Optimization strategy of crop cultivation in North China based on sample mean approximation

Jingjing Lin *

School of Economics and Management, Beijing Jiatong University, Beijing, China, 100044

* Corresponding Author Email: 22711051@bjtu.edu.cn

Abstract. Optimizing native resources can enhance productivity, improve farmers' living standards and promote ecological conservation. The aim of this paper is to analyze the characteristics of crop cultivation in North China and propose an optimized cultivation strategy to achieve maximum total returns during the period 2024-2030. The study adopts the Sample Average Approximation (SAA) method in stochastic optimization, combined with Latin hypercube sampling, to obtain the optimal planting plan using linear programming by considering constraints such as crop planting area, crop rotation requirements and legume planting. The results show that the optimized planting scheme can achieve an increase in average annual income compared with the ideal situation, and the total income is increased by 81.46% compared with the pre-optimization situation. The study shows that the optimized planting strategy not only improves the economic benefits but also provides a practical reference for the future revitalization of the countryside, which is of great potential for application and research value.

Keywords: Crop Planting Optimization Strategy, Linear Programming, Sample Average Approximation, Latin Hypercube Sampling.

1. Introduction

The development and utilization of local resources is a core component of the rural revitalization strategy. By optimizing local resources, rural areas can enhance productivity, improve farmers' living standards, and promote ecological conservation and environmental management. Zhang Hao et al. [1] studied the optimization of cropping structures in tropical savanna climate irrigation areas based on stochastic dynamic programming. They developed an optimization model to reduce the risks associated with water resource uncertainty in irrigation planning. Using the Louga irrigation area in Senegal as a case study, their research demonstrated that the model effectively optimized cropping structures, increasing economic benefits by an average of 137 million CFA per year and significantly improving economic returns under extreme drought conditions. Luo Dan et al. [2] proposed a multi-objective particle swarm-biogeography-based optimization algorithm to address tomato planting planning problems. They constructed three objective functions, and simulation results showed that the algorithm outperformed traditional evolutionary algorithms in planting planning, yielding reasonable planting schemes.

The North China region, located in northern China, is one of the country’s key agricultural production bases. The region's climate, soil, and water resource conditions are favorable for the growth of various crops. Wang Fuxin [3] pointed out that there is significant spatial imbalance in agricultural land-use efficiency across cities in North China, with the northeastern region showing higher growth rates in efficiency, while the southern region lags behind. However, few studies have focused on the characteristics of crop planting in mountainous areas of North China to optimize planting strategies.

This paper considers the crop types, time constraints, crop rotation requirements, and planting areas of a specific village in North China. It analyzes two scenarios for planting schemes. Scenario one assumes an ideal condition (planning based solely on the 2023 planting situation), while scenario two incorporates uncertainties in crop expected sales volume, yield per acre, planting costs, and sales prices. By employing the sample average approximation method in stochastic optimization and using Latin hypercube sampling to account for randomness and risk, the study ultimately derives an optimal planting scheme over a nearly seven-year period. The findings provide valuable references and practical insights for crop planting planning in North China.

2. Establishment of decision variables and objective functions

The data in this paper comes from 1,201 acres of cropland in a region of North China in 2023, with 34 plots of varying sizes. The data source is from https://www.mcm.edu.cn/html_cn/node/a0c1fb5c31d43551f08cd8ad16870444.html

For the 2023 crop planting and 2023 statistics, this paper correlates the data in the 2023 crop planting table and the 2023 statistics and calculates the expected 2023 sales production, total planting cost, average selling price, and sales unit price for each crop in each plot.

2.1. Establishment of decision variables and objective functions

In this paper, the decision variable is identified as x_{ijs}^{t}, i.e., the acreage of the crop in the jth of the i-th plot in the s-th season of the t-year.

This rural area contains three types of arable land, which are watered land, ordinary greenhouses, and smart greenhouses, and the number of seasons for crop cultivation varies among the different types of arable land, so this paper introduces the seasonal parameter s:

$$s=\begin{cases}0, & \text{Single-season planting}\\ 1, & \text{First season of two-season planting}\\ 2, & \text{Second season of two-season planting}\end{cases} \tag{1}$$

In order to optimize the cropping scheme for the crops in the village from 2024-2030, the total return should be maximized for the seven years from 2024-2030. In this paper, the total return is set to be W and $W=\sum_t W_t=\sum_t (Sale_t - Cost_t)$. When the total production of a crop exceeds the expected sales, the excess will be sold at a reduced price of 50% of the 2023 sales price. The formula for calculating sales is as follows:

$$\text{Sale}^{t}=\sum_j \begin{Bmatrix} \{min\left[\left(\sum_{i,s} x_{ijs}^{t}\cdot N_{ijs}-\sum_{i,s} x_{ijs}^{2023}\cdot N_{ijs}\right),0\right]+\sum_{i,s} x_{ijs}^{2023}\cdot N_{ijs}\}\cdot P_j \\ +\frac{1}{2}max\left[\sum_{i,s} x_{ijs}^{t}\cdot N_{ijs}-\sum_{i,s} x_{ijs}^{2023}\cdot N_{ijs}),0\right]\cdot P_j \end{Bmatrix} \tag{2}$$

