

Research on crop planting strategy based on improved genetic algorithm

Wenbo Zhang^{1, #}, Jingzhao Feng^{2, #, *}, Caizhi Tian^{3, #}

¹ Qingdao Institute of Software Technology, College of computer science and technology, China University of Petroleum (East China), Qingdao, China, 266580

² School of Marine and Spatial Information, China University of Petroleum (East China), Qingdao, China, 266580

³ College of New Energy, China University of Petroleum (East China), Qingdao, China, 266580

* Corresponding Author Email: 2209060408@s.upc.edu.cn

#These authors contributed equally.

Abstract. With the continuous development of China's national economy and the steady promotion of the rural revitalization strategy, optimizing crop planting strategies has become increasingly prominent. In this study, a crop planting strategy optimization method based on an improved genetic algorithm is proposed based on careful consideration of planting area, crop rotation system, crop growing season, and topography. Specifically, on the one hand, this method uses hyperbolic tangent function fitting, Pearson correlation coefficient, and hierarchical analysis to construct a correlation coefficient model, and specifies alternative and complementary constraints. On the other hand, the relationship between expected sales volume and sales price, and planting cost is analyzed by Lasso regression, the corresponding objective function is designed, and the proposed optimization scheme is solved by the improved genetic algorithm. The results of simulation experiments verified the effectiveness of the method and provided strong support for achieving scientific and rational agricultural cultivation. This paper innovatively proposed an improved genetic algorithm to optimize the planting strategy, effectively optimize the planting strategy, and improve the economic benefits of crops.

Keywords: Planting Strategies, Optimisation Algorithms, Genetic Algorithms, Statistical Learning.

1. Introduction

Crop planting strategy optimization is of great significance in agricultural production. It is conducive to promoting the sustainable development of agricultural production, improving the efficiency of agricultural resources, guaranteeing national food security, promoting agricultural technological innovation and industrial upgrading, and increasing farmers' income and employment opportunities. It is not only related to the livelihood of farmers but also to national food security, sustainable development of agriculture, and ecological environment protection.

There has been a lot of research on crop cultivation. For example, Feng Guangliang used a particle swarm optimization algorithm to solve the two-layer planting structure optimization model constructed and carried out the optimization calculation of the planting structure in the study area under the conditions of single-layer planning and double-layer planning [1]. Li Lixing et al. studied and established a multi-objective optimization model for regional crop structure scheduling based on the principle of "grain-oriented" and the consideration of fully applying agricultural big data to adjust and optimize the rural crop structure. The PSO algorithm based on the Spark platform is used to solve the model, which improves the running speed of the program [2]. Aiming at the scientific problems of planting structure optimization and rational allocation of agricultural water in the Manas River irrigation district under the constraint of total water resources, Pan Yue et al. constructed a multi-objective optimal allocation model to optimize and adjust crop planting structure and adopted the improved Gray Wolf optimization algorithm (DSF-GWO) to solve the problem [3]. Based on the calibrated crop growth model, Chen Lei carried out planting decisions on driving factors such as meteorological conditions, irrigation treatment, and fertilization management [4]. Based on the

swarm intelligence algorithm and crop model, the optimal planting decision is solved to meet the needs of agricultural production. Li Yanbin et al. introduced inertial weight attenuation and particle variation strategies to establish a multi-objective agricultural planting structure optimization model based on an improved particle swarm optimization algorithm, which can improve comprehensive benefits and yield while reducing irrigation water requirements [5]. Wang Lu constructed a multi-objective crop planting structure adjustment model based on NSGA-II to optimize the planting structure in the upper reaches of Baiyangdian Lake [6]. Based on the Matlab Global Search optimization algorithm, Yang Xiong established an optimization model for planting structure objectives, taking the maximum total crop output and the maximum total output value as two optimization objectives, taking the annual sown area of different crops as decision variables, and taking the type of cultivated land and planting system as constraint conditions to optimize the planting structure [7]. Wu Menghan and Wang Yi took Shachai Irrigation District in Xinjiang as an example and used a multi-objective comprehensive optimization model to optimize and analyze the planting structure of crops in the irrigation district from three perspectives: agricultural ecological benefit, economic benefit, and irrigation water consumption [8]. Using a systematic analysis method, Tu integrated the water footprint theory, fuzzy chance-constrained programming (FCCP), and multi-objective programming model (MOPM) into a coupled framework to optimize the allocation of agricultural water and soil resources and crop planting structure, and analyzed their effectiveness under different believability [9]. Taking Shanxi Province as the research area, Zhang Pengfei et al. analyzed the water consumption and energy consumption characteristics of major crops by using water footprint and energy consumption accounting models and built a multi-objective programming model with water resources, energy, and grain as constraints to optimize the planting structure of major crops [10]. Combining the advantages of deterministic dynamic programming and stochastic programming, Zhang Hao et al. established an optimization model of planting structure in tropical grassland climatic irrigation districts based on stochastic dynamic programming [11].

