Optimization of Crop Planting Strategies Based on Linear Programming and Monte Carlo Simulation

Hengtian Ma[a,*]

School of Finance, Renmin University of China, Beijing, China

[a]2022200213@ruc.edu.cn

*Corresponding author

Keywords: Crop planting strategy, Linear programming, Monte Carlo simulation, Agricultural optimization, Market volatility

Abstract: This study presents a comprehensive analysis and optimization of crop planting strategies for a village in the North China mountainous area, utilizing a multi-year planning model that integrates linear programming and Monte Carlo simulation. The research aims to maximize economic benefits by considering various factors such as land types, planting seasons, crop yields, and market demands. The dataset encompasses detailed information on arable land, greenhouses, and crop cultivation for the year 2023, which includes planting areas, seasons, yields per acre, planting costs, and sales prices. The study employs a multi-stage planning model that establishes an objective function to maximize profits, considering sales revenue and planting costs. The model is subject to constraints such as land availability, crop yield limitations, scale of planting, intercropping practices, and crop rotation requirements. A key aspect of the model is the inclusion of legumes in the planting plan every three years to maintain soil fertility. To address market uncertainties, a Monte Carlo simulation is applied to account for fluctuations in crop yields, cost prices, sales prices, and sales volumes. This simulation assists in deriving optimal planting plans under varying market conditions, ensuring adaptability and risk mitigation. The simulation process involves parameter initialization, random variation generation within specified ranges, and repeated trials to achieve statistically significant results.

1. Introduction

With the growth of the global population and changes in consumption patterns, sustainable development in the agricultural sector has become increasingly important. Particularly in the North China region, the unique geographical and climatic conditions present numerous challenges to crop cultivation, including limited land resources, seasonal climate changes, and the constant fluctuations in market demand. To enhance the economic benefits and sustainability of crop cultivation, it is necessary to optimize existing planting strategies to adapt to these challenges[1, 2].

This study focuses on a village in the North China mountainous area, collecting and analyzing data on arable land, greenhouses, and crop cultivation from the year 2023 to explore the optimization of crop planting strategies [3]. The core of the research lies in how to use scientific methods to reasonably allocate land resources, select appropriate crop planting patterns, and develop effective market response strategies, thereby maximizing farmers' economic benefits while ensuring crop yield and quality. In terms of methodology, this study employs two main tools: linear programming and Monte Carlo simulation [4, 5]. Linear programming, as a mathematical optimization method, helps us find the best combination of crop planting and the allocation of planting areas under a series of constraints. Monte Carlo simulation allows us to consider market uncertainty by simulating different market scenarios and evaluating the performance of various planting strategies under different conditions, thus providing a more flexible and robust basis for decision-making [6].

  DOI: 10.25236/icemeet.2024.031

2. Data Analysis

2.1. Dataset

The dataset we have includes detailed information on the arable land and greenhouses of a village in the North China mountainous area, as well as the crop cultivation data for the year 2023. The details of the land and greenhouses encompass the area of different types of plots, applicable planting methods, and restrictions on planting seasons. The plots are categorized as flat dry land, terraced fields, sloping land, and irrigated land, with the flat dry land, terraced fields, and sloping land typically only allowing for one crop season per year, while irrigated land offers flexibility, capable of supporting either one or two crop seasons. Additionally, there are ordinary greenhouses and smart greenhouses, which provide more planting flexibility; for instance, ordinary greenhouses can cultivate two seasons of crops annually, while smart greenhouses utilize solar energy to regulate the temperature inside, ensuring the growth of crops even during the winter.

The 2023 village crop cultivation data involves detailed information including planting area, planting season, yield per acre, planting cost, and sales price per unit. Integrating these datasets, it can be deduced that the village's crop planting strategy needs to take into account factors such as the seasonality of crops, types of plots, planting costs, and potential sales markets. Moreover, it is necessary to consider the substitutability and complementarity between different crops, as well as the interrelation between sales volume, price, and cost, in order to formulate a planting plan that maximizes economic benefits.

