Atlantis Highlights in Intelligent Systems, volume 2
Proceedings of the 2021 International conference on Smart Technologies and Systems for Internet of Things (STS-IOT 2021)

Optimization Model Research Based on Partial Least Squares Method

Huilan Jin, Zhenrong Zhang, Liyan Xu and Hong Fang*

College of Basic Sciences, Tianjin Agricultural University, Tianjin 300384, China
**Corresponding author. Email: fanghong681019@126.com*

ABSTRACT

In this paper, taking the preparation of C4 olefins by ethanol coupling as an example, we discuss an optimization model based on partial least squares method. First of all, build a univariate polynomial regression model, and analyse the effect of temperature on the selectivity of ethanol conversion and C4 olefins. Secondly, calculate regression equations using partial least squares, and conduct a cross-validity test. The fitting effect is satisfactory; at last, An optimization model is established according to the correlation between the variables, and the catalyst combination and temperature are calculated when the C4 olefin yield is as high as possible. This model is great significant for studying the interdependencies and maximality problems between multiple correlation variables.

Keywords: *Unary polynomial regression models, Partial least squares, Optimization model, Cross-validity testing*

1. INTRODUCTION

In practical work, it is often necessary to study the interdependencies between two sets of multiple related variables and to study the use of a set of variables to predict a set of variables. Methods such as multiple linear regression analysis and principal component regression analysis are usually used. When there are multiple correlations between two sets of variables and the number of observed data is small, Partial least squares has advantages that other regression analyses do not have. In this paper, taking the preparation of C4 olefins by ethanol coupling as an example[1]-[3], an optimization model based on partial small multiplication is studied. The catalyst combination and temperature in which ethanol-coupled preparation of C4 olefins is prepared is designed to maximize the yield.

2. ETHANOL CONVERSION,C4 OLEFIN SELECTIVITY VS. TEMPERATURE

2.1. Effect of Temperature on Ethanol Conversion, Selectivity of C4 Olefins

Data are screened(Data from Appendix 1 of The 2021 China Higher Education Society Cup National College Students Mathematical Modeling Competition B),Find the A11 catalyst combination to replace the catalyst carrier HAP with quartz sand, We believe that the usability of the A11 group is too low, interfering with subsequent data analysis, so we decide to exclude the A11 group. Use 20 sets of data to make line charts.

