Research on Crop Planting Strategy Optimization Based on Fuzzy C-Means Clustering and Generalized Additive Model

Jingjing Lin *

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

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

Abstract. Against the backdrop of global climate change and resource scarcity, optimizing farmland planting strategies is crucial for improving agricultural production efficiency and achieving sustainable development. Current research primarily focuses on crop planting structures, but discussions on crop substitutability and legume complementarity remain insufficient. This study aims to explore crop planting strategies in a village in the North China region, particularly how to flexibly adjust planting structures under market fluctuations to maximize profits. The study employs fuzzy C-means clustering to analyze crop substitutability and incorporates legume complementarity, using a generalized additive model to predict sales volumes. The results show that considering substitutability and complementarity increases profits by approximately 1.08 times compared to pre-optimization levels, demonstrating that flexible planting strategies can effectively enhance profitability. Furthermore, the optimized planting plan improves farmland utilization rates. The findings provide valuable references for agricultural production, highlighting the potential of rational planting structures to enhance economic benefits.

Keywords: Crop Planting Strategy Optimization, Linear Programming, Fuzzy C-Means Clustering, Generalized Additive Model.

1. Introduction

Optimizing farmland planting strategies is critical for improving agricultural production efficiency and achieving sustainable development. With the challenges posed by global climate change and resource scarcity, the rational allocation of crop varieties has become increasingly important. Xi Kaiyan et al. [1] developed an interval two-stage risk-averse stochastic programming model for optimizing oasis water resource allocation, investigating irrigation strategies under uncertain water availability and improving the lower bound of total revenue intervals. Bao Shuzhong [2] analyzed the impact of agricultural planting factors on planting structures and proposed optimization strategies, emphasizing the significance of agricultural planting structures for both economic and ecological outcomes, providing valuable references for related research. However, few studies have formulated suitable planting strategies that account for crop substitutability and legume complementarity, which could assist farmers in flexibly adjusting planting structures in response to market demands and environmental changes. Moreover, the complementarity of legumes contributes to improving soil quality and increasing yields.

In this context, this study focuses on farmland in a village in the North China region, considering market fluctuation factors and employing fuzzy C-means clustering to classify crop categories and explore their substitutability. The innovation of this study lies in the application of fuzzy C-means clustering, which enables more precise identification of similarities and substitutability among crops, thereby providing farmers with more targeted planting recommendations. Additionally, the study incorporates a per-unit yield function that accounts for legume complementarity and utilizes a generalized additive model to predict crop sales volumes, effectively capturing nonlinear relationships and improving prediction accuracy. By leveraging these models to investigate the impact of these factors on planting strategies, this research contributes to achieving efficient, ecological, and economically coordinated agricultural development.

2. Establishment of optimization models under basic conditions

2.1. Data source

The data in this article come from 1,201 acres of cultivated land in a village in North China in 2023, with 34 plots of different sizes. The data source comes from https://www.mcm.edu.cn/html_cn/node/a0c1fb5c31d43551f08cd8ad16870444.html

2.2. Establish decision variables, objective functions and constraints

Determining the optimal crop planting plan for 7 years is an optimization problem. The decision variable in this article is x_{ijs}^{t}, which represents the area to plant the $j-th$ crop in the t year, $s$season, and i plot. There are three types of cultivated land in the study area: irrigated land, ordinary greenhouses and smart greenhouses. Because of the different planting seasons, this article introduces seasonal parameters s: $s=0$: single-season planting, $s=1$: first season of two-season planting, and $s=2$: second season of two-season planting. In order to optimize the crop planting plan from 2024 to 2030, it is necessary to ensure that the total income in seven years is maximized. Assuming the total income is W, the calculation formula is:

$$W=\sum_{t} W_{t}=\sum_{t}\left(Sale_{t}-Cost_{t}\right) \tag{1}$$

If the total output of a certain crop exceeds the expected sales volume, the excess will be sold at a reduced price of 50% of the 2023 sales price. The formula for calculating sales is:

$$\mathrm{Sale}_{t}=\sum_{j}\left\{\begin{array}{l}\left\{min\left[\left(\sum_{i,s} x_{ijs}^{t}\cdot N_{ijs}-\sum_{i,s} x_{ijs}^{2023}\cdot N_{ijs}\right),0\right]+\sum_{i,s} x_{ijs}^{2023}\cdot N_{ijs}\right\}\cdot P_{j}\\+\frac{1}{2}max\left[\sum_{i,s} x_{ijs}^{t}\cdot N_{ijs}-\sum_{i,s} x_{ijs}^{2023}\cdot N_{ijs}),0\right]\cdot P_{j}\end{array}\right\} \tag{2}$$

The final objective function is:

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

Where x_{ijs}^{t} denotes the area planted with the j −th crop in year t, season s, and plot i. N_{ijs} is the acre yield of the j −th crop in plot i in season s, P_j is the average unit price of the j-th crop, and Q_{ijs}^{t} is the cost per acre of the j-th crop in plot i in season s in year t .

Based on the climate characteristics and crop growth patterns of North China, this article sets the following constraints:

i. Crop acreage requirements: the acreage of each crop in a plot shall be less than the total area of the plot and greater than 0. To prevent insufficient production per unit area, the acreage shall not be less than one-tenth of the total area.

$$\sum_{j} x_{ijs}^{t}\leq A_{i},\quad x_{ijs}^{t}\geq\frac{1}{10}A_{i} \tag{4}$$

ii. Crop rotation requirements: to maintain soil health, continuous planting of the same crop is prohibited. This paper sets the following constraints on smart greenhouses:

$$\begin{cases}x_{ijs=1}^{t}\times x_{ijs=2}^{t}=0\\x_{ijs=2}^{t-1}\times x_{ijs=1}^{t}=0\end{cases} \tag{5}$$

iii. Requirements for legume planting: In order to enhance soil health, it is stipulated that each plot of land should be planted with legumes at least once in three years.

$$\sum_{j=bean}\left(x_{ijs}^{t-1}+x_{ijs}^{t}+x_{ijs}^{t+1}\right)>0 \tag{6}$$

iv. Planting concentration requirements: To facilitate field management, it is assumed that the number of crops planted per plot is limited to three.

