

# Multi-source Uncertainty Quantification and Spatiotemporal Synergistic Allocation of Crops in Complex Agricultural Systems Based on Large-scale Mixed Integer Programming

Sihan Lyv<sup>1)</sup> Renwen Lu<sup>2)</sup> Yunyahui Zhao<sup>3)</sup>

<sup>1)</sup>School of Humanities and Social Sciences, Harbin Engineering University

<sup>2)</sup>National Characteristic Demonstration School of Software, Harbin Engineering University

<sup>3)</sup>School of Intelligent Systems and Science and Engineering, Harbin Engineering University

## Abstract

This study addresses the seven-year planting optimization problem for 41 crops across 54 land plots in a specific village. Aiming to transcend traditional static planning assumptions, we construct a decision-making framework incorporating market uncertainty and inter-crop linkage effects, providing theoretical support and empirical evidence for agricultural planting optimization.

In Problem 1, this paper employs Hierarchical Clustering to categorize the 41 crops based on 2023 data. Using Euclidean distance to measure economic similarity, the crops are divided into four groups of equivalent economic value. For the single-year planning of 2024, a Mixed Integer Linear Programming (MILP) model is established, incorporating constraints such as plot compatibility, seasonal limitations, and legume rotation. To balance economic efficiency and operational feasibility, a two-stage solution strategy is adopted: the first stage solves for the theoretical optimal profit; the second stage minimizes the number of planting state variables to enhance the scheme's implementation feasibility, under the premise that the profit remains no less than 95% of the optimal solution. Under single-season planting conditions, the optimal profit is found to be 57,459,362 CNY. Under multi-season planting conditions, by fully leveraging the seasonal advantages of irrigated land and greenhouses, the optimal profit reaches 98,462,192 CNY.

In Problem 2, after an in-depth analysis of the diverse variation trends of agricultural products, climate factors are introduced. Climate data from 2013 to 2023 for a region in North China, including key indicators such as temperature, precipitation, and sunshine duration, were collected via official government statistical reports. Combined with the sales volumes of wheat and corn recursively deduced for 2013-2023, Multiple Regression Analysis is performed to determine that climate is the primary driver of the diverse variations in agricultural products in the region. Accordingly, based on the economic categories defined in Problem 1, and considering weather impacts and environmental adaptability, crops are further classified into Cold-adaptive, Low-temperature Tolerant, and Cool-temperature Fragile types. Within the given fluctuation ranges, and combining climate and market influences to determine more specific fluctuation intervals and development trends for each category, a large-scale MILP model containing approximately 85,000 decision variables and 90,000 constraints is established. The solution yields a seven-year optimal total profit of 103,034,611 CNY.

In Problem 3, to overcome the limitations of deterministic models in ignoring market fluctuations, this paper constructs a structured prediction model based on the coupling of Geometric Brownian Motion (GBM) and a Cross-Price Elasticity Matrix. Based on historical data calibration of GBM parameters, a cross-price elasticity matrix for the 41 crops is constructed to characterize substitution and complementarity relationships. Through 1,000 Monte Carlo Simulations, random paths for price, cost, and yield are generated, and dynamic supply-demand coupling is realized using market clearing conditions. The optimization yields a seven-year total profit of 100,229,466 CNY, a decrease of 2.7% compared to the deterministic model. To evaluate model robustness, robustness tests are further conducted by introducing random noise of varying intensities into the Monte Carlo simulations. The results indicate that the contribution of price uncertainty to profit volatility is approximately four times that of yield uncertainty, and the model's reliability shows an inflection point when relative uncertainty reaches 10%.

The main contributions of this study lie in: proposing a framework coupling GBM with a Cross-Price Elasticity Matrix to break traditional static assumptions; quantifying the relative importance of price and yield risks to provide a scientific basis for the allocation of risk management tools; and identifying critical thresholds for model reliability to clarify the applicable boundaries of different optimization methods. The constructed methodology possesses strong generalizability and can be extended to other agricultural regions and crop combinations.

**Keywords:** Agricultural Planting Optimization; Hierarchical Clustering; Mixed Integer Programming; Multiple Linear Regression; Geometric Brownian Motion; Cross-Price Elasticity; Monte Carlo Simulation; Robustness Analysis

## 1 Problem Restatement

### 1.1 Background

Against the backdrop of the ongoing Rural Revitalization Strategy, the development of mountainous agriculture places higher demands on the rationality of crop planting structures. Given limited arable land resources and distinct natural conditions in mountainous areas, optimizing planting schemes under specific constraints directly impacts agricultural production efficiency and farmers' income levels. By selecting suitable crops adapted to local conditions and balancing economic benefits with risk control, it is possible to simplify field management processes and reduce planting uncertainty, thereby promoting sustainable agricultural development and the long-term stable growth of the rural economy.

### 1.2 Given Information

- Different types of cultivated land have varying cropping capacities within a year, which in turn affects crop planting arrangements.
- Planting decisions must align with the growth laws and seasonal characteristics of crops to ensure the stability of agricultural production and the sustainability of the rural economy<sup>[1]</sup>.
- The formulation of planting schemes must integrate economic factors such as expected sales volume, planting costs, yield per unit area, and selling prices.
- Crops must be sold within the current season. If actual production exceeds the expected sales volume, the surplus cannot be sold through normal channels.
- The appendices provide data on the village's existing arable land structure, previously planted crop types, and crop planting and sales situations for 2023. They also explicitly define the correspondence between different land types and suitable crops.

### 1.3 Problem Formulation

**Problem 1** Assuming that expected sales volume, planting costs, yield per unit area, and selling prices remain relatively stable, formulate optimal planting schemes for the next seven years under the following two scenarios: First, when total production exceeds expected sales volume, the surplus cannot be sold and is considered waste. Second, when total production exceeds expected sales volume, the surplus is sold at a discounted rate of 50

**Problem 2** Considering a more realistic market environment, introduce uncertainty into factors such as expected sales volume, planting costs, yield per unit area, and selling prices, and reformulate the optimal planting scheme for the next seven years.

**Problem 3** Further consider potential complementarity and substitutability among crops in actual production and markets, as well as the correlations between expected sales volume, selling prices, and planting costs. Based on this, analyze the simulated data to propose a new seven-year planting scheme, and compare and discuss the results with those obtained in Problem 2.

### 1.4 Model Assumptions

To ensure the operability of the model, this study is based on the following assumptions:

(1) Time and Sales Assumption: Sales periods for single-season crops are concentrated in September-October; for the first season crops in June-July; and for the second season crops in November-December. All crops are sold within the harvest year, and no cross-year inventory exists.

(2) Cost and Yield Assumption: Crop loss rates are fixed (based on appendix data), where Actual Unit Cost = Planting Cost / (1 - Loss Rate). Unit yield follows historical distributions, excluding extreme climate events.

(3) Planting Constraint Assumption: Land types (Dry Land, Terraced Fields, Irrigated Land, Greenhouses) remain unchanged during the planning period. Leguminous crops require rotation every three years on the same plot. The planting area of a single crop must not exceed 20% of the total area. Vegetables and edible fungi cannot be planted continuously for more than two years on the same plot.

(4) Market Assumption: The village acts as a price taker and does not influence market prices; farmers can sell all output at the market price (subject to the sales constraints defined in the problems).

The above assumptions are universal in agricultural planning problems, and the impacts of certain assumptions have been evaluated via noise analysis in the robustness test.

## 2 Data Preprocessing and Descriptive Statistical Analysis

Data quality serves as the cornerstone of model reliability. Prior to establishing the optimization model, we performed cleaning, standardization, and Exploratory Data Analysis (EDA) on the dataset provided in the appendices. This process aims to eliminate invalid noise and uncover inherent distributional patterns within the data.

### 2.1 Numerical Processing of Unstructured Price Data

In the original data (Attachment 2), the sales prices of crops are presented as interval strings (e.g., “2.50-4.00”), which cannot be directly used for linear programming computations. To convert interval-type data into deterministic parameters, we employ the mean approximation method.

Let the price interval for the  $i$ -th crop on plot  $j$  be  $[P_{\min}^{(i)}, P_{\max}^{(i)}]$ . The benchmark calculation price  $P_{\text{avg}}^{(i)}$  is defined as:

$$P_{\text{avg}}^{(i)} = \frac{P_{\min}^{(i)} + P_{\max}^{(i)}}{2} \quad (1)$$

Post-processing, the dataset was augmented with three numerical feature columns: Min\_Price, Max\_Price, and Avg\_Price, providing a quantitative basis for the subsequent construction of the revenue function.

### 2.2 Analysis of Agricultural Resource Structure

Through statistical analysis of the arable land data in Appendix 1 (see Figure 1), we observe that the land resources in this village exhibit significant heterogeneity.

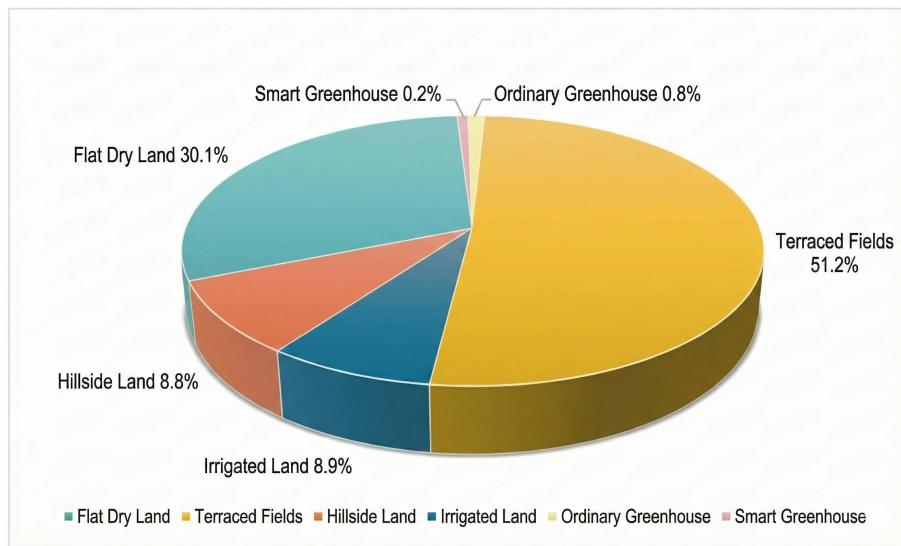


Fig 1 Distribution structure of village arable land types

The analysis indicates that high-efficiency facility agriculture lands, such as ordinary greenhouses and smart greenhouses, constitute a very low proportion, whereas traditional agricultural lands, such as dry flatlands and terraced fields, dominate. This resource constraint implies that maximizing the village's total revenue cannot rely solely on expanding the cultivation of high-value crops; instead, potential must be unlocked by optimizing crop rotation combinations on traditional land plots.

