ISSN 2664-4010 (Print) & ISSN 2664-6730 (Online)
**South Asian Research Journal of Agriculture and Fisheries**
*Abbreviated Key Title: South Asian Res J Agri Fish*

**| Volume-8 | Issue-5 | Nov-Dec-2026 |** **DOI: https://doi.org/10.36346/sarjaf.2026.v08i05.001**

**Original Research Article**

# Grain Self-Sufficiency, Not Market Price, Determines the Revenue-Maximising Cropping Structure of Chongming District, Shanghai, China: An XGBoost–SHAP–Linear Programming Analysis

Zixi Liu[1]*, Guang Zhang[1]

[1]Business School, University of Shanghai for Science and Technology, Shanghai, China

***Corresponding Author:** Zixi Liu
Business School, University of Shanghai for Science and Technology, Shanghai, China

**Article History**
Received: 09.09.2026
Accepted: 03.11.2026
Published: 05.11.2026

**Abstract:** Peri-urban agricultural districts must reconcile the grain area that food-security policy requires with the revenue that high-value crops can earn. Which constraint binds this trade-off is rarely identified. Prediction studies stop at accuracy; optimisation studies take the constraint set as given. Neither shows whether prediction quality or policy design governs the planting decision. We link XGBoost yield and price prediction with SHAP attribution and a revenue-maximising linear programme for five crops in Chongming District, Shanghai, China, over the ten modelled years 2015–2024, benchmark every prediction against a naive alternative, and solve the programme across a range of grain-area floors. The prediction layer is an input to the answer, not a contribution in itself. The yield model beats a per-crop historical mean for only one of five crops, and the price model is beaten by a last-observation rule on every metric. Yet the optimal structure is unchanged by ±40 % perturbations of vegetable price or ±30 % perturbations of vegetable yield, and responds only to the grain floor; the exception is maize, whose expansion rests on a single unsustained 2024 price observation. The district's official grain target of 18,667 ha is already met by its 2024 structure (18,703 ha), so food security costs nothing at the stated target. The 0.390 billion CNY (84,270 CNY per additional ha) is the cost of grain held above it. The revenue-maximising planting structure is set by the grain floor, not by market price.

**Keywords:** Crop Planting Structure, Food Security, Grain Self-Sufficiency, XGBoost, SHAP, Linear Programming.

## 1. INTRODUCTION

Cropland allocation is one of the few agricultural decisions that governments regulate directly, and in peri-urban China it is pulled between two objectives: the revenue growers can earn from high-value crops and the grain area that policy obliges them to maintain [1, 2]. Chongming District, the island district of Shanghai, concentrates that tension. It supplies roughly one third of the municipality's locally produced agricultural products from about one fifth of its land area, and its rice planting area reached about 18,280 ha in 2023 at a scale-operation rate of about 82 %. Its 2035 land-use plan requires the district to hold cultivated land at no less than 42,333 ha, of which 36,000 ha is designated permanent basic farmland, and the district has designated 44,220 ha of cultivated-land protection space in total, against an existing cultivated-land stock of 49,100 ha. Its annual grain-production target is a sown area of no less than 18,667 ha and an output of no less than 150,000 t. Whether such a floor or the market price governs what is planted determines how much revenue the district forgoes to remain self-sufficient in grain.

Two research lines have grown up around this allocation problem and rarely meet. One treats crop yield, and less often crop price, as a prediction target and applies machine learning to it; reviews of machine learning in agriculture map this line [3, 4]. The other casts the same problem as a mathematical programme over land and other resources, usually maximising profit or revenue and solved by linear, goal or multi-objective programming [5-7], including recent national- and provincial-scale applications that coordinate water saving, grain revenue and ecosystem services [8, 9].

**Copyright © 2026 The Author(s):** This is an open-access article distributed under the terms of the Creative Commons Attribution **4.0 International License (CC BY-NC 4.0)** which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited.

Two strands of this literature have developed separately, and each leaves the same question open. Prediction studies are organised around accuracy and stop there. Recent work applies genetic-algorithm-tuned deep networks to wheat yield [10], CNN–Bi-LSTM hybrids to sugarcane [11], spatio-temporal deep learning to county-level maize [12], multi-stage weather models [13] and soil-microbial biomarkers [14], to yield. In all of it prediction is the end point: once a satisfactory metric is reached the analysis stops, and the decision the prediction was meant to inform is never solved. Optimisation studies take the constraint set as given and rarely ask which element of it binds. Profit or revenue maximisation is standard [5, 6], and goal programming trades revenue against cost, labour and land [7], where food security appears, it usually provides motivation rather than a constraint [5]. The instructive exceptions find that the staple-area bound, not price, governs the solution: a village-level allocation in China admitted few feasible solutions under rigid staple-crop area constraints, with the grain-area bound rather than prices controlling the feasible space [15], and a partial-equilibrium model of Shanghai's districts attributed a reallocation of the cropping mix to an environmental policy target, with Chongming among the districts most affected [16]. Policy-econometric evidence points the same way, with farmland-protection policy shifting the grain share of planting structure [17], and non-grain land use tracking arable-land policy rather than local prices [18, 19]. What neither strand establishes is whether prediction quality changes the decision it feeds. Outside agriculture the question is settled: minimising prediction error is not the same as improving the downstream decision [20], even a perfect conditional-mean estimate can prescribe suboptimally [21], and a benchmark literature now compares methods for closing that gap [22], yet agricultural studies apply the two-stage pipeline without that examination [23].

The question this paper addresses is narrower than the usual one. We do not ask what the optimal cropping structure is; we ask which constraint determines it. In a district whose grain floor is administratively set while the price of its highest-value crop moves freely, does the revenue-maximising structure respond to price at all, and what does the grain floor cost per ha? Price volatility is often invoked as a source of income risk for growers in China, but risk is not the same as a binding constraint on land allocation. The setting itself limits the answer. The five modelled crops occupy between 71 % and 82 % of the district's total sown area, and three hard constraints bound the programme: a total sown-area budget fixed at the 2014–2024 mean of 51,238 ha, a floor G on the combined area of wheat, rice and maize, whose official value is 18,667 ha, and a rotation constraint preventing overwintering crops from exceeding the rice area. Revenue is the objective; the grain floor is the parameter of interest. Costing self-sufficiency has precedent: at national scale land is the binding constraint on food self-sufficiency and the resulting welfare loss falls mainly on farmers [24], and at district scale in the same municipality a policy target has been shown to reallocate the cropping mix [16]. What has not been done is to impose a grain-area floor on a sub-national cropping structure and price it at the margin.

Three features of the data and methods condition the design. The panel is small (ten annual observations per crop after the 2014 lag year is removed), so a pooled fit can be dominated by between-crop variance rather than model skill. The failure mode is documented, not hypothetical: on small agronomic panels, benchmark models have outperformed many of the machine-learning models tested in a multi-country grain forecasting study [25], a linear-trend baseline has matched or beaten a regularised regression for wheat [26], and a one-month last-observation rule has attained the lowest error for every commodity examined and resisted four tree-ensemble models [27]. Under leakage-free walk-forward validation, machine-learning and deep-learning models have failed to outperform the naive benchmark significantly for monthly commodity prices, with the first price lag supplying the bulk of the predictive gain [28], leakage and validation design are now recognised as systematic sources of inflated accuracy [29-31]. Judging the models here requires per-crop benchmarks rather than a single pooled coefficient of determination. Prediction error matters too, because predicted yields and prices enter the programme as coefficients: how far that error propagates into the optimum cannot be read off prediction metrics [20, 21], and has to be measured by perturbing the inputs and by re-solving the programme with naive coefficients. Attribution raises a related issue. The district needs to know what drives yield and price, but Shapley values do not remove correlation among features, and their game-theoretic axioms have been questioned for feature importance [32]. The attributions reported below rank drivers and are not read causally.

We assemble an annual panel for Chongming District on rice, wheat, maize, rapeseed and vegetables for 2014–2024, drawing on statistical yearbooks, wholesale price records and meteorological observations. XGBoost models predict yield and price, chosen because gradient-boosted trees suit small-sample tabular data of mixed scale and have proved competitive well beyond agronomy [14-34], and SHAP attributes both. The predictions are embedded as coefficients of a linear programme that maximises revenue subject to the three constraints above and is solved to its exact global optimum by the simplex method. Because the grain floor is a policy variable with an official value, the programme is solved across a range of floors spanning and exceeding it, and the resulting cost curve is reported. Every model is reported alongside naive benchmarks (a per-crop historical mean for yield and a last-observation rule for price), so that the value of the prediction layer is measured rather than assumed.

