Problem Chosen	2023	Team Control Number
B	ShuWei Cup Summary Sheet	2023091819500

Summary

Cotton straw has become an important part of renewable energy due to its rich biomass content such as cellulose (CE) and lignin (LG). The process of its pyrolysis activity and the yield of each product are greatly affected by factors such as catalysts . This paper aims at the mechanism of the catalytic reaction of cotton straw pyrolysis. Based on the interaction between the desulfurization ash mixing ratio and the yield of each pyrolysis product and the yield of pyrolysis gas, mathematical modeling methods and machine learning models are used to predict the performance of the catalyst in the pyrolysis process. to study its mechanism and effects.

For question 1, first preprocess the data in Appendix 1. Based on various statistical indicators, the overall distribution of pyrolysis products under different pyrolysis combinations was analyzed. The Pearson correlation coefficient was used to explore the correlation between pyrolysis combinations, pyrolysis products, and pyrolysis products. There was an obvious positive correlation between the data. Negative relationship and significance (for example, the correlation coefficient between the catalyst mixing ratio and water production in D FA/CS **is 0.98 , and** the correlation coefficient between water production and tar production in D FA/LS is -0.99). The changing characteristics and **trends** of each pyrolysis yield were analyzed through **the quadratic fitting** curve with good fitting effect (average R^2 is **0.87). Finally** , the catalyst mixing ratio of each pyrolysis yield was discussed based on the information entropy weight obtained by **the entropy weight method.** The degree of influence of pyrolysis products.

For question 2, first process the data in Appendix 2 into the form **of yield** , and understand the interactive relationship between the two based on the histogram of the overall gas yield changing with the desulfurization ash mixing ratio in the three pyrolysis combinations and the trend line with confidence intervals. , and then the distribution of the production of each gas product during the pyrolysis process is displayed through stacked bar graphs. Conduct a descriptive analysis of the pyrolysis process through various statistical indicators, and finally use the quadratic fitting results of each pyrolysis gas yield (average R^2 = **0.88**) to explore the impact of the mixing ratio on the pyrolysis gas yield, combined with the chart and review Relevant literature reviews the pyrolysis mechanism.

100 data points with a mixing ratio between 0 and 1 and an interval of 0.01 correspond to the yields and heat of CE and LG pyrolysis products. The yield of decomposed gas components was predicted for **data expansion** . **The Q -Q** plot and **Shapiro-Wilk method were used to test the normality** of the obtained data set . The yields of each pyrolysis product and pyrolysis gas **were all normal. Distribution** , based on the paired **sample t test** method, we judge whether there is a significant difference in the yield of each pyrolysis product and the yield of pyrolysis gas components under the catalysis of the same proportion of desulfurization ash and discuss the mechanism.

For question four, firstly, the catalytic reaction mechanism model of desulfurization ash of model compounds such as CE and LG was constructed through SEM structural equations, and based on the influence mechanism between each pyrolysis combination and pyrolysis products in the pyrolysis reaction revealed by each **effective path .** Reaction mechanism model. **The cubic fitting method** (average R^2 = **0.96)** was used to establish a reaction kinetic model to study and analyze the catalytic reaction mechanism of desulfurization ash.

For question five, first select random forest (RF), XGBoost, gradient boosting regression tree (GB R T) and K NN (K - nearest neighbor) algorithm for model training, and use the grid search method to optimize the parameters of each model. Based on the model simulation effect shown in **the Taylor diagram , the X** GB oost and R F algorithms were finally selected for further research. Combined with the reaction kinetics model constructed in question 3 **, an integrated model** of the three learners was constructed . The integrated model was used to analyze the pyrolysis in different pyrolysis combinations. The yield of products and the yield of pyrolysis gas are simulated and predicted.

Finally, the built model was discussed and evaluated, and the advantages and disadvantages of the model were analyzed.

Keywords: **cotton stalk pyrolysis, function fitting, SEM structural equation, machine learning, integrated model**

Content

1 Problem restatement

1.1 Problem background

As the global demand for renewable energy continues to increase, research and development of various renewable energy technologies have become the focus of common concern of all countries. Biomass energy, as a mature form of renewable energy, plays an important role in this. Cotton straw, as an agricultural waste, is widely recognized as a renewable biomass resource. It is rich in biomass components such as cellulose and lignin, so it has potential energy value. However, the process and products of cotton straw pyrolysis are affected by many factors, including pyrolysis temperature, catalysts, etc. Pyrolysis temperature has an important influence on the quality and yield of pyrolysis products. Therefore, in-depth study of the pyrolysis mechanism and properties of cotton straw, as well as the study of the mechanism and effect of catalysts in the pyrolysis process, are of great significance for achieving efficient utilization and sustainable development of cotton straw.

1.2 Question request

(1) Analyze the relationship between product yield and mixing ratio under different pyrolysis combination conditions in Appendix 1, and explore the impact of desulfurization ash as a catalyst on the pyrolysis process.

(2) For Appendix 2, draw images to analyze the effect of mixing ratio on pyrolysis gas yield under different pyrolysis combination conditions.

(3) Compare the differences in product yields and pyrolysis gas components between CE and LG pyrolysis combinations under the same catalytic conditions, and explain the reasons for the differences.

(4) Establish a desulfurization ash catalytic reaction mechanism model for model compounds such as CE and LG, and use the reaction kinetic model for further analysis.

(5) Use limited data to predict the yield or quantity of pyrolysis products based on mathematical models.

2 problem analysis

2.1 Question one

This question requires an analysis of the relationship between the yield of pyrolysis products for each pyrolysis combination in Appendix 1 and the mixing ratio of the corresponding combination, and to determine whether desulfurization ash affects the pyrolysis of cotton stalks, cellulose, and lignin. There is a significant promotion effect. First, preprocess the data to determine whether there are missing values and outliers in Appendix 1. In order to better display the distribution pattern of product yields of different pyrolysis combinations, statistical indicators including kurtosis and coefficient of variation were introduced to describe them, and the Pearson correlation coefficient was used to explore the relationship between pyrolysis combinations and pyrolysis products. Correlation between pyrolysis products, then use quadratic fitting to analyze the changing characteristics and trends of each pyrolysis product yield, and finally use the entropy weight method to calculate the information entropy weight to discuss the impact of the desulfurization ash mixing ratio on the yield of each pyrolysis product degree.

2.2 Question 2

This question requires an analysis of the impact of the mixing ratio of the pyrolysis combination on the gas yield of each group. First, based on the pyrolysis gas yield given in Appendix 2, calculate the yield of each component gas of each pyrolysis combination and draw a histogram. and a trend line with a confidence interval to describe the relationship between the mixing ratio of different pyrolysis combinations and the total gas production. Then descriptive statistics were performed on the gas production rate of each pyrolysis combination to analyze its statistical characteristics. Finally, quadratic fitting was used to analyze each The changing characteristics of the component gas yields, and the reaction mechanism is reviewed based on relevant literature.

2.3 Question three

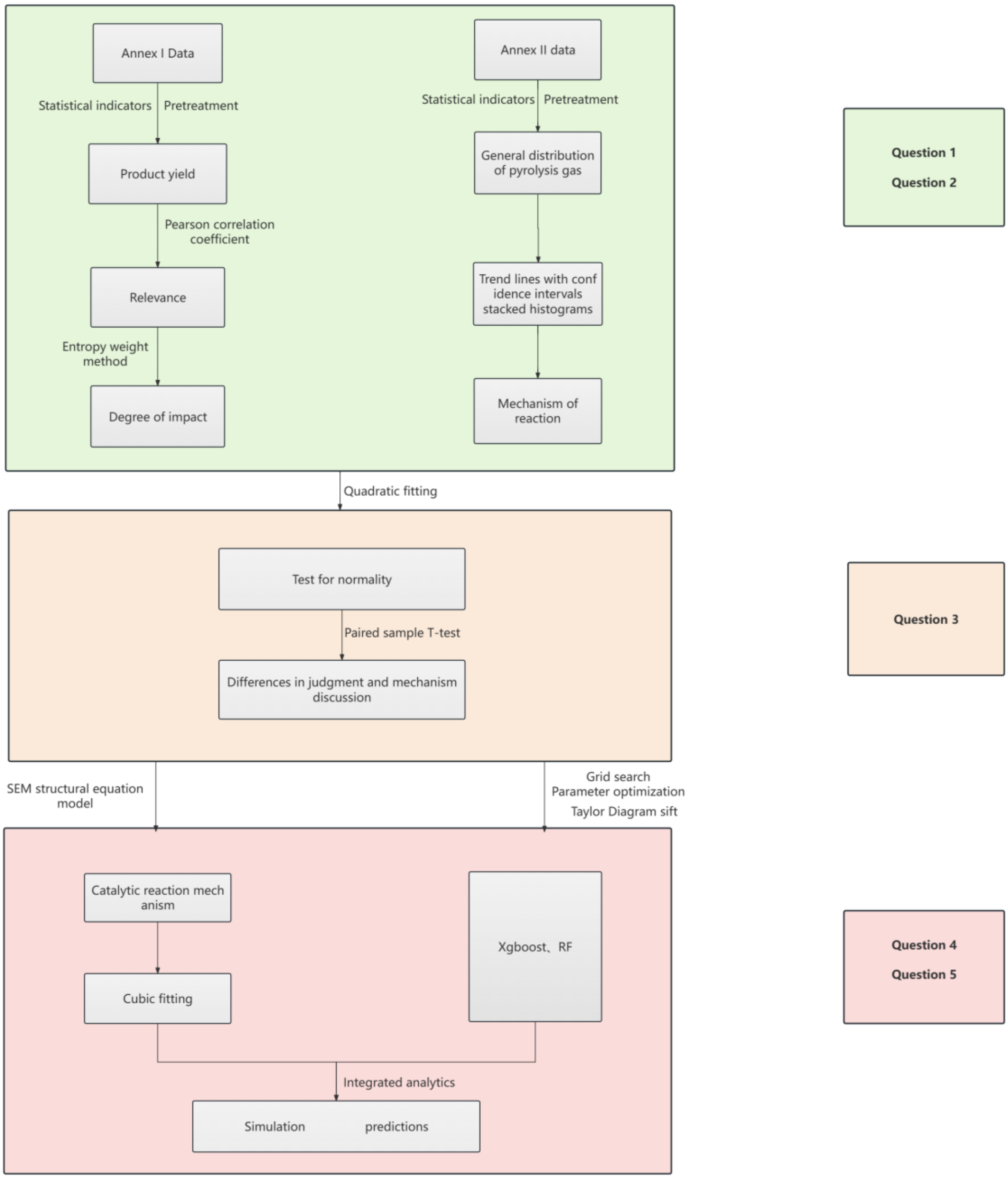
This question requires an analysis of whether there is a significant difference in the yields of CE and LG pyrolysis products and the yields of each component of pyrolysis gas under the catalytic action of the same desulfurization ash ratio . First , based on the secondary questions in question one and question two Fit the equation, the mixing ratio is 0 to 1, and the interval is 0.01 . Predict the yield of CE and LG pyrolysis products and the yield of pyrolysis gas components. Perform data expansion and use QQ plot and Shapiro-Wilk method. The predicted data set was tested for normality, and the yields of each pyrolysis product and pyrolysis gas were normally distributed. Finally, a paired sample t test was used to test the yield of each pyrolysis product under the catalytic action of the same proportion of desulfurization ash and Determine whether there is a significant difference in the yield of pyrolysis gas components and discuss the reaction mechanism.

2.4 Question 4

This question requires the establishment of a catalytic reaction mechanism model of desulfurization ash of model compounds such as CE and LG, and a mathematical model to analyze the reaction kinetic mechanism. First, the catalytic reaction mechanism of desulfurization ash of model compounds such as CE and LG is described through an SEM structural equation model . Finally, a three-dimensional fitting (average R^2 =0.96) was used to establish a reaction kinetic model to quantitatively analyze the catalytic reaction mechanism of desulfurization ash.

2.5 Question 5

This question requires predicting the yield or quantity of pyrolysis products under limited data conditions. First, four machine learning algorithms (random forest, XGBoost, gradient boosting regression tree and K-nearest neighbor) are selected for model training, and use Grid search for parameter optimization. Then the Taylor diagram can effectively display the simulation effect of the model. Finally, XGBoost and RF algorithms were selected and combined with the reaction kinetic model constructed in question 3 to construct an integrated model of three learners to predict the production of pyrolysis products in different pyrolysis combinations. The rate and pyrolysis gas production were simulated and predicted.



