Research on Crop Planting Strategy Decision-Making Model Based on Intelligent Optimization Algorithms

Yumeng Yang[1,a,*], Chongrun Wang[2,b], Liren You[3,c], Min Yang[4,d], Ruqian Xiao[1,e]

[1]School of Information Engineering, Xi'an Mingde Polytechnic University, Xi'an, China

[2]School of Higher Vocational and Technical Education, Xi'an Mingde Polytechnic University, Xi'an, China

[3]School of Economics and Management, Xi'an Mingde Polytechnic University, Xi'an, China

[4]School of Intelligent Manufacturing and Control Technology, Xi'an Mingde Polytechnic University, Xi'an, China

[a]1431422435@qq.com, [b]2350718147@qq.com, [c]3280044896@qq.com, [d]2381250973@qq.com, [e]2753677498@qq.com

*Corresponding author

Keywords: Crop Planting Strategy, Intelligent Optimization Algorithms, Linear Programming, Robust Optimization, Genetic Algorithms

Abstract: This paper investigates a decision-making model for crop planting strategies based on intelligent optimization algorithms. Initially, we collected basic information on arable land, crops, and planting conditions for the year 2023, and performed data preprocessing to ensure the accuracy and reliability of the results. We then established a decision-making model aimed at maximizing total revenue and used MATLAB to solve the optimization model, obtaining the optimal planting plans for scenarios where the excess expected sales volume is either unsold or sold at 50% of the 2023 price. The article also constructs a linear programming model considering the overstocking issue due to excess sales beyond expectations and proposes a new objective function. Additionally, we incorporated uncertainties such as sales growth rates, yields, planting costs, and sales prices, and developed a robust optimization model. By analyzing the substitutability and complementarity of crops, we established a multi-objective optimization model and solved the optimal planting strategy using genetic algorithms. Finally, we verified the rationality of the model by comparing the planting revenue before and after optimization, and found that the revenue increased after optimization.

1. Introduction

This paper aims to study a decision-making model for crop planting strategies based on intelligent optimization algorithms, in order to provide scientific and rational planting plans for agricultural production [1]. Firstly, we collected relevant data on arable land, crops, and planting situations for the year 2023, and conducted meticulous data preprocessing to ensure the accuracy and reliability of the analysis results [2]. On this basis, we constructed a decision-making model aimed at maximizing total revenue and introduced linear programming methods to evaluate the impact of different planting schemes on the final total revenue [3]. Additionally, we considered two scenarios when sales volume exceeds expectations: unsold and sold at 50% of the 2023 price, and established corresponding optimization models for each.

To more realistically simulate real-world data, we introduced uncertainty factors related to the expected growth rate of crop sales, sales volume, yield, planting costs, and sales prices, and constructed a robust optimization model. By analyzing the substitutability and complementarity between crops, we further established a multi-objective optimization model to determine the optimal planting strategy considering various complex factors [4].

The research in this paper not only provides a new decision support tool for crop planting strategies but also offers an effective solution for optimizing planting strategies through the application of genetic algorithms [5]. By comparing the planting revenue before and after optimization, we verified

  DOI: 10.25236/icmmct.2025.020

the rationality of the model and found that the revenue increased after optimization, thereby proving the practical application value of the model in agricultural decision-making.

2. Crop Planting Strategy Decision-Making Model

We first collected the basic information on the existing arable land, crops, and the crop planting situation for 2023, along with relevant data. To prevent the impact of abnormal or missing data on the accuracy and reliability of the results, we performed data preprocessing. Using SPSS, we conducted data cleaning and validation, and found that there were no missing or outlier values. Then, we merged the crop-related data and planting information, ensuring that the land type and sales data corresponded to each crop. We initially establish a decision-making model for crop planting strategies,with the objective function being the maximization of total revenue, and constraints including that the planting area of each plot must not exceed its actual area. We use MATLAB to solve the optimization model and obtain the optimal planting plans for scenarios where the excess expected sales volume is either unsold or sold at 50% of the 2023 price.

2.1. Linear Programming Model Construction

2.1.1. Excess Sales Beyond Expectations Resulting in Overstock

Crops with sales exceeding expectations cannot be sold but incur planting costs. Different planting schemes can be evaluated based on the final total revenue, thereby establishing the objective function.

$$max\, L = \sum_{i=1}^{n}(p_i \times q_i) - \sum_{i=1}^{n}(A_i \times S_i) \tag{1}$$

In this context, L represents profit, p_i represents the expected sales volume, q_i represents the selling price, A represents the cost, and s_i represents the area of land planted.