The paper makes three contributions. The first contribution is a negative result. Perturbing the vegetable price by ±40 % and the vegetable yield by ±30 % leaves the revenue-maximising allocation unchanged, and the structure responds

only to the grain-area floor; prediction accuracy, in this setting, is not what determines the planting decision. The second prices the grain constraint and locates its threshold. The district's official annual target of 18,667 ha is already met by its 2024 structure and is slack in the programme, so food security costs the district nothing at its stated target, and the cost attaches only to grain held above it; raising the floor from the district's 2024 actual grain area of 18,703 ha to a 23,333 ha reference level, 25 % above the official target, lowers maximum attainable revenue by 0.390 billion CNY, or about 84,270 CNY per additional ha of grain, and the full cost curve over 18,667–33,333 ha is reported. The third is a baseline audit of the prediction layer, in which the pooled yield coefficient of determination of 0.9974 is driven mainly by between-crop variance, the yield model beats a per-crop historical mean for only one of the five crops, and the price model is beaten by a last-observation rule on every metric. Materials and Methods develops the framework; the Results and Discussion section reports the optimisation results and their sensitivity analysis. In this respect the paper sits closer to the policy literature, in which supply constraints rather than prices shape cropping structure [1-16], than to the optimisation literature from which it borrows its method.

## 2. MATERIALS AND METHODS

### 2.1 Study Area

Chongming District lies at the mouth of the Yangtze River (Fig-1) and comprises the islands of Chongming, Changxing and Hengsha, of which Chongming Island is the third largest island in China and the largest estuarine alluvial island in the world[1], with a total land area of about 1,413 km² and about 49,100 ha of cultivated land; among the districts of Shanghai it has the largest cultivated-land reserve and the highest share of agricultural output. According to the Chongming Statistical Yearbook, the district contributes more than 20 % of the municipality's annual grain output and more than 30 % of its vegetables. Its designation as a "world-class ecological island" has also increased the tension among green agricultural transformation, tightening resource and environmental constraints, and adjustment of the planting structure. Demand for well-grounded crop-layout decisions has grown accordingly.

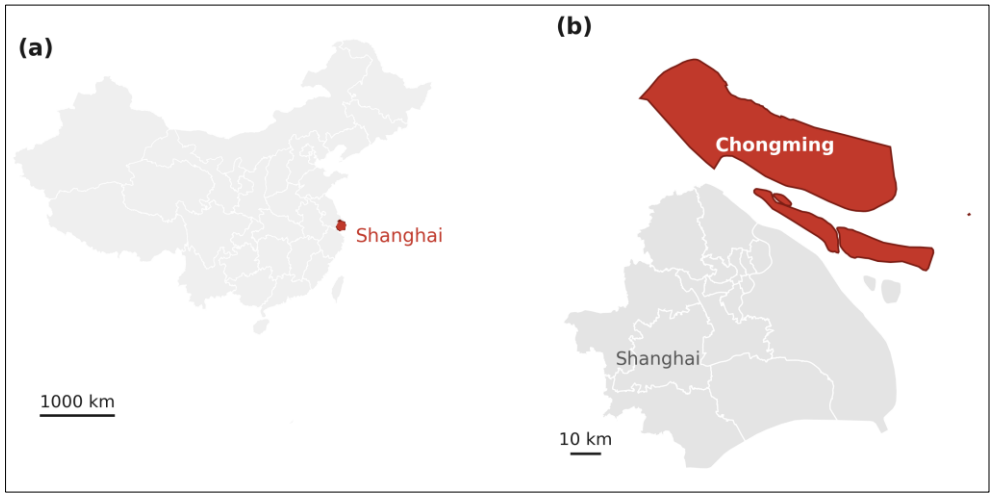



**Fig. 1: Location of the study area. (a) China; (b) Shanghai municipality, with Chongming, Changxing and Hengsha highlighted [2]**

### 2.2 Data

We assembled an annual panel for the eleven years 2014–2024 covering four groups of variables for the five crops: production, market, meteorology and regional economy. Crop yield (kg/ha) and fertiliser application rate (kg/ha) were taken from the Chongming Statistical Yearbook (2014–2025 editions). Crop prices (CNY/kg) were derived from the daily wholesale prices of the National Agricultural Products Wholesale Market Price Information System and of the main agricultural markets of Shanghai, averaged to annual means; the agricultural PPI (previous year = 100) came from the Shanghai Statistical Yearbook, and the minimum purchase prices of wheat and rice (CNY/50 kg) from the annual policy announcements of the National Development and Reform Commission. Growing-season mean temperature (°C) and mean

[1] Climate. Chongming has a northern subtropical monsoon climate with four distinct seasons, and rainfall is concentrated in the growing season. Over 2014–2024, mean temperature during each crop's growing season at Chongming station (No. 58366) averaged 14.3 °C for wheat (November–June), 17.6 °C for rapeseed (October–May), 21.7 °C for rice (June–October), 25.1 °C for maize (April–August) and 19.0 °C for year-round vegetables; the corresponding precipitation averaged 551, 534, 635, 521 and 1,358 mm. These daily observations from the National Meteorological Science Data Center (https://data.cma.cn) are the meteorological inputs to the yield models.

[2] Administrative boundaries are drawn from the public Aliyun DataV GeoJSON basemap (https://datav.aliyun.com/portal/school/atlas/area_selector; accessed 19 October 2026) at a latitude-corrected aspect ratio.

precipitation (mm) were obtained from daily observations at the Chongming station (No. 58366) of the National Meteorological Science Data Center, aggregated over the phenological window of each crop. Rural population, total agricultural output value and rural per-capita disposable income were also taken from the Chongming Statistical Yearbook [3].

### 2.3 Data Preprocessing

The raw data were preprocessed in five steps to limit the effect of noise, missing values and differences of scale on model performance. Step 1: Zero values were treated as missing and filled by linear interpolation. Step 2: Growing-season information was extracted by regular-expression matching to obtain the number of growing months, and crop types were one-hot encoded according to their ecological attributes. Step 3: One-period lagged terms were constructed for core variables such as yield and price to capture temporal dependence, and first-year samples were removed so that every lagged feature rests strictly on a true previous-period value and no data leakage arises. Step 4: Variables were combined on agricultural-economic and agronomic grounds to derive higher-order features, while contemporaneous derived features whose causality runs opposite to the prediction targets were removed: the contemporaneous per-unit-area revenue (= yield × price × 2) contains both prediction targets, so only its one-period lagged form was retained. Step 5: Min-Max normalisation compressed the numerical variables to the [0, 1] interval. Given only ten annual observations, Winsorisation would not reach its threshold and XGBoost is robust to multicollinearity, so neither Winsorisation nor variance-inflation-factor removal was applied.

### 2.4 Crop Yield and Price Prediction Using XGBoost

XGBoost is an efficient and scalable machine-learning algorithm built on the gradient-boosting framework: it combines multiple decision trees to improve prediction progressively, and it supports regularisation against overfitting together with automatic handling of missing values [35]. Because the annual data contain only ten time-series points and the predictors are somewhat collinear, XGBoost was used to build both the yield and the price model. The dataset was not randomly shuffled: the first 80 % of the chronological samples (2015–2022) formed the training set and the remaining 20 % (2023–2024) a held-out test set that took no part in model selection; 2025 lies outside the panel and is predicted forward. Hyperparameters were tuned by grid search with three-fold time-series cross-validation inside the training period, minimising the cross-validated root-mean-square error (RMSE). The selected configurations are given in Table-1. All models were implemented in Python 3.13.5 with xgboost 3.2.0, shap 0.51.0, scikit-learn 1.8.0 and SciPy 1.17.1 (HiGHS solver) under a fixed random seed of 42; the analysis code is provided in Supplementary Material S1.

**Table 1: Optimal hyperparameter configurations of the XGBoost yield- and price-prediction models**

| Parameter | Meaning | Yield model | Price model |
|---|---|---|---|
| learning_rate | Learning rate | 0.05 | 0.05 |
| n_estimators | Number of boosting rounds | 300 | 300 |
| max_depth | Maximum tree depth | 5 | 5 |
| subsample | Sample sampling ratio | 0.8 | 0.8 |
| colsample_bytree | Feature sampling ratio | 0.8 | 0.8 |
| reg_alpha | L1 regularisation coefficient | 0 | 1 |
| reg_lambda | L2 regularisation coefficient | 2 | 1 |

### 2.5 Model Interpretability Using SHAP

To identify the key drivers of crop yield and market price, SHAP was used to interpret the XGBoost models [36]. SHAP assigns each feature the value of its marginal contribution to the prediction, derived from Shapley values in cooperative game theory; it satisfies the properties of consistency and local accuracy, and can be computed quickly for tree models. The Shapley value of feature *i* is defined as:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!} [v(S \cup \{i\}) - v(S)] \quad (1)$$

In Eq. (1), $N$ is the set of all features, $S$ is any subset not containing feature $i$, $|S|$ is the number of features in subset $S$, and $v(S)$ and $v(S \cup \{i\})$ are the model outputs based on subset $S$ and on $S$ with feature $i$ added, respectively.

