Genetic Algorithm Based Optimal Strategy for Crop Planting Plots and Selections

Yangyuan Lai[1], Xibo Lin[1], Jingbo Yang[1], Hong Wang[2,*]

[1]College of Engineering Physics, Shenzhen Technology University, Shenzhen 518118, China
[2]College of Big Data and Internet, Shenzhen Technology University, Shenzhen 518118, China

*hitwanghong@163.com

Abstract. This study proposes a crop planting optimisation strategy based on a genetic algorithm that integrates agricultural production principles and market factors to provide a scientifically sound and rational approach to maximising profitability. Firstly, a linear programming model is constructed for the purpose of analysing common crops and market dynamics, thereby facilitating the scientific allocation of agricultural resources. Subsequently, in order to account for the volatility of key factors, such as market demand and crop yields, Spearman's correlation coefficient is employed for the purpose of analysing variable relationships. The introduction of complementarity and substitutability parameters enables the simulation of dynamic changes in dominant factors among crops, thereby generalising the model to adapt to different scenarios. The experimental results demonstrate that the proposed strategy exhibits stability, realism, and generalisability, thereby providing effective support for rural crop planting planning.

Keywords: Planting Strategy, Genetic Algorithm, Linear Programming, Spearman Rank Correlation Coefficient

1 Introduction

Implementing effective planting strategies can enhance agricultural yields, improve field management efficiency, and mitigate economic risks associated with various fluctuations. These strategies align with sustainable development goals and support the promotion of agricultural modernization and rural revitalization.

Effective agricultural planning continues to be a significant research focus. Liu Wei et al. utilized dynamic models to assess the impacts of climate change and market fluctuations on wheat-maize rotation systems, suggesting optimized land use strategies [1]. Zhang Xu et al. applied linear and objective programming to optimize crop planting structures under constraints of water, land, and labor, evaluating the economic benefits of various plans [2]. Xu Bin et al. developed a land use planning framework for rice production that considers resource efficiency, environmental impact, economic benefits, and social demands, proposing sustainable development strategies tailored to different regions [3]. Although these studies provide valuable insights into multi-factor

X. Li et al. (eds.), *Proceedings of the 2025 International Conference on Agriculture and Resource Economy (ICARE 2025)*, Advances in Biological Sciences Research 46,
https://doi.org/10.2991/978-94-6463-746-5_10

planting scheme optimization under specific conditions, their limitation lies in considering only a limited number of crops and constraints, making it challenging to apply their solutions directly to complex and dynamic real-world situations.

The objective of this study is to develop a linear programming model based on crop planting volumes on different plots, with the aim of optimising profitability while taking into account practical farming scenarios. The model considers a comprehensive range of constraints, including sales volumes, prices, production costs and yields. Furthermore, the model incorporates complementarity and substitutability coefficients derived from the interrelationships between crops. This allows for the simulation of real-world variations and enhances the applicability of the model by capturing the mutual substitutability and complementarity between crops. The model's constraints ensure the sustainable utilisation of economic, resource and environmental factors. Subsequently, a genetic algorithm is employed to resolve the model and ascertain the optimal cropping strategy.

2 Establish Crop Income Planning Model

This paper proposes an optimization scheme for agricultural structures aimed at maximizing planting income under realistic conditions. The model's formulation and solution are grounded in the following context, enhancing its general applicability: crops are cultivated on plots with varying production conditions, leading to differences in yields. Market fluctuations influence crop sales, selling prices, and cultivation costs, especially in rural areas with diverse crop types. Based on data from the China Grain Statistical Yearbook, land types are categorized into flat dry land, terraced land, hillside land, irrigated land, ordinary greenhouses, and smart greenhouses. Crop categories include grain, vegetables, edible fungi, pulses, and leguminous vegetables. Additionally, crop yields and prices are affected by temperature fluctuations, with two planting and marketing seasons annually. Specifically, flat dry land, terraced fields, and hillside plots can produce one season of food crops each year; irrigated land can produce two seasons of rice or vegetables; smart greenhouses can produce two seasons of vegetables (excluding Chinese cabbage, white radish, and carrot); and ordinary greenhouses are suitable for growing edible fungi in the second season.

2.1 Decision Variables

Decision variable X_{ijk} is the amount of the j type crops planted in the k season on the i plot.

