ISSN:2790-1688 DOI: 10.56028/aetr.1.1.168

Application of Multiple Linear Regression Analysis in the Preparation of C4 Olefins by Ethanol Coupling

Yiyue Zhang

Management Science and Engineering North China Electric Power University Hebei, China

E-mail: 220191061030@ncepu.edu.cn

Abstract. For the preparation of C4 olefins by ethanol coupling, two-dimensional spline interpolation was performed on the original data to reduce the temperature interval.Then the correlation analysis was carried out to determine that the ethanol conversion rate and C4 olefin selectivity are strongly correlated with temperature, and the two are not independent of each other.Subsequently, a multivariate linear regression model was established to fit the ethanol conversion rate and C4 olefin selectivity, and the fitting equation was solved by the least square method. The relationship between the ethanol conversion rate, C4 olefin selectivity and temperature was analyzed by the fitted graph, and then the fitting results were judged and analyzed by the determination coefficient . Finally, the relationship between the selectivity of ethanol conversion products and time was observed by linear interpolation fitting.

Keywords: Two-dimensional spline interpolation; correlation analysis; multivariate linear regression model; interpolation fitting

1. Introduction

Nowadays, with the further development of new economies such as the chemical industry, fossil energy such as coal and oil is increasingly depleted, and the production process conditions are more stringent, so the traditional production technology is difficult to meet the future development needs. Therefore, looking for new renewable energy to prepare industrial chemicals has become an urgent problem.

The production of C4 olefins, a chemical feedstock, is dependent on an increasing shortage of fossil energy.Therefore, from the perspective of sustainable development, the adoption of biomass ethanol for C4 olefins becomes an effective alternative route.In order to better advance the practical production, the exploration of process conditions for the catalytic coupling of ethanol for the preparation of C4 olefins is very far-reaching in terms of both environmental and economic values.

It is known in this paper that different catalyst combinations and temperatures have different effects on ethanol conversion and C4 olefin yield. Firstly, the correlation between ethanol conversion, C4 olefin selectivity and temperature was analyzed, and then a binary linear model was established to explore the functional relationship between the three variables.Finally, the relationship between ethanol conversion rate and product selectivity with time was observed by interpolation linear fitting method, and the change of the given catalyst combination with time in a test data at 350°C was analyzed.

2. Data Processing and Correlation Analysis

2.1 Experimental data

In this paper, the data from the literature[4] , the data in Annex I is the performance data table, the table ethylene, C4 olefins, acetaldehyde, carbon number of 4-12 fatty alcohols are reaction products ; the loading mode I was used in the catalyst experiment numbered A1 – A14, and the loading mode II was used in the catalyst experiment numbered B1 – B7.The data in Annex II are the test data for a given catalyst combination at 350 °C (i.e. a combination of Co loading, Co / SiO2 and HAP loading ratio, and ethanol concentration).

ISSN:2790-1688 DOI: 10.56028/aetr.1.1.168

2.2 Two-dimensinal spline interpolation

Due to the known data is less and the temperature gradient is too large, it can not be fully fitted, which affects the operation effect of data mining. Therefore, it is necessary to interpolate the annex I data reasonably. The two-dimensional interpolation method is selected for data processing, so that the goodness of fit is higher.

The commonly used two-dimensional interpolation methods are nearest neighbor interpolation, bilinear interpolation and cubic polynomial interpolation. Although the nearest neighbor interpolation method is easy to calculate, the accuracy is poor.The calculation using cubic polynomial interpolation method is too complicated. The bilinear interpolation is selected by comparison, and two directional linear interpolations are made according to the nearest neighbor points. The formula is as follows:

$$f(e_0,f) = f(e,f) + \gamma[f(e+1,f) - f(e,f)]$$
$$f(e_0,f+1) = f(e,f+1) + \gamma[f(e+1,f+1) - f(e,f+1)]$$
$$f(e_0,f_0) = f(e_0,f) + \delta[f(e_0,f+1) - f(e_0,f)]$$

