

Research on Crop Planting Strategies Based on Monte Carlo-Genetic Algorithm

Renze Wei^{1,#}, Boyue Zhang^{1,#}, Zhengjie Huang²

¹College of Information Engineering, Northwest A & F University, Yangling, 712100, China

²College of Water Resources and Architectural Engineering, Northwest A & F University, Yangling, 712100, China

[#]These authors contributed equally.

Abstract: The issue of global food security is becoming increasingly severe, and sustainable agricultural development has emerged as a focal point of global attention. This study takes rural areas in the mountainous regions of North China as the research object, investigating how to optimize planting schemes for terraces, sloped fields, irrigated lands, and greenhouses under cold climatic conditions and complex arable land environments through scientific crop rotation and intercropping strategies. The aim is to enhance farmland utilization efficiency, reduce environmental and market risks, and promote the sustainable development of rural economies. This research employs Monte Carlo simulations to estimate the expected sales volumes of crops. Based on two scenarios of handling surplus crop production beyond the expected sales volume—Scenario 1: disposal as waste; Scenario 2: surplus sold at a 50% discount—a linear programming model is constructed to optimize crop planting areas and types. The model is solved using a genetic algorithm. Comparative analysis of planting strategies under the two scenarios yields profitability forecasts for the period from 2024 to 2030. Under the surplus production scenario, the seven-year profit for Scenario 1 (disposal as waste) is 7,851,693 RMB, while for Scenario 2 (selling surplus at a 50% discount), the profit is 23,868,146 RMB.

Keywords: Monte Carlo Simulation, Linear Programming, Genetic Algorithm, Crop Planting Scheme

1. Introduction

In the context of agricultural modernization and sustainable development, planting strategies in mountainous rural areas are no longer confined to mere production functions but also strive to balance the efficient utilization of land resources, environmental protection, and economic benefits. By optimizing the selection of land and crops, it is possible to increase yields, enhance the long-term productivity of land, mitigate risks, and promote the sustainable development of rural economies.

Zhang Aihua et al. [1] developed a winter wheat irrigation yield model based on a genetic algorithm and optimized fertilization decision-making strategies to enhance wheat yield and irrigation efficiency. Case studies demonstrated that this technique significantly improved the net value of wheat, achieving high-yield and high-quality cultivation with promising application potential and research significance. Bi et al. [2] proposed a genetic algorithm (GA)-assisted deep learning method that combines global and local search to optimize crop yield prediction models, thereby overcoming the issues of local optima and gradient vanishing. The study demonstrated that this method outperformed traditional gradient-based approaches in terms of convergence speed and prediction accuracy.

Building upon the aforementioned studies, this paper takes a rural area in the mountainous regions of North China as an example and constructs a linear programming mathematical model. Combining Monte Carlo simulation-based expected sales volumes, the model optimizes the planting structure and area configuration of crops under two scenarios: surplus as waste and surplus sold at a discount. The model incorporates constraints such as land types, crop rotation, and dispersed planting. By solving the model with a genetic algorithm, optimal planting strategies are proposed for the period 2024 to 2030, aiming to improve agricultural production efficiency.

2. Establishment of the optimization model

The data used in this study are open-source and originate from agricultural planting data of a rural area in the mountainous regions of North China. The data were obtained from <https://www.mcm.edu.cn/>.

2.1 Data Preprocessing

The yield per acre, planting cost, and unit selling price were extracted from Attachment 2 and preprocessed to identify and handle outliers. The data were processed in Python using the 3σ principle, and an outlier detection plot shown in Figure 1 was generated.