[3] Data sources. Chongming Statistical Yearbook, 2014–2025 editions (Chongming District Bureau of Statistics, Shanghai), https://www.shcm.gov.cn; Shanghai Statistical Yearbook (Shanghai Municipal Bureau of Statistics), https://tjj.sh.gov.cn; National Agricultural Products Wholesale Market Price Information System (Ministry of Agriculture and Rural Affairs), https://pfsc.agri.cn; minimum purchase prices from the annual announcements of the National Development and Reform Commission, https://www.ndrc.gov.cn; meteorological observations from the National Meteorological Science Data Center, station 58366, https://data.cma.cn. All accessed 19 October 2026.

**2.6 Planting-Structure Optimisation Using Linear Programming**

The sown areas of the five crops (wheat, rice, rapeseed, maize and vegetables) were taken as the decision variables, denoted $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)$, in ha. The objective was to maximise total regional planting revenue:

$$\max Z = \sum_i \left(y_i \times p_i \times x_i\right) \quad (2)$$

In Eq. (2), $y_i$ is the predicted yield of crop $i$ (kg/ha) and $p_i$ is the predicted market price (CNY/kg). Five groups of constraints were imposed:

1. **Total Sown-Area Constraint**:

The sum of the crop areas equals 51,238 ha, the 2014–2024 average total sown area of the five crops, counting summer and winter cropping separately. This is why the sown area exceeds the district's cultivated-land stock of 49,100 ha[4]. That stock is in turn above the cultivated-land retention floor of 42,333 ha the district is required to maintain through 2035[5], and the sown area also exceeds the 44,220 ha of cultivated-land protection space the district has designated, of which 36,000 ha is permanent basic farmland[6]; that is, $\Sigma_i x_i = 51{,}238$;

2. **Food-Security Constraint**:

The total area of grain crops (wheat, rice and maize) is no less than a floor value $G$, i.e., $x_1 + x_2 + x_4 \geq G$ ($G$ is a policy variable with an official value: the district's annual grain-production target sets a sown area of no less than 18,667 ha, with an output target of no less than 150,000 t[7], and $G$ follows the area component of the target (18,667 ha). The programme is solved for $G$ over 18,667–33,333 ha, the reference solution being taken at $G$ = 23,333 ha, 25 % above the official target, and the resulting cost curve is reported in the sensitivity analysis below);

3. **Seasonal Rotation-Matching Constraint**:

After rice (summer, June–October) is harvested, wheat (November–June of the next year) or rapeseed (October–May of the next year) is rotated in winter on the same plot, forming "rice–wheat" and "rice–rapeseed" rotations, so that the total area of overwintering crops does not exceed the rice area, i.e., $x_1 + x_3 \leq x_2$; in the optimal solution this constraint is slack (wheat + rapeseed 1,644 ha against rice 18,438 ha) and does not bind; it is retained only to keep the solution agronomically admissible;

4. **Lower- And Upper-Bound Constraints**:

Each crop's area is no less than 50 % of its historical mean (preventing overwintering crops from being zeroed out and ensuring all five crops remain non-zero) and no more than its historical maximum sown area (preventing extreme monoculture); and

5. **Non-Negativity Constraints** $x_i \geq 0$.

Combining the objective with constraints 1–5, the linear programming model can be formulated as follows:

$$\max Z = \sum_{i=1}^{5} y_i p_i x_i \quad (3)$$

$$\text{s.t.} \sum_{i=1}^{5} x_i = 51{,}238$$

$$x_1 + x_2 + x_4 \geq G$$

$$x_1 + x_3 \leq x_2$$

$$0.5\bar{x}_i \leq x_i \leq x_i^{\max}, i = 1, 2, \ldots, 5$$

$$x_i \geq 0, i = 1, 2, \ldots, 5$$

[4] Miao, J. (2024, May 22). Hold the red line of cultivated-land protection and build a world-class ecological island to a high standard. China Natural Resources News. https://ghzyj.sh.gov.cn/hyxw/20240524/84ff62a726f242beaec54fff88fd507f.html

[5] Shanghai Municipal People's Government. (2018, May 11). Reply approving in principle the Chongming District Master Plan and Land Use Master Plan (2017–2035) (Hu Fu [2018] No. 40). https://www.shanghai.gov.cn/nw42844/20200823/0001-42844_56055.html

[6] Chongming District People's Government. (2023, May 15). Reply to Proposal No. 0276 of the First Session of the 14th Shanghai Municipal Committee of the CPPCC. https://shcm.gov.cn/ztzl/009007/009007003/20230515/c77f73d4-6ed0-4afc-84f8-f40562035686.html

[7] Chongming District People's Government. (2025, April 14). Chongming District holds an on-site meeting on spring agricultural production. https://shcm.gov.cn/bmpd/019006/019006002/20250414/35f9a0a0-fefb-4b73-8124-8dc088d4d5a4.html

In Eq. (3), the lower and upper bounds on each crop's area are those described under constraint 4, G is the grain-area floor, and the remaining symbols are as defined for Eq. (2).

Since the objective function and all constraints are linear, the problem is a linear programme, solved to its exact global optimum by the simplex method (HiGHS solver). The whole framework (the data, the two XGBoost models, the SHAP attribution and the linear programme, together with the two stages that qualify it) is summarised in Fig-2.