3 Symbols and Assumptions

3.1 Symbol Description

symbol	significance
μ_i	average value
σ_i	standard deviation
Max_i	maximum value
Min_i	minimum value
ξ_{ki}	kurtosis
ξ_{si}	Skewness
CV_i	coefficient of variation
R^2	decisive factor

3.2 Basic assumptions

(1) Assume that the data given in the question are accurate data obtained in the laboratory, and the error in the model can be ignored;

(2) It is assumed that the pyrolysis product yield and pyrolysis gas yield can completely reflect the state of pyrolysis activity;

(3) Assume that the pyrolysis reaction process is stable and the changes in the output and yield of each product are continuous;

(4) The role of other external influencing factors (such as reaction time, atmospheric pressure) during the pyrolysis process is not considered ;

(5) It is assumed that the proportion of desulfurization ash plays a decisive role in the pyrolysis reaction of substances, and other factors have no significant influence on the pyrolysis reaction.

4 Data preprocessing

(1) Convert the ratios of D FA/CS , DFA / CE, and DFA /LG in Annex 1 and Annex 2 into floating point numbers to facilitate subsequent equation fitting; convert the gas production of each component in Annex 2 into the corresponding yield.

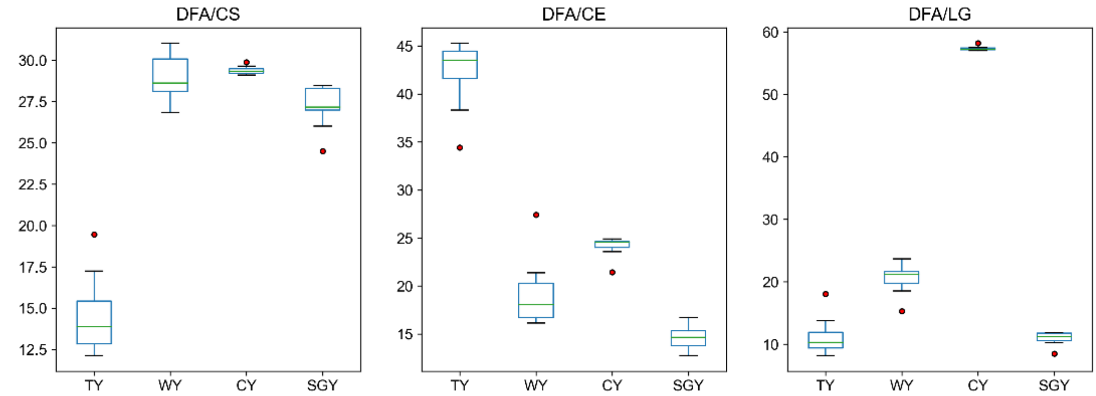
(2) Missing value processing

Use python 's pandas library to search for missing values in the data in Appendix 1 and Appendix 2. If there are missing values, remove the samples containing missing values or interpolate the missing values to avoid confusing the data mining process. Resulting in unreliable output. By searching, if there are no missing values or NAN values in the data, there is no need to perform missing value processing.

(3) Identification of outliers

The 3sigma principle, also known as the 3 σprinciple , is a statistical method based on the characteristics of the normal distribution, used to identify outliers or outliers. This principle means that under a normal distribution, approximately 68% of the data fall within plus or minus one standard deviation of the mean, 95% of the data fall within plus or minus two standard deviations of the mean, and 99.7% of the data fall within the range of plus or minus one standard deviation of the mean. Within plus or minus three standard deviations of the mean. If a data point is more than three standard deviations from the mean, that is, it falls outside the range of plus or minus three standard deviations from the mean, it can be regarded as an outlier or abnormal value [1]. Use the 3 σprinciples to identify outliers in the data and find that there are no outliers in the data.

Box plots are often used to display the dispersion of data. They are mainly used to reflect the distribution characteristics of the original data and can also be used to compare the distribution characteristics of multiple groups of data. Use box plots to describe the distribution characteristics of the data in Appendix 1. Part of the results are shown in Figure 2 , and the rest are in the Appendix.



Note: TY is tar production; WY is water production; CY is carbon yield; SGY is syngas yield.

Figure 2 Data distribution box plot

From Figure n, the tar production in DFA/CS , DFA /CE and D FA/LG is concentrated in the range of 1 2.6 ~ 15.1 , 4 2.1 ~ 43.4 and 9.7 ~ 12.1 respectively ; the water production is concentrated in the range of 2 8.3 ~ 29.2 respectively. , 17.0 ~ 20.5 and 19.5 ~ 21.4 ranges; the carbon yields are concentrated in the 2 8.6 ~ 29.2 , 2 4.1 ~ 24.8 and 5 7.2 ~ 57.9 ranges respectively; the synthesis gas yields are concentrated in the 2 6.6 ~ 28.0 and 1 3.4 ~ 15.6 ranges respectively. and the 11.2 ~ 12.3 interval.

5 Model

5.1 Establishment and solution of the model for Problem 1

5.1.1 Descriptive statistics

/CS , DFA /CE and DFA /LG pyrolysis $\mu_i$in Appendix 1 , and discuss their statistical rules, namely mean , standard deviation σ_i, maximum Max_i, kurtosis ξ_{ki}, minimum value Min_i, skewness ξ_{si}, and coefficient of variation CV_i.

①Kurtosis _

Kurtosis is a statistic used to describe the sharpness of a probability distribution or data shape distribution. It reflects the tail thickness and peak degree of the data distribution. The calculation formula is as follows:

$$\xi_{k_i} = \frac{1}{n}\sum_{i=1,j=1}^{n}\left[\frac{(x_{ij}-\mu_i)^4}{\sigma_i}\right] \tag{1}$$

The value range of kurtosis is [$-\infty, +\infty$]. The larger the kurtosis value, the steeper the peak part of the data distribution and the more concentrated peak value. The smaller the kurtosis value, the flatter the peak part of the data distribution and the more dispersed distribution.

② Skewness

Skewness is a statistic used to describe the skewness of a probability distribution or data distribution. It reflects the direction and degree of deviation of the data on the distribution curve. The calculation formula is as follows:

$$\xi_{s_i} = \frac{1}{n}\sum_{i=1,j=1}^{n}\left[\frac{(x_{ij}-\mu_i)^3}{\sigma_i}\right] \tag{2}$$

The value range of skewness is ($-\infty, +\infty$) , $\xi_{si} > 0$, indicating that the distribution is skewed to the right, with heavy tails on the right; $\xi_{si} < 0$, indicating that the distribution is skewed to the left, with heavy tails on the left; $\xi_{si} = 0$, indicating that the data distribution is relatively symmetrical, and the heavy tail distribution is relatively balanced.

③Coefficient of variation

The coefficient of variation, also known as the standard deviation rate, is a statistic used to measure the dispersion of data relative to its mean. The calculation formula is as follows:

$$CV_i = \frac{\sigma_i}{\mu_i} \times 100\% \tag{3}$$

When the coefficient of variation is small, the relative dispersion of the data is small, and the data values are closer to the mean; when the coefficient of variation is large, the relative dispersion of the data is large, and the data values are further away from the mean.

The descriptive statistics results are shown in Table 1 :

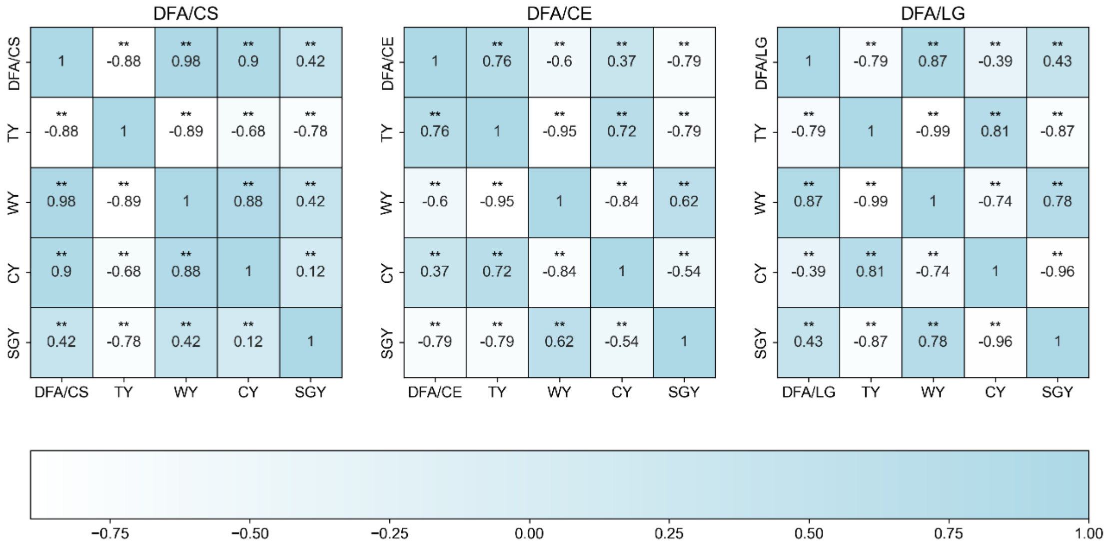
Table 1 Descriptive statistics results

Group	Index	$\bar{x} \pm s$	Max	Min	Kurtosis	Skewness	Coefficient
CFA/CS	Carbon yield	14.55 ±2.43	19.46	12.13	0.78	1.23	16.71
	Water production	28.91 ±1.42	31.02	26.84	-1.03	0.26	4.90
	Syngas yield	29.38 ±0.25	29.87	29.11	0.68	1.06	0.84
	Tar production	27.16 ±1.28	28.45	24.49	1.36	-1.17	4.71
CFA/CE	Carbon yield	42.08 ±3.76	45.28	34.42	1.64	-1.56	8.94
	Water production	19.26 ±3.76	27.42	16.14	3.11	1.71	19.51
	Syngas yield	24.07 ±1.14	24.91	21.43	5.04	-2.20	4.74
	Tar production	14.59 ±1.27	16.73	12.75	-0.11	0.18	8.71
CFA/LG	Carbon yield	11.30 ±3.21	18.06	8.19	2.42	1.58	28.38
	Water production	20.44 ±2.54	23.69	15.30	1.93	-1.19	12.42
	Syngas yield	57.34 ±0.37	58.17	56.98	4.04	1.89	0.65
	Tar production	10.93 ±1.15	11.88	8.47	2.63	-1.60	10.55

Generally speaking, the ratio of pyrolysis products of cotton straw is relatively uniform, and the importance of each pyrolysis product is equal; among the pyrolysis products of cellulose, tar production accounts for the highest proportion, which is more than three times the synthesis gas yield with the lowest proportion; in In the pyrolysis process of lignin, the carbon yield accounts for more than 50% and is the most important component of the pyrolysis process of lignin. During the pyrolysis process of cotton straw, the coefficient of variation of tar production is the largest, and the proportion fluctuates the most; the coefficient of variation of water production in the pyrolysis products of cellulose is large, and the coefficients of variation of the other three products are small; carbon in lignin The coefficient of variation of the yield is less than 1, which means that as the desulfurization ash ratio changes, there is almost no change in the carbon yield.

5.1.2 correlation analysis

Using Python 's seaborn library , P earson correlation analysis was performed on different ratios of D FA/CS , D FA/CE and DFA /LG and the corresponding yields of decomposition products obtained from pyrolysis, and significance analysis was performed. The analysis results are shown in the figure 3 shows:



Note: * means 0.01 < p <0.05 , * * means p <0.01 ; TY means tar production, WY means water production, CY means carbon yield, and SGY means syngas yield.

Figure 3 Correlation coefficient diagram

The different ratios of the three pyrolysis combinations of DFA/CS, DFA/CE and DFA/GL are significantly correlated with tar production, water production, char yield and syngas yield at the 99% confidence interval. For DFA/CS, its ratio has a strong and significant negative correlation with tar production, and a significant positive correlation with water production, char yield and syngas yield, but the correlation with syngas yield is obviously weak; tar production There is an obvious negative correlation with the other three products ; the water yield has a strong correlation with the char yield and a weak positive correlation with the syngas yield ; there is only a weak positive correlation between the char yield and the syngas yield. Correlation . For DFA/CE, its ratio has a significant negative correlation with water production and syngas yield, and a positive correlation with tar production and carbon yield, but the correlation with carbon yield is obviously weak; tar production is significantly related to water production and syngas yield. The yield shows an obvious negative correlation and a positive correlation with the carbon yield; the water yield has a strong negative correlation with the carbon yield and a positive correlation with the syngas yield; the carbon yield has a general negative correlation with the syngas yield. . For D FA/LG, its ratio is positively correlated with water production and syngas yield, but has a weak correlation with syngas yield, and is negatively correlated with tar production and carbon yield, but has a weak correlation with carbon yield. Weak; tar production has a significant negative correlation with water production and syngas yield, and a strong positive correlation with carbon yield; water production has a strong negative correlation with carbon yield, and a strong positive correlation with syngas yield ; There is a significant negative correlation between carbon yield and syngas yield.