The final objective function is as follows:

$$W=\sum_t \left\{\sum_j \left\{\begin{pmatrix} \{\min\left[\left(\sum_{i,s} x_{ijs}^{t}\cdot N_{ijs}-\sum_{i,s} x_{ijs}^{2023}\cdot N_{ijs}\right),0\right]+\sum_{ijs}^{2023}\cdot N_{ijs}\}\cdot P_j \\ +\frac{1}{2}\max\left[\left(\sum_{i,s} x_{ijs}^{t}\cdot N_{ijs}-\sum_{i,s} x_{ijs}^{2023}\cdot N_{ijs}\right),0\right]\cdot P_j-\sum_{i,s}\left(x_{ijs}^{t}\cdot Q_{ijs}^{t}\right)\end{pmatrix}\right\}\right\} \tag{3}$$

Based on the climate characteristics, cropland types and crop growth patterns in North China, the following constraints are formulated in this paper:

(1) Crop acreage requirements

Crops must be planted on a plot that is smaller than the area of the plot and the crop must be planted on an area larger than the 0.

$$\sum_j x_{ijs=0or1or2}^{t}\le A_i \tag{4}$$

$$x_{ijs}^{t}\ge 0 \tag{5}$$

And too small area planting may lead to insufficient yield per unit area to cover the production cost and reduce the overall economic efficiency. Therefore, this paper assumes that the planting area of each crop on each plot should not be less than one-tenth of the total area of the plot.

$$x_{ijs}^{t}\ge \frac{1}{10}A_i \tag{6}$$

(2) Crop rotation requirements

In order to maintain soil health, reduce pests and diseases, and improve crop yield and quality, agriculture requires that the same plant be grown without successive heavy cropping. Since the first and second seasons of ordinary greenhouses in this rural area as well as the single and double cropping arrangements in watered land satisfy the requirement of growing the same crop without successive re-cropping, this paper makes the following constraints only for smart greenhouses:

$$\begin{cases} x_{ijs=1}^{t} \times x_{ijs=2}^{t} = 0 \\ x_{ijs=2}^{t-1} \times x_{ijs=1}^{t} = 0 \end{cases} \tag{7}$$

(3) Requirements for growing legumes

Legume crops significantly improve soil health and agricultural sustainability by fixing nitrogen, improving soil structure, managing pests and diseases and increasing crop diversity. This paper therefore stipulates that all land should be planted with legumes at least once in three years.

$$\sum_{j=bean} \left(x_{ijs}^{t-1} + x_{ijs}^{t} + x_{ijs}^{t+1}\right) > 0 \tag{8}$$

(4) Planting concentration requirements

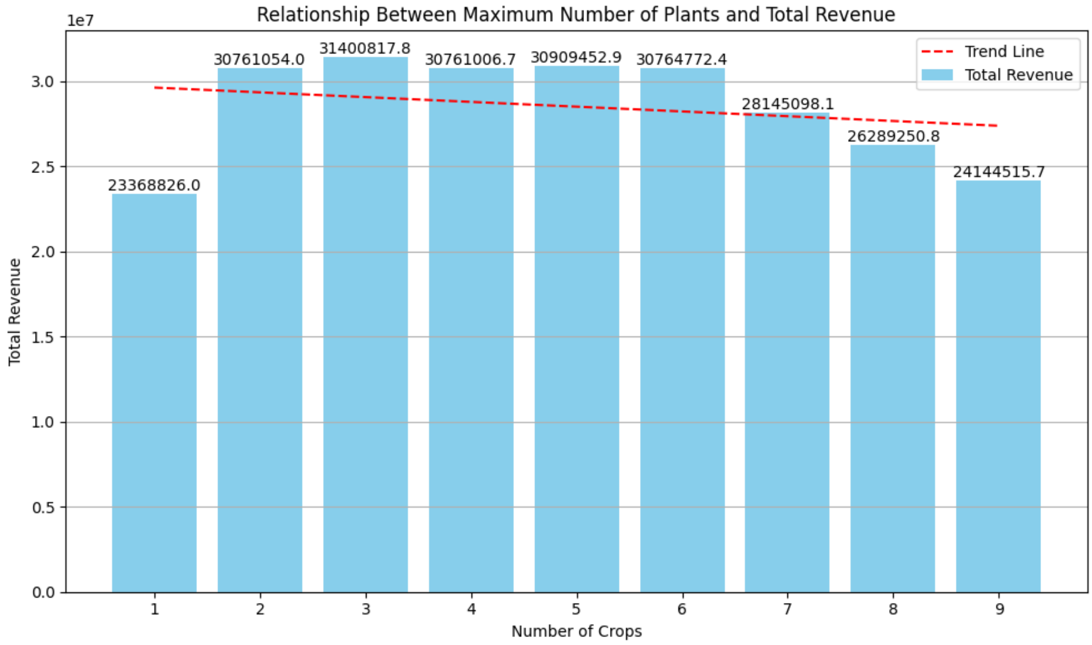

Figure 1. Relationship between the number of crops grown per plot and total returns

In Figure 1, it can be seen that the total return is maximum when the number of crops is 3, and the total return decreases more when the number of crops is 7-9. In order to facilitate farming and field management, the planting area of each crop in each season should be concentrated as much as possible, and the area planted on a single plot should not be too small, so this paper assumes that the number of crops planted on each plot is 3.

