Optimization of crop cultivation based on Monte Carlo simulation and linear programming

Yishuai Liu[1, *], Shuo Liu[2], Xiaoxi Ye[1]

[1]School of Mechanical Engineering, Yanshan University, Qinhuangdao, China, 066000

[2]School of Vehicle and Energy, Yanshan University, Qinhuangdao, China, 066000

*Corresponding author: lys387813261@outlook.com

Abstract. With the advancement of agricultural modernization, crop planting optimization faces challenges of price fluctuation risk and diversified constraints, necessitating the combination of Monte Carlo simulation with linear planning for scientific decision-making. This study quantifies crop planting rules based on 2023 data including expected sales volume, planting cost, yield, and sales price, establishing a comprehensive planning model. The sales price of each crop is obtained through Monte Carlo simulation method. By ensuring full land utilization and minimizing crop waste to maximize benefits, the model yields a maximum gain of 52.672 billion yuan. When adjusting the objective function to calculate excess production at 50% of the sales price, the maximum gain becomes 23.671 billion yuan. This research provides scientific decision support for agricultural production, optimizes land resource allocation, enhances risk resistance capacity, and promotes agricultural efficiency. By incorporating agronomic requirements such as legume crop planting, the model supports sustainable agricultural development and digital transformation, significantly contributing to agricultural modernization and rural economic development.

Keywords: Monte Carlo simulation, linear programming model, crop cultivation optimization.

1. Introduction

With the continuous progress of China's agricultural modernization, the issue of crop planting optimization is receiving increasing attention. Reasonable planting planning is not only related to farmers' income but also directly affects the efficient use of agricultural resources and the sustainable development of agricultural production. However, crop planting optimization is a complex systematic project, and its decision-making process is faced with a lot of uncertainties, as well as many realistic constraints, such as land resource limitations, crop rotation requirements[1], crop growing season[2] and so on, which increase the complexity of optimization decision-making. Traditional crop planting decisions mostly rely on experience, which makes it difficult to cope with the market environment of fluctuating prices and changing demand for agricultural products, as well as the variability of climate conditions in different seasons. To address this stochasticity, this study combines Monte Carlo simulation with linear programming, which can not only deal with the planting optimization problem of various plots and different types of vegetables, making crop planting more reasonable but also cope with the risk of revenue uncertainty caused by the price of[3], which provides a new research idea for the optimization of crop planting[4] The economic benefits of agricultural products can visually express the income and expenditure of the crops. It is of great significance to the analysis of growers' annual income and expenditure and the structural adjustment of the planting industry, and it is chosen to use the income as a measurement index to study the most reasonable scheme of new crop planting[5].

This study identifies that traditional research predominantly relies on static price assumptions or empirical judgments, which fail to effectively quantify market price fluctuation risks, resulting in limited applicability of decision-making models in dynamic market environments. Second, existing approaches inadequately integrate multiple constraints (e.g., land resources, seasonal crop rotation, labor allocation), often neglecting the interdependencies among complex real-world constraints, thereby oversimplifying models that cannot reflect realistic planting scenarios. Finally, deterministic models lack dynamic adaptability, with incomplete response mechanisms to uncertainties such as climate change and market demand volatility, leading to a disconnect between theory and actual production. To address these limitations, this study proposes systematic solutions through methodological innovations: First, integrating Monte Carlo simulation with multi-constraint linear programming to quantify price volatility risks via stochastic sampling and constructing a hybrid model incorporating seven constraint categories. Second, designing a dual-objective optimization framework that introduces a 50% price penalty mechanism for excess yields, achieving an expected revenue of CNY 52.67 billion under full land utilization while reducing overstock risks. This research bridges gaps in existing literature concerning dynamic adaptability, multi-constraint integration, and empirical validation, providing a replicable theoretical framework for agricultural digital transformation.

Firstly, the research data were processed through Python, and then the crop planting equations were listed according to the planting conditions of different plots and different types of vegetables, and then Monte Carlo simulation [6] was applied to determine the prices of different agricultural products in each year, and then the linear equations of the returns were established according to the principle of selling corresponding to the production of agricultural products, and finally, the most reasonable planting plan of the crops, as well as the final returns, were obtained The most reasonable planting program of crops and the final yield are finally obtained.

2. Data Sources and Preprocessing

The data for this article was obtained from https://www.mcm.edu.cn/html_cn/. All data were read from the relevant Excel files via the read_excel function of the panda's library and pre-processed to ensure data integrity and accuracy.

After the dataset was read, this study started with missing value processing. For all the missing values in the dataset, the forward fill (fill) method was used, which ensures the continuity of the data especially if there are missing values in the merged cells. In addition, for extraneous rows (the last four rows) in the table of Relevant Data for Statistics 2023, this study conducted a deletion process to remove redundant data, thus further optimizing the data structure.