5.1.3 quadratic fit

Quadratic fitting approximates known data points through a quadratic polynomial function , with the goal of finding a quadratic function equation that minimizes the error between the function and the known data points. Through quadratic fitting, the trends and relationships of data can be better understood and used in fields such as data analysis, regression analysis, and curve fitting. Assume $y_i = \hat{a}x^2 + \hat{b}x + \hat{c}$, $y_{ij}$is the yield of each decomposition product obtained from pyrolysis of different pyrolysis combinations, and $x_{ik}$is the different ratio of catalysts in different pyrolysis combinations . in:

$$\begin{cases} \hat{a} = \dfrac{(\overline{x_{ik}y_{ij}}-\overline{x_{ik}}\,\overline{y_{ij}})(\overline{x_{ik}^3}-\overline{x_{ik}}\,\overline{x_{ik}^2})-(\overline{x_{ik}^2y_{ij}}-\overline{x_{ik}^2}\,\overline{y_{ij}})(\overline{x_{ik}^2}-(\overline{x_{ik}})^2)}{(\overline{x_{ik}^3}-\overline{x_{ik}}\,\overline{x_{ik}^2})^2-(\overline{x_{ik}^4}-\left(\overline{x_{ik}^2}\right)^2\overline{x_{ik}^2})(\overline{x_{ik}^2}-(\overline{x_{ik}})^2)} \\ \hat{b} = \dfrac{\overline{x_{ik}y_{ik}}-\overline{x_{ik}}\,\overline{y_{ij}}-\hat{a}(\overline{x_{ik}^3}-\overline{x_{ik}}\,\overline{x_{ik}^2})}{\overline{x_{ik}^2}-(\overline{x_{ik}})^2} \\ \hat{c} = \overline{y_{ij}} - \hat{a}\overline{x_{ik}^2} - \hat{b}\overline{x_{ik}} \end{cases} \tag{4}$$

i=1,2,3; j=1,2,3,4; k=1,2,…,9

The coefficient of determination is used R^2as an index to evaluate the goodness of fit, and the calculation formula is as follows:

$$R^2 = 1 - \frac{\sum(\widehat{y_{ij}}-y_{ij})^2}{\sum(y_{ij}-\overline{y_{ij}})^2} \tag{5}$$

Python was used to fit the relationship between the proportions of catalysts in different pyrolysis combinations and the yields of each decomposition product obtained, and the goodness of fit was calculated. The results are shown in Figure 4 :

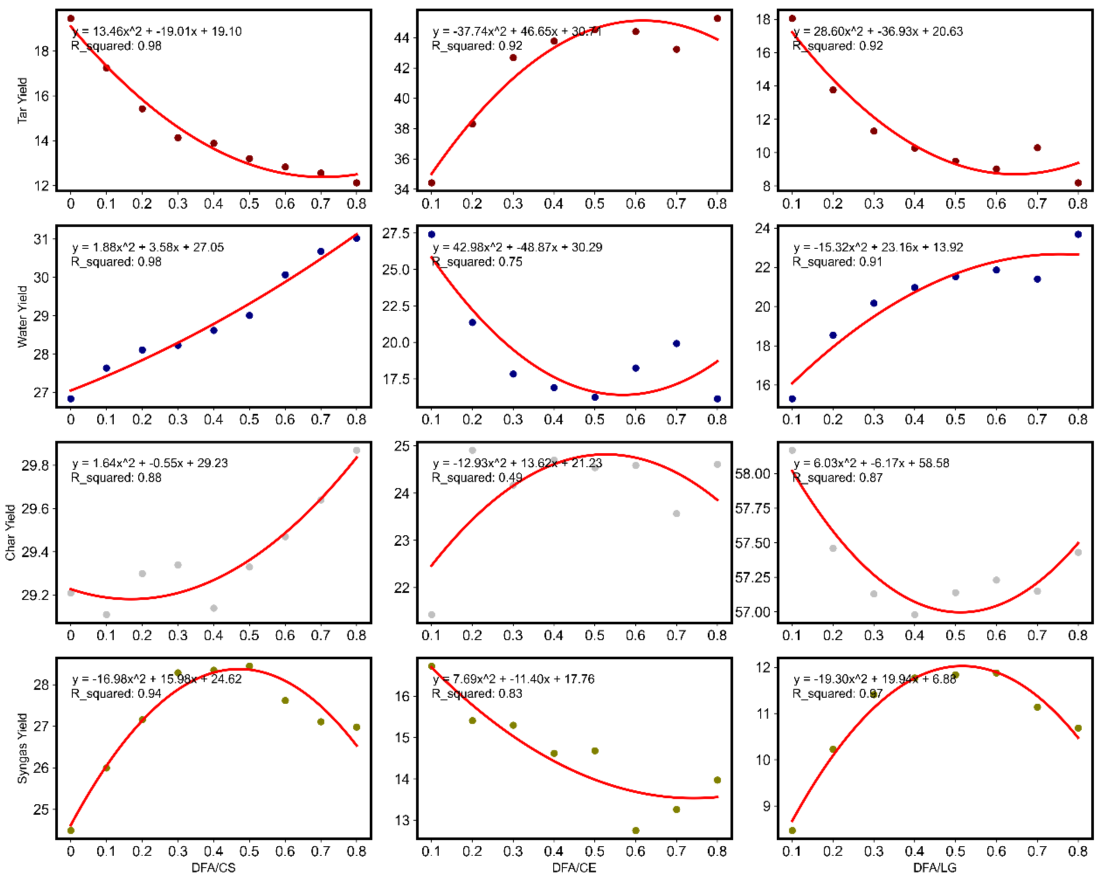


Figure 4 Quadratic fitting

As far as DFA /CS is concerned, the tar production and carbon yield continue to decrease as the ratio increases, the water production increases as the ratio decreases, and the syngas yield increases first with the ratio, reaching The maximum value (about 2 9) and then gradually decreases; in terms of D FA/CE , the tar production and carbon yield first increase with their ratio, reach the maximum value (about 4 6 and 2 5) and then gradually decrease. is small, the water production first decreases to the minimum value (about 1 6) and then gradually increases, and the synthesis gas yield gradually decreases with its ratio; in terms of D FA/LG , the tar production and carbon yield increase with its ratio. The ratio first decreases to the minimum value (about 8.2 and 5.7) and then gradually increases. The water production gradually increases with the ratio. The synthesis gas yield first increases with the ratio and reaches the maximum value (about 1.2). then gradually decreased.

5.1.4 entropy weight method

The entropy weight method is a multi-criteria decision-making method that uses the concept of information entropy to determine the weight of different influencing factors to help decision-making or evaluation [2] . The entropy weight method was initially used in the field of environmental impact assessment, and has since been widely used in various decision-making and assessment problems . The specific steps are as follows:

(1) Min-Max normalization

Using M in-Max standardization, the linear function linearizes the original data to the range of [0,1] , and the calculation result is the normalized data. The standardized formula is as follows:

$$y_{ijk}' = \frac{y_{ijk}-\min(y_{ijk})}{\max(y_{ijk})-\min(y_{ijk})} \text{(When the indicator is positive)} \tag{6}$$

$$y_{ijk}' = \frac{\min(y_{ijk})-y_{ijk}}{\max(y_{ijk})-\min(y_{ijk})} \text{(When the indicator is negative)} \tag{7}$$

Which represents the yield of different decomposition products obtained by different pyrolysis combinations at different ratios.

(2) Find the ratio of each indicator

In fact, it is to calculate the variation of various indicators. Calculated as follows:

$$P_{ijk} = \frac{y_{ijk}'}{\sum_{i=1}^{n} y_{ijk}'} \tag{8}$$

(3) Find the information entropy of each indicator

The amount of information is a measure of how much information is required to figure out an unknown thing, and the unit is bits. The formula is as follows:

$$I(y_{ijk}) = \log_2\left(\frac{1}{P_{ijk}}\right) = -\log_2 P_{ijk} \tag{9}$$

Information entropy is the expectation of the amount of information, which can be understood as the size of uncertainty. The greater the uncertainty, the greater the information entropy. The formula is as follows:

$$H(y_{ijk}) = \sum P_{ijk} \times \log_2\left(\frac{1}{P_{ijk}}\right) \tag{10}$$

(4) Calculate the weight of each indicator through information entropy

Calculated as follows:

$$w(y_{ijk}) = \frac{1-H(y_{ijk})}{k-\sum H(y_{ijk})} \tag{1 1 }$$

The entropy weight method is used to calculate the weight. The results are shown in Table 2 , and the visualization results are shown in Figure 5 .

Table 2 Weight results table

Group	Index	Information Entropy	Information Utility	Weight (%)
CAF/CS	Carbon yield	0.823	0.177	40.695
	Water production	0.887	0.113	25.854
	Syngas yield	0.929	0.071	16.206
	Tar production	0.925	0.075	17.245
CAF/CE	Carbon yield	0.931	0.069	21.371
	Water production	0.919	0.081	25.16
	Syngas yield	0.926	0.074	22.857
	Tar production	0.901	0.099	30.612
CAF/LG	Carbon yield	0.92	0.08	26.352
	Water production	0.925	0.075	24.856
	Syngas yield	0.923	0.077	25.312
	Tar production	0.929	0.071	23.481

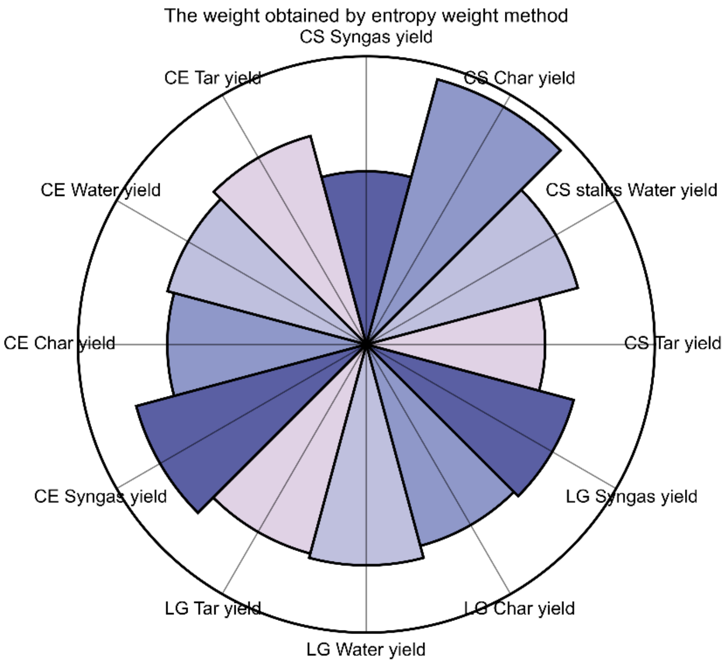


Figure 5 Weight rose chart

It can be seen intuitively from the weight result table and Nightingale rose diagram obtained by the entropy weight method that there are obvious differences in the effects of different pyrolysis combinations on the ratio of each pyrolysis product in cotton straw, cellulose and lignin. The maximum weight value appears in the carbon yield of cotton straw, and the syngas ratio of cotton straw is the least affected by changes in the pyrolysis combination; the weight of water production has a small difference in the proportions of the three substances, indicating that water molecules play a role in thermal It is a relatively stable component during the solution process.

5.2 Model establishment and solution of problem 2

5.2.1 Descriptive statistics

Perform descriptive statistics on the gas production obtained from the pyrolysis of D FA/CS, D FA / CE and DFA /LG in Appendix 2. The statistics are mean, standard deviation, maximum value, minimum value, kurtosis, skewness and variation respectively. Coefficients describe the basic characteristics of the data and discuss their statistical laws. The results are shown in Table 3 :

Table 3 Descriptive statistics

Group	Index	$\bar{x} \pm s$	min	max	Kurtosis	Skewness	Coefficient
DFA/CS	H2	2 5.66 ±14.38	6.48	43.62	-1.52	-0.14	0.56
	CO	2 2.39 ±.98	20.53	25.93	1.62	1.35	0.09
	CO2	5 4.15 ±3.96	50.91	60.40	-0.83	1.09	0.07
	CH4	2 9.80 ±3.49	23.41	33.11	2.38	-1.45	0.12
	C2H6	2.80 0.37 _±	2.09	3.08	3.95	-1.96	0.13
	C3H8	0.71 0.14 _±	0.55	0.88	-1.96	0.45	0.20
	C3H6	0.38 0.02 _±	0.35	0.41	0.52	0.33	0.05
	C2H4	0.76 0.04 _±	0.71	0.81	-0.64	-0.09	0.05
	C4H10	0.38 0.12 _±	0.25	0.56	-0.99	0.71	0.32
DFA/CE	H2	5 0.74 ±14.94	26.60	65.80	1.89	-1.27	0.29
	CO	8 .84 ±3.36	5.30	13.70	-0.63	0.69	0.38
	CO2	7.92 6.73 _±	2.00	18.70	1.30	1.28	0.85
	CH4	1.68 0.95 _±	0.70	3.20	1.69	1.20	0.57
	C2H6	6.68 4.54 _±	3.20	13.80	0.45	1.24	0.68
DFA/LG	H2	0.75 0.06 _±	0.67	0.81	-0.89	-0.85	0.08
	CO	1 7.66 ±2.36	15.23	20.94	-1.33	0.52	0.13
	CO2	3 7.26 ±7.53	24.35	43.56	3.55	-1.79	0.20
	CH4	3 0.98 ±2.56	27.67	33.86	-1.65	-0.06	0.08
	C2H6	0.07 0.02 _±	0.05	0.09	-1.20	0.00	0.23

Generally speaking, there are obvious differences in the pyrolysis gas products of cotton straw. The CO_2 production exceeds 5 0 ml / g, C_3H_8, C_2H_6, C_3H_6, C_2H_4, C_4H The proportion of $_{10}$ is less than 0.1 ml /g. This situation is different between cellulose and lignin: the production of H2 in cellulose is significantly lower than that in cotton straw (0.75 ml / g ±0.06) , and in lignin The yield of C_2H_6 is clearly small, as is the share of the other two substances. The $H_{2\ \text{yield}}$ in cotton straw has the largest change, with a range of 2 7.14 ml / g. The smallest fluctuations are C_2H_4 and C_3H_6. The coefficients of variation of both are 0.03 , indicating that they are affected by the desulfurization ash ratio. The impact of the changes is minor.