Combining the uncertainty of expected crop sales, acreage, planting costs, and selling prices, this paper introduces the following volatility factors:

i. Expected sales V_j^t:The annual growth rate R_{tj} for wheat and corn follows a normal distribution $N(7.5\%, (1.2\%)^2)$.

ii. Per-unit yield N_{ij}^t: Influenced by natural factors, the seasonal fluctuation rate Y_{tjs} of per-unit yield follows a normal distribution $N(0, (4\%)^2)$.

iii. Planting costs C_j^t :The annual growth rate E_{tj} follows a normal distribution $N(5\%, (0.5\%)^2)$. Sales price P_j^t.

a) Food crops: The price change rate F_j^t follows a normal distribution $N(0, (0.5\%)^2)$.

b) Vegetable crops: The annual growth rate F_j^t follows a normal distribution $N(5\%, (0.5\%)^2)$.

c) Edible fungi crops: Prices decrease by 1%-5% annually, with the change rate F_j^t following a normal distribution $N(3\%, (1\%)^2)$.When j is morels, the price change rate is fixed at 5%.

Scenario 1 is the optimal crop planting plan considering only the above mark et fluctuations:

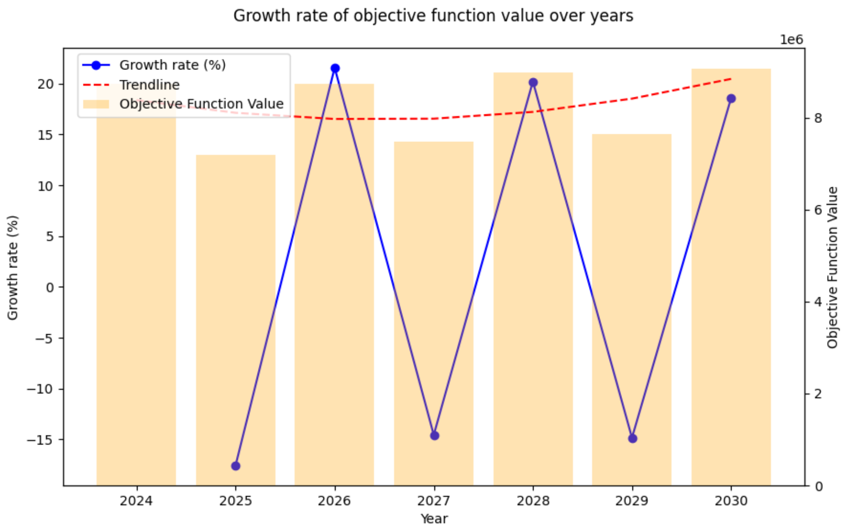

Fig. 1 Situation 1 Optimal planting plan

According to Figure 1, the total income in seven years is approximately 562,748,01.8276 yuan, considering only the above market fluctuations and constraints. The column chart in Figure 2 shows the income from crop planting in each year from 2024 to 2030, while the broken line represents the annual growth rate of income in each year.

3. Explore the substitutability, complementarity and relevance of crops

3.1. Analysis of crop substitutability based on fuzzy C-mean clustering

Fuzzy C-Means Clustering (FCM) is a soft clustering method that allows data points to belong to multiple clusters with varying degrees of membership. Unlike hard clustering algorithms such as K-Means, FCM assigns a membership value to each crop, indicating the degree to which it belongs to each cluster. This soft clustering characteristic makes FCM more flexible in handling data, especially when crop features are similar [3]. FCM performs exceptionally well when dealing with real-world fuzzy data and is widely applied in fields such as data mining, machine learning, and image processing. Liu Zhiguo et al. [4] studied the theory and experiments of the Fuzzy C-Means clustering algorithm, highlighting its simplicity in design and broad applicability while pointing out its tendency to fall into local optima. Liu Xiaona et al. [5] applied the Fuzzy C-Means clustering algorithm to the classification of university students' hardship levels. By calculating fuzzy similarity matrices and fuzzy equivalence matrices and validating the classification results with empirical methods, they enhanced the scientific rigor and accuracy of the classification process. Zhang Xiang et al. [6] applied the Fuzzy C-Means clustering algorithm to employee performance evaluation, demonstrating that the algorithm not only reduces management costs but also offers greater flexibility and accuracy, making it superior to traditional methods. Given the unsupervised nature, multi-indicator characteristics, and classification complexity of the data in this study, the Fuzzy C-Means clustering algorithm was also employed.

Given a data set $X = \{x_1, x_2, \cdots, x_n\}$, containing n data points, each data point x_i is a $d$dimensional feature vector. The goal of the FCM algorithm is to divide the data set into c clusters, with the cluster center being $V = \{v_1, v_2, \ldots, v_c\}$.

The goal of FCM is to minimize the following objective function:

$$J(U,V) = \sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij}^{m} \left\| x_i - v_j \right\|^2 \tag{7}$$

Where u_{ij} is the membership degree of data point x_i to cluster j. m is the blur factor, usually greater than 1. $\| x_i - v_j \|$ is the distance from the data point x_i to the cluster center v_j. The specific steps of the algorithm are as follows:

Input: number of clusters c, data set X, stopping threshold ϵ, blur factor m, maximum number of iterations T.

Initialization: Randomly select $c$samples as the initial clustering center V, and initialize the membership matrix U.