To solve the optimal planting scheme of crops, we use a new way to accelerate the convergence speed of genetic algorithms. Aiming at the shortcoming of slow convergence of the genetic algorithm, we increase mutation probability linearly and cross probability decrease in the iterative process. Under the assumption that various crops do not affect each other, the optimal planting plan can be solved by doing so. Considering the complementarity and substitutability between multiple crops and the relationship between sales volume, price, and cost, we need to optimize the above model from two dimensions: horizontal and vertical. Horizontally, depending on the complementarity and substitutability of different crops, the correlation between sales volume, price, and cost of other crops was discussed. The Pearson correlation coefficient was used to analyze the price of different crops to determine the correlation between crops. Using the analytic hierarchy process, the proportion weights of weather factors, sales price, and market demand are determined respectively, and the correlation coefficient model is established. Vertically, through the Lasso regression model, the relationship between the expected sales of crops, sales price, and planting cost is established. After modifying the previous objective function, adding complementarity constraint and substitutability constraint, the simulation data are solved, and the results without considering the interaction of various crops are analyzed.

2. Theory and method

Without considering the substitutability and complementarity between crops and the relationship between expected sales, selling price, and planting cost, the objective function is as follows:

$$\max Z = \sum_{t=2024}^{2030} \sum_{i=1}^n \sum_{j=1}^m (\min(x_{i,j,t} \cdot Y_{i,j,t}, S_{i,t}) \cdot P_{i,t} - x_{i,j,t} \cdot C_{i,t}) \quad (1)$$

The interaction between prices and sales of crops i and k is reflected by market correlation. If they are substitutes, the correlation coefficient shows the competition. If they are complementary, the

correlation coefficient is positive and sales and prices increase together. Market correlation is defined by the elasticity of market demand. When supply of crop i changes, market demand for crop k may change accordingly depending on the relationship.

$$R_{i,k}^s = \frac{1}{7} \sum_{t=2024}^{2030} \left(-\mu_{i,k} \cdot \frac{S_{k,t}}{\sum_{j=1}^{54} (x_{k,j,t} \cdot Y_{k,j,t})} + \nu_{i,k} \cdot \frac{S_{i,t}}{\sum_{j=1}^{54} (x_{i,j,t} \cdot Y_{i,j,t})} \right) \quad (2)$$

Where $R_{i,k}^s$ is market correlation coefficient. When $\mu_{i,k} > 0$, crop i and k are substitutable. When $\nu_{i,k} > 0$, they are complementary. Price-volume relationship is reflected in planted area $x_{k,j,t}$ and expected sales volume $S_{k,t}$. Hyperbolic tangent function is introduced to standardize and normalize correlation coefficients:

$$r_{i,k}^s = \tanh(R_{i,k}^s) \quad (3)$$

Where $r_{i,k}^s \in [-1, 1]$. When $r_{i,k}^s$ is negative and closer to -1, stronger i, k substitution. When positive and closer to 1, stronger i, k complementarity. Closer absolute value to 1 means stronger correlation. Also, under various factors, crop i and k prices are correlated. When i 's selling price changes, k 's price may rise due to complementarity. If directly competitive, prices may fall due to vicious competition. Based on Pearson correlation coefficient, judge correlation between sales prices of various crops. Formula as follows:

$$\rho_{i,k}^c = \frac{E(x_{k,j,t} \cdot x_{i,j,t}) - E(x_{k,j,t}) \cdot E(x_{i,j,t})}{\sqrt{E(x_{k,j,t}^2) - E^2(x_{k,j,t})} \cdot \sqrt{E(x_{i,j,t}^2) - E^2(x_{i,j,t})}} \quad (4)$$

Larger $\rho_{i,k}^c$ means stronger positive correlation and substitutability. Smaller $\rho_{i,k}^c$ means stronger negative correlation and complementarity. Also, climate change effects on different crops may be related. If crops I and K respond similarly to same climatic conditions, yields fluctuate similarly. Specifically, $r_{i,k}^q = 1$ for same response (positively correlated), $r_{i,k}^q = -1$ for opposite response (negative correlation), and $r_{i,k}^q = 0$ for no response. Next, build a correlation model. Divide complexity of $r_{i,k}$ into multiple parts reflecting different factors. Final correlation coefficient is a weighted sum of multiple parts:

$$r_{i,k} = w_1 \cdot r_{i,k}^s + w_2 \cdot \rho_{i,k}^c + w_3 \cdot r_{i,k}^q \quad (5)$$

Among them, w_1, w_2, w_3 are weighting coefficients, balanced by AHP to meet $w_1 + w_2 + w_3 = 1$. Establish Lasso model with planting cost and unit price as independent vars and expected sales volume as dependent var. Determine Lasso model hyperparameters by cross-validation. Determine variables based on regression coefficient results. Variables with zero normalization coefficient are eliminated by Lasso regression. Finally, obtain Lasso regression formula and prediction and list retained and excluded vars. Specific fitting formula as follows:

$$\hat{S}_{i,t} = \eta_0 + \eta_1 P_{i,t} + \eta_2 C_{i,t} + \varepsilon \quad (6)$$