5.2.2 The relationship between the mixing ratio of different pyrolysis combinations and the output gas

Among the three pyrolysis combinations DFA q_{ijk}/CS , DFA /CE and DFA/LG in Appendix 2 , the different gas productions at different mixing ratios are added up to obtain the total gas production of different pyrolysis combinations and different mixing ratios Q_{ik}, i represents different pyrolysis Combination, j represents different gases obtained by decomposition, and k represents different mixing ratios.

$$\begin{aligned} Q_{1k} &= \sum q_{1jk}, j = 1,2,\dots,9, k = 1,2,\dots,6 \\ Q_{2k} &= \sum q_{2jk}, j = 1,2,\dots,5, k = 1,2,\dots,5 \\ Q_{3k} &= \sum q_{3jk}, j = 1,2,\dots,5, k = 1,2,\dots,5 \end{aligned} \tag{12}$$

The relationship between the mixing ratio of different pyrolysis combinations and the total gas production is shown in Figure 6 :

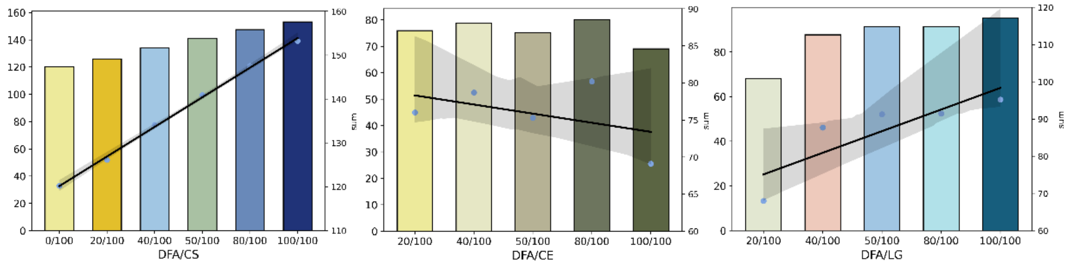


Figure 6 Total gas production of different pyrolysis combinations

In DFA /CS and DFA/LG, the total gas production increases with the increase in the proportion of desulfurization ash. In DFA/CE, the total gas production fluctuates slightly with the increase in the proportion of desulfurization ash. At 8 0 / Maximum is reached at 100 .

By drawing stacked histograms of different gases in three different pyrolysis combinations of DFA/CS, DFA/CE and DFA/LG in Appendix 2 , the interaction between the mixing ratio and the yield of different gases is analyzed, and the influence mechanism between the two is explored. , as shown in Figure 7 :

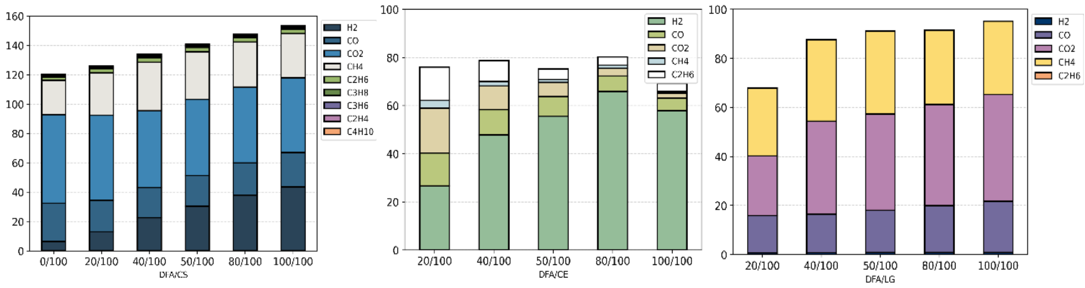


Figure 7 Stacked bar chart of gas production of each component

In DFA /CS , as the proportion of desulfurization ash increases, H_2 production increases, CO production first decreases and then increases, CO_2 decreases, CH_4 and C_2H_6 increase first and then decrease, There are also small amounts of C_3H_6, C_3H_8, C_2H_4 and C_4H_{10} gases in the DFA/CS pyrolysis gas; in DFA / CE , as the proportion of desulfurization ash increases, H2 The output increased significantly, and the output of CO, CO_2, CH_4 and C_2H_6 gradually decreased; in DFA/LG, the H_2 output was low and almost did not change with the increase in the proportion of desulfurization ash, and the C_2H_6 output was extremely Low, CO and CO_2 production increases with the increase of desulfurization ash, and CH_4 production increases first and then decreases.

5.2.3 Fitting curves of gas yields for different pyrolysis combinations

5.2.3.1 D FA /CS

Use Python to perform secondary fitting to explore the changing characteristics of CS pyrolysis gas yield under the catalysis of different proportions of desulfurization ash, capture the changing trend between CS pyrolysis gas yield and different proportions of desulfurization ash, and analyze the relationship between the two. The mutual influence relationship of , and the goodness of fit is calculated at the same time. The results are shown in Figure 8 :

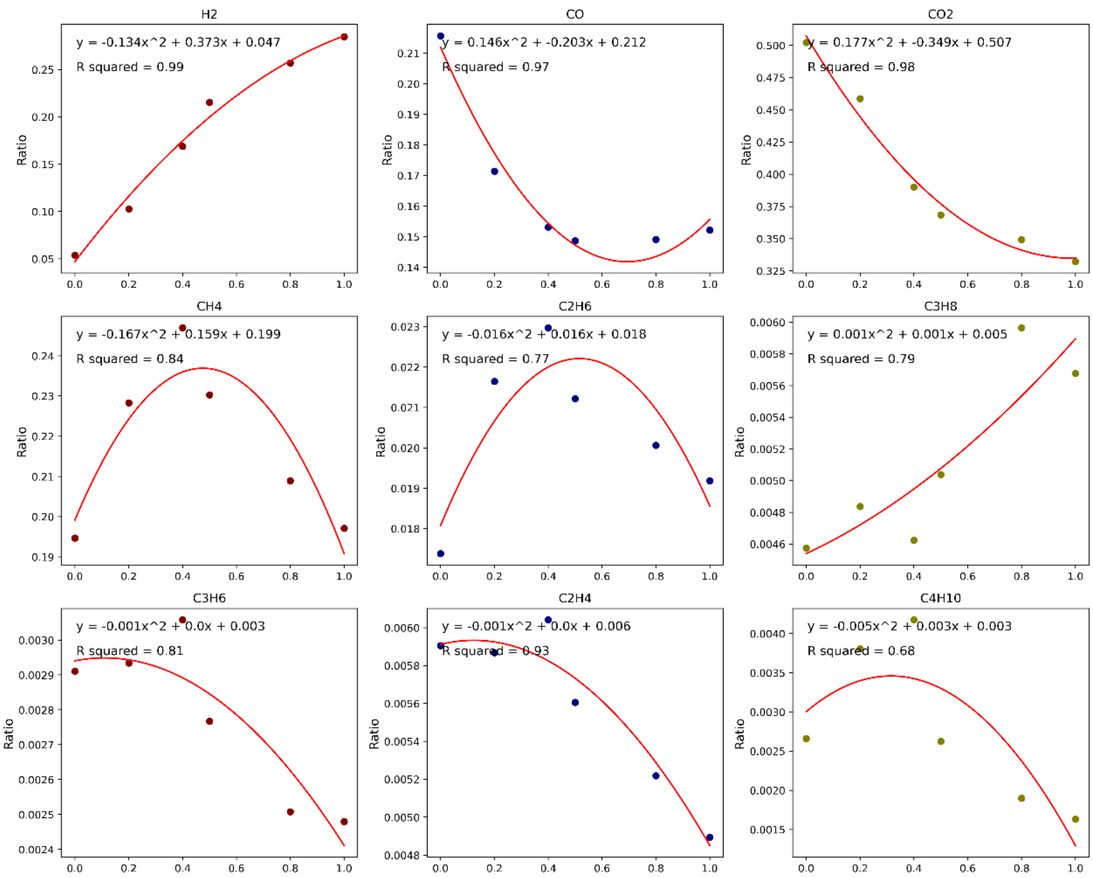


Figure 8 DFA/CS quadratic fitting

The H2 yield increases significantly with the increase in the ratio of desulfurization ash to CS; CO and CO_2 in the pyrolysis gas mainly come from the breakage of ether bonds and carboxyl groups [3], and the CO yield increases with the increase in the ratio of desulfurization ash. It first decreases, reaches the minimum value (about 0.14) at the 0.5 ratio , and then rises; the CO_2 production rate is inhibited as the desulfurization ash ratio increases, and the CH_4 production rises first, at the 0.5 ratio . reaches the maximum value (about 0.24) and then decreases; desulfurization ash has almost no impact on the yield of C_2H_6. In addition, there are very small amounts of C_3H_8, C_3H_6, and C_2H in the produced gas. Gases such as $_4$ and C_4H_{10}.

5.2.3.2 DFA /CE

Using Python to perform secondary fitting, the purpose is to study the changing characteristics of CE pyrolysis gas yield under the catalysis of different proportions of desulfurization ash, in order to capture the correlation trend between CE pyrolysis gas yield and proportion of desulfurization ash, and to conduct in-depth research on CE thermal The relationship between degassing yield and different proportions of desulfurization ash was calculated, and the goodness of fit was calculated to evaluate the quality and applicability of the model. The results are shown in Figure 9 :

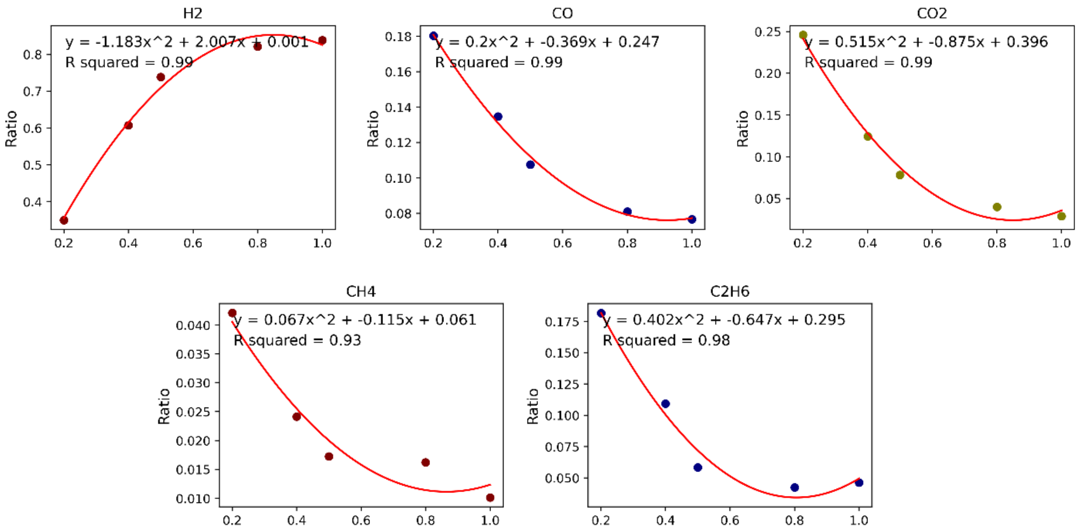


Figure 9 DFA/CE quadratic fitting

The H_2 yield increases with the increase in the ratio of desulfurization ash to CE, indicating that desulfurization ash promotes the reaction of CE pyrolysis to produce hydrogen radicals; however, the increase of desulfurization ash inhibits the reaction of CE decomposition into CO and CO_2, causing CO The yield of CO_2 continues to decrease; the yield of CH_4 is extremely small, and the increase in the proportion of desulfurization ash has a significant inhibitory effect on the generation of CH_4 ; the yield of C_2H_6 significantly decreases with the increase in the proportion of desulfurization ash, it can be seen that Under the catalysis of desulfurization ash, CE generates more free radicals with smaller molecular weights such as carbonyl groups and carboxyl groups, thereby inhibiting the process of generating C_2H_6 free radicals from CE structural units.