Update the membership matrix:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\|x_i - v_j\|}{\|x_i - v_k\|} \right)^{\frac{2}{m-1}}} \tag{8}$$

i. Update cluster center:

$$v_j = \frac{\sum_{i=1}^{n} u_{ij}^{m} x_i}{\sum_{i=1}^{n} u_{ij}^{m}} \tag{9}$$

ii. Check the convergence conditions: calculate the objective function values $J(t)$and $J(t-1)$. If $|J(t) - J(t-1)| \leq \epsilon$is met, stop iteration; otherwise, return to step 3.

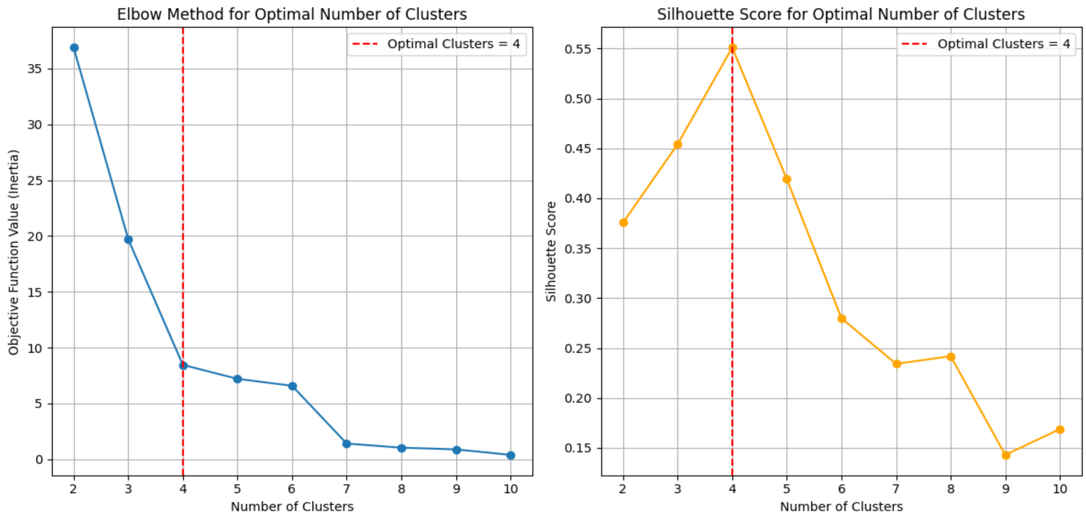

Fig. 2 Grain crop elbow plot (left) and contour coefficient plot (right)

The left side of Figure 2, using the Elbow Method, shows that when the number of clusters k=4, the "Within-Cluster Sum of Squares" (WSS) decreases more slowly, forming an "elbow." On the right side, the silhouette score reaches its highest value at k=4, indicating the best clustering performance. Combining both, this study selects k=4 as the optimal number of clusters.

This article sets the initialization fuzzy factor to 2, and the iteration allowable error ϵ to 1×10^{-5}. After iteratively calculating and updating the membership matrix and cluster center, the following grain clustering results are obtained (Figure 3).

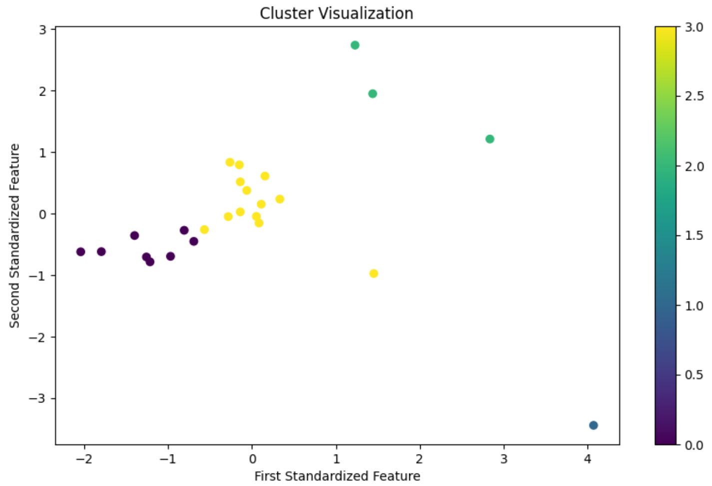

Fig. 3 Grain clustering renderings

The specific crops in each cluster and their related data are shown in Table 1:

Tab. 1 Grain clustering result table

Category	Total Planting Area	Average Planting Cost per Acre	Average Yield per Acre	Average Selling Price	Cluster
Soy	147	400	380	3.25	4
Green beans	96	350	340	7	3
Wheat	222	450	760	3.5	4
…	…	…	…	…	…
Sweet potato	18	2000	2000	3.25	1
Rice	42	680	500	7	3

Mushrooms and root vegetables share similar growing conditions, leading to comparable cooking methods. This article examines their potential as substitutes.

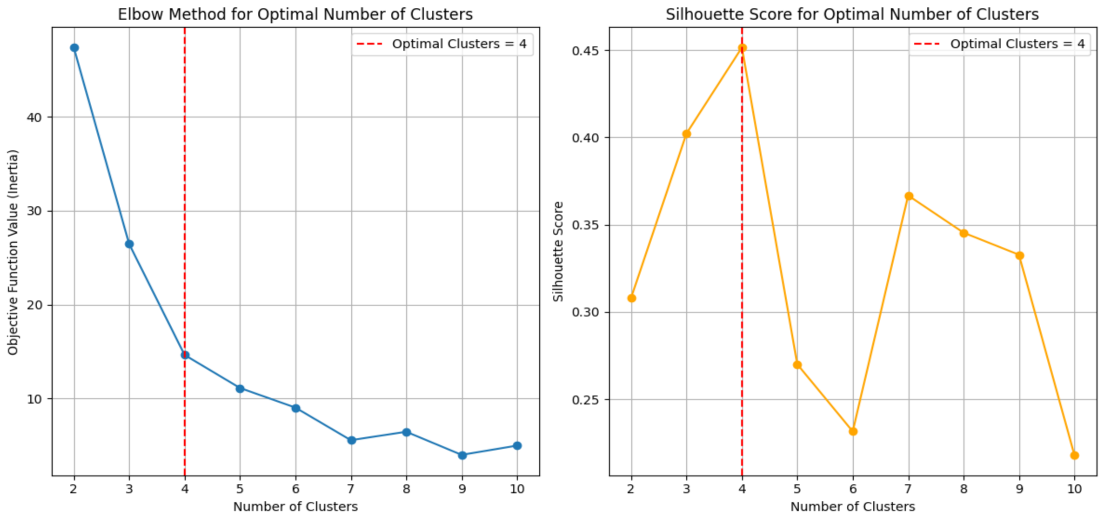

Fig. 4 Elbow plot (left) and profile coefficient plot (right) for vegetable and mushroom crops