Where $P_{i,t}$ represents the selling price of crop i in year t , and $C_{i,t}$ represents the planting cost of crop i in year t . By introducing the correlation coefficient model, the overall objective function can be modified as follows:

$$\begin{aligned} \max Z = & \sum_{t=2024}^{2030} \sum_{i=1}^n \sum_{j=1}^m \left[\left(\min(x_{i,j,t} \cdot Y_{i,j,t}, \hat{S}_{i,t}) \cdot P_{i,t} + E_{i,j,t} \cdot 0.5 \cdot P_{i,t} \right) - x_{i,j,t} \cdot C_{i,t} \right] \\ & + \sum_{i,k} r_{i,k} \cdot x_{i,j,t} \cdot x_{k,j,t} \end{aligned} \quad (7)$$

The model incorporates complex market supply-demand, crop substitution/complementarity, climate impact on yield, and land adaptability, making Problem 3's model more complex and realistic.

Solving this model yields a more optimized planting scheme. Considering actual planting, the following constraints are proposed:

(1) Alternative constraints:

$$x_{i,j,t} + x_{k,j,t} \leq U_{i,k,j,t} \quad (8)$$

(2) Complementary constraints:

$$x_{i,j,t} \times x_{k,j,t} \leq M_{i,k,j,t} \quad (9)$$

(3) General constraints:

$$s.t. \left\{ \begin{array}{l} \sum_{t=t_0}^{t_0+2} \sum_{i \in B} x_{i,j,t} \geq \epsilon_j, \quad \forall j = 1, 2, \dots, m, \quad \forall t_0 = 2024, 2025, 2026, 2027 \\ x_{i,j,t} \cdot x_{i,j,t-1} = 0, \quad \forall i, j, t \geq 2024 \\ \sum_{i=1}^{16} \sum_{j=27}^{54} x_{i,j,t} = 0, \quad \forall t = 2024, 2025, \dots, 2030 \\ \sum_{j=1}^{26} x_{16,j,t} + \sum_{j=35}^{54} x_{16,j,t} = 0, \quad \forall t = 2024, 2025, \dots, 2030 \\ \sum_{i=17}^{34} \sum_{j=1}^{26} x_{i,j,t} = 0, \quad \forall t = 2024, 2025, \dots, 2030 \\ \sum_{i=35}^{37} \sum_{j=1}^{26} x_{i,j,t} + \sum_{i=35}^{37} \sum_{j=35}^{54} x_{i,j,t} = 0, \quad \forall t = 2024, 2025, \dots, 2030 \\ \sum_{i=38}^{41} \sum_{j=1}^{34} x_{i,j,t} + \sum_{i=38}^{41} \sum_{j=51}^{54} x_{i,j,t} = 0, \quad \forall t = 2024, 2025, \dots, 2030 \\ \sum_{i=17}^{34} \sum_{j=27}^{50} \beta_{i,j,t} = 0, \quad \forall t = 2024, 2025, \dots, 2030 \\ \sum_{i=35}^{37} \sum_{j=27}^{34} \alpha_{i,j,t} = 0, \quad \forall t = 2024, 2025, \dots, 2030 \\ \sum_{i=38}^{41} \sum_{j=35}^{50} \alpha_{i,j,t} = 0, \quad \forall t = 2024, 2025, \dots, 2030 \\ x_{i,j,t} \leq A_j \quad \forall t = 2024, 2025, \dots, 2030 \\ \quad \quad \quad \forall j = 1, 2, \dots, 34 \quad \forall i = 1, 2, \dots, 16 \\ 0 \leq \alpha_{i,j,t} \leq A_j \quad \forall t = 2024, 2025, \dots, 2030 \quad \forall j = 27, 28, \dots, 34 \quad \forall i = 17, 18, \dots, 41 \\ 0 \leq \beta_{i,j,t} \leq A_j \quad \forall t = 2024, 2025, \dots, 2030 \quad \forall j = 27, 28, \dots, 34 \quad \forall i = 17, 18, \dots, 41 \\ 0 \leq \alpha_{i,j,t} \leq A_j \quad \forall t = 2024, 2025, \dots, 2030 \quad \forall j = 35, 36, \dots, 54 \quad \forall i = 17, 18, \dots, 41 \\ 0 \leq \beta_{i,j,t} \leq A_j \quad \forall t = 2024, 2025, \dots, 2030 \quad \forall j = 35, 36, \dots, 54 \quad \forall i = 17, 18, \dots, 41 \\ \sum_{i=1}^n I(x_{i,j,t} > 0) \leq 2, \quad \forall j = 35, 36, \dots, 54, \quad \forall t = 2024, 2025, \dots, 2030 \end{array} \right. \quad (10)$$

The above mixed integer programming is complex, so genetic algorithm is employed. Genetic algorithm simulates Darwinian evolution and natural selection. Matlab is used to design an improved adaptive genetic algorithm to solve the first and second sub-questions. Set max evolutions to 2000. Mutation prob. increases linearly from 0.2 to 0.4. Crossover prob. decreases linearly from 0.8 to 0.5 to accelerate convergence. Flow chart as in Figure 1.