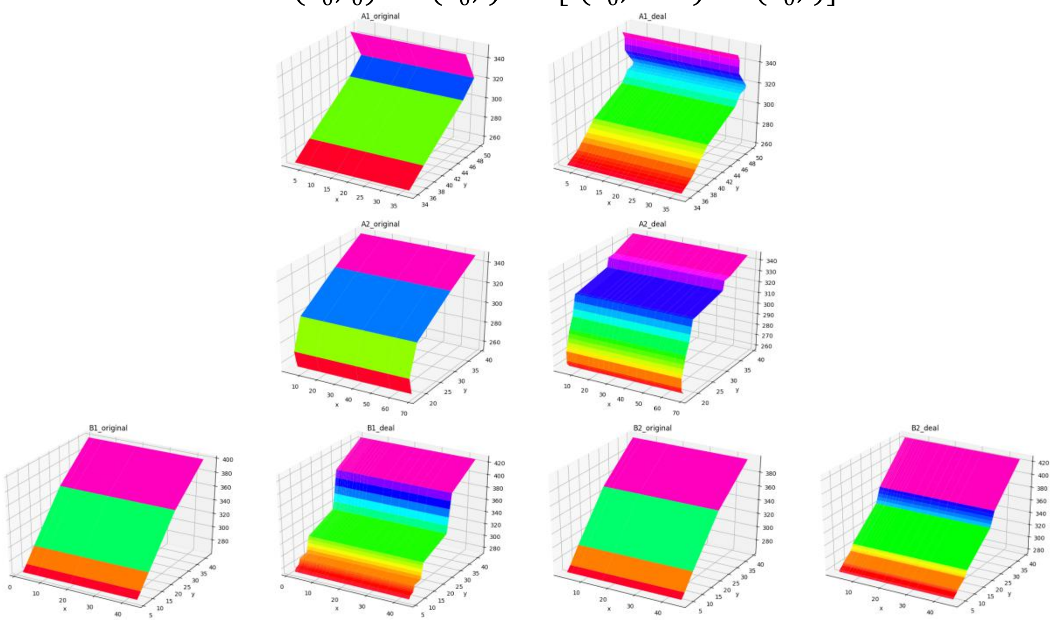

Figure. 1 A1 - A2 group and B1 - B2 group 3D comparison before and after interpolation

2.3 Correlation analysis

As can be seen in Annex 2, the ethanol conversion decreases from 43.5% to 29.9% with increasing time at 350°C. The selectivity of C4 olefins first decreases and then increases, with slight fluctuations during this period.From the analysis of the reactants and products of this reaction, it can be preliminarily judged that the two are not mutually independent with temperature. Then explore the relationship between the three, the need for correlation analysis.

Correlation analysis refers to the analysis of multiple variables with correlation, so as to measure whether there is a certain dependence between variables. In order to accurately describe the correlation between variables, the correlation coefficient r is used as an index to measure, and the calculation formula is :

$$r_{YC} = \frac{\sum[(d_i - \bar{d})(k_i - \bar{k})]}{\sqrt{\sum(d_i - \bar{d})^2} \times \sqrt{\sum(k_i - \bar{k})^2}}$$

$$r_{TC} = \frac{\sum[(t_i - \bar{t})(k_i - \bar{k})]}{\sqrt{\sum(t_i - \bar{t})^2} \times \sqrt{\sum(k_i - \bar{k})^2}}$$

$$r_{YT} = \frac{\sum[(d_i - \bar{d})(t_i - \bar{t})]}{\sqrt{\sum(d_i - \bar{d})^2} \times \sqrt{\sum(t_i - \bar{t})^2}}$$

ISSN:2790-1688 DOI: 10.56028/aetr.1.1.168

Using python to solve the correlation between various factors, the following data :

Table 1. Correlation coefficient between temperature and ethanol conversion

Catalyst combination number	A1	A2	A3	A4	A5	A6	...	B6	B7
r	0.9655	0.995	0.982	0.9975	0.9344	0.9836	...	0.9399	0.9361

Table 2. Correlation coefficients between temperature and C4 olefin selectivity

Catalyst combination number	A1	A2	A3	A4	A5	A6	...	B6	B7
r	0.8871	0.9143	0.9554	0.9578	0.9696	0.8854	...	0.9821	0.9944