Each plot's planting area must not exceed the actual area available.

$$\sum_{\mathrm{j}} \mathrm{S}_{\mathrm{ijt}} \leq \mathrm{E}_{\mathrm{i}} \tag{2}$$

In which, E_i represents the area of the i-th plot.

The land must be planted with legume crops at least once within three years.

$$\sum_{i\in B} 1(x_k^{t-1} = i) + \sum_{i\in B} 1(x_k^{t-2} = i) \geq 1 \tag{3}$$

In which, $1(\cdot)$ represents a function that takes the value 1 when crop i is planted, and 0 otherwise, and x_k^t represents the crop i planted in plot k in year t.

To facilitate management, the planting areas should not be too dispersed. The number of plots planted with each crop i on arable land type $l \in L$ should not exceed 1.

$$\sum_{k\in G_l} 1(x_k^t = i) \leq 1 \tag{4}$$

G_l represents the set of plots in arable land type l, and x_k^t represents the crop i planted in plot k in year t.

$$|Q_i^t - D_i^t| \leq \varepsilon_i \tag{5}$$

D_i^t represents the expected sales volume of crop i in year t, and Q_i^t represents the total yield of crop i in year t.

$$Q_i^t = \sum_{i=1}^{2} \sum_{k=1}^{N} \times A_{k,j}^t \tag{6}$$

The final model can be formulated as follows:

$$max\, L = \sum_{i=1}^{n}(p_i \times q_i) - \sum_{i=1}^{n}(A_i \times S_i) \tag{7}$$

$$
s.t.=\begin{cases} \sum_j x_{ijt} \le S_i \\ S_{ijt} \ge 0 \\ x_{i,t}^{t} \ne x_{i,t}^{t+1} \\ \sum_{i\in B} 1(x_k^{t-1}=i)+\sum_{i\in B} 1(x_k^{t-2}=i) \ge 1 \\ \sum_{k\in G_l} 1(x_k^{t}=i) \le 1 \\ |Q_i^t - D_i^t| \le \varepsilon_i \end{cases} \tag{8}
$$

2.1.2. Excess Sales at 50% of 2023 Prices

For the part of sales that exceeds the expected volume, it is sold at 50% of the 2023 selling price. After calculation, the sales revenue after the price reduction is still higher than the planting cost. We establish a new objective function:

$$
max\, L_2 = \sum_{i=1}^{n} [(p_i \times q_i) + 0.5 \times p_i \times max(Q_i^t - D_i^t, 0) - A_i^t] \tag{9}
$$

p_i represents the expected sales volume, q_i represents the selling price, Q_i^t represents the total yield of crop i in year t, D_i^t represents the expected sales volume of crop i in year t, and A represents the cost. The constraints are the same as for the excess unsold portion.

2.2. Model Solution

First, we analyze the data to visually understand the distribution of the land area occupied by each plot. Using MATLAB, we create a plot showing the distribution of land area for each plot as shown in Figure 1.

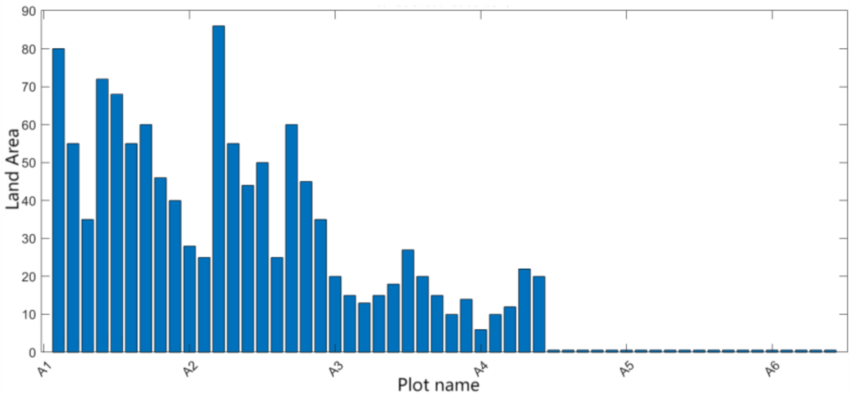

Figure 1 Plot of land area distribution

Using MATLAB to create a scatter plot to display the relationships between yield per acre, sales price, and planting cost, the results are shown in Figure 2. The scatter plot illustrates the relationships among yield per acre, sales price per unit, and planting cost. From the plot, it can be observed that there is a certain positive correlation between yield per acre, sales price per unit, and planting cost. That is, as the yield per acre increases, the planting cost and sales price per unit tend to be higher as well.