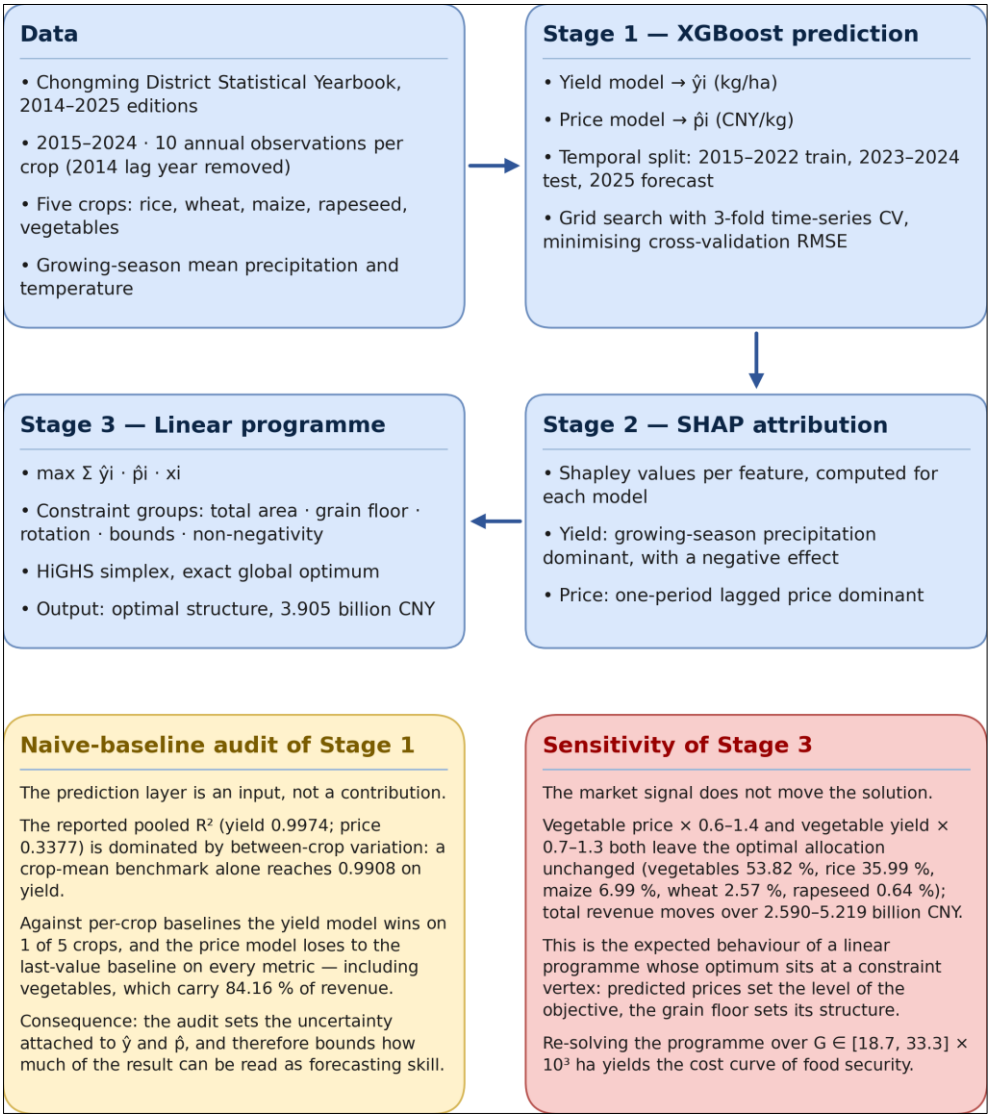



**Fig. 2: Analytical framework linking XGBoost prediction, SHAP attribution and linear-programming optimisation**

### 2.7 Scope of the Crop Menu

The land budget $\Sigma_i x_i$ = 51,238 ha is the average sown area of the five modelled crops, not of all crops grown in the district. In the years for which the yearbook publishes comparable subtotals (2014, 2018–2021 and 2024) the five crops account for 71–82 % of total sown area, with a mean of 76 %; the remainder consists of crops outside the model, including soybean, barley and other minor grains (which the yearbook counts inside its own grain subtotal and which exceeded the model's three grain crops by 232.3 ha in 2024), together with cotton, melons and other non-grain crops. All results reported below are conditional on this five-crop menu. Other crops, aquaculture, livestock and other land uses are treated as exogenous and fixed, and the area released by the contraction of the five modelled crops is assumed not to have been reallocated to crops outside the model. Because the omitted grain crops would also count towards the self-sufficiency target, leaving them out if anything raises the estimated cost of that target.

# 3. RESULTS AND DISCUSSION

## 3.1 Descriptive Statistics

Table-2 and Table-3 summarise the descriptive statistics of yield and price for the five crops over 2014–2024. Yield rose steadily over the period and accelerated after 2015, mainly through high-standard farmland construction, improved seed varieties and mechanisation. Rice and maize stayed high and stable over the long term. Vegetables had by far the highest absolute yield (29,197.95 kg/ha) and the smallest relative variability; wheat and rapeseed fluctuated the most in relative terms. Prices fluctuated around an upward trend, with pronounced differences between crops. By coefficient of variation, vegetables were the most volatile, followed by maize and rice; wheat and rapeseed remained relatively stable.

**Table 2: Descriptive statistics of yield for the five crops in Chongming District (2014–2024)**

| Crop | Mean (kg/ha) | SD (kg/ha) | Min (kg/ha) | Max (kg/ha) |
|---|---|---|---|---|
| Wheat | 4,694.70 | 978.75 | 3,225.00 | 6,456.00 |
| Rice | 8,282.85 | 239.55 | 7,995.00 | 8,715.00 |
| Maize | 6,746.10 | 269.70 | 6,160.95 | 7,050.00 |
| Rapeseed | 2,445.60 | 381.75 | 1,950.00 | 3,090.00 |
| Vegetables | 29,197.95 | 477.60 | 28,693.05 | 30,090.00 |

**Table 3: Descriptive statistics of price for the five crops in Chongming District (2014–2024)**

| Crop | Mean (CNY/kg) | SD (CNY/kg) | Min (CNY/kg) | Max (CNY/kg) |
|---|---|---|---|---|
| Wheat | 2.40 | 0.26 | 1.96 | 2.94 |
| Rice | 4.38 | 1.50 | 2.64 | 6.28 |
| Maize | 2.66 | 0.96 | 1.88 | 5.00 |
| Rapeseed | 5.28 | 0.50 | 4.46 | 6.22 |
| Vegetables | 4.30 | 1.96 | 2.30 | 7.82 |

## 3.2 Prediction Results

The yield model attains a pooled $R^2$ of 0.9974 and an RMSE of 484.95 kg/ha on the test years, and the price model a pooled $R^2$ of 0.3377 and an RMSE of 0.9940 CNY/kg. Because the five crops differ by an order of magnitude in yield level (rapeseed 2,445.60 against vegetables 29,197.95 kg/ha), these pooled statistics are dominated by between-crop variation and overstate forecasting skill: a benchmark that simply predicts each crop's training-period mean already attains $R^2 = 0.9908$ on the same test set. Evaluated crop by crop (Table-4), the yield model outperforms that benchmark for only one of the five crops (wheat, MAE 453.75 against 2,007.60 kg/ha) and is worse for the other four, most markedly for rapeseed (MAE 525.75 against 44.10 kg/ha). The price model is beaten on every pooled metric by a last-observation benchmark ($R^2$ 0.3377 against 0.5701; RMSE 0.4970 against 0.4004; MAE 0.4046 against 0.2910), and for vegetables, which contribute 84.16 % of the optimised revenue, its mean absolute error is 6.6 times that of the benchmark (0.86 against 0.26 CNY/kg). With only two test years per crop (n = 2; ten pooled crop-year observations), these MAE differences cannot support a significance test, so the comparison is made on point estimates. Removing the six crop-category indicators from the price model raises its pooled $R^2$ from 0.3377 to 0.5008, indicating that the crop identity is being used to memorise crop-level price levels rather than to predict their movement. Prediction feeds the framework; it is not a contribution of it. The optimisation results are read through the sensitivity analysis below, not as point estimates. Table-5 reports the predicted 2025 yields and prices of the five crops. Fig-3 presents the same comparison crop by crop, alongside the pooled statistics above, as a point plot.

**Table 4: Mean absolute error of the XGBoost models against naive benchmarks on the 2023–2024 test set, by crop**

| Crop | Yield MAE — XGBoost (kg/ha) | Yield MAE — Crop mean (kg/ha) | Yield MAE — Last value (kg/ha) | Price MAE — XGBoost (CNY/kg) | Price MAE — Last value (CNY/kg) |
|---|---|---|---|---|---|
| Wheat | 453.75 | 2,007.60 | 1,082.25 | 0.40 | 0.34 |
| Rice | 173.85 | 164.25 | 164.25 | 0.48 | 0.18 |
| Rapeseed | 525.75 | 44.10 | 559.50 | 0.26 | 0.92 |
| Maize | 273.90 | 107.25 | 438.75 | 1.18 | 1.20 |
| Vegetables | 596.40 | 309.60 | 321.00 | 1.72 | 0.26 |

**Table 5: Predicted 2025 yields and prices for the five crops in Chongming District**

| Crop | Yield (kg/ha) | Price (CNY/kg) |
|---|---|---|
| Wheat | 6,585.60 | 2.96 |
| Rice | 8,091.15 | 3.22 |
| Maize | 6,371.40 | 4.74 |

| Crop | Yield (kg/ha) | Price (CNY/kg) |
|---|---|---|
| Rapeseed | 2,943.45 | 5.52 |
| Vegetables | 28,774.65 | 4.14 |

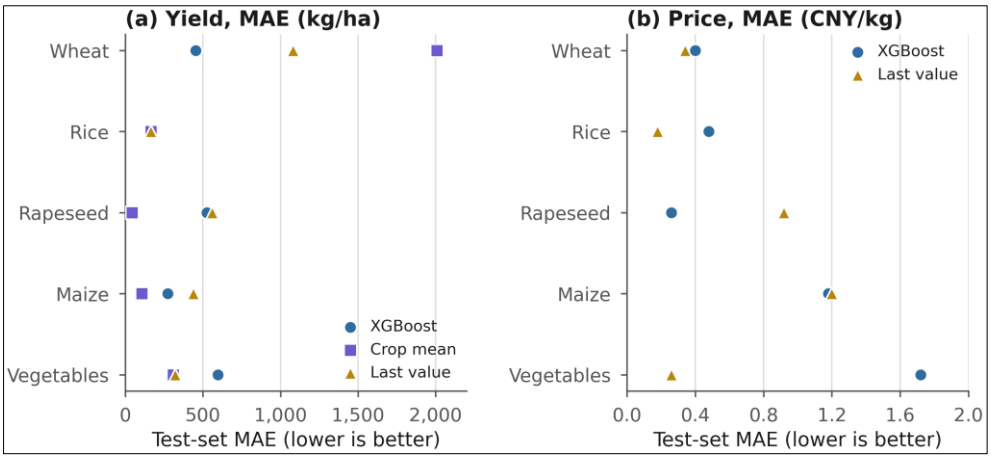



**Fig. 3: Mean absolute error of the XGBoost models against naive benchmarks on the 2023–2024 test set: (a) yield; (b) price**

### 3.3 Optimisation Results

Table-6 reports the optimal planting structure from the linear programme, at a total revenue of approximately 3.905 billion CNY. Fig-4 and Fig-5 show the area-share and revenue-share distributions of each crop. For reference, the same total sown area under the district's 2024 planting structure, valued with the same predicted 2025 coefficients, would return 4.065 billion CNY (see the sensitivity analysis below). Table-6 entries are rounded; totals are computed from unrounded values (see the note to the table).

