Research on multi-algorithm fusion and comparison of intelligent farmland planting strategies based on BRBMO-BiLSTM-KAN model prediction data

Rongsheng Chen *

Institue of mathematics and computer science, Northwest Minzu University, Lanzhou, China, 730030

* Corresponding Author Email: 1429283177@qq.com

Abstract. Intelligent farmland planting strategy is an important topic of intelligent planting. In order to improve the adaptability and accuracy of the strategy in the complex market environment, this paper generated simulated data based on the macro data of the National Bureau of Statistics of China, used BRBMO-BiLSTM-KAN to predict the expected sales volume and carried out cluster analysis, and then adopted the simulated annealing algorithm to obtain the crop planting strategy. Finally, a comprehensive evaluation model based on Entropy Weight-Grey Relation-TOPSIS was used to compare the crop planting strategies obtained by the LPO algorithm under dynamic environment.The results show that the model successfully improves the adaptability and accuracy of intelligent farmland planting strategy in complex market environment, and provides strong technical support for intelligent planting.

Keywords: Cluster analysis; BRBMO-BiLSTM-KAN model; simulated annealing algorithm; Lungs performance-based optimization (LPO).

1. Introduction

In the early and middle 1990s, developed countries put forward the concept of precision agriculture, which is based on information technology to promote agricultural production and management informatization and sustainable development [1]. Based on the development of fine agriculture, smart agriculture is a deep integration of modern information technology and agriculture. It not only needs the support of faster, higher accuracy, lower cost and diversified information sharing service mechanism, but also needs to analyze agricultural and biological resources, environment, ecology and socio-economic conditions in different regions [2]. Therefore, it is necessary and challenging to study the intelligent farmland planting strategy. Taking the mountainous area of North China as an example, under the challenges of resource constraints and planting efficiency, this paper innovatively integrates theories such as big data analysis and dynamic optimization, innovatively applies Kolmogorov-arnold networks to smart planting, and proposes for the first time an agricultural parameter verification method suitable for complex market environments, thus building a set of efficient and flexible farmland planting strategies.

2. Research method

2.1. Data acquisition and preprocessing

In this paper, cost and income data of China's agricultural products from National Bureau of Statistics of China from 1990 to 2023 are collected and information of China's agricultural products from rural areas in mountainous areas of North China is obtained from the open-source website (https://www.mcm.edu.cn).

Based on the national cost and income of agricultural products from 1990 to 2023 of the National Bureau of Statistics of China, the data table of sales volume, sales price, total output and planting cost is simulated. In order to achieve the accuracy and interpretability of regression prediction, this paper adopts an algorithm based on KAN model [3]. The research on the prediction algorithm based on KAN has verified its adaptability, effectiveness and feasibility [4].

Therefore, the Red-billed blue magpie optimizer algorithm (RBMO) [5] was adopted to optimize the following four parameters of BiLSTM-KAN model:

Lstm_units: The number of LSTM units, this parameter determines the number of neurons in each LSTM layer.

Layer_units: The number of neurons in the DenseKAN layer of KAN network.

l2_units: The intensity of L2 regularization is used to control the complexity of the model and prevent overfitting.

Learning_rate: The learning rate controls the pace of model weight updating.

The first 80% of the data table was taken as the training set, and the last 20% was taken as the test set, and the expected sales volume, selling price and planting cost were obtained by regression prediction. Then quantitative treatment of crops and plots in rural areas in mountainous areas of North China was carried out (Table 1):

Table 1. Quantitative results of crops and plots in mountainous areas of North China.

	Canavalia	Cabbage	Pumpkin	...	Cucumber	Soya beans	Broomcorn millet	Black bean
i=	1	2	3	...	38	39	40	41
	A1	A2	A3	...	F1	F2	F3	F4
j=	1	2	3	...	51	52	53	54

Then classify the data. After processing the data, the following Figure 1-3 is obtained, from which we can clearly see the macro-statistical rules of planting plots, planting costs, per mu yield and sales price of each plant.

