ISSN 2616-5880 Vol. 4, Issue 5: 36-43, DOI: 10.25236/AJMC.2023.040506

The Regression Analysis Model of C4 Olefin Yield Optimization Study

Aijia Chen[1,*]

[1]Faculty of Science, Shanghai Maritime University, Shanghai, 201306, China
[*]Corresponding author

Abstract: *Olefin is widely used in chemical and pharmaceutical production. In the process of producing C4 olefin, catalyst combination and temperature will affect the formation of C4 olefin. As one of the most important high value-added raw materials in the chemical industry, the synthesis of C4 Olefin by ethanol coupling was of great significance in the field of the chemical industry. Different catalysts and various conditions have different effects on the chemical reaction [1]. This paper mainly explores the relationship between selectivity of C4 olefin and catalyst combination and temperature, and how to optimize catalyst combination and temperature to make the yield of C4 olefin as high as possible. This article is based on relevant data sets. This article explores the relationship between variables such as the amount of Co loading, Co/SiO2 ratio, HAP loading ratio, ethanol concentration, catalyst loading method, temperature, and the selectivity of C4 olefins and the conversion rate of ethanol. It is based on relevant datasets and utilizes the method of controlling variables. Pearson coefficient is used to test the correlation between variables, and a multiple linear regression model is established. Finally, experiments are designed to verify the optimal conditions for achieving the highest yield of C4 olefins. Based on analysis and validity testing, the findings suggest that the proposed model is well-suited for chemical reactions in a wide range of industrial production contexts, facilitating efficient extraction of the desired product during actual manufacturing processes. Future improvements will focus on optimizing the model with a larger dataset, incorporating advanced deep learning techniques, and enhancing both its accuracy and precision.*

Keywords: *Regression analysis, Model Control variable method, C4 olefin yield*

1. Introduction and background

C4 Olefins are hydrocarbons that contain carbon-carbon double bonds (C=C bonds). Depending on the number of double bonds, they can be classified as mono-olefins, di-olefins, and so on. The general formula for a linear mono-olefin molecule is CnH2n. The double bond group serves as the functional group in olefin molecules, endowing them with reactivity. These molecules can undergo various addition reactions, including hydrogenation, halogenation, hydration, halo hydrogenation, secondary halogenation, sulfuric acid addition, epoxidation, and polymerization. They can also undergo oxidation reactions, leading to the cleavage of the double bond and the formation of aldehydes, carboxylic acids, and other products. Olefins, as fundamental chemical raw materials, significantly contribute to optimizing product structures and driving economic development. However, the olefin industry has encountered adverse external challenges in recent years, such as low oil prices and the global pandemic, intensifying uncertainties surrounding its future development. Given the substantial demand, technological innovation in olefin production has emerged as a crucial imperative.

The yield of C4 olefin is frequently limited by the intricate nature of the accompanying products. Discovering the optimal conditions for ethanol reactions necessitates iterative manual experiments, leading to significant resource consumption. Consequently, designing ethanol reaction conditions that yield the highest possible amount of C4 olefin poses a considerable challenge[2]. C4 olefins possess wide-ranging applications in the chemical and pharmaceutical industries, making them highly valuable. The catalyst combination and temperature employed during the production process exert a significant influence on the generation of C4 olefins. This study is dedicated to investigating the interplay between C4 olefin selectivity and the catalyst combination, along with temperature. Moreover, the study aims to optimize both the catalyst combination and temperature to maximize the yield of C4 olefins. Achieving a high yield of C4 olefin presents challenges due to the complexity of the associated products. The process of finding the optimal reaction conditions for ethanol conversion typically involves laborious and resource-intensive manual experiments. Consequently, designing ethanol reaction conditions that maximize the yield of C4 olefin is a formidable task. To address the challenges of optimizing the yield of C4 olefins, this study processes the data by using regression analysis.

2. Literature review

The production of C4 olefins through ethanol conversion is a topic of significant interest in the chemical and pharmaceutical industries. The yield of C4 olefins is influenced by various factors, such as catalyst combination and temperature. In recent years, there has been a growing focus on optimizing the production process to maximize the yield of C4 olefins. This literature review aims to explore the existing research on the optimization of C4 olefin yield in ethanol conversion processes, examining the key findings, methodologies, and advancements in this field.