As can be seen from Figure 4, the elbow method (left picture) shows that at 4 clusters, the objective function value drops significantly and levels off, indicating that this is a reasonable choice of the number of clusters. The silhouette coefficient method (right picture) also reaches the highest value when there are 4 clusters, which further supports the conclusion that 4 clusters are selected as the optimal number of clusters. The clustering results are shown in Figure 5 and Table 2:

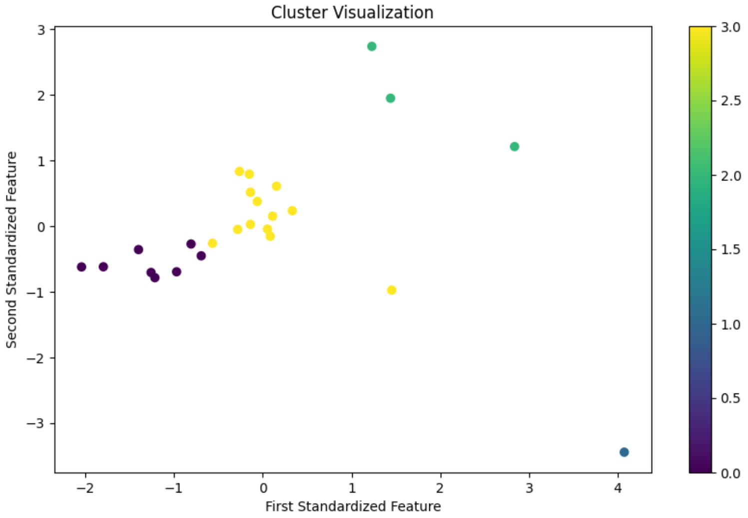

Fig. 5 Vegetables and mushrooms clustering results chart

Tab. 2 Vegetables and mushrooms clustering results table

Category	Total Planting Area	Average Planting Cost per Acre	Average Yield per Acre	Average Selling Price	Cluster
Cowpea	11.8	2360	3350	8.3	3
Sword bean	13.2	1133.33	2266.67	6.75	3
Potato	15	2000	2000	3.5	3
…	…	…	…	…	…
Water spinach	0.3	5500	11000	5.3	4
Celery	0.3	1200	6000	4.4	1

Based on the crop classification standards of the Ministry of Agriculture and Rural Affairs of China, clustering analysis results, and actual production and living conditions, this study ultimately classifies 41 crops into 6 categories, with the specific classification results shown in Table 3.

Tab. 3 Final crop classification table

Crop Classification	Crop Name
Beans	Soybeans, Black Beans, Red Beans, Mung Beans, Climbing Beans, Cowpeas, Sword Beans, Kidney Beans
Cereals	Wheat, Corn, Millet, Sorghum, Millet, Buckwheat, Oatmeal, Barley, Rice
Rhizome	Pumpkin, Sweet Potato, Potato, White Radish, Carrot
Leafy Vegetables	Spinach, Lettuce, Green Pepper, Water Spinach, Cabbage, Cabbage, Celery, Chinese Cabbage, Lettuce, Baby Greens
Fruit Vegetables	Tomatoes, Eggplants, Peppers, Cucumbers, Cauliflower
Mushrooms	Mushrooms, Shiitake Mushrooms, White Mushrooms, And Morels

Due to the substitution effect between similar crops, this article believes that excessive total output of similar crops of a crop will lead to a decrease in the price of the crop. That is, this article considers the overall impact of the output of similar crops on the crop. The influencing factors are as follows:

$$M_k^t = \frac{\sum_j^{j\in k} V_j^t}{\sum_j^{j\in k}\left(\sum_{i,s} x_{ijs}^t \cdot N_{ijs}^t\right)} \tag{10}$$

Among them, the $j-$th crop belongs to the $k-$th category, $\sum_{j}^{j\in k} V_j^t$ represents the total expected sales volume of the $k-th$ crop in t years, and $\sum_{j}^{j\in k} \left(\sum_{i,s} x_{ijs}^t \cdot N_{ijs}^t\right)$ represents the total output of the $k-$th crop in t years.

In addition to considering the overall impact of the yield of similar crops on the crop, in order to further consider the relationship between the sales price of the crop and the yield of the crop itself and the total yield of similar crops, this article makes the following assumptions:

If the total output of type $k-th$ crops is less than the expected sales volume of type $k-th$ crops, and the total output of type $j-th$ crops is less than the expected sales volume of type $j-th$ crops, then the sales unit price of type $j-th$ crops $P'_{tj}$is:

$$P'_{tj} = M_k^t \cdot P_j^t \tag{11}$$

Where P_j^t is the average unit price of the $j-th$ crop in year t.