**Table 6: Optimised crop planting structure in Chongming District**

| Crop | Area (ha) | Area share (%) | Revenue (billion CNY) | Revenue share (%) |
|---|---|---|---|---|
| Wheat | 1,315 | 2.57 | 0.026 | 0.66 |
| Rice | 18,438 | 35.99 | 0.479 | 12.27 |
| Rapeseed | 329 | 0.64 | 0.005 | 0.14 |
| Maize | 3,580 | 6.99 | 0.108 | 2.77 |
| Vegetables | 27,576 | 53.82 | 3.286 | 84.16 |
| Total | 51,238 | 100.00 | 3.905 | 100.00 |

**Note**: *totals are calculated from unrounded values; components rounded to the precision shown may not sum exactly to the stated totals.*

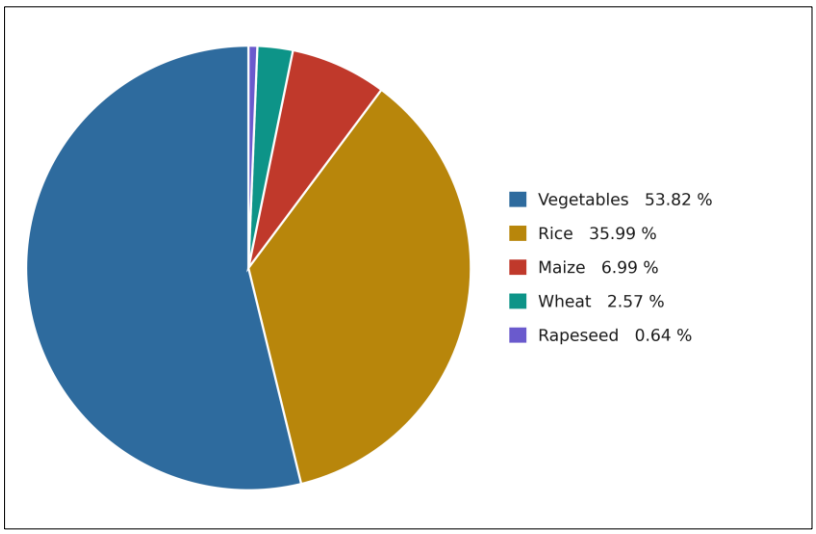



**Fig. 4: Distribution of planted area by crop in the optimised (linear-programme) structure.**

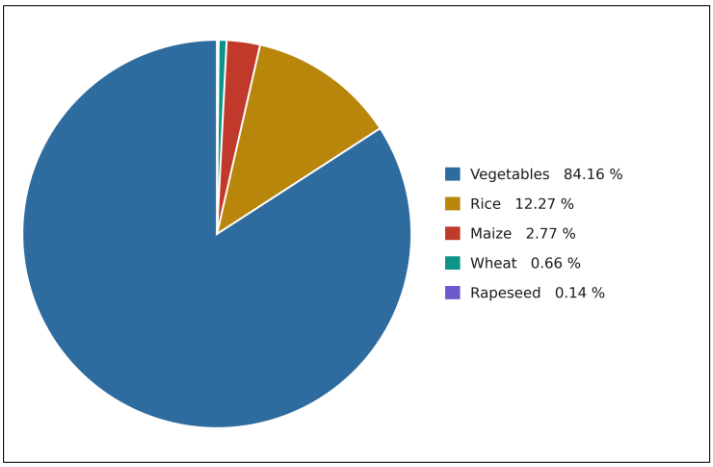



**Fig. 5: Revenue contribution by crop in the optimised structure**

Vegetables dominate both area and revenue: they occupy 53.82 % of the area and contribute 84.16 % of the revenue, at 119,175 CNY per ha, the highest of the five crops. That follows from their very high yield and from market-based pricing, and is consistent with Chongming's role as Shanghai's "vegetable basket" base. Their area is not capped by their own historical maximum (solving the programme without that upper bound leaves the vegetable allocation unchanged), but by the total-area identity together with the grain-area floor, which leaves them no additional land. Rice accounts for 35.99 % of the area and 12.27 % of the revenue and is the crop through which the food-security floor is met and the rotation constraint satisfied; its area (18,438 ha) far exceeds the combined wheat and rapeseed area (1,644 ha), because a share of Chongming's paddy land lies idle in winter. Maize accounts for 6.99 % of the area and 2.77 % of the revenue; at 30,195 CNY per ha it earns more than any other grain crop and so is allocated up to its historical maximum (3,580 ha), while the food-security floor is still met. Wheat and rapeseed earn the least per ha (19,545 and 16,230 CNY respectively) and are pushed to their lower bound, "no less than 50 % of the historical mean" (1,315 and 329 ha); that bound prevents overwintering crops from being driven to zero under pure revenue maximisation, and so preserves the rotation system and the diversity of the planting structure.

### 3.4 SHAP-Based Driver Analysis

In the yield model (Fig-6), growing-season mean precipitation is the leading driver of crop yield, followed by crop type (especially vegetables) and then by variables such as the one-period lagged yield and the one-period lagged price. Its SHAP values are negative in 32 of the 40 training observations (80 %), including all 24 grain-crop observations, and positive only for vegetables, so water conditions limit grain-crop growth strongly. In the price model (Fig-7), the one-period lagged price contributes most. The grain-crop indicator has a negative SHAP value in all 24 grain-crop observations (mean −0.310 CNY/kg) and a positive value in all 16 non-grain observations (mean +0.454 CNY/kg), consistent with grain and non-grain crops having structurally different price-formation mechanisms and with greater prediction uncertainty for non-grain crop prices.