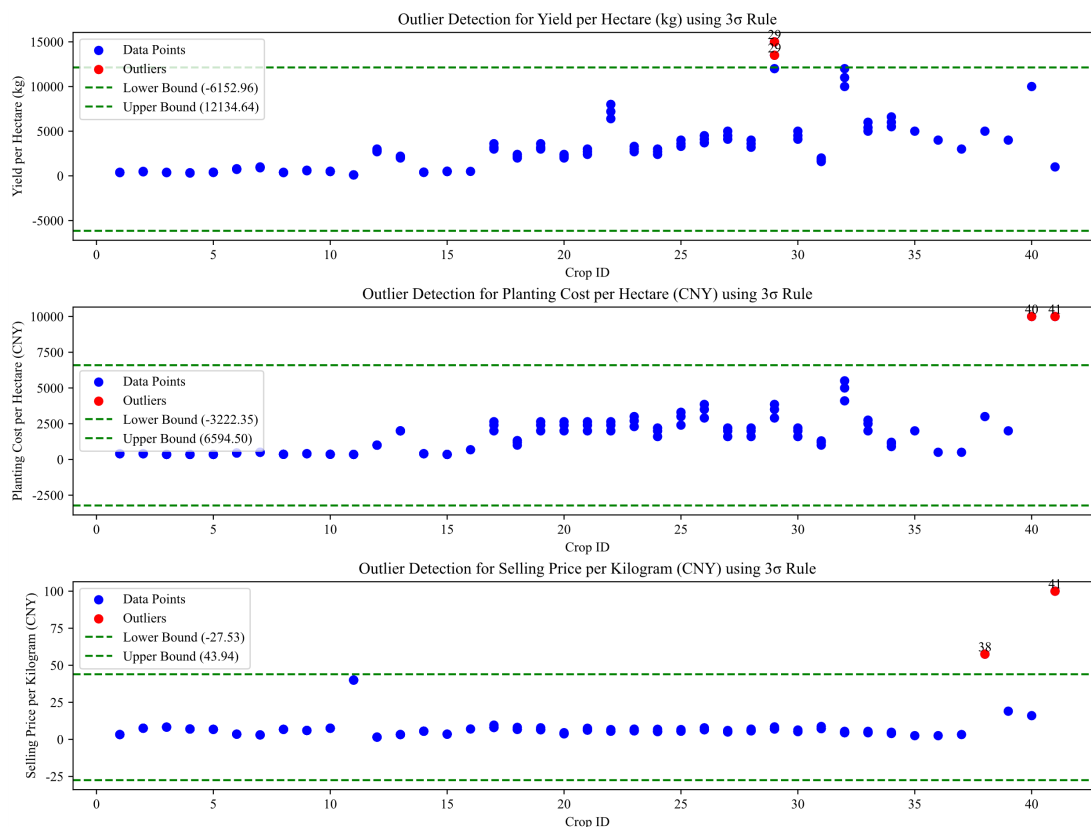


Figure 1. Outlier Detection Diagram

From the outlier detection diagram, it can be observed that the yield per acre for crop number 29 is anomalous. Additionally, the planting costs for crop numbers 40 and 41, as well as the selling prices for crop numbers 38 and 41, exhibit anomalies. The anomalous data are presented in Table 1.

Table.1. Details of Abnormal Crops

Crop Number	Crop Name	Yield per Acre	Planting Cost	Selling Price
29	Cucumber	12000	2900	7
38	Elm Mushroom	5000	3000	57.5
40	White Mushroom	10000	10000	16
41	Morchella	1000	10000	100

Through practical analysis, it can be observed that cucumbers, as a high-yield crop, can achieve a significantly high yield per acre under optimal cultivation conditions, which may lead to elevated production data. Therefore, although such data might appear anomalous, they align with the characteristics of cucumbers as a crop. Similarly, the golden oyster mushroom, being a premium edible fungus, inherently commands a high market price, resulting in a significantly higher unit sales price compared to ordinary crops. This price disparity is not anomalous but rather reflects the high economic value of golden oyster mushrooms.

In contrast, the cultivation of king oyster mushrooms incurs relatively high costs, particularly in modern greenhouse environments where expensive facilities and management fees are commonplace. These elevated cultivation costs are not unusual. Moreover, the sales price of king oyster mushrooms aligns with their high production costs, effectively covering these expenditures. Morel mushrooms, a rare and valuable edible fungus, have extremely high cultivation costs and market prices. Due to the significant difficulty in their cultivation and the robust market demand, both their production costs and sales prices far exceed those of other crops.

As such, the data flagged as anomalies are, in fact, reasonable and reflect the actual characteristics of these crops. Therefore, no outlier processing is performed on this data.

Since the data sources did not provide the expected sales volumes of various crops, this study estimated the expected sales volumes by multiplying the yield of each crop by 0.8, followed by Monte Carlo simulation [3]. Table 2 presents the simulated results for selected crops.

Table 2. The expected sales volumes of different crops.