2.2. Solution of Case I model and analysis of results

In this section, only the ideal case is considered, i.e., it is assumed that the expected sales volume, planting cost, acreage, and sales price remain stable, and that crops that exceed the expected sales volume are treated as being sold at a 50% price reduction. The model results are shown in Table 1:

Table 1. Income forecast statement

Year	2024	2025	2026	2027	2028	2029	2030
Annual yield	4390416.5	3798047.0	4060116.3	4040712.7	4937509.6	3948592.0	5853151.3

The total seven-year return for this planting scheme was 310,285,545.4, and the return changes is visualized in Figure 2:

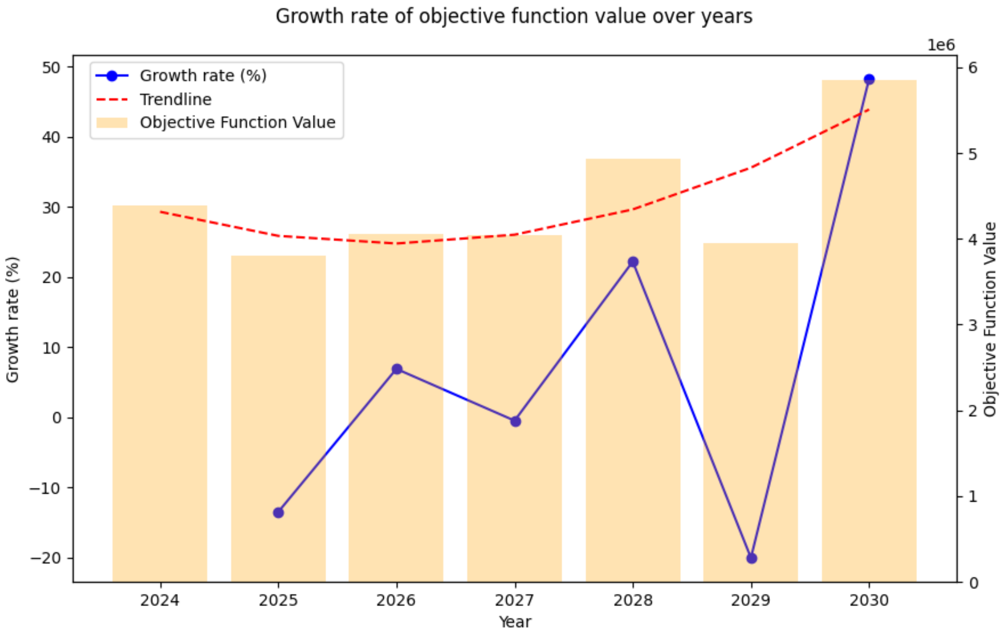

Figure 2. Diagram of the results of the optimal planting scheme for case 1

In Figure 2, the bar charts show the crop cultivation returns for each year from 2024 to 2030, while the line graphs show the growth rate of returns for each year of the same period. Based on these data, the following conclusions can be drawn:

(1) As shown in the trend line, there is a slight overall upward trend in crop yields for each year from 2024-2030;

(2) The annual growth rate of returns fluctuates considerably in each of the years 2024-2030, with the highest rate approaching 50 percent in 2030.

3. Solving Crop Planting Uncertainty Problem Based on Sample Average Approximation (SAA)

Sample Average Approximation (SAA) is an optimization technique widely used in mathematical optimization and operations research, primarily for solving stochastic optimization problems. It is particularly effective in situations where the objective function cannot be calculated precisely but can be estimated through stochastic simulation. SAA works by drawing samples from the probability distribution of uncertain parameters and using the empirical average of the finite samples to approximate the expected value, thereby transforming the stochastic problem into a deterministic one, which can be solved using traditional optimization techniques. This method has broad applications in fields such as stochastic programming, risk management, and engineering design [4-5]. Thomas Lew et al. [6] studied non-convex stochastic programming problems with expected value equality constraints and found that the SAA method does not necessarily guarantee asymptotic optimality as the sample size increases. To address this, they relaxed the equality constraints and demonstrated that under moderate smoothness and boundedness conditions, the modified SAA method could achieve asymptotic optimality. Wang Mingzheng et al. [7] proposed an SAA-based method that transforms stochastic problems into deterministic constrained optimization problems using a regularized gap function and proved the convergence of both the optimal value and the optimal solution. Ren Yonghong et al. [8] introduced an SAA method based on the Log-Sigmoid approximation to solve chance-constrained optimization problems. They demonstrated that the optimal value and the set of optimal solutions for the SAA problem exhibit good convergence properties when the sample size is sufficiently large. Given the stochastic, uncertain, and chance-constrained characteristics of the data in this study, the sample average approximation method is also employed in this paper.

