

Study on the process conditions of C4 olefin preparation based on statistical method

Zijian Hu*, Danyu Li, Linhao Qian

Northeast Forestry University, Harbin, China

* Corresponding Author Email: 2954491804@qq.com

Abstract. Due to the wide application of C4 olefins, it is of profound research significance and important value to explore the process conditions for the preparation of C4 olefins by catalytic coupling of ethanol. This paper calculates and analyzes the data generated in the C4 olefin preparation process by establishing a mathematical model and provides targeted suggestions based on statistical methods. First, one-dimensional linear regression was used to analyze the relationship between ethanol conversion, C4 olefin selectivity and temperature so as to obtain the regression coefficients of each group and test their significance; Secondly, SPSS is used for matrix scatter plot to observe the linear trend and spearman correlation coefficient is used for correlation analysis. Hypothesis test is conducted to draw relevant conclusions; Then the fuzzy synthesis is used to draw the conclusion that the catalyst has a greater impact on the reaction. Finally, the neural network is used to train the data and arrange them to obtain all the combination of catalyst and temperature. It is obtained that at 400°C, the combination of "200mg 2wt% Co/SiO₂-200mg HAP ethanol concentration 2.1ml/min" can obtain 45.259% of the maximum C4 olefin yield without limiting the temperature; If the temperature is lower than 350°C, the catalyst is "50mg 5wt% Co/SiO₂-50mg HAP ethanol concentration 0.3ml/min" and the temperature is 275°C. The maximum C4 olefin yield is 30.75%.

Keywords: Regression analysis; Significance test; Spearman correlation coefficient; Fuzzy comprehensive evaluation; BP neural network prediction.

1. Introduction

In production, C4 olefins have been widely used in many chemical and pharmaceutical products. It is well known that ethanol is the raw material for the production and manufacture of C4 olefins. In the chemical production process, catalysts and temperature are both important factors that affect the chemical reaction. Therefore, how catalysts are used and how temperature is controlled would have a great impact on the preparation of C4 olefins, including the selectivity of C4 olefins and the yield of C4 olefins. Therefore, the analysis of the design of catalyst combinations and the selection of appropriate temperature are of great significance and value in production.

2. Model construction and solving

2.1. Analysis of influence on catalyst combination

2.1.1. One-dimensional linear regression model

By Observing all the data provided, it was determined that a univariate linear regression model was used for all of them. A univariate linear regression model refers to a set of x independent variables and a set of y dependent variables. If the linear relationship between x and y is approximately satisfied, a regression model can be built to represent these n sets of data^[1].

Each catalyst set is separated out to find the temperature, ethanol conversion, and C4 olefin selectivity of each set. Then the regression equation is established and the regression coefficients for each group are obtained by using stata. See TABLE1-2 for specific data.(Where β_0 represents the constant term of the univariate linear regression equation and β_1 represents the regression coefficient of the univariate linear regression equation.)

Table 1. Regression coefficients of ethanol conversion versus temperature for different catalyst combination

Catalyst Portfolio	β_0	β_1
A1	-84.08	0.33
A2	-161.89	0.66
A3	-95.88	0.42
A4	-144.57	0.58
A5	-97.62	0.41
A6	-119.83	0.50
A7	-74.26	0.38
A8	-83.78	0.34
A9	-65.67	0.25
A10	-49.65	0.18
A11	-56.50	0.21
A12	-74.81	0.29
A13	-67.37	0.25
A14	-86.69	0.34
B1	-73.14	0.28
B2	-70.95	0.27
B3	-36.16	0.13
B4	-56.81	0.21
B5	-72.27	0.27
B6	-100.00	0.38
B7	-109.34	0.42

Table 2. Regression coefficients of C4 olefin selectivity versus temperature for different catalyst combination

Catalyst Portfolio	β_0	β_1
A1	-3.24	0.15
A2	-41.55	0.22
A3	-59.19	0.26
A4	-52.41	0.23
A5	-57.81	0.23
A6	-50.75	0.20
A7	-44.26	0.19
A8	-57.26	0.24
A9	-59.10	0.25
A10	-12.36	0.05
A11	-13.31	0.05
A12	-47.73	0.20
A13	-35.89	0.16
A14	-35.12	0.14
B1	-56.73	0.24
B2	-61.07	0.24
B3	-28.30	0.12
B4	-22.21	0.10
B5	-34.60	0.15
B6	-45.76	0.19
B7	-56.45	0.23

According to the data in the table, the coefficients and the constant term in each set can be seen. Then we test the significance of this one-dimensional linear regression model.