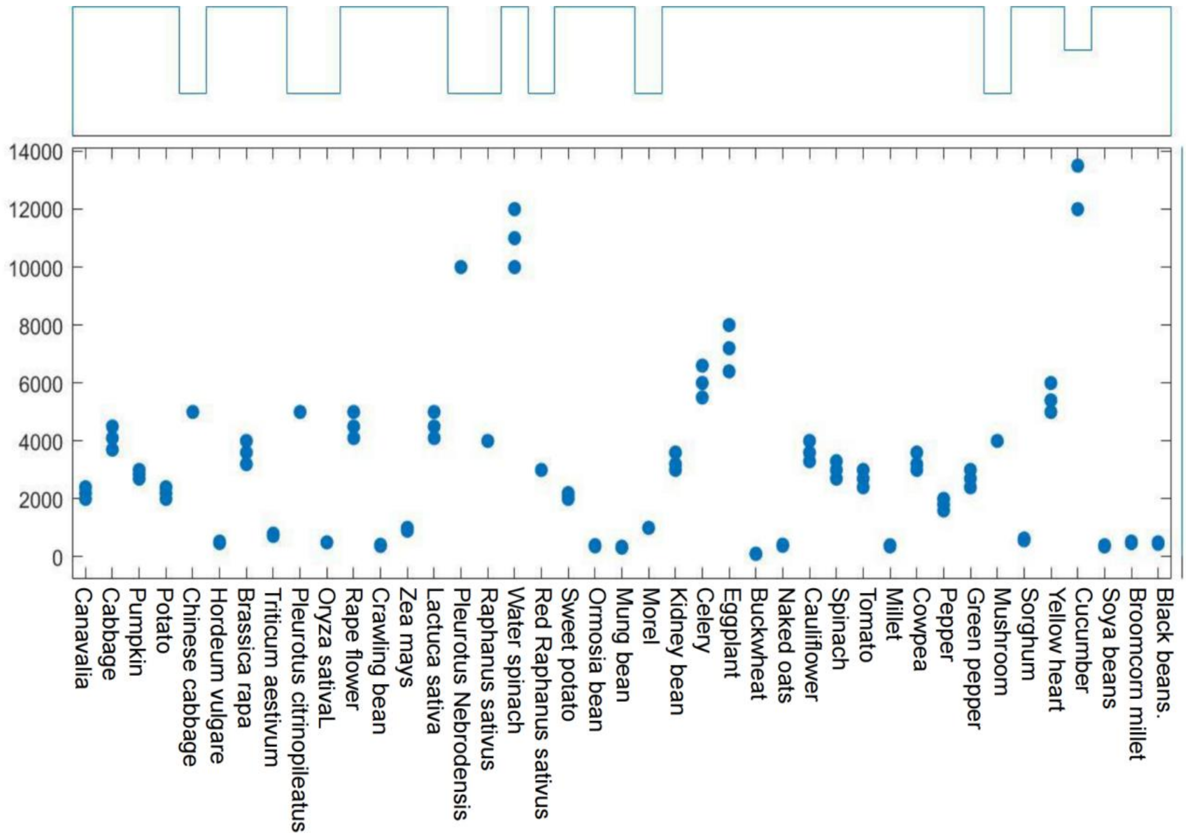

Figure 1. Point column chart of agricultural yield per mu and planting cost.

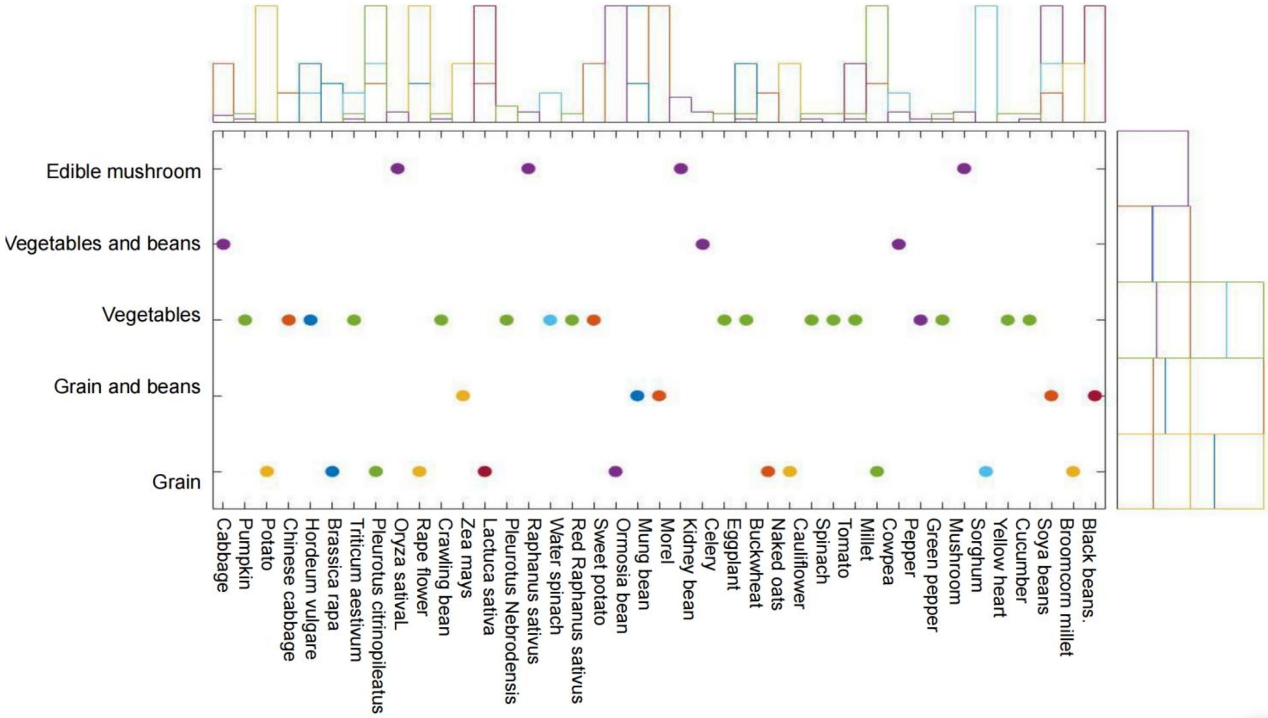

Figure 2. Dot bar chart of crop types and sales volume.