The formula for the sample means approximation can be expressed as:

$$f_n(x) = \frac{1}{n}\sum_{k=1}^{n} f(x, \xi_k) \tag{9}$$

Where $f_n(x)$ is a sample-based approximation function. ξ_k is the kth sample drawn from the distribution of the random variable ξ. With SAA, this paper transforms the original stochastic optimization problem into the following deterministic optimization problem:

$$\min_{x \in D} f_n(x) \tag{10}$$

Where D is the set of feasible solutions. By solving this deterministic problem, this paper can obtain an approximate solution to the original stochastic problem.

Sampling method is a key component of Sample Average Approximation, and Latin Hypercubic Sampling is chosen in this paper. Latin hypercubic sampling improves the accuracy and efficiency of the simulation results by dividing the range of values for each dimension into a number of equal-width intervals and ensuring that only one sample point is taken in each interval [9]. Compared with traditional random sampling, it effectively avoids the bias problem and ensures that there are enough samples in different regions of the sample space. Due to the smaller sample size compared to traditional Monte Carlo methods, it can also greatly save computational resources while considering more uncertainties and risks [10]. Xu et al [11] investigated how to utilize the Latin Hypercubic Sampling (LHS) method for parametric uncertainty analysis to simulate the uncertainty of a random variable and consider the correlation between the variables in order to improve the computational accuracy under the condition of small samples and the efficiency. Yan Zhang et al [12] proposed a stratified Latin hypercube sampling method, which proved to be advantageous in improving estimation accuracy and shrinking Monte Carlo variance. The data in this paper is characterized by multi-dimensionality, high optimization difficulty and constraints, so the Latin hypercube related method is also used in this paper.

The basic steps of Latin hypercube sampling are as follows:

(1) Delineation of intervals:

Suppose that this paper has an N-dimensional random variable with a range of values in each dimension divided into M equal-width intervals. The length of each interval is $\frac{1}{\mathrm{M}}$.

(2) Generate a sample

One sample point is randomly selected in each interval of each dimension. Specifically, for the i-th dimension($i = 1,2, \dots, N$), this paper randomly selects a sample point P_{ij} from j intervals ($j = 1,2, \dots, M$), and combines these sample points into a matrix A.

$$A = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1M} \\ P_{21} & P_{22} & \cdots & P_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ P_{N1} & P_{N2} & \cdots & P_{NM} \end{bmatrix} \tag{11}$$

(3) Confuse the order:

To ensure the randomness of the samples, each row of matrix A needs to be randomly disrupted to obtain the final sample matrix B. Each column of matrix B represents a sample point:

$$B = \begin{bmatrix} P_{1x} & P_{1y} & \cdots & P_{1z} \\ P_{2o} & P_{2p} & \cdots & P_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ P_{Nr} & P_{Ns} & \cdots & P_{Nt} \end{bmatrix} \tag{12}$$

Where $x, y, z, o, p, q, r, s, t$ are the indexes of the sample points[13-15].

Based on the above ideas, the uncertainties faced by the expected sales volume, acreage, planting costs and selling prices of various crops in the real situation are combined with the possible risks of planting. This paper introduces the following fluctuations based on the ideal situation in the previous section:

Expected sales $\boldsymbol{V_j^t}$: For food crops, the average annual growth rate of wheat and corn relative to 2023 is between 5% and 10%. This article believes that the distribution of annual growth rate R_{tj} of wheat and corn is a normal distribution $N(7.5\%, (1.2\%)^2)$.

$$V_j^{2023} = \sum_{i,s} x_{ijs}^{t=2023} \cdot N_{ijs}^{t=2023} \tag{13}$$

$$V_j^t = (1 + R_{tj}) \cdot V_j^{t-1} \tag{14}$$

Other crops: Since the expected annual sales volume in the future will change by about ±5% relative to 2023, this article treats the changes in other crops R_{tj} as a normal distribution $N(0, (3\%)^2)$:

$$V_j^t = (1 + R_{tj}) \cdot V_j^{t=2023} \tag{15}$$

Yield per mu $\boldsymbol{N_{ij}^t}$: The yield per mu of crops is significantly affected by natural factors such as weather, pests and diseases, and soil conditions. Crops of the same type usually respond similarly to these factors because they grow in similar ecological environments. Therefore, this paper uses the same change in yield per mu for the same type of crops to describe fluctuations in each season.

$$N_{ijs}^t = \left(1 + Y_{tjs}\right) N_{ijs}^{t=2023} \tag{16}$$

$$Y_{tjs} \sim N(0, (4\%)^2) \tag{17}$$

Planting cost $\boldsymbol{C_j^t}$: Due to the influence of market conditions, the planting cost of crops increases by about 5% every year on average. Therefore, this article believes that the average growth rate of the planting cost of each crop is $E_{tj} \sim N(5\%, (0.5\%)^2)$.

$$C_j^{2023} = \sum_{i,s} \left(x_{ijs}^t \cdot Q_{ijs}^t\right) \tag{18}$$

$$C_j^t = \left(1 + E_{tj}\right) C_j^{t-1} \tag{19}$$

Among them, Q_{ijs}^t is the cost per acre of the j crop in the i plot in t.