2.2 Objective Function

As the objective function, the maximum income of the village is actually the total sales amount - the total production cost. The maximum return is defined as

$$maxS = \sum_{i}\sum_{j}\sum_{k}\left(q_{ijk} \cdot a_{ijk} - X_{ijk} \cdot c_{ijk}\right) \tag{1}$$

The benefit is set as S. We use q_{ijk} represents the sales volume of the j type crops planted in the k season on the i plot in North China Mountain area, a_{ijk} represents the selling price of the j type crops planted in the k season on the i plot, and c_{ijk} represents the planting cost per unit area of the j type crops planted in the k season on the i plot.

2.3 Constraints

Let n be the number of types for all crops. m =1, 2, 3, …

St.1: Cultivated land area constraint

The planting area cannot be greater than the area of all plots, b_{ijk} will be set as the area of all plots in the mountain area. The expression is:

$$\sum_{j=1}^{41} X_{ijk} \geq b_{ijk} \tag{2}$$

St.2: legume crop constraint

Since planting legumes improves soil structure, legumes must be planted at least once every three years in each plot and must be present in each plot at least once in three years, where $j \in H$ (H is the set) all represent legumes and the expression is:

$$count\left(\sum_{k=1}^{k+5}\sum_{i}\sum_{j} X_{ijk}\right) \geq 1, j \in H \tag{3}$$

St.3: Discontinuous planting constraints

In order to avoid affecting the crop yield and ensure the yield, each crop can not be continuously planted in the same plot. It is only necessary to ensure that the two planting times are different on the same plot. Then GX_{ijk} can be set as a binary variable [4], planting is 1, not planting is 0. The expression is:

$$GX_{ijk} \cdot GX_{ij(k-1)} = 0 \tag{4}$$

St.4: Flat drylands, terraced fields and hillsides can only grow one season crop constraints

Because the land in these three plots is relatively poor, only one crop can be grown per year. Flat drylands are represented by $i \in A$, terraced fields are represented by $i \in B$ and hillsides are represented by $i \in C$ (A, B and C are all sets). The expression is:

$$\begin{cases} flat\ drylands\text{：}\ X_{ijk} = 0, i \in A, k \in \{2,4,6 \cdots 2m\} \\ terraced\ fields\text{：}\ X_{ijk} = 0, i \in B, k \in \{2,4,6 \cdots 2m\} \\ hillsides\text{：}\ X_{ijk} = 0, i \in C, k \in \{2,4,6 \cdots 2m\} \end{cases} \tag{5}$$

St.5: annual mid-single season grain crop (except rice) constraints on flat dry-lands, hillsides and terraces

Rice production requires special conditions, is not suitable for flat dry land, slope land and terrace of these three plots, can only be planted in irrigated land, and the cultivation of other food crops also need better land fertility, for sustainable development, flat dry land, slope land and terrace in a year to plant single-season food crops. i =a indicates the number of the rice crop (a is a constant) and food crops are represented by $j \in Z$ (Z is set), which is converted into a mathematical formula:

$$\begin{cases} X_{ijk} = 0, i \in A, i \notin \{a\}, j \in Z, k \in \{2,4,6 \cdots 2m\} \\ X_{ijk} = 0, i \in B, i \notin \{a\}, j \in Z, k \in \{2,4,6 \cdots 2m\} \\ X_{ijk} = 0, i \in C, i \notin \{a\}, j \in Z, k \in \{2,4,6 \cdots 2m\} \end{cases} \tag{6}$$

St.6: Irrigated fields grow two seasons of rice alone per year or two seasons of vegetable crop restraint

Irrigated land is suitable for the growth of rice and vegetables, and most of them are fine seeds, so two seasons of rice or two seasons of vegetable crops can be planted on irrigated land each year. $\mathrm{V}_{\mathrm{ijk}}^{\max}$ denotes the maximum planting amount of the j type crops planted in the k season on the i plot. Irrigated plots are represented by $j \in D$ (D as the set). Convert the mathematical formula to

$$\begin{cases} 0 \le \mathrm{X}_{\mathrm{ijk}} \le \mathrm{V}_{\mathrm{ijk}}^{\max}, \mathrm{i} \in \mathrm{D}, \mathrm{j} \in \{1,2 \cdots \mathrm{n}\}, \mathrm{k} \in \{1,3 \cdots 2\mathrm{m}-1\} \\ 0 \le \mathrm{X}_{\mathrm{ijk}} \le \mathrm{V}_{\mathrm{ijk}}^{\max}, \mathrm{i} \in \mathrm{D}, \mathrm{j} \in \{1,2 \cdots \mathrm{n}\}, \mathrm{k} \in \{2,4 \cdots 2\mathrm{m}\} \end{cases} \tag{7}$$