In terms of data merging, this study matched the 'plot name' and 'plot type' in the 'Existing arable land in villages' table with the relevant fields in the 'Crops grown in 2023' table by using the 'Plot name' field. Crops in 2023 table. This operation ensured that the planting information for each plot was complete and merged into a single dataset. After the merger, all rows containing missing values were deleted, thus ensuring data integrity.

For the unit sales price field, this study first converted it to string format for further processing. For unit sales price intervals (e.g., "2-4" yuan/catty), this study calculates the mean value of the interval to obtain the average unit sales price for the crop. If the unit sales price was a single value (e.g., "$3" per pound), that value was used directly. This treatment ensures that the unit sales price is flexible enough to adapt to market fluctuations, which in turn ensures the consistency of the price data.

A probabilistic statistical approach was used to analyze the risk level of major crops in China, i.e. by using historical yield data, the mu yield of each crop was calculated by removing changes in crop yields caused by socio-economic and scientific and technological developments, etc.[7] . Then for processing the acre yield data, this study adjusted the acre yield of each crop by dividing it by 2 to obtain an approximate value. This adjustment ensured that the data were reasonable and provided an accurate basis for subsequent yield calculations.

The study also calculated the total production of each crop based on the "area under cultivation" and processed acreage data from Crops under Cultivation in 2023. This calculation provided the necessary basic data for subsequent analyses.

In terms of data storage, this study used the numpy library to create several 3D arrays of shape (41, 54, 2), which stored information such as planting cost, selling unit price, and acre yield for each crop.

With these 3D arrays, this study provides a structured and efficient data storage format for subsequent model analysis.

All data processing steps focus on data cleansing, pre-processing, calculation, and storage, ensuring the accuracy and consistency of the data before it enters the subsequent analysis and optimization process.

3. Optimization Model Construction and Solution

3.1. Objective function construction

The goal is to maximize the income from crops in the village in the years 2024~2030. Through analysis, if the actual production of crops is more than the expected sales volume, the excess part will be stagnant, and the excess part will be a loss, taking into account the fact that the land can be utilized as much as possible so that there is no wastage of spatial resources, the production of all types of crops in the year 2023 will be taken as the expected sales volume of all types of crops in the years 2024~2030 so that they can just be sold out. The production of each type of crop in 2023 will be used as the expected sales volume of each type of crop in 2024~2030 so that it can be sold out just[8].
The total gain from selling all the produce is:

$$\text{Max } Z=\sum_{t=1}^{7}\sum_{k=1}^{2}\sum_{i=1}^{41}\left(c_{i,j,r}\cdot\min\left(\sum_{j}k_{i,j,r,t}\cdot p_{i,j,r}\cdot M_i\right)\right)-n_{i,j,r}\cdot\sum_{j}k_{i,j,r,t} \tag{1}$$

Where $P_{i,j,r}$: acreage of the crop planted in j the plot in r the quarter; $c_{i,j,r}$: selling price of i the crop planted in j the plot in the quarter; $k_{i,j,r}$: acreage of the crop planted in j the plot in r a quarter of the year.

3.2. Monte Carlo simulation formula

Assuming that the density of distribution of the random variable $f(x)$ is known to be, the mathematical expectation of the variable $f(x)$ is.

$$E=\int_{x_0}^{x_1}f(x)\varphi(x)dx \tag{2}$$

N sample points xi are randomly selected according to the distribution density function $\varphi(x)$ and the arithmetic mean of the function values $f(x_i)$ corresponding to the sample points is used as the integral estimate.

$$\bar{E}_N=\frac{1}{N}\sum_{i=1}^{N}f(x_i) \tag{3}$$

According to the variable's probability distribution density function sequentially randomly selected variable values, after a large number of repeated independent simulations of the values of the variable, you can obtain the probability density distribution of the sales price of various crops, to achieve the variable random sampling calculation process[9].

3.3. Constraints

(1) Area constraint: the sum of the areas planted with each type of crop is equal to the sum of the rural land areas, i.e.

$$\sum_{i=1}^{41}A_i=1201 \tag{4}$$

At the same time, it should be ensured that the area planted for each type of crop is non-negative and does not exceed the land area.

$$\sum_{i=1}^{41} k_{i,j,r,t} \le A_j \tag{5}$$

$$\sum_{i=1}^{41} k_{i,j,r,t} \ge 0 \tag{6}$$

(2) Discontinuous Stubble Constraints

For the requirement of not being able to continuously re-crop on flat drylands, terraces, and hillsides, the need to meet the requirement of growing different crops on the same plot for two consecutive years.

$$k_{i,j,r,t} \cdot k_{i,j,r,t+1} = 0 \quad (i = 1, 2, \ldots, 15, 38, \ldots, 41, r = 1) \tag{7}$$