5.2.3.3 D FA/LG

LG pyrolysis gas production rate under the catalysis of different proportions of desulfurization ash , and to conduct an in-depth study of the relationship between LG pyrolysis gas production rate and different proportions of desulfurization ash. We fit the experimental data and use a quadratic polynomial model to describe the correlation trend between the two, and evaluate the degree of agreement between the fitted model and the actual data by calculating the goodness of fit, thereby determining the reliability and applicability of the model. properties , the results are shown in Figure 10 :

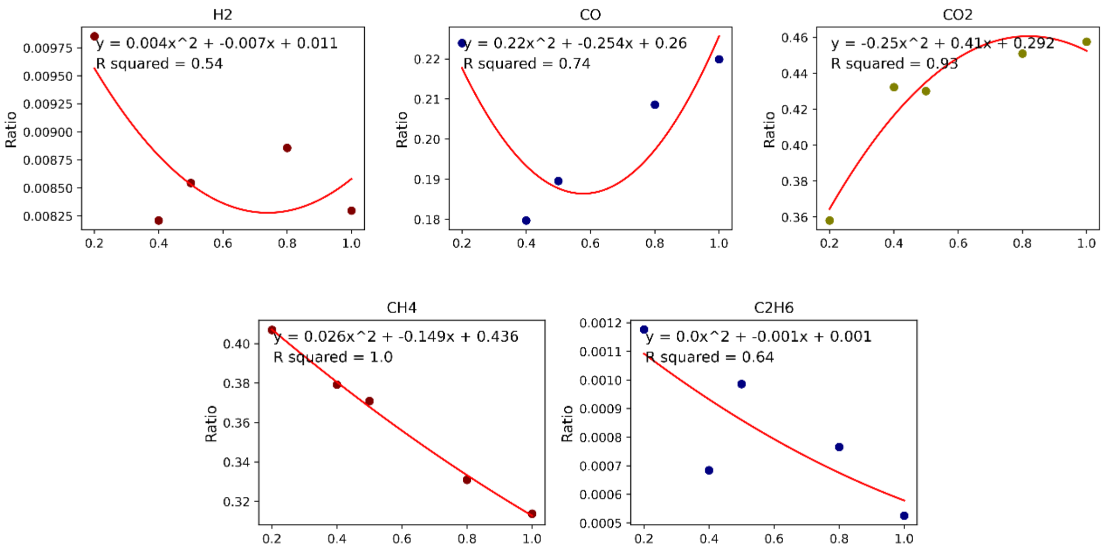


Figure 10 DFA/LG quadratic fitting

The H2 yield is at an extremely low level. It can be seen that H2 mainly comes from the decomposition process of CE; the CO yield first decreases with the increase in the proportion of desulfurization ash, reaching the minimum value (about 0.19) at a ratio of 0.6 . then rises; the catalytic effect of desulfurization ash promotes the reaction of LG pyrolysis to generate CO_2, and the CO2 yield increases with the increase in the proportion of desulfurization ash; the CH_4 yield decreases with the increase of desulfurization ash, but its value is higher, it can be seen that CH_4 mainly comes from the break of the methoxy side chain contained in LG; it is difficult to generate C_2H_6 in DFA/LG , because the C_2H_6 generated by LG mainly comes from the break of - CC- bond, while the aromatic ring structure in LG is difficult to destroy.

5.3 Model establishment and solution of problem three

5.3.1 Normality test

For Annex 1 and Annex 2, the pyrolysis of CE and LG under the catalytic action of different desulfurization ash ratios yields products such as tar and water, as well as gas products such as H 2 and CO respectively. In order to select a suitable differential analysis method, the production of these products is The normality test is performed on the yield. Since the sample size is less than 5000 and it is a small sample, the Shapiro-Wilk test is used . The test results are shown in Table 4. It can be seen that the p-value of the yield of all products is less than 0.05 , so the original criterion is rejected. Assume and use MATLAB to draw the QQ diagram, part of which is shown in Figure 1 1. See the appendix for detailed results. The distribution is a straight line, indicating that it basically conforms to the normal distribution.

Table 4 SW inspection table

Group	Variable	Sample size	SW test
DFA/CE	Tar yield	100	0.793(0.0564**)
	Water yield	100	0.819(0.095**)
	Char yield	100	0.709(0.153***)
	Syngas yield	100	0.978(0.951)
	H2	100	0.909(0.461)
	CO	100	0.954(0.767)
	CO2	100	0.889(0.354)
	CH4	100	0.904(0.435)
	C2H6	100	0.836(0.153)
DFA/LG	Tar yield	100	0.843(0.081*)
	Water yield	100	0.906(0.325)
	Char yield	100	0.795(0.125**)
	Syngas yield	100	0.828(0.057*)
	H2	100	0.907(0.453)
	CO	100	0.943(0.685)
	CO2	100	0.813(0.102)
	CH4	100	0.929(0.591)
	C2H6	100	0.987(0.967)

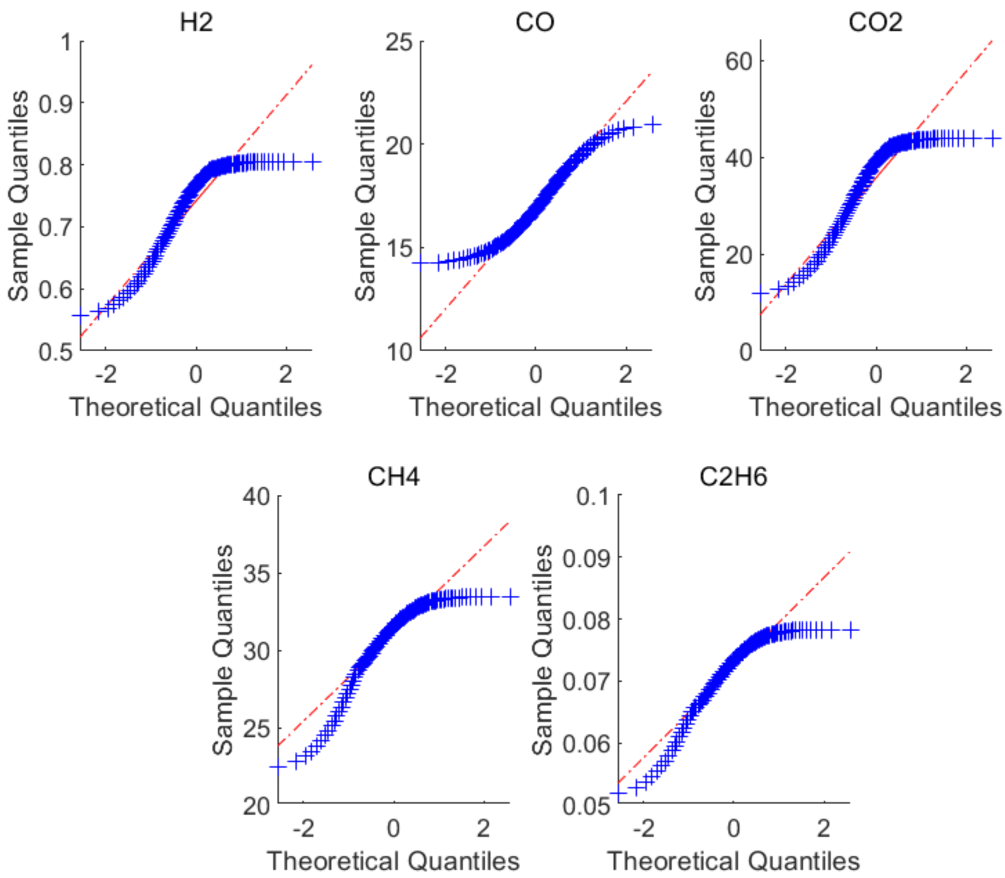


Figure 11 Q - Q diagram

5.3.2 Paired samples T-test

If all product yields pass the normality test, paired samples T test can be used to conduct difference analysis to explore whether there are significant differences between the yields of CE and LG pyrolysis products and each component gas under the same catalytic ratio of desulfurization ash. , the results are shown in Table 5 :

Table 5 Paired sample test results table

Pairing variable	X±S			t	P	Cohen's d
	DFA/CE	DFA/LG	Pairing difference			
Tar yield	42.082±3.764	11.3±3.207	30.782±0.557	12.52	0.000***	4.426
Water yield	19.262±3.759	20.435±2.539	-1.173±1.22	-0.538	0.607	2.642
Char yield	24.065±1.141	57.336±0.373	-33.271±0.768	-64.763	0.000***	22.897
Syngas yield	14.59±1.271	10.929±1.153	3.661±0.118	4.621	0.002***	1.634
H2	0.671±0.201	0.009±0.001	0.662±0.2	7.509	0.002***	3.358
CO	0.116±0.043	0.204±0.019	-0.088±0.024	-3.497	0.025**	1.564
CO2	0.104±0.088	0.426±0.04	-0.322±0.048	-4.629	0.010***	2.07
CH4	0.022±0.012	0.36±0.038	-0.338±-0.025	-21.177	0.000***	9.471
C2H6	0.088±0.059	0.001±0	0.087±0.059	3.259	0.031**	1.457

Note: ***, **, and * represent the significance levels of 1%, 5%, and 10% respectively.

The visualization results are shown in Figure n:

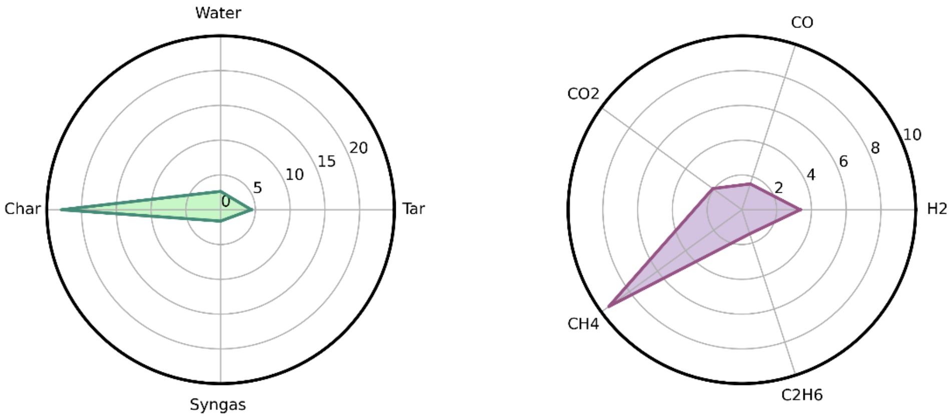


Figure 12 Significant difference radar chart

Cohen's d is a commonly used effect size measure to measure the size of the difference between two samples. It is used to compare the difference between two means and is combined with the standard deviation to express the size of the difference. Its value less than 0.2 indicates that the magnitude of the difference is very small . Between 0.2 and 0.5 indicates a moderate difference, and greater than 0.5 indicates a large difference. A P value less than 0.05 indicates a significant difference. Tar yield, carbon yield, syngas yield, H_2, CO_2 and CH_4 yields show significant differences at the 1% level. Among them, the strongest difference is the carbon yield, with a difference of 2 2.897 , the weakest difference is the synthesis gas yield, the difference is only 1.634 ; the yields of CO and C_2H_6 show significant differences at the 5% level, the differences are similar, but both are relatively strong; in Under the catalytic action of the same desulfurization ash ratio, there is no significant difference in water yield between CE and LG pyrolysis .

5.3.3 Result analysis

Different reaction conditions such as reaction temperature, pressure, and catalyst characteristics will lead to different reaction pathways and product selections. Different reaction materials have different cracking and polymerization properties, resulting in significant differences in tar yields .

Factors that may affect carbon yield include reaction temperature, catalyst characteristics, and selection of reaction materials. Higher reaction temperatures and higher concentration ratios of catalysts may promote hydrocarbon cracking and polymerization, resulting in high carbon yields. In addition, the choice of reactive species may also lead to different carbon yields, because the composition and reaction rate of hydrocarbons produced during the cracking process may be different for different reactive species.

Syngas is typically produced by methane (CH_4) reforming reactions or gasification reactions. Therefore, the synthesis gas yield may be affected by factors such as catalyst concentration , reaction temperature, and selection of reaction materials . The choice of catalysts with different concentrations may lead to different reaction pathways and product selectivities, thereby affecting the synthesis gas yield.

The yields of CO and C_2H_6 may be affected by factors such as catalyst selection, reaction temperature, and selection of reaction materials. The production of CO is usually associated with the partial oxidation of hydrocarbons, while the production of C2H6 may involve the cracking and recombination of hydrocarbons . Therefore, different catalysts and reaction conditions may lead to differences in CO and C_2H_6

yields.