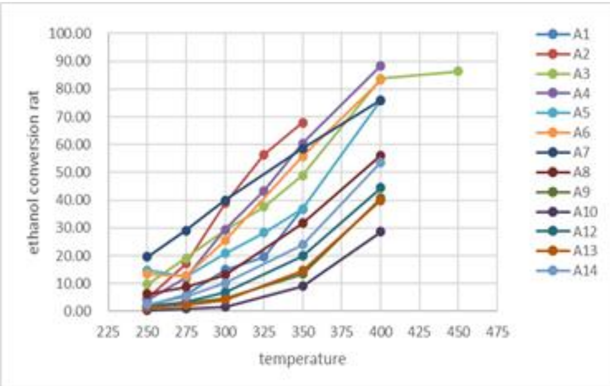

Figure 1. Charging method I Ethanol conversion rate vs. temperature

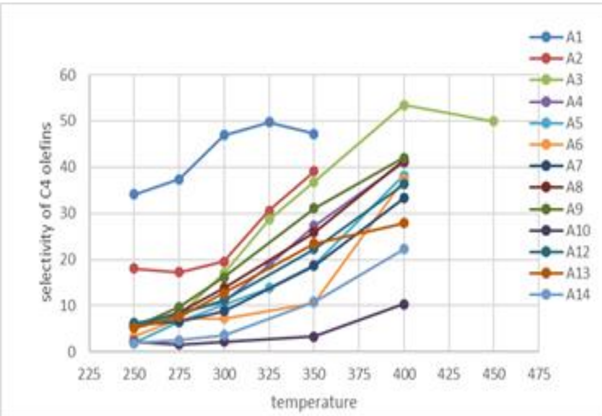

Figure 2. Charging method IC4 Selectivity of olefins vs. temperature

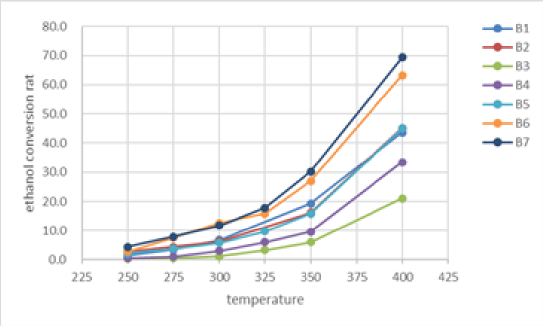

Figure 3. Charging method II Ethanol conversion rate vs. temperature

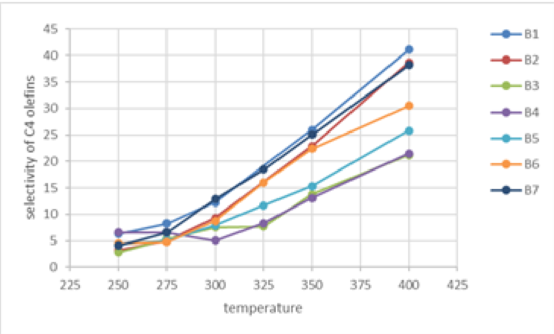

Figure 4. Charging method II Selectivity and temperature of C4 olefins

Observation Figures 1 to 4, in most catalyst combinations, Ethanol conversion shows an increasing trend with increasing temperature, The influence of the catalyst charging method on the ethanol conversion rate is not obvious. In order to obtain a more accurate ethanol conversion, C4 olefin selectivity vs. temperature, a univariate polynomial regression model is established.

2.2. Ethanol conversion ,C4 olefin selectivity vs. temperature

Based on MATLAB's toolbox debugging, we select 3 to the highest power.

The unary polynomial regression model is

$$y = a_1x^3 + a_2x^2 + a_3x + a_0 \qquad (1)$$

In the formula, a_1, a_2, a_3 are the regression coefficients, a_0 is an intercept. Bring the data in Annex 1 into the univariate polynomial regression model to obtain the regression coefficient, intercept and value of each equation R^2 . The results R^2 are close to 1, indicating that the effect of temperature on ethanol conversion and C4 olefins is significant. Therefore, Model (1) is reliable as a whole. Select representative scatter plots and regression curves for further analysis.

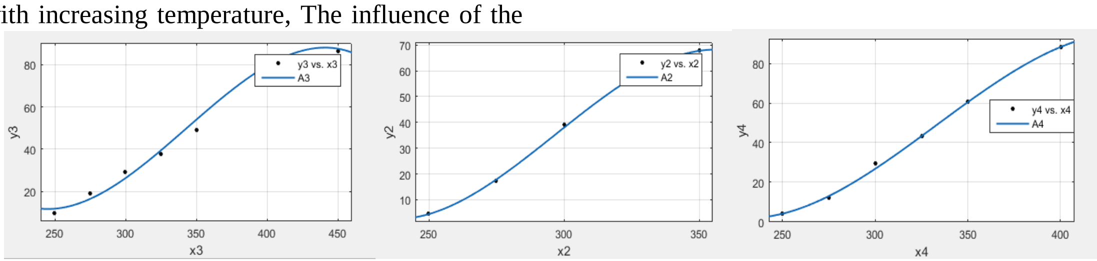

Figure 5. Catalyst A2、A3、A4 Regression curve of ethanol conversion versus temperature

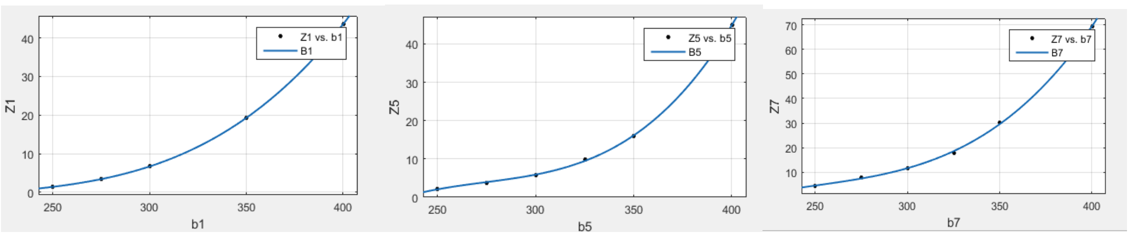

Figure 6. Catalyst B1、B5、B7 Regression curve of ethanol conversion versus temperature

As we can see from Figures 5 and 6: at 350 degrees below, the ethanol conversion rate grows slowly; Ethanol conversion between 350 and 400 degrees increases rapidly; A3 has the highest point between 400 degrees and 450 degrees.