C4 olefins, hydrocarbons containing carbon-carbon double bonds, have widespread applications in the chemical and pharmaceutical industries. Their versatility and reactivity make them valuable building blocks for the synthesis of various products. C4 olefins play a crucial role in optimizing product structures and promoting economic development. Their demand is steadily increasing due to their applications in the production of plastics, synthetic rubbers, solvents, and other essential materials[4].

The catalyst combination used in ethanol conversion processes significantly affects the selectivity and yield of C4 olefins. Different catalysts, such as Co/SiO2 and HAP, have been studied for their effectiveness in promoting C4 olefin production. Temperature is another critical parameter that influences the yield of C4 olefins. Research has shown that higher temperatures can enhance the conversion of ethanol and increase the yield of C4 olefins. However, finding the optimal temperature range is crucial to balance selectivity and yield.

Regression analysis is commonly employed to investigate the relationship between independent variables (catalyst combination and temperature) and dependent variables (ethanol conversion and C4 olefin selectivity). This statistical technique allows for the identification of significant factors and the quantification of their impact. Given the complexity of the system and the presence of multiple independent variables, the controlled variable method is often used to explore the effects of each independent variable on the function. This approach allows for a systematic examination of the influence of catalyst combination and temperature on the yield of C4 olefins. Numerous studies have investigated the effects of different catalyst combinations on C4 olefin yield. The results indicate that specific catalyst combinations can significantly enhance the selectivity and yield of C4 olefins, offering promising strategies for process optimization. Optimizing the temperature plays a vital role in achieving high yields of C4 olefins. Researchers have explored the temperature range and identified the optimal conditions that balance ethanol conversion and C4 olefin selectivity. The development of mathematical models, such as regression models, has facilitated the prediction and optimization of C4 olefin yield. These models provide a systematic approach to analyze the impact of catalyst combination, temperature, and other variables on the process.

In conclusion, the optimization of C4 olefin yield in ethanol conversion processes is a topic of significant research interest. Catalyst combination, temperature, and other factors have been identified as critical parameters influencing the yield and selectivity of C4 olefins. The literature review highlights the importance of these factors and presents key findings regarding their optimization. The use of regression analysis and controlled variable methods has contributed to a better understanding of the relationship between independent and dependent variables. Future research should focus on further exploring the synergistic effects of catalyst combinations, temperature, and other process variables to maximize the yield of C4 olefins while ensuring optimal selectivity.

This literature review provides a comprehensive overview of the research conducted on the optimization of C4 olefin yield in ethanol conversion processes. By examining the key findings and methodologies employed in previous studies, it establishes a solid foundation for further advancements in this field. The hierarchy and logical flow of the review ensure that the reader gains a clear understanding of the significance of C4 olefins, the factors influencing their yield, and the research methodologies employed to study this phenomenon.

The review highlights the economic importance of C4 olefins and their role in promoting product optimization and economic development. It then delves into the factors that affect C4 olefin yield, emphasizing the significance of catalyst combination and temperature. The research methodologies section elucidates the statistical techniques, such as regression analysis and controlled variable methods, used to investigate the relationships between variables. Finally, the review presents the key findings and advancements in the field, emphasizing the relationship between catalyst combination, temperature, and C4 olefin yield.

Overall, this literature review serves as a valuable resource for researchers and practitioners in the field of ethanol conversion and C4 olefin production. It provides a comprehensive understanding of the factors influencing C4 olefin yield and offers insights into the optimization strategies employed in previous studies. By building upon the existing knowledge, future research can further advance the field and contribute to the development of efficient and sustainable processes for C4 olefin production.

3. Establishment and solution of model

3.1. Data Analysis

Firstly, we use the method of cluster analysis to process the existing data. Cluster analysis can be applied in the process of data preprocessing[3].

Many clustering techniques have been developed in different application fields. These techniques are used to describe data, measure the similarity between different data sources, and classify data sources into different clusters. For multi-dimensional data with complex structure, cluster analysis can be used to gather the data and standardize the data with complex structure. Cluster analysis can also be used to discover the dependencies between data items, to remove or merge data items with close dependencies.