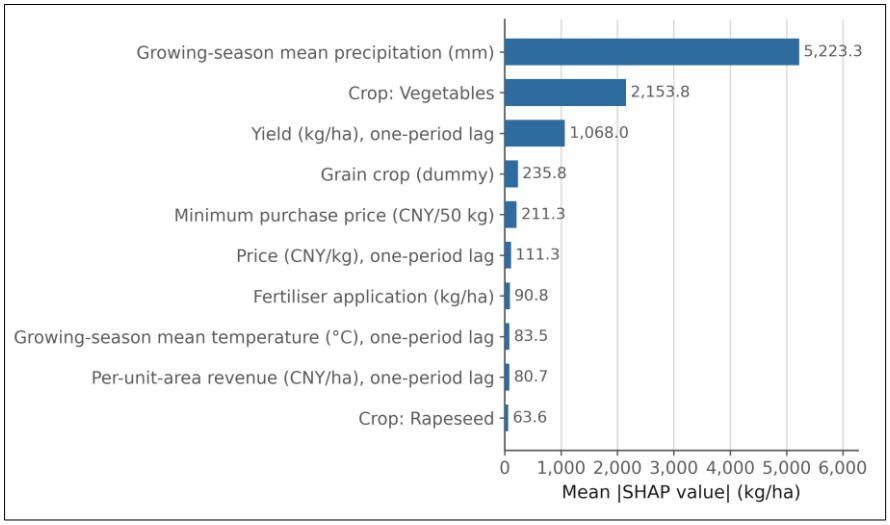



**Fig. 6: SHAP feature-importance ranking for the crop yield-prediction model (mean absolute SHAP value)**

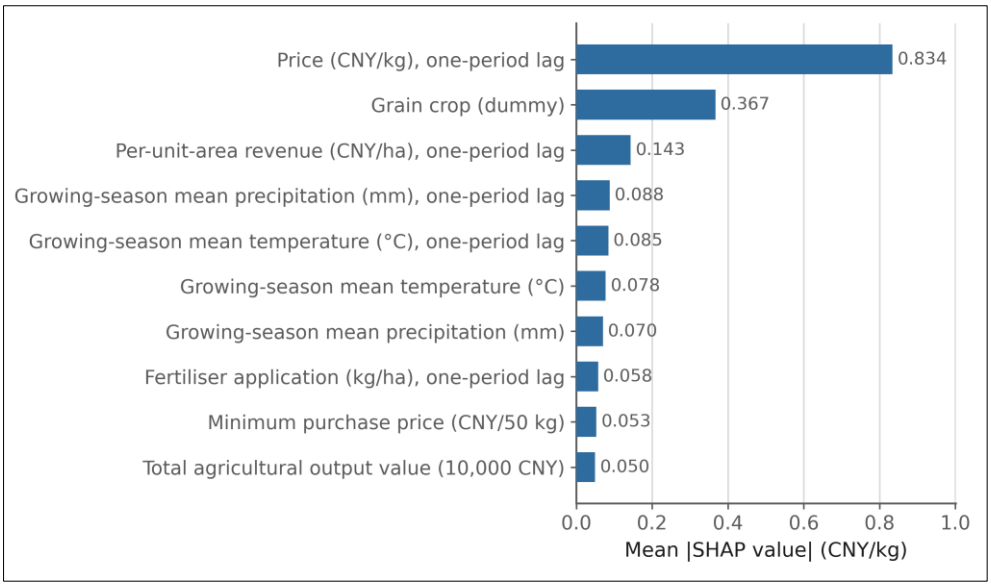



**Fig. 7: SHAP feature-importance ranking for the crop price-prediction model (mean absolute SHAP value)**

The attributions help explain the optimisation outcome. As cash crops, vegetables are far more exposed than grain crops to price swings in market supply and demand, yet their very high yield and the premium from market-based pricing earn them the largest allocation under revenue maximisation. That allocation is limited not by their own historical maximum, since removing that bound leaves it unchanged, but by the fixed total sown area together with the grain floor, which leaves them no further land; the price risk vegetables carry does not restrain their area. Maize's growing season (summer to autumn) does not coincide with Chongming's precipitation peak, so it is less sensitive to extreme rainfall than rice; as a grain crop its price is policy-supported and stable. Among the three grain crops, maize combines the mildest precipitation risk with the highest revenue per ha, so it is allocated up to its historical maximum. Wheat and rapeseed are both overwintering crops, complementary to rice in growing season and forming the agronomic basis of the rotation constraint; they earn the least per ha and would approach zero under revenue maximisation alone, so the lower bound holds them at a non-zero allocation within the rotation framework.

### 3.5 Sensitivity Analysis and the Cost of Food Security

Because the objective coefficients are point predictions from models whose out-of-sample skill is limited (see Prediction results), the linear programme was re-solved under systematic perturbations to establish which inputs determine the optimised structure (Fig-8).

Multiplying the vegetable price by factors from 0.6 to 1.4 (−40 % to +40 %) leaves the optimal allocation completely unchanged (vegetables 53.82 %, rice 35.99 %, maize 6.99 %, wheat 2.57 % and rapeseed 0.64 %), while total revenue ranges from 2.590 to 5.219 billion CNY; perturbing the vegetable yield over 0.7×–1.3× has the same effect. A linear programme whose optimum lies at a vertex defined by the crop-area bounds and the grain constraint behaves in exactly this way: as long as the ranking of per-unit-area revenues is preserved, the optimum cannot move. The prediction stage governs the level of the objective but not the structure of the solution. The one exception is maize: its 2024 price (5.00 CNY/kg) is both a doubling on the previous year's 2.60 CNY/kg and the eleven-year maximum, and the predicted 2025 value of 4.74 CNY/kg is what lifts maize to its historical maximum area. If the maize price reverted to 0.5× its predicted value, the optimum would shift to rice 41.06 % and maize 1.78 %, with total revenue falling only marginally, to 3.879 billion CNY. The maize-expansion recommendation rests on a single, unsustained price observation.

In contrast, the grain constraint governs the solution directly (Fig-8b): raising the floor reallocates area monotonically from vegetables to grain (Table-7). The floor the district itself sets is its annual grain target of 18,667 ha, which its 2024 structure already meets (18,703 ha, within 0.2 %). At that level the constraint does not bind: the revenue-maximising solution carries 19,146 ha of grain, 2.57 % above the target, at 4.295 billion CNY, the value it also takes when no grain floor is imposed. The district pays nothing for food security at its own stated target. Between the district's 2024 grain area (18,703 ha) and the 23,333 ha reference floor, the district forgoes 0.390 billion CNY, equal to 9.09 % of its maximum attainable revenue and to about 84,270 CNY per additional ha of grain; that figure is the marginal cost of grain held above the official target. Because the floor is a policy variable, the programme answers a conditional question, namely the cheapest crop mix that meets a given grain target, rather than an unconditional one about the best crop mix in itself.

**Table 7: Cost of the grain self-sufficiency floor (total sown area 51,238 ha). The district's official annual grain target (18,667 ha) is met by the 2024 actual level (18,703 ha) and is slack; the solution is identical at both, and the cost appears only above the official target.**

| Grain floor (ha) | Maximum revenue (billion CNY) | Vegetables (%) | Rice (%) |
|---|---|---|---|
| 18,703 (2024 actual) | 4.295 | 61.99 | 27.81 |
| 20,000 | 4.215 | 60.32 | 29.48 |
| 21,333 | 4.091 | 57.72 | 32.08 |
| 23,333 (reference) | 3.905 | 53.82 | 35.99 |
| 26,667 | 3.589 | 47.31 | 41.06 |
| 30,000 | 3.257 | 40.81 | 41.06 |
| 33,333 | 2.925 | 34.30 | 41.06 |

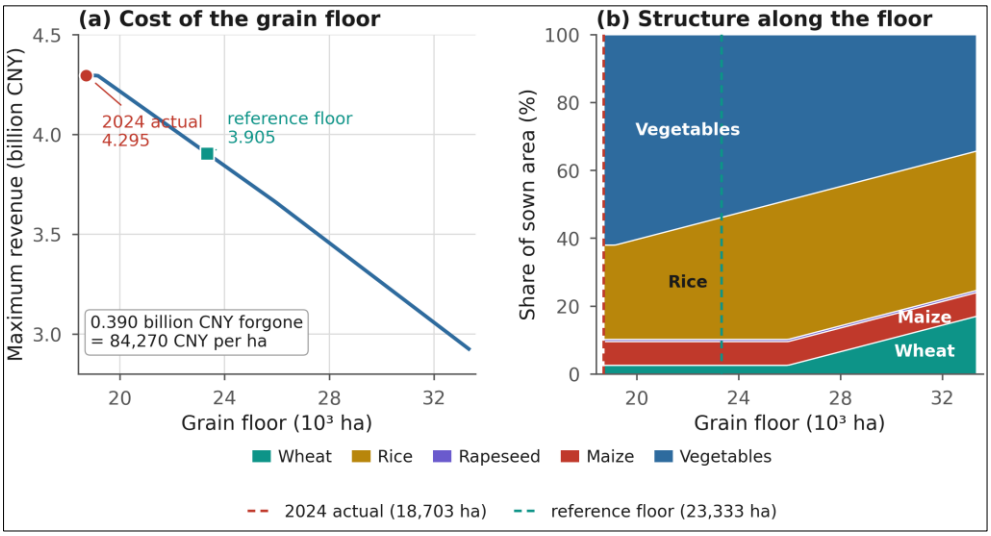



**Fig. 8: Cost of the grain self-sufficiency floor. (a) Maximum attainable revenue against the floor; (b) composition of the revenue-maximising structure**

### 3.6 Comparison with the Current Planting Structure

The optimised scheme is best compared with Chongming's current structure, not with an eleven-year average. Averaged over 2014–2024 the area shares were vegetables 53.14 % and rice 36.89 %, close to the optimised 53.82 % and 35.99 %; but the structure has since shifted (Fig-9), and in 2024 the actual shares were vegetables 57.23 %, rice 40.58 %, maize 1.42 %, wheat 0.64 % and rapeseed 0.14 %, with a grain area of 18,703 ha out of a total sown area of 43,870 ha for the five modelled crops. Measured against the current structure, the constrained optimum differs by −3.41 percentage points of vegetables, −4.59 of rice and +5.57 of maize.