Table 3. P-value of regression coefficient between ethanol conversion and temperature under different catalyst combination

Catalyst combination	P value of β_0	P value of β_1	Catalyst combination	P value of β_0	P value of β_1
A1	0.013**	0.008***	A12	0.015**	0.008***
A2	0.001***	0.000***	A13	0.031**	0.019**
A3	0.001***	0.000***	A14	0.015**	0.008***
A4	0.000***	0.000***	B1	0.015**	0.009***
A5	0.017**	0.006***	B2	0.038**	0.022**
A6	0.006***	0.003***	B3	0.028**	0.017**
A7	0.000***	0.000***	B4	0.025**	0.015**
A8	0.009***	0.004***	B5	0.021**	0.011**
A9	0.043**	0.026**	B6	0.011**	0.005***
A10	0.039**	0.025**	B7	0.012**	0.006***
A11	0.053*	0.036**			

*** p<0.01 ** p<0.05 * p<0.1

Table 4. P-value of regression coefficient between ethanol conversion and temperature under different catalyst combination

Catalyst combination	P value of β_0	P value of β_1	Catalyst combination	P value of β_0	P value of β_1
A1	0.832	0.045**	A12	0.007***	0.003***
A2	0.094*	0.030**	A13	0.004***	0.002***
A3	0.005***	0.001***	A14	0.018**	0.010**
A4	0.009***	0.003***	B1	0.005***	0.002***
A5	0.003***	0.001***	B2	0.005***	0.002***
A6	0.081*	0.046**	B3	0.004***	0.001***
A7	0.016**	0.007***	B4	0.053*	0.016**
A8	0.002***	0.001***	B5	0.002***	0.001***
A9	0.000***	0.000***	B6	0.001***	0.000***
A10	0.115**	0.061*	B7	0.000***	0.000***
A11	0.003***	0.001**			

*** p<0.01 ** p<0.05 * p<0.1

β_1 in the linear regression equation of one variable is known to obey normal distribution,

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum x_i^2}\right) \tag{1}$$

The following statistics can be constructed by replacing real σ^2 with unbiased estimator,

$$t = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{\sigma}^2 / \sum x_i^2}} = \frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} \sim t(n-2) \tag{2}$$

For β_0 in the univariate linear regression equation, we construct the following t statistics for significance test,

$$t = \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{\sigma}^2 \sum X_i^2 / n \sum x_i^2}} = \frac{\hat{\beta}_0}{S_{\hat{\beta}_0}} \sim t(n-2) \quad (3)$$

Next, the original hypothesis is put forward. $H_0: \beta_1(\beta_0)=0$; $H_1: \beta_1(\beta_0) \neq 0$ and then the t statistic is constructed. After the stata calculation, the corresponding p value is obtained.

It can be seen from TABLE 3 that among the p values of the regression coefficients of ethanol conversion and temperature under different catalyst combinations except for the A11 combination, all the data are less than 0.05. In other words, at the 95% confidence level, most of the regression coefficients are significantly different from 0. So we reject H_0 and accept H_1 , the significance test is passed.

It can be seen from TABLE 4 that the p-value of the regression coefficient between C4 olefin selectivity and temperature under different catalyst combinations is mostly less than 0.05. After removing the extreme data of A1, A2, A6, A10 and B4 can be different from 0 at the 90% confidence level and the regression coefficients of all the remaining data are different from 0 at the 95% confidence level. On the whole, we reject H_0 and accept H_1 at 90% confidence level, thus passing the significance test.