5.4 Model establishment and solution of problem 4

5.4.1 structural equation modeling

(1) Establishment of latent variables and observed variables

The measurement model is composed of latent variables and observed variables. The latent variables are the yield of pyrolysis products of CE and LG under the catalytic action of different desulfurization ash ratios and the yield of each component gas. The observed variables are tar yield, water yield, Carbon yield and syngas yield as well as the yields of H_2, CO, CO_2, CH_4 and C_2H_6.

(2) Construction of structural model

The regression coefficient table of the model is shown in Table 6. If the P value is less than 0.05 and significant at the level, the null hypothesis is accepted. There is a direct impact between the latent variables and the manifest variables, that is, the path exists.

Table 6 Model regression coefficient table

Latent Variable	→	Explicit Variable	P	Path Presence	Influence Coefficient
CEPPY	→	CEPGY	0.024**	Yes	0.864
CEPPY	→	LGPPY	0.015***	Yes	-0.694
CEPPY	→	LGPGY	0.009***	Yes	0.763
CEPGY	→	CEPPY	0.308	No	
CEPGY	→	LGPPY	0.004***	Yes	-0.905
CEPGY	→	LGPGY	0.708	No	
LGPPY	→	CEPPY	0.018**	Yes	-0.821
LGPPY	→	CEPGY	0.002***	Yes	-0.659
LGPPY	→	LGPGY	0.009***	Yes	-0.987
LGPGY	→	CEPPY	0.531	No	
LGPGY	→	CEPGY	0.006***	Yes	0.735
LGPGY	→	LGPPY	0.288	No	

Note: ***, **, and * represent the significance levels of 1%, 5%, and 10% respectively ; CEPPY represents the yield of pyrolysis products of DFA/CE, and CEPGY represents the yield of each component gas of DFA/CE. ; LGPPY represents the yield of the pyrolysis product of DFA/LG, and LGPGY represents the yield of each component gas of DFA/LG.

CEPPY has a unidirectional influence on both CEPGY and LGPGY, LGPGY has a unidirectional influence on CEPGY, and LGPPY also has a unidirectional influence on LGPGY; there are bidirectional influences between CEPPY and LGPPY and between CEPGY and LGPPY. Under the catalysis of different desulfurization ash ratios, the yield of LG pyrolysis products negatively affects the yield of gases obtained by LG pyrolysis; the yield of LG pyrolysis products is related to the yield of each component gas obtained by CE pyrolysis and The yield of CE pyrolysis products has a mutual inhibitory effect; at the same time, the yield of CE pyrolysis products has a promoting effect on the gas yield of each component obtained by CE pyrolysis and the gas yield of each component obtained by LG pyrolysis; LG pyrolysis The obtained gas production of each component also has a positive promoting effect on the gas production of each component obtained by CE pyrolysis.

(3) Result analysis

The SEM structural picture is shown in Figure 13 . It can be seen from Figure 13 that first, the yield of LG pyrolysis products negatively affects the yield of gas obtained by LG pyrolysis. In LG pyrolysis, the catalyst DFA plays a catalytic role to promote the cracking and recombination of long-chain hydrocarbon molecules. Higher product yields indicate more carbon converted to tar, char, and other organic compounds. These macromolecular products are relatively stable and are not easily converted into gaseous products. Therefore, high product yields reduce the number of available carbon atoms and thus reduce gas production.

Secondly, there is a mutual inhibitory effect between the yield of LG pyrolysis products, the yield of each component gas obtained from CE pyrolysis, and the yield of CE pyrolysis products. Under different catalysts, LG pyrolysis and CE pyrolysis often have different reaction mechanisms and active sites. The higher LG pyrolysis product yield may consume part of the reaction raw materials and energy, thereby limiting the CE pyrolysis reaction path and product generation, resulting in the suppression of CE pyrolysis products and gas production.

has a promoting effect on the gas production of each component obtained by CE pyrolysis and the gas production of each component obtained by LG pyrolysis . This may be due to the higher activity and selectivity of CE pyrolysis reaction under the action of catalyst. Catalysts can provide active sites that are more suitable for reactions and are more conducive to rearrangement products, which can dissociate and desorb from the catalyst surface and release them into gaseous products.

In addition, the gas production of each component obtained by LG pyrolysis may also have a positive promoting effect on the gas production of each component obtained by CE pyrolysis. This may be because the gases generated in both processes may involve common reaction intermediates or reactants. Certain reactants or intermediates may be generated during the LG pyrolysis process and further participate in the CE pyrolysis reaction, increasing the generation of gas components from CE pyrolysis.

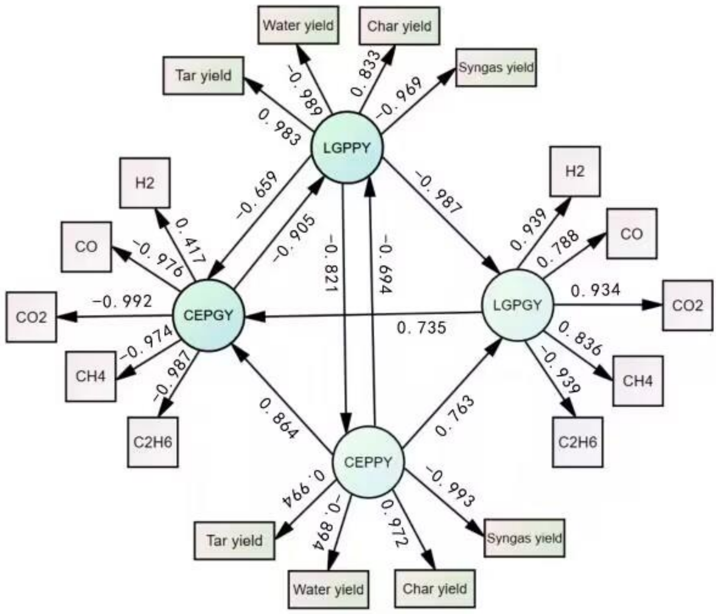


Figure 13 SEM structure diagram

5.4.2 Three-dimensional fitting reaction mechanism analysis

Based on the quadratic fitting in question 1, the cubic term is introduced to conduct the calculation of the yield of pyrolysis products of CE and LG under the catalysis of different proportions of catalysts in Annex 1 and Annex 2, as well as the yield of each component gas obtained from pyrolysis. Three fittings were performed to deeply explore the changing characteristics between the catalyst ratio, the yield of pyrolysis products and the yield of pyrolysis gas of each component, and then analyze the desulfurization ash catalytic reaction mechanism of CE and LG. The results are shown in Figure 14. Fitting The equation and goodness of fit are shown in Table 7.

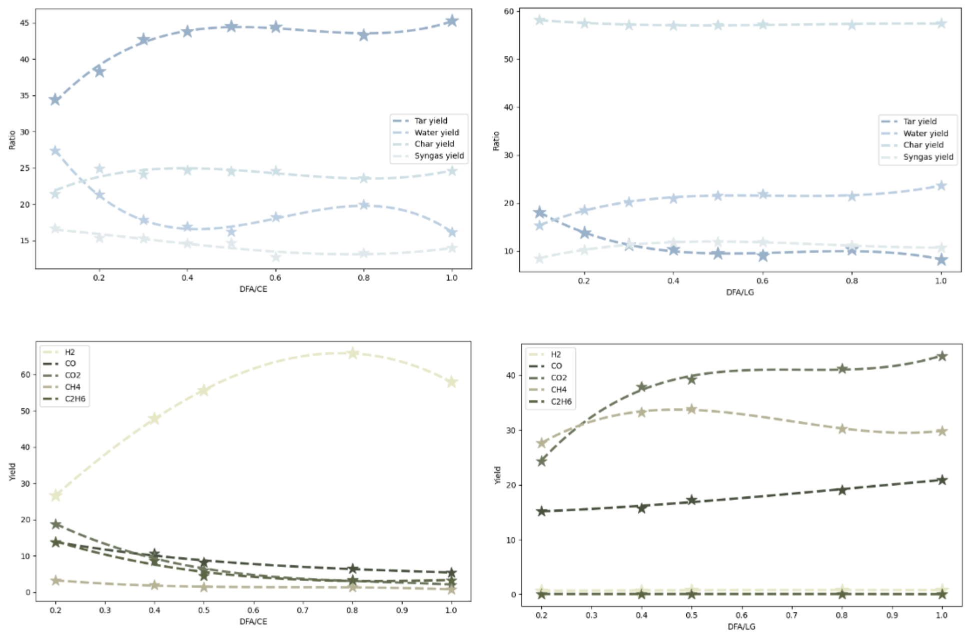


Figure 14 Cubic fitting curve

In the DFA/CE group, as the proportion of desulfurization ash increases, tar production increases and remains at a high level, water production and carbon yield are low, and syngas yield is the lowest, among which carbon yield and syngas yield are the lowest. The rate is relatively stable, which may be due to the increase in the proportion of desulfurization ash leading to the reduction of active sites on the catalyst surface, thereby affecting the cracking and rearrangement reactions of long-chain hydrocarbons. This results in more carbon atoms being converted into tar rather than gaseous products, thus leading to an increase in tar production and a decrease in water production ;

The H_2 production is high and gradually increases as the proportion of desulfurization ash increases, with the peak appearing at DFA/CE of 0.8 . The production of CO, CO_2, CH_4 and C_2H_4 is low and increases as the proportion of desulfurization ash increases. The downward trend may be because when the proportion of desulfurization ash is moderate, the reaction of rearrangement of cracked products to form H2 is relatively enhanced, while the formation of other compounds is suppressed ;

As the proportion of desulfurization ash increases in DFA/LG, the tar yield, water yield, carbon yield and syngas yield do not change much. The carbon yield is stable at a high level, the water yield is low, and the tar yield is low . and syngas yield is the lowest, which may be due to the fact that under low-temperature pyrolysis gas conditions, it is not sensitive to changes in the proportion of desulfurization ash, resulting in little change in yield ;

CO and CO_2 increases with the increase in the proportion of desulfurization ash. The output of H_2, CH_4 and C_2H_6 is relatively stable. Among them, the output of CO_2 and CH_4 is higher, the output of CO is lower, and the output of H_2 and C_2 The low production of H6 may be due to the insignificant effect of the catalyst on the rearrangement of long-chain hydrocarbons, resulting in the production of a higher proportion of CO2 and CH4 , while the production of CO is relatively small .

Table 7 Cubic fitting equation

Group	Index	Equation	R2
DFA/CE	Tar yield	69.88x^3 + -138.75x^2 + 87.41x + 26.63	0.988
	Water yield	-118.5x^3 + 217.49x^2 + -120.31x + 37.52	0.993
	Char yield	38.52x^3 + -69.68x^2 + 36.95x + 18.86	0.778
	Syngas yield	10.1x^3 + -9.06x^2 + -4.05x + 16.99	0.879
	H2	-50.26x^3 + -30.2x^2 + 137.67x + 0.69	1.000
	CO	-10.46x^3 + 30.57x^2 + -34.27x + 19.51	0.983
	CO2	-44.47x^3 + 115.8x^2 + -104.73x + 35.46	0.996
	CH4	-15.73x^3 + 32.23x^2 + -22.31x + 6.52	0.986
	C2H6	-17.52x^3 + 59.42x^2 + -62.92x + 24.32	0.969
DFA/LG	Tar yield	-58.52x^3 + 114.8x^2 + -72.18x + 24.18	0.992
	Water yield	46.14x^3 + -86.35x^2 + 52.99x + 10.86	0.995
	Char yield	-7.6x^3 + 15.95x^2 + -9.91x + 58.94	0.912
	Syngas yield	19.98x^3 + -44.39x^2 + 29.1x + 6.02	0.996
	H2	-0.42x^3 + 0.36x^2 + 0.23x + 0.61	0.965
	CO	-3.44x^3 + 8.89x^2 + 0.76x + 14.68	0.982
	CO2	122.25x^3 + -262.88x^2 + 187.82x + -3.58	0.996
	CH4	87.63x^3 + -183.71x^2 + 114.56x + 11.38	0.998
	C2H6	-0.32x^3 + 0.49x^2 + -0.23x + 0.11	0.580

5.5 **Model establishment and solution of problem five**

5.5.1 Establishment of prediction model for pyrolysis product yield or quantity

Different proportions of catalysts are used as dependent variables, and the yield of pyrolysis products and the yield of each component of the gas obtained from pyrolysis are used as target variables. Machine learning is used to establish a prediction model to complete the prediction of the yield or quantity of pyrolysis products.

The proximity algorithm (KNN) is one of the most basic classification methods. It is used in many fields. It requires a small number of samples and only relies on nearby samples to make decisions. There is no need to distinguish the class domain when determining the category.

Random forest is an ensemble algorithm composed of multiple decision trees. Each node in the tree represents a test, and each value represents the test result.