2.2. Data Analysis

We need to utilize the assumption that "farmers are rational in their planting, meaning that in 2023, the yield of crops is basically the same as market demand" to calculate the production and sales volume of each crop, and then calculate the total profit of each agricultural product. To make the results more credible, we need to conduct a correlation analysis of the planting area and yield per acre of each crop, as shown in Figure 1, most crops' planting area and yield per acre do not have a correlation, and only a few non-bean crops have a positive relationship between planting area and yield per acre, which may be related to the scale of the crop.

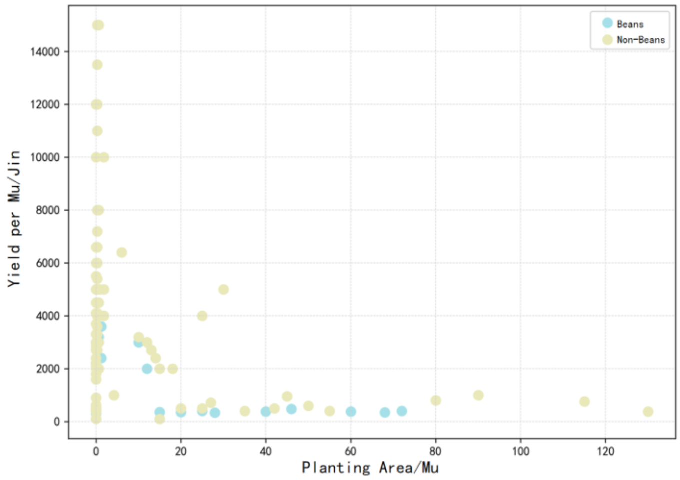

Figure 1 The Relationship between Planting Area and Yield per Mu

Then, we statistically analyze the crops of different types separately. First, we use box plots to depict the distribution of sales prices for different types of crops, as shown in Figure 2. It can be observed that the sales price of edible fungi is the highest, while the sales price of vegetables is relatively lower. Next, we use bar charts to illustrate the planting area of crops in different seasons, as shown in Figure 3. It is known that only in greenhouses can edible fungi be cultivated, while flat dry land, terraces, and hillsides are only suitable for single-season grain crops.

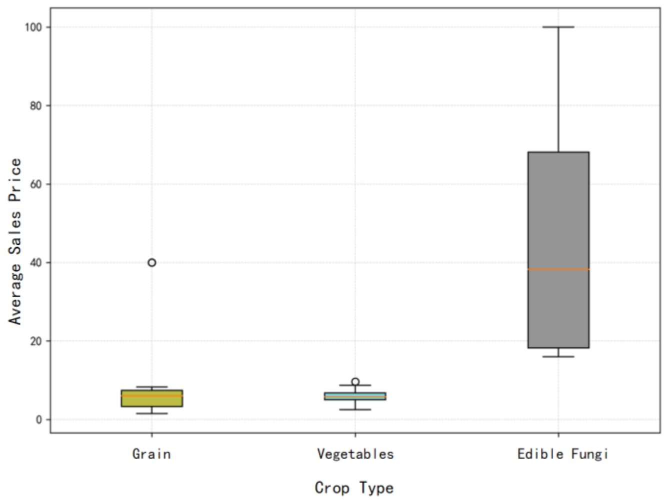

Figure 2 The Relationship between Planting Area and Yield per Mu

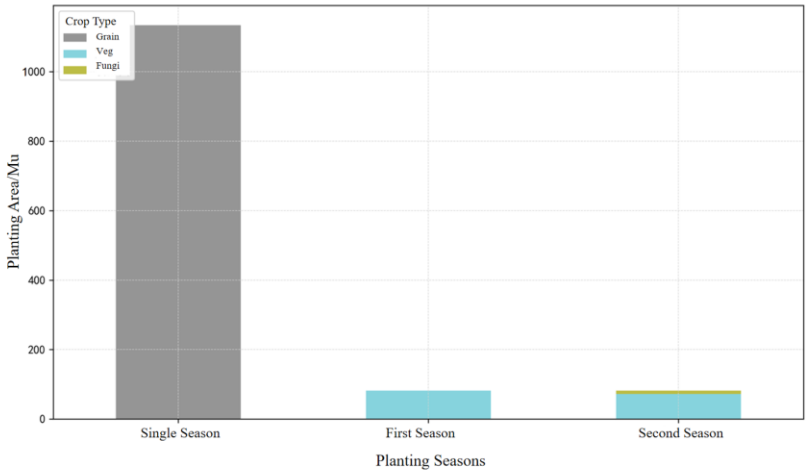

Figure 3 Planting Area of Crops in Different Seasons

Finally, we calculated the average profit and total profit for each type of crop and compared them, as shown in Figure 4. From the chart, we can see that for total profit, grain is the highest and edible fungi are the lowest; for average profit, edible fungi are the highest and vegetables are the lowest.

