Optimization of Ethanol Coupling to C_4 Olefins Based on Regression Analysis

Di Shao [1, *], Mengqi Lu [2], Xin Xu [2]

[1] College of Electrical Engineering and Automation, Fuzhou University, Fuzhou, China

[2] Maynooth International Engineering College, Fuzhou University, Fuzhou, China

* Corresponding Author Email: 1750938512@qq.com

Abstract. C_4 olefin is a basic organic chemical raw material, and ethanol is the raw material for its production. In this paper, the effects of different catalyst combinations and temperatures on the target product were analyzed, and a mathematical model was established to explore better preparation conditions. In this paper, the conversion of ethanol and the selectivity of C_4 olefins were used as independent variables, respectively, and the temperature was used as a dependent variable to carry out correlation analysis, and the Pearson coefficient was higher. Then it is necessary to study whether there is a significant difference under different cross levels of different control variables (catalyst combination, temperature), and then to determine whether multiple factors have a significant impact on the observed variables. The yield of C_4 alkene was taken as the objective function, and the optimal model was established under the constraint condition. Firstly, the regression models were established with Co loading, Co/SiO_2 and HAP loading ratio, ethanol concentration and temperature as independent variables, and ethanol conversion and C_4 olefin selectivity as dependent variables.

Keywords: Multivariate fitting, regression analysis, multivariate analysis of variance, control variable method, optimization model.

1. Introduction

C_4 olefin is a basic organic chemical raw material [1], which is widely used in chemical products and pharmaceutical production. Ethanol is a basic organic chemical raw material for the production of C_4 olefins, which is widely used in chemical products and pharmaceutical production. To speed up the preparation process and get more products, it is necessary to add catalysts to improve the reaction rate. A commonly used catalyst combination is a combination of Co loading, Co/SiO_2, HAP loading ratio, and ethanol concentration. Temperature and catalyst combination has a certain impact on the selectivity and yield of C_4 olefins, so it is of great practical significance to select the appropriate catalyst combination and explore the process conditions of catalytic coupling of ethanol to C_4 olefins. An important raw material for olefins. To speed up the preparation process and get more products, it is necessary to add catalysts to improve the reaction rate. A commonly used catalyst combination is a combination of Co loading, Co/SiO_2, HAP loading ratio, and ethanol concentration. Temperature and catalyst combination has a certain impact on the selectivity and yield of C_4 olefins, so it is of great practical significance to select the appropriate catalyst combination and explore the process conditions of catalytic coupling of ethanol to C_4 olefins. In this paper, the conversion of ethanol and the selectivity of C_4 olefins were used as independent variables, the temperature was used as the dependent variable to analyze the correlation, and the Pearson coefficient was higher. Then, to study whether there is a significant difference under different cross levels of different control variables (catalyst combination, temperature), and then judge whether multiple factors [2] have a significant impact on the observed variables. Through the multi-factor analysis of variance, the significance level was judged according to the sign given by the F distribution table, and through the profile, it was judged that Co loading, ethanol concentration and the temperature had significant effects on ethanol conversion and C_4 olefin selectivity. The yield of C_4 alkene was taken as the objective function, and the optimal model was established under the constraint condition. Firstly, regression models were established with Co loading, Co/SiO_2 and HAP loading ratio, ethanol concentration and temperature as independent variables, and ethanol conversion and C_4 olefin selectivity as dependent variables.

2. Modeling of different catalyst Combinations

2.1. Build a model

A linear regression equation $\hat{y} = \hat{a} + \hat{b}x$ was established with ethanol conversion and C_4 olefin as independent variables x and temperature as dependent variable y, the expression is as formula (1):

$$\begin{gathered} \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad S_{xx} = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \\ S_{yy} = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 \quad S_{xy} = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} \end{gathered} \tag{1}$$

The following test statistic as shown in formula (2).

$$T = \frac{\hat{b}}{\hat{\sigma}}\sqrt{S_{xx}} \tag{2}$$

Where: $\hat{\sigma} = \sqrt{\frac{1}{n-2}(S_{yy} - \hat{b}^2 S_{xx})}$

When $x = x_0$, the confidence level of y is $1-\alpha$ and the prediction interval as shown in formula (3).

$$\begin{gathered} (\hat{a} + \hat{b}x_0 - t_{\frac{\alpha}{2}}(n-2)\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \\ \hat{a} + \hat{b}x_0 + t_{\frac{\alpha}{2}}(n-2)\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}\ ) \end{gathered} \tag{3}$$

When the significance test meets the requirements [3], the linear regression equations of ethanol conversion and C_4 olefin selectivity with temperature can be successfully established for each catalyst combination.