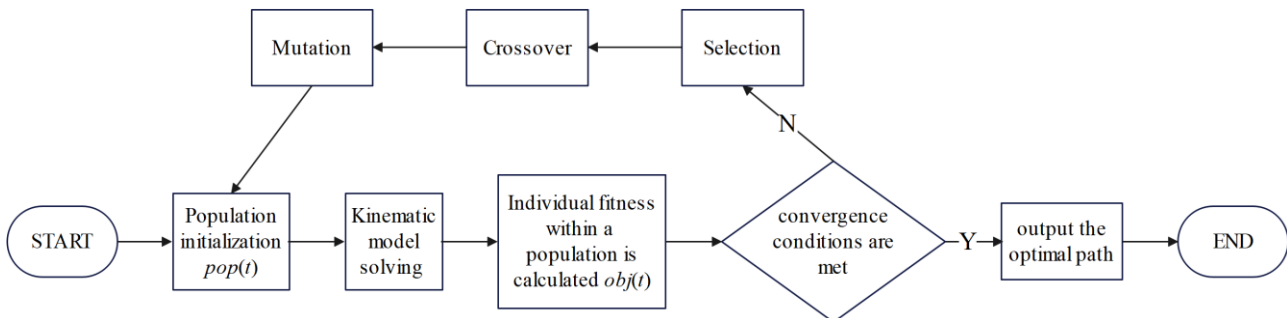


Figure 1. Schematic diagram of optimization process

3. Real data analysis

According to the data of a village located in the mountainous area of North China given by the National College Students Mathematical Modeling Competition, we first preprocessed and observed the data. There are five types of plots, of which terraced fields account for the largest proportion (51%), followed by flat dry land (30.1%), hill land (9.0%) and irrigated land (8.9%), and ordinary greenhouses (0.8%) and smart greenhouses (0.2%) account for the least. In the data obtained by the author, each crop has the highest unit price and the lowest unit price. In this paper, the average value of the highest and lowest unit price is taken as the unit price of this crop. Carefully observing the data, it can be found that the yield per mu of the same crop in different plots is also different. Generally speaking, the yield of planting on flat and dry land is greater than that of planting on terrace land, greater than that of planting on hillside land, and the yield of planting in greenhouses is greater than that of planting on irrigated land.

3.1. Correlation between sales prices

Affected by many factors, the price of crops has a certain correlation with that of crops. The seven-year sales prices of various crops are processed. The Pearson correlation coefficient can be used to measure the degree of correlation between two variables (formula (4)). Visualize it using heat maps.

It can be seen from figure 2 that the larger $\rho_{i,k}^c$ (red) is, the stronger the positive correlation between the two, and the two will crowd out the market space and have substitutability; the smaller $\rho_{i,k}^c$ (yellow) is, the stronger the negative correlation between the two, and the two are complementary. Based on this, this paper makes the following assumptions:

$\rho_{i,k}^c > 0.85$ There is substitution between the two crops.

$\rho_{i,k}^c < -0.85$ There is complementarity between the two crops.