2.1.2. Spearman Correlation Analysis

2.1.2.1. Linear Trend Judgment

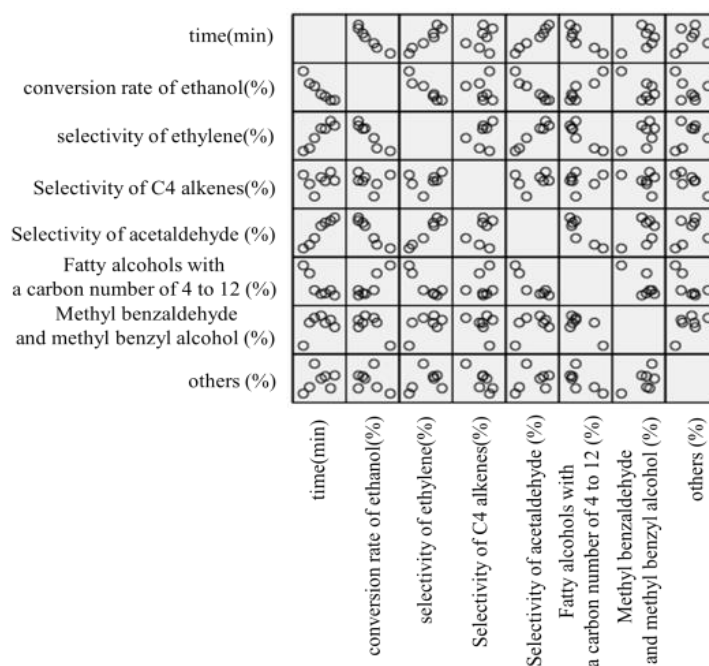


Fig. 1 Scatterplot of matrix between variables

It is assumed that there is a linear relationship between the variables in the combination of catalysts at 350 degrees so the hypothesis is tested by plotting the matrix scatter plot between the variables by SPSS. It can be seen that most of the variables do tend to be linearly correlated with each other and the linear correlation between some of them is very obvious. While, all groups of data need to satisfy normal distribution if Pearson correlation coefficient is used. Therefore, the Jarque-Bera test for normal distribution was performed next^[2]. Scatterplot of matrix between variables is shown in Fig. 1.

2.1.2.2. Jarque-Bera test for normal distribution

The Jarque-Bera test for normal distribution was run in Matlab on eight groups of data. It can be seen from the output that only the data of "methylbenzaldehyde and methylbenzyl alcohol" among the eight groups of data has an output h of 1, which means that this group of data accepts the hypothesis of normal distribution at 95% confidence level. The other seven groups of data have an

output h of 0, which means that the hypothesis of normal distribution is rejected at 95% confidence level. The hypothesis of normal distribution is rejected at 95% confidence level. The data do not conform to the normal distribution completely, so the Pearson correlation coefficient cannot be used for correlation analysis^[3].

The closer the Spearman coefficient nears to 1 or -1, the greater the correlations between the variables are^[4]. Watching the lower left side of the diagonal (the lower left and upper right side of this table are symmetric), it can be seen that six groups of variables have a Spearman coefficient of 0.9 and above. Three groups of variables are in the interval of [0.8,0.9) and two groups are in the interval of [0.7,0.8). To determine the significance of the data, hypothesis testing of the Spearman coefficient was performed next.

Taking the lower left of the diagonal as an example (the lower left and upper right in this table are symmetrical relations) for analysis. When p-value is 0.05, the Spearman coefficient is considered not significantly different from 0^[5]. Combining, it can be obtained that the data with Spearman's coefficient at the interval of [0.8,1) and 0.79 passed the Spearman's hypothesis test.

Analyzing TABLE 5 and TABLE 6 together, we found that for combination of catalysts at 350 degrees, significant correlations include time with ethanol conversion, ethylene selectivity, acetaldehyde selectivity, ethylene conversion with carbon number 4-12 fatty alcohol selectivity, ethylene selectivity with acetaldehyde selectivity, which are all positively correlated. And time with carbon number 4-12 fatty alcohol selectivity, ethanol conversion rate and ethylene selectivity, ethanol conversion rate and acetaldehyde selectivity, ethylene selectivity and carbon number for 4-12 fatty alcohol selectivity, and acetaldehyde selectivity and carbon number for 4-12 fatty alcohol selectivity, all of which are negatively correlated.

Table 5. Spearman's coefficient between variables

	Time	Ethanol conversion rate	Ethylene	C4 Olefins	Acetaldehyde	Carbon number of 4-12	Methylbenzyl alcohol	Other
Time	1.00	-0.99	0.93	0.21	1.00	-0.86	0.07	0.39
Ethanol conversion	-0.99	1.00	-0.95	-0.27	-0.99	0.81	-0.13	-0.32
C4 olefin	0.93	-0.95	1.00	0.39	0.93	-0.79	0.32	0.21
Ethylene	0.21	-0.27	0.39	1.00	0.21	-0.07	-0.18	-0.71
Acetaldehyde	1.00	-0.99	0.93	0.21	1.00	-0.86	0.07	0.39
Carbon number of 4-12	-0.86	0.81	-0.79	-0.07	-0.86	1.00	-0.21	-0.54
Methyl benzyl alcohol	0.07	-0.13	0.32	-0.18	0.07	-0.21	1.00	0.29
Others	0.39	-0.32	0.21	-0.71	0.39	-0.54	0.29	1.00