Table 3. Correlation coefficient between ethanol conversion and C4 olefin selectivity

Catalyst combination number	A1	A2	A3	A4	A5	A6	...	B6	B7
r	0.7712	0.9029	0.977	0.969	0.9861	0.8986	...	0.933	0.9599

From the above, it can be seen that the correlation coefficients between temperature, ethanol conversion and C4 olefin selectivity are all greater than 0.7, indicating that they are highly correlated with each other. Therefore, when studying the relationship between ethanol conversion, C4 olefin selectivity and temperature, any two could not be selected for separate discussion due to the interaction between the three. A binary linear regression method was used to model the relationship between ethanol conversion, C4 olefin selectivity, and temperature.

3. Multiple Linear Regression Model

3.1 Establishment of the model

The ethanol conversion Y (%) and C4 olefin selectivity C (%) were used as independent variables and temperature T as dependent variable. Because the two independent variables were not independent of each other, it was simple to consider the influence of the self-multiplication degree of each variable and the combination of each variable. Therefore, the multivariate linear regression analysis model is as follows :

$$T = \beta_1 Y + \beta_2 C + \beta_3$$

Using multiple linear regression model fitting temperature and ethanol conversion, C4 olefin selectivity of three-dimensional surface figure is as follows :

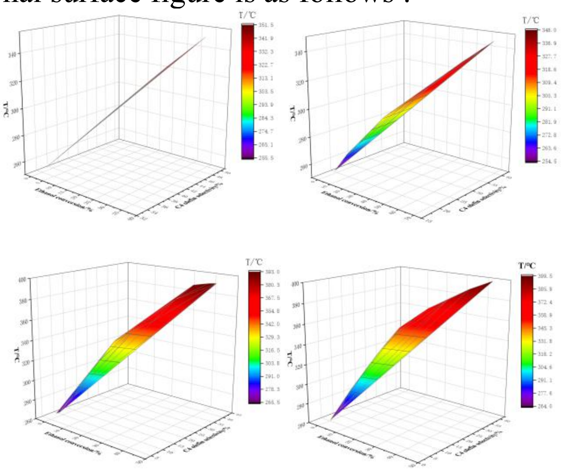

Figure 2 Three-dimensional fitting maps of group A1-A2 and group B1-B2

Note : Since there are 21 sets of images, only four sets of fitted images A1 ~ A2 and B1 ~ B2 are listed here

It can be seen from the above diagram that the change trends of ethanol conversion and C4 olefin selectivity under various catalyst groups are consistent.

ISSN:2790-1688 DOI: 10.56028/aetr.1.1.168

The effect of temperature on ethanol conversion and C4 olefin selectivity was investigated when the reaction temperature was 250-350 °C and the catalyst combination was A1-B7 type. Based on the raw experimental data, it was concluded that the three-dimensional surface plots of the two with reaction temperature were not accurate.The data are preprocessed by two-dimensional spline interpolation, and the experimental data are operated by binary linear fitting method to obtain the three-dimensional fitting surface map, as shown in the above figure

3.2 Solution of the model

As can be seen from A1 in Figure 3, the ethanol conversion and the selectivity of C4 olefins are clearly correlated with the reaction temperature.As the reaction temperature increases from 250 °C to 350 °C, the ethanol conversion increases from 2.07 % to 36.80 %, and the upward trend remains unchanged. The C4 olefin selectivity increases from 34.05 % to 47.21 %, and then the change is not obvious. Figure 3 color changes can be analyzed :

Table 5 Correlation analysis of ethanol conversion rate and C4 olefin selectivity with temperature

Catalyst combination number	Ethanol conversion rate	C4 Olefin selectivity
A1	Positive correlation	Positive correlation
A2	Positive correlation	Positive correlation
A3	Positive correlation	Positive correlation
...	...	...
A10	Positive correlation	Negative correlation
A11	Negative correlation	Positive correlation
...	...	...
B7	Positive correlation	Positive correlation

Using the relevant data in Annex 1, using the least squares optimization algorithm, through matlab to fit its parameters, we can get the following results :