Sales Price $\boldsymbol{P_j^t}$:

(1) The sales price of food crops is basically stable. This article believes that when $j$is a food crop, the sales price change rate F_j^t obeys the normal distribution $N(0, (0.5\%)^2)$:

$$P_j^t = \left(1 + F_j^t\right) P_j^{2023} \tag{20}$$

(2) The sales price of vegetable crops has an increasing trend, with an average annual increase of about 5%. This article believes that when $j$is a vegetable crop, the sales price change rate F_j^tobeys the normal distribution $N(5\%, (0.5\%)^2)$:

$$P_j^t = \left(1 + F_j^t\right) P_j^{t-1} \tag{21}$$

(3) The sales price of edible fungi is stable but declining and can decrease by approximately 1%-5% per year, among which the sales price of morels decreases by 5% per year. This article believes that when $j$is other edible mushrooms except morels, the sales price change rate F_j^tobeys the normal distribution $N(3\%, (1\%)^2)$; and when j is morels, the sales price change rate is $F_j^t = 5\%$.

$$P_j^t = \left(1 - F_j^t\right) P_j^{t-1} \tag{22}$$

Based on case 1 and combined with the above change equation, the objective function is as follows:

$$W = \sum_t \left\{ \sum_j \left\{ \begin{array}{c} \left\{ \begin{array}{c} \min\left[\left(\sum_{i,s} x_{ijs}^t \cdot N_{ijs}^t - V_j^t\right), 0\right] \\ + \sum_{i,s} x_{ijs}^{2023} \cdot N_{ijs}^t \end{array} \right\} \cdot P_j^t \\ + \frac{1}{2} \max\left[\left(\sum_{i,s} x_{ijs}^t \cdot N_{ijs}^t - V_j^t\right), 0\right] \cdot P_j^t - C_j^t \end{array} \right\} \right\} \tag{23}$$

4. Case 2 model results

According to the SAA method, this paper approximates the distribution of a random variable by generating a large number of samples. In order to better determine the number of samples for the SAA method, this paper sets the maximum number of samples to be 10,000, and increases 100 samples per iteration to examine the effect of the change in the number of samples N on the SAA method.

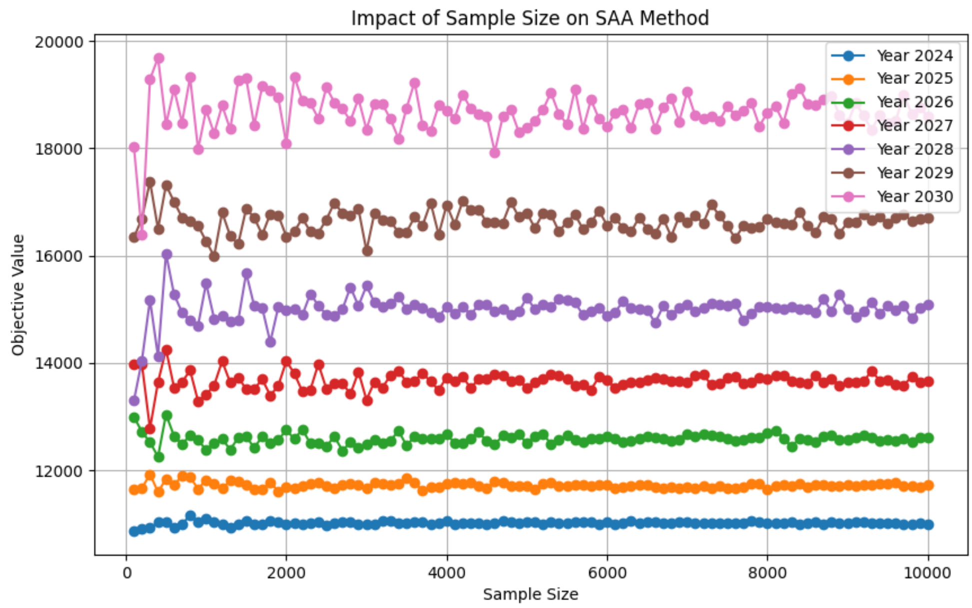

Figure 3. Sample Average Approximation convergence analysis

It can be seen in Figure 3 that as the sample size increases, the returns for each year converge at sample sizes greater than 4000. For each random variable, the paper performs the following steps:

Use the SAA sampling method to generate N samples ($N = 4000$)samples of random variables, such as sales volume, yield per mu, cost and price.

The Latin hypercube sampling method was applied to ensure sample homogeneity and representativeness.

Calculate the average of these samples as the final estimate $\hat{V}_j^t = \frac{1}{N}\sum_{i=1}^{N} V_{j,i}^t$, where $V_{j,i}^t$ is the sales volume of the i sample. Other variables (yield per mu, planting cost, sales price) are estimated in a similar way.