3.2. Symbol description

Symbols used for model establishment are described in Table 1.

Table 1: Symbol Description

Symbol	Description	Unit
X1	Temperature	degree
X2	Co loading capacity	wt
X3	Co/SiO2 and HAP ratio	mg
X4	Ethanol concentration	ml /min
X5	Catalyst charging method	/
Y1	Ethanol conversion rate	%
Y2	C4 olefin selectivity	%
H(X)	C4 olefin yield	%

3.3. The relationship between ethanol conversion and C4 olefin selectivity and temperature

To analyze the intricate relationship between ethanol conversion rate, C4 olefin selectivity, and temperature for each catalyst combination provided in Appendix 1, a rigorous approach involving simple linear regression analysis can be employed. The following step-by-step procedure outlines the analytical strategy:

Data visualization: Visualize the dataset by plotting scatter plots of ethanol conversion rate (Y1) and C4 olefin selectivity (Y2) against temperature (X1). Carefully examine the scatter plots to identify any discernible trends or patterns.

Determination of the regression equation: Based on the visual analysis of the scatter plots, derive the appropriate regression equations that can effectively model the relationship between the variables.

3) Utilization of the "fit" method: Utilize the powerful "fit" method provided by the linear modeling capabilities of the Matlab Statistics Toolbox to perform model fitting. This method will enable the estimation of regression coefficients, significance levels, and other essential statistical measures.

Refinement of the model: Enhance the model's accuracy and reliability by identifying and eliminating any outliers or anomalous data points that might adversely affect the regression analysis.

Residual analysis: Conduct thorough residual analysis to assess the goodness of fit and validate the regression model. This analysis allows for the identification of any deviations or discrepancies that may undermine the model's performance, providing insights into potential improvements or necessary adjustments.

For the second step, advanced functionalities such as nlinfit, nlparci, nlpredci, and nlintool from the Matlab Statistics Toolbox can be employed. These tools provide comprehensive regression coefficient estimates, confidence intervals, predicted values, and additional statistical measures necessary for an in-depth analysis.

3.4. Explore the ethanol conversion and C4 selectivity of different catalyst combinations and temperatures

To investigate the impact of different catalyst combinations and temperatures on ethanol conversion and selectivity of C4 olefins, the following methodologies were employed to establish the relationship between independent variables (catalyst combination and temperature) and dependent variables (ethanol conversion and selectivity of C4 olefins).

Initially, a multiple regression analysis model was constructed, and the regression coefficients were estimated using the least square method. Subsequently, residual analysis was performed to evaluate the goodness of fit of the model. The linear test of the regression model was conducted to assess its validity. The entire model-solving process was carried out utilizing Matlab, which facilitated the derivation of optimal solutions. The main challenge encountered in this study was the identification of variables and their complex interrelationships within the given contexts. To address non-unique dependent variables, the controlled variable method was employed to investigate the individual effects of various independent variables on the overall function.

This approach allowed for a comprehensive analysis of the effects of different catalyst combinations and temperatures on ethanol conversion and selectivity of C4 olefins. The utilization of multiple regression analysis, combined with the controlled variable method, enhanced the understanding of the intricate relationships between independent and dependent variables. The application of Matlab as a computational tool provided a robust framework for solving the model and deriving optimal solutions.

3.4.1. The independent variable is temperature

We investigated the impact of different catalyst combinations and temperatures on the ethanol conversion rate and C4 olefin selectivity. The independent variables included catalyst combination (consisting of Co loading amount, Co/SiO2, HAP packing ratio, and ethanol concentration) as well as temperature. The dependent variables were ethanol conversion rate and C4 olefin selectivity.

3.4.2. The independent variable is Co load x

Table 2: Experimental data processing2

Catalyst combination (Under 250°C)	Co loading capacity /wt%	Ethanol conversion rate (%)	C4 olefin selectivity (%)
A1	1	2.07	34.05
A2	2	4.60	18.07
A4	0.5	4.0	9.62
A6	5	13.4	3.3

The effects of Co loading amount x2 on ethanol conversion y1 and selectivity y2 of C4 olefin are investigated. We conducted analysis and processing on the existing data, resulting in the following Table 2:

Based on the data from the table, a scatter plot is generated to illustrate the relationship between ethanol conversion rate and C4 olefin selectivity with respect to Co loading amount. Based on the analysis, we established a simple linear regression model:

$$F(x) = a + bx \tag{1}$$

By applying this simple linear function, we can simulate the relationship between X and Y. The key is that this function is not only linearly related to the input variable X but also linearly related to the parameters a and b.