Crop Name	2023	2024	2025	2026	2027	2028	2029	2030
Soybean	31954.72	32047.08	32042.19	31979.33	31856.25	32108.43	32111.3	32124.32
Black Bean	40116.73	40151.54	40094.57	39908.14	39933.74	39909.92	40002.34	39937.94
Red Bean	32045.05	31991.23	32286.06	32054.92	32104.84	32214.88	31914.89	31883.08
Mung Bean	27908.63	28026.26	28045.82	27881.8	27990.51	27928.72	28057.95	28027.27

2.2 Definition of Decision Variables

To address Scenario 1, an optimized planting scheme is proposed to maximize rural crop profits while considering the varying demands of different crops and the constraints of specific plots. The model accounts for two distinct situations: (1) surplus yield exceeding the expected sales volume leads to unsold waste, and (2) the surplus can be sold at half the original price.

The decision variable $x_{i,j,k,t}$ is defined as follows: $i \in \{1,2, \dots, N\}$ represents the i -th plot, $j \in \{1,2, \dots, M\}$ represents the j -th crop type, $k \in \{1,2\}$, indicates the season (where $k = 1$ corresponds to the first growing season and $k = 2$ to the second growing season), and t denotes the year.

$$x_{i,j,k,t} = \begin{cases} 1 & \text{The } i\text{-th plot is planted with crop } j \\ 0 & \text{The } i\text{-th plot is not planted with crop } j \end{cases} \quad (1)$$

2.3 Establishing the Objective Function

To determine the optimal planting scheme that maximizes the total profit of crop cultivation from 2024 to 2030, this study establishes a linear programming model [4], which considers two scenarios: unsold waste and discounted sales.

For Scenario 1, where the portion of yield exceeding the expected sales volume cannot be sold and results in waste, the objective function is defined as shown in Equation (2):

$$\max Z = \sum_{t \in T} \sum_{i \in I} \sum_{j \in J} (\min(y_{j,k,t}, D_j) \times P_j - A_{i,j,t} \times C_j) \quad (2)$$

Where $A_{i,j,t}$ represents the area of the i -th plot planted with crop j in year t , and Y_j denotes the yield per unit area of crop j . P_j is the selling price of crop j , and C_j is the planting cost of crop j . The term $\min(y_{j,k,t}, D_j)$ indicates the smaller value between the yield of crop j and the market demand, representing the actual sales volume.

For Scenario 2, where the surplus yield exceeding the expected sales volume is sold at a 50% discount, the objective function is defined as shown in Equation (3) to Equation (5):

$$\max Z = \sum_{t \in T} \sum_{i \in I} \sum_{j \in J} (\min(y_{j,k,t}, D_j) \times P_j + \max(y_{j,k,t} - D_j, 0) \times P'_j - A_{i,j,t} \times C_j) \quad (3)$$

$$y_{j,k,t} = \sum_i x_{i,j,k,t} \times Y_j \quad (4)$$

$$P'_j = 0.5 \times P_j \quad (5)$$

$\min(y_{j,k,t}, D_j)$ represents the actual sales volume of crop j grown on the i -th plot during the k -th season, which is the smaller value between the total yield and market demand.

$\max(y_{j,k,t} - D_j, 0)$ denotes the surplus yield of crop j that exceeds the expected sales volume.

2.4 Constraints

(1) Plot Area Constraint

The total planting area of the i -th plot in any given year cannot exceed its available area L_i . This constraint is expressed as shown in Equation (6):

$$\sum_{j \in J} A_{i,j,t} \leq L_i \quad \forall i \in I, \forall t \in T \quad (6)$$

(2) Crop Yield and Sales Constraint

The actual sales volume of each crop cannot exceed its total yield. This constraint is expressed as shown in Equation (7):

$$\sum_{i \in I} A_{i,j,t} \times Y_j \geq D_j \quad \forall j \in J, \forall t \in T \quad (7)$$

(3) Plot Type Constraint

Different types of plots are restricted to specific crops. Flat dryland, terraced fields, and slopes can only grow one season of grain crops; irrigated land can grow either one season of rice or two seasons of vegetables; ordinary greenhouses can grow one season of vegetables and one season of edible fungi; smart greenhouses can grow two seasons of vegetables. If crop j is not suitable for planting on plot type R_i , the planting area of crop j on the plot must be 0, as expressed in Equation (8):

$$A_{i,j,t} = 0 \quad (8)$$

Additional planting constraints are applied based on different plot types. Flat dryland, terraced fields, and slopes can only grow one season of grain crops. For these plot types $R_i \in \{\text{Flat Dryland, Terraced Fields, Slopes}\}$, only one season of grain crops is allowed, as expressed in Equation (9).

