

Multi-source Data-driven Cause Identification and Risk Early Warning for Unsaleable Agricultural Products

Liu Zhanhong¹, Chen Mengfei¹, Wang Zhanyu¹, Zhang Jiahang¹, Li Mingxu²

¹*School of Artificial Intelligence and Big Data, Henan University of Technology, Zhengzhou, Henan, China*

²*IFLYTEK Co., Ltd., Hefei, Anhui, China*

Abstract: Farm products may not be sold due to a mismatch in harvesting time, buyer demand, transportation, information and market organization; excessive output is not the only reason for this problem. We take the monthly data of eleven products in nine Henan cities from 2019 to 2025 and use a city-product-month working unit. Analyze the construction of the seven-factor risk profile, derive a ten-variable index with entropy weighting, attribute observations to five relative risk groups, and examine five categories of causes. The Cost Structure is relatively rigid. Chinese cabbage has a high cost risk and resulted in a loss in 2023 and 2024; this combination also occurred for pork in 2023. Maize was profitable and low-risk in 2023, but it became loss-making and high-risk the following year. Diesel and road-freight indicators were relatively high in 2022, but there were still some fluctuations month-to-month. Online attention is also uneven, and chili has 73 anomalies and is the most numerous in the sample. City results do not have a single channel story. Shangqiu has high e-commerce penetration and a relatively high reported frequency; Sanmenxia also has high penetration but a much lower frequency. Only a few product-year cases of the cost signal have been identified: Chinese cabbage in 2023-2024, pork in 2023, and maize in 2024. Freight pressure was highest in 2022. The two cities do not have the same coverage; although both are highly online, the frequency of reported cases in Shangqiu and Sanmenxia is different. Therefore, these records were selected for inspection but not included in the closed-won. Search anomalies are used to identify which records need to be verified. They are not considered evidence of a transaction or that a product has gone out of stock. Therefore, the corresponding response will be taken to address the

bottleneck in the data. A cost measure is suitable for one product, organisation of transportation needs to take place before a sensitive marketing window, and an online signal should be compared with market data prior to action. Frequent Combinations can help to find cases for investigation, but they do not establish causality or predict a particular person's behaviour.

Keywords: Unsaleable Agricultural Products; Causes of Unsaleable Surplus; Multi-Source Data; Risk Early Warning; E-Commerce-Enabled Agriculture; Henan Province.

1. Introduction

1.1 Research Background

Farm production supports food security and rural income, but the production decision occurs long before the sale. Seed, fertiliser, land and labour are committed months in advance; purchases react later to income changes, shifts in demand, holidays, substitution by other products, etc. When production, demand and distribution are out of sync, prices will drop, the sales cycle will be extended, and inventory will build up. The slow adjustment of agricultural supply disturbs market equilibrium [1]. Fresh food has a short shelf life and is easily damaged during transport and storage. Cold-chain coordination can extend the sales period and reach remote consumers [2, 3]. E-commerce is another distribution channel for logistics, and grading and production-sales coordination have been established [4]. Therefore, we take unsaleability to be the sum of frictions in cost, logistics, demand, policy, information and sales channels [1-6]. Henan is a large area of agriculture and a distribution centre; thus, the sampled cities have different operating conditions. Industrial structure, purchasing power, logistics, e-commerce activity and policy capacity differ

among the nine cities. Separate these places and products to keep the local differences in the analysis, and do not include them in the provincial average.

1.2 Research Questions and Objectives

Unsalability refers to a mismatch among the production, demand, circulation, information and institutions in this study. First, we will determine the extent of pressure on growers from high production costs and when freight and cold chain problems start to occur. Whether the demand for consumption or policy support can absorb the supply is still unknown, not assumed. We will also add search and public opinion records to the ordinary market data. A narrower question refers to a reduction in the regional mismatch of an online channel, rather than an expansion [3-6].

The foundation is one city, one product, and one month. Combine cost, circulation, demand, policy, public opinion and e-commerce data in a seven-dimensional fact table, then use ten entropy-weighted indicators to calculate a warning score. The items with high scores are shown in the dashboard. The five levels are ranks of the sample and not probabilities of a future unsold event.

1.3 Research Content and Technical Route

The workflow is in the order that the evidence is obtained. First, build and verify the source tables; then, assess the risk and check for supporting evidence; finally, only after completing these checks will recommendations be made. Table 1 is the four-stage process.

First, collect the source files and align the dates, locations, product names and indicator definitions. The aligned records are passed to ETL, and then the entropy-weighted index is computed and assigned to the five relative classes. Then we will examine the visual evidence and link a recommendation to a product, time and place where it can be used. Since the source files do not have the same unit or time interval, they are not statistically comparable. Standardisation provides a common scale for the indicators, and the entropy weights are obtained by calculating the variation in the sample. Figure 1 is the obtained route.

Table 1. Research Modules and Contents

Module	Main content
Data collection and	Start with the cost, circulation, demand, policy, public-opinion, and e-commerce files; align their dates, locations,

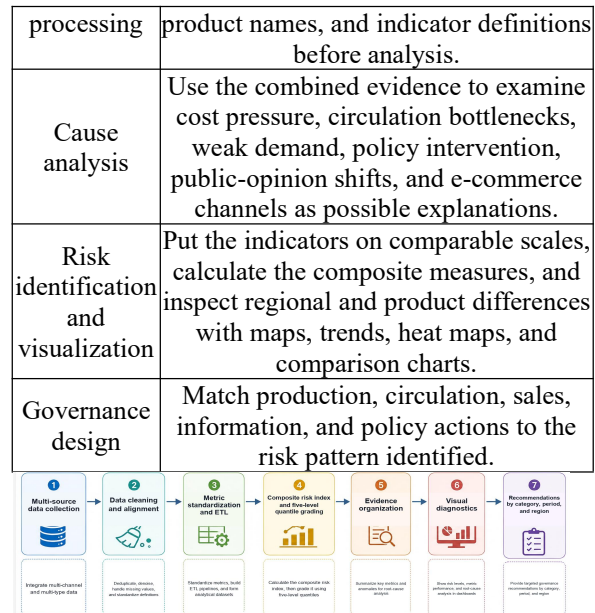


Figure 1. Technical route

Source: Compiled from the project dashboard.

1.4 Contributions and Research Boundaries

The study presents the following reasons for production costs, circulation, consumer demand, policies, public opinion and e-commerce. The sources of the evidence are generally examined separately: official statistics, cost-benefit records, high-frequency logistics data, search activity, public opinion and policy documents, and e-commerce indicators. The dashboard will show the user how to find the supporting materials for a high score and who was involved in the response.