Table 6. The p-value of the Spearman coefficient between the variables

	Time	Ethanol conversion rate	Ethylene	C4 Olefins	Acetaldehyde	Carbon number of 4-12	Methylbenzyl alcohol	Other
Time	1.0000	0.0008	0.0067	0.6615	0.0004	0.0238	0.9063	0.3956
Ethanol conversion	0.0008	1.0000	0.0040	0.5508	0.0008	0.0381	0.7929	0.4746
C4 olefin	0.0067	0.0040	1.0000	0.3956	0.0067	0.0480	0.4976	0.6615
Ethylene	0.6615	0.5508	0.3956	1.0000	0.6615	0.9063	0.7131	0.0881
Acetaldehyde	0.0004	0.0008	0.0067	0.6615	1.0000	0.0238	0.9063	0.3956
Carbon number of 4-12	0.0238	0.0381	0.0480	0.9063	0.0238	1.0000	0.6615	0.2357
Methyl benzyl alcohol	0.9063	0.7929	0.4976	0.7131	0.9063	0.6615	1.0000	0.5560
Others	0.3956	0.4746	0.6615	0.0881	0.3956	0.2357	0.5560	1.0000

2.2. Exploring the degree of influence of temperature and catalysts

2.2.1. Overall data description

The data were first subjected to descriptive statistics as TABLE 7.

Table 7. Test data statistics for a given combination of catalysts at a temperature of 350 degrees

	Minimum value Maximum value	Minimum value Maximum value	Average value	Standard deviation
Time	20.00	273.00	153.29	91.68
Ethanol conversion	29.85	43.540	34.58	5.010
Ethylene	4.230	4.7600	4.520	0.200
C4 olefin	36.72	40.320	39.00	1.170
Acetaldehyde	5.170	8.7900	7.190	1.460
Carbon number is 4-12	30.86	39.700	33.64	3.460
Methyl benzyl alcohol	2.580	4.8000	4.140	0.740
Others	8.420	15.430	11.50	2.360

2.2.2. Charting

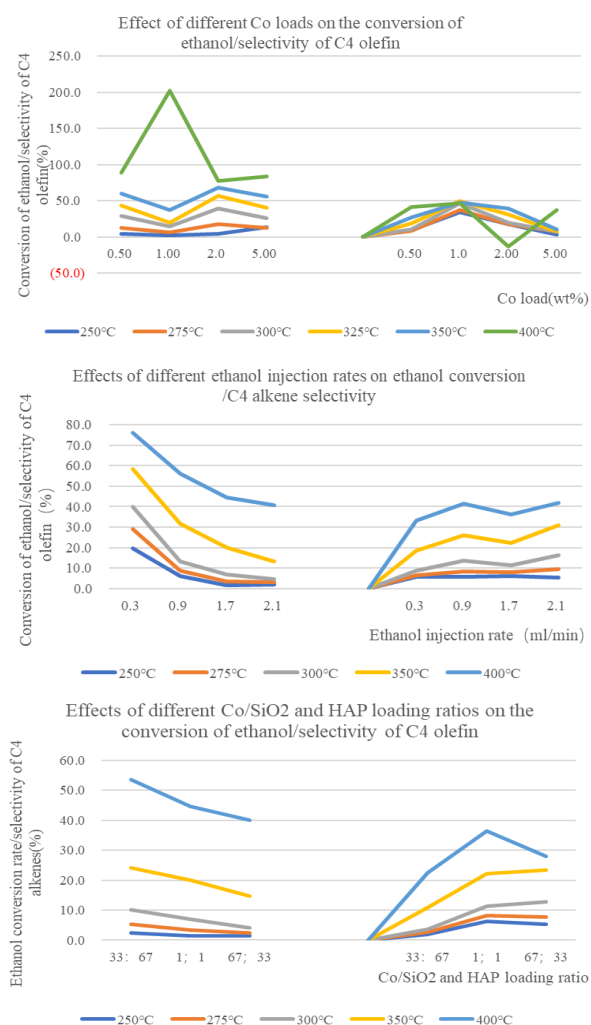


Fig. 2 Effect of different factors in catalysts on ethanol conversion/C4 olefin selectivity at different temperatures

Here the effect of temperature on the conversion of ethanol and the magnitude of C4 olefin selectivity is studied qualitatively and three lines graphs like Fig. 2 are drawn.

As can be seen from the three graphs above, all catalyst combination shows an increase in ethanol conversion and C4 olefin selectivity with the increase of temperature excluding few extreme data.