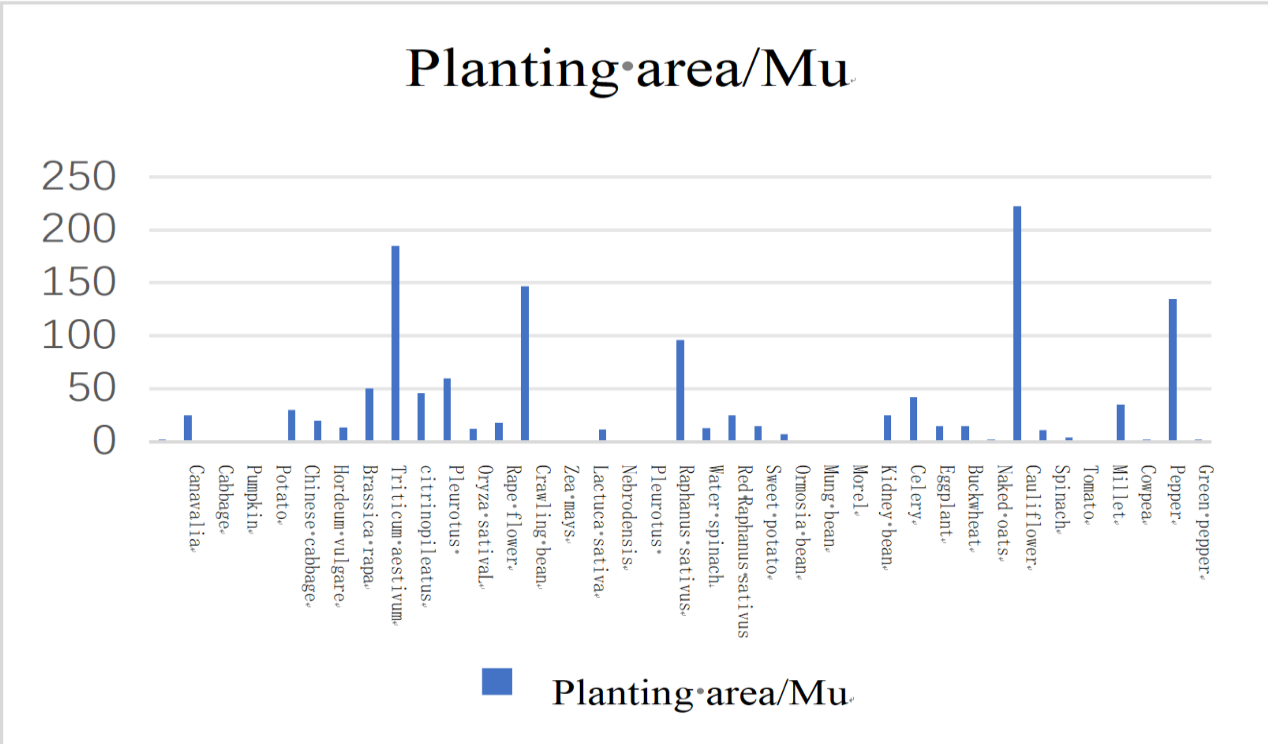

Figure 3. Histogram of planting area by crop (Excerpt).

2.2. Method introduction

The simulated annealing algorithm has a relatively complete and highly mature system in solving optimization problems [6].The parameters of this model are set as follows: temperature attenuation coefficient is 0.95, initial temperature is 100, final temperature is 0, and iteration times are 100.

Lungs performance-based optimization, which aims to complete a comprehensive and robust optimization algorithm, and the experimental results also precisely prove its practicability and value [7].Compared with genetic algorithm, ant colony algorithm and other algorithms, it shows strong evolutionary ability and can adapt to the problem environment more quickly and find the optimal solution. The specific steps of iteration are shown in the literature [7].

3. Model building and solving

3.1. The establishment of constraint conditions

Let I represent the index set composed of all plots and J represent the index set composed of all crops($i \in I, j \in J$); T is the set of all natural numbers from 1 to 7.

According to literature [8], it is assumed that (expected) sales volume and total production are roughly in a certain proportion, the proportion is 0.95 to 0.90. In this paper, logarithmic mean is used to define selling price and price value Figure 4 is obtained through aggregate analysis (the ratio of expected sales volume to total crop production is denoted as k). Its advantage is that it is more conducive to modeling when the relationship between the original data and response variables may be non-linear, and the probability of skewness can be effectively reduced.

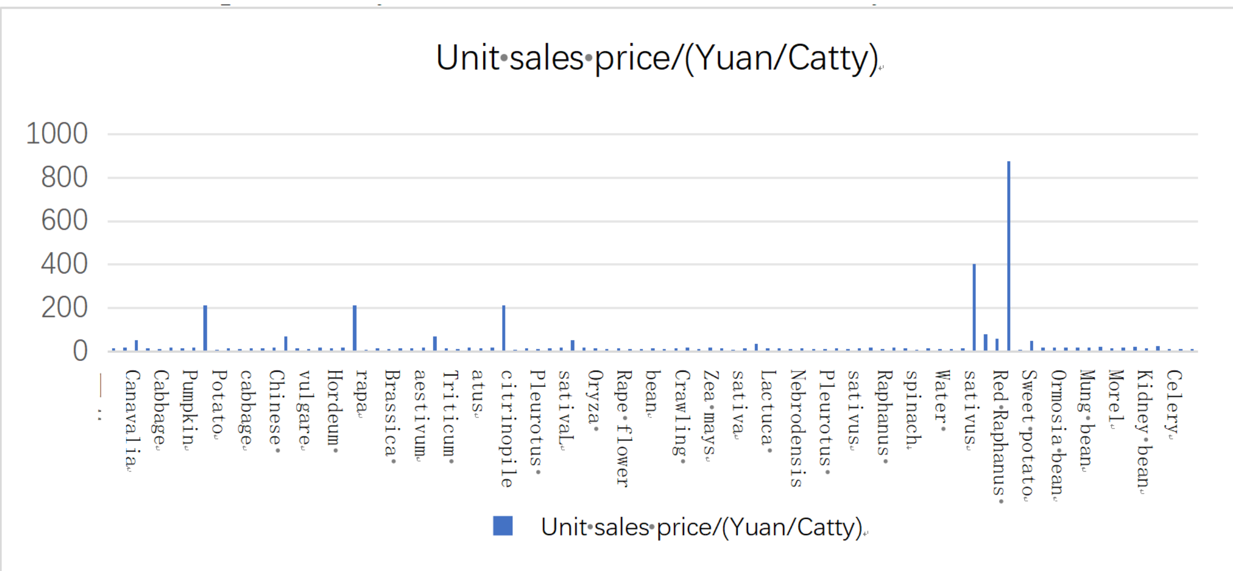

Figure 4. The logarithmic mean of the selling price of each crop.

Since the planting area of crops on each plot cannot be negative, and the total planting area of all different crops on the same plot is the area of the plot, there is

$$\begin{cases} 0 \le x_{ij,t}, \\ \sum_{ij} x_{ij,t} = A_i, \end{cases} \tag{1}$$

Where $x_{ij,t}$ represents the planting area of crops on the ground of block i .

According to the literature [9], assuming that the crop species of the combined species shall not be less than 10%, then there are

$$\sum_j H\left(\frac{x_{ij,t}}{x_{ij,t}+1}\right) \le 10, \tag{2}$$

Where H is the Heaviside function, i.e. H of x is equal to 0, for x is less than or equal to 0 and H of x is equal to 1, for x greater than 0.

According to the growth laws of crops, each crop in the same plot (including greenhouses) cannot be continuously planted, otherwise it will reduce production

$$x_{ij,t} x_{ij,t+1} = 0, \tag{3}$$

Since soil containing rhizobacteria of legume crops is conducive to the growth of other crops, starting from 2023, all land in every plot (including greenhouses) will be required to plant legume crops at least once in three years

$$\sum_{i=1}^{3} B_{ij} \geq 1 , \tag{4}$$

Where B_{ij} represents the binary variable of whether beans are planted in a given season.