Pearson correlation coefficient was used to measure the degree of correlation between ethanol conversion and temperature, as well as between C_4 olefin selectivity and temperature. As shown in formula (4).

$$\rho(x, y) = \frac{E[(X - \mu x)(Y - \mu y)]}{\sqrt{\sum_{i=1}^{n}(X_i - \mu x)^2}\sqrt{\sum_{i=1}^{n}(Y_i - \mu y)^2}} \tag{4}$$

When r>0, the two variables are positively correlated; when r<0, they are negatively correlated; and when r=0, they are not linearly correlated. When r=1 and -1, it means that the two variables x and y can be well described by a straight line equation, and all the sample points are well on a straight line, and the closer to 1 and -1, the stronger the correlation between the two variables.

2.2. Model solving

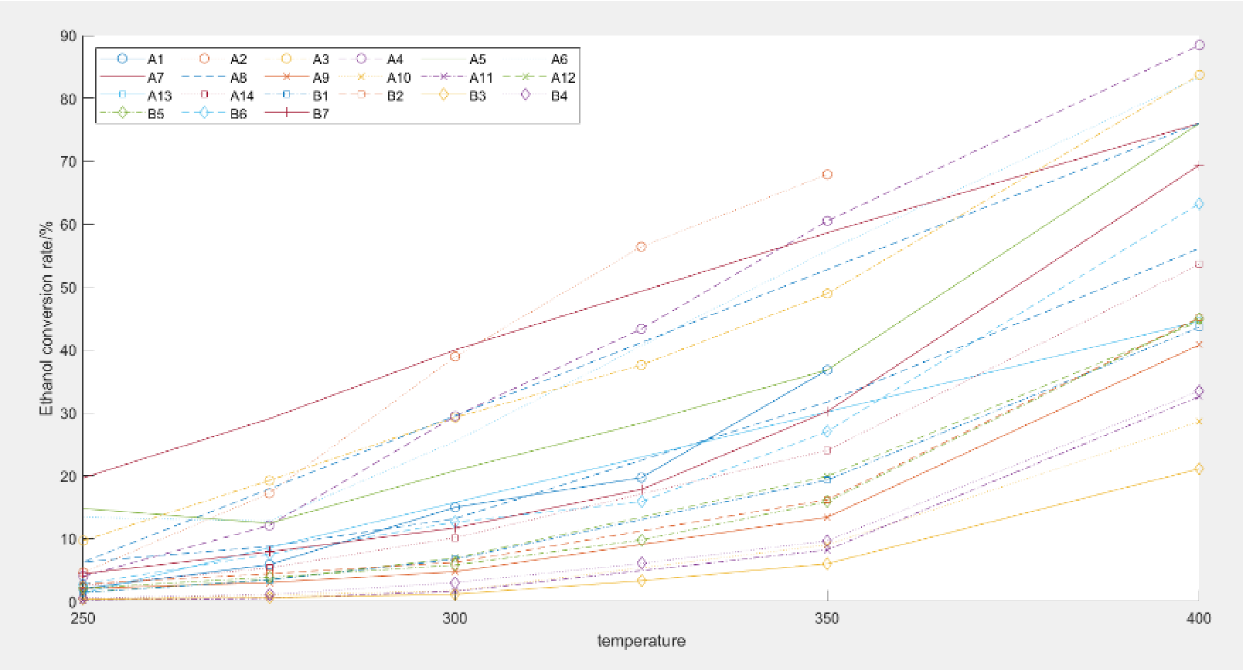


Figure 1. Ethanol Conversion Rate vs. Temperature.

It can be seen from the figure 1 that the conversion rate of ethanol has a significant positive correlation [4] with temperature under various catalyst combinations, that is, the conversion rate of ethanol also increases with the increase of temperature.