2.2.3. Fuzzy integrated evaluation model

- (1) The set of factors: $U=[\text{Co loading, Co/SiO}_2 \text{ and HAP loading ratio, ethanol injection rate}]$.
- (2) The set of comments: $V=[\text{high impact, low impact, average impact, low impact}]$.

(3) The set of weights: $A=[a_1,a_2,\dots a_m]$.

(the sum of each weight needs to be satisfied as 1).

According to some sources, the highest selectivity of C4 olefin is achieved when the loading ratio of Co/SiO₂ and HAP is 1:1 [6]. Therefore, we took Co/SiO₂ and HAP loading ratios as the factors with the greatest weight and Co loading and ethanol injection rate as the two factors with less weight. The final weight matrix is determined as $A=[0.25,0.25,0.25]$.

(4) The fuzzy comprehensive judgment matrix.

For the index u_i , the affiliation degree of each rubric is a fuzzy subset of the rubric set. The judgment of the indicator is noted as $R=[r_{i1},r_{i2},\dots r_{im}]$ [7]. Afterwards, the affiliation of each set of factors corresponding to the four sets of rubrics is determined from the image of the qualitative analysis plotted in the figure.

The graph shows that the ethanol conversion/C4 olefin selectivity does not show a monotonic increasing or decreasing trend with increasing Co loading at different temperatures but fluctuates up and down. It indicates that the factor of Co loading does not particularly affect the final result. So the affiliation of this factor for the four rubric sets can be given as $R1=[0.1,0.4,0.3,0.2]$. On the other hand, with the increase of ethanol injection rate, the ethanol conversion rate decreases obviously and the C4 olefin selectivity does not change much. So the affiliation of this factor for the four rubric sets can be given as $R2=[0.5,0.2,0.2,0.1]$. Finally, with the increase of the loading ratio, the ethanol conversion rate also decreases but the C4 olefin selectivity shows a fluctuating trend and the affiliation given by the first two factors can be given as $R3=[0.4,0.3,0.2,0.1]$. The fuzzy comprehensive judgment matrix of each index is finally derived as,

$$R = \begin{bmatrix} 0.1 & 0.4 & 0.3 & 0.2 \\ 0.5 & 0.2 & 0.2 & 0.1 \\ 0.4 & 0.3 & 0.2 & 0.1 \end{bmatrix} \quad (4)$$

(5) Fuzzy integrated evaluation. Since $B=A \times R$, the matrix multiplication is calculated by matlab to obtain $B=[0.375,0.275,0.225,0.125]$. The largest value of 0.375 is chosen as the result of the comprehensive evaluation and the corresponding impact degree is "high impact".

2.3. Exploration of optimal combination of catalyst and temperature based on neural network

We consider using the existing catalyst and temperature combination to make predictions for unknown catalyst and temperature combination and eventually find the combination with the highest c4 olefin yield. If the required temperature is below 350 degrees, the data for catalyst combinations and temperatures above 350 degrees are removed.

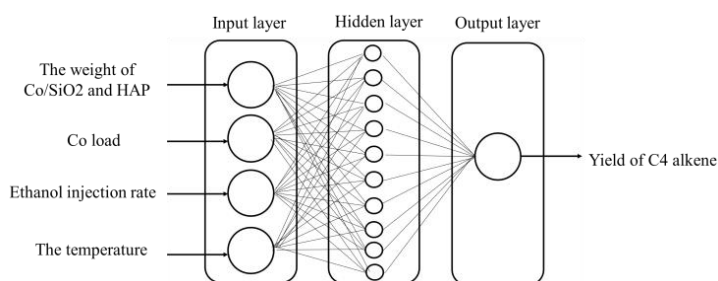


Fig. 3 Input layer, hidden layer, output layer in neural network

Mathematical forecasting methods include regression analysis forecasting, trend extrapolation forecasting, gray forecasting models and BP neural network forecasting [8]. BP neural network is a modern intelligent algorithm and a more widely used forecasting model that contains an input layer, hidden layer and output layer. The construction steps include preprocessing of data, number of nodes in the hidden layer, creation of forward neural network and BP network training [9]. The relationship between the input and output layers is obtained by simulating the training data and making a regression image while visualizing the good or bad results of the training data.

The Neural Net Fitting toolbox was used in matlab to construct the neural network model and the input layer was set to the mass of Co/SiO₂ and HAP, Co loading, ethanol injection rate and temperature to transmit our data information into the hidden layer. The output layer was set to the c4 olefin yield to transmit the results to the outside world as shown in Fig. 3.