### 2.3 Baseline Analysis of the 2023 Planting Structure

To understand the village's existing planting habits and patterns, we conducted a statistical analysis of the actual planting data from 2023.

As shown in Figure 2, the village cultivates a diverse range of crops, covering food crops (legumes, cereals) and cash crops (vegetables, edible fungi). Notably, food crops occupy a large planting area, reflecting the foundational status of ensuring food security.

Regarding seasonal distribution characteristics, statistics based on Appendix 2 reveal a balanced trend: there are 30 entries for the second season, 29 for the first season, and 28 for single-season crops. This data structure indicates that the village maintains relatively stable production intensity across different seasons. Furthermore, most plots suitable for multi-season planting (e.g., irrigated land, greenhouses) have been fully utilized, with no significant seasonal idleness observed.

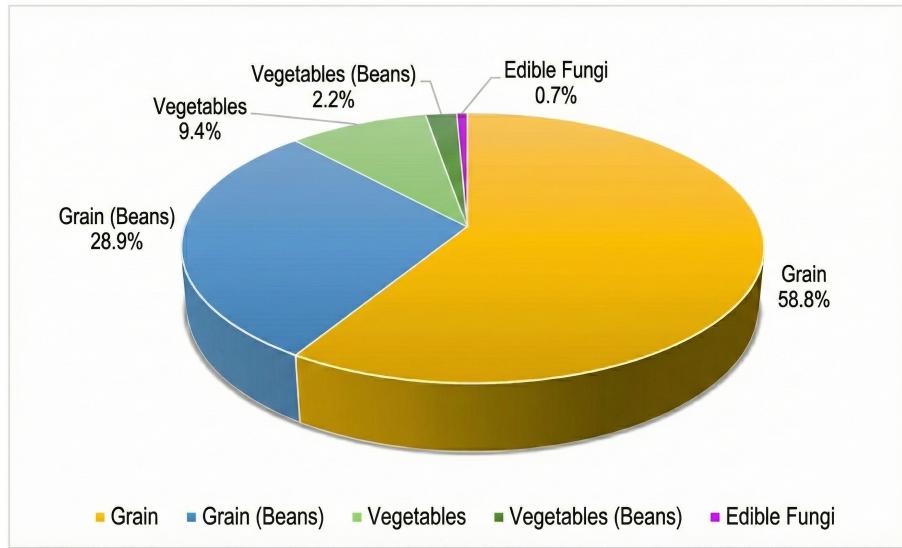


Fig 2 Distribution of crop types in 2023

## 2.4 Statistical Characteristics of Key Economic Indicators

To understand the economic potential of different crops, this paper conducts a descriptive statistical analysis on the 107 cleaned crop data entries. Table 1 presents core statistics for yield per mu, planting cost, and average selling price.

Table 1 Descriptive Statistics of Key Crop Economic Indicators (2023 Benchmark)

Indicator (Unit)	Mean	Std Dev	Min	Median	Max
Yield (jin/mu)	2990.84	3047.93	100.00	2400.00	15000.00
Planting Cost (CNY/mu)	1686.07	1636.14	350.00	1320.00	10000.00
Avg. Price (CNY/jin)	8.20	11.91	1.50	6.00	100.00

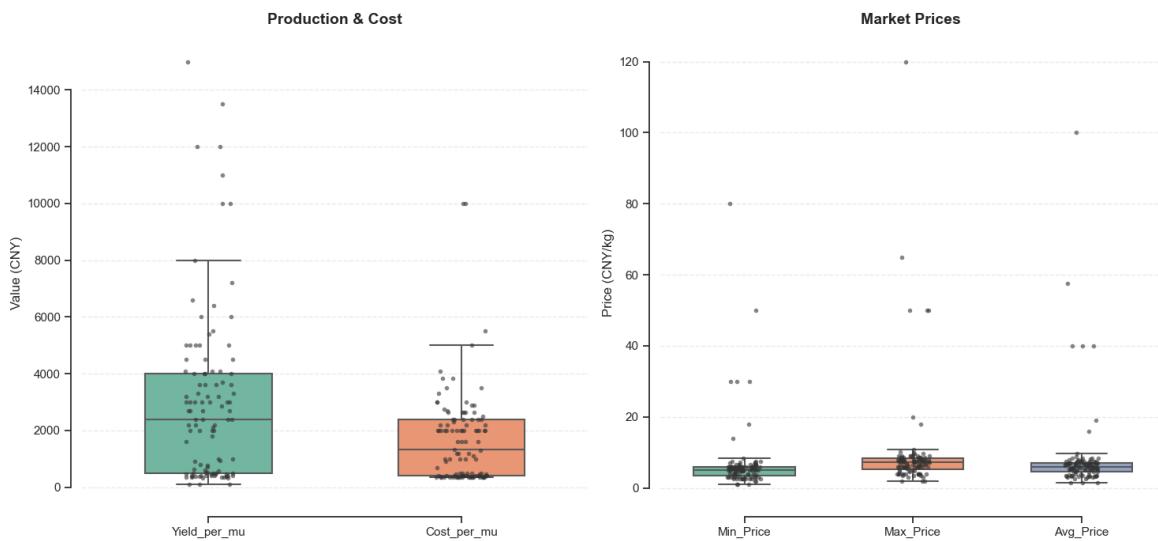


Fig 3 Boxplot distribution of yield, planting cost, and market price

Combined with the boxplot analysis in Figure 3, the data exhibits the following characteristics:

- High Dispersion: The high dispersion of yield and cost confirms the lack of homogeneity among crops. The huge standard deviation implies that the sample covers a variety of production modes, ranging from extensive food crops to intensive cash crops.
- Long-tail Distribution: The price distribution shows significant asymmetry. Although the main body of prices is concentrated at the low end, the existence of extreme values breaks the normal distribution assumption. This indicates that the agricultural region features a product mix dominated by "bulk low-price crops, supplemented by a few high-net-value crops."

## 2.5 Outlier Detection and Validity Assessment

During outlier detection, we noticed anomalies in the data for "Water Spinach" under "Smart Greenhouses," with yields reaching 10,000-12,000 jin/mu and planting costs reaching 4,100-5,500 CNY/mu. Upon consulting agricultural literature, water spinach is identified as a multi-harvest vegetable with a short growth cycle and high biomass in greenhouse environments; thus, the characteristic of high input corresponding to high output is reasonable. Therefore, we deem this data point a genuine and valid "strong leverage point." In subsequent modeling, such crops will serve as key breakthroughs for increasing total profit, though the risk of excessive resource occupation on a single plot must be noted.

## 3 Problem 1: Planting Planning Model under Stable Market Conditions

Given that expected sales volume, planting costs, yield per mu, and selling prices remain stable relative to 2023 for the next seven years, it is necessary to solve for expected sales volume to complete the comparison between Scenario 1 and Scenario 2.

Analyzing the planting requirements, the cultivation of crops on different arable land types must comply with the figure below:

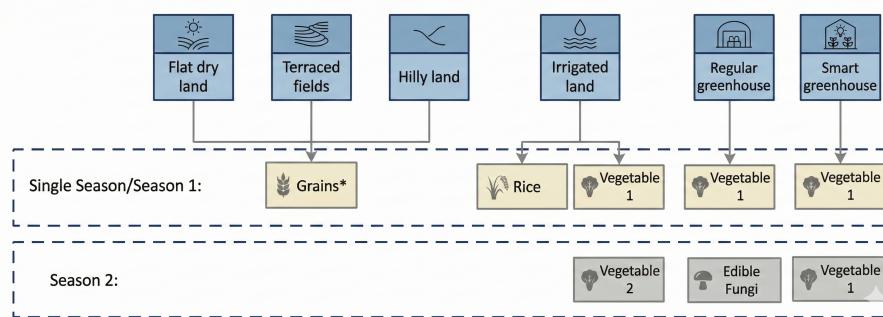


Fig 4 Cultivated Land Planting Type Diagram

In the figure, "Grain" excludes rice; "Vegetable 1" refers to vegetables other than Chinese cabbage, white radish, and carrots; "Vegetable 2" refers to Chinese cabbage, white radish, and carrots.

### 3.1 Economic Feature Extraction of Crops Based on Hierarchical Clustering

To explore the performance differences and value hierarchies of different crops across the economic dimension, this paper applies hierarchical clustering to a sample set of 107 crops, implemented via the `scipy.cluster.hierarchy` module. The workflow described below encompasses feature construction and selection, variable preprocessing, selection of distance metrics and linkage criteria, and objective criteria for cluster number determination.

#### 3.1.1 Construction of Economic Indicators and Feature Selection

Based on crop yield per mu, selling price, and planting cost, four types of comprehensive economic indicators were constructed as clustering features to measure absolute return, unit output efficiency, relative profitability, and risk-adjusted return capability, respectively. The variables are defined as follows:

$$P_i = Y_i \cdot \pi_i - C_i, \quad (\text{Profit per mu, Absolute Return}) \quad (2)$$

$$p_i^{\text{profit}} = \pi_i - \frac{C_i}{Y_i}, \quad (\text{Profit per jin, Unit Output Efficiency}) \quad (3)$$

$$M_i = \frac{P_i}{Y_i \cdot \pi_i} \times 100\%, \quad (\text{Profit Margin, Relative Profitability}) \quad (4)$$

$$S_i = \frac{P_i}{\pi_{i,\max} - \pi_{i,\min}}, \quad (\text{Sharpe-like Ratio, Risk-adjusted Return}) \quad (5)$$

Where  $Y_i$ ,  $\pi_i$ , and  $C_i$  represent the yield per mu (jin/mu), average selling price (CNY/jin), and unit cost (CNY/mu) of crop  $i$ , respectively;  $\pi_{i,\max}$  and  $\pi_{i,\min}$  represent the historical maximum and minimum prices of the crop.

During the feature selection phase, correlation analysis among variables revealed a strong linear correlation between Profit Margin and other features, particularly the Sharpe-like ratio. Consequently, this highly redundant variable was excluded to prevent repetitive information from disproportionately influencing the distance-based clustering results. The final feature set adopted is  $\{P_i, p_i^{\text{profit}}, S_i\}$ , which describes crop economic attributes from three complementary dimensions: absolute return, efficiency, and risk.