Flat drylands, terraces and hillsides are suitable for planting grain crops one season a year. Irrigated land is suitable for growing one season of rice or two seasons of vegetables per year. Ordinary greenhouses are suitable for planting one season of vegetables and one season of edible fungi every year, and smart greenhouses are suitable for planting two seasons of vegetables every year. The same plot (including greenhouses) can be combined with different crops each season, then

$$\mathrm{P_{ij}} = C_{ij} = S_{ij} = 0 \quad \begin{cases} 0 \neq \left|\mathrm{sgn}(26-i)\,\mathrm{sgn}(j-16)\right| \leq 2, \\ i = 26,\ j = 16, \end{cases} \tag{5}$$

Among them, S_{ij}, P_{ij} and C_{ij} respectively represent the unit price of sales of j crops on the ground of block i, the yield per mu of j crops on the ground of block i, and the planting cost of j crops on the ground of block i.

The statistical data were visually analyzed, and the crops were classified according to the plots, such as soybean flat dry land, soybean terrace, soybean hill land, etc. Because of the randomness of the plot, the heat map of the three columns of random variables and the yield per mu can be analyzed that the correlation between the yield per mu and the plot is very low.

First, the data is divided into two categories of grain + edible fungi and vegetables for statistics. First, for grain + edible fungi, two random variables, planting cost and per mu yield were taken to draw heat maps. Take two random variables, the unit price of sales and the yield per mu to draw a heat map. Take two random variables, planting cost and unit sales price to draw a heat map. In the same way, for vegetables, two random variables, planting cost and per mu yield were taken to draw heat maps. Take two random variables, the unit price of sales and the yield per mu to draw a heat map. Two random variables, planting cost and unit sales price were taken to draw heat maps.

Compare Figure 5-7, thus, it is easy to find the following constraints, namely:

i) The planting cost and selling unit price of the same crop on flat dry land, terrace and hillside are the same

ii) The selling price of the same crop on irrigated land and ordinary greenhouses is the same

$$\begin{cases} P_{ij} = P_{(i+1)j} & \mathrm{j} < 50, \\ \mathrm{C_{ij}} = C_{(i+1)j} & \mathrm{j} < 15, \\ S_{ij} = S_{(i+1)j} & \mathrm{j} < 15, \end{cases} \tag{6}$$

Paraphrase i)-ii) into mathematical language as follows:

3.2. The mathematical language expression of each risk parameter

In real life, risk is characterized by high uncertainty, randomness and volatility. These factors together lead to a variety of crops in the expected sales, yield per mu, planting cost and selling price there are significant uncertainties, and hidden complex planting risks. Therefore, the parameters should be selected according to the actual situation and the reliability test should be carried out.

The expected future sales of wheat and corn have a trend of growth, with an average annual growth rate of $5\%\sim10\%$, and the expected annual sales of other crops have a change of about $\pm5\%$ compared with 2023. That is, the expected sales volume of each crop each year thereafter is:

$$\left\{rand_1^t(a_j,b_j)+H[(j-6)(j-7)](1\pm 5\%)\right\}kP_{ij}x_{ij,t}, \tag{7}$$

The non-measured coefficient is defined as the risk parameter of the expected sales volume.

The yield of crops per mu is often affected by climate and other factors, and there is a $\pm 10\%$ change every year. That is, the annual yield of each crop per mu is

$$(1\pm 10\%)^t P_{ij}, \tag{8}$$

The non-measured coefficient is defined as the risk parameter of mu yield.

The cost of growing crops increases by about 5% a year on average. That is, the annual planting cost of each crop is

$$rand_2^t(a_j,b_j)C_{ij}, \tag{9}$$

The non-measured coefficient is defined as the risk parameter of planting cost.

The selling prices of food crops were basically stable. The sales price of vegetable crops has a growing trend, with an average annual increase of about 5%. The sales price of edible mushrooms is stable and has fallen, about $1\%\sim5\%$ per year, especially the sales price of morel mushrooms has fallen by 5% per year. That is, the annual sales price of each crop is

$$rand_3^t(a_j,b_j)S_{ij}, \tag{10}$$

The non-measured coefficient is defined as the risk parameter of the selling price.

The set of all comprehensive risk parameters $:= k_T^{'} \otimes rand_2^T$

P.S Description of the $rand_\mu^t(a_j,b_j)$ function:

$$\mu=1\begin{cases} j\neq 6,7 \;\; a_j=b_j=0, \\ j=6,7 \;\; a_j=1+5\% \quad b_j=1+10\% \end{cases} \tag{11}$$

$$\mu=2 \;\; \forall j\in J, a_j=1+4.5\% \quad b_j=1+5.5\% \tag{12}$$

$$\mu=3\begin{cases} j\leq 16 \;\; a_j=b_j=0, \\ 16<j<38 \;\; a_j=1+4.5\% \quad b_j=1+5.5\%, \\ 38\leq j\leq 40 \;\; a_j=1-1\% \quad b_j=1-5\%, \\ j=41 \;\; a_j=b_j=1-5\%, \end{cases} \tag{13}$$

3.3. Standardize the data matrix of each risk parameter

Considering the expected sales volume, yield per mu, uncertainty of planting cost and selling price of various crops and potential planting risks, it is necessary to determine the risk parameters. Firstly, four types of risk parameters are planned, and then their respective autocorrelation matrices are constrained. That is, check and screen the risk parameters of expected sales volume, yield per mu, planting cost and selling price in section 3.2.On the basis of self-constraints, Pearson correlation coefficients are mutually constrained according to Table 2. Finally, comprehensive parameter values are formulated according to the box plot constructed by the data matrix, as shown in Figure 5.

Table 2. Pearson correlation coefficient reference table.