The optimised structure is also not more profitable than the current one. On a common sown area of 51,238 ha, valued with the same predicted 2025 coefficients, the 2024 structure implies a grain area of 21,843 ha and would return 4.065 billion CNY (79,320 CNY per ha) against the optimised 3.905 billion (76,200 CNY per ha), so the constrained optimum is 3.93 % less productive per unit area. Two constraints account for this, and neither is evidence of allocative failure. Most of the gap comes from the grain floor: reaching 23,333 ha of grain requires converting land out of vegetables, and the district's own 2024 grain area (18,703 ha out of its actual 43,870 ha) lies 19.8 % below that floor, although it coincides with the district's official grain target of 18,667 ha to within 0.2 %, so the shortfall is measured against the above-target reference rather than against policy. The rest comes from the lower bound that protects overwintering crops, which holds wheat and rapeseed at a minimum of 1,644 ha. Evaluated at the district's own grain level (21,843 ha on a 51,238-ha basis) the LP optimum returns 4.043 billion CNY, 0.52 % below the district's actual structure, because the district has effectively abandoned winter cropping (wheat 280 ha and rapeseed 59 ha in 2024, against lower bounds of 1,315 and 329 ha) and places that land under vegetables instead.

Two further observations bear on this reading. The seasonal rotation constraint (wheat + rapeseed ≤ rice) is slack by a factor of eleven in the solution (1,644 against 18,438 ha), so it does not bind. What has changed over the decade is cropping intensity rather than the allocation of a fixed area: total sown area fell from 65,754 to 43,870 ha (−33 %), and the fall is concentrated in winter cropping. Wheat dropped from 8,806 to 280 ha and rapeseed from 1,346 to 59 ha, so the overwintering area shrank from 10,152 ha (50 % of the rice area) to 339 ha (2 %). Every crop's absolute area declined (vegetables −21 %, rice −12 %, maize −83 %); the rise in the vegetable and rice shares to 57.23 % and 40.58 % is an effect

of the shrinking denominator rather than evidence that the district has been converging on the revenue-maximising allocation.

The constraint that governs Chongming's planting structure is the administratively set grain-area target, which the district meets to within 0.2 % and which binds the revenue-maximising solution as soon as it is raised above the official level. The withdrawal of winter cropping is a second factor. Neither the target nor the withdrawal is a matter of price information; that trade-off is the one quantified in the sensitivity analysis above. The possible reasons for the residual difference from the unconstrained revenue maximum (historical inertia, machinery reconfiguration costs, the labour intensity of vegetables, and farmers' market-risk perception) remain relevant, but they bear on the gap between the current structure and an unconstrained optimum, not on the gap between the current structure and the grain-constrained optimum reported here.

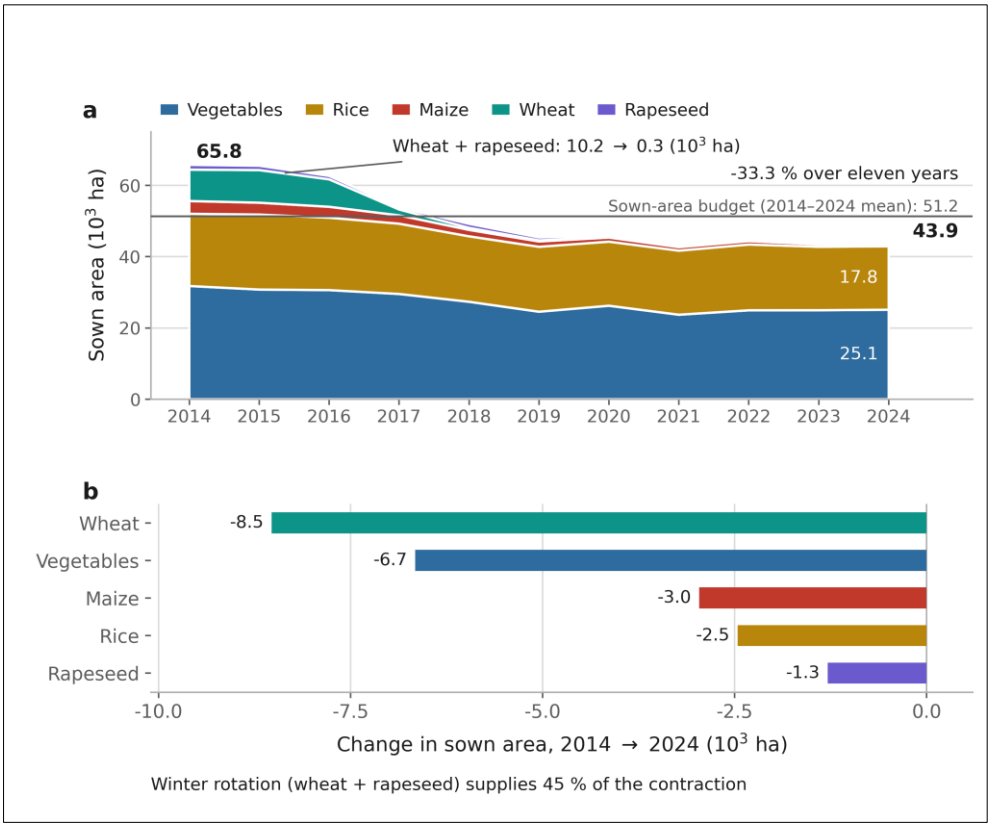



**Fig. 9: Structural change in the five modelled crops, Chongming District, 2014–2024. (a) Annual sown area by crop; (b) change in sown area between 2014 and 2024**

## 4. CONCLUSION

We set out to establish which constraint governs the crop planting structure of Chongming District. The central finding is that the market signal is not the binding constraint there. We built an integrated pipeline of XGBoost prediction, SHAP attribution and multi-constraint linear programming, applied it to five crops over 2014–2024, and benchmarked the prediction stage against naive alternatives. The yield model outperformed a per-crop mean benchmark for only one of the five crops, and the price model was beaten by a last-observation benchmark on every metric. Sensitivity analysis showed that the optimised structure is invariant to a ±40 % perturbation in vegetable price and a ±30 % perturbation in vegetable yield, and is governed instead by the grain-area floor. At the district's official grain target of 18,667 ha, which its 2024 structure already meets (18,703 ha), the constraint is slack and the cost of food security is nil. The 0.390 billion CNY, or approximately 84,270 CNY per additional ha of grain, is the marginal cost of grain held above the official target, between the district's 2024 grain area and the above-target reference floor. Relative to the district's actual 2024 structure, which sits at the official target but below the modelled reference floor, the constrained optimum is a lower-revenue solution: it is the price of grain self-sufficiency beyond the stated target rather than a gain from better allocation. In practice, then, a peri-urban district whose planting structure already sits close to the unconstrained revenue maximum should treat further structural adjustment as an explicit purchase of grain self-sufficiency, and the policy question becomes the level of that purchase, not the crop mix itself. Future work could replace the deterministic single-period formulation with stochastic or robust programming that propagates price uncertainty into the objective. Ecological objectives such as fertiliser reduction

and carbon emissions would be a natural extension, as would applying the framework to contrasting districts to test how the shape of the food-security cost curve depends on local resource endowments.

## DECLARATIONS

**Funding:** This research received no external funding.

**Competing Interests:** The authors declare no competing interests.

**Data Availability:** The annual crop panel assembled here, the XGBoost, SHAP and linear-programming code, and the scripts that generate the figures are provided as Supplementary Material S1. The panel was compiled from the Chongming Statistical Yearbook and from published district agricultural reports, which are publicly available; the compiled panel and code are also available from the corresponding author on reasonable request.

**Acknowledgements:** The authors thank the Editor and the anonymous reviewers for their constructive comments, which improved the manuscript.