The original problem is transformed into

$$max \frac{1}{N}\sum_{i,j=1}^{N} W\left(x_{ijs}^t, \hat{V}_j^t, \hat{N}_{ij}^t, \hat{C}_j^t, \hat{P}_j^t\right) \tag{24}$$

The result is shown below:

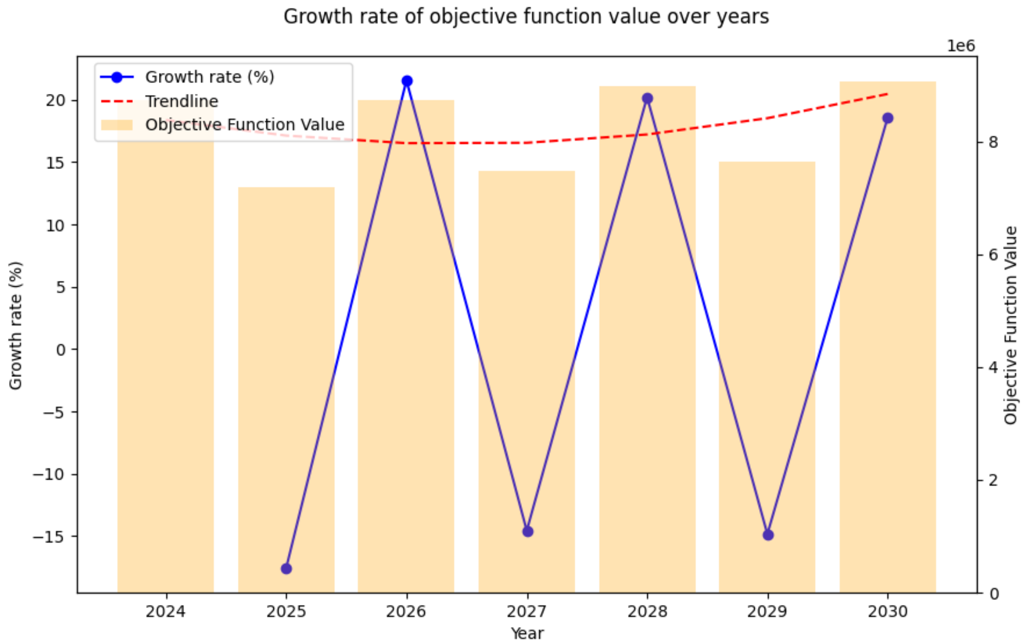

Figure 4. Diagram of the results of the optimal planting scheme for case 2

In Figure 4, the total income over seven years is approximately 56,274,801.83 yuan. The columns represent annual crop planting income (2024–2030), while the line shows the annual growth rate. Key conclusions:

(1) The annual crop planting income from 2024 to 2030, as shown by the trend line, generally shows a fluctuating upward trend.

(2) There will be certain fluctuations in the annual growth rate of earnings from 2024 to 2030.

The planting situation of each plot of land in 2024 is shown in Figure 5 and Figure 6.