The current objective is to determine the values of parameters a and b that best fit the training data.

This can be achieved by measuring the mismatch between the actual target values y and the model F(x) for each input x and minimizing this mismatch. This minimized mismatch, also known as the error function, has various options to choose from, but the simplest one is the RSS. It calculates the sum of squared errors between the model F(x) and the target values y for each data point X.

$$RSS = \sum(F(x) - y)^\wedge 2 \tag{2}$$

Using the concept of error function, we replace "determine parameters a and b that best fit the training data" with "determine parameters a and b", to minimize the error function.

Calculate the error function of the training data by the following formula:

$$E(a, b) = \sum(F(x) - y)^\wedge 2 \tag{3}$$

The above equation is an error function requiring a minimum value. But how to find the parameters a and b to get the minimum value of this function? We visualize the function and compute the solution on Matlab.

$$\mathrm{\sigma/\sigma a(E(a, b)) = 0, \sigma/\sigma b(E(a, b)) = 0} \tag{4}$$

Thus, it shows that the error is within a reasonable range, so the model is reasonable.

3.4.3. The independent variable is Co/SiO2 and HAP content

Analyze and process the existing data, and the table after processing is shown as follows. Then a simple linear regression model is established with reference to the above models, and the relationship between them is calculated by the least square method and correlation. We conducted analysis and processing on the existing data, resulting in the following Table 3:

Table 3: Experimental data processing3

Catalyst combination (250℃)	Co/SiO2 to HAP content ratio	Ethanol conversion (%)	C4 olefin selectivity (%)
A12	1	1.40	6. 17
A13	2.03	1.30	5. 19
A14	0.49	2.50	1.89

3.4.4. The independent variable is Ethanol concentration

In a similar way, analyze and process the existing data, and the table after processing is shown as follows. Then a simple linear regression model is established with reference to the above models, and the relationship between them is calculated by the least square method and correlation. We conducted analysis and processing on the existing data, resulting in the following Table 4:

Table 4: Experimental data processing 4

Catalyst combination (250℃)	Ethanol concentration	Ethanol conversion (%)	C4 olefin selectivity (%)
A7	0.3	19.7	5.75
A8	0.9	6.3	5.63
A9	2.1	2.1	5.4
A12	1.6	1.4	6. 17

3.5. Considering the optimal catalyst combination and temperature

To develop a comprehensive understanding of the factors influencing the yield of olefins, a multiple linear regression model was constructed, incorporating all relevant independent variables into a single equation. By considering the optimal combination of multiple independent variables, the model aims to predict or estimate the dependent variable more accurately compared to a single independent variable approach.

The selection of independent variables for the multiple regression model was determined based on the correlation matrix, which assesses the relationships between variables. This analysis enables the identification of variables that exhibit strong correlations with the dependent variable, thereby enhancing the predictive power of the model. By considering multiple independent variables, the model can capture the combined effects of various factors on the yield of olefins, providing a more comprehensive and accurate prediction or estimation.

The principle of establishing multiple function regression model:

1) The independent variable must have a significant impact on the dependent variable and is closely linear correlation.

2) The linear correlation between the independent variable and the dependent variable must be real, not formal.

3) Independent variables should have complete statistical data, and their predictive values are easy to determine.

3.5.1. Correlation analysis among independent variables

Pearson coefficient was used to test the correlation between variables, including independent variables temperature(X1,), Co loading(X2), CoSiO2/HAP weight(X3), ethanol concentration(X4), catalyzer loading mode(X5) and dependent variable olefins yield.

The correlation coefficient for h(x) varies from -1 to +1, with $r>0$ indicating that the two variables are positively correlated, that is, higher values of one variable correspond to higher values of the other variable. A value of $r<0$ indicates that the two variables are negatively correlated, meaning that higher values for one variable are lower values for the other.