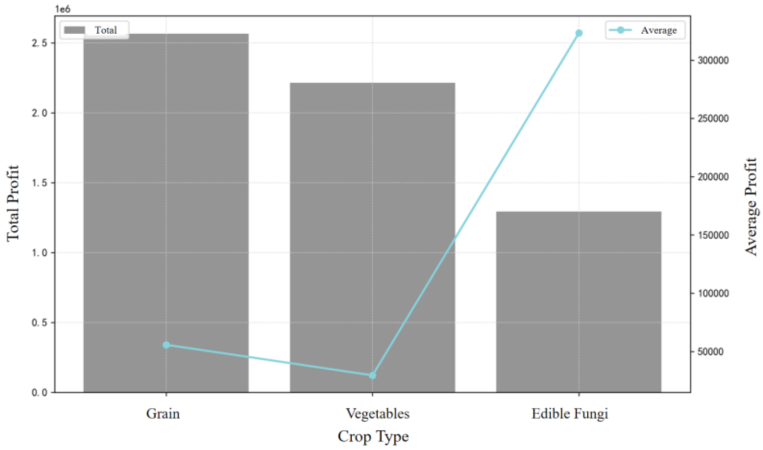

Figure 4 Total and Average Profits of Crops

3. Multi-Year Crop Planting Planning Optimization Model

Assuming that the data from 2023 remains unchanged, we have designed a planting planning model for the years 2024-2030. Under two scenarios (either wasting the surplus or selling the excess at half price), we have determined the optimal planting plan for crops such as grains, vegetables, and edible fungi through linear programming solutions.

3.1. Multi-stage Planning Model

First, we establish the objective function as follows:

$$f_i = \sum_{j=1}^{M} x_{ij} y_{ij} a_{ij} \tag{1}$$

In the following formulas, $i = 1,2,\ldots,15$ represents 15 types of crops such as soybeans and wheat, and $j = 1,2,\ldots,26$ represents the types of land sequentially. xij represents the area of crop i planted on land j, and yij is a binary variable, where 1 indicates that crop i is planted on land j, and 0 indicates that crop i is not planted on land j. a_{ij} represents the unit yield of crop i on land j. b_{ij} represents the unit planting cost of crop i on land j. $c_j$represents the selling price per unit of crop. d_i represents the estimated sales volume of crop$i$. e_j represents the maximum planting area for land j. f_i represents the total production volume of crop i, as follows:

The sales revenue for crop i is denoted as g_i:

$$g_i = f_i c_i \tag{2}$$

The planting cost is cost, that is:

$$cost = \sum_i^N \sum_j^M x_{ij} y_{ij} b_{ij} \tag{3}$$

$$\max Z = \sum_i^N g_i - cost \tag{4}$$

The establishment of the constraint function is as follows:

Condition 1: The area planted with crops must not exceed the area of arable land available for cultivation.

$$\sum_i^N x_{ij} y_{ij} \le e_j, j = 1,2 \cdots ,26 \tag{5}$$

Condition 2: The total yield of the crops must not exceed their sales volume.

$$f_i = \sum_{j=1}^{M} x_{ij} y_{ij} a_{ij} \le d_i, i = 1,2, \cdots 15 \tag{6}$$

Condition 3: Each crop planted should have a certain scale.

From our dataset, it is known that all arable land is utilized, hence the area planted on each plot of land must not be less than the minimum value, and the scale of each crop planted should be realistic and meet certain standards.

$$\begin{cases} x_{ij} \ge 35, j = 1,2,\cdots,6 \\ x_{ij} \ge 20, j = 7,8,\cdots,20 \\ x_{ij} \ge 13, j = 21,22,\cdots 26 \end{cases} \tag{7}$$

Condition 4: The intercropping of crops should be appropriate.