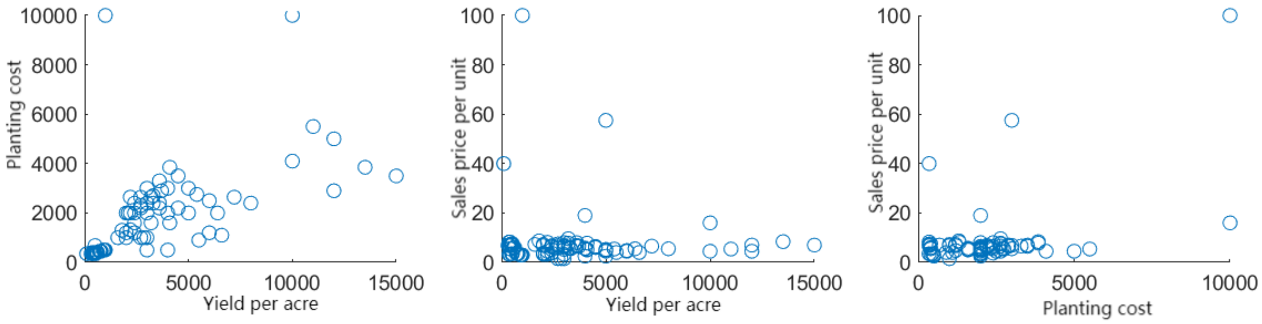

Figure 2 Scatter plot of the three indicators

The expected sales volume, planting cost, yield per acre, and sales price of crops are expected to stabilize compared to 2023. Our dataset provides the planting cost, yield per acre, and sales price. For the expected sales volume, it is obtained by multiplying the crop planting area by the yield per acre to get the expected sales volume of crops for 2023. The expected sales volume for the future is assumed to remain stable relative to 2023, providing a mathematical basis for the optimization of subsequent decision-making models.

Using MATLAB to solve the optimization model, we can obtain the optimal planting plan for the part of the expected sales volume that is unsold, such as planting 3.5 mu of soybeans in plot A3 in the first season of 2024. Similarly, using MATLAB to solve for the optimal planting plan where the excess part is sold at 50% of the 2023 price, such as planting 3.5 mu of soybeans in plot A3 in the first season of 2024.

3. Model Optimization

3.1. Robust Optimization Considering Uncertainty

In order to more accurately simulate real-world data, we incorporate uncertainties related to the annual growth rate of expected sales for wheat and corn, the sales volume of other crops, yield per acre, crop planting costs, the sales prices of vegetable crops, and the sales prices of edible fungi. By establishing a robust optimization model, we aim to develop the optimal crop planting plan for rural areas from 2024 to 2030. We define uncertain parameters as follows: the expected sales growth rate for wheat and corn is between 5% and 10%, denoted as α_1 and α_2, respectively. The expected sales volume of other crops is subject to $\pm 5\%$ change relative to 2023, with a change coefficient β_i. The yield per acre of crops can vary by $\pm 10\%$, with a change coefficient γ_i. Planting costs are expected to increase by about 5% annually, with a growth coefficient δ. The sales prices of vegetable crops are expected to increase by about 5% annually, with a growth coefficient λ_i (where i refers to vegetable crops). The sales prices of edible fungi are expected to decrease by 1% to 5% annually, with a decline coefficient ζ_i (where i refers to edible fungi).

The objective function is to maximize the total revenue.

$$maxL = \sum_{t=2024}^{2023} \sum_i \sum_j x_{ijt} q_{it} y_{it} - \sum_{t=2024}^{2030} \sum_i \sum_j x_{ijt} A_{it} \tag{10}$$

In which, q_{jt} represents the sales price of the \j-th crop in 2023, r_{qjt} is the growth rate of its sales price, y_{jt} is the yield per acre, r_{yjt} is the growth rate of the yield per acre, A_{jt} is the planting cost, and r_{Ajt} is the growth rate of the planting cost.