The Dashboard is to filter and display data. A high search index, an abnormal public opinion record or a composite score show that the analyst should direct their attention to another place; it does not indicate how many units of produce were unsold. Sales, orders, inventory, and confirmed event labels are needed for this. Given the above results, panel, time-series, machine-learning or causal methods can be employed to investigate the proposed mechanisms.

2. Research Design and Data Foundation

2.1 Study Area, Sample and Analytical Framework

The sample includes Henan Province and the nine cities: Zhengzhou, Luoyang, Nanyang, Zhoukou, Shangqiu, Xinyang, Anyang, Puyang, and Sanmenxia. They are consumer centres, large-scale production areas and cities with a

relatively balanced production-consumption structure. The 11 products in the list are: Chinese cabbage, cucumber, tomato, eggplant, chili, garlic, wheat, maize, watermelon, apple and pork. Records from January 2019 to December 2025; the main risk calculation is based on January 2021 to December 2025.

The two classifications are in connection. The fact table includes price, supply, weather, cost,

circulation, demand and public-opinion risk. The Business Dashboard shows the diagnosis of costs, logistics, market, channels and public opinion. Weather is present in the fact table but not included in the root-cause categories. Price, supply and demand are observed in a market; e-commerce provides channel data. Figure 2 is the view connection diagram.

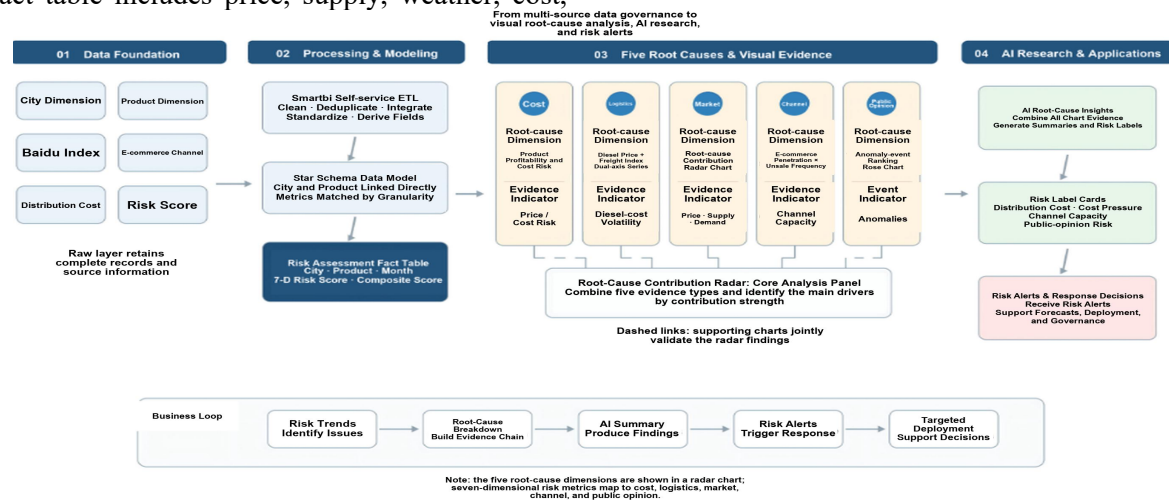


Figure 2. Overall Research Framework and Root-Cause Evidence Chain

Source: Copied from the project dashboard.

Table 2 is the crosswalk of each root-cause dimension and the data used for presentation.

2.2 Data Sources and Variables

The source layer has 34,810 records in the seven thematic tables. We will keep these records in

their original form for filtering, aggregation and joining of the input data. The Analysis Layer sets the study period, groups the records according to the grain requirements of the model, and joins the resulting tables. Table 3 shows the data sources and variables.

Table 2. Relationship between Root-Cause Dimensions and Underlying Data

Root-cause dimension	Underlying data	Processing
Cost	Cost risk score, planting cost, product profitability	Direct selection combined with cost structure and profit/loss status
Logistics	Circulation risk score, freight index, diesel price	Rename circulation as logistics and add time-series indicators
Market	Price, supply, and demand risk scores	Aggregate by market-business theme
Channel	E-commerce penetration, online retail indicators	Supplement with e-commerce channel data
Public opinion	Public-opinion risk score, Baidu anomaly records	Direct selection plus anomaly frequency
Weather (not displayed)	Weather risk score	Retained in the fact table; excluded from dashboard root-cause charts

Table 3 Data Sources and Variable System

Data table	Records	Grain	Core variables	Main source
City dimension	9	Static	City role, population, urbanization, wholesale markets, coordinates	Henan-focused land-use and socioeconomic literature [7]
Product dimension	11	Static	Category, storability, main season, price-volatility ranking	Agricultural price-volatility literature [8]
Baidu Index	28, 127	Daily	Search index, 30-day moving average, Z-score	Online-search and agricultural price-forecasting literature [9]

E-commerce channel	45	Annual	Online retail, e-commerce penetration, express coverage	Agricultural e-commerce literature [10]
Circulation cost	623	Weekly/monthly /adjustment date	Freight index, changes, diesel price	Road-freight pricing and agricultural-cost literature [11,12]
Planting cost	55	Annual	Output value, total cost, cost structure, net profit, cost-profit ratio	Agricultural production-cost literature [12]
Risk fact table	5,940	Monthly	Seven risk scores, composite score, level, top drivers	Calculated in this study

2.2.1 City and Product Dimensions

The city name is used as the key for the city table, and functional role, population, urbanization, wholesale-market conditions and

WGS84 coordinates are also stored here. The attributes of the product table are category, storage life, peak season, production cycle, main -producing area and price history. For the last item, we calculate the coefficient of variation (CV) of weekly prices from 2019 to 2024:

$$CV = \sigma / \mu \quad (1)$$

σ is the standard deviation of price and μ is the mean price in equation (1). Therefore, the coefficient will be relatively larger in the case of a large dispersion in prices. We read the storage life and the main selling season of the same short-lived surplus, as it is difficult to absorb when storage space is limited.