0	ρ_{13}	ρ_{14}	ρ_{23}	ρ_{24}
$>0\wedge\neq\pm1$	ρ_{12}	ρ_{34}		

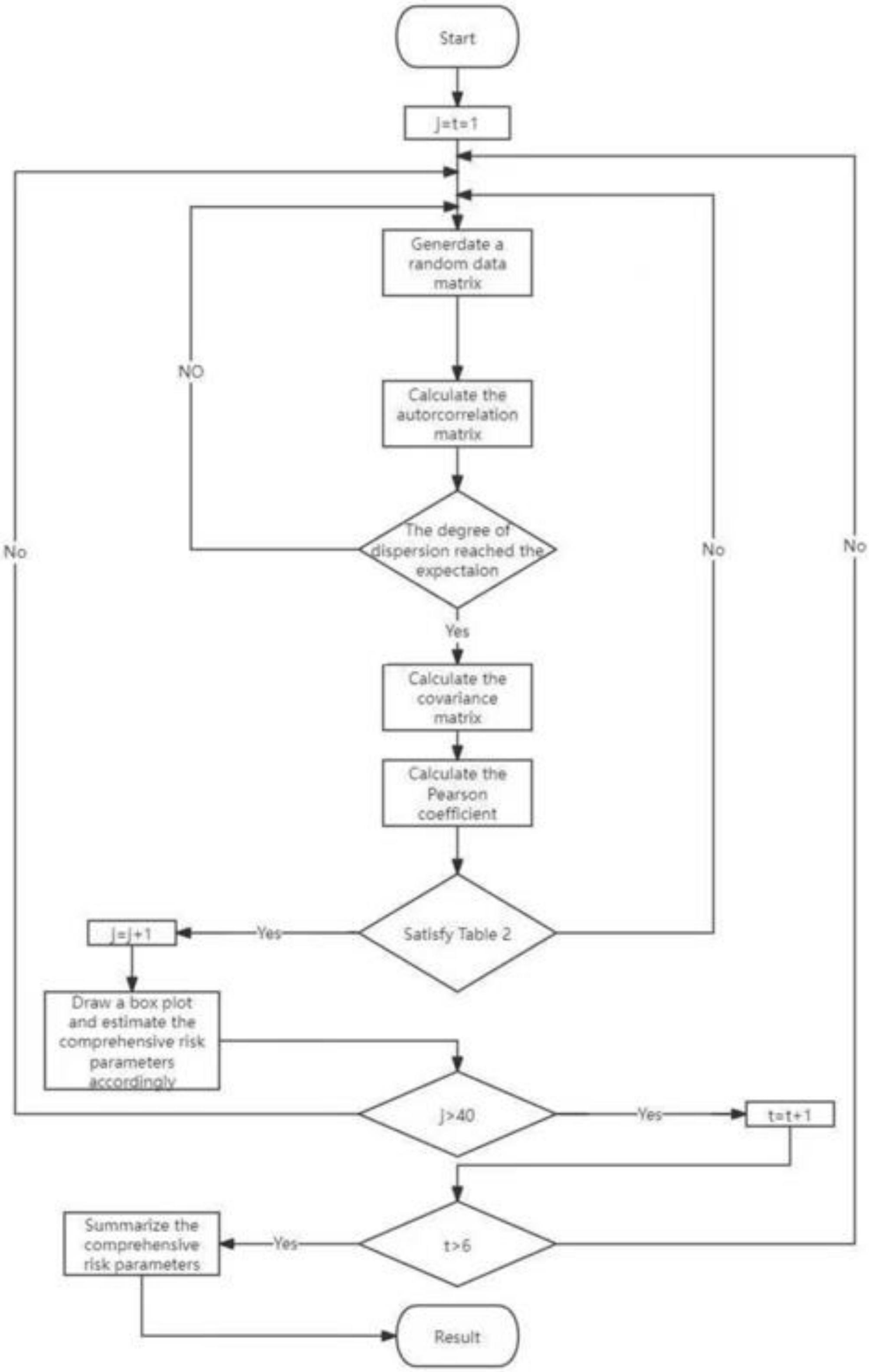

Figure 5. Flowchart of each risk parameter of the specification.

3.4. Cross price elasticity and complementarity coefficient

Because of the substitutability and complementarity between crops, cross price elasticity and complementarity coefficient should be introduced. Under the demand elasticity, the demand rate can be described by the differential equation of crop stocks (see reference [10] for the formula). In order to facilitate modeling, this paper considers that the demand changes randomly, and the price elasticity coefficient is denoted as E_{jn}, which is used to represent the substitution between crops j and n. If the price of crop j falls, the demand for crop n will fall, and vice versa. The mathematical language is as follows:

$$D_{ij,t}^{'} = D_{ij,t}(1+E_{jn}\frac{\sum_i \Delta S_{in}}{\sum_i S_{in}}), \tag{14}$$

If two crops are complementary, planting on the same plot may increase their yield or reduce planting cost. The complementarity coefficient θ_{jn} can be considered to represent the complementarity of crops j and n, and the P_{ij} value of crop j can be adjusted:

$$P_{IJ}^{'} = P_{ij}(1+\theta_{jn}), \tag{15}$$

3.5. Construction of objective function

Because if the total output of crops in each season exceeds the corresponding expected sales volume, the excess part cannot be sold normally. Taking the total crop revenue from 2024 to 2030 as the objective function, the optimization problem under the dynamic environment can be obtained as follows:

$$\max \sum_{t=2024}^{2030}\sum_{i,j}(k_t^{'}kP_{ij}x_{ij,t}S_{ij} - rand_2^t C_{ij}x_{ij,t}) \tag{16}$$

$$\begin{cases} 0 \le x_{ij,t}, \\ x_{ij,t}x_{ij,t+1} = 0, \\ \sum_j H(\frac{x_{ij,t}}{x_{ij,t}+1}) \le 10, \\ \sum_{ij} x_{ij,t} = A_i, \\ P_{ij} = P_{(i+1)j} & \mathrm{j} < 50, \\ \mathrm{C_{ij}} = C_{(i+1)j} & \mathrm{j} < 15, \\ S_{ij} = S_{(i+1)j} & \mathrm{j} < 15, \\ \mathrm{P_{ij}} = C_{ij} = S_{ij} = 0 & \begin{cases} 0 \ne |\mathrm{sgn}(26-i)\,\mathrm{sgn}(j-16)| \le 2, \\ i = 26, j = 16, \end{cases} \\ \sum_{i=1}^{3} B_{ij} \ge 1 & \forall i \in I, \end{cases} \tag{17}$$

If the complex market environment is considered, cross-price elasticity and complementarity coefficient need to be introduced.