After the neural network is fitted to the data, the neural network model "net" can be obtained in the matlab workspace. Fig. 4 shows that the neural network fit is good and the fit is good. Then we used the sim function in matlab to predict all possible combinations of catalyst and temperature and obtained the values of C4 olefin yield for size comparison and maximum value^[10].

Observing the results predicted by matlab, we found that a catalyst-temperature combination with a Co/SiO₂ and HAP mass ratio of 200 mg:200 mg, a Co loading of 2 wt%, an ethanol injection rate of 2.1 ml/min, and a temperature of 400 degrees should be chosen if the temperature size is not restricted. In other words, the combination of "200mg 2wt% Co/SiO₂-200mg HAP-ethanol concentration 2.1ml/min" and temperature of 400 degrees should be selected without temperature control and the maximum yield of C4 olefin was 45.259%.

If the temperature is lower than 350 ° C, remove the input layers at 350 ° C and 400 ° C, and predict again. The final result is that the catalyst is selected as "50mg 5wt% Co/SiO₂-50mg HAP ethanol concentration 0.3ml/min", the temperature is selected as 275 ° C, and the C4 olefin yield is 30.75%, which is the maximum.

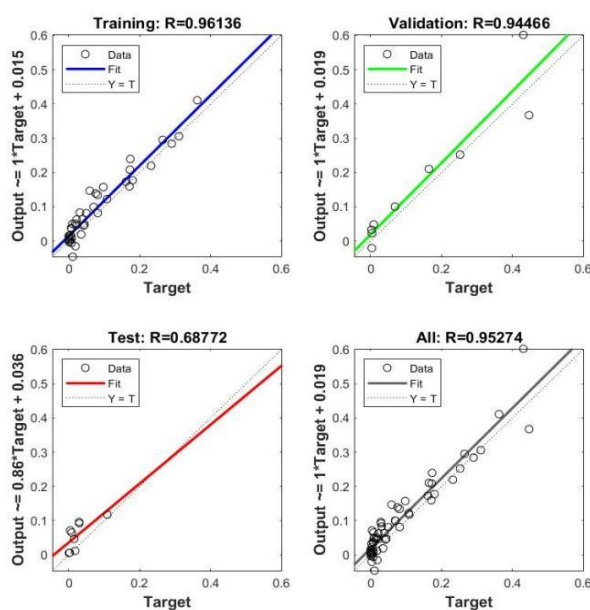


Fig. 4 The fitting effect of neural network is illustrated

3. Conclusion

In this paper, the regression analysis and correlation analysis between the independent variables and the dependent variables were conducted by using the one-dimensional linear regression model and Spearman's correlation coefficient and the corresponding significance test and goodness-of-fit test were conducted to quantitatively analyze the patterns existing between the data. We also drew the corresponding graphs by spss, excel, and matlab to observe the trend and pattern of data qualitatively and visually. We analyzed the data in both qualitative and quantitative dimensions to make the model more persuasive.

References

- [1] Wang P, Zhang QH, Jin B, et al. A rudder fault diagnosis algorithm based on one-dimensional linear regression analysis[J]. Firepower and Command Control, 2009, 34(7):20-23.

- [2] Zhang Wenyao. Measuring degree correlation of networks with Spearman coefficients[D]. Hefei, University of Science and Technology of China, 2016.
- [3] Sedgwick P. Pearson's correlation coefficient[J]. Bmj, 2012, 345.
- [4] Xu Wichao. A review of correlation coefficient research[J]. Journal of Guangdong University of Technology, 2012, 29(3): 12-17.
- [5] Hauke J, Kossowski T. Comparison of values of Pearson's and Spearman's correlation coefficient on the same sets of data[J]. 2011.
- [6] Lv S-P. Preparation of butanol and C₄ olefins by ethanol coupling[D]. Dalian University of Technology.
- [7] Shoukui Si. Mathematical Modeling Algorithms and Applications Exercise Solutions [M]. National Defense Industry Press,2015.
- [8] Zhu F. An introduction to prediction methods in mathematical modeling [J]. Science and Technology Information, 2010, 000(035): 842+862.
- [9] Zhang ZB. Research on probabilistic interval prediction of water demand in Liuzhou based on BP neural network[J]. People's Pearl River,2021,42(09):91-97.
- [10] Cong MY, Wang LP. Modern heuristic algorithm theory research[J]. High Technology Communication, 2003(05):105-110.