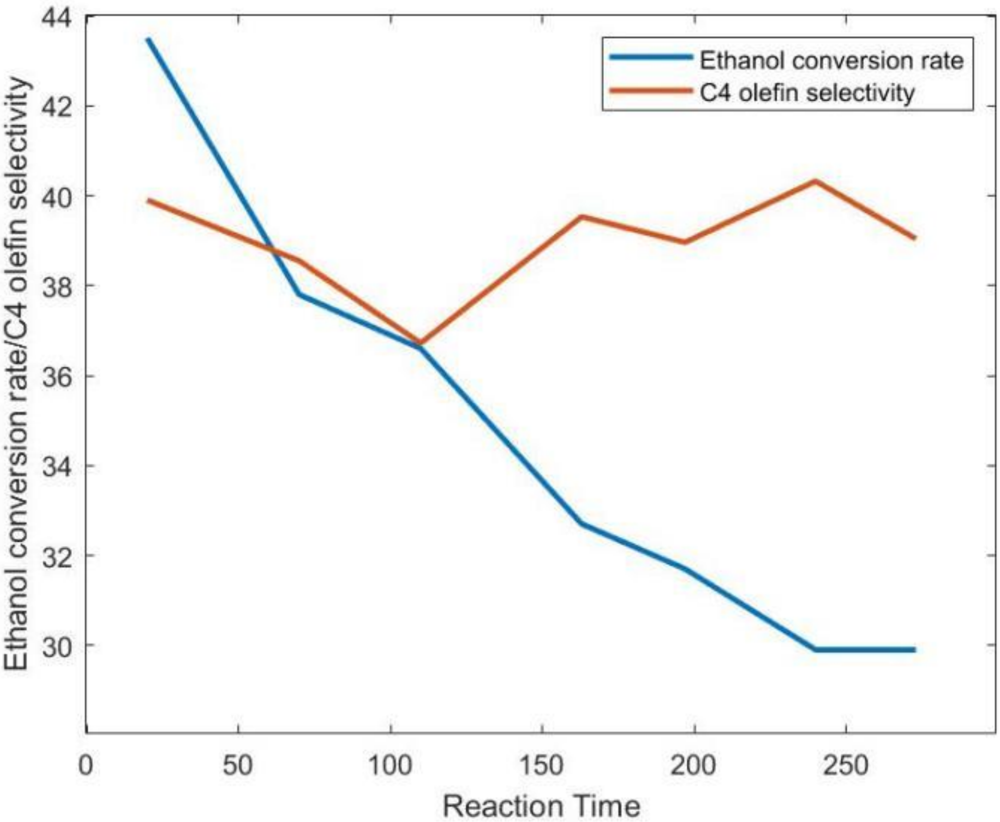


Figure 2. Temperature dependence of C_4 olefin selectivity.

It can be seen from the figure 2 that the C_4 olefin selectivity changes little with time, while the ethanol conversion rate decreases gradually with the increase of reaction time.

3. Variable Analysis Based on Multifactor Analysis of Variance

3.1. Build a model

Catalyst combination and temperature affect C_4 olefin selectivity and C_4 olefin yield. The problem of determining whether multiple factors have a significant effect on an observed variable applies to a multivariate analysis of variance. The mathematical expression for the sum of squares of deviations of the dependent variable, as shown in formula (5)-(8).

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{r}\sum_{k=1}^{n_{ij}}\left(x_{ijk} - \overline{x}\right)^2 \tag{5}$$

x_{ijk} is the k sample value at the i level of the control variable A and the j level of the control variable B; n_{ij} is the number of samples at the i level of control variable A and the j level of control variable B, and $\overline{x}$ is the mean value of the observed variable.

$$SSA = \sum_{i=1}^{k}\sum_{j=1}^{r} n_{ij}\left(\overline{x}_i^A - \overline{x}\right)^2 \tag{6}$$

$\overline{x}_i^A$ Is the mean of the observed variable at the i level of the control variable A

$$SSB = \sum_{i=1}^{r}\sum_{j=1}^{k} n_{ij}\left(\overline{x}_j^B - \overline{x}\right)^2 \tag{7}$$

$\overline{x}_j^B$ Is the mean of the observed variable at the j level of the control variable B

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{r}\sum_{k=1}^{n_{ij}}\left(x_{ijk} - \overline{x}_{ij}^{AB}\right)^2 \tag{8}$$

$\overline{x}_{ij}^{AB}$ Is the mean value of the observed variables at the i, j level of and for the control variables A and B

The ratio obtained by the F test determines whether the influence of the independent variable on the dependent variable reaches the standard of significant level difference, and improves the experimental process.

Firstly, the effects of various catalyst combinations and temperature on the conversion rate of ethanol were studied.

The significance test of the calibration model [5] is less than the significance level of 0.05, so the null hypothesis is rejected, so it is effective to use linearity to fit the model. Co loading, ethanol concentration, temperature, and the interaction properties of Co loading and temperature, ethanol concentration and temperature are all $p=0.000<0.05$, indicating that Co loading, ethanol concentration, temperature, Co loading and temperature, ethanol concentration and temperature all had significant effects on ethanol conversion. Variables with significance greater than 0.05 were excluded, and the model test was reconstructed. The results are as Table 1.

Table 1. Inter-subject effect test after the first elimination.

Source	Class III sum of squares	Degrees of freedom	Mean Square	F	Significance
Modified model	49202.483a	41	1200.061	10.043	.000
Intercept	39223.178	1	39223.178	328.246	.000
Co Load * Temp	2700.226	14	192.873	1.614	.098
Ethanol Concentration * Temperature	2128.794	15	141.920	1.188	.303
Co loading	4568.033	3	1522.678	12.743	.000
Ethanol Concentration	3442.863	3	1147.621	9.604	.000
Temperature	18475.846	6	3079.308	25.770	.000
Error	8006.044	67	119.493		
Total	112891.149	109			
Total after Correction	57208.527	108			

Catalyst combination and temperature [6] affect C_4 olefin selectivity and C_4 olefin yield. The problem of determining whether multiple factors have a significant effect on an observed variable applies to a multivariate analysis of variance. The mathematical expression for the sum of squares of deviations of the dependent variable.