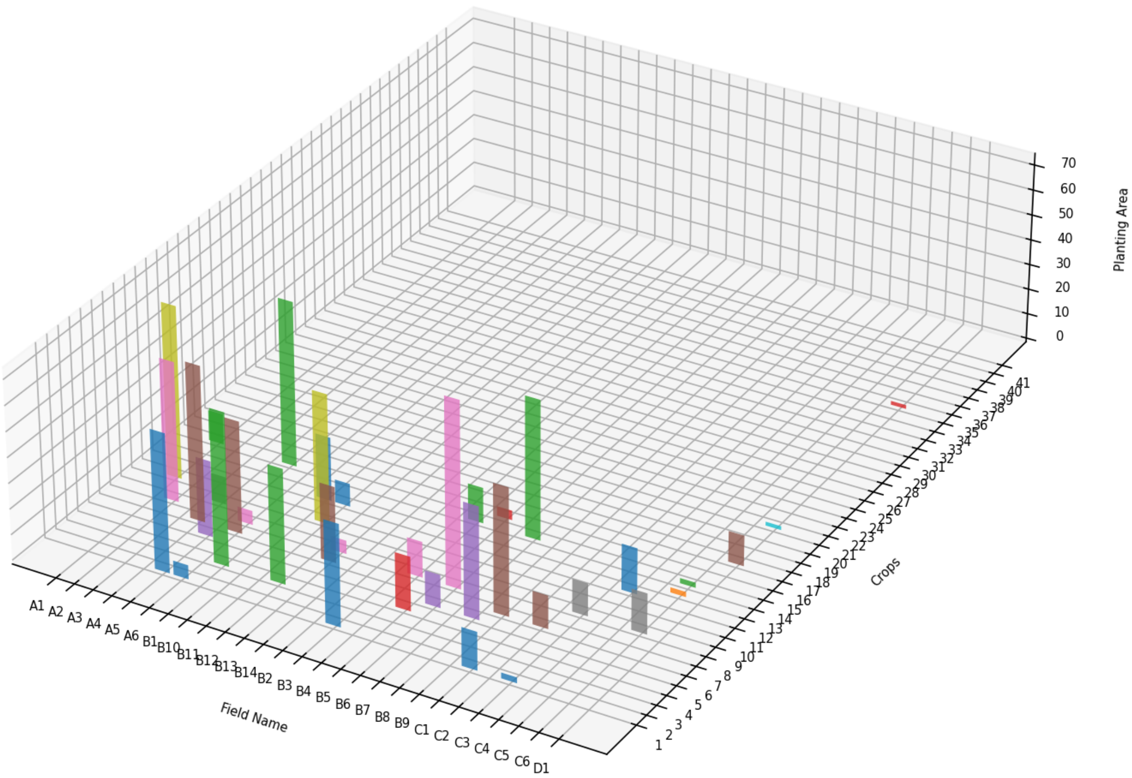

Figure 5. 3D Crop Planting Scheme (Plot number A1-C6) in 2024

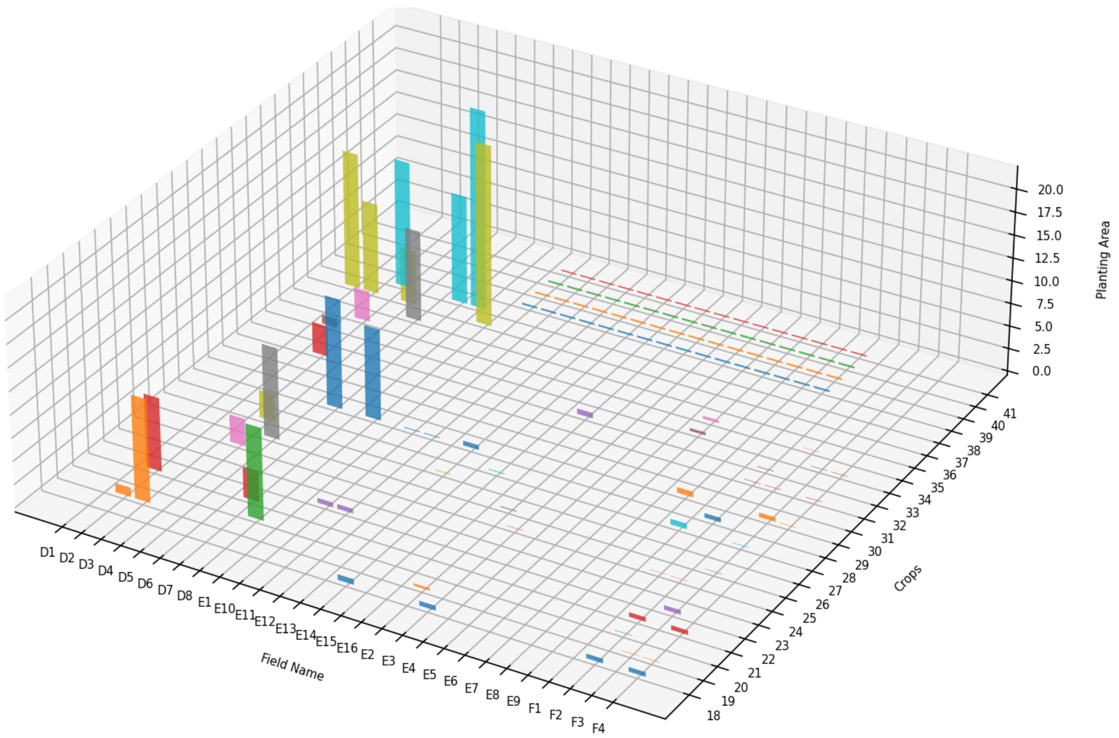

Figure 6. 3D Crop Planting Scheme (Plot number D1-F4) in 2024

Cropland numbers and corresponding types are: A1-A6-plain dry land, B1-B14-terraces, C1-C6-sloping land, D1-D8-irrigated land, E1-E16-ordinary greenhouses, and F1-F4-smart greenhouses. The crop numbers and corresponding crop names are shown in Table 2.

Table 2. Crop number and corresponding crop name

ID	Crop Name	ID	Crop Name	ID	Crop Name	ID	Crop Name
1	Soybean	12	Pumpkin	23	Spinach	34	Celery
2	Black Bean	13	Sweet Potato	24	Green Pepper	35	Chinese Cabbage
3	Red Bean	14	Oat	25	Cauliflower	36	White Radish
4	Green Bean	15	Barley	26	Cabbage	37	Red Radish
5	Pod Bean	16	Rice	27	Oilseed Lettuce	38	Elm Mushroom
6	Wheat	17	Cowpea	28	Baby Bok Choy	39	Shiitake Mushroom
7	Corn	18	Sword Bean	29	Cucumber	40	White Pearl Mushroom
8	Millet	19	Kidney Bean	30	Lettuce	41	Morel Mushroom
9	Sorghum	20	Potato				
10	Foxtail Millet	21	Tomato				
11	Buckwheat	22	Eggplant				

5. Conclusion

This study focuses on the optimization of local resources in the North China region. By employing the Sample Average Approximation (SAA) method combined with Latin Hypercube Sampling, it analyzes multiple factors influencing crop cultivation and proposes optimized planting strategies tailored to local climate and soil conditions. The results indicate that the optimized planting strategies can achieve significant economic benefits during the period from 2024 to 2030, with annual average profits showing a fluctuating upward trend. The findings demonstrate strong practical value and serve as a meaningful reference, providing scientific support for the rural revitalization strategy and promoting the sustainable development of agricultural production.