(1) If the total yield of the k-th crop is less than its expected sales volume, but the total yield of the j-th crop exceeds its expected sales volume, then the selling price per unit P'_{tj} for the portion of the j-th crop's total yield that does not exceed its expected sales volume is:

$$P'{}_{tj} = M_k^t \cdot P_j^t \tag{12}$$

The excess sales price P'_{tj} is:

$$P'{}_{tj} = 0.5 M_k^t \cdot P_j^t \tag{13}$$

The corresponding excess output is:

$$L_j^t = \sum_{i,s} x_{ijs}^t \cdot N_{ijs}^t - V_j^t \tag{14}$$

(2) If the total yield of the k-th crop exceeds its expected sales volume, but the total yield of the j-th crop is less than its expected sales volume, then the selling price per unit P'_{tj} for the j-th crop is:

$$P'_{tj} = M_k^t \cdot P_j^t \tag{15}$$

(3) If the total yield of the k-th crop exceeds its expected sales volume, and the total yield of the j-th crop also exceeds its expected sales volume, then the selling price per unit P'_{tj} for the portion of the j-th crop's total yield that does not exceed its expected sales volume is:

$$P'{}_{tj} = M_k^t \cdot P_j^t \tag{16}$$

For the part of the total crop output per season that exceeds the corresponding expected sales volume, this article divides it into two categories: a. The part that does not lead to the excess of the total output of the k-th crop; b. The part that causes the total output of the k-th crop to exceed. There is a difference in the price of the crop in the two scenarios:

(1) The sales unit price P'_{tj}that does not result in an excess of the total output of category $k-th$ crops is:

$$P'{}_{tj} = 0.7 M_k^t \cdot P_j^t \tag{17}$$

The corresponding output of this part is:

$$L_j^t = \sum_{i,s} x_{ijs}^t \cdot N_{ijs}^t - V_j^t - \left[\sum_{j}^{j\in k} \left(\sum_{i,s} x_{ijs}^t \cdot N_{ijs}^t\right) - \sum_{j}^{j\in k} V_j^t\right] \cdot \frac{max\left(\sum_{i,s} x_{ijs}^t \cdot N_{ijs}^t - V_j^t, 0\right)}{\sum_{j}^{j\in k} max\left(\sum_{i,s} x_{ijs}^t \cdot N_{ijs}^t - V_j^t, 0\right)} \tag{18}$$

(2) The sales unit price P'_{tj}that results in the excess of the total output of category $k-th$ crops is:

$$P_{tj}' = 0.5 M_k^t \cdot P_j^t \tag{19}$$

The corresponding output of this part is:

$$L_j^t = \left[\sum_j^{j\in k} \left(\sum_{i,s} x_{ijs}^t \cdot N_{ijs}^t\right) - \sum_j^{j\in k} V_j^t\right] \cdot \frac{max\left(\sum_{i,s} x_{ijs}^t \cdot N_{ijs}^t - V_j^t, 0\right)}{\sum_j^{j\in k} max\left(\sum_{i,s} x_{ijs}^t \cdot N_{ijs}^t - V_j^t, 0\right)} \tag{20}$$

3.2. Complementarity analysis of pulses

This study explores the positive impact of legume crops on the per-unit yield of other plants. Through intercropping with other plants, legume crops can significantly enhance the yield of other plants by means such as nitrogen fixation, improving soil structure, reducing pests and diseases, increasing resource utilization efficiency, and suppressing weeds[7]. To facilitate model calculations, this study assumes that planting legumes on the same plot of land increases the per-unit yield of other crops, with the increase being related to the area planted with legumes. A per-unit yield function considering complementarity is constructed as follows:

$$N'_{ijst} = N_{ijs}^t \cdot \left(1 + \frac{\beta_{it}^3}{1+\beta_{it}^3}\right) \tag{21}$$

Here, N'_{ijst} represents the sales price of the j-th crop on the i-th plot in year t, considering complementarity. β_{it} denotes the proportion of legume planting area on the i-th plot in year t, specifically:

$$\beta_{it} = \frac{\sum_j^{j\in beans} x_{ijs}^t}{A_i} \tag{22}$$

3.3. Correlation Analysis Based on the Generalized Additive Model

To better analyze the expected sales volume of each crop, this paper adopts the Generalized Additive Model (GAM) to explore the impact of sales price, planting cost, and average yield per acre on expected sales volume. GAM is a flexible regression model capable of capturing the nonlinear relationships between the response variable and explanatory variables, making it suitable for addressing complex real-world problems. Zhao Weining et al., based on[8] highway accident data, constructed a Generalized Additive Model (GAM) and found that it outperformed the Generalized Linear Model (GLM) in terms of goodness-of-fit and predictive accuracy, while also revealing the nonlinear relationships of accident influencing factors. Ye Jianhong et al.[9] used GAM to study the competition intensity and influencing factors between shared electric bicycles and conventional buses, demonstrating that GAM has significant value in uncovering the competition patterns and regional differences of transportation modes. Dang Fujia et al.[10] proposed a GAM-based photovoltaic power station output prediction method, verifying its effectiveness in addressing the nonlinearity and uncertainty of power generation. Given the nonlinearity, multidimensionality, and uncertainty of the data in this study, the Generalized Additive Model is also employed for analysis.

The mathematical expression of the GAM model used in this article is as follows:

$$V_j^t = \alpha_0 + s\left(P_j^t\right) + s\left(Q_{ijs}^t\right) + s\left(N_{ijs}^t\right) + \varepsilon \tag{23}$$

In this model, $s\left(P_j^t\right)$, $s\left(Q_{ijs}^t\right)$and $s\left(N_{ijs}^t\right)$represent the smooth functions of sales price, planting cost and average mu yield respectively. This indicates a non-linear relationship between expected sales and these three variables. The intercept of the model $\alpha_0$is 30678, which represents the expected sales volume when all independent variables are zero, and its p-value is 1.84e-14, indicating that it is statistically significant.

Tab. 4 Smoothing item significance analysis table

Smooth term	degrees of freedom (edf)	F value	p value	Significance
s (average selling price)	1	7.559	0.0107	*
s (average planting cost)	8.405	34.39	<2e-16	***
s (average yield per mu)	4.125	17.725	9.46E-07	***

As shown in Table 4, the p-values for the smoothing terms s (Average Planting Cost) and s (Average Yield per Acre) are both less than 0.001, indicating that these two variables have a highly significant impact on the model. Meanwhile, the p-value for s(Average Selling Price) is 0.0107, suggesting that it is also statistically significant, but its impact is relatively weaker.