The mapping of the target variable D_{ij} (expected sales) obtained by RBMO-BiLSTM-KAN algorithm for the three characteristic variables of $P_{ij}x_{ij,t}$、C_{ij}、S_{ij} (total output, planting cost and selling price) is recorded as a function:

$$F: P_{IJ}^{'}x_{IJ,T} \oplus C_{IJ} \oplus S_{IJ} \to \{D_{IJ,T}^{'}\} \text{ via } (P_{ij}^{'}x_{ij,t}, C_{ij}, S_{ij}) \mapsto D_{ij,t}^{'} \tag{18}$$

Than

$$\max \sum_{t=2024}^{2030}\sum_{i,j}(\, k_t^{'}F(P_{ij}^{'}x_{ij,t}, C_{ij}, S_{ij})S_{ij} - rand_2^t C_{ij}x_{ij,t}) \tag{19}$$

$$\begin{cases} 0 \le x_{ij,t} \\ x_{ij,t}x_{ij,t+1} = 0 \\ \sum_j H\left(\frac{x_{ij,t}}{x_{ij,t}+1}\right) \le 10 \\ \sum_{ij} x_{ij,t} = A_i \\ P_{ij} = P_{(i+1)j} \qquad \mathrm{j} < 50 \\ \mathrm{C}_{\mathrm{ij}} = C_{(i+1)j} \qquad \mathrm{j} < 15 \\ S_{ij} = S_{(i+1)j} \qquad \mathrm{j} < 15 \\ \mathrm{P}_{\mathrm{ij}} = C_{ij} = S_{ij} = 0 \qquad \begin{cases} 0 \ne \left|\mathrm{sgn}(26-i)\mathrm{sgn}(j-16)\right| \le 2 \\ i = 26, j = 16 \end{cases} \\ \sum_{i=1}^{3} B_{ij} \ge 1 \qquad \forall i \in I \end{cases} \tag{20}$$

3.6. Solution of the model

For the model (16-17) under dynamic environment, the lung function optimization algorithm was used to solve the optimal solution, and SOM diagram was made after classifying the plots. Part of the results were shown in Figure 6 and. Figure 7.

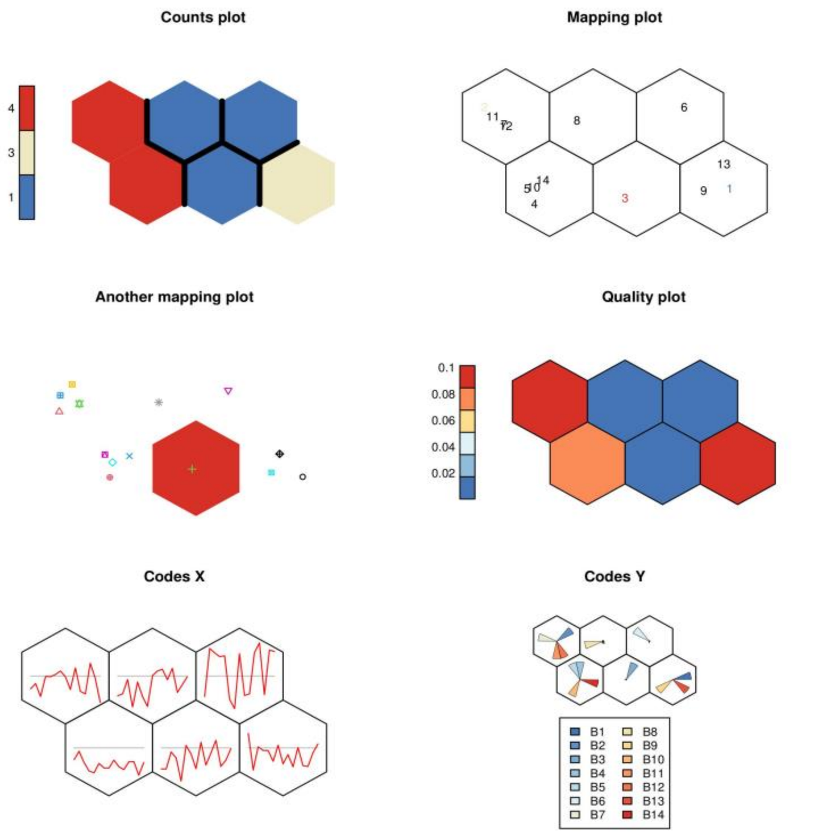

Figure 6. SOM diagram of flat dry land.

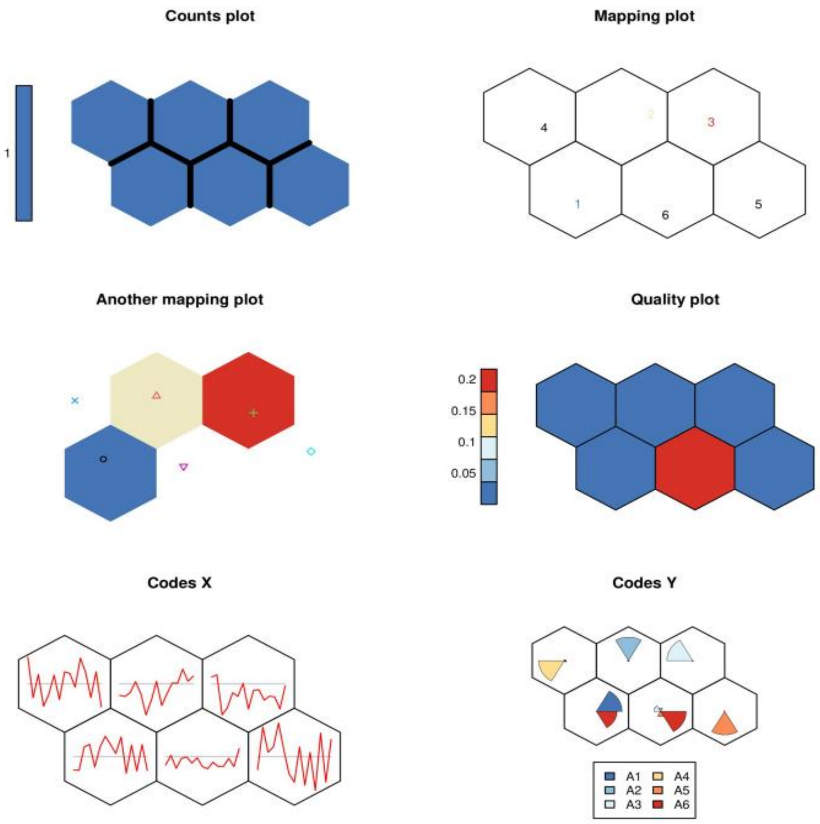

Figure 7. SOM diagram of terraced field.