The larger the absolute value of r, the more correlated the two variables are. A value of $r=0$ means that the two variables are not linearly r correlated but may be correlated in other ways.

In general, when $|r| \geq 0.8$, we can confidently conclude that the two variables exhibit a strong correlation. For values between $0.5 \leq |r| < 0.8$, we can consider the correlation to be moderate. When the correlation coefficient falls within the range of $0.3 \leq |r| < 0.5$, we observe a weak correlation between the variables. Lastly, if $|r| < 0.3$, the variables can be deemed essentially unrelated. Since the correlation coefficient obtained from the analysis using SPSS software is less than 0.3, it provides compelling evidence to suggest that no significant correlation exists among the variables. Therefore, employing a multiple linear regression model is appropriate for this scenario.

3.5.2. Solution process and results

In this problem, there are independent variables x1 temperature, x2Co loading, x3CoSiO2/HAP weight, x4 ethanol concentration, x5 catalyst loading mode and dependent variable olefinic yield h (x). The model is established as follows.

$$h(x) = p_0 + p_1x_1 + p_2x_1 + p_3x_3 + p_4x_4 + p_4x_5 \tag{5}$$

The matrix is further expressed as:

$$h(x) = x_m \tag{6}$$

The function h (x) is a 5*1 vector, θ is a 5*1 vector, and there are five algebraic model parameters in it. X is a 5x5 dimensional matrix. 5 represents the number of samples and 5 represents the number of features of the sample.

Now that we have our model, we need to figure out the loss function we need. For linear regression, we use mean squared error as the loss function. The algebraic representation of the loss function is as follows.

$$J(p_1, p_2, p_3, p_4, p_5) = \sum(h(x_1, x_2, x_3, x_4, x_4) - y_i)\text{^}2 \tag{7}$$

Using the least square method, we obtain

$$p = (X\text{^}TX)\text{^}(-1)X\text{^}TY \tag{8}$$

In summary, the yield of C4 olefin can be as high as possible under the condition of 400°C by choosing A charging method of 75mg 0.5wt % Co/SiO2 to 75mg HAP-ethanol concentration of 0.3ml/min and 400.

3.6. Design experiments to verify the results under the best conditions

Based on the above problems, the catalyst combination used in this experiment and the best temperature at the time when the yield of C4 olefin is the highest have been calculated, and the experimental evaluation can be designed to verify whether the yield of C4 olefin obtained under these conditions will be the best.

When designing the experiment, we still adhere to the principle of controlling variables. We devised five groups of experiments, allowing for complementary contrasting experiments. In each group, one variable is systematically altered while keeping the other variables at their optimal experimental conditions. This approach allows us to further validate the rationality of the optimal solution proposed in question three and subject it to a secondary examination.

Design experiment 1: 75mg 1wt%Co/SiO2- 75mg HAP-ethanol concentration 0.3ml/min at 400°C

Design experiment 2: 75mg 0.5wt%Co/SiO2- 75mg HAP-ethanol concentration 0.3ml/min at 350°C

Design experiment 3: 50mg 0.5wt%Co/SiO2- 50mg HAP-ethanol concentration 0.3ml/min at 400°C

Design experiment 4: 75mg 0.5wt%Co/SiO2- 75mg HAP-ethanol concentration 0.3ml/min at 400°C

Design experiment 5: 75mg 0.5wt%Co/SiO2- 75mg HAP-ethanol concentration 1.68ml/min at 400°C

4. Evaluating the Model

4.1. Advantages of the model

1) The model effectively visualizes a large volume of tabular data and employs different data processing techniques for each specific problem. The utilization of Matlab for generating graphical representations ensures intuitive and clear visualization.

2) Given the complexity arising from multiple independent variables and non-unique dependent variables, the control variable method is employed to explore the influential pathways of each independent variable on the function[5].

3) The model exhibits strong operability and can be easily implemented and applied in various scenarios, thus facilitating widespread adoption.

4) The error analysis of the established regression model was tested to ensure the accuracy and completeness of the model.

5) This paper primarily utilizes the regression model to expand the modeling approach, enabling convenient establishment, solution, and testing of the model while maintaining a thorough understanding of the underlying principles.