Tab. 5 Model fitting goodness index table

Norm	Adjusted R^2 (Adjusted R^2)	Deviance Explained	Generalized Cross Validation (GCV)
Value	0.895	93%	2.65E+08

Table 5 shows that the adjusted R^2 of the model is 0.895, indicating that approximately 89.5% of the variation in the response variable can be explained by the independent variables, showing a good fitting effect. The deviance explained was 93%, further supporting the validity of the model and demonstrating its excellent performance in capturing the variability in the data.

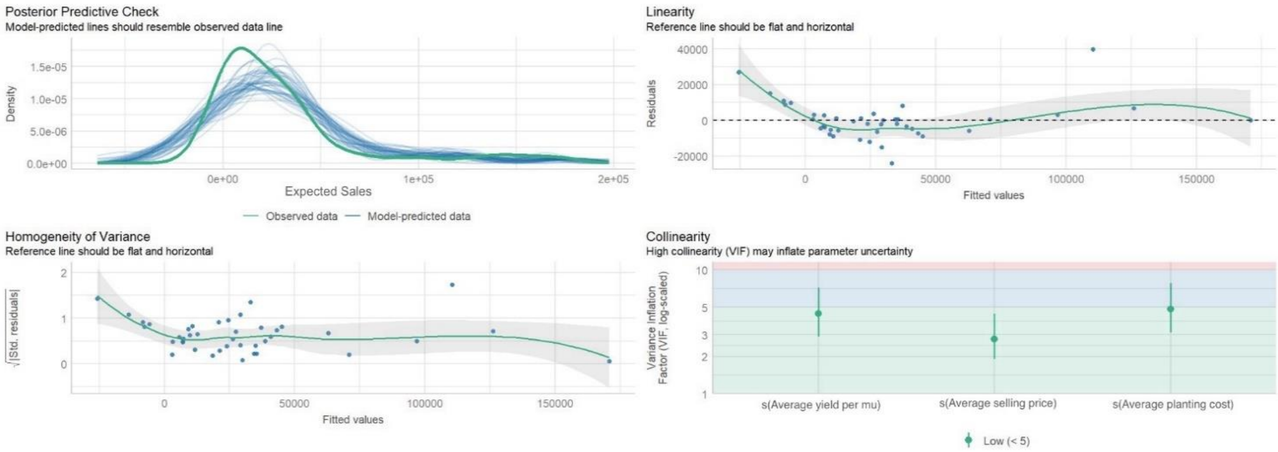

Fig. 6 Posterior prediction check (top left) and homogeneity of variance analysis (bottom left) Test of model linearity (top right) and collinearity (VIF) analysis (bottom right)

Figure 6 demonstrates the results of the model diagnostic metrics assessment: the a posteriori prediction check shows that the observed values are consistent with the predicted values, indicating that the model captures the data distribution well; the variance-alignment plot supports the assumption of variance-alignment; the linear relationship plot shows that the residuals fluctuate around zero, with a slight tendency; and the covariance analysis shows that the VIF values of all smoothing terms are lower than 5, with no serious covariance. Overall, the model is shown to be well fitted.

To evaluate the performance of the Generalized Additive Model (GAM) compared to the traditional linear regression model in predicting expected sales volume, this paper conducted a comparison of the two models. The following presents the output results of the model performance:

Tab. 6 Model performance comparison

Model	adj.R2	MSE
GAM	0.8948	110638452
Linear Regression	0.4747	772394409

Table 6 shows that GAM outperforms linear regression with a higher adjusted R^2 (0.895 vs. 0.475) and lower MSE (110638452). Figure 7 confirms this, as GAM predictions align more closely with actual values, while linear regression shows significant deviation. This highlights GAM's ability to capture nonlinear relationships and provide more reliable analysis.

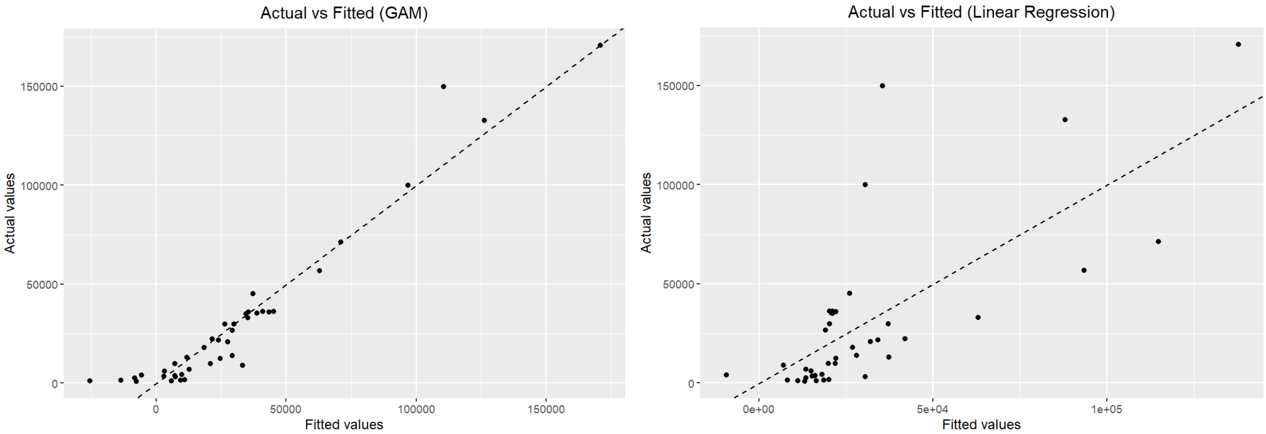

Fig. 7 Comparison of actual values and fitted values (left: GAM model, right: linear regression model)