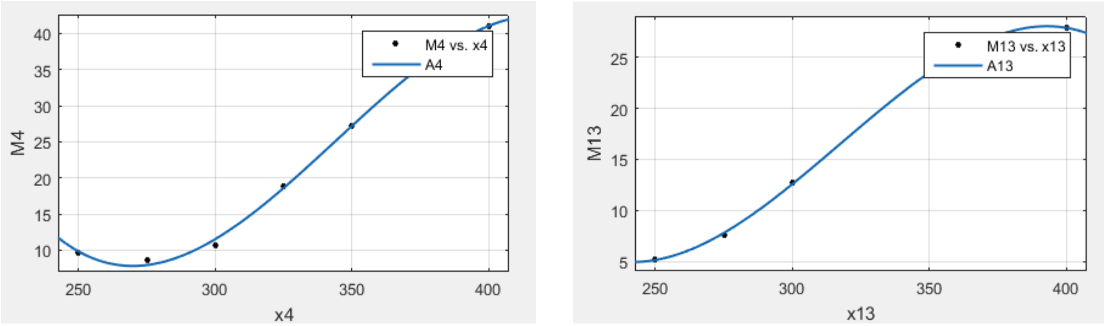

Figure 7. Catalyst A4、A13 Selectivity of C4 olefins vs. Regression curve for temperature

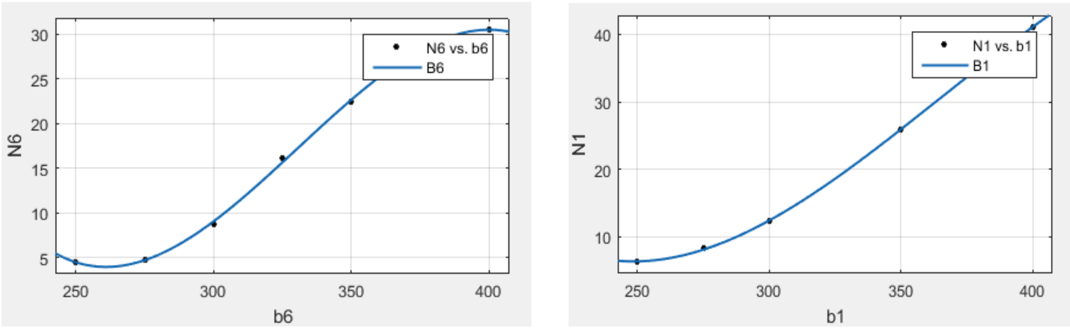

Figure 8. Catalyst B6、B1 the selectivity of C4 olefins is a regression curve for temperature

As we can see from Figure 7 and Figure 8: The selectivity of C4 olefins decreases with the increasing of temperature when the reaction temperature is in the range of 250 degrees and 275 degrees under the action of most catalysts; When the temperature is between 275 degrees and 325 degrees, the selectivity of C4 olefins slowly increases with the increase of temperature; When the temperature is between 325 degrees and 400 degrees, the selectivity of C4 olefins increases rapidly with increasing temperature. When most catalyst temperatures are around 400 degrees, the selectivity ofC4 olefins is taken to the maximum.

From the above analysis, it can be seen that the catalyst combination and temperature have a very large influence on ethanol conversion and the selectivity of C4 olefins. The effect of catalyst charging method on ethanol conversion and C4 olefin selectivity is not obvious, Therefore, the charging method of the catalyst is not discussed separately. In addition to the catalyst combination and temperature, the reaction time also affects the ethanol conversion as well as the selectivity of C4 olefins. According to the extensive information we know[4-5]: The effect of reaction time on the change in C4 olefin yield is much smaller than the effect of catalyst combination and temperature on the change in C4 olefin yield. Therefore, this article focuses on the effects of catalyst combination and temperature on ethanol conversion and selectivity of C4 olefins.

