DOI: 10.23977/agrfem.2025.080109
ISSN 2616-2202 Vol. 8 Num. 1

Research on Cultivation Optimization Model in Northern China Based on Genetic Algorithm

Wanhui Chen[1], Luoqin Xiang[2], Weixuan Zhang[3]

[1]*School of Control and Computer Engineering, North China Electric Power University, Beijing, 102206, China*
[2]*School of Energy, Power and Mechanical Engineering, North China Electric Power University, Beijing, 102206, China*
[3]*School of Water Resources and Hydroelectric Engineering, North China Electric Power University, Beijing, 102206, China*

Keywords: Genetic Algorithm, Robust Optimization, MAE, Spearman's Rank Correlation, Cultivation Decision Model

Abstract: North China has various terrains and complex farming conditions, including flat and dry land, terraced land, hillside land, watered land and other arable land types, which can be planted with annual or biannual crops. In this paper, based on the cultivation situation in a typical mountainous area in North China in 2023, taking into account the economic benefits, crop characteristics, management convenience and other factors, we use genetic algorithms and robust optimisation to initially establish a cultivation decision model to derive the optimal planting strategy of crops for the next seven years, and then further explore the substitutability between crops, as well as the correlation between the expected sales volume and the sales price by using the MAE and Spearman's correlation coefficient, The model is optimised by exploring the correlation between expected sales volume and sales price, and planting cost through MAE and Spearman correlation coefficient. The model is conducive to improving the utilisation efficiency of arable land, flexibly adjusting the planting strategy according to the market demand, and increasing economic income.

1. Introduction

North China is rich in natural resources, including grasslands, arable land, forests and water bodies, etc., and possesses the topographic features of plains, plateaus and mountains, of which the mountains are mainly distributed in the west and the north, and the development of organic planting industry by the local conditions is of great significance to the sustainable development of the countryside economy [1]. The countryside of a mountainous area in North China has low temperatures all year round, and the flat dry land, terraced land and hillside land are suitable for planting one season of grain crops every year, the watered land is suitable for planting one season of rice or two seasons of vegetables every year, the common greenhouses are suitable for planting one season of vegetables and one season of edible fungi every year, and the intelligent greenhouses are suitable for planting two seasons of vegetables [2-3]. According to the growth law of crops, each crop should not be planted continuously in the same plot, otherwise the yield will be seriously reduced. The legume crop is suitable to be planted at least once in three years because its root symbiotic colonies can increase soil fertility. For ease of ploughing operations and field management, each crop should not be planted too sparsely in each season and the area planted in a single plot should not be too small [4-5].

Shah F's team proposed a variety of innovative solutions such as genetic engineering, precision agriculture, and vertical farming, which can help improve resource use efficiency, however, it is difficult to scale up due to challenges such as technological barriers [6]. Tommaso's team used intercropping systems, constraint programming modelling, and the development of seeding robotic arms to optimise planting, however, the measure suffers from a high degree of complexity in research and implementation as well as a lack of long-term data support and other shortcomings [7]. Maoxun's team incorporated the resilience of the water resource system into the constraint set, developed a multi-objective optimisation model with rice, soybean and maize planting area as decision variables and used the NSGA-III algorithm to optimise and analyse the agricultural cropping structure, which was effective, however, the model complexity was high and required a large amount of data support [8].

To fully consider the production constraints of the site and the limitations of the current study, this paper combines the planting situation of the site in 2023 with the use of genetic algorithms and robust optimisation to initially establish an optimisation model for farming decisions to derive the optimal planting strategy of crops for the next seven years, with a view to maximising profits. On this basis, the MAE was used to study the possible substitutability between crops, and the Spearman correlation coefficient was used to explore the correlation between the expected sales volume and the sales price and planting cost, to further improve the model and derive the optimal planting strategy.