The planting area of each plot must not exceed the actual area available.

$$\sum_j x_{ijt} \leq S_i \tag{11}$$

s_i represents the area of the i-th plot.

Considering the uncertainty in yield per acre, the total output of each crop fluctuates within a certain range of the expected sales volume.

$$(1-\gamma_i) y_i^t \sum_j x_{ijt} \leq D_i^t (1+\beta_i) \tag{12}$$

$$D_i^t (1-\beta_i) \leq (1+\gamma_i) y_i^t \sum_j x_{ijt} \tag{13}$$

D_i^t represents the expected sales volume of the i-th crop in year t.

To meet the requirements of discontinuous planting and planting legume crops at least once every three years, we introduce a binary variable z_{ijt} to indicate whether the i-th crop is planted in year t, and an auxiliary variable h_{ij}^t to represent the last time the i-th crop was planted on the j-th plot.

$$z_{ijt} + z_{ijt-1} + z_{ijt-2} \leq 2 \tag{14}$$

$$\sum_{t=2023}^{2025} \sum_{i \in D} z_{ijt} \geq 1 \tag{15}$$

$$h_{ijt} = (t-1)(1-z_{ijt}) + z_{ijt}\, h_{ijt-1} + 3z_{ijt}(1-z_{ijt-1}) \tag{16}$$

In which, D represents the set of legume crops.

Considering the uncertainty in sales prices and planting costs, we define upper and lower bounds for them:

$$\underline{p_{it}} \le p_{it} \le \overline{p_{it}} \tag{17}$$

$$\underline{c_{it}} \le c_{it} \le \overline{c_{it}} \tag{18}$$

In which, $\overline{p_{it}}$ and $\underline{p_{it}}$ represent the lower and upper bounds of the sales price for the i-th crop in year t, respectively, and $\overline{c_{it}}$ and $\underline{c_{it}}$ represent the lower and upper bounds of the planting cost for the i-th crop in year t, respectively.

3.2. Analysis of Crop Substitutability and Complementarity

Based on practical life experience and research, the existence of substitutability among crops[5] allows for more flexible and diverse planting strategies, which can better reduce planting risks. For example, when the expected sales volume of a certain crop decreases or the planting cost increases, one can opt to plant a substitutable crop. The substitution relationships between crops under different circumstances can be represented as follows:

$$min\{\frac{c_{ij}^{(1)}}{c_{ij}^{(1)}+c_{ij}^{(2)}}, \frac{c_{ij}^{(2)}}{c_{ij}^{(1)}+c_{ij}^{(2)}}\} \tag{19}$$

Where, $c_{ij}^{(1)}$ and $c_{ij}^{(2)}$ represent two crops that are substitutable in the j-th quarter of the i-th year.

After importing the data on sales volume, selling price, and cost into SPSS for normality testing, it can be concluded that the expected sales volume, selling price, and cost are normally distributed. Furthermore, by conducting a Spearman correlation analysis on the same data in SPSS [6], we obtain a correlation coefficient table and a corresponding heatmap, as illustrated in Table 1.

Table 1 Correlation Coefficient Table

Example column 1	Planting Cost (Yuan/Mu)	Sales Price (Yuan/Jin)	Expected Sales Volume
Planting Cost (Yuan/Mu)	1(0.000***)	0.458(0.000***)	-0.629(0.000***)
Sales Price (Yuan/Jin)	0.458(0.000***)	1(0.000***)	0.011(0.921)
Expected Sales Volume	-0.629(0.000***)	0.011(0.921)	1(0.000***)

The table above shows the parameter results of the model test, including the correlation coefficient and the significance P-value. To test whether there is a statistically significant relationship between X and Y, we check if the P-value is significant (P < 0.05). This indicates the presence of a correlation between the two variables, and the sign and magnitude of the correlation coefficient further describe the direction and strength of the relationship.

Table 2 Linear Regression Analysis Results Table (n=87)

Example text 1	Unstandardized Coefficients		Standardized Coefficients	t	P	VIF	R ²	Adjusted R ²	F
	B	Standard Error	Beta						
Constant	18.75	14.206	-	1.32	0.191	-	0.707	0.697	F=66.82 P=0.000 ***
Planting Cost	-0.011	0.005	-0.24	-2.36	0.020 **	2.931			
Sales Price	1.651	0.44	0.349	3.74	0.000 ***	2.452			
Planting	4.667	0.404	0.789	11.5	0.000 ***	1.322			
Dependent Variable: Expected Sales Volume									
Note: ***, **, * represent the significance levels of 1%, 5%, and 10%, respectively.									