The significance of Co loading, ethanol concentration and temperature was less than 0.05, but the significance probability level of the interaction between Co loading and temperature, ethanol concentration and temperature was greater than 0.05, so they were excluded. The significance of Co loading, ethanol concentration and temperature was less than 0.05.

The independence of the variables is checked by the model, and the normality test and the homogeneity of variance can be seen by the residual plot. As shown in figure 3.

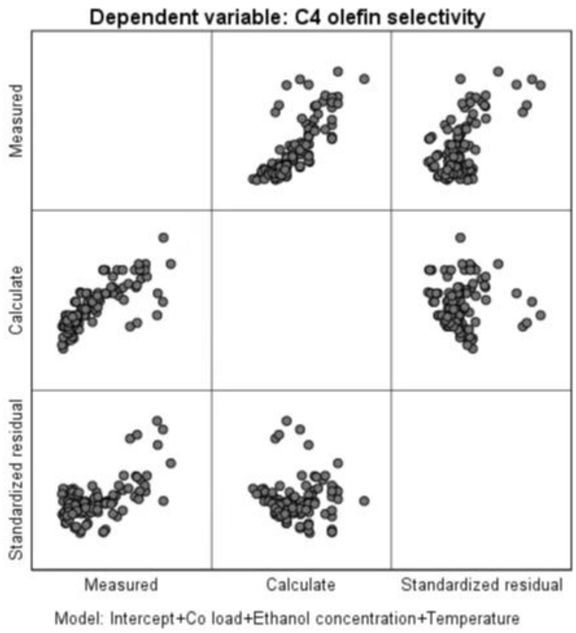
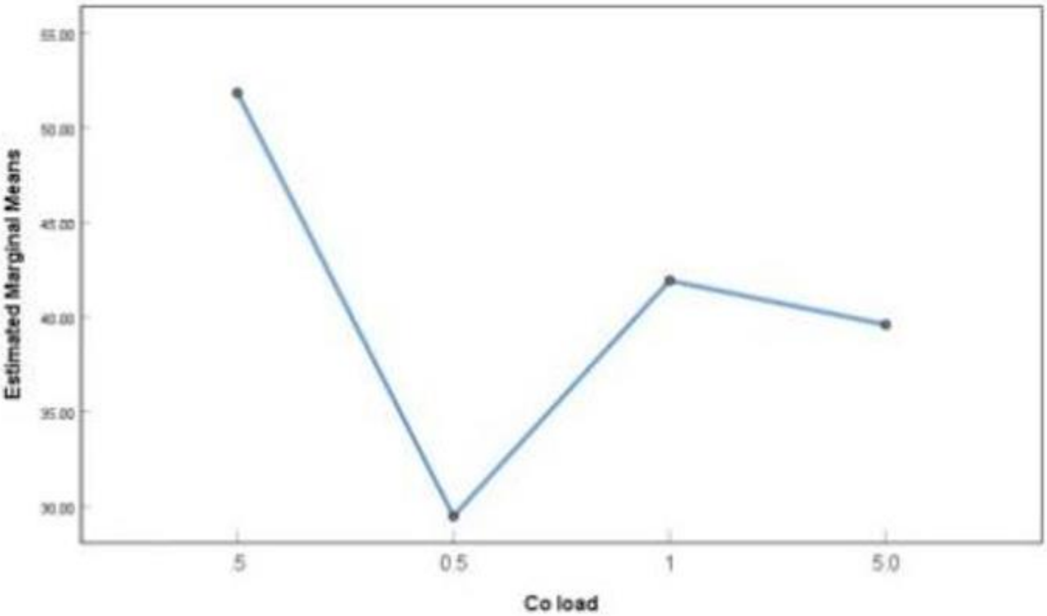
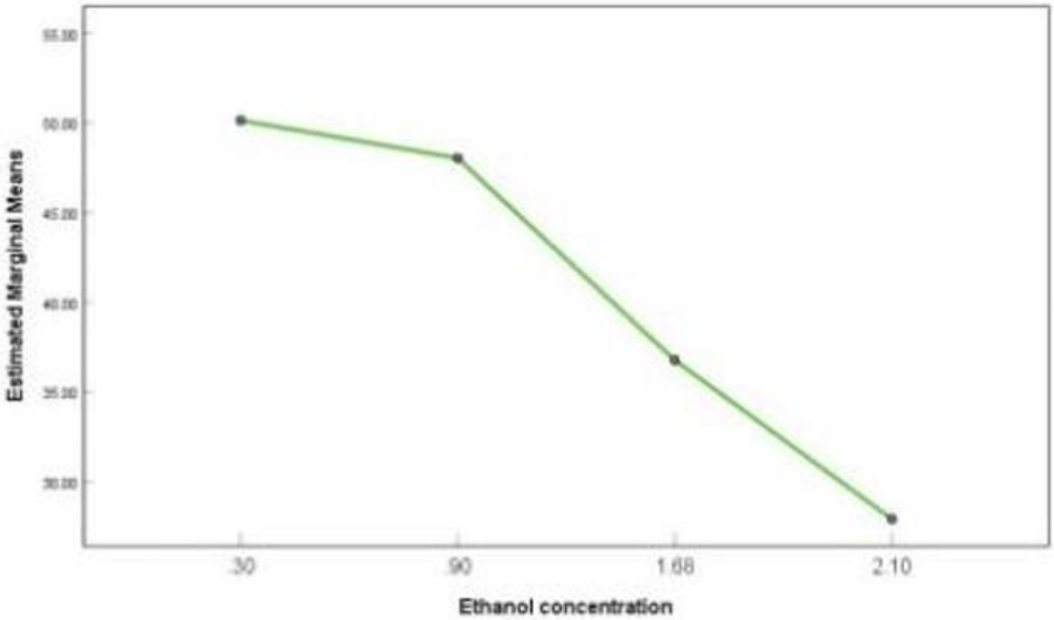