St.7: Edible fungi can only be confined in the second season of ordinary greenhouses

The cultivation conditions of edible fungi are more special, need to be planted on the medium, and the temperature requirements are more stringent, so edible fungi can only be planted in the second season of ordinary greenhouses. The edible mushroom species is represented by $j \in Y$ and common greenhouse is represented by $i \in E$ (E, Y as the set). Convert the mathematical formula to

$$0 \le X_{ijk} \le V_{ijk}^{max}, i \in E, j \in Y, k \in \{2,4 \cdots 2m\} \tag{8}$$

St.8: Wisdom greenhouses are constrained by two seasons of vegetables per year (except Chinese cabbage, white radish and carrot)

Chinese cabbage, white radish and carrot are less profitable, so they are only grown in irrigated fields, and a maximum of one season a year, and they are represented by $j \in U$. Vegetables other than these are denoted by $j \in W$, which can be grown in vegetable sheds. Wisdom greenhouses is represented by $i \in F$ (F and W are sets). Convert to the mathematical formula as

$$\begin{cases} 0 \le X_{ijk} \le V_{ijk}^{max}, i \in D, j \in U, k \in \{2,4 \cdots 2m\} \\ 0 \le X_{ijk} \le V_{ijk}^{max}, i \in F, j \in W, k \in \{1,2,3 \cdots 2m\} \end{cases} \tag{9}$$

2.4 Complementarity and Substitutability

To describe the relationship among the unit selling price, sales volume, and cultivation cost of crops, a corresponding correlation coefficient is constructed. The expression is given as:

$$P_{i,k,t} = P_{j,0} \cdot \left(1 - \rho_{S,P} \cdot \frac{S_{j,k,t} - \bar{S}}{\bar{S}}\right) \tag{10}$$

$$C_{i,k,t} = C_{j,0} \cdot \left(1 + \rho_{S,C} \cdot \frac{S_{j,k,t} - \bar{S}}{\bar{S}}\right) \tag{11}$$

$$P_{i,k,t} = P_{j,0} \cdot \left(1 + \rho_{S,P} \cdot \frac{C_{j,k,t} - C_{j,0}}{C_{j,0}}\right) \tag{12}$$

In this paper, $\rho_{S,P}, \rho_{S,C}, \rho_{P,C}$ are defined as the correlation coefficient between sales volume and price, sales volume and cost, and price and cost, respectively.

This study builds on the model to further analyse the impact of substitutability, complementarity and variable correlations between different crops and varieties on the optimal cropping strategy, exploring diversified cropping schemes. Spearman's correlation analysis is employed to evaluate the interrelationships between crop categories by calculating the pairwise correlation coefficients (δ) for sales volume, sales price and production cost between crops. Spearman's correlation coefficient is defined as follows:

$$\gamma = 1 - \frac{6\sum x_i^2}{s(s^2 - 1)} \tag{13}$$

Here x_i represents the rank difference for the i-th data point, calculated as $x_i = rank(X_i) - rank(Y_i)$, where X_i and Y_i denote the values of the i-th sample for two variables. The Spearman correlation coefficient ranges between -1 and +1. A positive value indicates a positive correlation, i.e. when one variable increases, the other also increases. Conversely, a negative value indicates a negative correlation, meaning that as one variable increases, the other decreases.