Furthermore, we conducted a linear regression analysis on the expected sales volume, selling price, and cost using SPSS, and the results are shown in Table 2 as follows:

The analysis of the F-test results indicates a significance P-value of 0.000***, which is significant at the conventional levels. This allows us to reject the null hypothesis that the regression coefficients are zero, suggesting that the model essentially meets the requirements. Regarding the issue of variable collinearity, all Variance Inflation Factors (VIFs) are below 10, indicating that there is no problem of multicollinearity in the model. Therefore, the model is well-constructed.

The formula for the model is as follows:

$$y = 18.751 - 0.011 \times A_i + 1.651 \times q_i + 4.677 \times S_i \tag{20}$$

A_i represents the planting cost, q_i represents the sales price per unit, and S_i represents the planting area of the crop.

From Figure 3, the fitting effect diagram can be seen, indicating a good fit.

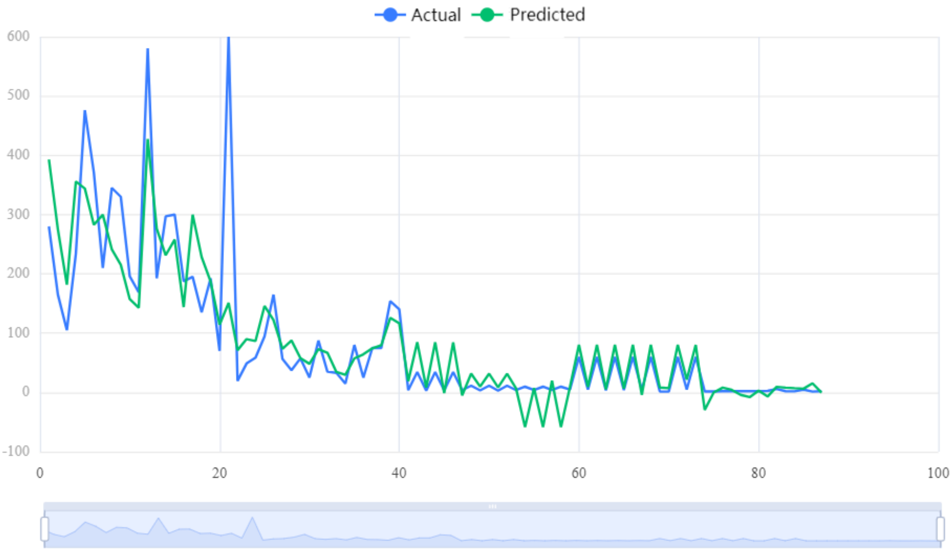

Figure 3 Fitting effect diagram

3.3. Construction of Multi-Objective Optimization Models

Considering the substitutability between crops, the correlation between expected sales volume, selling price, and cost, and taking into account various complex factors to determine the optimal planting strategy, we establish a multi-objective optimization model.

The objective function 1 is to maximize the total revenue.

$$maxL_{ij} = (Q_k \times S_{ik} \times q_i) - (A_i \times S_{ik}) \tag{21}$$

In which, $Q_k \times S_{ik}$ represents the total yield of crop i on plot k, q_i represents the sales price per unit of crop i, and A_j represents the planting cost.

The objective function 2, aiming to minimize risk, is as follows:

$$Z_2 = \sum_{i=1}^{n} \sum_t \left[\sum_j x_{ijt} y_{it} (1 + \beta_i^{t-2023}) - d_{it}(1 + \gamma_i^{t-2023})\right]^2 \tag{22}$$

The constraints are the same as those in section 3.1.

4. Optimal Planting Strategy Based on Genetic Algorithms

Using genetic algorithms to solve the established model, we perform operations such as individual encoding, selection, crossover, and mutation on the population, iterating and optimizing continuously to obtain the optimal planting strategy. Some of the optimal planting plan data are shown in Table 3.