### 3.1.2 Robust Preprocessing

Given the significant long-tail distribution of economic indicators, with sample spans reaching tens of thousands, we employ a RobustScaler based on the median and Interquartile Range (IQR) to mitigate the impact of extreme values on distance-based algorithms:

$$z_{ik} = \frac{X_{ik} - Q_{50}(X_k)}{Q_{75}(X_k) - Q_{25}(X_k)}, \quad (6)$$

where  $X_{ik}$  represents the original value of variable  $k$  in sample  $i$ , and  $Q_{25}$ ,  $Q_{50}$ ,  $Q_{75}$  represent the 25th, 50th, and 75th percentiles, respectively. This processing effectively suppresses the influence of outliers on distance calculations, thereby preserving relative differences among medium and low-value samples, which facilitates the identification of economically meaningful groupings.

### 3.1.3 Distance Metric and Hierarchical Clustering Algorithm

In the standardized feature space, the similarity between samples is measured using Euclidean distance:

$$d_{ij} = \sqrt{\sum_{k=1}^3 (z_{ik} - z_{jk})^2}. \quad (7)$$

Cluster merging adopts Ward's linkage method, which aims to minimize the increment of the within-cluster sum of squares (WSS) at each step. Given clusters  $U$  and  $V$ , the distance after merging is characterized by:

$$D(U, V) = \frac{|U||V|}{|U|+|V|} \|\mu_U - \mu_V\|^2, \quad (8)$$

where  $|U|$  and  $|V|$  are the number of samples in the clusters, and  $\mu_U$  and  $\mu_V$  are the centroids of the respective clusters. Ward's method tends to generate compact, relatively balanced clusters, making it suitable for the hierarchical analysis and interpretation of economic numerical features.

### 3.1.4 Selection of Cluster Number

To objectively determine the optimal number of clusters, the Silhouette Score is used as a measure of clustering validity. The silhouette coefficient for sample  $i$  is defined as:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, \quad (9)$$

where  $a(i)$  is the average distance from sample  $i$  to other samples within the same cluster (intra-cluster compactness), and  $b(i)$  is the average distance from sample  $i$  to all samples in the nearest neighboring cluster (inter-cluster separation).  $s(i) \in [-1, 1]$ , with a higher average silhouette coefficient indicating better clustering quality. After evaluating different  $k$  values, the average silhouette coefficient reaches an optimal value ( $\bar{s} \approx 0.751$ ) at  $k = 4$ , and the clustering results demonstrate good interpretability in both economics and agricultural practice. The visualization of clustering results is shown in Figure 5.

Based on the above analysis, the samples are ultimately divided into four classes, described as follows:

- First-tier Value: Represented by rare fungi such as Enoki mushrooms and Morels, featuring extremely high profit per mu ( $\gtrsim 9 \times 10^4$  CNY/mu) and "High Risk-High Return" attributes. Based on business decision needs, a higher target return on investment can be set for this category.
- Second-tier Value: Mainly consists of smart greenhouse or facility cultivation crops (e.g., White Ling mushrooms, facility cucumbers), with profit per mu in the medium-high range (approx.  $7 - 15 \times 10^4$  CNY/mu). These crops rely heavily on facilities and management inputs, suggesting differentiated cost and revenue management.
- Third-tier Value: Includes various bulk vegetables (e.g., Chinese cabbage, Cowpeas, Water spinach), with profit per mu in the medium-low range (approx.  $0.8 - 5.4 \times 10^4$  CNY/mu). Characterized by stable markets and mature technology, they serve as the primary income source for rural households or cooperatives.

- Fourth-tier Value: Represented by low-cost grain crops like Buckwheat, with lower profit per mu (approx. several thousand CNY/mu). However, they possess attributes of easy mechanization and good storage/transportation, contributing to food security and production stability.

The above clustering results reveal the economic value stratification of crops, providing an analytical framework for the subsequent interpretation of optimization results. By statistically analyzing the area allocation and profit contribution of crops at each level, the formation mechanism of optimization strategies and their economic implications can be better understood.

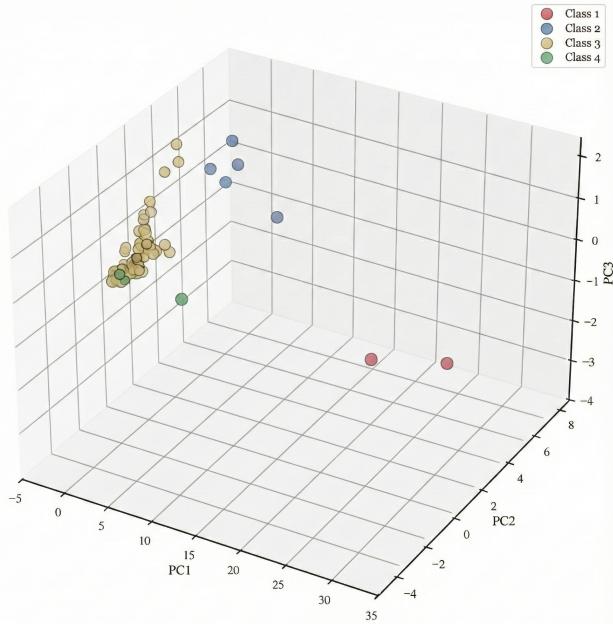


Fig 5 Visualization of clustering results for different crop samples ( $k=4$ )

### 3.2 Construction of the Optimization Model for the Optimal Planting Scheme

Based on the previous crop clustering analysis and the 2023 benchmark data, this paper constructs a Mixed Integer Linear Programming (MILP) model<sup>[2-3]</sup>. The model aims to maximize the overall profit of rural agricultural production from 2024 to 2030, while satisfying multiple constraints such as land resources, ecological rotation, and market demand.

#### 3.2.1 Basic Parameter Settings

The sources and definitions of the key parameters required for the optimization model are as follows:

(1) Crop Economic Parameters: Provided by Appendix 1, including the yield per mu  $y_{j,c}$  (jin/mu), planting cost  $C_{j,c}$  (CNY/mu), and average selling price  $P_c$  (CNY/jin) for each crop on different plot types.

(2) Plot Characteristics: Appendix 1 provides the area  $A_j$  (mu) and cultivation type for each plot.

(3) Market Capacity Upper Limit: Assuming market demand remains relatively stable, the actual yield of 2023 is taken as the annual sales cap  $Q_c$  for each crop. The specific calculation is:

$$Q_c = \sum_{j,s} \text{Area}_{2023}^{j,s,c} \times y_{j,c} \quad (10)$$

where  $\text{Area}_{2023}^{j,s,c}$  is the planting area of crop  $c$  in season  $s$  of plot  $j$  in 2023 (from Appendix 2). This method is based on two assumptions: first, the 2023 yield was fully absorbed by the market, proving that demand reached at least this level; second, market demand will remain stable in the short term (2024-2030). Statistics show that the total yield of the village in 2023 was 2,833,300 jin, covering 41 crops.

(4) Legume Crop Set: According to the crop classification in Appendix 1, the legume crop set  $\mathcal{B}$  is determined for rotation constraints.

#### 3.2.2 Definition of Decision Variables

To accurately describe the planting behavior for each plot and each season, we define the continuous variable  $x_{t,j,s,c}$  to represent the planting area (unit: mu) of crop  $c$  in season  $s$  on plot  $j$  in year  $t$  ( $t \in \{2024, \dots, 2030\}$ ).

Furthermore, to handle logical constraints such as crop rotation, we introduce the 0-1 binary variable  $H_{t,j,s,c}$  as a planting status indicator:

$$H_{t,j,s,c} = \begin{cases} 1, & \text{if crop } c \text{ is planted at spatiotemporal point } t, j, s \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

### 3.2.3 General Constraints

Regardless of the market sales scenario, the model must adhere to the following four fundamental agronomic and resource constraints:

#### I. Land Resource Capacity Constraint

For any year  $t$ , any plot  $j$ , and any planting season  $s$ , the sum of the planting areas allocated to different crops  $c$  must not exceed the actual available area  $\text{Area}_j$  of that plot:

$$\sum_{c \in C} x_{t,j,s,c} \leq \text{Area}_j, \quad \forall t, j, s \quad (12)$$

#### II. Variable Logical Association Constraint

To ensure logical consistency between the continuous variable  $x$  and the binary variable  $H$  (i.e., "if there is area, the status must be planting; if the status is non-planting, the area must be 0"), we introduce the classic Big-M method constraint:

$$\varepsilon \cdot H_{t,j,s,c} \leq x_{t,j,s,c} \leq M \cdot H_{t,j,s,c} \quad (13)$$

where  $\varepsilon$  is the minimum planting area threshold (set to 0.01 mu in this paper) to avoid generating excessively fragmented micro-plots, and  $M$  is a sufficiently large positive number (e.g., the upper limit of the total plot area).

#### III. Ecological Continuous Cropping Obstacle Constraint

To prevent soil fertility depletion and the accumulation of pests and diseases caused by continuous planting of the same crop, the model strictly prohibits the planting of the same crop for two consecutive seasons. That is, for any crop  $c$ , the sum of its planting states in two adjacent time steps must not exceed 1:

$$H_{t,j,s,c} + H_{t,j,s+1,c} \leq 1 \quad (14)$$

Specifically, this constraint also applies to cross-year situations, i.e., between the 2nd season of year  $t$  and the 1st season of year  $t+1$ .

#### IV. Legume Crop Rotation Constraint

According to soil improvement requirements, each plot must plant at least one season of legume crops within any consecutive 3 years (i.e., 6 planting seasons). Let  $C_{bean}$  be the set of legume crops (such as soybeans, adzuki beans, etc.); the constraint is expressed as:

$$\sum_{\tau=t}^{t+2} \sum_{s \in S} \sum_{c \in C_{bean}} H_{\tau,j,s,c} \geq 1, \quad \forall j, \forall t \in \{2024, \dots, 2028\} \quad (15)$$

### 3.3 Scenario 1: Model of Waste Caused by Unsalable Surplus

#### 3.3.1 Problem Analysis and Logic Transformation

In this scenario, the problem stipulates that "portions exceeding the expected sales volume cannot be sold normally," implying they become waste. From an economic perspective, this means that for production exceeding the threshold  $Q_c$ , the marginal revenue is 0, while the marginal cost remains positive. Therefore, a rational profit-maximizing model would never actively produce more than  $Q_c$ .