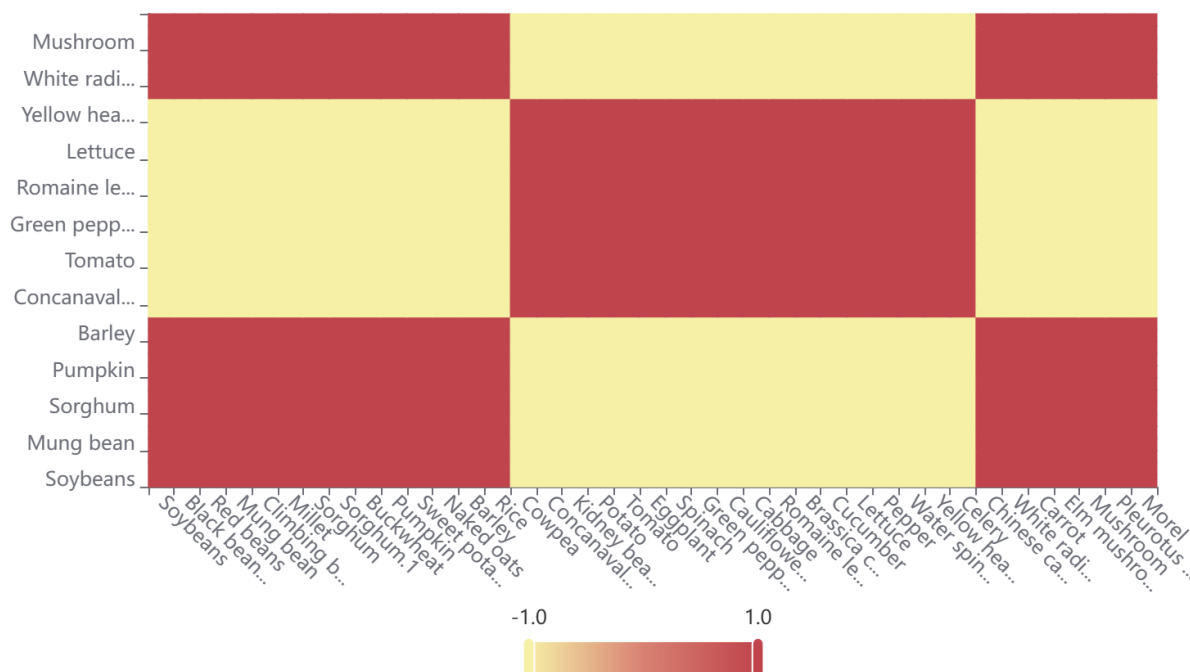


Figure 2. Pearson correlation coefficient heat map

3.2. The correlation between expected sales volume, selling price and planting cost

With the planting cost and selling unit price of each crop as independent variables and the expected sales volume as dependent variables, the Lasso model (shown in figure 3) was established for analysis.

Step1: The λ value was determined by cross-validation method. The λ value is chosen to minimize the mean square error of the Lasso model.

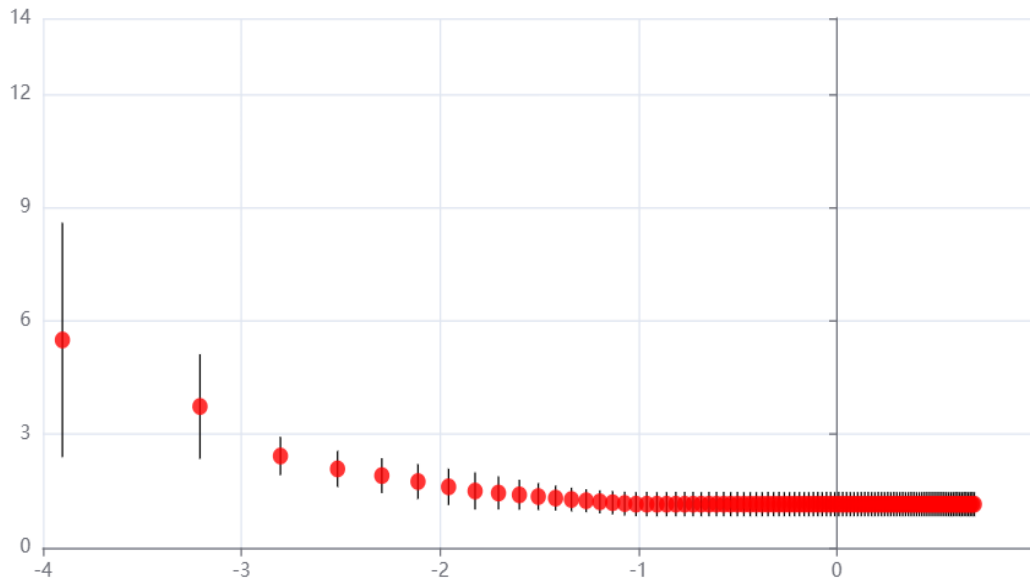


Figure 3. Lasso regression cross validation graph

To minimize the mean square error, determine $\lambda = 2.0$, $(\log(\lambda)) = 0.693$.

Step2: λ and regression coefficient plots were used to determine the variables selected by the model, in which variables with zero normalization coefficient could be considered to be eliminated by the Lasso regression model (shown in figure 4).