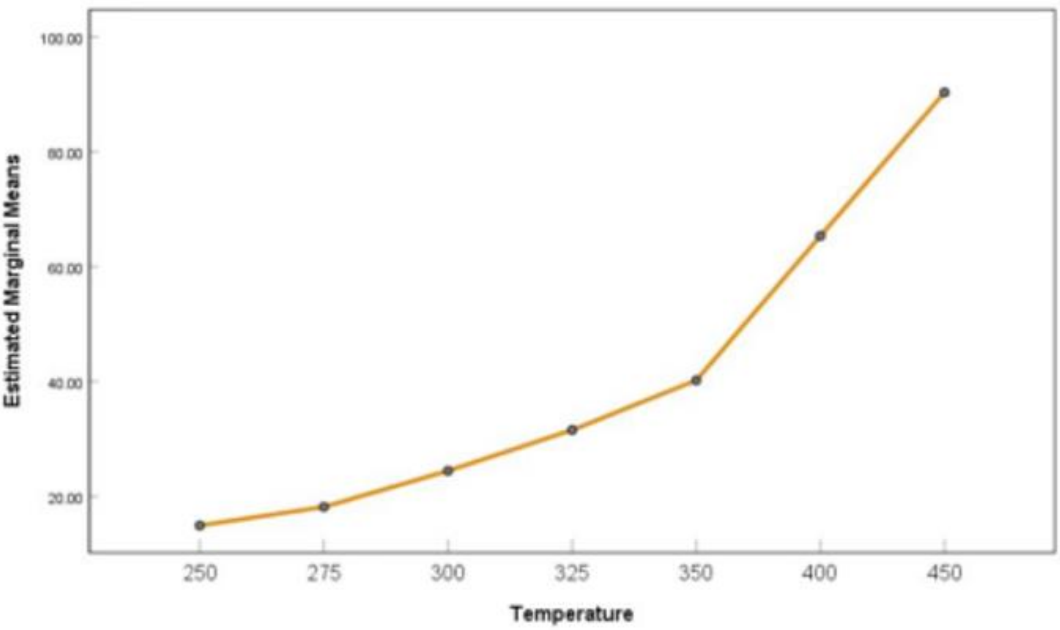
Figure 3. Residual Plot of Normality Test and Homogeneity of Variance.

All the points in the residual chart are included, and the normal test and homogeneity of variance are satisfied, so the multi-factor analysis model is applicable.

Finally, the marginal mean is estimated, that is, after controlling other factors, the change of the dependent variable is only under the use of one factor [7]. In ordinary analysis, the change of the dependent variable is the result of several factors. A contour plot based on the estimated marginal mean is as figure 4.


Estimated Marginal Means
55.00
50.00
45.00
40.00
35.00
30.00
.5
0.5
1
5.0
Co load


Estimated Marginal Means
55.00
50.00
45.00
40.00
35.00
30.00
.30
.90
1.68
2.10
Ethanol concentration


Estimated Marginal Means
100.00
80.00
60.00
40.00
20.00
250
275
300
325
350
400
450
Temperature