GBRT uses gradient boosting techniques to train models. Gradient boosting is a residual-based optimization method that gradually improves the accuracy of the model by iteratively fitting the residuals of the previous round of prediction results. In each iteration, GBRT calculates the gradient of the residual based on the loss function (such as mean square error), and uses the gradient as the target to train a new regression tree.

The X GBoost extreme gradient boosting model is optimized based on the gradient boosting tree (GBDT). Its basic idea is the same as GBDT, and strives to maximize speed and efficiency. In order to avoid overfitting, a regularization term needs to be added to the objective function to Reduce model complexity

5.5.2 Model optimization

(1) Initialize the training model and train an initial machine learning prediction model using default parameters .

(2) Perform cross-validation, divide the training data into a training set and a validation set, and repeat this process multiple times to obtain the performance of the model on different data subsets.

(3) Adjust the model parameters, perform hyperparameter tuning, and use grid search for the best learning rate, number and maximum depth of trees, regularization parameters, number of iterations, and feature subsampling ratio combinations to achieve the best classification effect of the model. excellent.

The optimized parameters of the model are shown in Table 8 :

Table 8 Model parameter table

	KNN	GBRT	RF	XGBoost
feature1	K=5	Loss= " ls "	N_estimators=80	Learning_rate=0.10
feature2	Weights= " uniform "	Alpha=0.90	Max_depth=2	Max_depth=1
feature3		max_depth=1	Max_features=1	N_estimators=200
feature4		m in_sample_split=0.10	Min_samples_leaf=2	Boostor= " gbtree "
feature5		Min_sample_leaf=2	Min_samples_split=2	

5.5.3 Model integration

The Taylor diagram is often used to evaluate the accuracy of the model. It is an effective method widely used in model evaluation and verification. It expresses the difference between the model prediction results and the observed values as the length and direction of the radial to identify the simulation results. Correlation, standard deviation and root mean square deviation. The correlation coefficient is reflected in the Taylor diagram through the length of the trajectory. A longer trajectory indicates a stronger correlation between the simulation results and the observed values. The standard deviation is represented in a Taylor plot by the distance of the trajectory from the reference point, with smaller distances meaning smaller differences between the variables of the simulated results and the observed values. The root mean square deviation is represented in the Taylor diagram by the radial distance from the trajectory to the reference point, with shorter radial distances indicating smaller overall differences between the simulation results and the observed values. Taylor diagrams can visually compare the performance of different models or models under different parameter settings, and comprehensively evaluate the accuracy and reliability of the model . The results are shown in Figure 15 :

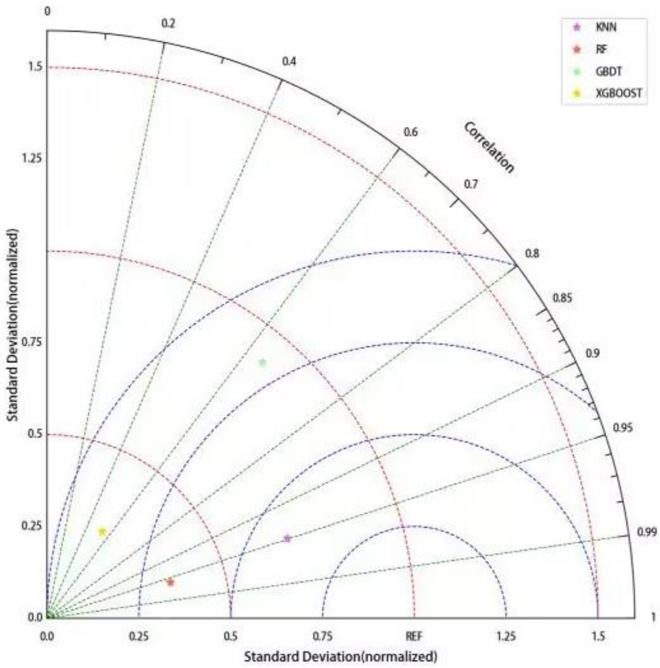


Figure 15 Taylor diagram

Ensemble modeling is a technique that builds more powerful predictive models by combining multiple independent base learners. It utilizes the principle of collective intelligence to combine multiple weak learners (also called base learners) to achieve higher accuracy and better generalization capabilities. Voting is a common combination strategy in ensemble models to combine the predictions of multiple base learners. This method is suitable for classification problems, where each base learner corresponds to a category, and the final prediction result is determined based on the voting results of each base learner.

From Figure 16 , the top two machine learning models in model accuracy, namely X GBoost and RF, are selected and integrated with the cubic fitting model in the fourth question to build a more reliable and universal learner. The result is shown in Figure n :

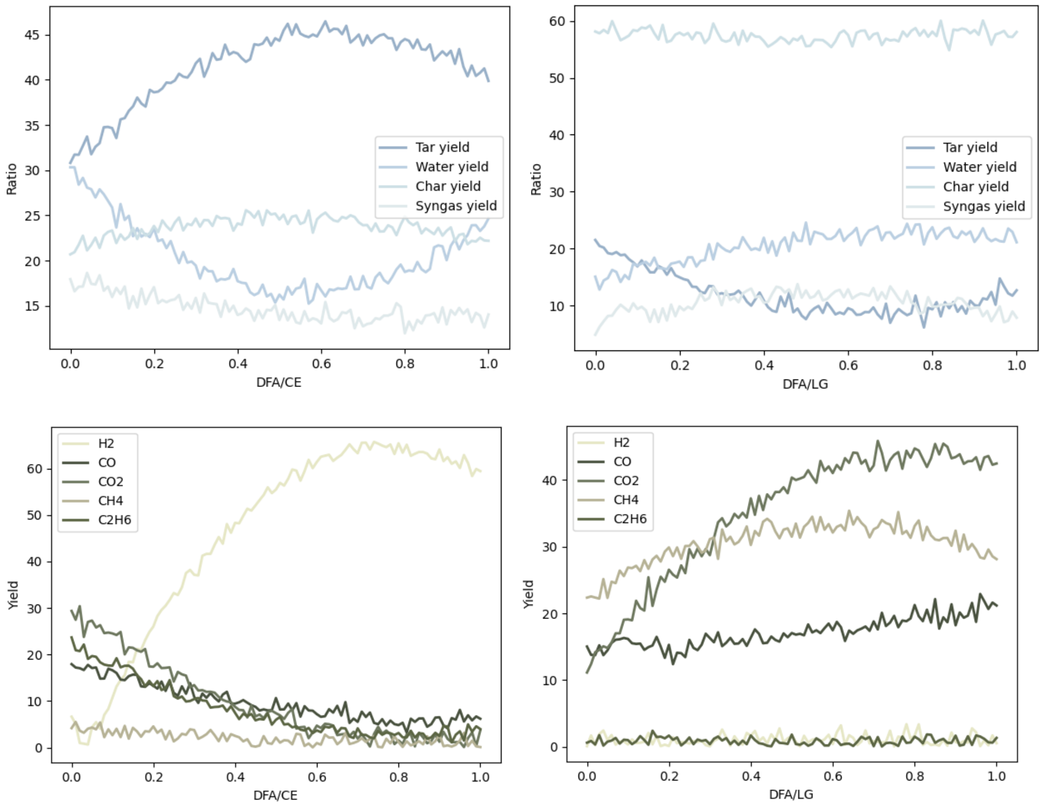


Figure 16 Integrated model fitting diagram

In DFA/CE, as the proportion of desulfurization ash increases, the tar yield is larger and first increases and then decreases, the aquatic product first decreases and then increases, the carbon yield and synthesis gas yield are relatively stable; H 2 production gradually increases , The production of CH_4 is very low and relatively stable, and the production of CO, CO_2 and C_2H_6 is low and gradually decreases.

In DFA/LG, as the proportion of desulfurization ash increases, the carbon yield remains basically stable and has a large value, while the tar production, water production and syngas yield do not change much and have low values ; CO , CO_2 The production of CH_4 and H 4 gradually increased, and the production of H_2 and C_2H_6 was small and basically remained stable .

6 Advantages and disadvantages

6.1 advantage

(1) When constructing the catalytic reaction mechanism of desulfurization ash, Pearson correlation coefficient and SEM structural equation were used to conduct a second round of discussion on the relationship, that is, the linear relationship was discussed, and the causal relationship was also explored;

(2) When predicting the yield of pyrolysis products, an integrated model of the reaction kinetics of the fourth question and the machine learning algorithm through parameter optimization and screening was constructed, taking into account both mathematical model methods and artificial intelligence learning methods;

(3) When exploring the impact of desulfurization ash ratio on pyrolysis behavior, the entropy weight method with objective weighting is used to fully tap the inherent information of the data.

6.2 shortcoming

(1) Due to the limited known data in this question, there is less data used to construct the reaction kinetic model and machine learning model, and the simulation effect of the model may deviate from the real situation;

(2) Due to the lack of data such as temperature, the construction of the reaction kinetic model lacks the consideration of other objective influencing factors.

7 references

[1] FRIEDRICH P.The three sigma rule[J].The American Statistician,1994,48(2):88-91.

[2] Zhang Ailin, Zhang Xiuying, Xu Youjun, etc. Research on Fuzzy Comprehensive Evaluation of Urban Comprehensive Pipeline Construction Progress Risk Based on Improved AHP-Entropy Method [J]. Construction Technology, 2017, 48(09):

[3] Dong CQ, Zhang ZF, Lu Q, et al. Characteristi and mechanism study of analytical fast pyrolysis of poplar wood [J]. Energy Conversion & Management, 2012, 57: 49-59.

[4] Wang Aijun, Wang Yafei, Li Zheming, etc. Low-voltage distribution network topology verification based on AHP-GRA analysis and KNN algorithm [J]. Science and Technology Innovation and Application , 2023,13(18):31-36.