2. Fundamentals of Genetic Algorithms

From the above, it can be seen that cultivation in this place needs to satisfy more constraints at the same time, and the conventional methods are computationally intensive or even unsolvable, so this paper adopts a heuristic algorithm - Genetic Algorithm to solve the problem, and robustly optimises the worst-case scenario of the model to ensure the stability and reliability of the solution, and initially establishes the cultivation decision-making model.

2.1 Genetic Algorithm

A genetic algorithm is an optimisation algorithm based on the principles of natural selection and genetics. It stimulates the process of biological evolution and searches for the optimal solution in the solution space through operations such as selection, crossover and mutation, which is especially suitable for complex multi-peak function optimisation problems and large-scale combinatorial optimisation problems, and the main steps are as follows:

Step 1: Population initialisation: a certain number of individuals are randomly generated to form an initial population of 100, constituting a three-dimensional solution vector.

Step 2: Find the fitness of each individual: bring the solution x into the objective function to find the function value $f(x)$, the purpose is to select the good individuals from the generation population $P(t)$ and inherit the genes to the next generation population $P(t+1)$. A roulette wheel is often used for selection:

Roulette method:

1) Calculate the probability of each being selected $P(x)$, which is proportional to its fitness value.

$$P(x_i) = \frac{f(x_i)}{\sum_{j=1}^{N} f(x_j)} \tag{1}$$

2) Calculate the cumulative probability of each q_i which is the sum of the selection probabilities from the first individual to the current individual.

$$q_i = \sum_{j=1}^{i} P(x_i) \tag{2}$$

3) Randomly generate an array *bet* with elements taking values between 0 and 1 and sort the elements from smallest to largest.

4) If the cumulative probability of the individual, q_i, is greater than *bet(i)*, then the individual is selected, and *bet(i)* to *bet*（*i+1*）. Otherwise, select *bet(i+1)* individual to compare with *bet(i)* until an individual is selected.

5) Repeat Step 4 until the number of individuals equal to the population size is selected (where some individuals are selected more than once).

Step 3: Crossover operation: Among all individuals selected, crossover operation is performed between two individuals. The crossover probability is set to 90% in this paper to determine whether to perform a crossover operation, as shown in Figure 1.

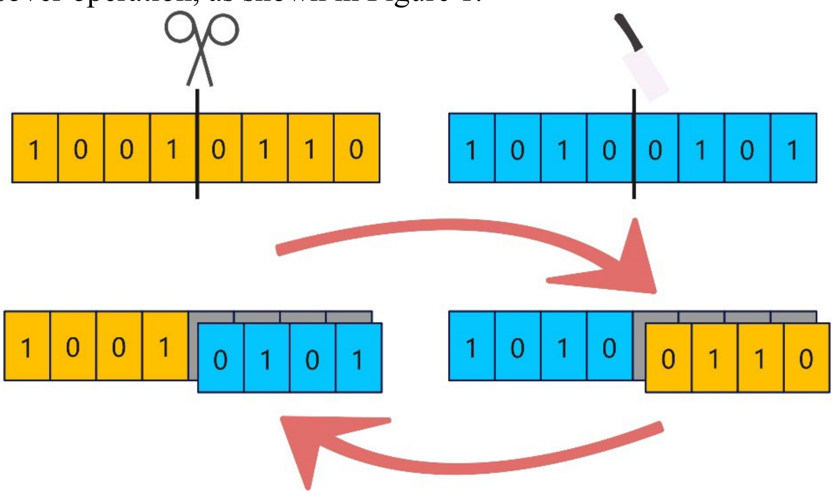

Figure 1: Schematic diagram of crossover operation

Step 4: Mutation operation: Randomly change part of the genes of individuals to increase the diversity of the population, take the inverse of the mutated genes, 0 to 1, 1 to 0. In this paper, the probability of mutation is set to 1%, as shown in Figure 2.