For crops, considering the practical situation, having too many types of crops planted is not conducive to planting and management. However, taking into account that appropriate intercropping may potentially increase yield, it is stipulated here that a maximum of three types of crops can be planted on a single piece of arable land.

$$\sum_i^N y_{ij} \le 3, j = 1,2 \cdots ,26 \tag{8}$$

Condition 5: The same crop should not be planted continuously on the same piece of land.

Continuous replanting of the same crop on the same land can lead to a reduction in yield due to soil fertility issues. Therefore, let the indicator variable $z_{ij}(t)$ represent the situation of crop i on j in year t, that is:

$$z_{ij}(t) = \begin{cases} 0, \text{fallowing} \\ 1, \text{planting} \end{cases} \tag{9}$$

Where, $t = 2023,2024,\cdots 2030$.

Let the indicator variable $zz_{ij}(t)$ represent whether crop i can be planted on land j in year t, with the following restrictions:

$$z_{ij}\begin{cases} zz_{ij}(t+1) = 0, z_{ij}(t) = 1 \\ zz_{ij}(t+1) = 1, z_{ij}(t) = 0 \end{cases} \tag{10}$$

Substituting the aforementioned formula into Conditions 1 and 2, we obtain:

$$\sum_i^N x_{ij} y_{ij} zz_{ij}(t) \leq e_j, j = 1,2,\cdots,26 \tag{11}$$

$$f_i = \sum_{j=1}^M x_{ij} y_{ij} zz_{ij}(t) \cdot a_{ij} \leq d_i, i = 1,2,\cdots 15 \tag{12}$$

Condition 6: Legumes should be planted at least once every three years.

The rhizobia in the root nodules of legumes are beneficial for nitrogen fixation in the soil, which is advantageous for the growth of other crops. As per the problem statement, legumes should be planted at least once every three years. Therefore, let the indicator variable $q_{ij}(t)$ represent the situation of legume crop i on land j in year t.

Let the indicator variable $qq_{ij}(t)$ represent the necessary condition for legume crop i to be planted on land j in year t.

Then $q_{ij}(t)$ and $qq_{ij}(t)$ satisfy the following relationship:

$$q_{ij}(t) + q_{ij}(t+1) = qq_{ij}(t+2) \tag{13}$$

$qq_{ij}(t)$ can be used to constrain y_{ij}.

$$y_{ij} = \begin{cases} 0, qq_{ij}(t) = 1 or\ 2 \\ 1, qq_{ij}(t) = 0 \end{cases}, i = 1,2,\cdots 15, j = 1,2,\cdots,26 \tag{14}$$

Integrating the above models allows us to derive the final predictive model. Here, we discuss two scenarios separately.

For the year 2024:

$$\max Z = \sum_i^N \left[\sum_{j=1}^M x_{ij} y_{ij} zz_{ij}(2024) \cdot a_{ij}\right] \cdot c_{ij} - \sum_i^N \sum_j^M x_{ij} y_{ij} zz_{ij}(2024) \cdot b_{ij} \tag{15}$$

For the year 2025 and beyond:

$$\max Z = \sum_i^N \left[\sum_{j=1}^M x_{ij} y_{ij} zz_{ij}(t) \cdot a_{ij}\right] \cdot c_{ij} - \sum_i^N \sum_j^M x_{ij} y_{ij} zz_{ij}(t) \cdot b_{ij} \tag{16}$$

The main difference between the planning of vegetables, rice, and edible fungi and that of grain lies in the types of crops and their cultivation conditions. Vegetables, rice, and edible fungi have specific seasonal and environmental requirements for planting. For instance, vegetables are typically grown in irrigated fields and greenhouses, rice, as a single-season crop, is mainly cultivated in irrigated land, and edible fungi are predominantly grown in greenhouses. Moreover, the sales prices, yields per acre, and planting costs of these crops may differ from those of grain crops. Therefore, when establishing an optimization model, it is necessary to consider the unique growth cycles and market conditions of these crops.