Table 6 Least squares fitting results

ISSN:2790-1688 DOI: 10.56028/aetr.1.1.168

Catalyst combination number	β_1	β_2	β_3
A1	2.234212153	1.330076226	208.6640173
A2	1.266774761	0.352506875	245.2359841
A3	1.752550878	0.785171056	234.0383872
A4	1.495915909	0.218194817	253.0409631
A5	1.584717008	1.818036846	247.9861108
A6	1.694842271	0.384924342	249.7117392
A7	2.261536876	0.418612194	204.9912947
A8	1.915509069	1.527111041	249.9613561
A9	-0.435864795	4.491098497	238.8288402
A10	9.841108247	-17.96030338	307.0004769
A11	-2.457370141	29.31387895	262.0323248
A12	3.505170587	-0.021849544	266.8159587
A13	0.736575823	4.901016704	238.5645042
A14	4.516244729	-4.487646069	264.4804494
B1	3.893149004	-0.31052925	269.1422725
B2	1.078237144	2.927000795	263.0429767
B3	0.778600166	8.257252935	245.6850137
B4	5.08277266	0.227179435	278.1312727
B5	-0.642706887	8.411162138	238.6984666
B6	1.11599294	3.177194872	253.5298809
B7	0.642402327	3.278165301	252.973788

From the above solution the fitted equation can be obtained as :

T_{A1}=2.2342121529766947*Y_{A1}+1.330076226008581*C_{A1}+ 208.66401733884032

T_{A2}=1.26677476066133*Y_{A2}+0.352506874913693*C_{A2}+ 245.235984121621

T_{B1}=3.89314900441018*Y_{A1}+-0.31052924999165*C_{A1}+ 269.142272473557

T_{B2}=1.07823714351427*Y_{B2}+2.92700079529242*C_{B2}+ 263.042976654036

Note : Since there are 21 groups of fitting equations, only A1-A2 and B1-B2 fitting equations are listed here.

4. Testing of Results

The relationship between the two factors explored in this problem and temperature is calculated by establishing a binary linear fitting equation. Therefore, the accuracy of the fitting function is the key to the problem, so the fitting results are tested and analyzed below.

Decision coefficient R^2 : If the predicted value is greater than the actual value, the decision coefficient is greater than 1 ; if the predicted values are less than the actual values, the determination coefficient is less than 1. The closer the determination coefficient is to 1, the smaller the difference between the predicted value and the actual value is, and the better the fitting effect is.

$$R^2 = \frac{\sum(\hat{T}_i - \bar{T}_i)^2}{\sum(T_i - \bar{T}_i)^2}$$

The correction coefficient of the fitting surface is as follows : Table 6 shows that the correction coefficient of each group is close to 1, indicating that the goodness of fit is high. In order to further increase the overall goodness of fit, partition fitting can be considered.

Table 7 Correction coefficient of reaction under different catalyst combinations

ISSN:2790-1688 DOI: 10.56028/aetr.1.1.168

Catalyst combination number	A1	A2	A3	A4	A5	A6	...	B6	B7
R^2	0.9216	0.9186	0.9903	1.0146	0.9588	0.9326	...	0.9299	0.9369

5. Results Analysis

When the reaction temperature was 350 °C and the catalyst combination was given, the change of ethanol conversion rate with the reaction (from 0 to 273 min) was investigated. According to the experimental data, the broken line diagram of ethanol conversion rate with the change of reaction time was made by linear interpolation.