In Scenario 2, after comprehensively considering the substitutability and complementarity among various crops, as well as the correlation between expected sales volume, sales price, and planting cost, the following objective function is derived:

$$W = \sum_t \left\{ \sum_j \left\{ \begin{matrix} \{min[(\sum_{i,s} x_{ijs}^t \cdot N'_{ijst} - V'_{tj}), 0] + \sum_{i,s} x_{ijs}^{2023} \cdot N'_{ijst}\} \cdot P'_{tj} \\ +\frac{1}{2} max[(\sum_{i,s} x_{ijs}^t \cdot N'_{ijst} - V_j^t), 0] \cdot P'_{tj} - C_j^t \end{matrix} \right\} \right\} \tag{24}$$

4. Results

Tab. 7 Revenue forecast table

Year	2024	2025	2026	2027	2028	2029	2030	Total
Net Income	20837163.07	13021789.51	19913864.23	12287379	19337525	13983132	17540528	116921381.3

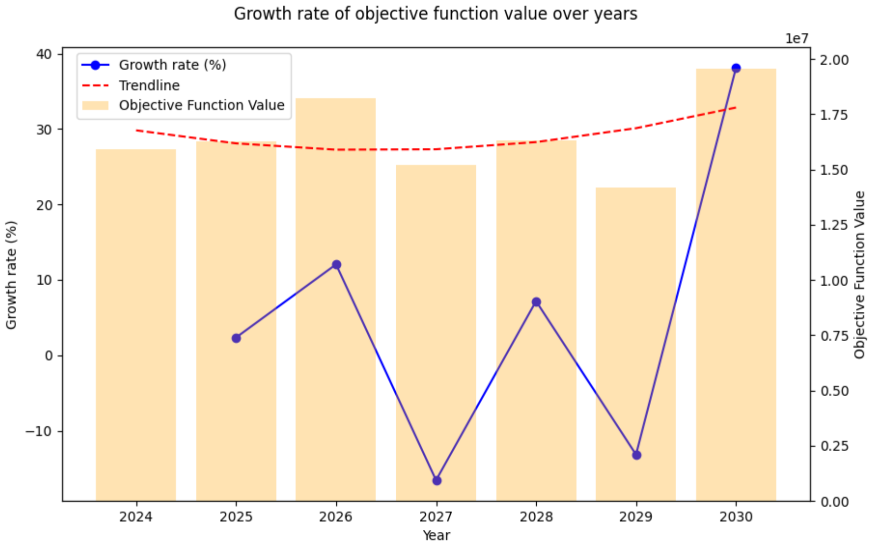

Fig. 8 Results chart of the optimal planting plan for case 2

As shown in Table 7 and Figure 8, the total revenue from crop cultivation over seven years reached 116,921,381.30 yuan. Comparing this with the total revenue under the baseline scenario, this study found that the total revenue, when considering substitutability, similarity, and complementarity, was significantly higher than the baseline. This phenomenon can be attributed to two main factors:

First, in the analysis of Scenario 2, this study fully accounted for the substitutability of similar crops. This means that when selecting crops, the planting strategy could be adjusted more flexibly to allocate resources to crops with higher profit potential. This flexibility allowed for the expansion of planting areas for certain high-revenue crops, thereby significantly increasing overall revenue.

Second, the model incorporated the cultivation of legume crops. These crops not only improve soil fertility but also promote the growth of other crops. By reasonably integrating legume crops into the same land, the total yield of crops can be effectively increased, thereby enhancing overall economic returns[11]. This diversified planting strategy not only optimized the growing environment for crops but also brought higher income to farmers.