3.2. Model Solving

To simulate real-world conditions, we assume that the sales volume of each crop will fluctuate within a range of plus or minus 10%. Our strategy is to set the target sales volume for each year based on the highest increase from the previous year's sales, prioritizing overproduction over the risk of customers being unable to purchase agricultural products and ensuring that the people do not face food shortages. In the objective function, we use the maximum decrease value as the actual sales volume, incorporating it to consider the maximum reasonable amount of loss within an acceptable range. For the nth year in the future, we need to evaluate the crops planted in the last three years, ensuring that every plot of land has been sown with legumes at least once every three years to maintain soil fertility for crop cultivation. Since the ultimate goal of the model is to maximize profits, we can establish two models: one where "excess production is left unsold and wasted directly," and another where "the excess is sold at 50% of the sales price."

Below is a categorized discussion of the situation. For the sake of brevity, let $f_j(x) = \sum_{i=1}^{M}\sum_{t=1}^{2} r_j \cdot A_i \cdot x_{i,j,t}$, representing the production of crop j, where d_j denotes the sales volume of product j.

For scenario one, the excess is directly wasted.

$$g_i = f_i(t) - d_i \tag{17}$$

For scenario two, the excess is sold at 50% of the sales price.

$$g_i = d_i c_i + \frac{1}{2}[f_i(t) - d_i] \cdot c_i \tag{18}$$

In summary, we can conclude that:

$$g_i = f_i(t)c_i + \frac{1}{2}\max[0, f_i(t) - d_i] \cdot c_i \tag{19}$$

Regarding the solution for grain crops, there are some changes in the objective function.

For the year 2024:

$$\max Z = \sum_i^N f_i(2024) \cdot c_{ij} + \frac{1}{2}\max[0, f_i(2024) - d_i] - \sum_i^N \sum_j^M x_{ij} y_{ij} zz_{ij}(2024) \cdot b_{ij} \tag{20}$$

$$\text{s.t.}\begin{cases} \sum_i^N x_{ij} y_{ij} zz_{ij}(2024) \leq e_j, j = 1,2,\cdots,26 \\ f_i = \sum_{j=1}^{M} x_{ij} y_{ij} zz_{ij}(2024) \cdot a_{ij} \\ \begin{cases} x_{ij} \geq 35, j = 1,2,\cdots,6 \\ x_{ij} \geq 20, j = 7,8,\cdots,20 \\ x_{ij} \geq 13, j = 21,22,\cdots 26 \end{cases} \\ \sum_i^N y_{ij} \leq 3, j\ 1,2\cdots,26 \\ \begin{cases} zz_{ij}(2024) = 0, z_{ij}(2023) = 1 \\ zz_{ij}(2024) = 1, z_{ij}(2023) = 0 \end{cases} \end{cases} \tag{21}$$

For the year 2025 and beyond:

$$\max Z = \sum_i^N f_i(t) \cdot c_{ij} + \frac{1}{2}\max[0, f_i(t) - d_i] - \sum_i^N \sum_j^M x_{ij} y_{ij} zz_{ij}(t) \cdot b_{ij} \tag{22}$$

$$\text{s.t.}\begin{cases} \sum_i^N x_{ij} y_{ij} zz_{ij}(t) \leq e_j, j = 1,2,\cdots,26 \\ f_i = \sum_{j=1}^{M} x_{ij} y_{ij} zz_{ij}(t) \cdot a_{ij} \leq d_i, i = 1,2,\cdots 15 \\ \begin{cases} x_{ij} \geq 35, j = 1,2,\cdots,6 \\ x_{ij} \geq 20, j = 7,8,\cdots,20 \\ x_{ij} \geq 13, j = 21,22,\cdots 26 \end{cases} \\ \sum_i^N y_{ij} \leq 3, j\ 1,2\cdots,26 \\ \begin{cases} zz_{ij}(t-1) = 0, z_{ij}(t) = 1 \\ zz_{ij}(t-1) = 1, z_{ij}(t) = 0 \end{cases} \\ q_{ij}(t-2) + q_{ij}(t-1) = qq_{ij}(t) \\ y_{ij} = \begin{cases} 0, qq_{ij}(t) = 1 \text{ or } 2 \\ 1, qq_{ij}(t) = 0 \end{cases}, i = 1,2,\cdots 15, j = 1,2,\cdots,26 \end{cases} \tag{23}$$

The solution for vegetables, rice, and edible fungi crops is as follows:

For the year 2024:

$$\max Z = \sum_i^N f_i(2024) \cdot c_i + \frac{1}{2}\max[0, f_i(2024) - d_i] - \sum_{k=1}^{2}\sum_i^N \sum_j^M x_{ij} y_{ij} z_{ij}(2024) \cdot b_{ijk} \tag{24}$$

$$\text{s.t.}\begin{cases} \sum_i^N x_{ijk}y_{ijk}zz_{ij}(2024) \le e_j, j=1,2,\cdots,26, k=1,2 \\ f_{ik} = \sum_{j=1}^M x_{ijk}y_{ijk}zz_{ij}(2024) \le d_{ik}, i=1,2,\cdots 15, k=1,2 \\ \begin{cases} x_{ijk} \ge 6, j=1,2,\cdots,8, k=1,2 \\ x_{ijk} \ge 0.3, j=9,10,\cdots,24, k=1,2 \\ x_{ijk} \ge 0.3, j=25,26, k=1,2 \end{cases} \\ \sum_{i=1}^N y_{ijk} \le 3, j=1,2,\cdots,26, k=1,2 \\ x_{ij1} = y_{ij1} = a_{ij1} = 0, i \in I, j \in J \\ b_{ij1} = +\infty, i \in I, j \in J \\ b_{ij2} = +\infty, i \in I, j \in J \end{cases} \tag{25}$$

For the year 2025 and beyond:

$$\max Z = \sum_i^N f_i(t)\cdot c_i + \frac{1}{2}\max[0, f_i(t) - d_i] - \sum_{k=1}^2 \sum_i^N \sum_j^M x_{ij}y_{ij}z_{ij}(t)\cdot b_{ijk} \tag{26}$$

$$\text{s.t.}\begin{cases} \sum_i^N x_{ijk}y_{ijk}zz_{ij}(t) \le e_j, j=1,2,\cdots,26, k=1,2 \\ f_{ik} = \sum_{j=1}^M x_{ijk}y_{ijk}zz_{ij}(t) \le d_{ik}, i=1,2,\cdots 15, k=1,2 \\ \begin{cases} x_{ijk} \ge 6, j=1,2,\cdots,8, k=1,2 \\ x_{ijk} \ge 0.3, j=9,10,\cdots,24, k=1,2 \\ x_{ijk} \ge 0.3, j=25,26, k=1,2 \end{cases} \\ \sum_{i=1}^N y_{ijk} \le 3, j=1,2,\cdots,26, k=1,2 \\ \begin{cases} zz_{ij}(t) = 0, z_{ij1}(t-1) + z_{ij2}(t-1) = 1 \\ zz_{ij}(t) = 1, z_{ij}(t-1) + z_{ij2}(t-1) = 0 \end{cases} \\ q_{ij1}(t-2) + q_{ij2}(t-2) + q_{ij1}(t-1) + q_{ij2}(t-1) = qq_{ij}(t) \\ y_{ijk} = \begin{cases} 0, qq_{ij}(t) \ge 1 \\ 1, qq_{ij}(t) = 0 \end{cases}, i=1,2,\cdots 15, j=1,2,\cdots,26, k=1,2 \\ x_{ij1} = y_{ij1} = a_{ij1} = 0, i \in I, j \in J \\ b_{ij1} = +\infty, i \in I, j \in J \\ x_{ij2} = y_{ij2} = a_{ij2} = 0, i \in I, j \in J \\ b_{ij2} = +\infty, i \in I, j \in J \end{cases} \tag{27}$$

Based on the constraint functions, we can calculate the results for the two scenarios using Python.

4. Planting Strategies in Volatile Markets

4.1. Model Optimization

Building on the aforementioned model, we have incorporated considerations for market uncertainties. The purpose is not only to devise an optimal planting plan under known stable conditions but also to dynamically adjust planting strategies by simulating and analyzing potential future market fluctuations, such as changes in prices and costs. This approach aims to maximize profits in a market environment with higher uncertainty while effectively managing and mitigating risks.

Compared to the original model, the difference lies in the fact that the unit price of crop j, the yield per unit area of crop j, and the unit cost of planting crop j on plot i are no longer constants but variables. The biggest challenge in the solution process is the uncertainty brought to the calculations by changes in crop yields per acre due to weather and fluctuations in cost and sales prices influenced by market factors. To address this issue, we will employ Monte Carlo simulation.