Accordingly, the coefficients of complementarity and substitutability between crops are defined as follows:

(1) Complementarity Coefficient:

If a specific parameter (e.g., sales volume a_{i1k}) of one crop changes Δa_1, the sales volume of the other crop also fluctuates accordingly.

$$a_{i2k} = a_{0\,i2k} + \frac{\Delta a_1}{\delta_{12}} \tag{14}$$

$$\varepsilon_{12} = \frac{a_{i2k}}{a_{0\,i2k}} \tag{15}$$

$$\varepsilon_2 = \sum_j \varepsilon_{j2}\,. \tag{16}$$

Let $a_{0_{i2k}}$ denote the original sales volume of the other crop, and let a_{i2k} represent the sales volume after the change. The complementarity coefficient of crop $j = 1$ with respect to crop $j = 2$, denoted by ε_{12}, represents the extent to which crop $j = 1$ is complementary to crop $j = 2$. Similarly, the self-complementarity coefficient of crop $j = 2$, represented by ε_2, quantifies the degree to which crop $j = 2$ is self-complementary.

(2) Substitutability Coefficient:

If a specific parameter (e.g., sales volume a_{i1k}) of one crop changes Δa_1, the sales volume of the other crop also fluctuates accordingly.

$$a_{i2k} = a_{0_{i2k}} - \delta_{12} \cdot \Delta a_1 \tag{17}$$

$$\theta_{12} = \frac{a_{i2k}}{a_{0_{i2k}}} \tag{18}$$

$$\theta_2 = \sum_j \theta_{j2}\,. \tag{19}$$

Let $a_{0_{i2k}}$ denote the original sales volume of the alternative crop, and let a_{i2k} represent the sales volume subsequent to the change. The substitutability coefficient of crop $j = 1$ with respect to crop $j = 2$, represented by θ_{12}, quantifies the extent to which crop $j = 1$ can be substituted for crop $j = 2$.Similarly, the self-substitutability coefficient of crop $j = 2$, represented by θ_2, quantifies the extent to which crop $j = 2$ can be substituted for itself.

Considering complementarity and substitutability, the objective function is:

$$maxS_0 = \sum_i \sum_j \sum_k \frac{1}{2}\left(q_{ijk} \cdot a_{ijk} - X_{ijk} \cdot c_{ijk}\right) \cdot \left(\varepsilon_j + \theta_j\right). \tag{20}$$

The maximum profit, represented by the variable maxS_0, is contingent upon the substitutability and complementarity coefficients between crops.

3 Solve by Genetic Algorithm

3.1 Principle and Idea of Algorithm

A Genetic Algorithm (GA) is an algorithm that is based on Darwinian theory of biological evolution, natural selection, and the principles of genetics. It is used to search for the optimal solution to a problem. It has been widely applied to address problems related to the optimization of crop planting schemes[5]. Genetic algorithms simulate the process of natural selection, whereby the fittest organisms survive and reproduce, thus gradually evolving towards an optimal solution over multiple generations[6]. In computer simulations, this method transforms the problem into a genetic problem in the biological domain, whereby selection, reproduction, and evolution are performed on randomly generated planting schemes, ultimately yielding the optimal planting plan. The specific process is illustrated in Figure 1.

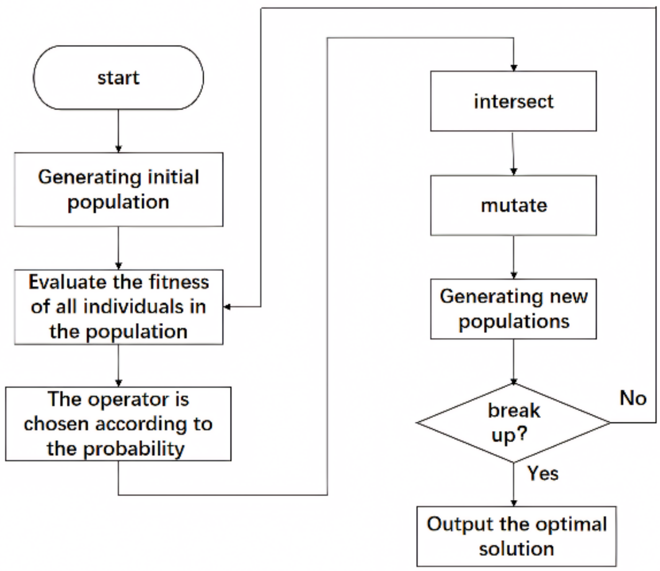

Fig. 1. Genetic algorithm flowchart

1. The initial population should be generated. The initial step is to ascertain the number of individuals that comprise the population. In accordance with the specified encoding method, each individual within the initial population should be randomly generated. To illustrate, if binary encoding is employed, each individual can be represented as a fixed-length binary string.