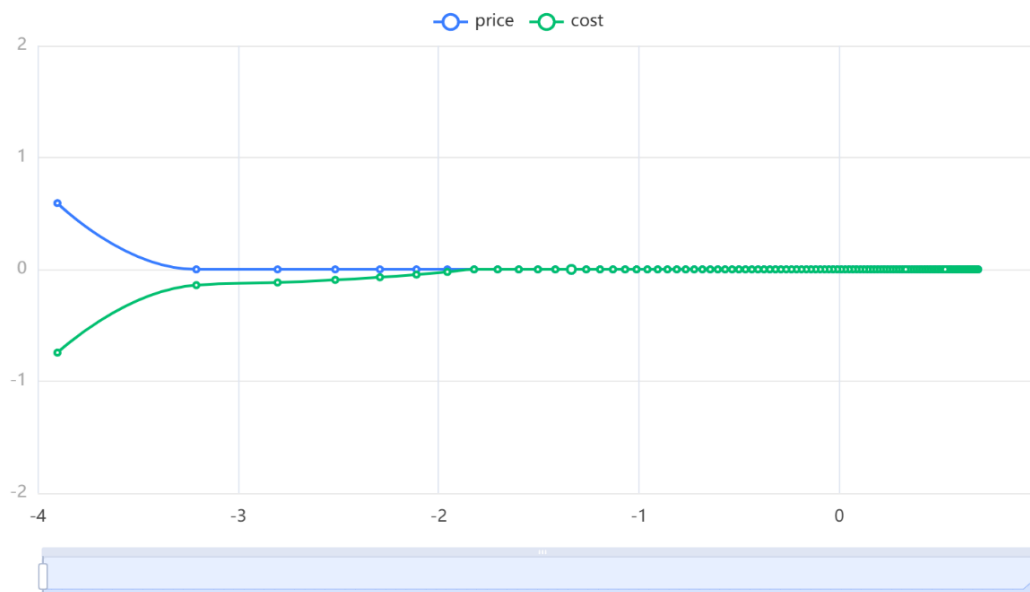


Figure 4. λ and model regression coefficient graph

Step3: The formula (formula (6)) and prediction of the Lasso regression model are obtained, and the retained and excluded variables are listed. The regression results are shown in table 1:

Table 1. Lasso returns results

Variable name	Standardization coefficient	Nonnormalized coefficient	R ²
Intercept	574966.728	581270.982	0.159
cost	-4.062	-4.013	
price	1707.475	1684.776	

As shown in the figure 5, after 1000 iterations of the genetic algorithm, the fitness has been relatively stable, and the optimal planting scheme solved is represented by a bar chart. As shown in the figure 6 and figure 7, the results showed that compared with the results without considering the

mutual influence of various crops, the planting amount of vegetables increased by 15.6%, and the planting amount of the same kind of crops had a large difference, and there was a trend of polarization.

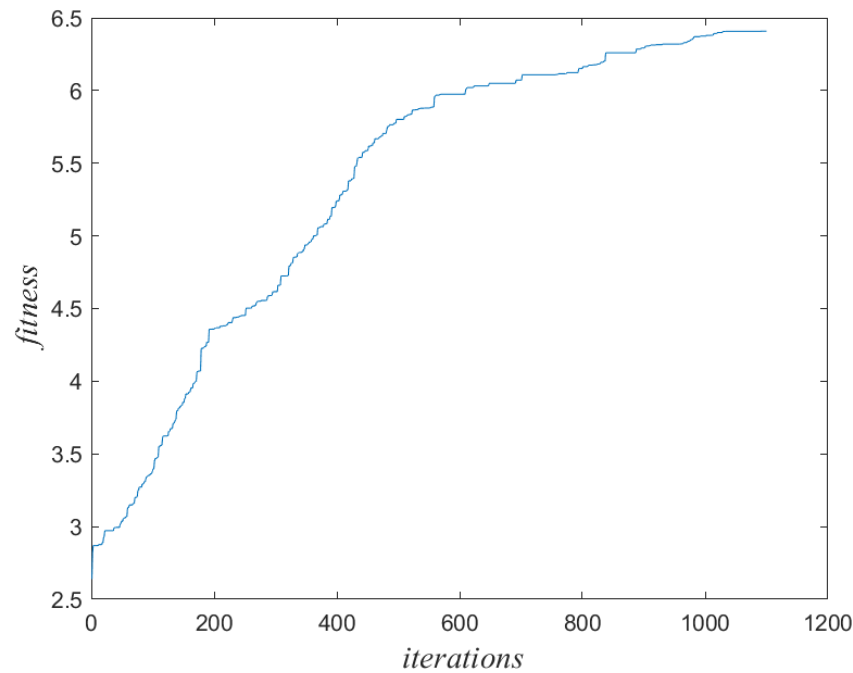


Figure 5. The change of fitness of genetic algorithm with the increase of iteration times