2.2.2 Baidu Index and E-commerce Channel Data

Each of the eleven products in the Baidu series has a keyword. It supplies 28,127 daily observations for 2019-2025. First, a 30-day moving average is calculated for each keyword, and then this observation is compared with the previous distribution of the same keyword using a Z-score:

$$Z_t = (X_t - \mu) / \sigma \quad (2)$$

If the value is greater than 2, it will be added to the review; otherwise, $Z > 3$ for the extreme_anomaly flag in the ETL output. The first is to detect motion; the second will trigger an alarm if motion is detected. The e-commerce table shows online retail, penetration and last-mile delivery for the nine cities over five years. As there is no official series of city-level agricultural online-retail data in this dataset, its proportion has been estimated through linear interpolation of public benchmarks. It is suitable for comparison in this sample, and an absolute city statistic would require an official series [10].

2.2.3 Circulation, planting cost, and risk-score data

The circulation-cost file contains a weekly road freight index, monthly freight data and dated diesel-price adjustments. As these series have different base periods, they will be considered individual scales and interpreted independently. The planting-cost table has 55 product-year records from 2021 to 2025 and includes output value, total cost, material and service inputs, labour and imputed family labour, land, net profit, and the cost-profit ratio. An additional 5,940 city-product-month records are in the risk fact table, and this table has seven component risks, a composite score and class, and the three leading drivers.

2.3 Data Preprocessing and ETL

SmartBI receives the standardised data records from the self-service ETL flow, deals with missing values, eliminates duplicates, separates date components, and calculates derived indicators. The original rows and transformation path remain available. Static dimensions, time series and model outputs are processed according to different rules, and thresholds and formulas are recorded. City and product keys are validated before incremental loading. Table 4 shows the treatments.

Table 4. ETL Rules for the Seven Data Tables

Data table	Key processing	Loading
City dimension	Select columns, remove invalid keys, deduplicate by city key	Validated incremental append
Product dimension	Select columns, remove invalid keys, deduplicate by product key	Validated incremental append
Baidu Index	Keep rows from 2021 onward; derive year and month, flag $Z > 3$ anomalies, and create record IDs.	Full overwrite
E-commerce channel	Standardize field types, remove duplicate keys, and sort by year and city.	Insert/update

Circulation cost	Clean numeric fields, remove duplicate dates, branch by time grain, and derive changes.	Full overwrite
Planting cost	Clean key cost fields, deduplicate, remove anomalies; derive cost shares, profitability and cost risk	Full overwrite
Risk fact table	Unpivot seven risk columns, remove nulls and duplicates	Insert/update

Baidu rows from 2021 onwards are kept; year and month are taken from the date. Circulation Rows keep the original frequency; weekly data contain freight changes and have a year/month field; annual data store the year, and daily adjustment data save event attributes. Reshape the risk columns into city-product-month-dimension-value rows for the dashboard.

Cleanse planting-cost records, deduplicate, selectively impute and align by key. Then the

cleaned data are used to calculate cost shares, profitability and cost-risk classes. If the city-product price is not available, it will be assigned a price-risk score of 0.5 and a missing-data flag, and no quote will be interpolated. Min-Max Normalisation is performed before loading into SmartBI, so it is not necessary in the platform's ETL. Indicators that are larger in value for a higher risk:

$$z_j = (x_j - x_{j, \min}) / (x_{j, \max} - x_{j, \min}) \quad (3)$$

A relatively high number shows a low risk; therefore, the standardised value is inverted as follows:

$$z_j = (x_{j, \max} - x_j) / (x_{j, \max} - x_{j, \min}) \quad (4)$$

If the maximum and minimum values of the indicator are the same, then a standardised value of 0 will be assigned to the batch because there is no difference among the observations. For a batch of processing, the range, time window and direction of the risk are kept constant. Figure 3 is the whole ETL process.

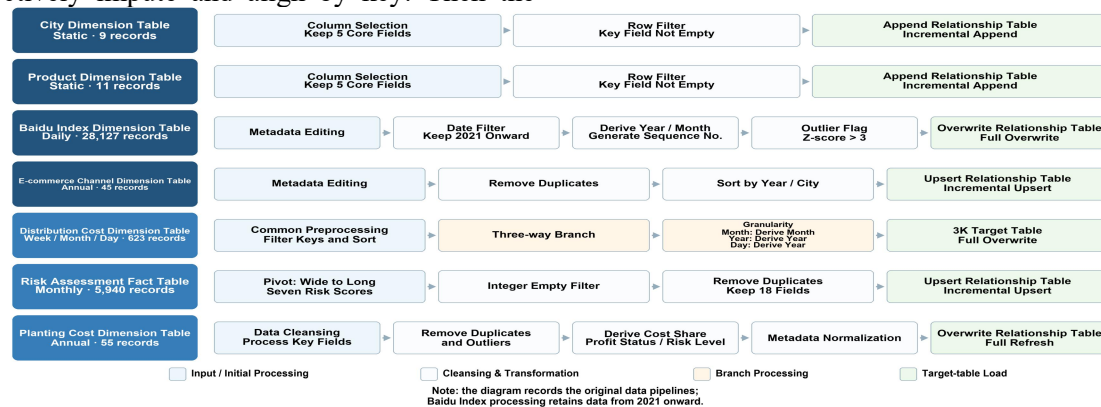


Figure 3. ETL Process for the Seven Data Tables

Source: Copied from the project dashboard.

2.4 Risk Measurement, Grading and Quality Control

All seven fact-table dimensions are in the range [0, 1], are combined, and then used to calculate the weighted score shown in the risk-profile dashboard:

$$R = w_P P + w_S S + w_W W + w_C C + w_L L + w_D D + w_O O \quad (5)$$

The symbols are P, S, W, C, L, D and O for price, supply, weather, cost, circulation, demand and public-opinion risk, respectively. This seven-point score is not the same as the five-category root-cause diagram. Weights were given according to the conceptual importance of the factors and the team's project experience; however, an expert survey or regression analysis of historical data would still need to be conducted to verify them. Table 5 is the Grading

Rule.

The quality check field does not specify the source of the data it depends on. Official statistics and authoritative platforms provide most city, product, Baidu, circulation, and planting-cost records. Some e-commerce values are estimates, and the risk table is a model output; both facts have been identified. The 100% complete rate only includes the required fields in the final tables, and population coverage is not included in this calculation. Some of the 2021 indicators are also missing data from the previous year, so setting them at 0.5 will reduce the impact of the first year.

