultivation Strategy Based on Robust Optimization Model and Genetic Algorithm[C]//2024 IEEE 2nd International Conference on Electrical, Automation and Computer Engineering (ICEACE). IEEE, 2024: 1066-1070.

[7] Honglin HUI, Hao GONG, Jiawen GAO, et al. Research on the Optimization of Crop Planting Strategies Based on a Multi-Factor Comprehensive Analysis Model[J]. Transactions on Environment, Energy and Earth Sciences, 2024. DOI:10.62051/qfst4t37.

[8] Maa H. Optimization of Crop Planting Strategies Based on Linear Programming and Monte Carlo Simulation[J]. 2024.

[9] Katoch S, Chauhan S S, Kumar V. A review on genetic algorithm: past, present, and future[J]. Multimedia tools and applications, 2021, 80: 8091-8126.

[10] Wealer B, Bauer S, Hirschhausen C V, et al. Investing into third generation nuclear power plants-Review of recent trends and analysis of future investments using Monte Carlo Simulation[J]. Renewable and Sustainable Energy Reviews, 2021, 143: 110836.