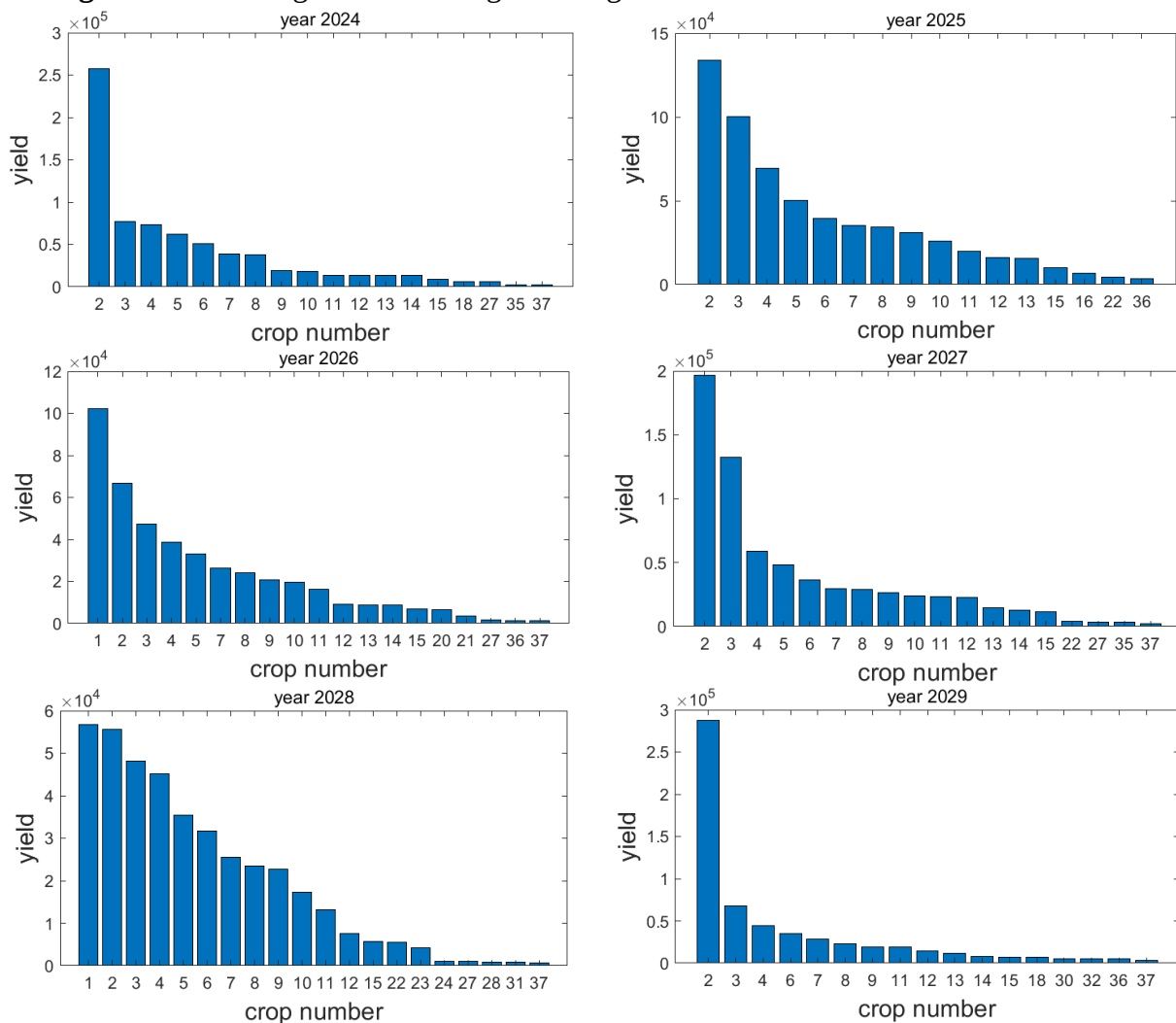


Figure 6. Run result (by year)

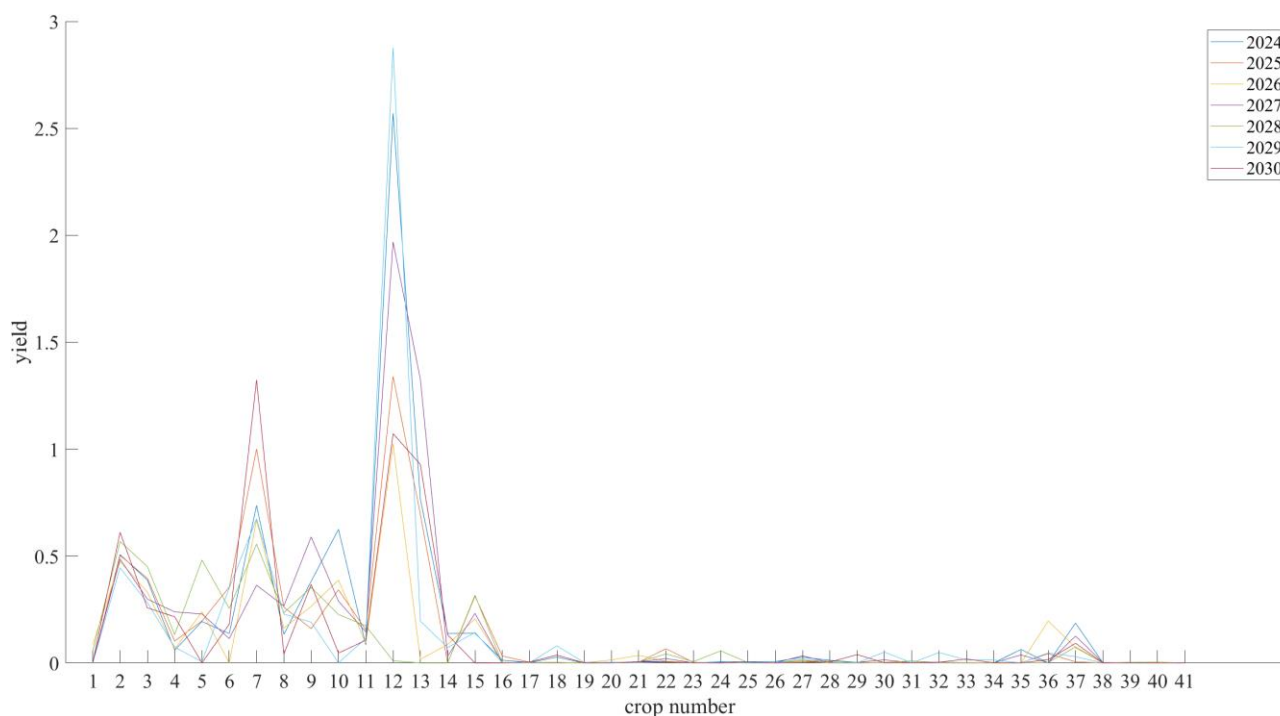


Figure 7. Operation result (subplot)