3. PARTIAL LEAST SQUARES REGRESSION MODEL

Partial least squares regression as a novel multivariate data analysis method is suitable for solving problems in multiple regression models where there are multiple correlations between variables and a small number of observations[6]-[11]. Our purpose is to explore the effects of different catalyst combinations and temperatures on ethanol conversion and selectivity of C4 olefins. The quality of the Co/SiO2,Co load, the quality of HAP, Ethanol concentration, temperature are set to be independent variables x_1, x_2, x_3, x_4, x_5 .Ethanol conversion rat, selectivity of C4 Olefin are set to be dependent variables y_1, y_2 .And then perform partial least squares regression.

3.1. Correlation Coefficient Matrix

First of all, We calculate the correlation coefficient matrix about the quality of Co/SiO2,Co load, the quality of HAP, Ethanol concentration, temperature, Ethanol conversion rat and the selectivity of C4 olefins.

TABLE 1 RELATIONAL NUMBER MATRIX

	Co/SiO_2 quality	Co load	HAP quality	Ethanol concentration	temperature	Ethanol conversion rat	Selectivity of C4 olefins
Co/SiO_2 quality	1	0.2003	0.9894	-0.2913	-0.0022	0.3861	0.3613
Co load	0.2003	1	0.2003	0.1486	-0.0148	0.0327	-0.1866
HAP quality	0.9894	0.2003	1	-0.2913	-0.0022	0.3925	0.3495
Ethanol concentration	-0.2913	0.1486	-0.2913	1	-0.0393	-0.3267	-0.0921
temperature	-0.0022	-0.0148	-0.0022	-0.0393	1	0.7774	0.7241
Ethanol conversion rat	0.3861	0.0327	0.3925	-0.3267	0.7774	1	0.7315
Selectivity of C4 olefins	0.3613	-0.1866	0.3495	-0.0921	0.7241	0.7315	1

As we can see from the correlation coefficient matrix, ethanol conversion and ethanol concentration, selectivity of C4 olefins and Co load, the selectivity of C4 olefins and ethanol concentration are inversely correlated. Ethanol conversion is strongly positively correlated with the selectivity of C4olefins.

3.2. Establishment the Partial Least Squares Regression Model

(1) Suppose that the variables obtained after standardizing the original data of 2 dependent variables y_1, y_2 and 5 independent variables x_1, x_2, x_3, x_4, x_5 are expressed as $\tilde{y}_1, \tilde{y}_2$ and $\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4, \tilde{x}_5$ respectively, and the two standardized observation data arrays are recorded as $\boldsymbol{F}_0$ and $\boldsymbol{E}_0$, $X = (\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4, \tilde{x}_5)^{\mathrm{T}}$, $Y = (\tilde{y}_1, \tilde{y}_2)^{\mathrm{T}}$ respectively.

(2) Find the eigenvector w_1 corresponding to the maximum eigenvalue of the matrix $\boldsymbol{E}_0^{\mathrm{T}} \boldsymbol{F}_0 \boldsymbol{F}_0^{\mathrm{T}} \boldsymbol{E}_0$, obtain the component $t_1 = w_1^{\mathrm{T}} X$, and calculate the score vector $\hat{t}_1 = \boldsymbol{E}_0 w_1$ and residual matrix $\boldsymbol{E}_1 = \boldsymbol{E}_0 - \hat{t}_1 \alpha_1^{\mathrm{T}}$ of the component t_1, where $\alpha_1 = \frac{\boldsymbol{E}_0^{\mathrm{T}} \hat{t}_1}{\|\hat{t}_1\|^2}$.

(3) Find the eigenvector w_2 corresponding to the maximum eigenvalue of the matrix $\boldsymbol{E}_1^{\mathrm{T}} \boldsymbol{F}_0 \boldsymbol{F}_0^{\mathrm{T}} \boldsymbol{E}_1$, obtain the component $t_2 = w_2^{\mathrm{T}} X$, and calculate the score vector $\hat{t}_2 = \boldsymbol{E}_1 w_2$ and residual matrix $\boldsymbol{E}_2 = \boldsymbol{E}_1 - \hat{t}_2 \alpha_2^{\mathrm{T}}$ of the component t_2, where $\alpha_2 = \frac{\boldsymbol{E}_1^{\mathrm{T}} \hat{t}_2}{\|\hat{t}_2\|^2}$.

(4) Repeat the above operation, find the eigenvector w_r corresponding to the maximum eigenvalue of the matrix $\boldsymbol{E}_{r-1}^{\mathrm{T}} \boldsymbol{F}_0 \boldsymbol{F}_0^{\mathrm{T}} \boldsymbol{E}_{r-1} (r \le 5)$, find the component $t_r = w_r^{\mathrm{T}} X$, and calculate the score vector $\hat{t}_r = \boldsymbol{E}_{r-1} w_r$ of the component t_r.