Based on this logic, we can transform "unsalable" into a hard production cap constraint, thereby avoiding the use of complex non-linear objective functions.

#### 3.3.2 Model Construction

Objective Function: Maximize total net profit  $Z_1$ .

$$\max Z_1 = \sum_{t \in T} \sum_{c \in C} \left[ P_c \cdot \left( \sum_{j,s} Y_{j,c} \cdot x_{t,j,s,c} \right) - \sum_{j,s} Cost_{j,c} \cdot x_{t,j,s,c} \right] \quad (16)$$

where  $P_c$  is the average selling price of crop  $c$ ,  $Y_{j,c}$  is the yield per mu on plot  $j$ , and  $Cost_{j,c}$  is the corresponding cost per mu.

Market Capacity Constraint: Let  $TotalYield_{t,c}$  be the total yield of crop  $c$  in the village in year  $t$ . This yield must not exceed the expected sales cap  $Q_c$  calculated previously:

$$\text{TotalYield}_{t,c} = \sum_{j \in J} \sum_{s \in S} Y_{j,c} \cdot x_{t,j,s,c} \leq Q_c, \quad \forall t, c \quad (17)$$

### 3.4 Scenario 2: Model of Discounted Sales

#### 3.4.1 Introduction of Piecewise Revenue Function

In this scenario, total production is allowed to exceed the expected sales volume  $Q_c$ , but the surplus is sold at 50% of the original price. This means the total sales revenue  $R_c$  for crop  $c$  is no longer a linear function of yield but exhibits piecewise linear characteristics.

Let  $y$  be the total yield of a certain crop in a given year. Its revenue function  $R_c(y)$  is defined as:

$$R_c(y) = \begin{cases} P_c \cdot y, & 0 \leq y \leq Q_c \quad (\text{Normal Sales Interval}) \\ P_c \cdot Q_c + 0.5P_c \cdot (y - Q_c), & y > Q_c \quad (\text{Discounted Sales Interval}) \end{cases} \quad (18)$$

#### 3.4.2 Derivation of Objective Function Linearization

Since the piecewise function contains logical decision points (break points), it is unfavorable for direct use in linear programming solvers. To solve this non-convexity problem, this paper employs binary variables and the Big-M method for linearization.

We introduce three types of auxiliary variables:

- $y_{\text{norm}}^{(t,c)}$ : The portion of yield sold at the normal price;
- $y_{\text{surplus}}^{(t,c)}$ : The portion of surplus yield sold at a discount;
- $\lambda^{(t,c)} \in \{0, 1\}$ : Overproduction status indicator (1 indicates overproduction, 0 indicates no overproduction).

Constraints:

$$\text{TotalYield}_{t,c} = y_{\text{norm}}^{(t,c)} + y_{\text{surplus}}^{(t,c)} \quad (19)$$

$$0 \leq y_{\text{norm}}^{(t,c)} \leq Q_c \quad (20)$$

$$\text{TotalYield}_{t,c} - Q_c \leq M \cdot \lambda^{(t,c)} \quad (21)$$

$$y_{\text{surplus}}^{(t,c)} \leq M \cdot \lambda^{(t,c)} \quad (22)$$

where  $M$  is a sufficiently large constant (Big-M). This set of constraints ensures:

- When total yield does not exceed  $Q_c$ ,  $\lambda^{(t,c)} = 0$ , thereby forcing  $y_{\text{surplus}}^{(t,c)} = 0$ , and all yield is sold at the normal price;
- When total yield exceeds  $Q_c$ ,  $\lambda^{(t,c)} = 1$ , allowing  $y_{\text{surplus}}^{(t,c)} > 0$ , and the excess portion is sold at the discounted price.

Objective Function:

$$\max Z_2 = \sum_{t,c} \left[ P_c \cdot y_{\text{norm}}^{(t,c)} + 0.5P_c \cdot y_{\text{surplus}}^{(t,c)} - \sum_{j,s} \text{Cost}_{j,c} \cdot x_{t,j,s,c} \right] \quad (23)$$

This method transforms the original piecewise non-linear function into a standard Mixed Integer Linear Programming (MILP) form via explicit binary variables and Big-M constraints, ensuring the precision of the model and the efficiency of the solver.

### 3.5 Two-stage Solution Strategy

Considering the practical needs of agricultural production management, purely pursuing profit maximization may lead to the excessive fragmentation of the same plot, thereby increasing the costs of operation organization and management. To balance economic efficiency and operational feasibility, this paper adopts a two-stage optimization strategy for both Scenario 1 and Scenario 2.

#### 3.5.1 Stage 1: Obtaining the Theoretical Optimal Solution

Subject to the aforementioned five types of constraints, the Mixed Integer Programming (MIP) model is solved with  $Z_1$  or  $Z_2$  as the objective function, respectively, to obtain the theoretical optimal total profit, denoted as  $Z_{\text{opt}}^*$ . This stage aims to provide an upper bound on profit, serving as a benchmark for the subsequent regularization of the scheme.

### 3.5.2 Stage 2: Optimization for Scheme Regularization

In this stage, without significantly compromising economic benefits, the planting scheme is regularized to reduce the dispersion of plot types and improve management feasibility.

Objective Function  $Z_{stage2}$ : Minimize the total sum of planting status variables  $H$ :

$$\min Z_{stage2} = \sum_{t,j,s,c} H_{t,j,s,c} \quad (24)$$

Additional Constraint: The optimal profit  $Z'$  in the second stage is required to be no less than 95% of the optimal profit from the first stage:

$$Z' \geq 0.95 \times Z_{opt}^* \quad (25)$$

Through this two-stage solution strategy, the model can generate a more regularized and highly executable planting layout scheme while maintaining total profit close to the optimal level, achieving an effective balance between theoretical optimization and practical management requirements.

## 3.6 Results and Analysis

Upon executing the optimization model, the Gurobi solver successfully identified the optimal solutions for both scenarios and outputted information regarding planting decisions, including plot IDs, crop IDs, years, seasons, and planting proportions.

The general overview of the solution under Scenario 1 is as follows: The output indicates an objective function value of 57,459,362, meaning the maximum planting profit optimized by the linear programming model is 57,459,362 CNY.

The general overview of the solution under Scenario 2 is as follows: The output indicates an objective function value of 98,462,192, meaning the maximum planting profit optimized by the linear programming model is 98,462,192 CNY.

The optimal planting schemes for Scenario 1 and Scenario 2 have been populated into the appendix tables result1\_1.xlsx and result1\_2.xlsx, respectively.

## 4 Problem 2: Comprehensive Planting Planning Based on Hybrid Monte Carlo Simulation

### 4.1 Problem Analysis and Modeling Strategy

Compared to the static deterministic optimization in Problem 1, Problem 2 introduces significant characteristics of dynamic evolution and stochastic perturbation. The core challenge of this problem lies in formulating a robust medium-to-long-term optimal planting scheme for 2024-2030 under the combined influence of multidimensional uncertain factors, such as expected crop sales volume, yield per mu, planting costs, and selling prices.

#### 4.1.1 Analysis of the Impact of Climate Factors on Crop Sales

Before constructing the prediction model, we need to clarify the differences in growth characteristics of various crops to match the most suitable models for each category.

Given that the crops grown in Hebei Province are highly consistent with those listed in the appendix, this paper is supported by the climate bulletins of the Hebei Provincial Meteorological Bureau from 2013 to 2023. We have compiled data on temperature, precipitation, and sunshine over the past decade.

The surveyed values are presented in the table below:

Table 2 Climate Data of Hebei Province (2013 - 2023)

Year	Avg. Temp. (°C)	Summer Temp. (°C)	Winter Temp. (°C)	Avg. Precip. (mm)	Summer Precip. (mm)	Winter Precip. (mm)	Avg. Sunshine (h)	Summer Sunshine (h)	Winter Sunshine (h)
2013	11.9	25.4	-4.3	559.3	424.7	21.3	2288.9	586.2	398.2
2014	13.0	25.2	-1.4	393.3	217.2	8.4	2285.6	646.7	448.8
2015	12.9	25.8	-2.8	437.6	288.5	6.9	2373.3	637.2	442.5
2016	12.6	25.4	-2.2	608.9	449.7	14.5	2358.8	603.7	513.1
2017	13.0	25.5	-1.0	484.2	326.1	14.4	2473.3	637.2	442.5
2018	11.8	24.9	-2.9	529.6	393.1	13.5	2367.6	640.4	483.2
2019	12.9	25.8	-2.8	437.6	288.5	6.9	2373.3	637.2	442.5
2020	12.6	25.1	-2.1	512.6	342.3	9.4	2341.6	640.8	506.8
2021	12.9	25.1	-1.8	861.2	487.0	15.8	2260.9	563.9	581.0
2022	12.6	25.6	-1.8	569.3	428.2	10.1	2488.3	624.9	562.5
2023	13.2	26.8	-2.2	648.1	449.1	11.9	2549.4	759.3	577.6

Based on the provided information, grain cultivation accounts for as much as 90% of the village's planting area. Wheat and corn are two typical grain crops exhibiting a distinct upward trend in sales volume. Consequently, the following predictions for expected sales volume will utilize wheat and corn as representative examples.

Prior to constructing the regression model, this study first tests the linear relationship between climate variables and crop sales volume. Given that most climate characteristics (temperature, precipitation, sunshine duration) are continuous variables, and assuming their impact on crop yield is primarily linear, the Pearson correlation coefficient—the most commonly used metric for linear correlation—is selected.

The Pearson correlation coefficient assumes a bivariate normal distribution. The Shapiro-Wilk test ( $\alpha = 0.05$ ,  $n = 11$ ) indicates that 9 out of 11 climate variables satisfy normality. The two variables with slight deviations exhibit weak correlations with sales volume ( $|r| < 0.20$ ), while all significantly correlated variables passed the test. Therefore, the prerequisites for Pearson analysis are satisfied.

The Pearson coefficient quantifies the strength of the linear relationship between two continuous variables, with a range of  $[-1, 1]$ . An  $r \approx 1$  indicates a strong positive correlation,  $r \approx -1$  indicates a strong negative correlation, and  $r \approx 0$  implies a weak or non-existent linear correlation. This metric is not only suitable for Exploratory Data Analysis (EDA) but is also widely used for variable selection, dimensionality reduction, and pre-regression testing in studies concerning climate factors and crop yields.