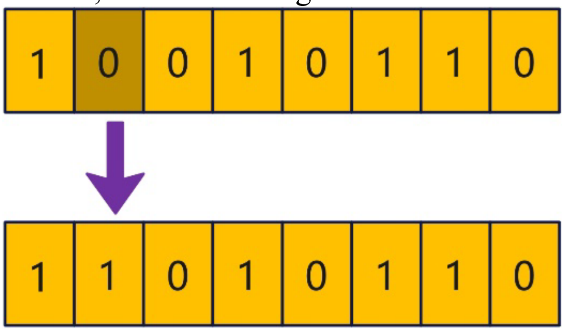

Figure 2: Schematic diagram of mutation operation

Step 5: Iterative cycle: select the new generation of the population (solution set) after the cross mutation, record the optimal individual and the optimal fitness (i.e. max *Qt*) in this round of iteration, and repeat steps 2-4 until the iteration is full 100 times.

2.2 Establishment of the objective function

Pulses are suitable to be planted at least once in three years, so taking three years as a cycle, the planting scheme is considered optimal when the cumulative profit of all plots reaches the maximum value in three years. Of the 41 crops that can be planted on the site, 54 plots are numbered, split each year into two quarters and sequentially arranged as 1-6 quarters. When the actual yield exceeds the expected yield, the excess is sold at 50 per cent of the selling unit price, so that the three-year cumulative profit of the crop is:

$$\sum_{i=1}^{6} Q_t = P_t - C_t \tag{3}$$

where Q_t is the total profit of all plots in season *t*, P_t is the total revenue of all plots in season *t*, and C_t is the total cost of all plots in season *t*.

Total revenue and total cost can be expressed as respectively:

$$P_t = \sum_{i=1}^{41}((\min\{\sum_{j=1}^{54} M_{ijt} \cdot S_{ijt}, N_{it}\} + \left|\sum_{j=1}^{54} M_{ijt} \cdot S_{ijt} - N_{it}\right| \cdot \alpha_p) \cdot B_{it}) \tag{4}$$

$$C_t = \sum_{i=1}^{41}\sum_{j=1}^{54} M_{ijt} \cdot S_{ijt} \cdot D_{it} \tag{5}$$

$$\alpha_p = \begin{cases} 0.5, & if \quad \min\{\sum_{j=1}^{54} M_{ijt} \bullet S_{ijt}, N_{it}\} = N_{it} \\ 0, & if \quad \max\{\sum_{j=1}^{54} M_{ijt} \bullet S_{ijt}, N_{it}\} = N_{it} \end{cases} \tag{6}$$

where M_{ijt} is the yield per unit area of the i-th crop in the j-th plot in the t-th quarter (the corners *ijt* all have the same meaning), S_{ijt} is the planted area, N_{it} is the expected sales volume of the i-th crop in the t-th quarter, B_{it} is the unit sales price, D_{it} is the cost of planting, and α_p is a price factor that can only take on the value of 0 or 1.

For a single-season crop, it is defined as a crop that can only be planted in the first quarter and no crop can be planted on this plot once it is planted in the second quarter of the year.

2.3 Constraint establishment

Variable constraints: based on previous data, wheat and maize average annual growth rates range from 5 to 10 per cent, and expected sales of other crops vary by approximately ±5 per cent per year relative to 2023. Yields per acre will vary ±10% per year due to climate and other factors. Cropping costs will increase on average by about 5 per cent per year due to market conditions. Sales prices for grain crops are generally stable; sales prices for vegetable crops increase by about 5% per year on average. Edible mushroom sales prices can decrease by about 1 to 5 per cent per year. To improve robustness, the most extreme case is considered, i.e., the growth rate is taken to be the lowest, the acreage declines year by year, etc., and the specific formulas are as follows:

$$
\begin{cases}
N_{i(t+1)} = \begin{cases} 1.05N_i & i=6,7 \\ 0.95N_i & i=others \end{cases} \\
M_{i(t+1)} = 0.9M_{it} \\
D_{i(t+1)} = 1.05D_{it} \\
B_{i(t+1)} = \begin{cases} B_{it} & i\in[1,16] \\ 1.05B_{it} & i\in[17,37] \\ 0.95\mathrm{B}_{it} & i\in[38,41] \end{cases}
\end{cases}
\tag{7}
$$

Area constraint: all crops planted per plot cannot exceed the area of that plot:

$$
\sum_{i=1}^{41} S_{ijt} \le Rjt \tag{8}
$$

where Rjt is the area planted in the j-th plot.