For the requirement of no continuous heavy cropping on watered land, it has been analyzed that this requirement can be met as long as the rice is not cropped continuously on the same plot.

$$k_{16,j,r,t} \cdot k_{16,j,r,t+1} = 0 \quad (i = 27, ..., 34, r = 1) \tag{8}$$

For the intelligent greenhouse not be continuous crop planting requirements, needs to meet two situations, first, the same year of the two seasons can not be planted with the same kind of vegetable crops: second, the second season of a year and the first season of the second year can not be planted with the same kind of vegetable crops, that is, can not be across the year continuous crop.

$$\begin{cases} k_{i,j,1,t} \cdot k_{i,j,2,t} = 0 \quad (i = 17, \ldots, 34, j = 51, \ldots, 54) \\ k_{i,j,2,t} \cdot k_{i,j,1,t+1} = 0 \end{cases} \tag{9}$$

(3) Regarding the requirement of joint planting, this study refers to the 2023 planting strategy and considers that non-canopy plots cannot be jointed, and the canopies that can be jointed should meet the requirement that the area of each type of plant grown on the plot is equal to the area of the canopies.

$$\begin{cases} k_{i,j,r,t} = A_j \quad (j = 1, \ldots, 34) \\ \sum_{i=17}^{34} k_{i,j,1,t} = A_j \quad (j = 35, \ldots, 54) \\ \sum_{i=38}^{41} k_{i,j,2,t} = A_j \quad (j = 35, \ldots, 50) \\ \sum_{i=17}^{34} k_{i,j,2,t} = A_j \quad (j = 51, \ldots, 54) \end{cases} \tag{10}$$

(4) Yield constraint: To maximize returns and reduce waste, the actual production of each crop type in 2023 is used as the projected sales volume from 2024 to 2030.

$$\sum_{j=1}^{54} \sum_{r=1}^{2} k_{i,j,r,t} \cdot P_{i,j,r} = M_i \tag{11}$$

(5) The greenhouses adopt a combined planting strategy, so it is stipulated that the area of crops planted in the greenhouses is not less than 50 percent of the area of the greenhouses as the area is not too small, and at the same time, to avoid the remaining land area being less than 50 percent of the area of the greenhouse

$$\begin{cases} k_{i,j,1,t} \ge 0.5A_j \quad (j = 35,\ldots,54, i = 17,\ldots,34) \\ k_{i,j,2,t} \ge 0.5A_j \quad (j = 35,\ldots,50, i = 38,\ldots,41) \\ k_{i,j,2,t} \ge 0.5A_j \quad (j = 51,\ldots,54, i = 17,\ldots,34) \end{cases} \tag{12}$$

(6) The requirement that all land in each plot (including greenhouses) be planted with legumes at least once in three years. Constraints are imposed on each land type The legume cropping requirement for flat dry land, terraced land, and hillside land is satisfied when there is legume cropping in any three consecutive years[10].

$$\sum_{i=1}^{5}\sum_{t}^{t+2} k_{i,j,r,t} \ge A_j (j = 1,\ldots,26) \tag{13}$$

For the requirement to grow pulses on watered land, since the only pulses grown on watered land are three of the vegetable crops, it is sufficient to satisfy the presence of pulses in any three consecutive years.

$$\sum_{i_3=17}^{18} k_{i_3,j,r,t} + k_{i_3,j,r,+1} + k_{i_3,j,r,+2} \ge A_j \quad (j = 27,\ldots,34, r = 1) \tag{14}$$

For the legume requirement for common greenhouses, since common greenhouses only have legumes in the first season's crop, it is only necessary to satisfy the requirement that the area of the same plot planted with legumes in three consecutive years is greater than or equal to the area of the plot:

$$\sum_{t=17}^{t+2}\sum_{i=17}^{19} k_{i,j,r,t} \ge A_j \quad (j = 51, 52, 53, 54) \tag{15}$$

For smart greenhouses, the requirements for growing pulses in the same plot for six seasons in any three consecutive years are only to be satisfied if the area of pulses grown in the same plot is greater than or equal to the area of the plot:

$$\sum_{r=1}^{2}\sum_{t}^{t+2}\sum_{i=17}^{19} k_{i,j,r,t} \ge A_j \quad (j = 51, 52, 53, 54) \tag{16}$$

(7) Watered land has two options each year, growing one season of rice or two seasons of vegetables. The constraints are as follows:

$$0.5A_j < k_{16,j,1,t} + \frac{k_{1,j,1,t} + k_{2,j,2,t}}{2} \le A_j \quad (i_1 = 17,\ldots,34, i_2 = 35, 36, 37, j = 27,\ldots,34) \tag{17}$$

Through Python code, the above model was solved to obtain the maximized 2024-2030 yield as well as the optimal planting scheme. Due to the large amount of data, only the optimal planting scenarios for selected plots in 2027 are shown, as shown in Table 1.