Table 5. Composite Risk Grading

Level	Risk-score interval	Interpretation
Low risk	[0, 0.2)	Routine monitoring
Moderately	[0.2, 0.4)	Continuous

low risk		observation
Medium risk	[0.4, 0.6)	Monitor trend
Moderately high risk	[0.6, 0.8)	Early warning recommended
High risk	[0.8, 1.0]	Priority verification and response

2.5 Data Model and Platform Implementation

At the centre of the extended Star Schema is a city-product-month risk fact table, and city and product are master dimensions. Baidu, e-commerce, circulation-cost and planting-cost data maintain their original collection frequencies. Product keywords link searches to the product dimension; e-commerce joins by city and year, and planting cost by product and year. Generally, the circulation cost is at the provincial level, and assigning it to individual cities would introduce a degree of detail that is not supported by the data. Retain the native grains to prevent the false precision and retain all seven risk dimensions.

Smartbi links records at annual, monthly, and daily frequencies and lets users filter by city, product, date, and risk dimension. The second dashboard page groups the evidence as cost, logistics, market, channel, and public opinion. Each visual states its grain, time span, and whether an estimate is involved. Driver fields connect a change in the composite score to the component measures. Figure 4 shows the table links.

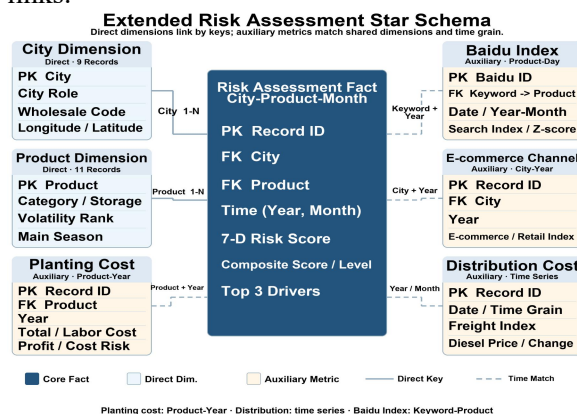


Figure 4. Extended Star-Like Data Model and Table Relationships

Source: Compiled from the project dashboard.

3. Agricultural Product Unsaleability Risk Early-Warning Model

The three steps in the index calculation are: selection of indicators, acquisition of their entropy weights, and classification of scores according to their position in the sample. Figure 5 shows the sequence.



Figure 5. Risk Early-Warning Model Construction Process

Source: Compiled from the project dashboard.

3.1 Indicator System

Ten indicators enter the warning score. They are demand absorption, e-commerce sales capacity, production returns, circulation costs, and the market and policy environment. Table 6 shows the definitions and risk directions of all indicators.

The sign of the indicator is used to change the score. A larger value lowers risk for a protective measure and raises risk for a pressure measure. Demand variables approximate potential absorption; E-commerce variables describe online reach and delivery support. Production-return variables indicate whether revenue exceeds costs, and circulation variables are the costs of transporting goods between areas. Search and policy counts are indirect: a search may occur before or after a purchase or news of unsold output, and a policy can be introduced at any time. We will keep them for reference. Comparable measures of cold-chain coverage and current inventory were not available.

Table 6. Risk Early-Warning Indicator System

Dimension	Indicator	Meaning	Risk direction
Demand absorption capacity	Urban residents' per-capita consumption	Purchasing power of the target market	Lower

	expenditure		
Demand absorption capacity	Online retail sales	Scale of online absorption capacity	Lower
E-commerce sales capacity	E-commerce penetration	Coverage of digital sales channels	Lower
E-commerce sales capacity	Express delivery volume	Logistics support for online sales	Lower
Production-return pressure	Planting cost per unit area	Input intensity at the production stage	Higher
Production-return pressure	Cost-profit ratio	Ability of returns to cover costs	Lower
Circulation-cost pressure	Road freight index	Burden of interregional transportation	Higher
Circulation-cost pressure	Diesel price	Fuel-cost transmission pressure	Higher
Market environment and support	Baidu search index	Change in online market attention	Lower
Market environment and support	Policy-event frequency	External support from government intervention and production-sales coordination	Lower

3.2 Entropy Weighting and Index Calculation

The units of the indicators are different; therefore, they need to be standardised before entropy calculation. The weighted entropy method limits subjective judgment, but the resulting weights still depend on the current distribution of the sample [13, 14]. For observation i and indicator j , we first calculate the share of the indicator in the total:

The standardised value of observation i is denoted as z_{ij} . Its share p_{ij} is calculated as follows:

$$p_{ij} = (z_{ij} + \varepsilon) / \sum_i (z_{ij} + \varepsilon) \quad (6)$$

A small positive epsilon is added to prevent the logarithm of zero. The entropy of indicator j is then calculated as:

$$e_j = -k \sum_i p_{ij} \ln(p_{ij}), k = 1/\ln(m) \quad (7)$$

The remaining information is shown as the divergence d_j ; normalising these divergences gives the weights w_j :

$$d_j = 1 - e_j \quad (8)$$

$$w_j = d_j / \sum_j d_j \quad (9)$$

Here m is the number of observations. The normalised weights sum to 1, and their estimates are given in Table 7.

Table 7. Entropy-Weight Calculation Results

Dimension	Indicator	Entropy	Divergence	Weight
Demand absorption	Per-capita consumption expenditure	0.9715	0.0285	0.0916
Demand absorption	Online retail sales	0.9933	0.0067	0.0214
E-commerce sales	E-commerce penetration	0.9930	0.0070	0.0224
E-commerce sales	Express delivery volume	0.9499	0.0501	0.1608
Production-return pressure	Planting cost per unit area	0.9464	0.0536	0.1721
Production-return pressure	Cost-profit ratio	0.9827	0.0173	0.0555
Circulation-cost pressure	Road freight index	0.9554	0.0446	0.1433
Circulation-cost pressure	Diesel price	0.9448	0.0552	0.1771
Market environment and support	Baidu search index	0.9926	0.0074	0.0239
Market environment and support	Policy-event frequency	0.9589	0.0411	0.1320

The resulting weights are: diesel price at 0.1771, planting cost per unit area at 0.1721, and express delivery volume at 0.1608. Together, they make up 51.0% of the whole. This is an expression of separation in the current sample, not one of causal effect. Online retail sales, e-commerce penetration and the Baidu search index are almost identical in these records, so they are assigned a lower weight. The Warning Index is shown below:

$$E_i = \sum_j w_j z_{ij} \quad (10)$$

A large E_i indicates that the observation has a relatively high combined pressure of demand, sales channels, production cost, circulation and the market environment.