Although this study has achieved preliminary results, there are still some limitations, such as insufficient sensitivity analysis of market demand changes and constraints in model assumptions. In the next steps, the plan is to further improve the model by incorporating more external factors, such as climate change and market fluctuations, to enhance its adaptability and accuracy. Additionally, field investigations will be conducted to collect more empirical data to validate and refine the research outcomes, ensuring the feasibility and effectiveness of the optimized strategies. This will provide a more robust foundation for supporting rural revitalization.

References

[1] Zhang Hao, ZHAO Shengwei, Qian Jun, et al. Based on stochastic dynamic programming of savannah climate irrigation area planting structure optimization study [J]. Water saving irrigation, 2024 (10): 15 - 21.

[2] Luo Dan, Jiang Bingbing. A multi-objective particle swarm optimization and biogeography algorithm for tomato planting planning problem [J]. Computer Applications and Software, 2019, 40 (07): 294 - 299.

[3] Wang Fuxin. Time and space differences of north China agricultural land use efficiency evaluation and analysis [D]. Hebei economic and trade university, 2022.

[4] OSMAN ALTUN, ENGIN KARATEPE. Stochastic Generation and Transmission Expansion Planning using Sample Average Approximation[C]//2023 13th International Symposium on Advanced Topics in Electrical Engineering: ATEE 2023, Bucharest, Romania, 23 - 25 March 2023, [v.2]. 2023: 1 - 5.

[5] HU S., DONG Z.S., DAI R. A machine learning based sample average approximation for supplier selection with option contract in humanitarian relief [J]. Transportation research, Part E. Logistics and transportation review, 2024, 186 (Jun.): 1. 1 - 1. 26.

[6] Lew T, Bonalli R, Pavone M. Sample Average Approximation for Stochastic Programming with Equality Constraints [J]. SIAM Journal on Optimization, 2024, 34 (4): 3506 - 3533.

[7] WANG MINGZHENG, LIN GUIHUA, GAO YULI, et al. Sample average approximation method for a class of stochastic variational inequality problems [J]. Journal of Systems Science and Complexity, 2011, 24 (6): 1143 - 1153.

[8] Ren Yonghong, Wang Jia, Wang Yu, Ma Yanni. Sample average approximation method for solving the Log-Sigmoid approximation problem of chance-constrained optimization [J]. Journal of Liaoning Normal University (Natural Science Edition), 2014, 37 (2): 153 - 156.

[9] BOUSHEHRI, REZA, MOTAMED, RAMIN, ELLISON, KIRK, et al. Estimating epistemic uncertainty in soil parameters for nonlinear site response analyses: Introducing the Latin Hypercube Sampling technique [J]. Earthquake Spectra: The Professional Journal of the Earthquake Engineering Research Institute, 2022, 38 (4): 2422 - 2450.

[10] Hou Z S, Liao W X, Wang C X, et al. Caprock Safety Evaluation Method in CCUS Based on Latin Hypercube Sampling: A case study of a block reservoir [J]. Journal of Physics: Conference Series, 2024, 2834 (1): 012165 - 012165.

[11] ZHAOXIA XU, XIUZHEN WANG. Global sensitivity analysis of the reliability of the slope stability based on the moment-independent combine with the Latin hypercube sampling technique [J]. Stochastic Environmental Research and Risk Assessment, 2023, 37 (6): 2159 - 2171.

[12] Zhang Yan, Yang Jinyu. Stratified Latin hypercube sampling method and its application[J]. Statistics and Decision, 2023, 39 (15): 48 - 51.

[13] KHAIRUL UMAM SYALIMAN, ADLI ABDILLAH NABABAN, MIFTAHUL JANNAH, et al. Latin Hypercube Sampling Approach to Improve K-Nearest Neighbors Performance on Imbalanced Data[C]//2023 International Conference of Computer Science and Information Technology: ICOSNIKOM 2023, Binjia, Indonesia, 10-11 November 2023. 2023: 1 - 6.

[14] LUALDI, PIETRO, STURM, RALF, SIEFKES, TJARK. Exploration-oriented sampling strategies for global surrogate modeling: A comparison between one-stage and adaptive methods [J]. Journal of computational science, 2022, 60 (Apr.): 101603.1 - 101603.20.

[15] XIN PEI, FEI MEI, JIAQI GU. The real-time state identification of the electricity-heat system based on Borderline-SMOTE and XGBoost [J]. IET Cyber-Physical Systems: Theory & Applications, 2022, 8 (4): 236 - 246.