$$\sum_{j \in \text{Grain}} x_{i,j,1,t} = 1 \quad R_i \in \{\text{Flat Dryland, Terraced Fields, Slopes}\} \quad (9)$$

Additionally, planting in the second season is prohibited for these plot types, as expressed in Equation (10):

$$\sum_{j=1}^M x_{i,j,2,t} = 0 \quad R_i \in \{\text{Flat Dryland, Terraced Fields, Slopes}\} \quad (10)$$

Irrigated land can either grow one season of rice or two seasons of vegetables. For irrigated land, one season of rice or two seasons of vegetables must be planted, as expressed in Equation (11):

$$\sum_{j \in \text{Rice}} x_{i,j,1,t} + \sum_{j \in \text{Two-Season Vegetables}} x_{i,j,1,t} = 1 \quad R_i = \text{Irrigated Land} \quad (11)$$

Ordinary greenhouses must grow one season of vegetables and one season of edible fungi. For ordinary greenhouses, one season of vegetables is required, as expressed in Equation (10), and one season of edible fungi is required, as expressed in Equation (12):

$$\sum_{j \in \text{One-Season Vegetables}} x_{i,j,1,t} = 1 \quad R_i = \text{Ordinary Greenhouse} \quad (12)$$

$$\sum_{j \in \text{Edible Fungi}} x_{i,j,2,t} = 1 \quad R_i = \text{Ordinary Greenhouse} \quad (13)$$

Smart greenhouses are suitable for planting two seasons of vegetables each year. This is expressed in Equations (14) and (15):

$$\sum_{j \in \text{Two-Season Vegetables}} x_{i,j,1,t} = 1 \quad R_i = \text{Smart Greenhouse} \quad (14)$$

$$\sum_{j \in \text{Two-Season Vegetables}} x_{i,j,2,t} = 1 \quad R_i = \text{Smart Greenhouse} \quad (15)$$

(4) Crop Rotation Constraint

To prevent excessive depletion of soil nutrients, the same type of crop cannot be planted consecutively on the same plot. This constraint is expressed as shown in Equation (16):

$$A_{i,j,t} \times A_{i,j,t-1} = 0 \quad \forall i \in I, \forall j \in J, \forall t \in T \quad (16)$$

(5) Legume Crop Constraint

Each plot or greenhouse must plant legume crops at least once within three years to ensure crop rotation requirements are met. This constraint is expressed as shown in Equation (17):

$$\sum_{t=t_0}^{t_0+2} \sum_{j \in \text{Legumes}} A_{i,j,t} \geq 1 \quad \forall i \in I \quad (17)$$

(6) Minimum Planting Area Constraint

To prevent scattered planting, the planting area of each crop on any plot must not be too small and must meet a threshold T_j . In this study, the threshold is set to 0.1. This constraint is expressed as shown in Equation (18):

$$A_{i,j,t} \geq 0.1 \quad \forall i \in I, \forall j \in J, \forall t \in T \quad (18)$$

2.5 Model Establishment

Firstly, based on the 2023 production data of various crops, Monte Carlo simulation was used to estimate their expected sales volumes, which were further utilized to project the expected sales volumes of different crops from 2024 to 2030. Secondly, genetic algorithms were employed, considering several constraints to optimize the crop planting plan [5]. Due to the limited land area of each plot, it is necessary to constrain the total planting area of each crop within each plot. Based on practical experience, sales volumes cannot exceed production, requiring constraints on their relationship. Furthermore, since not all plots are suitable for all crops, constraints were established to define the types of crops that can be planted on each plot. To preserve soil nutrients, constraints were set to prevent continuous cropping. Additionally, the rhizobia in leguminous crops aid nitrogen fixation, improve soil composition, and promote the yield of other crops, making it essential to introduce constraints to retain this effect. To facilitate farming operations, reducing the dispersion of crop planting across plots is advisable, which requires the establishment of planting area threshold constraints. These measures collectively optimize the crop planting strategy for the period from 2024 to 2030.