3.3 Relative Risk Grading

Quartiles are the 20th, 40th, 60th and 80th percentiles. The five groups of observations were divided approximately equally, and they were: low, relatively low, medium, relatively high, and high. These labels locate an observation in the current dataset and do not assign a probability to an unsaleable event. Add

a new observation, then move the boundary; otherwise, it will be difficult to compare with the previous period.

4. Empirical Results and Visual Analysis

The results are at the level of product, year, city and date. First, observe that there is a relatively large difference, then determine which dimension is related to it, and check whether the underlying series move in the same direction. As the above data are generally distributions or changes over time, the language of discussion uses association and warning. Causal interpretation would need a reasonable Design and control.

4.1 Overall Root-Cause Structure

Figure 6 is the five scores—supply, weather, cost, circulation and public opinion—of the eleven products. The profiles are not uniform; some products are at the ends of the cost, circulation or public opinion ranges, while others fall in the middle of the observed range. A chart is a way to choose the next evidence of observation, but it does not show which factor led to an unsaleable result.

More than one pressure can exist for a product, and the same pressure does not need to have the same meaning in different products. Compare the radar profile with the cost record, public opinion ranking, channel scatter plot and diesel and freight series. Radius and Shaded Area are indices of relative risk. Neither one gives unsold tonnage or monetary loss.

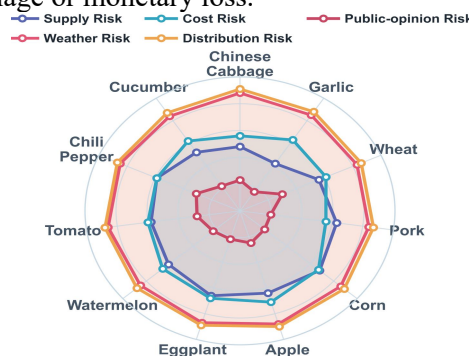


Figure 6. Five-Dimensional Root-Cause Contribution of Agricultural Products

Source: Copied from the project dashboard.

4.2 Production Cost and Product Profitability

When the selling price is low, production costs limit how much a producer can reduce costs. Expenditures for labour, land, materials and operations have filled the room. Most product-year observations in Figure 7 remain profitable

and are placed in the low cost-risk class. Chinese cabbage in 2023 and 2024, and pork in 2023, are exceptions; each has a high cost risk and resulted in a loss. Maize is moving in the same direction; it was profitable and low-risk in 2023 but has now become unprofitable and high-risk in 2024.

Cost risk and profitability by product, 2021-2025 (ascending year)

Product	2021 Cost Risk	2021 Profitability	2022 Cost Risk	2022 Profitability	2023 Cost Risk	2023 Profitability	2024 Cost Risk	2024 Profitability	2025 Cost Risk	2025 Profitability
Chinese Cabbage	Low	Profit	High	Loss	High	Loss	High	Loss	High	Loss
Garlic	Medium	Profit	Low	Profit	Low	Profit	Low	Profit	Medium	Profit
Wheat	Medium	Profit	Low	Profit	High	Loss	Medium	Profit	High	Loss
Pork	Medium	Profit	Medium	Profit	High	Loss	Medium	Profit	Medium	Profit
Corn	Medium	Profit	Low	Profit	Low	Profit	High	Loss	Medium	Profit
Apple	Low	Profit	Low	Profit	Low	Profit	Low	Profit	Low	Profit
Eggplant	Low	Profit	Low	Profit	Low	Profit	Low	Profit	Low	Profit
Watermelon	Low	Profit	Low	Profit	Low	Profit	Low	Profit	Low	Profit
Tomato	Medium	Profit	Low	Profit	Low	Profit	Medium	Profit	Medium	Profit
Chili Pepper	Low	Profit	Low	Profit	Low	Profit	Low	Profit	Low	Profit
Cucumber	Low	Profit	Low	Profit	Low	Profit	Low	Profit	Low	Profit

Figure 7. Recent Profitability Status and Cost Risk by Product

Source: Copied from the project dashboard.

Figure 8 shows the different cost structures and is not uniform. Land is a large cost for pork, maize and wheat. Material spending moves more sharply over time and is high for pork and Chinese cabbage in several years. Chinese cabbage also has a higher family and total labour cost than the less labour-intensive products. Profit falls or trends downwards in many of the 2021-2025 observations, and Chinese cabbage and pork reach low or negative values in some years. The figures are close; based on the data, it cannot be ruled out that the cause of the loss is only cost risk. Output and purchase price of raw materials also affect the profit.

4.3 Public-Opinion Anomalies

The public opinion data in Figure 9 reflect the concentration of market interest and abnormal information. Most of the abnormalities occur in a few places. Chili has 73 events, tomato 57, wheat 43, and Chinese cabbage 41. Maize and apple have 24 and 21. Cucumber, eggplant, watermelon and garlic have only a few, and pork has none in the sample. Therefore, most of the attention signals are concentrated in a small number of products.

The count can identify the products to be promoted; however, it does not display unsold stock. Figure 10 adds the time and amplitude of the signals. The number will be affected by how widely the news is spread and other reasons. Apple, wheat, chili and some other products exceed 3 in some months from 2021 to 2025. The above dates need to be verified with news,

prices, sales, inventory and field reports before being used in an operational warning.