For the model under complex market environment (19-20), simulated annealing algorithm was used to solve the optimal solution, and the following results were obtained:

Table 3. Optimal crop planting strategies under complex market environment (part).

Season	Plot name	Soya beans	Black bean	Ormosia bean	...	Broomcorn millet	Sweet potato	Avena chinensis
1	A1	8	0	8	...	8	0	8
1	A2	5.5	0	5.5	...	0	5.5	0
1	A3	3.5	0	0	...	3.5	0	3.5
1	A4	7.2	7.2	0	...	7.2	0	0
1	A5	6.8	6.8	6.8	...	6.8	6.8	0
1	A6	5.5	5.5	5.5	...	0	0	5.5
1	B1	6	0	0	...	6	6	6
1	B2	4.6	4.6	4.6	...	4.6	4.6	0
1	B3	4	4	0	...	4	4	4
1	B4	2.8	0	0	...	2.8	2.8	0
1	B5	2.5	0	2.5	..	0	2.5	2.5
1	B6	8.6	0	0	...	8.6	8.6	8.6
1	B7	5.5	5.5	5.5	...	0	5.5	0
1	B8	4.4	4.4	4.4	...	4.4	0	0
1	B9	5	0	5	...	0	0	5
...	...	...	...	...	...	...	...	...
2	E14	0	0	0	...	0	0	0
2	E15	0	0	0	...	0	0	0
2	E16	0	0	0	...	0	0	0
2	F1	0	0	0	...	0	0	0
2	F2	0	0	0	...	0	0	0
2	F3	0	0	0	...	0	0	0
2	F4	0	0	0	...	0	0	0

The table shows that terrace is the main type of cultivated land, accounting for 52%, with a total area of 619 mu, distributed in 14 plots, ranging in size from 20 mu to 86 mu. Slope land and irrigated land accounted for 9% each, the slope land area of 108 mu, divided into 6 plots, the area range between 13 mu and 27 mu: irrigated land area of 109 mu, divided into 8 plots, the area between 6 mu and 22 mu.

4. Evaluation of model

The results of mathematical models are compared indirectly by using the comprehensive evaluation model based on Entropy Weight-Grey Relation-TOPSIS.

Firstly, an evaluation system is given for mathematical models in dynamic environment and complex market environment, and a corresponding evaluation index system is constructed based on this system. There are four indexes in the evaluation system, which are described in Table 4 below:

Table 4. Evaluation indicators.

Index	Description of the metric
Biodiversity	Number of species per unit plot area
Economic benefit	The comparative relationship between gains and inputs in an economic activity
Technical feasibility	An evaluation process to determine whether a project, plan, or program can be successfully implemented within specific technical and resource constraints
Market risk assessment	The process of analyzing and evaluating the various risks that crop sales face in the market

Firstly, for the model under dynamic environment (total seven years), entropy weight method is used to calculate the weights of four indicators through information entropy, as shown in Figure 8(a). Similarly, for the model under complex market environment (total seven years), entropy weight method is used to calculate the weights of four indicators through information entropy, as shown in Figure 8(b).

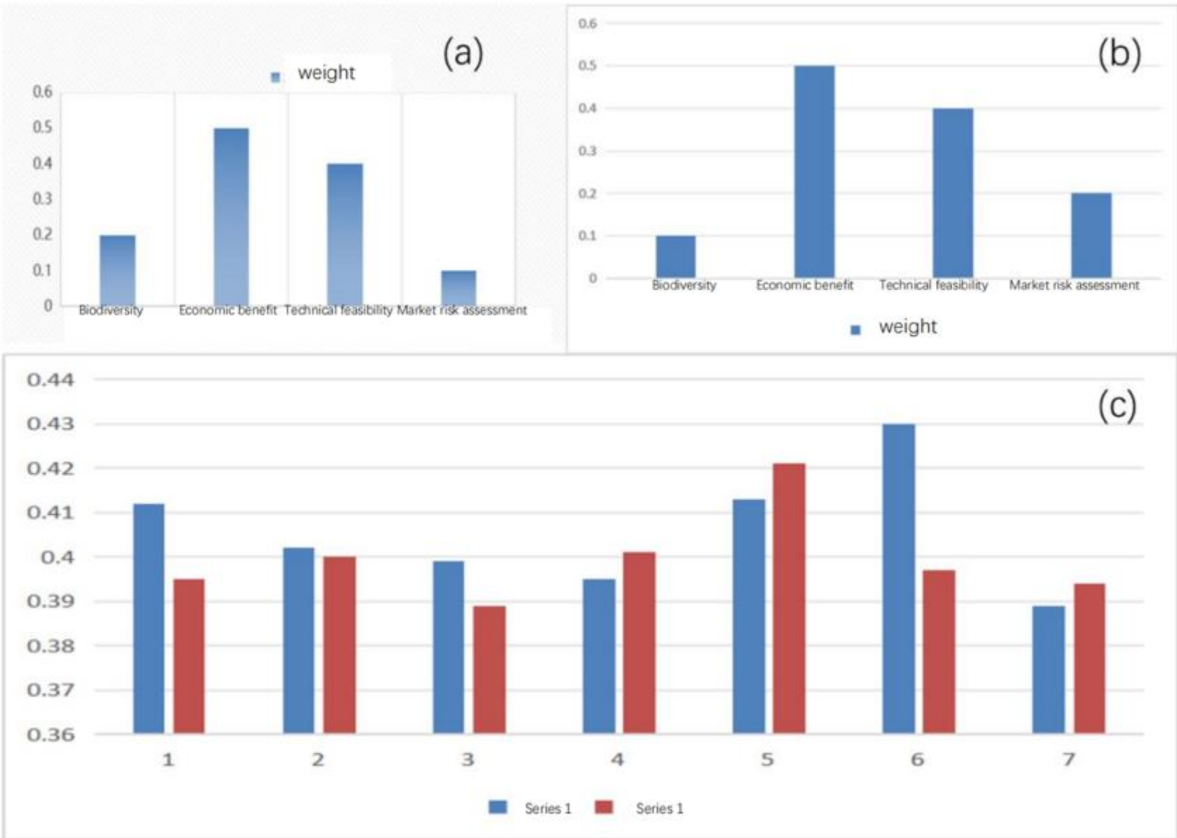

Figure 8. Weight bar chart.