2. The fitness of all individuals in the population must then be evaluated. A fitness function must be defined that accepts an individual as input and returns a value representing its fitness. The higher the value returned by the fitness function, the more optimal the individual. Subsequently, each individual in the population should be traversed and the fitness function applied to each.

3. The selection operator is based on probability. The selection operator is employed to identify and select individuals from the current population to serve as parents for the subsequent generation. The selection process is typically probabilistic, with the likelihood of selection being proportional to the individual's fitness. Individuals exhibiting superior fitness are more likely to be selected. The most commonly employed selection methods are roulette wheel selection and tournament selection.

4. The process of crossover is defined as follows: One or more positions are selected from two or more parent individuals, and the genes at these positions are then exchanged or replaced in order to generate new offspring. The most common crossover methods are single-point crossover and two-point crossover. This process emulates the natural phenomenon of hybridization, whereby new traits are introduced into the offspring through genetic recombination, thereby enhancing genetic diversity.

5. Mutation: The objective is to introduce random alterations to specific gene values within an individual's genetic sequence, with the aim of increasing population diversity and exploratory behaviour. This contributes to the prevention of premature convergence of the algorithm and enhances global search performance. This operation emulates the process of gene mutation observed in nature, whereby new genotypes are introduced to enhance population diversity.

6. Generate New Population: Subsequent to the assessment of individual fitness following genetic operations, it is necessary to ascertain whether the requisite number of iterations has been completed. In the event that the stipulated conditions have been met, the optimal solution should be outputted. Conversely, if the aforementioned conditions have not been satisfied, the iteration should be continued, with the newly generated offspring being incorporated into the original population, thereby forming the subsequent generation. This process should then be repeated, with the aforementioned steps (evaluating fitness, selecting parents, crossover, and mutation) being undertaken until the stipulated stopping condition is met.

3.2 Implementation of the Algorithm

In accordance with the objective function of maximum profit and the aforementioned constraints, including those pertaining to arable land area, legume crop limitations, non-contiguous planting restrictions, and eight additional constraints, an initial population is randomly generated. Each member of the population represents a specific crop planting scheme. Furthermore, an appropriate fitness function, designated as $f(x)$, is established. The probability of an individual being selected is proportional to its fitness value, as determined by the roulette wheel selection operator [7]. The probability of selection is directly proportional to the fitness value, thus increasing the likelihood of passing on to the next generation. Subsequently, a series of genetic operations is employed to generate and iteratively refine new planting schemes. Once the profit from the optimal planting scheme has reached a stable equilibrium after 200 iterations, the process is terminated, and the optimal planting scheme is output.

In the design of algorithms for solving crop planting schemes, it is also necessary to consider more complex factors. In practical applications, a comprehensive assessment of the uncertainty inherent in the anticipated sales volume of crops, the yield per acre, the costs associated with planting, and the prices at which the crops will be sold is essential, along with an evaluation of the potential risks associated with planting. Monte Carlo simulation is a valuable tool for estimating mathematical functions and simulating the operation of complex systems through random sampling and statistical modeling. Its application in the generation of random variables is therefore a beneficial one. Furthermore, consideration must be given to the substitutability and complementarity of crops. Ultimately, the model can be optimized and solved using a genetic algorithm, thereby deriving the profit for different regions and the planting scheme for each crop over the next seven years.

4 Numerical Experiment

This study examines the historical conditions pertaining to the planting and sale of crops in a mountainous region of northern China. The model and algorithm developed in this paper were employed to analyze the historical data of the region, resulting in the optimal planting scheme and maximum profit for the subsequent seven years. Through iterative optimization via computer simulation, the final scheme's profit exceeds that of the 2023 planting scheme, thereby demonstrating the exceptional efficacy of genetic algorithms in addressing the challenge of optimal crop planting strategies.

In this study, a Spearman correlation analysis was conducted on the survey data in order to obtain the correlation coefficients between the selling price, sales volume and planting cost of different crops. The results are presented in Figures 2–6.