For Scenario 1, the portion exceeding the expected sales volume is considered waste, resulting in the final optimization model as shown in Equation (19):

$$\max Z = \sum_{t \in T} \sum_{i \in I} \sum_{j \in J} (\min(y_{j,k,t}, D_j) \times P_j - A_{i,j,t} \times C_j) \quad (19)$$

For Scenario 2, the portion exceeding the expected sales volume is sold at a 50% discount, resulting in the final optimization model as shown in Equation (20):

$$\max Z = \sum_{t \in T} \sum_{i \in I} \sum_{j \in J} (\min(y_{j,k,t}, D_j) \times P_j + \max(y_{j,k,t} - D_j, 0) \times P'_j - A_{i,j,t} \times C_j) \quad (20)$$

Where:

$$y_{j,k,t} = \sum_i x_{i,j,k,t} \times Y_j \tag{21}$$

$$P'_j = 0.5 \times P_j \tag{22}$$

$$s. t \left\{ \begin{array}{l} \sum_{j \in J} A_{i,j,t} \leq L_i \quad \forall i \in I, \forall t \in T \\ \sum_{i \in I} A_{i,j,t} \times Y_j \geq D_j \quad \forall j \in J, \forall t \in T \\ A_{i,j,t} = 0 \\ \sum_{j \in \text{Grain}} x_{i,j,1,t} = 1 \quad R_i \in \{\text{Flat Dryland, Terraced Fields, Slopes}\} \\ \sum_{j=1}^M x_{i,j,2,t} = 0 \quad R_i \in \{\text{Flat Dryland, Terraced Fields, Slopes}\} \\ \sum_{j \in \text{Rice}} x_{i,j,1,t} + \sum_{j \in \text{Two-Season Vegetables}} x_{i,j,1,t} = 1 \quad R_i = \text{Irrigated Land} \\ \sum_{j \in \text{One-Season Vegetables}} x_{i,j,1,t} = 1 \quad R_i = \text{Ordinary Greenhouse} \\ \sum_{j \in \text{Edible Fungi}} x_{i,j,2,t} = 1 \quad R_i = \text{Ordinary Greenhouse} \\ \sum_{j \in \text{Two-Season Vegetables}} x_{i,j,1,t} = 1 \quad R_i = \text{Smart Greenhouse} \\ \sum_{j \in \text{Two-Season Vegetables}} x_{i,j,2,t} = 1 \quad R_i = \text{Smart Greenhouse} \\ A_{i,j,t} \times A_{i,j,t-1} = 0 \quad \forall i \in I, \forall j \in J, \forall t \in T \\ \sum_{t=t_0}^{t_0+2} \sum_{j \in \text{Legumes}} A_{i,j,t} \geq 1 \quad \forall i \in I \\ A_{i,j,t} \geq 0.1 \quad \forall i \in I, \forall j \in J, \forall t \in T \end{array} \right. \tag{23}$$

3. Results

After constructing the linear programming model, this study employs a genetic algorithm to solve the optimization problem. The initial population is defined as $X = [x_1, x_2, \dots, x_{7 \times N}]$, where N represents the number of plots. The algorithm parameters are set as follows: maximum iteration count $T = 7$, crossover probability $P_c = 0.9$, mutation probability $P_m = 0.1$, and generation gap $GAP = 0.7$. The algorithm steps are outlined as follows:

- Step 1: Generate the initial population with a size of $7 \times N$.
- Step 2: Calculate the fitness of each individual.
- Step 3: Perform tournament selection based on fitness to retain the best individuals.
- Step 4: Use single-point crossover to produce new offspring.
- Step 5: Apply random mutation to some individuals to increase population diversity.
- Step 6: Evaluate the fitness of the new offspring.
- Step 7: If the termination condition is met, output the optimal individual; otherwise, return to Step 3 and continue iterating.

Following the steps outlined above, the problem was solved using MATLAB, where the fitness function corresponds to the objective function of each specific scenario. Finally, the genetic algorithm fitness curve for Scenario 1 was plotted, as shown in Figure 2.