4.2. Shortcomings and improvements of the model

The use of the model in this paper is relatively single, mainly the application of regression model, and the workload is huge and tedious when processing data, and it is prone to errors. Therefore, the following improvements are made to the model:

1) The method of clustering analysis can be added based on regression analysis in question 1. Clustering can be used as an independent tool to obtain the distribution of data, observe the characteristics of each cluster of data, and focus on further analysis of specific clusters, to effectively simplify data and carry out data modeling[6].

2) On the basis of problem binary regression, neural network model can be added to optimize the processing of multiple types of data such as image, speech, text, and sequence to realize classification, regression, and prediction.

3) There are many uncertain factors in the actual experiment, and the influence of side reactions is ignored.

4.3. Generalization of the model

The primary approach employed in this study is regression analysis, which allows for a comprehensive exploration of the relationship between variables. Specifically, the univariate regression model is utilized to examine the individual impact of each independent variable on the dependent variable, providing valuable insights into their significance[7].

Furthermore, the multiple regression analysis model serves as a versatile tool for investigating the intricate connections between various relevant variables. By extending the model to incorporate additional factors related to industrial production and our daily lives, it becomes possible to unravel the complex web of relationships that influence these domains. Through this comprehensive analysis, a deeper understanding of the interplay between multiple factors and their collective impact on industrial production and service provision can be gained. This approach offers valuable insights for informed decision-making and the development of strategies to optimize outcomes in these areas.

5. Conclusion

This study focuses on investigating the relationship between the selectivity of C4 olefins and the catalyst combination as well as temperature. The experimental results of a specific catalyst combination at 350°C are quantitatively analyzed at different time intervals within a single experiment. Additionally, the effects of different catalyst combinations and temperatures on ethanol conversion and C4 olefin selectivity are discussed. By selecting the appropriate catalyst combination and temperature, it is possible to maximize the yield of C4 olefins under the same experimental conditions.

Furthermore, this study explores the optimization of C4 olefin yield when the temperature is below 350°C by selecting suitable catalyst combinations and temperatures. To validate the effectiveness of the optimal conditions, an experiment is designed to achieve the highest possible yield of C4 olefins. Additionally, five sets of experiments are designed to further increase the yield of C4 olefins, with detailed rationale provided. The findings of this research demonstrate that temperature has a significant positive correlation with the generation of C4 olefins. Moreover, the variables of Co loading, Co/SiO2 and HAP charging ratio, ethanol concentration, catalyst charging method, and temperature have a significant impact on both the selectivity of C4 olefins and ethanol conversion, showing a positive correlation.

In conclusion, this research provides valuable insights into the relationship between catalyst combination, temperature, and the selectivity of C4 olefins. The results highlight the importance of temperature and the influence of various variables on the production of C4 olefins. The designed experiments and their outcomes validate the efficacy of the proposed optimization approach in enhancing the yield of C4 olefins.

References

[1] Shaopei Lv. Preparation of butanol and C_4 olefin by coupling ethanol [D]. Dalian University of Technology, 2018.

[2] Jianfei Leng, Xu Gao, Jiaping Zhu. Application of multivariate linear regression statistics to predictive models [J]. Statistics and decision-making, 2016, 07:82-85.

[3] Liang Dai, Hongke Xu, Ting Chen, Chao Qian, Dianpeng Liang. Multiple linear regression prediction Model based on Map Reduce [J]. Computer application, 2014, 07:1862- 1866.

[4] Feng Ye. Application of multiple linear regression in forecasting economic and technical output [J]. China and Foreign Energy, 2015, 02:45-48

[5] Xingren Jin. Analysis of influence factors of institutional investors' shareholding preference in pharmaceutical industry [D]. Yunnan University of Finance and Economics, 2021.

[6] Li Shaowei, Wang Yujie, Xiong Lang, Huang Shengqi. Optimization model for the preparation of C4 olefins by ethanol coupling [J]. Journal of Taizhou University, 2021, 43(06): 26-32+77. DOI: 10. 13853/j. cnki. issn. 1672-3708. 2021. 06. 004.

[7] Wu Shaojun. Development Status and Optimization Measures of Chemical Engineering Technology [J]. Chemical Design Communication, 2021, 47(07):122-123.