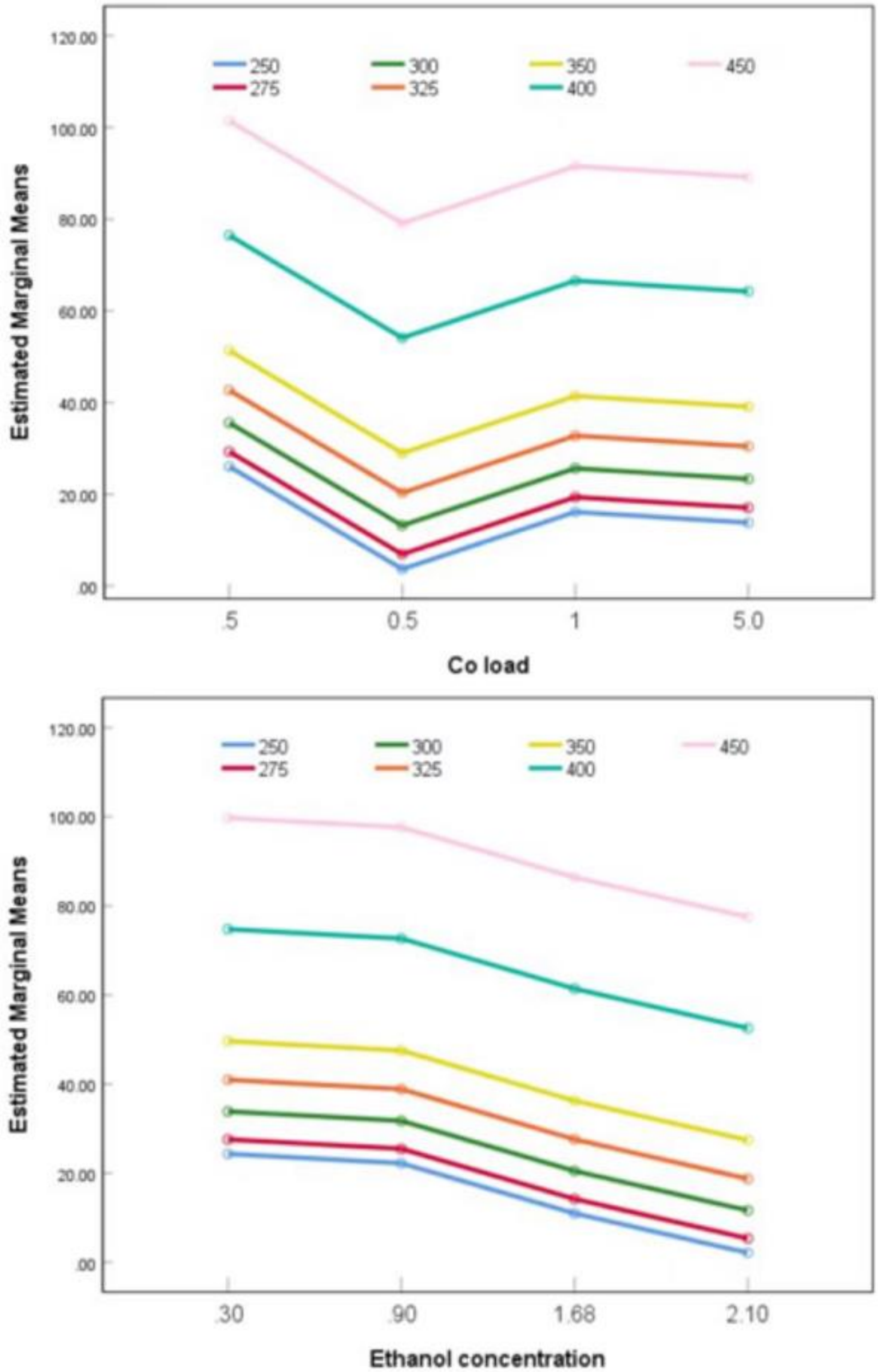


Figure 4. Estimated Marginal Mean Profile.

The effects of loading, ethanol concentration and temperature on ethanol conversion are consistent with the real experimental data and literature data. In the same way, the effects of various catalyst combinations and temperatures on the selectivity of C_4 olefins were studied. Finally, it is concluded that 400℃ is the suitable temperature for the catalytic reaction.

4. C4 Olefin Based on Optimal Model

4.1. Establishment of the model

According to the title [8], the catalyst combination and temperature should be selected to make the yield of C_4 olefins as high as possible. Therefore, an optimization model was established with C_4

olefin yield as the objective function, and the constraint conditions were obtained by analyzing the data, and then the model was solved. The optimization model is as figure 5.