If it is determined that a satisfactory prediction model can be obtained by extracting a total of r components $t_1, t_2, \cdots, t_r$ according to the cross effectiveness, the ordinary least squares regression equation $\boldsymbol{F}_0$ on $\hat{t}_1, \hat{t}_2, \cdots, \hat{t}_r$ is obtained. It is

$$\boldsymbol{F}_0 = \hat{t}_1 \beta_1^{\mathrm{T}} + \hat{t}_2 \beta_2^{\mathrm{T}} + \cdots + \hat{t}_r \beta_r^{\mathrm{T}} + \boldsymbol{F}_r$$

Put $t_k = w_{k1}^* \tilde{x}_1 + w_{k2}^* \tilde{x}_2 + \cdots + w_{k5}^* \tilde{x}_5 (k = 1, 2, \cdots, r)$ into $Y = t_1 \beta_1 + t_2 \beta_2 + \cdots + t_r \beta_r$,

The partial least squares regression equation of two dependent variables can be obtained as follows:

$$y_j = a_{j1} x_1 + a_{j1} x_1 + \cdots + a_{j5} x_5 \quad (j = 1, 2)$$

3.3. Solve the Model

According to the cross validity, two components t_1, t_2 can be proposed。

The regression model of standardized variables $\tilde{y}_k$ with respect to components t_1, t_2 is as follows:

$$\tilde{y}_k = r_{1k} t_1 + r_{2k} t_2 \quad (k = 1, 2)$$

Since the component t_h can be written as a function of the standardized variable $\tilde{x}_j$ of the original variable, that is:

$$t_h = w_{1h}^* \tilde{x}_1 + w_{2h}^* \tilde{x}_2 + w_{3h}^* \tilde{x}_3 + w_{4h}^* \tilde{x}_4 + w_{5h}^* \tilde{x}_5 \quad (h = 1, 2)$$

Normalize the regression equation between the indicator variables as:

$$\tilde{y}_1 = 0.2145\tilde{x}_1 - 0.1252\tilde{x}_2 + 0.2104\tilde{x}_3 - 0.1023\tilde{x}_4 + 0.7422\tilde{x}_5 \quad (2)$$

$$\tilde{y}_2 = 0.1599\tilde{x}_1 - 0.1469\tilde{x}_2 + 0.1549\tilde{x}_3 - 0.0626\tilde{x}_4 + 0.7500\tilde{x}_5 \quad (3)$$

As we can see from equation (4), except that the coefficients x_2 , x_4 are negative, the other coefficients are positive, that is, the quality of CO / SiO2, the quality HAP , temperature and the ethanol conversion rate are positively correlated. The Co loading and ethanol concentration are negatively correlated with the ethanol conversion rate. Among them, the ethanol concentration has the greatest impact on the ethanol conversion rate, and the coefficient ratio x_1 with x_3 is close to 1.

Similarly, it can be found that in equation (5), except that the coefficient of x_2 , x_4 are negative, the other coefficients of x are positive. That is, the quality of CO / SiO2, the quality of HAP , temperature and the selectivity of C4 olefins are positively correlated .Co loading is negatively correlated with the ethanol conversion. Among them, Co loading has the greatest impact on the ethanol conversion, and the coefficient ratio x_1 with x_3 is close to 1.

3.4. Interpretation and Testing of Models

To test the model, we obtained a histogram of the regression coefficients and a prediction plot of ethanol conversion and C4 olefin selectivity[12].

Therefore, the partial least squares regression model established by the component t_1, t_2 is:

$$\begin{aligned}\tilde{y}_k &= r_{1k}(w_{11}^*\tilde{x}_1 + w_{21}^*\tilde{x}_2 + \cdots + w_{51}^*\tilde{x}_5) + r_{2k}(w_{12}^*\tilde{x}_1 + w_{22}^*\tilde{x}_2 + \cdots + w_{52}^*\tilde{x}_5) \\ &= (r_{1k}w_{11}^* + r_{2k}w_{12}^*)\tilde{x}_1 + (r_{1k}w_{21}^* + r_{2k}w_{22}^*)\tilde{x}_2 \\ &\quad + \cdots + (r_{1k}w_{51}^* + r_{2k}w_{52}^*)\tilde{x}_5\end{aligned}$$