## REFERENCES

1. Xie, W., Zhu, A., Ali, T., Zhang, Z., Chen, X., Wu, F., Huang, J., & Davis, K. F. (2023). Crop switching can enhance environmental sustainability and farmer incomes in China. *Nature*, 616(7956), 300–305. https://doi.org/10.1038/s41586-023-05799-x
2. Su, J., & Hong, Q. (2025). Balancing food production and farmer welfare: Assessing the impact of China's cropland de-grainization rectification policy. *Human Ecology*, 53(6), 1161–1174. https://doi.org/10.1007/s10745-025-00649-y
3. Dorneanu, B., Zhang, S., Ruan, H., Heshmat, M., Chen, R., Vassiliadis, V. S., & Arellano-Garcia, H. (2022). Big data and machine learning: A roadmap towards smart plants. *Frontiers of Engineering Management*, 9(4), 623–639. https://doi.org/10.1007/s42524-022-0218-0
4. Attri, I., Awasthi, L. K., & Sharma, T. P. (2023). Machine learning in agriculture: A review of crop management applications. *Multimedia Tools and Applications*, 83(5), 12875–12915. https://doi.org/10.1007/s11042-023-16105-2
5. Custodio, J. M., Billones, R. K., Concepcion, R., & Vicerra, R. R. (2024). Optimization of crop harvesting schedules and land allocation through linear programming. *Process Integration and Optimization for Sustainability*, 8(1), 123–134. https://doi.org/10.1007/s41660-023-00357-4
6. Pechibilski, N. W., Brandes, L. A., Knop, M. L., Ramos, F. M., & Cembranel, P. (2024). Production optimization through linear programming in agricultural properties. *Environment, Development and Sustainability*, 27(11), 26109–26127. https://doi.org/10.1007/s10668-024-04757-5
7. Veena, K., Pushpa, C. N., Thriveni, J., & Venugopal, K. R. (2024). Intercropping system optimization using Python-based goal and fuzzy goal programming. *SN Computer Science*, 5(8), 1120. https://doi.org/10.1007/s42979-024-03368-1
8. Liu, W., Fan, Y., & Yu, T. E. (2026). Optimizing crop planting structures for water savings and grain revenue in China considering different spatial scales. *Resources, Environment and Sustainability*, 24, 100323. https://doi.org/10.1016/j.resenv.2026.100323
9. Luo, C., Ou, D., Ma, H., Wu, K., Yao, X., Jing, S., & Xi, M. (2026). Optimization of the planting structure of major grain crops on cultivated land in China for coordinated food production, ecosystem service value, and irrigation water consumption. *Agriculture*, 16(16), 1711. https://doi.org/10.3390/agriculture16161711
10. Lokeshwari, M., Jha, G. K., Praveenkumar, A., Kumari, J., Navathe, S., Singh, G. P., & Prasad, P. V. V. (2026). A novel deep learning framework for field-scale wheat yield prediction. *Theoretical and Applied Genetics*, 139(3). https://doi.org/10.1007/s00122-026-05166-0
11. Saini, P., Nagpal, B., Garg, P., & Kumar, S. (2023). CNN-BI-LSTM-CYP: A deep learning approach for sugarcane yield prediction. *Sustainable Energy Technologies and Assessments*, 57, 103263. https://doi.org/10.1016/j.seta.2023.103263
12. Zhong, R., Zhu, Y., Wang, X., Li, H., Wang, B., You, F., Rodríguez, L. F., Huang, J., Ting, K. C., Ying, Y., & Lin, T. (2023). Detect and attribute the extreme maize yield losses based on spatio-temporal deep learning. *Fundamental Research*, 3(6), 951–959. https://doi.org/10.1016/j.fmre.2022.05.006
13. Aravind, K. S., Vashisth, A., Krishnan, P., Kundu, M., Prasad, S., Meena, M. C., Lama, A., Das, P., & Das, B. (2025). Development of multistage crop yield estimation model using machine learning and deep learning techniques. *International Journal of Biometeorology*, 69(2), 499–515. https://doi.org/10.1007/s00484-024-02829-9
14. Ma, L., Niu, W., Li, G., Du, Y., Sun, J., & Siddique, K. H. M. (2024). Crop yield prediction based on bacterial biomarkers and machine learning. *Journal of Soil Science and Plant Nutrition*, 24(2), 2798–2814. https://doi.org/10.1007/s42729-024-01705-0
15. Lei, J., Bai, D., Xu, Z., Li, L., Bai, X., & Zhou, W. (2026). Village-level multi-objective optimization of wheat and maize area constraints: Balancing irrigation water use, carbon intensity, and farm income. *Agricultural Systems*, 238, 104862. https://doi.org/10.1016/j.agsy.2026.104862

16. Zhang, Y. W., McCarl, B. A., Cao, Z., Li, Q., Yang, S., & Liu, H. (2022). Operating pesticide use reduction within the boundary of food security in peri-urban settings. *Fundamental Research*, 2(4), 635–647. https://doi.org/10.1016/j.fmre.2022.04.003
17. Liu, Y., Zhang, J., Wang, J., Hu, Y., & Hu, N. (2026). The impact of the farmland protection policy on the adjustment of grain planting structure: Evidence in China. *Land*, 15(3), 425. https://doi.org/10.3390/land15030425
18. Li, D., & Wu, X. (2025). Understanding the structural imbalance in non-grain land utilization: Insights from China's arable land policy. *Applied Geography*, 181, 103673. https://doi.org/10.1016/j.apgeog.2025.103673
19. Lu, D., Wang, Z., Su, K., Zhou, Y., Li, X., & Lin, A. (2024). Understanding the impact of cultivated land-use changes on China's grain production potential and policy implications: A perspective of non-agriculturalization, non-grainization, and marginalization. *Journal of Cleaner Production*, 436, 140647. https://doi.org/10.1016/j.jclepro.2024.140647
20. Elmachtoub, A. N., & Grigas, P. (2022). Smart "predict, then optimize". *Management Science*, 68(1), 9–26. https://doi.org/10.1287/mnsc.2020.3922
21. Bertsimas, D., & Kallus, N. (2020). From predictive to prescriptive analytics. *Management Science*, 66(3), 1025–1044. https://doi.org/10.1287/mnsc.2018.3253
22. Mandi, J., Kotary, J., Berden, S., Mulamba, M., Bucarey, V., Guns, T., & Fioretto, F. (2024). Decision-focused learning: Foundations, state of the art, benchmark and future opportunities. *Journal of Artificial Intelligence Research*, 80, 1623–1701. https://doi.org/10.1613/jair.1.15320
23. Han, J., Sun, Y., Lin, N., Liu, X., & Ruan, J. (2026). A predict-then-optimize framework for multi-stage agricultural production and distribution planning. *Applied Soft Computing*, 200, 115224. https://doi.org/10.1016/j.asoc.2026.115224
24. Slater, Y., Finkelshtain, I., Troen, A. M., & Kan, I. (2025). Economic analysis of food self-sufficiency: Modeling and application to the case of Israel. *Food Policy*, 136, 102979. https://doi.org/10.1016/j.foodpol.2025.102979
25. Meroni, M., Waldner, F., Seguini, L., Kerdiles, H., & Rembold, F. (2021). Yield forecasting with machine learning and small data: What gains for grains? *Agricultural and Forest Meteorology*, 308–309, 108555. https://doi.org/10.1016/j.agrformet.2021.108555
26. Kopishynska, O., Fedorchenko, M., Utkin, Y., Sliusar, I., Liashenko, V., Svitlychna, A., & Pysarenko, S. (2026). Evaluating machine learning and baseline methods for crop yield forecasting using small datasets. *Environmental Research Communications*, 8(8), 085008. https://doi.org/10.1088/2515-7620/ae91bb
27. Mahajan, P., Muley, A. A., & Fegade, M. R. (2026). Forecasting agricultural commodity prices using machine learning models. *Discover Informatics*, 1(1), 15. https://doi.org/10.1007/s44564-026-00015-0
28. Kadir, D. H., Khalil, D. M., & Khudhur, A. M. (2026). Forecasting multivariate time series: A comparison of machine learning, statistical and deep learning models. *Forecasting*, 8(4), 73. https://doi.org/10.3390/forecast8040073
29. Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, 4(9), 100804. https://doi.org/10.1016/j.patter.2023.100804
30. Jang, H. J., Ng, W., Chen, S., Setia, R., & Minasny, B. (2026). Illusions of accuracy: Widespread data leakage distorts machine learning results in biochar research. *GCB Bioenergy*, 18(9), e70173. https://doi.org/10.1111/gcbb.70173
31. Zhu, J., Yang, M., & Ren, Z. J. (2023). Machine learning in environmental research: Common pitfalls and best practices. *Environmental Science & Technology*, 57(46), 17671–17689. https://doi.org/10.1021/acs.est.3c00026
32. Verdinelli, I., & Wasserman, L. (2024). Feature importance: A closer look at Shapley values and LOCO. *Statistical Science*, 39(4), 623–636. https://doi.org/10.1214/24-STS937
33. Moharana, P. C., Yadav, B., Malav, L. C., Biswas, H., & Patil, N. G. (2025). Machine learning-based crop suitability prediction: An emerging technique for sustainable agricultural production in the desert region of India. *Communications in Soil Science and Plant Analysis*, 56(3), 376–395. https://doi.org/10.1080/00103624.2024.2419994
34. Wu, H., Zhang, Z., & Dias, D. (2024). Prediction on compression indicators of clay soils using XGBoost with Bayesian optimization. *Journal of Central South University*, 31(11), 3914–3929. https://doi.org/10.1007/s11771-024-5681-9
35. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp.785–794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785
36. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30* (pp. 4765–4774). Curran Associates.