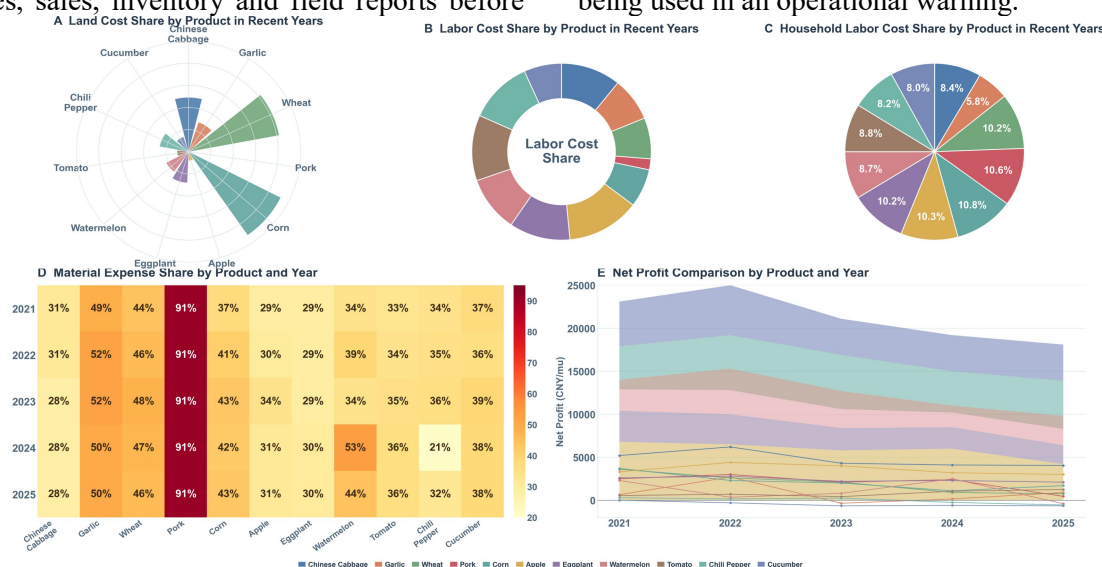


Figure 8. Planting-Cost Structure and Net-Profit Changes

Source: Compiled from the project dashboard

Anomaly Event Ranking (2021-2025)

Anomaly definition: Search Index Z-score > 3; petal radius represents the number of anomaly events

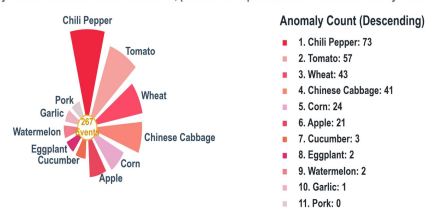


Figure 9. Ranking of Public-Opinion Anomaly Events by Product

Source: Compiled from the project dashboard.

4.4 E-Commerce Capacity and City-Level Differences

Online sales can connect producers with buyers in other places. Only when fulfilment, branding, grading, refrigeration and production-sales coordination work together will that connection be effective. Figure 11 shows the results; although some of the sampled cities have a high penetration rate, their reported unsaleable frequencies are not low. Both are relatively high in Shangqiu. Sanmenxia has good coverage and is relatively infrequently used.

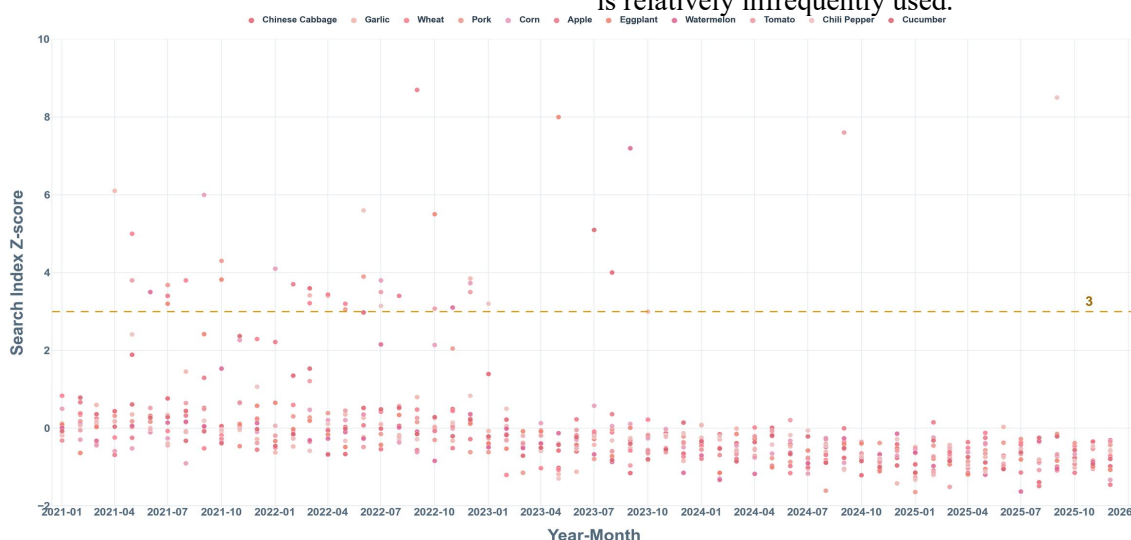


Figure 10. Public-Opinion Anomaly Warning Over Time

Source: Compiled from the project dashboard. The dashed line at 3 is the alarm reference value.

A scatter plot of the unusual city-level combinations does not show that a higher e-commerce penetration is related to lower unsaleability. A wider scope will gain more attention and may also attract more attention to

the local problem. The channel will still fail during the peak harvest due to poor grading or cold chain facilities. With nine cities and no control over production and distribution, a net penetration effect cannot be realised via the plot.



Figure 11. E-Commerce Penetration and Unsellable Frequency in Sampled Cities
Source: Copied from the project dashboard.

4.5 Fuel Prices, Freight Rates and Circulation Pressure

Perishable goods are prone to loss due to delays or extra delivery fees; thus, they are less profitable. Figure 12 shows a wide co-movement of diesel prices and the road freight index from 2019 to 2025; both fell in early 2020, rose by 2022, and then declined unevenly. Their short-term changes are not identical. Freight rates are affected by vehicle availability, traffic, road conditions, seasonal demand and the general economy.

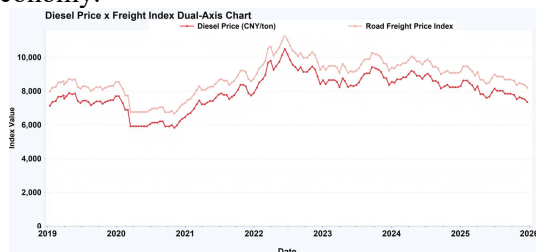


Figure 12. Diesel Price and Road Freight Index

Source: Copied from the project dashboard.

The monthly pattern in Figure 13 is different each year. Freight will generally be higher in the latter half of 2019-2025, but the peak month is not known. The maximum value is 1059, occurring in December 2024, and most 2025 observations are also relatively high.

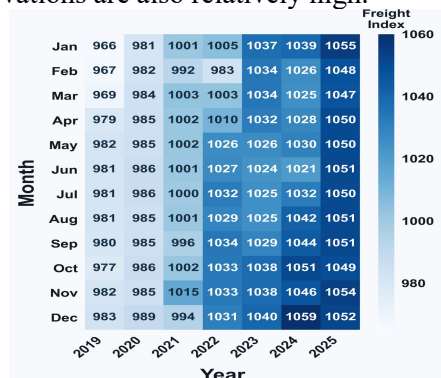


Figure 13. Monthly Freight Index Heatmap by Year

Source: Compiled from the project dashboard. In the annual comparison, diesel falls in 2020 and reaches its highest level in 2022 before easing during 2023-2025. The freight index changes much less. Since the two series use separate axes, Figure 14 only describes their movements; estimating the relationship would require a correlation or regression.