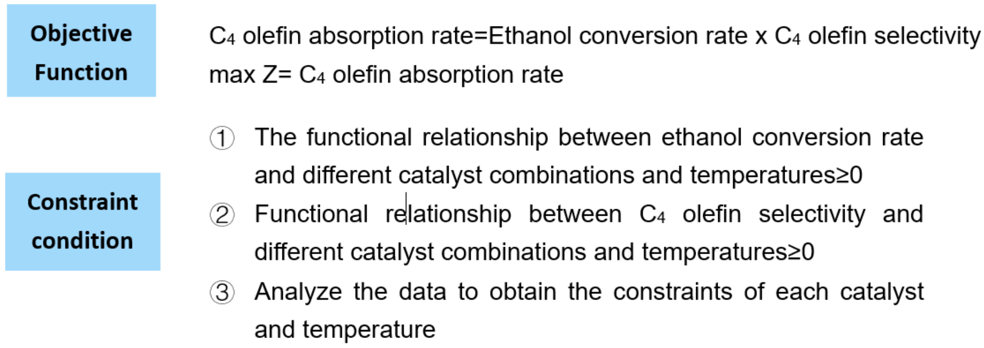


Figure 5. Optimization Model of Reaction Conditions for Finding the Highest Olefin Yield.

For the relationship of catalyst combination and temperature with ethanol conversion and C_4 olefins in the optimization model, the data are processed on the basis of the data preprocessed in the second problem. It is considered that the mass of Co/SiO_2 and the mass of HAP in the catalyst combination can be expressed by the loading ratio of Co/SiO_2 and HAP. Therefore, the Co loading, the loading ratio of Co/SiO_2 to HAP, ethanol concentration and temperature were used as independent variables $x_1, x_2, \cdots, x_5$ and the ethanol conversion and C_4 olefins were used as dependent variables y, respectively, to perform regression analysis. The steps of regression analysis as formula (9).

For the observation data $(y_i, x_{i1}, x_{i2}, \cdots, x_{i5}), \quad (i = 1, 2, \cdots, 109)$.

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{109} \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{15} \\ 1 & x_{21} & x_{22} & \cdots & x_{25} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1091} & x_{1092} & \cdots & x_{1095} \end{pmatrix}, \quad B = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_5 \end{pmatrix} \tag{9}$$

It is estimated by the least square method: $\hat{B} = (X^T X)^{-1} X^T Y$

So call $\hat{y} = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \cdots + \hat{b}_5 x_5$ the linear regression equation of the variable with respect to the variable $x_1, x_2, \cdots, x_5$.

That is, testing the hypothesis $H_0 : b_1 = b_2 = \cdots = b_5 = 0$.

Let $S_R = \sum_{i=1}^{109} (\hat{y}_i - \bar{y})^2$, $S_e = \sum_{i=1}^{109} (y_i - \hat{y}_i)^2$, then test statistic $F = \frac{S_R / 5}{S_e / (103)} \sim F(5,103)$ .To a small probability α, if $F > F_\alpha(5,103)$, reject H_0, it is considered that the established regression equation is correct.

With Co loading, ethanol concentration, temperature, and loading ratio as independent variables, marked as x_1, x_2, x_3, x_4, and with ethanol conversion and C_4 olefin selectivity as dependent variables, marked as y_1 and y_2, respectively. The relationships between the conversion rate of ethanol and the selectivity of C_4 olefins and four independent variables were established respectively, and two functional relationships $y_1 = f(x_1, x_2, x_3, x_4, x_5)$ and $y_2 = g(x_1, x_2, x_3, x_4, x_5)$ were obtained. The extraction Co loading, ethanol concentration, temperature, loading ratio, ethanol conversion and olefin selectivity were analyzed by SPSS.

The linear correlation between the conversion rate of ethanol and the four catalyst elements R^2=0.692, so the stool tool in MATLAB is used for pure quadratic multiple binomial regression. The root means square error is 12.4, which is relatively small for linear regression. The regression equation as formula (10):

$$y_1 = 60.75 - 0.96x_1 + 0.175x_2 - 0.52x_3 + 11.3x_4 + 0.5x_1^2 - 5.67x_2^2 + 0.0013x_3^2 - 6.2x_4^2 \quad (10)$$

The linear correlation between the selectivity of olefin and the ratio of four catalyst elements R^2=0.791. Therefore, the method of linear regression is adopted. The final regression equation as formula (11).

$$y_2 = -1.926x_1 + 1.024x_2 + 0.286x_3 + 1.893x_4 - 39.355 \quad (11)$$

4.2. Solution of the model

To solve the optimization model [9] successfully, we use the screening function of EXCEL and find out the interval of the optimal solution [10] of the corresponding variables.

(1) Discuss the Co optimal loading range, combined with the literature [1] and the fact that before 350 degrees, the main product is fatty alcohol with a carbon number of 4-12, so the temperature can be fixed at 400 degrees, the Co optimal loading is about 1 wt%, thus fixing the optimal range.

(2) Discuss the charge ratio interval of the optimal sum of Co/SiO_2 and HAP. According to the analysis and conclusion in (1), the optimal charge ratio of the optimal Co/SiO2 and HAP is about 1, so the optimal interval is fixed.