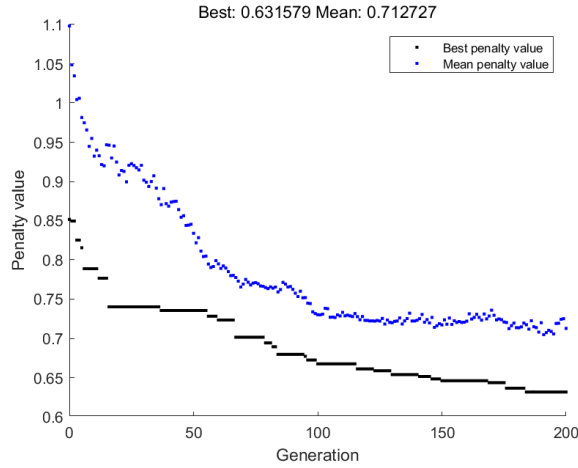


Figure 2. Fitness Curve of the Genetic Algorithm for Scenario 1

In Scenario 1, the portion exceeding the expected sales volume cannot be sold as usual. The fitness curve of the genetic algorithm shows that the solid black line represents the penalty value of the optimal individual in each generation. It is evident that the penalty value is relatively high at the initial stage and gradually decreases as the number of generations increases. By the end of the iterations, the penalty value of the optimal solution reaches 0.631579, indicating that the algorithm has identified a relatively optimal solution. Throughout the process, the penalty value decreases in a stepwise manner, demonstrating that the designed genetic algorithm can continuously improve the optimal solution. The dashed blue line represents the average penalty value of all individuals in each generation. During the first 50 generations, the average penalty value decreases rapidly, indicating a gradual improvement in the quality of most individuals in the population. From approximately the 100th generation, the average penalty value stabilizes with minor fluctuations, suggesting the presence of multiple local optima within the population at this stage. In the final dozens of generations, the average penalty value remains around 0.63, indicating minimal variation in the overall solution quality and that the population has largely converged.

The crop planting conditions for certain plots in different years are calculated and illustrated in Figures 3 to 5.

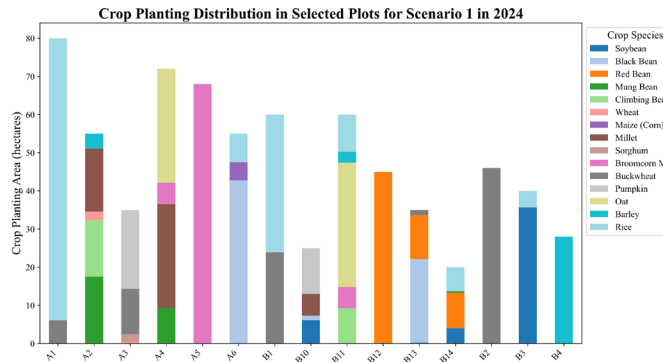


Figure 3. Crop planting distribution for certain plots under Scenario 1 in 2024.

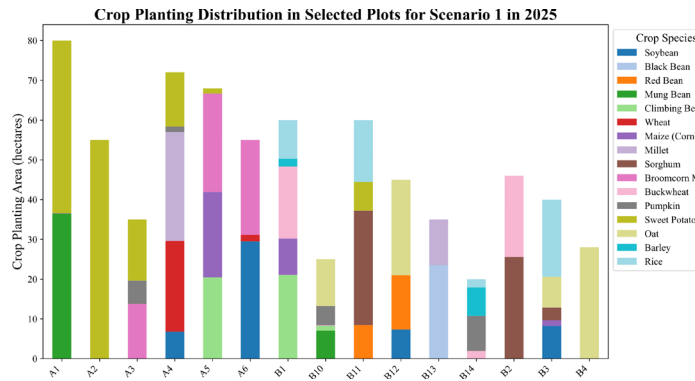


Figure 4. Crop planting distribution for certain plots under Scenario 1 in 2025.

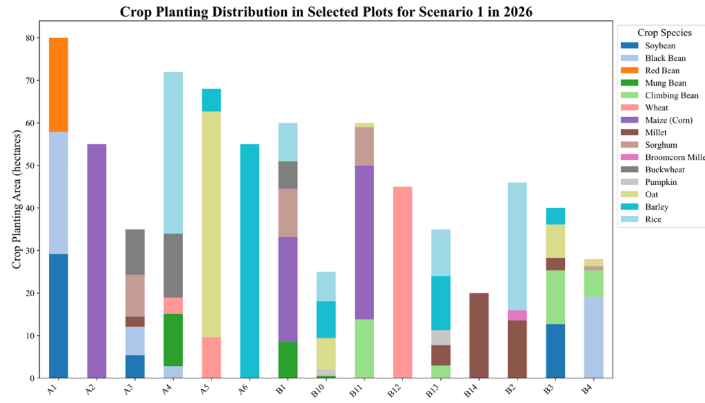


Figure 5. Crop planting distribution for certain plots under Scenario 1 in 2026.