The formula for the Pearson coefficient is as follows:

$$r_{XY} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}, \quad (26)$$

where  $X_i$  and  $Y_i$  are the  $i$ -th observations,  $\bar{X}$  and  $\bar{Y}$  are the means of the variables, and  $n$  is the sample size. This formula measures the ratio between the covariance of the two variables and the product of their standard deviations, thereby eliminating the influence of unit dimensionality and making comparisons of correlations across different climate variables comparable.

In practice, this paper calculates the Pearson coefficients between all climate features from 2013 to 2023 and the annual sales volumes of wheat and corn, forming a correlation matrix. Subsequently, a heatmap is plotted based on this matrix to visually demonstrate the linear correlation strength between various climate factors and sales volume, providing a basis for subsequent feature selection and regression modeling:

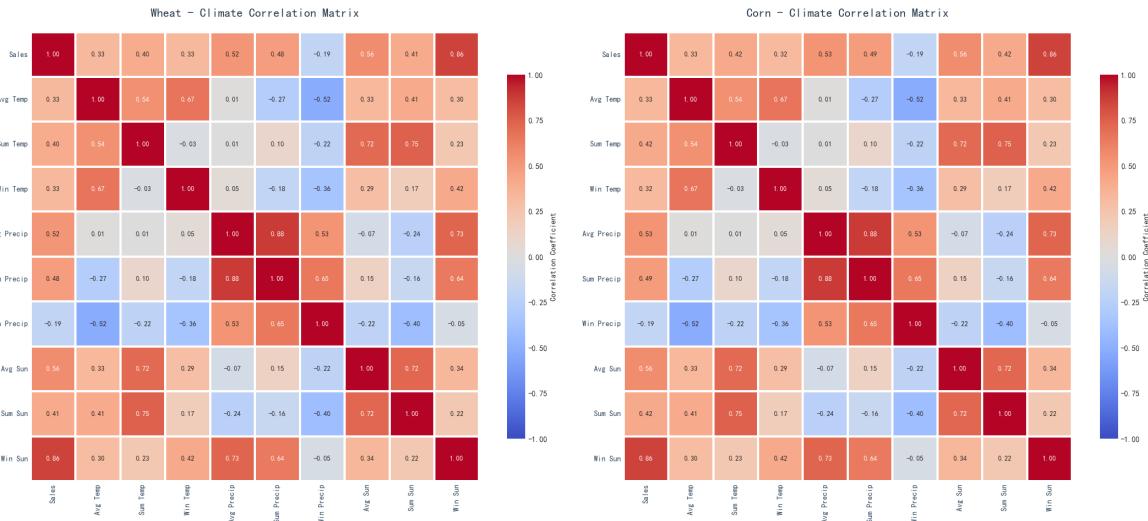


Fig 6 Heatmap of correlation coefficients for wheat and corn

In the preliminary correlation analysis, the heatmap reveals that the expected sales volumes of both wheat and corn exhibit significant linear correlations with multiple climate indicators. High correlations are primarily concentrated in the three dimensions of temperature, precipitation, and sunshine. For ease of comparison, the Pearson correlation coefficients between the two crops and each climate variable are ranked by absolute value, as shown in Table 3.

It can be observed that the sensitivity patterns of wheat and corn to climate factors are highly consistent: when sorted by the absolute value of correlation coefficients, Winter Sunshine Duration, Annual Average Sunshine Duration, Annual Average Precipitation, Summer Precipitation, and Summer Sunshine Duration consistently rank in the top five for both crops. Specifically, the corresponding  $|r|$  values for wheat are 0.8561, 0.5556, 0.5224, 0.4821, and 0.4118, while for

Table 3 Correlation Coefficients and Ranking of Climate Variables for Wheat and Corn

Variable	ID	$ r_{\text{Wheat}} $	Rank (Wheat)	$ r_{\text{Corn}} $	Rank (Corn)
Avg. Annual Temp.	$X_1$	0.3266	8	0.3284	7
Avg. Summer Temp.	$X_2$	0.4022	6	0.4153	6
Avg. Winter Temp.	$X_3$	0.3266	7	0.3197	8
Avg. Annual Precip.	$X_4$	0.5224	3	0.5254	3
Summer Precip.	$X_5$	0.4821	4	0.4871	4
Winter Precip.	$X_6$	-0.1928	9	-0.1858	9
Avg. Annual Sunshine	$X_7$	0.5556	2	0.5594	2
Summer Sunshine	$X_8$	0.4118	5	0.4198	5
Winter Sunshine	$X_9$	0.8561	1	0.8578	1

corn they are 0.8578, 0.5594, 0.5254, 0.4871, and 0.4198. Based on these results, the subsequent multiple linear regression model selects these five most highly correlated climate variables as independent variables for each crop to characterize the linear impact of climate conditions on expected sales volume.

#### 4.1.2 Validation of Climate Sensitivity for Wheat and Corn via Multiple Linear Regression

To quantitatively assess the specific contributions of various climate factors to grain crop sales volumes and to verify the effectiveness of the prediction model, we established Multiple Linear Regression models for both wheat and corn.

**1. Model Construction** We designate the annual expected sales volume of wheat and corn as the dependent variable  $Y$ . The independent variables  $X_i$  comprise the five key climate factors selected via correlation analysis: Winter Sunshine Duration ( $X_9$ ), Annual Average Sunshine Duration ( $X_7$ ), Annual Average Precipitation ( $X_4$ ), Summer Precipitation ( $X_5$ ), and Summer Sunshine Duration ( $X_8$ ).

The mathematical expression of the regression model is:

$$Y = \beta_0 + \beta_1 X_9 + \beta_2 X_7 + \beta_3 X_4 + \beta_4 X_5 + \beta_5 X_8 + \varepsilon, \quad (27)$$

The meanings of the symbols are as follows:

- $Y$  : Annual expected sales volume of the crop (tons)
- $X_9, X_8, X_7$  : Winter, Summer, and Annual Avg. Sunshine Duration, respectively (h)
- $X_5, X_4$  : Summer and Annual Avg. Precipitation, respectively (mm)
- $\beta_0$  : Intercept term
- $\beta_i$  : Regression coefficients for each climate variable ( $i = 1, \dots, 5$ )
- $\varepsilon$  : Random error term,  $\varepsilon \sim N(0, \sigma^2)$

**2. Solution and Result Analysis** To evaluate the magnitude of the impact of climate factors on grain sales, this study employs the Ordinary Least Squares (OLS) method to fit historical data from 2013 to 2023. Prior to regression, all independent variables were standardized to eliminate dimensional discrepancies that could bias weights, while also enhancing the comparability of coefficients. Since the model contains only five key meteorological features selected via correlation, multicollinearity issues among variables were controlled before construction, ensuring the stability of model estimation.

The regression results indicate that both the wheat and corn models achieved a high degree of fit. Specifically, the coefficient of determination for the wheat model is:

$$R^2_{\text{wheat}} = 0.8225, \quad \text{Adj. } R^2_{\text{wheat}} = 0.6450,$$

This indicates that the model explains over 82.25% of the sales volatility, retaining substantial explanatory power even after adjustment. The corn model exhibits similar performance:

$$R^2_{\text{corn}} = 0.8274, \quad \text{Adj. } R^2_{\text{corn}} = 0.6549,$$

This demonstrates that corn sales are also significantly driven by climate variables, with an explanatory power slightly higher than that of the wheat model.

**3. Conclusions and Implications** Analyzing the coefficient values, variables in the wheat model such as Winter Sunshine ( $X_9$ ), Annual Avg. Sunshine ( $X_7$ ), and Annual Avg. Precipitation ( $X_4$ ) exhibit significant positive effects, suggesting that sufficient sunlight and mild, moist climatic conditions contribute to increased sales volume. The coefficient for Summer Precipitation ( $X_5$ ) is negative, implying that excessive summer rainfall may have an inhibitory effect on yield. The coefficient structure of the corn model is highly consistent with that of wheat, with the direction of impact for each climate factor being identical, further corroborating their shared dependence on light resources and precipitation conditions.

To assess model applicability, we further conducted a residual diagnosis. The residual plot (see Figure 7) reveals no systematic trends in the residual distribution and no patterned structure relative to fitted values, validating the homoscedasticity assumption of linear regression. Simultaneously, the residuals exhibit an approximately symmetric distribution without significant skewness or heavy tails, suggesting the error term approximately satisfies the normality assumption.

In summary, the multiple linear regression models for wheat and corn demonstrate strong fitting capabilities and favorable statistical properties, effectively capturing the primary laws by which climate change impacts sales volume. Given the consistency of key meteorological drivers and their directional effects, the conclusions of this study possess high robustness across crops, providing an important basis for understanding the potential impact of climate change on food supply.



Fig 7 Fitted regression analysis of climate factors on expected wheat sales

Based on these findings, we reasonably infer that: climate factors exert a significant impact on other open-field crops in the region as well, causing random volatility in their yields and sales; conversely, crops in facility agriculture (e.g., greenhouses) are less affected by climate and are more likely to follow growth trends driven by market demand<sup>[4-5]</sup>.

#### 4.1.3 Crop Classification Strategy

Integrating the aforementioned climate sensitivity analysis with the market characteristics of crops provided in the problem, this paper classifies the 41 crops into three categories and constructs targeted prediction models for each. The specific strategies are as follows:

- Cold-adaptive (Type G1) Encompasses varieties such as wheat, corn, soybeans, and tubers. These crops are predominantly cultivated in open fields, are highly sensitive to climate conditions, and exhibit significant inter-annual yield volatility. Given their status as necessities with rigid market demand growth trends (particularly wheat and corn), this study adopts a composite prediction model for their sales volume (trend-based for wheat/corn, fluctuation-based for others), while introducing a Random Walk Model for their yields to simulate the uncertainty caused by climate risks.
- Low-temperature Tolerant (Type G2) Includes varieties such as Chinese cabbage, tomatoes, and peppers. These

crops partially rely on greenhouse facilities and are relatively less impacted by direct extreme climate shocks. Driven by consumption upgrading, their prices exhibit an average annual growth trend of approximately 5%. Consequently, the model applies compound growth trend extrapolation for their prices and costs, while retaining stochastic fluctuation terms for yield and sales to reflect short-term dynamic adjustments in market supply and demand.

- Cool-temperature Fragile (Type G3) Mainly comprises mushrooms and morels. These crops rely entirely on greenhouse environments and are almost unaffected by natural climate fluctuations, representing typical "High Risk-High Return" varieties. Their price volatility is primarily driven by specific market supply-demand relationships (e.g., the cyclical decline in morel prices). Accordingly, the model focuses on specific trend predictions for prices and models their sales volume as a high-volatility stochastic process to capture market risk.