It is found that the weight of economic effect indicators is the largest, so the weight of economic effect indicators is used as the reference frame to construct the weight comparison chart of economic effect indicators for seven years (Figure. 8(c)). It can be found that the weight of economic effect indicators of dynamic environment is higher than that of complex market in 4 years, while the weight of economic effect indicators of dynamic environment is slightly lower than that of complex market in the other 3 years. Finally, the comprehensive evaluation of entropy weight, grey correlation and topsis was carried out (Figure. 9(a), Figure. 9(b)).

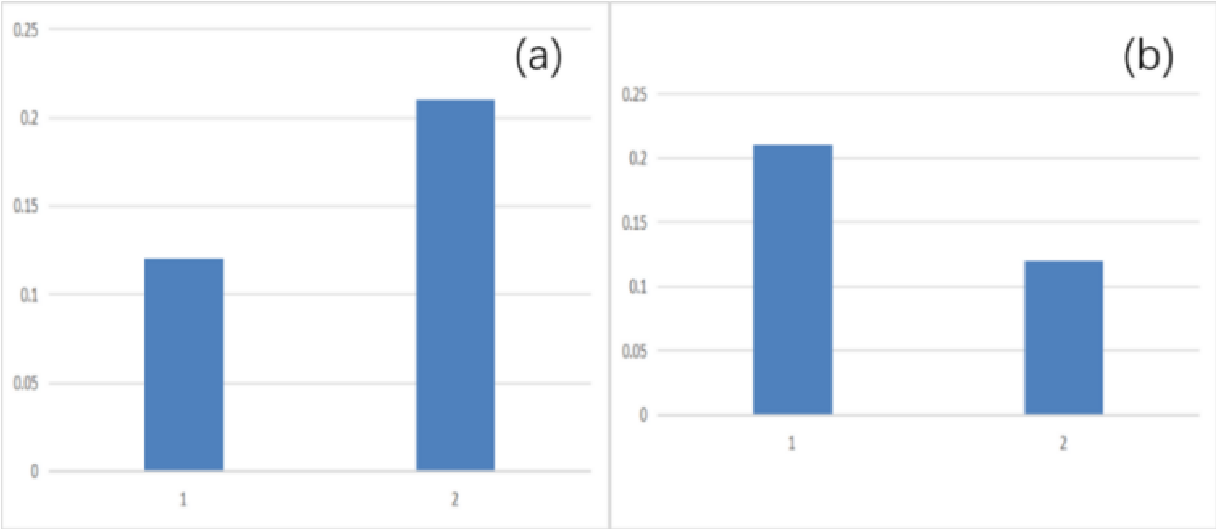

Figure 9. Euclidean distance (a) of positive ideal solution and Euclidean distance (b) of negative ideal solution.

5. Conclusion

According to the above statistical phenomenon, for dynamic environment, economic effect index has a relatively high weight among the four indicators, while for complex market environment, economic effect index has a relatively low weight among the four indicators. This is because complex market environment involves more diversified factors, such as policy changes, consumer preference changes, international situation, etc. These factors make the economic effect no longer the only or dominant measure, but need to be integrated with other indicators. Moreover, the model in the complex market environment is closer to the positive ideal solution and farther away from the negative ideal solution in the measure space of Euclidean distance. In conclusion, the model in the complex market environment has a better, more realistic and more interpretable solution algorithm for the possible substitutability and complementarity between market risks and crops, and can better describe the actual situation, make decisions with low error rate, and find the optimal solution in line with the goal. References

References

[1] IDOJE G, DAGIUKLAS T, IQBAL M. Survey for smart farming technologies: Challenges and issues [J].Computers & Electrical Engineering, 2021, 92

[2] LI Li, LI Minzan, LIU Gang, ZHANG Man, WANG Maohua. Goals, key technologies, and regional models of smart farming for field crops in China [J]. Smart Agriculture, 2022, 4(4): 26-34.

[3] Liu Z, Wang Y, and Vaidya S, et al. Kan: Kolmogorov-arnold networks [J]. arxiv preprint arxiv:2404.19756, 2024.

[4] Han X, Zhang X, Wu Y, et al. KAN4TSF: Are KAN and KAN-based models Effective for Time Series Forecasting? [J]. arxiv preprint arxiv: 2408.11306, 2024.

[5] Fu S, Li K , Huang H ,et al.Red-billed blue magpie optimizer: a novel metaheuristic algorithm for 2D/3D UAV path planning and engineering design problems[J].Artificial Intelligence Review, 2024, 57(6).

[6] Guillot J, Restrepo-Leal D, Carlos Robles-Algarín, et al. computation search for global maxima in multimodal functions by applying numerical optimization algorithms: a comparison between golden section and simulated annealing [J].Springer US, 2019

[7] Ghasemi M, Zare M, Zahedi A, et al. Optimization based on performance of lungs in body: Lungs performance-based optimization (LPO) [J]. Computer Methods in Applied Mechanics and Engineering, 2024, 419: 116582.

[8] HAN Zhipeng, LI Weiguang, FENG Yingying, HUANG Lanxin, NIUJing. Research on the scientific ratio and marketing status of crops [J].Intelligent Computer and Applications, 2021, 11(11):153-156.

[9] Guoe Li.Effects of Soil Heterogeneity and SpeciesComposition on Biomass and Competitiveability of Three Forages [D] Lanzhou University, 2023.

[10] P. N. Ghare, G. F. Schrader. A model for exponentially decaying inventories [J]. Journal of Industrial Engineering, 1963, 15: 238-243.