The regression equation is obtained by reducing the standardized variables $\tilde{y}_k (k = 1, 2)$ and $\tilde{x}_i (i = 1, 2, 3, 4, 5)$ to the original variables y_1, y_2 and x_1, x_2, x_3, x_4, x_5 respectively:

$$y_1 = -84.2335 + 0.0696x_1 - 2.4345x_2 + 0.0682x_3 - 4.4588x_4 + 0.3269x_5 \quad (4)$$

$$y_2 = -45.3355 + 0.0304x_1 - 1.6762x_2 + 0.0295x_3 - 1.6026x_4 + 0.1939x_5 \quad (5)$$

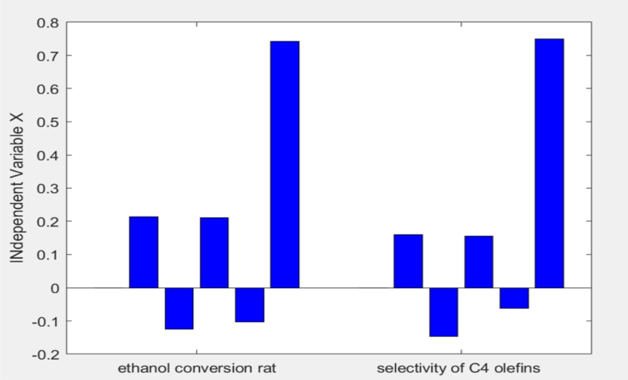

Figure 9. Histogram of the regression coefficients

It is immediately observable from the regression coefficient plot that temperature plays an extremely important role in explaining the 2 regression equations(2)and(3).However, the interpretive power of the charge ratio of Co/SiO2 to HAP differs little between the two regression equations. In order to investigate the accuracy of the regression equation, the prediction diagram is drawn for all sample points based on the coordinate values $(\tilde{y}_{ik}, y_{ik})$. $\tilde{y}_{ik}$ is the predicted value y_{ik} of the i -th sample point and the k -th variable. The forecast figure is shown in Figures 10 and 11 below.

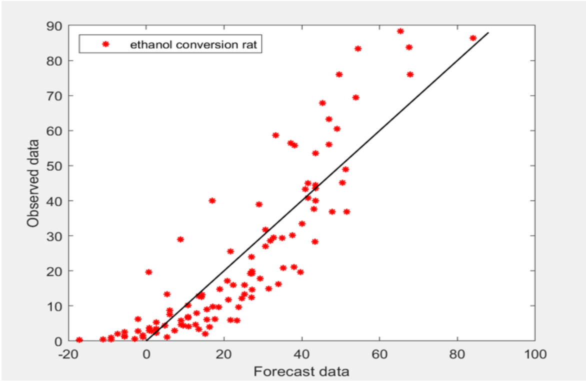

Figure 10. Predicted plot of ethanol conversion

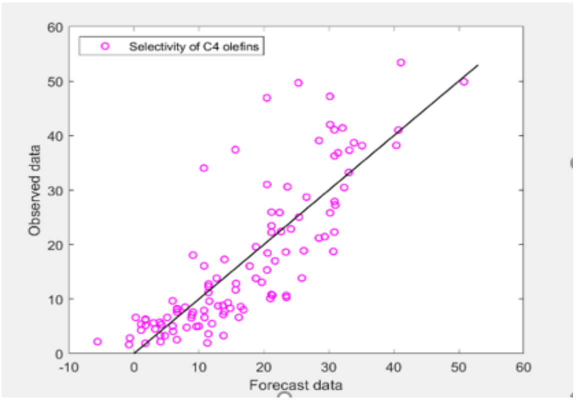

Figure 11. Prediction plot of C4 olefin selectivity

On this prediction plot, it can be seen that all points are basically evenly distributed near the diagonal of the plot, and the fitted value of the equation differs little from the original value, Note The fitting effect of this equation is satisfactory.