To facilitate an intuitive understanding of the above classification results, Figure 8 illustrates the specific categorization of the 41 crops and their distribution structure across different adaptability levels.

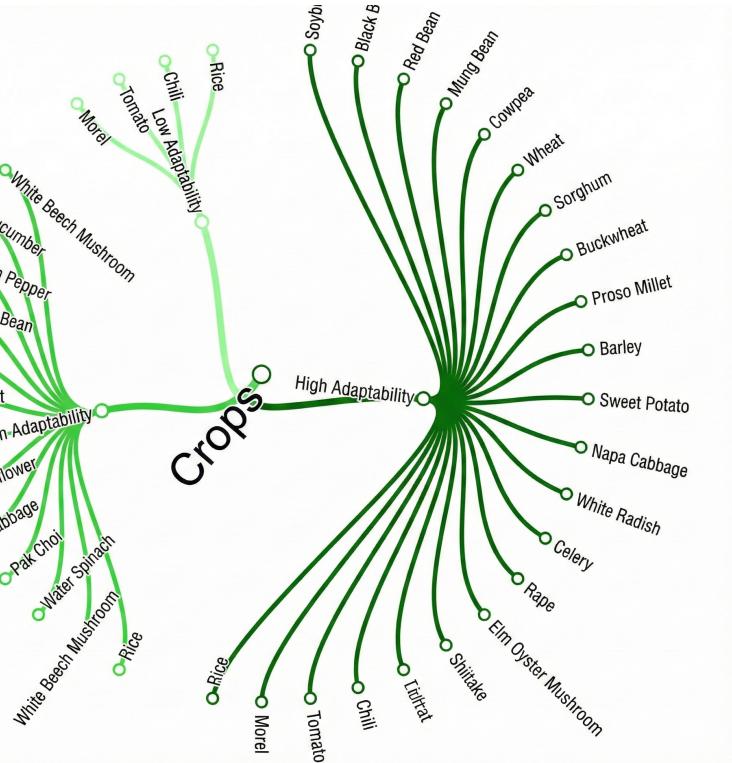


Fig 8 Schematic diagram of the classification structure for the 41 crop samples

## 4.2 Prediction of Uncertainty Indicators: Hybrid Monte Carlo Model

To accurately characterize the heterogeneous evolution laws of different agricultural indicators, this paper abandons the assumption of simple linear extrapolation and constructs a Hybrid Monte Carlo simulation model comprising two subsystems.

### 4.2.1 Sub-model A: Compound Trend Model

This sub-model specifically targets indicators described in the problem as having an "average annual growth rate" or being "basically stable." Specifically, these include sales price ( $p$ ), planting cost ( $c$ ), and the expected sales volume ( $X$ ) of wheat and corn. These indicators are primarily driven by long-term macroeconomic factors, where volatility is mainly reflected in the uncertainty of the long-term growth rate rather than drastic inter-annual oscillations.

In terms of the simulation mechanism design, for each simulation scenario, the model first randomly draws a long-term growth rate parameter  $r^{(k)}$  from a given growth rate interval and assumes this rate remains constant over the next seven years. The indicator value  $V_t^{(k)}$  for each year is derived recursively using the compound interest formula:

$$V_t^{(k)} = V_{2023} \times (1 + r^{(k)})^t, \quad r^{(k)} \sim U(r_{\min}, r_{\max}) \quad (28)$$

This method effectively simulates "trend uncertainty," meaning that while the overall upward trend is determined,

the specific slope of growth follows a probability distribution.

#### 4.2.2 Sub-model B: Annual Random Walk Model

This sub-model is suitable for describing indicators characterized by "annual changes," covering the yield per mu ( $q$ ) of all crops and the sales volume ( $X$ ) of crops other than wheat and corn. These indicators are significantly affected by stochastic perturbations such as climate change, pests, diseases, and short-term market sentiment, exhibiting clear random walk characteristics.

During the simulation process, at every step  $t$  of each year, the model independently redraws the volatility rate  $r_t^{(k)}$  for that year from a given change interval and applies it to the value of the previous year. Its dynamic evolution process can be expressed as:

$$V_t^{(k)} = V_{t-1}^{(k)} \times (1 + r_t^{(k)}), \quad r_t^{(k)} \sim U(a, b) \quad (29)$$

By introducing independent year-by-year stochastic shocks, this mechanism generates time series with sawtooth fluctuation characteristics, thereby more realistically reflecting the high-frequency risks faced by agricultural production.

Finally, by running  $N = 1000$  hybrid Monte Carlo simulations, we statistically aggregate the massive amount of generated scenario data and calculate the expected mean of each indicator for every year, which serves as the input parameter for the subsequent deterministic optimization model.

### 4.3 Optimization Model for the Optimal Planting Scheme

Based on the expected data obtained from the Monte Carlo simulation, we construct a multi-stage linear programming model.

#### 4.3.1 Model Assumptions and Notation

To accurately describe planting decisions, we define  $x_{i,j,t,s}$  as a 0-1 decision variable, indicating whether crop  $i$  is planted on plot  $j$  in season  $s$  of year  $t$ . In particular, to address the issue of potential overproduction caused by yield randomness, this paper introduces a key continuous auxiliary variable  $S_{i,t}$ , representing the actual sales volume of crop  $i$  in year  $t$ . This setting mathematically realizes the decoupling of "production behavior" and "sales behavior":

- Planting costs depend on planting area and inputs;
- Sales revenue depends on the output actually absorbed by the market (i.e., the smaller value between production yield and market demand).

This mechanism ensures that the model can automatically handle the balance between "production based on sales" and "inventory backlog," avoiding model infeasibility caused by rigid production constraints.

#### 4.3.2 Objective Function

The objective of the model is to maximize the total expected profit from 2024 to 2030. Total profit is defined as total sales revenue minus total planting costs. The mathematical expression is as follows:

$$\max Z = \sum_{t=2024}^{2030} \left( \sum_{i \in I} P_{it} \cdot S_{it} - \sum_{i \in I} \sum_{j \in J} \sum_{s \in \Omega} C_{it} \cdot A_j \cdot x_{i,j,t,s} \right) \quad (30)$$

where  $P_{it}$  and  $C_{it}$  are the predicted price and predicted cost obtained from the Monte Carlo simulation, respectively, and  $A_j$  is the area of the plot. This objective function reflects the decision logic of rational farmers in a risk environment: pursuing the maximization of effective output rather than simply maximizing production yield.

#### 4.3.3 Constraints

The constraints of the model in this problem are identical to those in Problem 1, consisting of the four general categories: land capacity, logical variable association, ecological continuous cropping obstacles, and legume crop rotation.

### 4.4 Analysis of Solution Results

Using the Gurobi large-scale optimization solver to solve the above model, we obtained the global optimal planting strategy for 2024-2030. Under the scenario that fully considers future market fluctuations and climate risks, the expected total profit for the village in the next seven years is 103,034,611.63 CNY.

#### 4.4.1 Implementation of the Crop Rotation System

The model results indicate that all plots across Dry Flatlands, Terraced Fields, and Hillslopes strictly enforced the "once every three years" legume planting constraint. Notably, crops such as soybeans and black beans were not concentrated in a single year for large-scale planting; instead, they were intelligently distributed across different years and plots for rotation. This "distributed" spatiotemporal synergistic strategy avoids the sharp decline in annual profit caused by concentrated legume planting while evenly satisfying soil conservation needs over the long term. This reflects a deep integration of ecological benefits and economic efficiency.

#### 4.5 Model Robustness Verification

To verify the reliability of the formulated strategy and quantify the economic cost of ecological protection, we designed a constraint relaxation experiment. The results show that even if the strict legume rotation constraint is relaxed—i.e., allowing land to go without legume planting for consecutive years to pursue short-term interests—the total profit only increases to 106.2 million CNY, representing a negligible increase.

The results of this sensitivity analysis carry significant policy implications. It demonstrates that adhering to an eco-friendly rotation system does not impose a significant burden on economic benefits. Sacrificing less than 1% of profit in exchange for the long-term maintenance of soil fertility and a significant reduction in pest and disease risks proves to be a highly cost-effective sustainable development strategy from a long-term perspective. Our model proves that green agriculture and efficient agriculture are not a zero-sum game; rather, a win-win situation is entirely achievable.

### 5 Problem 3: Strategy Optimization Based on Crop Substitutability and Complementarity

In the solution to Problem 2, the prediction of sales volume for each crop was primarily based on independent time-series trends, ignoring the interdependencies prevalent in the agricultural market. In reality, there are significant substitution and complementarity effects among crops. For example, when the price of a staple grain rises, consumers tend to purchase functionally similar but lower-priced substitutes; conversely, certain vegetables and edible fungi exist in a complementary relationship due to dietary habits requiring paired purchases. The problem requires us to further consider these substitutability and complementarity factors on the basis of Problem 2 to re-optimize the planting strategy for 2024-2030.

The core of the modeling in this section lies in: how to endogenize the correlation of markets and their price fluctuations into the prediction data, thereby capturing potential profit margins or risk hedging opportunities brought about by market linkages in the optimization model. Therefore, this section constructs a structural demand elasticity model based on crop features, describes the substitution and complementary relationships between crops via a Cross-Price Elasticity Matrix<sup>[7-8]</sup>, and on this basis, employs Monte Carlo simulation to generate a set of price-sales scenarios reflecting real market linkages, thereby solving for a more robust planting strategy.

#### 5.1 Construction of Feature-based Structural Demand Model

##### 5.1.1 Substitution Effect: Elasticity Construction Based on Price Similarity

To quantify the substitution strength between similar crops, price structure is adopted as the primary characterization of substitutability. For any crops  $i$  and  $j$  belonging to the same category, their substitution elasticity is defined as:

$$e_{ij}^{sub} = \alpha \cdot \exp \left( -\lambda \cdot \frac{|P_i^{base} - P_j^{base}|}{(P_i^{base} + P_j^{base})/2} \right) \quad (31)$$

where  $P_i^{base}$  is the benchmark year (2023) price of crop  $i$ ,  $\alpha > 0$  represents the maximum substitution intensity, and  $\lambda$  regulates the decay rate of substitutability with respect to price difference. The closer the prices, the larger  $e_{ij}^{sub}$  becomes (positive value), reflecting a stronger competitive substitution relationship.