Heavy cropping constraint: a crop cannot be planted consecutively in the same plot, for food crops:

$$
\forall i\in[1,16] \quad S_{ijt} \bullet S_{ij(t+2)} = 0 \tag{9}
$$

For smart greenhouse crops:

$$
\forall j\in[51,54] \quad S_{ijt} \bullet S_{ij(t+1)} = 0 \tag{10}
$$

There is no such constraint as the crops grown on watered land and ordinary greenhouses grow different types in different seasons.

Rotation constraints: all plots are planted with legumes at least once in three years and legumes are not allowed to be mixed:

$$
\sum_{t=1}^{6} E_{jt} \ge 1 \tag{11}
$$

where E_{jt} is a 01 variable indicating whether the j-th plot of land is planted with a legume crop in the quarter, 1 if planted, and 0 if not planted.

Management constraints: the crop can not be planted too dispersed, the area is too small, the requirements of ordinary greenhouses and intelligent greenhouses of the same crop must be planted numbered consecutively in two plots:

$$
\forall j\in[35,33] \quad S_{ijt} = S_{i(j+1)t} \tag{12}
$$

3. Results

3.1 Preliminary results and analysis of the model

A Python program is written to solve the model when the variable constraints are not considered, i.e., the expected future sales volume, planting cost, yield per unit area and sales price of various crops are considered to remain stable with respect to the year 2023, and the total production per year and the total profit under the maximum profit of one cycle (three years) can be obtained as shown in Figure 3.