4.2. Monte Carlo Simulation

The principle of Monte Carlo simulation is based on random sampling and probability statistics, used to solve complex mathematical or physical problems, especially when they cannot be solved directly. Its core idea is to approximate the solution of a problem through a large number of random trials. This method simulates probabilistic processes by relying on the generation of a vast number of random numbers to represent possible input data, and then performing calculations and statistical analysis on these data to obtain an approximate solution.

A typical application of Monte Carlo simulation is in multi-dimensional integration or complex financial derivatives pricing. In these cases, traditional analytical methods are difficult or impossible to solve, but through a large number of random simulations, an approximation to the solution can be achieved. Additionally, the Monte Carlo method can be used in fields like physics, engineering, and biology to solve complex system behaviors. Monte Carlo simulation is a powerful tool for obtaining approximate solutions by simulating uncertainty and randomness. It provides effective solutions for problems that are complex and difficult to analyze directly.

4.3. Model Solving

We utilize Monte Carlo simulation to model the acreage yield, cost price, sales price, and sales volume of crops from 2024 to 2030. By examining the dataset we have collected, it is observable that these parameters fluctuate within a certain range each year. Through the use of multiple Monte Carlo simulations, we can derive the optimal planting plan.

The solution process is as follows:

1) Parameter Initialization: We need to initialize the crop yield per acre, cost price, sales price, and sales volume based on the given data and actual conditions. Subsequently, the range of variation for these initialized data is determined, such as "The sales price of vegetable crops tends to increase, with an average annual growth of about 5%."

2) Conducting Monte Carlo Simulation: First, we will generate random variation values for the yield per acre, cost price, selling price, and sales volume of crops for each year, with these variations being randomly generated within specific restricted ranges. Then, using these randomly generated data, we will calculate the total profit according to the previously established model and record the corresponding planting methods. To ensure the reliability of the results, we will repeat the Monte Carlo simulation experiments 1000 times per year to obtain statistical results under different conditions. Through this method, we can gain a more comprehensive understanding of the potential economic benefits of different planting strategies under various market conditions.

3) Drawing Conclusions: On an annual basis, the planting plan that maximizes total revenue is selected based on multiple simulation experiments. Statistical analysis is performed, considering the two scenarios from the first question, to arrive at the results.

5. Conclusion

The study presents a robust multi-year crop planting optimization model for a village in North China's mountainous region, leveraging linear programming and Monte Carlo simulation to address the intricate dynamics of agricultural planning. The model adeptly maximizes economic benefits while considering land constraints and market uncertainties, demonstrating its effectiveness through the comprehensive analysis of 2023's dataset. The model's success in balancing profitability with practical limitations highlights its potential for broader application. Future research could expand its scope by integrating emerging factors such as climate trends and technological innovations, enhancing its adaptability to the evolving agricultural sector. In summary, this study underscores the value of sophisticated modeling in shaping sustainable and economically viable agricultural strategies, setting a foundation for ongoing innovation in agricultural planning.

References

[1] Bu R, He H S, Hu Y, et al. Using the LANDIS model to evaluate forest harvesting and planting strategies under possible warming climates in Northeastern China[J]. Forest ecology and management, 2008, 254(3): 407-419.

[2] Guo X X, Li K L, Liu Y Z, et al. Toward the economic-environmental sustainability of smallholder farming systems through judicious management strategies and optimized planting structures[J]. Renewable and Sustainable Energy Reviews, 2022, 165: 112619.

[3] Sajid S S, Hu G. Optimizing crop planting schedule considering planting window and storage capacity[J]. Frontiers in Plant Science, 2022, 13: 762446.

[4] Dantzig G B. Linear programming[J]. Operations research, 2002, 50(1): 42-47.

[5] Bonate P L. A brief introduction to Monte Carlo simulation[J]. Clinical pharmacokinetics, 2001, 40: 15-22.

[6] Wang L, Xiao T, Liu S, et al. Quantification of model uncertainty and variability for landslide displacement prediction based on Monte Carlo simulation[J]. Gondwana Research, 2023, 123: 27-40.