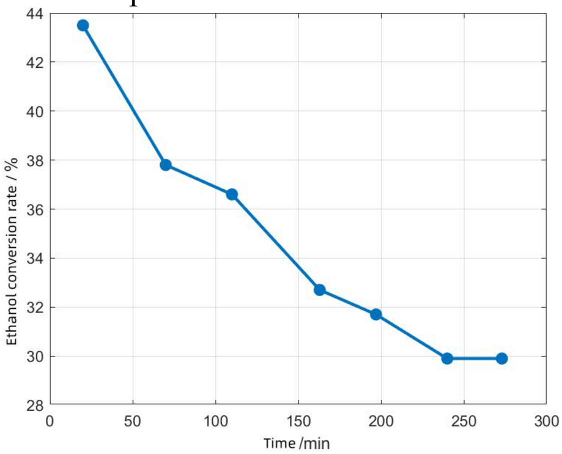

Figure 3 Breakline diagram of ethanol conversion rate changing with time

It can be seen from Fig. 4 that the reaction time has a significant effect on the ethanol conversion rate. With the extension of reaction time, the ethanol conversion rate gradually decreases from 43.5 % in 20 min to 29.9 % in 240 min, and then tends to be stable, that is, the ethanol conversion rate reaches a trough saturation value. It can also be seen from the slope change of the image that the conversion rate of ethanol decreases rapidly at the initial stage of the reaction, and the decrease rate gradually slows down to zero with the extension of the reaction time.

It can be inferred from the graphical trend and literature that the reaction is reversible, and there is a dynamic equilibrium establishment process under certain conditions. At about 240 min, the reaction achieves a dynamic equilibrium, so that the ethanol conversion rate does not change.

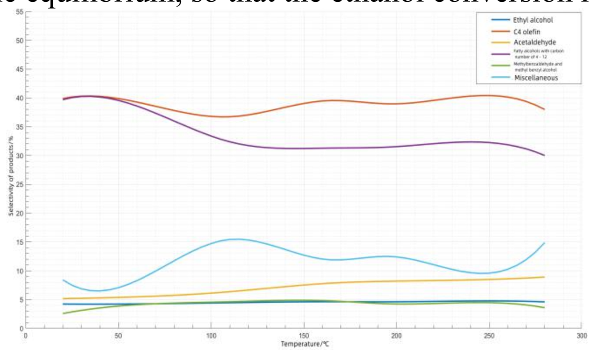

Figure 4 Fitting graph of selectivity of each product versus temperature

When the reaction temperature was 350°, the effect of reaction time on the selectivity of each product was investigated under the given catalyst combination. (1) The selectivity of acetaldehyde increased with the increase of temperature, while the selectivity of fatty alcohol with carbon number of 4 – 12 decreased ; (2) There is a maximum selectivity of C4 olefins in the reaction process, and the selectivity of C4 olefins in 50-150 min is the lowest compared with other periods, and the overall trend is also declining. Therefore, prolonging the reaction time may reduce the selectivity of C4 olefins. At the same time, the literature shows that the increase of the loading amount may lead to the decrease of the pore size of the catalyst. With the extension of the reaction time, the carbon

ISSN:2790-1688 DOI: 10.56028/aetr.1.1.168

on the surface area of the catalyst will lead to the gradual loss of activity and the selectivity of C4 olefins decreases gradually, which is consistent with the fitting curve, indicating that the image fitting is better.

6. Conclusion

In this paper, the relationship between ethanol conversion, C4 olefin selectivity and temperature for 21 different catalyst combinations was explored by developing a multivariate linear model. In order to verify the fitting effect, this paper verifies and analyzes, and obtains that the correction coefficient of each group is close to 1, indicating that the goodness of fit is high, but the constraint range of the parameters is not accurate enough. Then, the relationship between ethanol conversion and product selectivity over time was studied by interpolation fitting at 350° C. It can provide a reference for selecting catalysts to design the optimal performance reaction and play a positive role in the production of C4 olefins.

References

[1] Lu,S.P. Ethanol coupling to prepare butanol and C4 olefins [D]. Dalian : Dalian University of Technology, 2008 : 2-3.

[2] Sang.L, Shu.G P, Relationship between conversion and yield of Shenhua coal direct liquefaction and temperature and time [J]. Modern chemical industry, 2019,39 (5) : 230-231.

[3] FANG.Z, XIANG.H Y, ZHOU,H.J.Application of Multivariate linear Ｒegression Analysisin Grape Wine Quality Evaluation[J].Journal of Hubei University for Nationalities(Natural Science Edition),2014,32(4):427-429

[4] Organization Committee of National Undergraduate Mathematical Modeling Contest. Data of Group B of National Undergraduate Mathematical Modeling Contest in 2021 ［EB /OL］.[2014 － 09 － 09].http://www.mcm.edu.cn/node/4d73a36cc88b35bd4883c276afe39d89.html