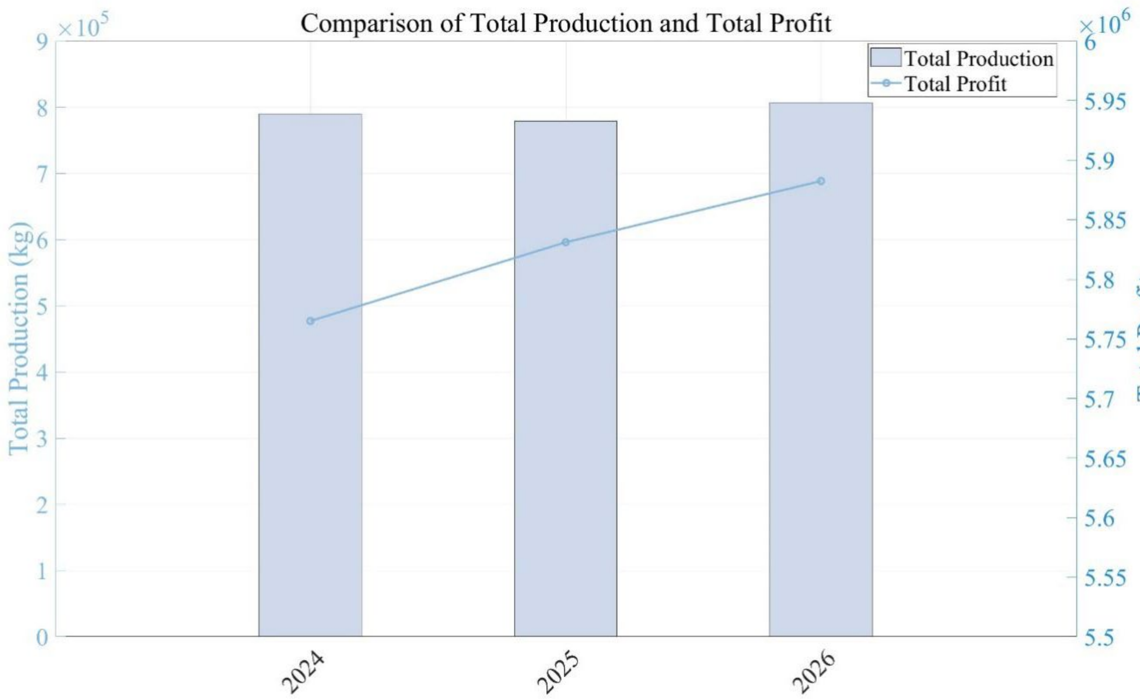

Figure 3: Comparison of Total Production and Total Profit

It can be seen that the total production is about 800,000 kg per year and the total profit is around \$5.8 million per year, which remains relatively stable, which is caused by the variables being constant. If variable constraints are considered, the total production and total profit per year from 2024-2030 are shown in Figure 4.

The total production and total profit per year decrease every year, and after 6 years the total production decreases by 53.3% and the total profit decreases by 59.2%, which is in line with the performance of the extreme case.

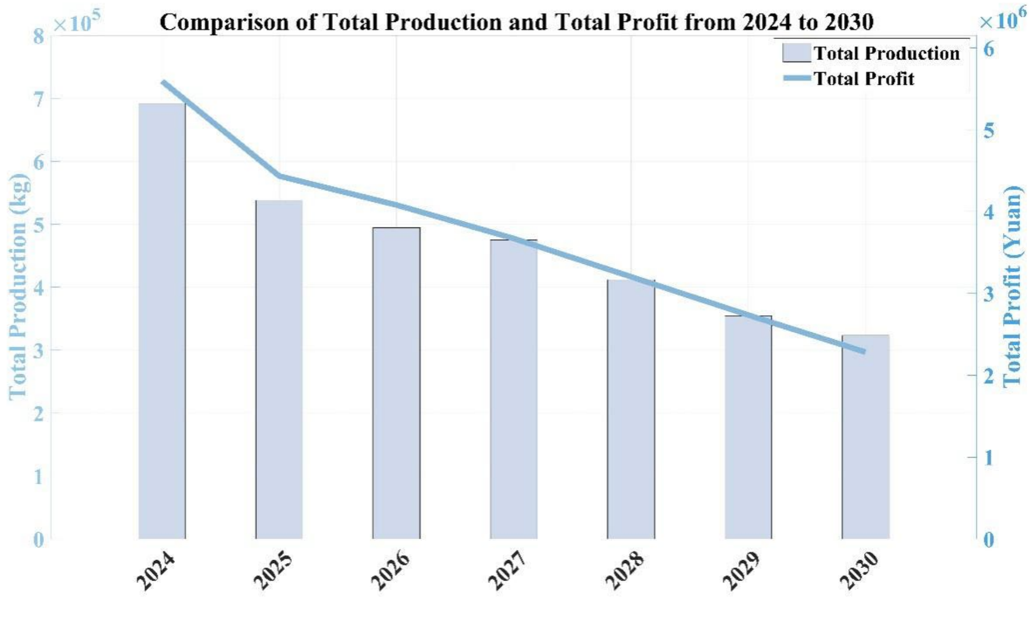

Figure 4: Comparison of Total Production and Total Profit from 2024 to 2030

3.2 Model refinement

3.2.1 Potential substitutability between crops

Since the eight crops grown on the site, such as rice, Chinese cabbage and edible mushrooms, have more stringent requirements on plot type and planting season, their substitutability is not considered for the time being, and at the same time, there are large differences between single-season crops and two-season crops in terms of the living environment, market adaptability, etc., so single-season crops and two-season crops are discussed separately.

The mean absolute error (MAE) is used to describe the deviation of a parameter between the two crops in terms of the expected sales volume N. The mean absolute error (MAE) is the average of the mean absolute errors of the two crops:

$$MAEN(x,y)=\frac{1}{3}\sum_{n=1}^{3}\left|N(x_n)-N(y_n)\right| \tag{13}$$

where $MAEN(x,y)$ is the mean absolute error of crop *x* over crop *y*, $N(x_n)$ is the expected sales volume of crop *x* in the n-th plot type, and the sales price B is the same as the cost of cultivation D.