From the figures, it is evident that the types of crops planted on different plots vary and adhere to the constraints of each plot type. Additionally, the planting scheme satisfies the requirements for crop diversity and legume crop rotation outlined in the problem, particularly avoiding consecutive planting of the same crop (repeated cropping). The genetic algorithm fitness curve under Scenario 2 is shown in Figure 6.

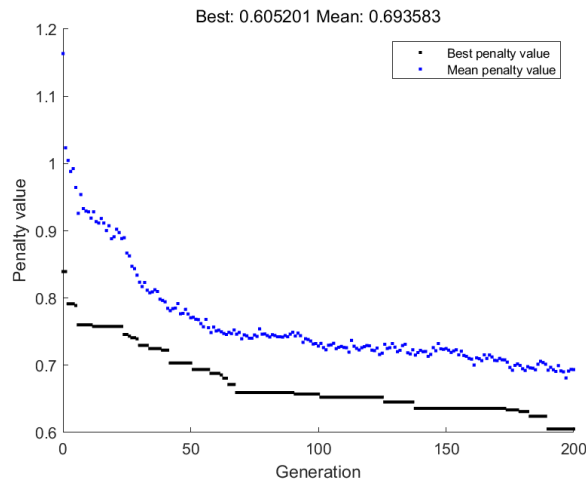


Figure 6. Genetic algorithm fitness curve for Scenario 2.

The figure shows a significant decline in the optimal penalty value at the early stages of the algorithm, particularly in the first 50 generations, where the optimal value drops rapidly from approximately 1.1 to around 0.7. This indicates that most of the poorer solutions were quickly eliminated in the initial phase, and the algorithm identified higher-quality solutions. Furthermore, the average solution quality stabilizes toward the end, with most solutions converging to similar quality levels. This demonstrates that the genetic algorithm achieved notable optimization performance and successfully identified a near-optimal solution. The crop planting distributions for certain plots in different years under Scenario 2 are shown in Figures 7 to 9:

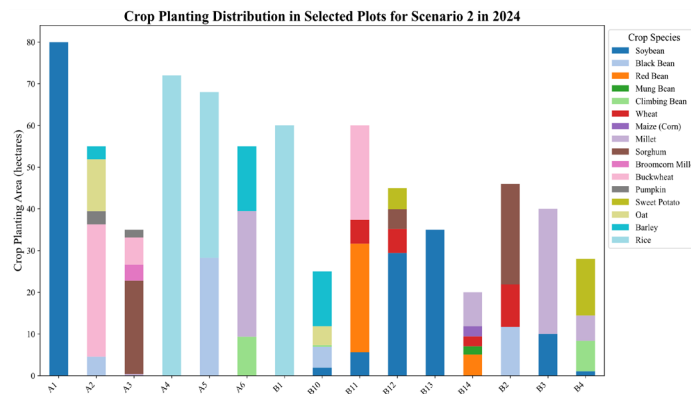


Figure 7. Crop planting distribution for certain plots under Scenario 2 in 2024.

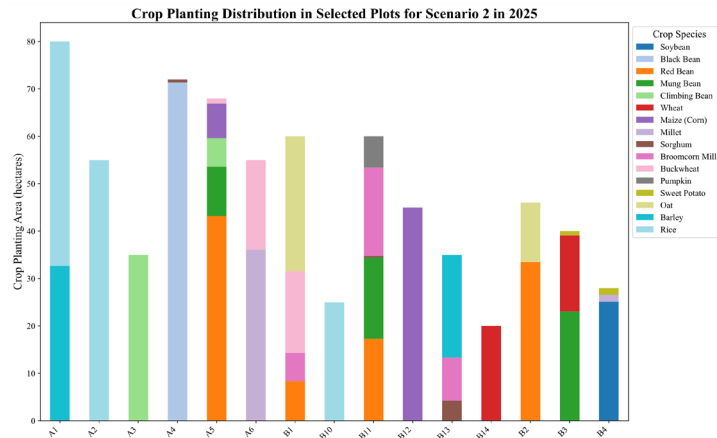


Figure 8. Crop planting distribution for certain plots under Scenario 2 in 2025.