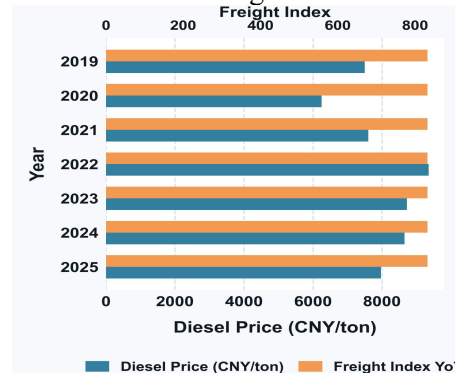


Figure 14. Annual Diesel Price and Freight Index Comparison

Source: Copied from the project dashboard. Figure 15 shows that out of 363 observations in the weekly series, 288 increased (79.3%) and 75 decreased (20.7%). The imbalance is shown in the record above: It is not expected that the same division will be employed again.

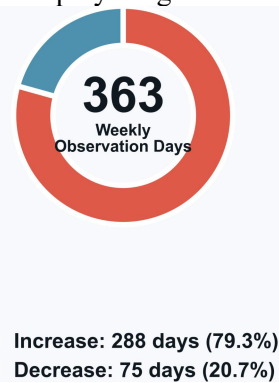


Figure 15. Weekly Freight Index Changes

Source: Copied from the project dashboard.

4.6 Integrated Findings and Analytical Boundary

Three observations emerge from the figures. First, the requirements of the various products are different; therefore, a single index cannot be applied to all goods. Second, there can be a combination of cost, freight and channel constraints; a high land or material bill reduces the margin, diesel increases delivery costs, and a poorly organised channel limits access to external buyers. Thirdly, there is a time lag; the signals of costs, freight and public opinion are concentrated in certain months or years, and

cities with similar e-commerce penetration may report such events at different frequencies.

The interpretations are not perfect. Public-opinion events and e-commerce penetration are proxies, and the figures do not cover the same observations or periods. Price, output, weather and policy are not formally regulated. Therefore, the patterns can help identify potential risks but should not be considered causes of these risks.

5. Empirical Analysis of Unsaleability Causes

The following sections link the values in Figures 6-15 to possible reasons for unsaleability. Cost, Circulation, Public Opinion and E-commerce are shown directly in the dashboard. The current data do not directly show the demand and policy; therefore, these two are employed as the instruments of investigation rather than the causes themselves. We will keep talking about this difference.

5.1 Cost Pressure: Shrinking Profit Space

A producer has a longer reaction time to rising production costs. After covering the costs of production, labour, land and management, the remaining margin at the market price is too small to justify the cost of capital. Extending the growing season will also be costly. Chinese cabbage and pork have a high cost risk and suffered losses in some years in the sample, and maize dropped sharply in 2024. The origins of the pressures for different goods are not the same. Land-based support would fail to address the problems of materials and labour in production for this product.

5.2 Circulation Bottlenecks: Transport, Storage and Cold Chain

Before reaching the buyer, a new product is collected, sorted, stored, transported and then sold. Fuel and freight increase the cost of reaching a distant market, and a lack of storage or refrigeration reduces the time available for sale. The diesel and freight records have the same broad period of 2019-2025, are relatively high around 2022, and show seasonal fluctuations within the years. This information can guide advance arrangements for transport, pre-cooling, and origin storage. Fuel is the only input for the freight rate.

5.3 Demand contraction: Limited market absorption

Demand will be relatively weak. Households

may buy in smaller quantities, wish to purchase different items, or have other requirements. If the planting decisions continue to meet the earlier expectations and consumption grows more slowly, arrivals may exceed purchases for a time. Currently, there are no corresponding series of income, consumption volume and product orders available to test the mechanism. A full-scale test will compare the supply with regional expenditure, wholesale and retail prices, platform orders, and repeat purchases.

5.4 Policy Intervention: Support and Expectation Effects

Although measures to improve production-sales coordination and consumption support, build a cold chain, reduce logistics and e-commerce can lower transaction costs, their actual effectiveness depend on when, where and how well they address real bottlenecks. The current data lack a standard index of policy intensity or results. Therefore, we will not study the effect of this policy; it would be necessary to connect implementation dates, target groups, funds, and subsequent sales, prices and losses.

5.5 Public-Opinion Diffusion: Attention and Expectations

An increase in searches can have many reasons. It may be that a purchase has been made or discussed in the market due to lower prices or overstock of production. The 73 chili anomalies and the repeated crossings for apple, wheat and chili show the dates of the follow-ups. Before issuing a warning, the relevant dates need to be compared with orders, inventory, prices and field data.

5.6 E-Commerce Channels: Conditional Risk Buffering

Online sales have brought the producer and consumer closer, but many of these consumers are not local residents. Shangqiu and Sanmenxia both have good coverage, but the frequency of incidents is very different. The reasons for the above differences may be grading, cold chain distribution, supply continuity, brand awareness or a slow order response. Therefore, the practical problem is how to cooperate among producers, platforms and logistics enterprises, rather than just covering.

5.7 Six-Factor Interaction Mechanism

All of the above reasons will be shown in the

dashboard. High-cost risk is a state of high cost and low return. Freight pressure in the period of intensive marketing reflects transport time. A cluster of public opinion anomalies has been reported, and due to their repeated occurrence in a highly urbanised area, a supply-chain investigation has been launched. Demand and policy are based on weak evidence; therefore, more data need to be collected before they can be combined with the other factors for analysis.

6. Conclusions and Recommendations

6.1 Main Conclusions

In short, the results show that the risk varies by product, year and city, and aggregate surplus is not suitable. Cucumber, Chinese cabbage, apple and eggplant are typical cases of the selected dimensions, and maize has a relatively simple root cause. The two main measured pressures are cost and circulation; Chinese cabbage and pork have a high cost risk and losses in the selected years, and diesel and freight values are relatively high around 2022. Public opinion is not the same; chili, tomato, wheat and Chinese cabbage were mentioned more often in the attention anomaly than sales or inventory losses. E-commerce is not uniformly associated with the frequency of events; Shangqiu and Sanmenxia are both highly developed but have different frequencies. For online access to be effective, the standards, cold chain services, supply organisation, branding and demand feedback work together.