Combining the above three parameters, the mean absolute error average is introduced to describe the substitutability between the two crops:

$$\overline{MAE}(x,y)=\frac{1}{3}[MAE(x,y)+MAEB(x,y)+MAED(x,y)] \tag{14}$$

Using the 2023 data for the two crops as an example, a heat map was used to visualise the substitutability between the different parameters of the two crops, with lighter colours indicating greater substitutability between the two crops, which is shown in Figure 5.

Observation of the heat map reveals that pumpkin and sweet potato have larger mean absolute errors than other crops, suggesting that these two crops are difficult to substitute with other crops and that their economic and production attributes are more independent. The mean absolute errors of legumes such as red beans and soya beans are smaller than those of other crops, indicating that these crops are good substitutes for other crops.

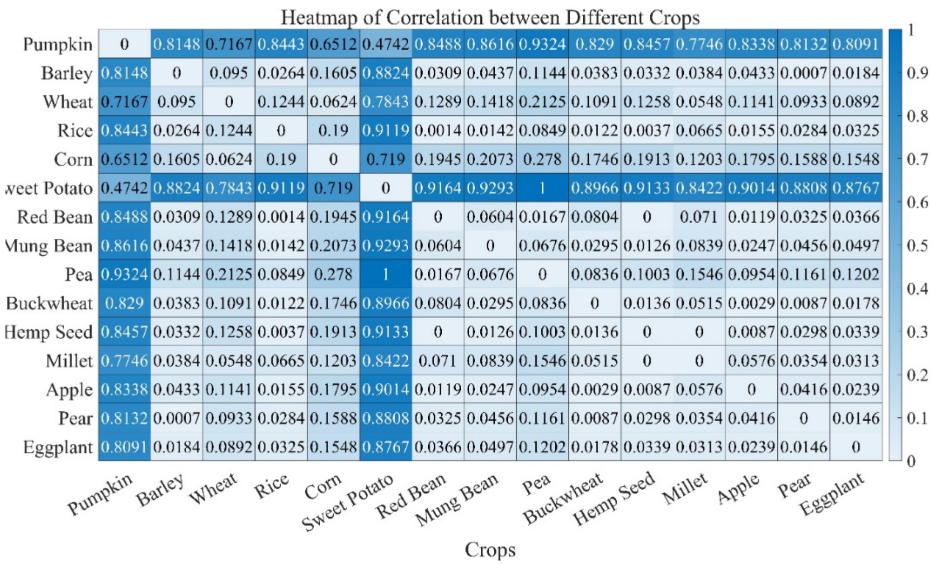

Figure 5: Heatmap of Correlation between Different Crops

3.2.2 Correlation between model variables

The correlation between expected sales volume, sales price, and planting cost was investigated by Spearman's correlation coefficient, and the correlation matrix is shown in Figure 6.

From Figure 6, it can be found that the planting cost has a strong positive correlation with the selling unit price, an average level of positive correlation with the expected sales volume, and no correlation between the selling unit price and the expected sales volume.