4. Conclusions

In summary, the model constructed in this study has excellent performance in terms of reality, algorithm effectiveness, and reliability of experimental verification. In terms of reality, the model fully considers the fungibility and complementarity between crops, and at the same time includes practical factors such as market demand and climate change, so that the obtained results are highly consistent with crop planting planning in the real world. In terms of algorithm effectiveness, the introduction of a genetic algorithm successfully solves complex linear and nonlinear programming problems under multiple constraints, in which the adaptive crossover and mutation probability significantly improve the convergence speed and the accuracy of the algorithm. In terms of experimental verification, to ensure the accuracy and stability of the results, we conducted several experiments and used the control variable method to conduct an in-depth analysis of the optimal solution distribution under different disturbance terms, which fully confirmed the reliability of the scheme. Therefore, this model can provide scientific, rigorous, and feasible decision support for crop planting planning.

Future research directions can be further expanded from the following key dimensions. First of all, in the detailed consideration of climate factors, it is necessary to build more complex climate disturbance models. Specifically, the probability of extreme weather events is included in the analysis, and their unique impact on different crop yields is explored in depth to more accurately reflect the potential threat of climate change to crops and lay the foundation for the development of scientific and effective response strategies. Secondly, in terms of market price prediction, more complex time series models or machine learning models can be used to accurately predict the future trend of market price changes through rich historical data, reduce the market volatility simplify assumptions, and improve the accuracy and reliability of market prediction. Thirdly, a multi-objective optimization model should be developed from the perspective of multi-objective optimization. Based on the current profit maximization goal, environmental protection, and social benefits are introduced promptly to promote a more comprehensive and sustainable crop planting program. Fourth, in the field of application of the model, it should be widely expanded in different climatic zones and geographical conditions of crop planting planning. For example, in the southern rice-growing area and the

northwest dry-farming area, the model can be widely used by adjusting crop parameters and climate conditions reasonably. Finally, in terms of cross-domain application exploration, the feasibility of this model in resource optimization in other fields such as energy distribution and manufacturing production planning is actively explored, and its universal advantages in considering resource substitution and complementarity are fully utilized, providing new ideas and methods for resource optimization in different fields.

References

- [1] Feng Guangliang. Research on agricultural planting structure allocation in the lower reaches of the Tarim River based on a double-layer planting optimization model [J]. Technical Supervision in Water Resources, 2024, (05): 213-217.
- [2] Li Liuxing, Tang Li. Multi-objective optimization model and PSO parallel algorithm for planting structure scheduling [J]. Journal of Chongqing University of Science and Technology (Natural Science Edition), 2020, 22 (01): 60-62.
- [3] Pan Yue, Yang Guang, Xue Lianqing, et al. Application of DSF-GWO algorithm in optimization of planting structure in Manas River Irrigation District [J]. Journal of Drainage and Irrigation Machinery Engineering, 2023, 41 (09): 943-951.
- [4] Chen Lei. Research on optimal planting decision-making based on swarm intelligence algorithm and crop model [D]. Nanchang University, 2024.
- [5] Li Yanbin, Ma Jiatong, Li Daoxi, et al. Application of improved particle swarm optimization algorithm in optimization of agricultural planting structure [J]. Journal of Irrigation and Drainage, 2022, 41 (01): 62-71.
- [6] Wang Lu. Water consumption characteristics and optimization of planting structure under different cropping systems in the upper reaches of Baiyangdian Basin [D]. Hebei Agricultural University, 2021.
- [7] Yang Xiong. Research on optimization of planting structure in hilly areas of eastern Sichuan [D]. Sichuan Normal University, 2019.
- [8] Wu Menghan, Wang Yi. Multi-objective planting optimization and adjustment of crops in Shache Irrigation District, Xinjiang [J]. Yellow River, 2024, 46(01): 120-125+131.
- [9] Tu Jiating. Optimization allocation of agricultural water and soil resources in river basins based on water footprint theory [D]. Huazhong University of Science and Technology, 2023.
- [10] Zhang Pengfei, Xiao Menglin, Zhang Zhaorui, et al. Optimization and adjustment of agricultural planting structure in Shanxi Province based on the water-energy-food nexus [J]. Water Saving Irrigation, 2023, (01): 55-63+70.
- [11] Zhang Hao, Zhao Shengwei, Qian Jun, et al. Research on optimization of planting structure in tropical savanna climate irrigation districts based on stochastic dynamic programming [J]. Water Saving Irrigation, 2024, (10): 15-21.