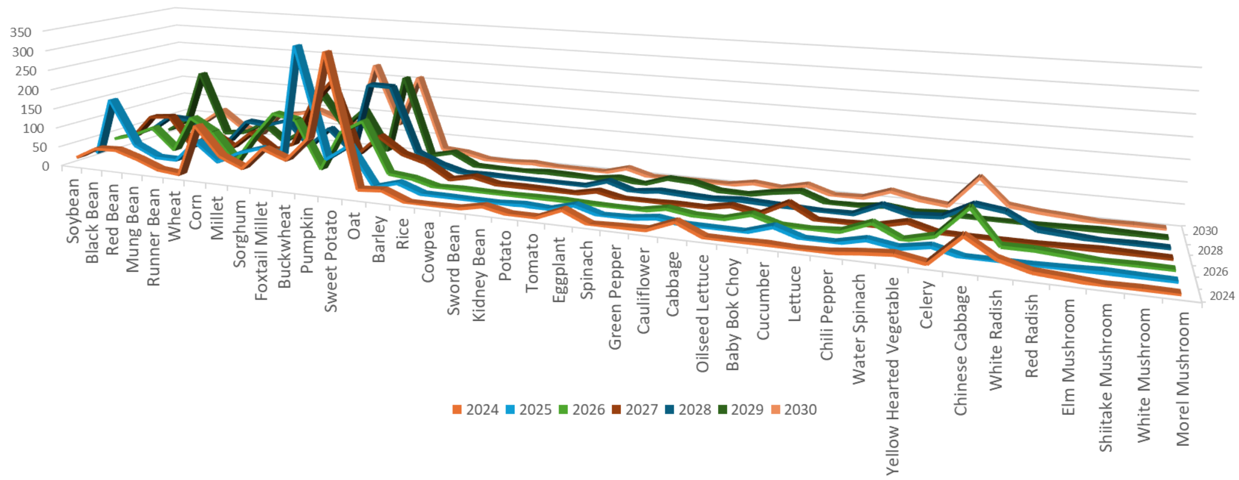

Fig. 9 Scenario 1 Three-dimensional line chart of crop planting area

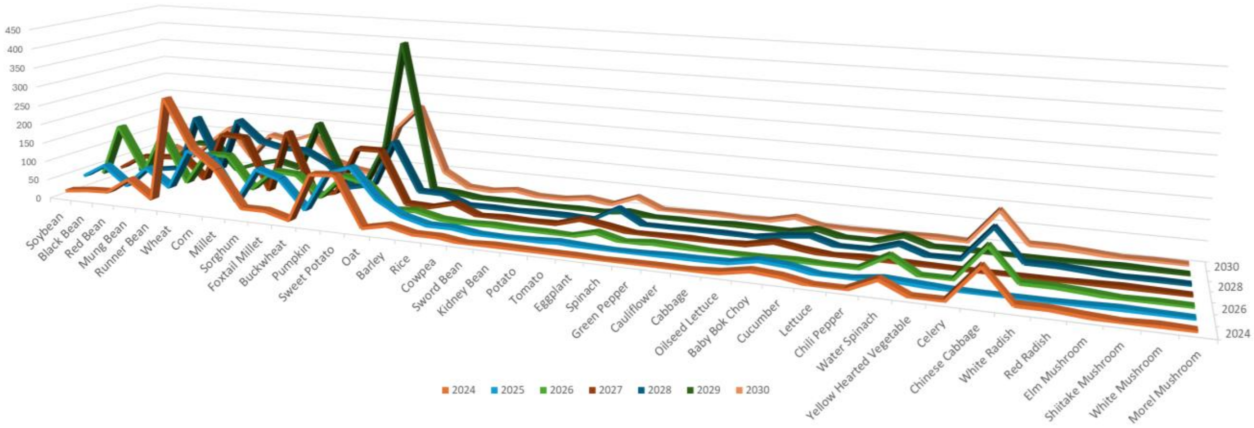

Fig. 10 Scenario 2 Three-dimensional line chart of crop planting area

According to Figures 9 and 10, in scenario 1, which only considers market fluctuations, soybean, corn, sweet potato and pumpkin acreage is consistently higher, reflecting their importance in the market. When considering the substitutability of products and the complementarity of beans, the planting area of some cereal crops and root crops remained high and stable, and the fluctuation decreased, while the planting area of cowpea and knife bean increased, indicating the flexible adjustment in the optimization plan. In Figure 9, the planting area fluctuates greatly. In particular, soybean and corn have significant peaks and valleys in some years, which may lead to greater risks. As shown in Figure 10, the fluctuation range of planting area decreased significantly and tended to be stable overall, indicating that the planting decision was more rational and effective after optimization.

5. Conclusions

In this study, the application of fuzzy C-mean clustering and generalized additivity model is used to conduct an in-depth analysis of crop cultivation schemes in a village in North China, demonstrating how to achieve a coordinated development of high efficiency, ecology and economy in a complex agricultural environment. The results of the study show that considering the substitutability between crops and the complementarity of legumes can optimize the allocation of resources, improve the utilization rate of arable land, and reduce the risk of cultivation. In summary, the practical value of this study is to provide farmers with a scientific prediction model to obtain a planting plan, which helps them flexibly adjust the planting structure according to the market demand and environmental changes, so as to realize higher economic and ecological benefits.

Although this study has achieved certain results, there are still some shortcomings. For example, the applicability of the model may be limited by the climatic and soil conditions in specific regions, and future research could consider validating the model in a wider geographical area. In addition, further empirical studies could incorporate more environmental factors and market dynamics to improve the predictive ability and adaptability of the model.

References

[1] Xi Kaiyan, Yu Jialu, Zhang Min, et al. Optimal allocation of oasis water resources based on interval two-stage risk-averse stochastic programming [J]. Agricultural Research in the Arid Areas, 2021, 39(6): 224-229,265.

[2] Bao Shuzhong. Influence of agricultural planting factors on planting structure and optimization strategies [J]. Agriculture and Technology, 2020, 40(13): 90-91.

[3] Lei Guangyuan, Zhang Tao, Tang Yongcong, Liang Te, Shu Kexin. Analysis method of electricity consumption characteristics of distribution area users based on optimized fuzzy C-means clustering [J]. Proceedings of the CSU-EPSA, 2024, 36(1): 99-105.

[4] Liu Zhiguo. A fuzzy C-means clustering algorithm and its implementation [J]. Modern Navigation, 2020, 11(2): 122-125.

[5] Liu Xiaona, Wang Kai, Xu Yanqiang. Application of fuzzy C-means clustering algorithm in the identification of college students' difficulty levels [J]. Journal of Lanzhou University of Arts and Science (Natural Science Edition), 2023, 37(4): 64-67.

[6] Zhang Xiang, Li Guochang. Application of fuzzy C-means clustering algorithm in employee performance evaluation: A case study of sales staff in YH Company [J]. Journal of Anhui Polytechnic University, 2020, 35(2)78-84.

[7] Man Benju, Wang Jianpeng, Liu Jili, et al. Study on nitrogen absorption and utilization characteristics of potato-legume intercropping system [J]. Agricultural Research in the Arid Areas, 2022, 40(2): 153-162.

[8] Zhao Weining, Huo Xiaoyan, Gu Fangde, et al. Analysis of influencing factors of highway traffic accidents based on generalized additive model [J]. Highway Traffic Technology, 2020, 36(3): 125-131.

[9] Ye Jianhong, Bai Jiahao, Qin Yifei. Competition intensity and influencing factors between shared electric bicycles and conventional buses based on generalized additive model [J]. Transportation and Transport, 2024, 40(6): 6-13.

[10] Dang Fujia, Lai Jieheng. Photovoltaic power station prediction method based on generalized additive model [J]. Hongshui River, 2022, 41(6): 77-81.

[11] Xu Xiuyuan, Zhou Hao. Optimization of land use in the Pearl River Delta region based on grey prediction and linear programming [J]. Land and Resources Herald, 2024, 21(2): 79-85.

[12] Song Xuefen. Study on high-yield and high-efficiency planting strategies for crops in the sandy ginger black soil test area of Mengcheng [J]. Anhui Agricultural Science, 1995,23(3): 229-231.