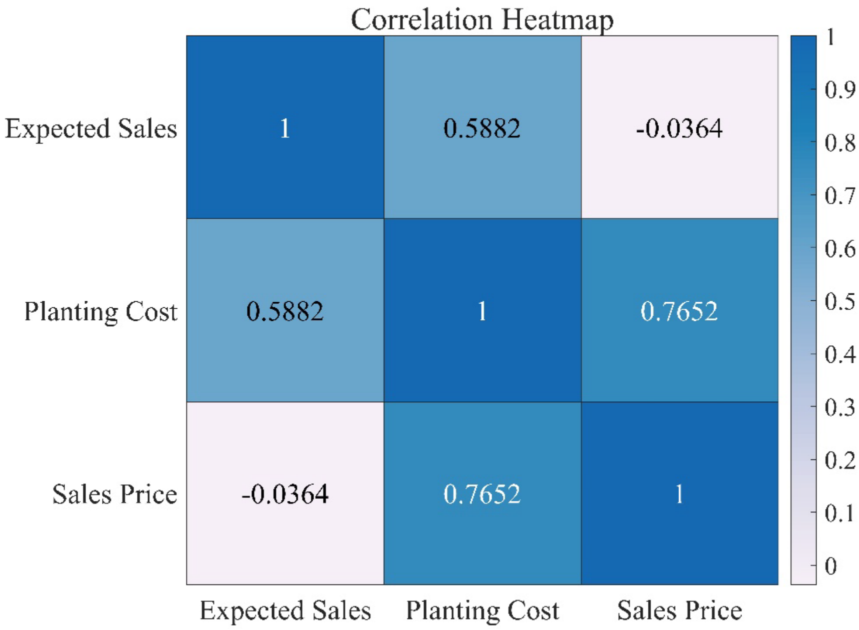

Figure 6: Correlation Heatmap

3.2.3 Synthesis

The correlation between planting costs unit sales price and expected sales volume was connected using a cause-and-effect relationship, whereby for every 2 per cent change in planting costs, there was a 1 per cent change in unit sales price and a 0.5 per cent change in expected sales volume.

$$\begin{cases} V_B = \frac{1}{2} \cdot V_D \\ V_N = \frac{1}{4} \cdot V_D \end{cases} \tag{15}$$

where V_B , V_N and V_D are the rates of change in unit sales price, expected sales volume and planting costs, respectively.

For substitutable crops, a penalty factor is applied to their planting costs when they increase by 2 per cent to increase the probability of planting the corresponding substitute crop.

$$DV_{it} = D_{it} \cdot \alpha^{\frac{V_B}{2\%}} \tag{16}$$

where α is the penalty factor and DV_{it} is the virtual planting cost of the i-th crop in the quarter, which is only involved in judging whether to plant or not to increase the probability of planting, but not in the final profit calculation. Similarly, for land planted with legume crops in the previous quarter, the planting cost of vegetable crops will set a reward factor to increase the probability of planting them in the next quarter.

3.3 Results and analysis of the final model

The relationship between profit and rate of change over time before and after model optimisation is shown in Figure 7.