##### 5.1.2 Complementarity Effect: Cross-category Synergistic Structure

For crop pairs  $(i, j)$  belonging to different categories, we set their demand to be complementary, written as:

$$e_{ij}^{comp} = -\beta \cdot \xi, \quad (32)$$

where  $\beta > 0$  is the baseline strength of the complementary relationship, and  $\xi \sim U(0.8, 1.2)$  is a random perturbation term for preference differences. Complementary elasticity is negative, indicating that when the price of one crop rises, the sales volume of its complement decreases.

##### 5.1.3 Structural Characteristics of the Cross-Price Elasticity Matrix

The visualization of the derived cross-price elasticity matrix presents the following structural characteristics:

- Strong Substitution Zone (Red): Concentrated near the diagonal of similar crops. For instance, the elasticity among various vegetables is generally high, reflecting significant market competition. Notably, the substitution elasticity between Wheat and Barley reaches 0.56, indicating high interchangeability on the consumer side.
- Complementary Zone (Blue): Mainly distributed in the cross-regions of Grain-Vegetable and Vegetable-Fungi. Dark blue values represent complementarity. For example, the elasticity between Wheat and Green Pepper is -0.35, reflecting a typical staple-side dish synergistic consumption structure.

##### 5.1.4 Analysis of Market Structure Characteristics

Based on the cross-price elasticity matrix calculated by the above model, this study presents the crop combinations with the strongest correlations. The results are shown in Table 4 and Table 5.

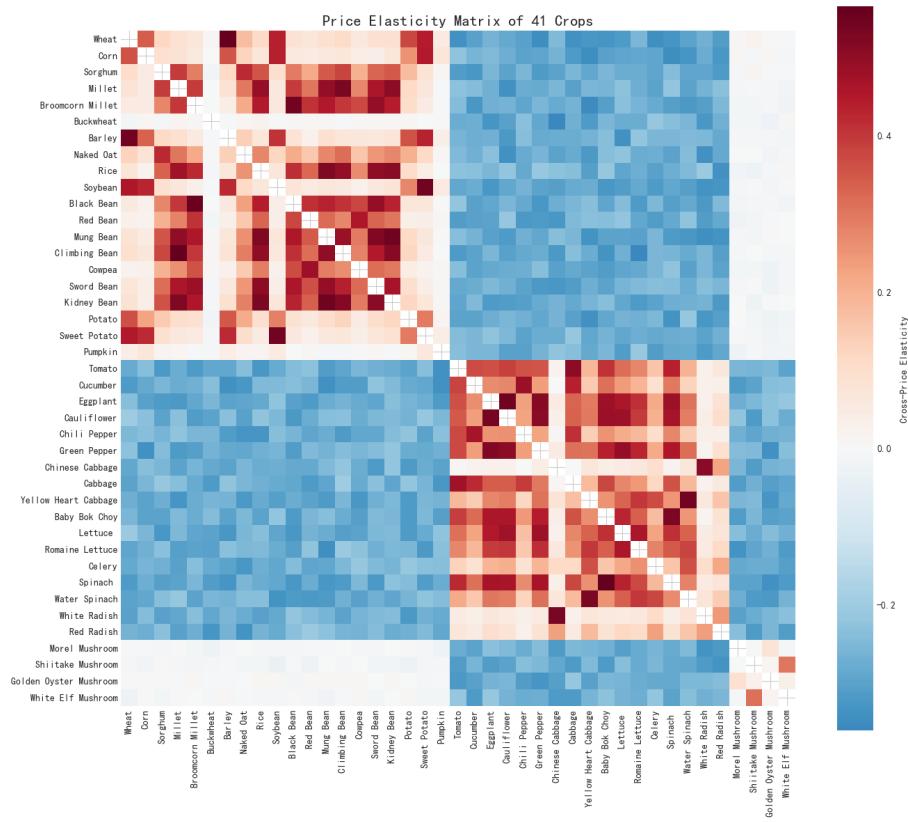


Fig 9 Heatmap of the Cross-Price Elasticity Matrix

### 1. Analysis of Strong Substitute Crop Combinations

As shown in Table 4, the strongest substitution combinations are mainly concentrated between miscellaneous grains and beans. For example, the cross-elasticity coefficients between Millet and Climbing Bean, and Broomcorn Millet and Black Bean both exceed 0.55. This indicates a strong competitive relationship within grain crops among miscellaneous grains and beans that have similar functions and prices. In addition, Spinach and Bok Choy, as well as Eggplant and Cauliflower in the vegetable category, also show significant substitutability, which aligns with the random selection behavior of consumers when purchasing leafy greens or brassica vegetables.

Table 4 Top 5 Strong Substitute Crop Pairs Derived Based on Price Similarity ( $e_{ij} > 0$ )

Rank	Crop A	Crop B	Avg. Cross Elasticity	Category Feature
1	Millet	Climbing Bean	0.5572	Grain-Grain
2	Broomcorn Millet	Black Bean	0.5544	Grain-Grain
3	Mung Bean	Red Bean	0.5461	Grain-Grain
4	Wheat	Barley	0.5446	Grain-Grain
5	Soybean	Sweet Potato	0.5428	Grain-Grain

### 2. Analysis of Strong Complementary Crop Combinations

Table 5 shows the combinations with the strongest complementarity, mainly manifesting as the "Vegetable + Miscellaneous Grain" pattern. For example, Romaine Lettuce and Climbing Bean, as well as Chinese Cabbage and Buckwheat, exhibit extremely strong complementarity. This negative correlation reflects the diversified demand in the resident dietary structure: an increase in the consumption of miscellaneous grains as a staple food often drives the consumption of leafy vegetables as side dishes. Furthermore, combinations such as Pumpkin with Cucumber, and Lettuce with Barley, also verify the natural consumption synergy existing across categories.

Table 5 Top 5 Strong Complementary Crop Pairs Derived Based on Dietary Structure ( $e_{ij} < 0$ )

Rank	Crop A	Crop B	Avg. Cross Elasticity	Category Feature
1	Romaine Lettuce	Climbing Bean	-0.3613	Veg-Grain
2	Chinese Cabbage	Buckwheat	-0.3594	Veg-Grain
3	Yellow Heart Cabbage	Broomcorn Millet	-0.3506	Veg-Grain
4	Yellow Heart Cabbage	Corn	-0.3497	Veg-Grain
5	Spinach	Corn	-0.3484	Veg-Grain

### 5.1.5 Structured Price-Sales Coupled Prediction Model

To avoid the irrationality caused by independent sales forecasting which ignores substitution and complementarity relationships among crops, a structured mechanism where price changes drive sales fluctuations is adopted.

#### Price Evolution

We use Geometric Brownian Motion<sup>[6]</sup> to simulate the price path of crop  $i$  in year  $t$ :

$$P_{it} = P_{i,t-1} \cdot \exp(\mu_p + \sigma_{macro} Z_t + \sigma_{idio} \varepsilon_{it}), \quad (33)$$

where:

- $\mu_p = 0.02$ : Annualized inflation rate
- $\sigma_{macro} = 0.05$ : Macroeconomic shock volatility (affecting all crops)
- $\sigma_{idio} = 0.03$ : Crop-specific idiosyncratic volatility
- $Z_t \sim N(0, 1)$ : Macroeconomic shock
- $\varepsilon_{it} \sim N(0, 1)$ : Independent disturbance

#### Sales Response

The rate of price change is defined as:

$$\Delta \ln P_{it} = \ln(P_{it}) - \ln(P_{i,t-1})$$

This is transmitted to sales changes via the cross-price elasticity matrix  $\mathbf{E} = (e_{ij})$ :

$$\Delta \ln \mathbf{Q}_t = \mathbf{E} \cdot \Delta \ln \mathbf{P}_t + \mu_q \mathbf{1}, \quad (34)$$

where  $\mu_q = 0.01$  is the long-term demand growth rate. Expanding to a single crop:

$$Q_{it} = Q_{i,t-1} \cdot \exp\left(\mu_q + \sum_{j=1}^{41} e_{ij} \cdot \ln\left(\frac{P_{jt}}{P_{j,t-1}}\right)\right), \quad (35)$$

#### Evolution of Cost and Yield

Costs are influenced by input prices:

$$C_{it} = C_{i,t-1} \cdot (1.01 + \xi_{it}), \quad \xi_{it} \sim U(-0.03, 0.03) \quad (36)$$

Yields are influenced by climate uncertainty:

$$Y_{it} = Y_{i,t-1} \cdot (1 + \eta_{it}), \quad \eta_{it} \sim U(-0.10, 0.10) \quad (37)$$

#### Monte Carlo Simulation and Expected Values

We run  $K = 1000$  independent scenario simulations, each generating a complete path for 2024-2030. The expected values ultimately used for optimization are:

$$\bar{P}_{it} = \frac{1}{K} \sum_{k=1}^K P_{it}^{(k)}, \quad \bar{Q}_{it} = \frac{1}{K} \sum_{k=1}^K Q_{it}^{(k)}, \quad \bar{C}_{it} = \frac{1}{K} \sum_{k=1}^K C_{it}^{(k)}, \quad \bar{Y}_{it} = \frac{1}{K} \sum_{k=1}^K Y_{it}^{(k)} \quad (38)$$

These expected values encapsulate both long-term trends and the average effects of uncertainty, serving as inputs for the subsequent optimization model.

## 5.2 Optimization Model Solution and Result Analysis

### 5.2.1 Optimization Model Based on Scenario Expectations

In Problem 3, the objective function remains maximizing total profit; however, its inputs are no longer prices and sales with fixed growth, but rather the expected values derived from the 1000 market scenarios generated by the structured model described above:

$$\max Z = \sum_{t=1}^7 \sum_i (\mathbb{E}[P_{it}] \cdot \min(Y_{it}, \mathbb{E}[Q_{it}]) - \mathbb{E}[C_{it}] \cdot A_{it}), \quad (39)$$

where  $A_{it}$  is the planting area of crop  $i$  in year  $t$ , and  $Y_{it}$  is the actual production determined by the area and yield per mu.

### 5.2.2 Solving the Optimization Model

Utilizing the generated prediction data containing market linkage information—specifically, the expected prices and sales volumes after accounting for substitution and complementarity effects—we re-ran the linear programming model. The objective function and constraints of the model remain consistent with Problem 2 to ensure the comparability of results.