(3) Discuss the optimal range of ethanol concentration: the rest variables can be fixed from the first two analyses, to observe the ethanol concentration that maximizes the C_4 olefin yield. The optimal ethanol concentration interval was approximately 0.9 ml/min, thus fixing the optimal interval.

According to the model 2[2], the catalyst combination and temperature are selected to make the yield of C_4 olefins as high as possible under the same experimental conditions. A nonlinear optimization model was established with the yield of C_4 olefin as the objective function, and the two regression equations established above and the optimal solution interval were sought as the constraint conditions to solve. The optimization model established for unlimited temperature is as formula (12).

$$\max y = y_1 y_2$$
$$\begin{cases} y_1 = 60.75 - 0.96x_1 + 0.175x_2 - 0.52x_3 + 11.3x_4 + 0.5x_1^2 - 5.67x_2^2 + 0.0013x_3^2 - 6.2x_4^2 \\ y_2 = -1.926x_1 - 1.024x_2 + 0.286x_3 + 1.893x_4 - 39.355 \\ y_1 \geq 0 \\ y_2 \geq 0 \\ 1 \leq x_1 \leq 2 \\ 0.3 \leq x_2 \leq 2.1 \\ 250 \leq x_3 \leq 400 \\ 1 \leq x_4 \leq 2 \end{cases} \quad (12)$$

The maximum yield of 4516.787 was obtained by Lingo. Based on the optimization model of the first problem, the temperature is limited to below 350 degrees by the constraint condition. The maximum yield of 2404.631 was obtained by Lingo. As shown in figure 6.

```
Local optimal solution found at iteration:          67
Objective value:                                  4516.787

                    Variable           Value        Reduced Cost
                          Y1        60.90613            0.000000
                          Y2        74.15981            0.000000
                          X1        1.000000            0.000000
                          X2       0.9000000            0.000000
                          X3        400.0000            0.000000
                          X4        1.036668            0.000000
```

Figure 6. Solution of an optimization model for maximum yield of C_4 olefin.

5. Conclusions

This article through the literature material, the understanding of chemistry knowledge, and the utilization of chemistry commonly used tools carry on the correlation analysis, the research has the significant difference in the different control variables (catalyst combination, temperature) under the different cross-level, and take the C_4 alkene yield as the goal function, seeks the restraint condition to establish the most superior model. Ethanol conversion and olefin selectivity were used as dependent variables to establish regression models respectively. By repeating five experiments, the relationship between ethanol conversion and C_4 olefin yield with time was observed, and the time for C_4 olefin yield to reach the maximum was determined. In this paper, the model is established based on chemistry, and the established optimization model has strong applicability and is suitable for popularization and utilization. However, this model uses a small amount of data, in future work, data can be augmented by interpolation, making the degree of fitting better.

References

[1] Lv Shaopei. Preparation of Butanol and C_4 Olefin by Coupling of Ethanol [D]. Dalian University of Technology, 2018.

[2] Si Shoukui and Sun Zhaoliang, Mathematical Modeling Algorithms and Applications (SeCond Edition). National Defense Industry Press, 2016

[3] Application of Lingo and Excel in Mathematical Modeling, edited by yuan Xinsheng, Shao Dahong and Yu Shilian. Science Press, 2007

[4] Ge Yuena. Study on selective polymerization of C4 olefins [D]. East China University of Science and Technology, 2019.

[5] Li Ruige, Zhao Zhonglin, Zhang Shuli, Li Xin, Li Jianwei, yuan Zhihua, yuan Chao. Application of SPSS software in experimental design and data processing [J]. Agricultural Network Information, 2016 (11): 127-129.

[6] Wang Peigang, Liang Jing, Zhang Gangming. Multivariate statistical analysis and SAS implementation. Wuhan University Press, 2020, 11.338.

[7] Zhang Wenjie, yuan Hongping. Research on selection of energy-saving service Companies based on multi-objective weighted grey target decision model [J]. China Management Science, 2019, 27 (02): 179-186. DOI: 10.16381/J. CNKI. Issn1003-207x.2019.02.018.

[8] Liu Lin. Matrix-based introduction to multivariate data analysis, by KoheiAdachi 2nd edition. Singapore: Springer Nature, 2020. pp. 457. [J]. Biometrics, 2021, 77(4).

[9] Wei Cui, Junyan Huang, Huifang Song, Ming Xiao. Development of two new anti-washout grouting materials using Multi-way ANOVA in Conjunction with Grey relational analysis [J]. Construction and Building Materials, 2017,156.

[10] Li Ruige, Zhang Mengjiao, Wu Lulu, Lv Dongcan, Shi Lijun, An Wankai, Zhao Zhonglin, yuan Chao. Application of multivariate analysis of variance base on SPSS software in chemical experiment. Agricultural Network Information, 2017 (04): 22-26.