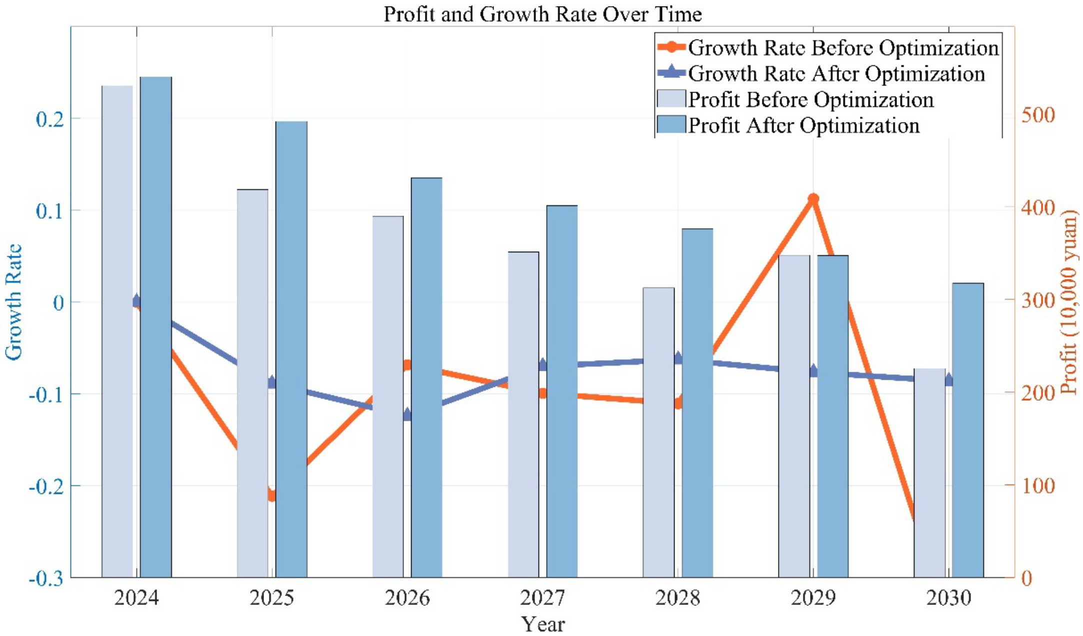

Figure 7: Profit and rate of change before and after optimization

In the first year of the pre-optimisation period, the rate of change surged to 21.11 per cent and then declined abruptly to 6.84 per cent, which was an ‘oscillatory period’ in the market's profit response to agricultural cultivation, suggesting that profits were weakly resilient to shocks to agricultural cultivation. As time goes by, the rate of change still increases gradually, indicating that profits are gradually heading towards destabilisation and collapse with the deterioration of agricultural cultivation. After optimisation, the agricultural cultivation structure is adjusted according to the substitutability between crops. It is shown that the final model after optimisation is still robust and highly profitable even in the most extreme cases, which will not be so extreme in reality, so the planting strategy given by the model can maximise profits.

4. Conclusions

In this paper, taking economic benefits, crop characteristics, management convenience and other factors into consideration, the planting scheme of three years as a planting cycle is reasonably proposed, and the optimal planting strategy for the next seven years is derived from the preliminary establishment of the farming decision model using genetic algorithm and robust optimisation. Further exploring the substitutability between crops through MAE, it is concluded that a variety of legumes can be good substitutes for other crops; analysed by Spearman's correlation coefficient, it is found that the planting cost has a strong positive correlation with the unit price of the sale and has a generally positive correlation with the expected sales volume, and there is no correlation between the unit price of the sale and the expected sales volume.

After considering substitutability and correlation, the final model shows that the first year of the 'oscillatory period' of market profit is clearly stabilised, with the rate of change gradually decreasing to 6.27% over the first four years, and profits gradually levelling off. Even in the most extreme cases, the model is still robust and highly profitable, while the actual situation will not be so extreme, so the planting strategy given can maximise profits.

In this paper, the more typical and decisive influencing factors in production are selected as the limiting conditions, while more unpredictable or difficult-to-intervene situations will be encountered in the actual production life, and more required conditions can be further considered to get a planting strategy more in line with the actual production and improve the applicability.

References

[1] Zhang Z, Zhang H, Liu X, et al. Dynamic change characteristics of comprehensive regionalization of natural resources in North China[J]. Geological Survey of China, 2021, 8(02): 92-99.

[2] De la Cruz V Y V, Cheng W, Tawaraya K. Yield gap between organic and conventional farming systems across climate types and sub-types: A meta-analysis[J]. Agricultural Systems, 2023, 211: 103732.

[3] Yang Q, Wu M, Al Mamun A, et al. Organic food production among Chinese urban botanists[J]. Humanities and Social Sciences Communications, 2024, 11(1): 1-14.

[4] Shah K K, Modi B, Pandey H P, et al. Diversified crop rotation: an approach for sustainable agriculture production[J]. Advances in Agriculture, 2021, 2021(1): 8924087.

[5] LING N, WANG Q, YANG X, et al. Control of Fusarium wilt of watermelon by nursery application of bio-organic fertilizer[J]. Journal of Plant Nutrition and Fertilizers, 2009, 15(5): 1136-1141.

[6] Shah F, Wu W. Soil and crop management strategies to ensure higher crop productivity within sustainable environments[J]. Sustainability, 2019, 11(5): 1485.

[7] Adamo T, Colizzi L, Dimauro G, et al. Crop planting layout optimization in sustainable agriculture: A constraint programming approach[J]. Computers and Electronics in Agriculture, 2024, 224: 109162.

[8] Li M, Zhang Z, Liu D, et al. Optimization of agricultural planting structure in irrigation areas of Heilongjiang province considering the constraints of water resource system resilience[J]. Journal of Cleaner Production, 2024, 434: 140329.