Table 3 Partial Data of the Optimal Planting Plan

Season	Plot Name	Soybeans	Black Beans	Red Beans	Mung Beans	Cowpeas
1	A1	8	8	8	8	8
1	A2	0	0	0	5.5	0
1	A3	0	0	3.5	0	3.5
1	A4	0	7.2	7.2	7.2	0
1	A5	6.8	6.8	6.8	0	0
1	A6	5.5	0	0	5.5	5.5
1	B1	6	6	0	6	0
1	B2	0	0	0	4.6	4.6
1	B3	4	4	4	0	0
1	B4	2.8	2.8	0	2.8	0
1	B5	2.5	0	0	2.5	2.5
1	B6	0	8.6	8.6	8.6	0
1	B7	0	5.5	5.5	0	0
1	B8	0	0	4.4	4.4	4.4

Table 3 presents partial data of the optimal planting plan under the condition that the excess sales volume is sold at 50% of the 2023 sales volume. For example, in the first season of 2024, soybeans are planted on 8 mu of Plot A1.

Randomly select a crop for single-objective and multi-objective optimization analysis. First, calculate the average planting cost of the crop for each plot type, then determine the maximum cost for the sample crop's predicted land area in 2024. Since cost and revenue are negatively correlated, we can approximately determine the relationship between the planting revenue of soybeans before and after optimization in 2024, as shown in Figure 4.

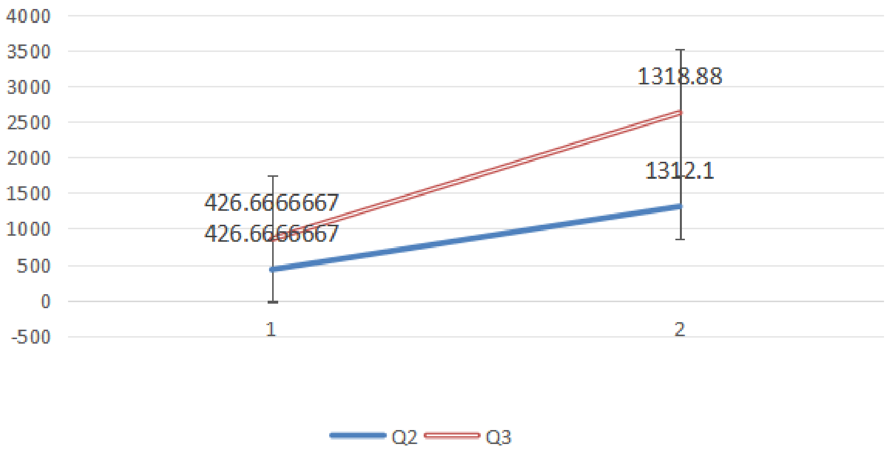

Figure 4 Sample revenue comparison chart

In Figure 4, Q2 represents the revenue fitting curve of soybeans before optimization, and Q3 represents the revenue fitting curve of soybeans after optimization. From the figure, it can be seen that Q3 is greater than Q2, which indicates that the model is reasonable. The revenue result after optimization is higher than the revenue result before optimization, with an increase of 6.78.

5. Conclusion

This study successfully developed and validated a decision-making model for crop planting strategies based on intelligent optimization algorithms. By integrating data on arable land, crops, and planting conditions, as well as uncertainties in sales volume, costs, and prices, the model effectively optimizes planting plans to maximize total revenue. The application of genetic algorithms further enhances the model's capability in handling complex multi-objective optimization problems, and experimental results indicate that the optimized planting strategies can significantly improve agricultural profits. Future research can explore a wider range of intelligent optimization algorithms to adapt to a broader range of agricultural environments and conditions.

References

[1] Bogdanović S, Hadžić M. Strategic multicriteria decision-making process in agriculture[J]. Економика пољопривреде, 2019, 66(1): 89-106.

[2] Castañeda-Vera A, Leffelaar P A, Álvaro-Fuentes J, et al. Selecting crop models for decision making in wheat insurance[J]. European Journal of Agronomy, 2015, 68: 97-116.

[3] Dantzig G B. Linear programming[J]. Operations research, 2002, 50(1): 42-47.

[4] Gunantara N. A review of multi-objective optimization: Methods and its applications[J]. Cogent Engineering, 2018, 5(1): 1502242.

[5] Reeves C R. Genetic algorithms[J]. Handbook of metaheuristics, 2010: 109-139.

[6] De Winter J C F, Gosling S D, Potter J. Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data[J]. Psychological methods, 2016, 21(3): 273.