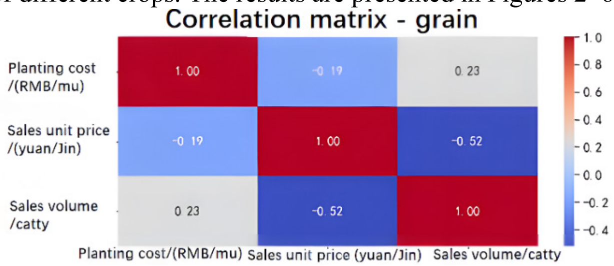

Fig. 2. The heat map of correlation coefficients for the sales volume, sales price, and sales cost of grain

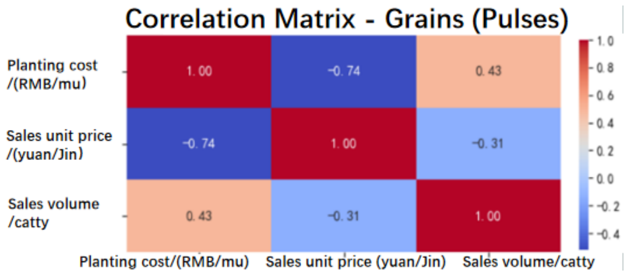

Fig. 3. The heat map of correlation coefficients for the sales volume, sales price, and sales cost of Grains (Pulses)

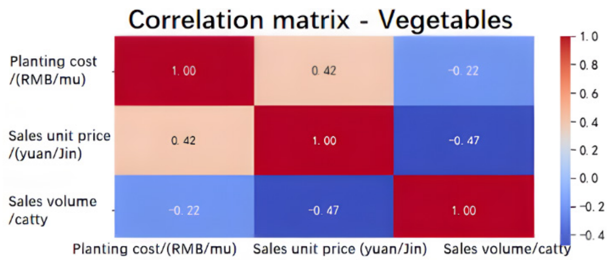

Fig. 4. The heat map of correlation coefficients for the sales volume, sales price, and sales cost of Vegetables

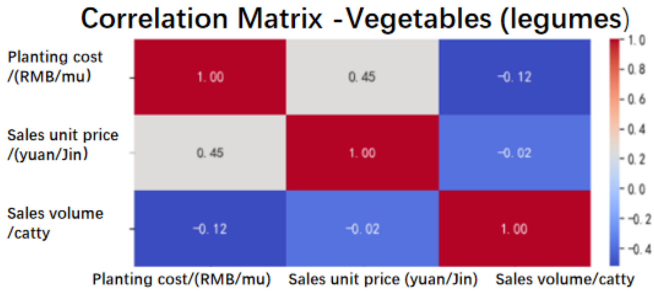

Fig. 5. The heat map of correlation coefficients for the sales volume, sales price, and sales cost of Vegetables (legumes)

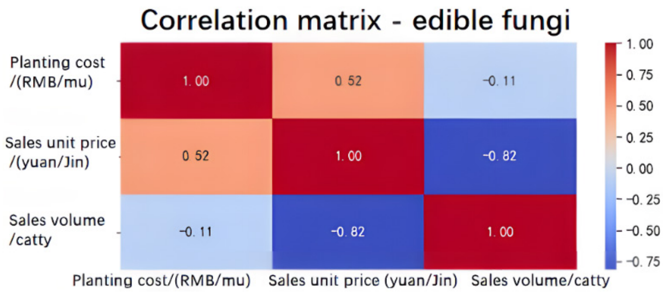

Fig. 6. The heat map of correlation coefficients for the sales volume, sales price, and sales cost of edible fungi

The sales volume of the remaining four crop types, with the exception of vegetable legumes, is inversely proportional to their selling price. In the case of grain crops (including legumes), a negative correlation is observed between selling price and planting cost. In contrast, the other three crop types exhibit a positive correlation. Furthermore, the sales volume of grain crops (including legumes) is positively correlated with planting cost, whereas the other three crop types demonstrate a weak negative correlation. Consequently, an increase in the planting cost of non-grain crops will result in higher prices and reduced sales volumes, which will in turn lead to a decrease in the planting area for non-grain crops and an increase in the planting area for other crops.