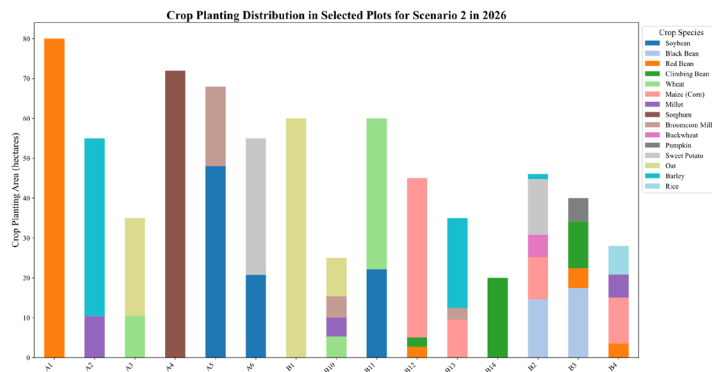


Figure 9. Crop planting distribution for certain plots under Scenario 2 in 2026.

From the figures, it is evident that Scenario 2 also satisfies the relevant constraints. According to the scenario's settings, surplus yields beyond expected sales volumes are sold at a discounted price. This means that for high-yield crops, the excess yield results in reduced profits but still contributes to total sales. Under this sales strategy, high-yield crops have a significant proportion of surplus sold at a 50% discount, reducing per-unit profit but contributing to overall sales volume. Medium-yield crops produce less surplus compared to high-yield crops but still generate some additional sales. For low-yield crops, there is almost no surplus beyond the expected sales volume, allowing them to be sold at full price and maintain a high profit margin.

The total profit over 7 years for Scenario 1 is calculated to be CNY 7,851,693, with the annual profit results shown in Table 3:

Table.3. Annual Total Profit for Scenario 1

Year	2024	2025	2026	2027	2028	2029	2030
Profit(CNY)	1147824	1022894	1181217	1147824	1022894	1181217	1147823

The total profit over 7 years for Scenario 2 is calculated to be CNY 23,868,146, with the annual profit results shown in Table 4:

Table.4. Annual Total Profit for Scenario 2

Year	2024	2025	2026	2027	2028	2029	2030
Profit(CNY)	3993191	3304977	3986200	2772734	3761774	2637214	3412057

4. Conclusions

This study takes rural areas in the mountainous regions of northern China as an example to develop a linear programming model that integrates Monte Carlo simulation and genetic algorithms. The model is designed to optimize crop planting structures and areas under different surplus sales scenarios, aiming to enhance agricultural production efficiency and maximize profits. By fully accounting for plot characteristics and planting constraints, the model demonstrates high practicality and accuracy. It

effectively reduces unsold waste, increases profits, and adapts to diverse regions and climatic conditions, making it highly transferable. Additionally, by incorporating weather forecasting and disaster prevention mechanisms, the model helps producers address challenges posed by extreme weather due to climate change, thereby improving the adaptability and sustainability of agricultural systems. It is applicable to optimizing production, processing, and sales across all stages of the agricultural value chain.

While the linear programming model, combined with genetic algorithms and Monte Carlo simulation, significantly improves the precision of planting schemes, it heavily depends on meteorological data and market demand. Furthermore, as the number of crop types and years increases, the solution space grows exponentially, resulting in high computational complexity. The model also exhibits certain limitations in responding to extreme weather and performing real-time dynamic adjustments. To address the rapid expansion of the solution space with the increase in plots, crop types, and planning years, parallel computing or distributed optimization techniques can be employed to decompose the problem into sub-problems for simultaneous processing. This approach can significantly reduce computation time and enhance the efficiency of large-scale agricultural planning.

References

- [1] Zhang Aihua. *High-yield and high-quality winter wheat cultivation technology based on genetic algorithm [J]*. *China Agricultural Abstracts - Agricultural Engineering*, 2023, 35(02): 86-90.
- [2] Bi, L., Hu, G. *A genetic algorithm-assisted deep learning approach for crop yield prediction [J]*. *Soft Computing*, 2021, 25: 10617-10628.
- [3] Yang, Z.H. *Analysis and Prediction of Factors Influencing Soybean Production in Heilongjiang Province [D]*. Northeast Agricultural University, 2012.
- [4] Wu, J.J. *Research on Agricultural Planting Structure Optimization Based on Linear Programming Model [J]*. *New Farmer*, 2024, (03): 28–30.
- [5] Chen, C.Y., Wang, Y.C., Shi, H.Y., et al. *Application of Genetic Algorithm in Irrigation Decision Optimization of *Isatis tinctoria* in the Hexi Oasis [J]*. *Agricultural Engineering*, 2023, 13(06): 53–59.