## 5.3 Result Analysis and Comparison

### 5.3.1 Overall Profit Analysis

The model solution indicates that, after comprehensively considering the substitutability and complementarity among crops, the expected total profit for 2024–2030 is 100,229,466.07 CNY. Compared to the 103,034,611.63 CNY obtained in Problem 2, the total profit decreased by approximately 2.7%. This marginal difference suggests that under the settings of this study, the optimal planting strategy possesses high robustness against varying market assumptions. This finding carries three significant implications:

- (1) **Strategy Robustness:** Whether the market is assumed to have independent fluctuations or price-demand linkages, the core decision portfolio—characterized by grain dominance, supplemented by fungi, and legume rotation—remains unchanged. This indicates that the scheme can maintain near-optimal economic performance when facing structural market changes, thereby demonstrating strong risk resilience.
- (2) **Enhanced Model Realism:** The lower profit in Problem 3 reflects the constraint effect on the demand side after introducing market linkages. For instance, a rise in wheat price is typically accompanied by a contraction in its own demand due to price elasticity. Although the demand for certain substitute crops (e.g., barley) increases, the overall addressable market size tends to tighten—this mechanism makes the prediction in Problem 3 closer to real-market behavior.
- (3) **Dominance of Land Constraints:** The profit difference of only 2.7% indicates that the hard constraints of the 54 plots within the study area play a dominant role in determining the profit ceiling. Compared to land resource constraints, the adjustment space provided by market elasticity is relatively limited, highlighting the decisive role of supply-side resource constraints in agricultural planning.

### 5.3.2 Practical Implications

For farmers and policymakers, Problem 3 provides a more conservative yet more achievable profit expectation. Meanwhile, the high consistency of planting layouts under both scenarios suggests that the recommended strategy requires no drastic adjustments under short-term price fluctuations or market linkage shocks. Based on this, we suggest:

- Farmers, when formulating annual planting plans, should prioritize adopting the core portfolio identified in this paper to reduce decision risks;
- In resource allocation and land policy design, priority should be given to optimizing supply-side constraints such as land scale and crop rotation systems to gain more significant room for output improvement;
- Future research could integrate risk measures and dynamic price processes into the model to assess the robustness of strategies under long-term and extreme scenarios.

## 6 Robustness Check

To evaluate the robustness of the optimization model against input uncertainty<sup>[9]</sup>, this section tests the sensitivity of the model output to perturbations by introducing relative additive noise into the price and yield inputs and conducting Monte Carlo simulations. Let the structured predicted values from Problem 3 be  $\tilde{P}_{it}$  and  $\tilde{Y}_{it}$ . In the noisy case:

$$\tilde{P}_{it} = \bar{P}_{it}(1 + \varepsilon_{p,it}), \quad \varepsilon_{p,it} \sim \mathcal{N}(0, \sigma_p^2), \quad (40)$$

$$\tilde{Y}_{it} = \bar{Y}_{it}(1 + \varepsilon_{y,it}), \quad \varepsilon_{y,it} \sim \mathcal{N}(0, \sigma_y^2), \quad (41)$$

where  $\sigma_p$  and  $\sigma_y$  represent the relative noise intensities of price and yield, respectively. We use the Mean Squared Error (MSE) between the profit under noisy conditions  $\tilde{\Pi}$  and the noise-free baseline profit  $\Pi_0$  as the robustness metric:

$$\text{MSE}(\sigma) = \mathbb{E}[(\tilde{\Pi}(\sigma) - \Pi_0)^2] \approx \frac{1}{N} \sum_{k=1}^N (\tilde{\Pi}_k - \Pi_0)^2,$$

where  $N$  is the number of Monte Carlo repetitions (preliminarily set to  $N = 20$  in this paper; it is recommended to increase to  $N \geq 100$  in the final manuscript to provide confidence intervals).

### 6.1 Main Results

Table 6 summarizes the MSE comparison under several noise levels (Unit:  $10^8$  CNY<sup>2</sup>). At the same noise level, the MSE caused by price noise is significantly larger than that caused by yield noise. When  $\sigma_p = \sigma_y = 0.20$ , the price MSE is approximately 176.84, while the yield MSE is approximately 46.02, with a ratio of about 3.84. Furthermore, the MSE

exhibits non-linear growth with noise intensity, showing an inflection point near  $\sigma \approx 0.10$ . This indicates the existence of a critical reliability threshold for the model at approximately 10% relative uncertainty.

Table 6 Comparison of MSE for Price and Yield Noise (Unit:  $10^8$  CNY $^2$ )

Noise Intensity $\sigma$	Price MSE	Yield MSE	Ratio
0.02	1.43	1.26	1.13
0.05	4.51	3.02	1.49
0.10	20.42	11.12	1.84
0.20	176.84	46.02	3.84

## 6.2 Mechanism Explanation

The reason for the dominance of price noise lies in the systematic amplification effect of market transmission. In the model, price affects demand through the  $41 \times 41$  cross-price elasticity matrix  $\mathbf{E} = [e_{ij}]$ :

$$\Delta \ln Q_{it} = \sum_{j=1}^{41} e_{ij} \Delta \ln P_{jt}.$$

Price fluctuations transmit across multiple crops and generate systematic impacts; whereas yield perturbations are treated in the model as more crop-specific and can be partially dispersed through diversified allocation, resulting in a weaker impact on total profit.

## 6.3 Model Evaluation and Generalization

### 6.3.1 Advantages of the Model

(1) Innovation of Structured Prediction Framework: The coupled model of Geometric Brownian Motion and Cross-Price Elasticity constructed in this paper organically integrates the stochasticity of price evolution with the linkage of market demand. It breaks through the static assumption of "future equals historical mean" in traditional agricultural planning, providing a theoretical basis for decision-making under uncertain environments.

(2) Realism of Multi-Constraint Integration: The model comprehensively considers agronomic constraints (legume rotation, continuous cropping limits), economic constraints (diversity, market demand caps), and physical constraints (plot compatibility, seasonal limitations). This ensures that the optimization scheme is technically feasible and economically reasonable, avoiding the pitfall of pursuing pure mathematical optimality while detaching from reality.

(3) Quantitative Basis for Risk Management: Through noise sensitivity analysis, this paper clearly identifies price risk as the dominant factor in fluctuations of farmers' income (MSE is approximately 4 times that of yield risk) and quantifies the critical threshold for model reliability (uncertainty  $\sigma \approx 0.10$ ). This provides a scientific basis for policymakers to allocate risk management resources.

(4) Refined Treatment of Loss Rates: The introduction of loss rates to correct the actual unit cost (Cost = Planting Cost / (1 - Loss Rate)) makes the expected sales volume closer to agricultural production reality, avoiding profit overestimation caused by ignoring post-harvest losses.

### 6.3.2 Limitations of the Model

(1) Theoretical Nature of the Elasticity Matrix: The cross-price elasticity matrix is derived based on substitution elasticity theory and historical price correlations, lacking empirical econometric estimation based on sales data. Upon obtaining sufficient market data, the elasticity parameters should be re-calibrated using demand system models (such as the Almost Ideal Demand System, AIDS).

(2) Simplification of Risk Preferences: The model assumes farmers are risk-neutral and have no financial constraints, solely maximizing expected profit. In reality, smallholder farmers are often risk-averse and face liquidity constraints. Future research could introduce Mean-Variance objective functions or Conditional Value at Risk (CVaR) constraints to explicitly consider downside risk.

(3) Static Compatibility Assumption: The assumption that plot types and crop compatibility remain unchanged over seven years ignores the possibilities of land improvement (e.g., converting dry land to irrigated land) and greenhouse expansion. For large-scale farms with long-term investment capabilities, infrastructure investment decisions should be incorporated into the optimization model.

### 6.3.3 Generalization and Extension of the Model

The core methodology of this model—structured stochastic prediction and market-coupled optimization—possesses broad generalizability. In the spatial dimension, the framework can be extended to other agricultural regions by simply

re-calibrating climate parameters, market elasticities, and constraints. In the temporal dimension, the model supports rolling horizon planning: updating predictions annually based on the latest market data and re-solving for the optimal scheme for subsequent years, thereby transforming the 7-year static plan into a dynamic adjustment mechanism to enhance adaptability to policy changes and market fluctuations.

Directions for technical extension include: (1) Machine Learning Enhancement: Adopting deep learning models such as LSTM or Transformer<sup>[10]</sup> to replace GBM for price prediction, leveraging their ability to capture non-linear patterns and long-term dependencies; (2) Real-time Data Integration: Developing a decision support system interfaced with agricultural big data platforms (e.g., meteorological monitoring, market price APIs) to realize automatic parameter updates and dynamic scheme recommendations; (3) Multi-objective Extension: Incorporating environmental objectives such as minimizing carbon emissions and conserving water resources alongside profit maximization to construct a Pareto frontier for sustainable agriculture, providing quantitative tools for green development.

Through these improvements, this model is expected to evolve from a single-village case study into a regional, intelligent agricultural decision-making platform, providing theoretical support and technical assurance for the precision management and sustainable development of modern agriculture.

## Reference

- [1] Jiang C, Jiang C. Promoting supply-side structural reform in agriculture and implementing the rural revitalization strategy[J]. Issues in Agricultural Economy, 2018(03): 4–12.
- [2] Glen J J. Mathematical models in farm planning: A survey[J]. Operations Research, 1987, 35(5): 641-666.
- [3] Hazell P B R, Norton R D. Mathematical Programming for Economic Analysis in Agriculture[M]. New York: Macmillan, 1986.
- [4] Xiao D, Tao F. Response of phenology and yield of winter wheat and summer maize to climate change in the North China Plain over the past 30 years[J]. Chinese Journal of Eco-Agriculture, 2012, 20(11): 1539-1545.
- [5] Lobell D B, Field C B. Global scale climate–crop yield relationships and the impacts of recent warming[J]. Environmental Research Letters, 2007, 2(1): 014002.
- [6] Schwartz E S. The stochastic behavior of commodity prices: Implications for valuation and hedging[J]. The Journal of Finance, 1997, 52(3): 923-973.
- [7] Deaton A, Muellbauer J. An almost ideal demand system[J]. The American Economic Review, 1980, 70(3): 312-326.
- [8] Qian J, Zhao Z. Analysis of the mechanism and effect of price support policy on grain prices[EB/OL]. China Academy for Rural Development (CARD), Zhejiang University, 2019-09-20.
- [9] Hardaker J B, Huirne R B, Anderson J R, et al. Coping with Risk in Agriculture[M]. Wallingford: CABI Publishing, 2004.
- [10] Liakos K G, Busato P, Moshou D, et al. Machine learning in agriculture: A review[J]. Sensors, 2018, 18(8): 2674.