Moreover, an analysis of the projected profit curves over the next seven years for scenarios that consider crop substitutability and complementarity revealed that selecting crops with these characteristics presents a greater challenge. Although these choices are designed to achieve higher profits, they also carry an increased risk of market volatility. This strategy is applicable in scenarios where higher potential returns are sought through the use of complex crop combinations, with the understanding that greater annual fluctuations may be an inherent consequence. In general, the former approach results in higher average profits. The results of numerical experiments indicate that, in the majority of cases, the selection of appropriate substitutable crops typically leads to enhanced profitability. As illustrated in Figure 7, the projected profit curves for the subsequent seven years are presented for the two scenarios, one of which incorporates the consideration of crop subplaceability and complementarity, while the other does not.

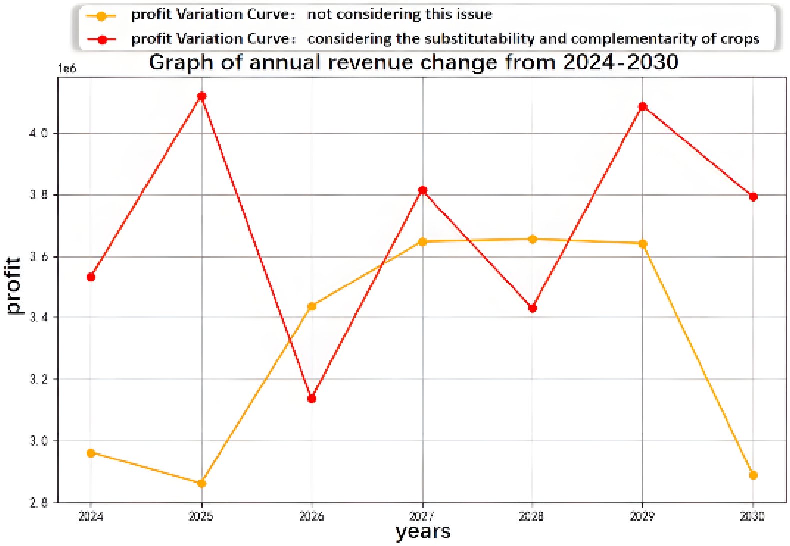

Fig. 7. Comparison chart of returns in the next 7 years considering the substitutability and complementarity of crops and not considering this issue

5 Conclusions

This paper investigates the practical constraints of multi-crop planting and develops a linear programming model to maximize profits. An enhanced genetic algorithm is employed to solve this model, determining the optimal crop planting strategy. By optimizing variable relationships within multiple constraints and mitigating conflicts from excessive restrictions, the enhanced algorithm improves both the model's versatility and computational accuracy. Numerical results indicate that, in comparison to traditional genetic algorithms, the proposed optimization strategy provides greater precision and stability, effectively realizing the optimal crop planting plan.

Acknowledgements

This work was supported by the Natural Science Foundation of Top Talent of SZTU (grant no. GDRC202136).

References

1. Liu, W., Zhang, X., & Zhou, Y. (2020). Dynamic land use planning in agricultural production systems: A case study of wheat-corn rotation in China. Field Crops Research, 248, 107752.

2. Zhang, X., & Li, H. (2018). Optimization of crop planting structure based on resource constraints and economic benefits. Agricultural Systems, 161, 98-107.
3. Xu, B., Li, J., & Zhang, Q. (2021). Sustainable agricultural land use planning: A systematic review and case study of rice production. Sustainability, 13(12), 6567.
4. Al-Madi, N., Faris, H. & Mirjalili, S. (2019) Binary multi-verse optimization algorithm for global optimization and discrete problems. Int. J. Mach. Learn. & Cyber. 10, 3445–3465
5. Mirjalili, S., Song Dong, J., Sadiq, A.S., Faris, H. (2020) Genetic Algorithm: Theory, Literature Review, and Application in Image Reconstruction. In: Mirjalili, S., Song Dong, J., Lewis, A. (eds) Nature-Inspired Optimizers. Studies in Computational Intelligence, vol 811. Springer, Cham.
6. Alhijawi, B., Awajan, A. (2024) Genetic algorithms: theory, genetic operators, solutions, and applications. Evol. Intel. 17, 1245–1256.
7. Kumar, R. (2012). Blending roulette wheel selection & rank selection in genetic algorithms. International Journal of Machine Learning and Computing, 2(4), 365.