4. OPTIMIZE THE MODEL

4.1. Establishment of the Model

Because the value of C4 olefin yield is the product of ethanol conversion and C4 olefin selectivity, the maximum value of C4 olefin yield is taken as the target function, and an optimization model is established.

$$\max \quad z = y_1 \times y_2$$

$$\text{s.t.}$$

$$\begin{cases} y_1 = -84.2335 + 0.0696x_1 - 2.4345x_2 + 0.0682x_3 - 4.4588x_4 + 0.3269x_5 \\ y_2 = -45.3355 + 0.0304x_1 - 1.6762x_2 + 0.0295x_3 - 1.6026x_4 + 0.1939x_5 \\ 50 \le x_1 \le 200, \\ 0.5 \le x_2 \le 5, \\ x_1 / x_3 = 1, \\ 0.3 \le x_4 \le 1.68, \\ 350 \le x_5 \le 450. \end{cases}$$

Based on the previous analysis of the effects between variables, Co/Sio2 is charged at values between 50mg and 200mg, The value of the Co load amount is between 0.5wt% and 5wt%.Ethanol concentrations are between 0.3ml/min and 1.68ml/min, The Co/SiO2 to HAP charging ratio is 1:1,The temperature limit is between 3and 50 degrees and 450 degrees.

4.2. Solve the Model

Using lingo ,the solutions are $x_1 = 200$, $x_2 = 0.5$, $x_3 = 200$, $x_4 = 0.3$, $x_5 = 450$ and the maximum C4 olefin yield is $z = 45.89\%$. Namely, when the quality of CO / SiO2 and HAP are 200mg, the loading of CO is 0.5wt%, the concentration of ethanol is 0.3ml/min and the reaction temperature is 450 ℃, the performance of the catalyst is the best and the yield of C4 olefins is the highest.

5. CONCLUSION

The optimal model based on the least squares method was used to obtain c4 olefin yields The largest catalyst combination and temperature are basically consistent with the experimental data, which shows that our model is trustworthy. You can use this model to study the interdependencies and maximality problems between multiple correlated variables.

REFERENCES

[1] Xia W., Wang L. X., Wang J. G.,Qian C., Luo Z.S.,Chen K.. Comprehensive experiment on the conversion of ethanol into light of olefins over P/ZSM-5 zeolite catalyst. Experimental technology and management, 2020,37(02):201-204.

[2] Lv S. P.. Ethanol-coupled preparation of butanol and C4 olefins[D].Dalian University of Technology,2018.

[3] Ding D., Lu W.D.,Hou L.,Lu A.H..Preparation of fibrous BPO4/SiO2 catalyst for oxidative dehydrogenation of propane .Journal of Chemical Engineering: 2021, 72(11):5590-5597.

[4] Lian Q.Y., Yan P.,Jiang J.Y.,Ye C.X.. Correlation Analysis of Factors Affecitng Seed Yield of Introduced Xanthoceras Sorbifolia Germplasm. Modern agricultural science and technology, 2020(14): 113-116.

[5] Zhang J.W.. Studies on the ethanol to mixed hydrocarbons[D].Lanzhou University,2015.

[6] Ren W.. Optimization Applications of Multi-Objective Drug Mixture Data Based On Hyperspherical Transformation and Partial Least Squares [D].Shanxi Medical University,2016.

[7] J u, C.Y., Cai, T.J,. Li, W.H,. Sun, G,. Lei, C.L.. Improving forest burn severity estimations with partial least squares regression and orthogonal

signal correction methods in Daxing'an. Journal of Forestry Research,2021,32(03):1157-1165.

[8] Li Y.K., Shao X.G.,Cai W.S.. Partial Least Squares Regression Method Based on Consensus Modeling for Quantitative Analysis of Near-Infrared Spectra. Chemical Journal of Colleges and Universities, 2007, 28(2):246-249.

[9] Ni H.M., He Y. M.,Yang X. M.. Optimization model research on major underlying factors in the subhealth condition evaluation in 1 City and 7 Provinces in China. World Journal of Traditional Chinese Medicine, 2019, 5(2):88-94.

[10] Zhong Y. C., Zhang Z. Q., Chen J., Niu J., Shi Y. L.,Wang Y.,etc. Physicochemical properties, content, composition and partial least squares models of A. Trifoliata seeds oil.2021,12.

[11] Mou Y., Zhou L., Chen W. Z., Fan J. J.,Zhao X.. Maximum correntropy criterion partial least squares.2018, 165:137-147.

[12] Fan Y. D.. A Summary of Cross-Validation in Model Selection. Shanxi University, 2013.