Table 1 Plant planted area in each plot (partial)

Plot name	Name of occupational plant	season (sports)	Plant type	plantation
F3	cowpea	1	Vegetables (pulses)	0.6
F3	broccoli	2	fruits	0.3
F3	greens	2	fruits	0.3
F4	kidney bean	1	Vegetables (pulses)	0.6
F4	celery (Apium graveolens)	2	fruits	0.3
F4	broccoli	2	fruits	0.3

From Table 1, it can be concluded that 0.6 acres of cowpeas were planted in the first quarter, 0.3 acres of baby bok choy and 0.3 acres of lettuce were planted in the second quarter in plot F3, 0.6 acres of kidney beans were planted in the first quarter, 0.3 acres of celery and 0.3 acres of spinach are planted in the second quarter in plot F4, which is in line with the requirements of the smart greenhouse's plots of crop cultivation, in line with the condition of not being able to continuously cultivate a heavy crop constraint, and meet the conditions of legume crop cultivation as well as the planting land can not be too scattered requirements. The optimal crop planting scheme obtained by applying Monte Carlo simulation and linear programming analysis in this study not only has a large planting area, which is convenient for cultivation and field management, but also reasonably plants legume crops, uses legume rhizobium to nourish the land, improves the crop yield, and increases the income, at the same time, taking into account the impact of continuous planting reduction and the phenomenon of reduced income caused by too much yield, and ultimately obtains the crop planting scheme with the largest income.

4. Conclusions

This paper proposes an optimization method of agricultural planting strategy based on Monte Carlo simulation and linear programming model, which can generate scientific and reasonable planting schemes under the constraints of complex plot topography and various crop planting conditions. Through the cleaning of actual data, the establishment and optimization of the Monte Carlo model as well as the linear revenue model of crop revenue, the research results show that the method significantly improves the planting revenue and optimizes the allocation efficiency of land resources. Compared with traditional optimization methods, this study provides more practical decision support based on the comprehensive consideration of planting costs, yield fluctuations, and market prices, and the planting requirements of continuous cropping and legume crops are also of great significance for the conservation of land resources and subsequent planting.

This study provides a theoretical basis and practical reference for intelligent decision-making and efficient use of resources in agriculture. However, the dynamic adaptability of the model still needs to be improved, this study did not discuss the impact of economic benefits caused by yield reduction due to heavy crop planting, and at the same time, to facilitate the study, we restrict the types of crops planted in the plots, and carry out some idealized treatment of crop planting, which has certain deviations from the actual crop planting, and the subsequent study can investigate the benefits of crop diversification in the plots, and can also be further expanded by combining real-time climate change, market dynamics, and other factors to cope with more complex and dynamic agricultural production environments, providing a broader application prospect for the refined management of modern agricultural production.

References

[1] NI Xuezhi,YU Xiaoyuan. Cropland rotation, agricultural cropping structure and lasting food security in China[J]. Exploration of Economic Issues,2018,(07):78-88.

[2] Tian Guoping. Research on the impact of temperature change on China's agricultural economy[D]. Southwest University of Finance and Economics,2021.

[3] Mao Yu, Mechanism of price influence on feed grain variety substitution in the context of food consumption structure evolution(D). Southwest University of Finance and Economics,2023.

[4] Su Xiaofeng. Research on crop planting decision-making model based on big data analysis technology[J]. Heilongjiang Grain,2023,(10): 88-90.

[5] WU Menghan, WANG Yi. Optimal adjustment of multi-objective planting of crops in Shache Irrigation District of Xinjiang[J]. People's Yellow River,2024,46(01):120-125+131.

[6] Li Yaqi. Agribusiness value assessment based on Monte Carlo simulation[D]. Central University of Finance and Economics,2022.

[7] WANG Jing, FANG Feng, WANG Suping, LI Zhenqi. Risk assessment of crop production in China based on probabilistic statistical methods[J]. Meteorological and Environmental Science,2023,46(02):9-18.

[8] LI Guizhou, ZHOU Xinguang. Bayesian autoregressive prediction analysis of concrete carbonation depth[J]. Journal of Yantai University (Natural Science and Engineering Edition),2013,26(04):282-286+291.

[9] YU Guo,LI Haitao,FANG Yizhu. Construction of target uncertainty prediction model for conventional gas planning production in Sichuan Basin[J] Petrochemical Applications,2024.43(07):30-35+51.

[10] Li Junxian. Effects of legume crop rotation on ammonia balance and productivity of farming systems in semi-arid areas (D). Gansu Agricultural University,2019.