6.2 Recommendations

- (1) Before the next planting cycle, farmers and cooperatives should bring current demand into their decisions. Consumption trends, platform orders, price movements, and each producer's sales history are useful inputs. Where costs are high and profit is falling, contract production, agreed purchasing, staggered planting, or joint procurement can be considered early.
- (2) Producers should focus first on the expenses they can influence. Joint purchasing, shared machinery, practical training, and services delivered at scale can reduce unit cost. Where the cost-profit ratio is weak, agricultural insurance and workable price-protection arrangements should supplement these measures.
- (3) Ahead of a large harvest, especially when diesel and freight are costly, producing areas should line up transport before the crop reaches

market. Pre-cooling, sorting, packaging, storage, refrigeration, and shared delivery need to work as a chain; adding one facility on its own has limited value. Temporary transport arrangements and cross-regional buyer links can be put in place before the harvest arrives.

(4) The demand side needs more than one outlet. Community purchasing, school and canteen procurement, catering suppliers, and direct supermarket links can give the product additional uses. Consistent grading, suitable packaging, and a recognizable brand make it easier for those buyers to accept the product.

(5) Use search activity as a reason to inspect the market, not as proof of unsaleability. A persistent keyword anomaly, simultaneous movement across products, or a wider spread of negative reports deserves follow-up. When signals cluster in one period, compare them with prices, orders, inventories, logistics records, and local observations before issuing a warning.

(6) Judge e-commerce by the sales it creates and the order information it returns, not by penetration alone. Regional brands, live-streaming, short videos, and direct shipment from the origin can reach new buyers. A city with high penetration and recurring events should first check standards, fulfillment, refrigeration, continuity of supply, and the return of order information.

(7) Government agencies can bring the six evidence streams into one warning process. Cost, logistics, consumption, policy, public opinion, and e-commerce data need to move from monitoring to verification, response, and later review. Until actual outcome data are available, a combination such as high cost with low profit, unusually expensive freight, clustered online attention, or weak channel coordination can be used to select cases for checking. The thresholds and predictive models should be recalibrated when sales, prices, inventories, and confirmed event labels become available.

6.3 Limitations and Future Research

First, the analysis is limited by the difference in the time period and statistical definition of the source tables. A more serious defect is the absence of continuous records for confirmed unsaleable events, inventories, transaction volumes and wholesale prices; without such data, neither the cause nor the prediction accuracy can be properly verified. Baidu searches and policy-event counts are also

indirect indicators of attention and external conditions, not observations of sales or policy effects. Future work will build a region-product-time panel of prices, sales, inventories, weather, cold-chain facilities, platform orders and verified events. Interviews, surveys and case studies can be used to determine the practical feasibility of the proposed mechanisms. Panel models, event studies, random forests and time-series methods can be employed for causal analysis and forecasting.

References

- [1] Hao L, Lee K S. The impact of economic policy uncertainty on agricultural prices: Evidence from China. *Asia and the Global Economy*, 2024, 4(2): 100100. DOI: 10.1016/j.aglobe.2024.100100.
- [2] Luo M, Zhou G, Xu H. A differential game model research on dynamic pricing and coordination of fresh agricultural products supply chain based on freshness. *Economic Research-Ekonomika Istraživanja*, 2023, 36(2): 2177696. DOI: 10.1080/1331677X.2023.2177696.
- [3] Zhu H, Liu C, Wu G, Gao Y. Cold Chain Logistics Network Design for Fresh Agricultural Products with Government Subsidy. *Sustainability*, 2023, 15(13): 10021. DOI: 10.3390/su151310021.
- [4] Guo J, Jin S, Zhao J, Li Y. E-commerce and supply chain resilience during COVID-19: Evidence from agricultural input e-stores in China. *Journal of Agricultural Economics*, 2023, 74(2): 369-393. DOI: 10.1111/1477-9552.12518.
- [5] Fu L, Zhang H. Analysis of factors influencing small-scale agricultural product prices from the perspective of the online public—a case study of China. *Frontiers in Sustainable Food Systems*, 2024, 8: 1355853. DOI: 10.3389/fsufs.2024.1355853.
- [6] Liu S, Jin Y, Zheng F. Did product certification and e-commerce benefit agricultural producers in China?. *Food Policy*, 2024, 127: 102688. DOI: 10.1016/j.foodpol.2024.102688.
- [7] Jiang L, Zhang Y. Modeling Urban Expansion and Agricultural Land Conversion in Henan Province, China: An Integration of Land Use and Socioeconomic Data. *Sustainability*, 2016, 8(9): 920. DOI: 10.3390/su8090920.
- [8] Tan Y, Zeng H. Price transmission, reserve regulation and price volatility. *China Agricultural Economic Review*, 2019, 11(2): 355-372. DOI: 10.1108/CAER-04-2017-0062.
- [9] Ling L, Zhang D, Chen S, Muger A W. Can online search data improve the forecast accuracy of pork price in China?. *Journal of Forecasting*, 2020, 39(4): 671-686. DOI: 10.1002/for.2649.
- [10] Lu H, Sun G, Zhou L, Zhang W, Zou J. The impact and mechanism of E-commerce of agricultural products on nongrain production of cultivated land in rural China. *Journal of Environmental Management*, 2025, 373: 123639. DOI: 10.1016/j.jenvman.2024.123639.
- [11] Liu Z, Zhao K, Ma J, Wang C. A Price Index Model for Road Freight Transportation and Its Empirical analysis in China[C]. *MATEC Web of Conferences*, 2017, 100: 05017. DOI: 10.1051/mateconf/201710005017.
- [12] Tian X, Yi F, Yu X. Rising cost of labor and transformations in grain production in China. *China Agricultural Economic Review*, 2019, 12(1): 158-172. DOI: 10.1108/CAER-04-2018-0067.
- [13] Zhang T, Feng Q. Multidimensional Risk Assessment of China's Grain Supply Chain with Entropy Weight TOPSIS Method. *Procedia Computer Science*, 2024, 242: 853-858. DOI: 10.1016/j.procs.2024.08.213.
- [14] Wang H, Lu L, Liu Z, Sun Y. Study on measurement and prediction of agricultural product supply chain resilience based on improved EW-TOPSIS and GM (1,1)-Markov models under public emergencies. *PLOS ONE*, 2025, 20(5): e0321248. DOI: 10.1371/journal.pone.0321248.