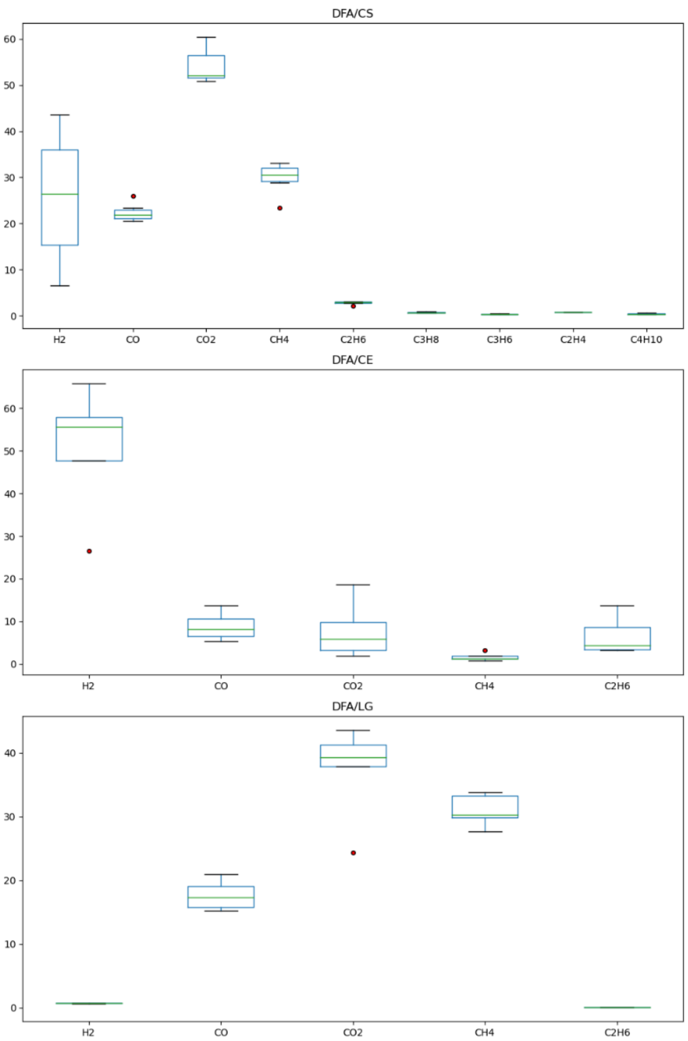
appendix

(1) Box plot of distribution characteristics of the data in Appendix 2

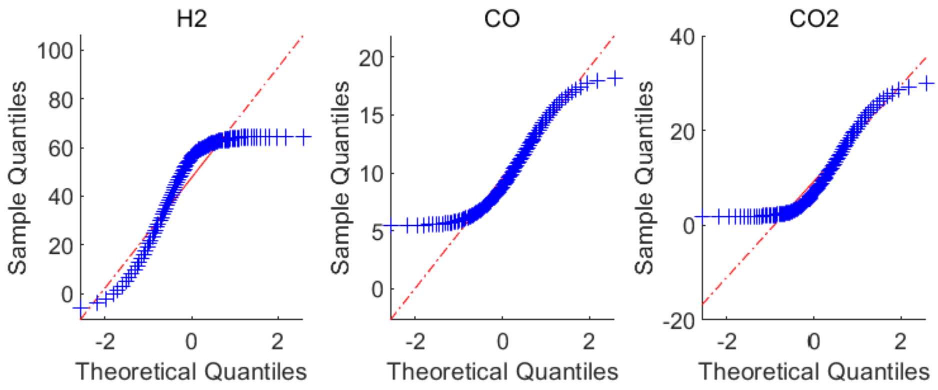
(2) QQ chart

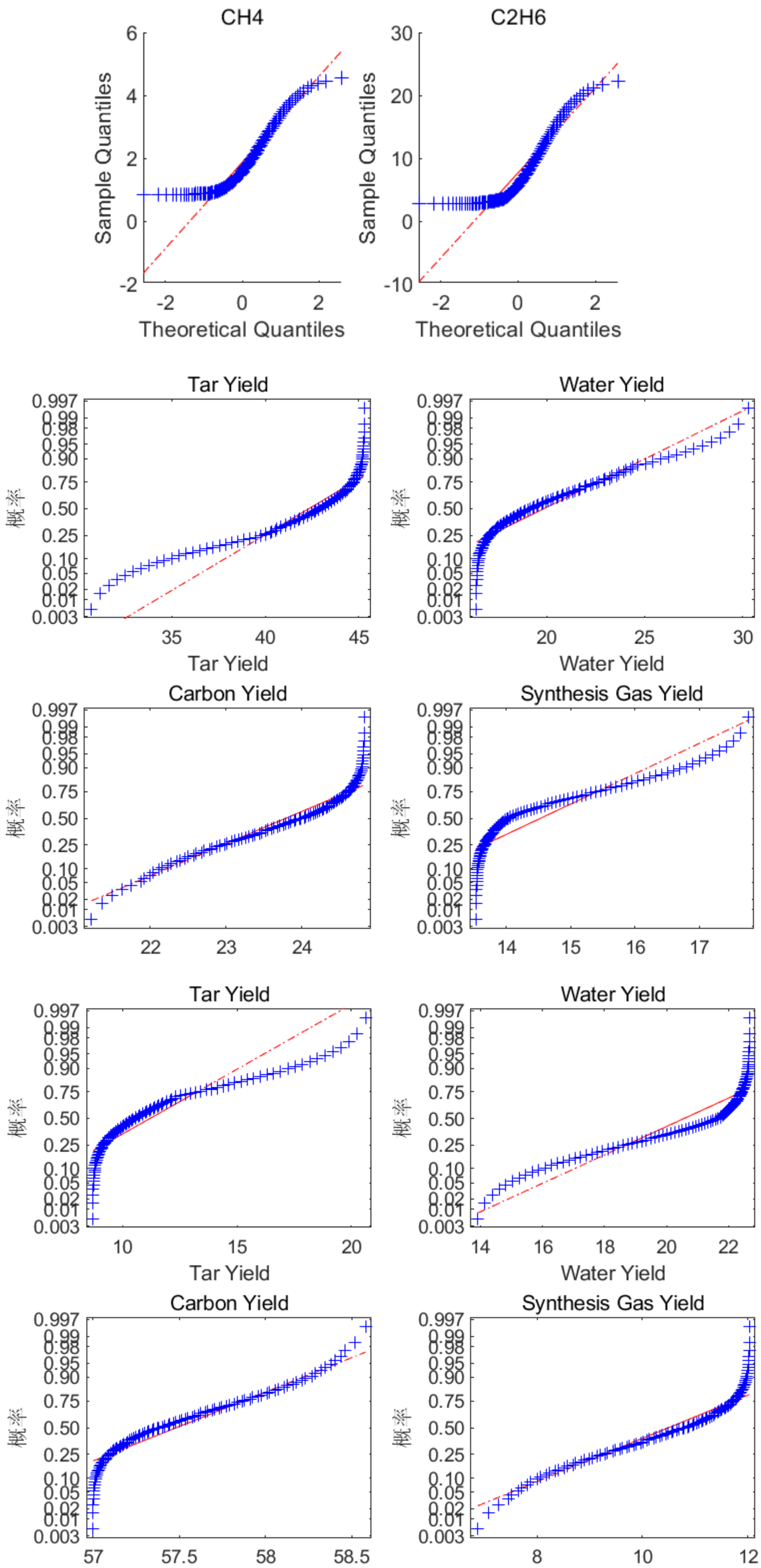
(3) Code

(1) Box plot of distribution characteristics of the data in Appendix 2


DFA/CS
H2
CO
CO2
CH4
C2H6
C3H8
C3H6
C2H4
C4H10
DFA/CE
H2
CO
CO2
CH4
C2H6
DFA/LG
H2
CO
CO2
CH4
C2H6

(2) QQ chart


H2
CO
CO2
Sample Quantiles
Theoretical Quantiles


CH4
Sample Quantiles
Theoretical Quantiles
C2H6
Sample Quantiles
Theoretical Quantiles
Tar Yield
概率
Tar Yield
Water Yield
概率
Water Yield
Carbon Yield
概率
Synthesis Gas Yield
概率
Tar Yield
概率
Tar Yield
Water Yield
概率
Water Yield
Carbon Yield
概率
Synthesis Gas Yield
概率

(3) Code

```
#题一
import pandas as pd
import matplotlib.pyplot as plt
data1 = pd.read_excel('F:/数维杯/DACS.xlsx', sheet_name='Sheet1')
data2 = pd.read_excel('F:/数维杯/DFACE.xlsx', sheet_name='Sheet1')
data3 = pd.read_excel('F:/数维杯/DFALG.xlsx', sheet_name='Sheet1')
```

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import rcParams

LABELS = {
    'DFA/CS': 'DFA/CS',
    'DFA/CE': 'DFA/CE',
    'DFA/LG': 'DFA/LG',
    '焦油产量': 'Tar Yield',
    '水产量': 'Water Yield',
    '炭收率': 'Char Yield',
    '合成气收率': 'Syngas Yield',
}

rcParams['font.family'] = 'Arial'

data1 = pd.read_excel("F:/数维杯/题一/DACS.xlsx", sheet_name='Sheet1')
data2 = pd.read_excel("F:/数维杯/题一/DFACE.xlsx", sheet_name='Sheet1')
data3 = pd.read_excel("F:/数维杯/题一/DFALG.xlsx", sheet_name='Sheet1')

fig, axs = plt.subplots(4, 3, figsize=(15, 12))

datasets = [data1, data2, data3]

for i, data in enumerate(datasets):
    x = data['DFA/CS'] if i == 0 else data['DFA/CE'] if i == 1 else data['DFA/LG']
    y1 = data['焦油产量']
    y2 = data['水产量']
    y3 = data['炭收率']
    y4 = data['合成气收率']

    for j, ax in enumerate(axs[:, i]):
        ax.scatter(x, [y1, y2, y3, y4][j], color=['Maroon', 'Navy', 'Silver', 'Olive'][j])
```

```
    for j, ax in enumerate(axs[:, i]):
        ax.scatter(x, [y1, y2, y3, y4][j], color=['Maroon', 'Navy', 'Silver', 'Olive'][j])

        # 使用拟合方程生成更多的 x 值
        x_smooth = np.linspace(x.min(), x.max(), 100)

        coef = np.polyfit(x, [y1, y2, y3, y4][j], 2)
        poly1d_fn = np.poly1d(coef)

        ax.plot(x_smooth, poly1d_fn(x_smooth), color="red", linewidth=2.5)

        ax.set_xticks(x)

        equation_text = f'y = {coef[0]:.2f}x^2 + {coef[1]:.2f}x + {coef[2]:.2f}'
        ax.text(0.05, 0.91, equation_text, ha='left', va='top', transform=ax.transAxes, fontsize=12)

        residuals = np.array([y - poly1d_fn(x_val) for x_val, y in zip(x, [y1, y2, y3, y4][j])])
        ss_res = np.sum(residuals**2)
        ss_tot = np.sum((np.mean([y1, y2, y3, y4][j]) - [y1, y2, y3, y4][j])**2)
        r_squared = 1 - (ss_res / ss_tot)
        r_squared_text = f'R_squared: {r_squared:.2f}'
        ax.text(0.335, 0.83, r_squared_text, ha='right', va='top', transform=ax.transAxes, fontsize=12, color='black')

        # 只给第一列子图添加y轴标题
        if i == 0:
            axs[j, i].set_ylabel([LABELS['焦油产量'], LABELS['水产量'], LABELS['炭收率'], LABELS['合成气收率']][j], fontsize=12)

    # 设置x轴标题
    axs[3, i].set_xlabel(LABELS['DFA/CS'], fontsize=12) if i == 0 else axs[3, i].set_xlabel(LABELS['DFA/CE'], fontsize=12) if i == 1 else axs

axs[0, 0].set_xticklabels(["0", '0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8'])
axs[1, 0].set_xticklabels(["0", '0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8'])
axs[2, 0].set_xticklabels(["0", '0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8'])
axs[3, 0].set_xticklabels(["0", '0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8'])

axs[0, 1].set_xticklabels(['0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8'])
axs[1, 1].set_xticklabels(['0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8'])
axs[2, 1].set_xticklabels(['0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8'])
axs[3, 1].set_xticklabels(['0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8'])
```

```
axs[0, 2].set_xticklabels(['0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8'])
axs[1, 2].set_xticklabels(['0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8'])
axs[2, 2].set_xticklabels(['0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8'])
axs[3, 2].set_xticklabels(['0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8'])

for i in range(4):
    for j in range(3):
        axs[i, j].spines['top'].set_linewidth(2.5)
        axs[i, j].spines['right'].set_linewidth(2.5)
        axs[i, j].spines['bottom'].set_linewidth(2.5)
        axs[i, j].spines['left'].set_linewidth(2.5)

for i in range(4):
    for j in range(3):
        axs[i, j].tick_params(axis='x', labelsize=14)
        axs[i, j].tick_params(axis='y', labelsize=14)

plt.tight_layout()
plt.savefig("F:/数维杯/1.png", dpi=400)
plt.show()
```

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

LABELS = {
    'DFA/CS': 'DFA/CS',
    'DFA/CE': 'DFA/CE',
    'DFA/LG': 'DFA/LG',
    '焦油产量': 'TY',
    '水产量': 'WY',
    '炭收率': 'CY',
    '合成气收率': 'SGY',
}

# 设置全局默认字体为 Arial
plt.rcParams['font.family'] = 'Arial'
plt.rcParams['font.size'] = 12

# 读取数据
data1 = pd.read_excel('F:/数维杯/DACS.xlsx', sheet_name='Sheet1')
data2 = pd.read_excel('F:/数维杯/DFACE.xlsx', sheet_name='Sheet1')
data3 = pd.read_excel('F:/数维杯/DFALG.xlsx', sheet_name='Sheet1')

# 计算皮尔逊相关系数
corr1 = data1.corr()
corr2 = data2.corr()
corr3 = data3.corr()
p_values1 = np.zeros_like(corr1)
p_values2 = np.zeros_like(corr2)
p_values3 = np.zeros_like(corr3)

# Create custom color map from light blue to pink
colors = ["white", "lightblue"]
cmap = plt.cm.colors.LinearSegmentedColormap.from_list("CustomMap", colors)
fig, axes = plt.subplots(1, 3, figsize=(15, 5))

# 绘制热力图，并为每个子图添加内部框线
sns.heatmap(corr1, annot=True, cmap=cmap, cbar=False, ax=axes[0], xticklabels=[LABELS[label] for label in corr1.columns], yticklabe
axes[0].set_title('DFA/CS')

sns.heatmap(corr2, annot=True, cmap=cmap, cbar=False, ax=axes[1], xticklabels=[LABELS[label] for label in corr2.columns], yticklabe
axes[1].set_title('DFA/CE')

sns.heatmap(corr3, annot=True, cmap=cmap, cbar=False, ax=axes[2], xticklabels=[LABELS[label] for label in corr3.columns], yticklabe
axes[2].set_title('DFA/LG')
```

```
# 添加外部框线
for ax in axes:
    for _, spine in ax.spines.items():
        spine.set_visible(True)

# 在相关位置上添加星号标记
for i, ax in enumerate(axes):
    for i in range(len(corr1.columns)):
        for j in range(len(corr1.columns)):
            if i != j and p_values1[i, j] < 0.01:
                ax.text(j + 0.5, i + 0.30, "**", fontsize=12, ha='center', va='center')
            elif i != j and p_values1[i, j] < 0.05:
                ax.text(j + 0.5, i + 0.3, "*", fontsize=12, ha='center', va='center')

# 手动创建颜色条并放置在底部中央
cax = fig.add_axes([0.125, -0.2, 0.775, 0.15])  # 指定颜色条的位置和尺寸
cbar = fig.colorbar(axes[0].collections[0], cax=cax, orientation='horizontal')  # 创建颜色条并放置在指定的位置
# 保存图片
plt.savefig('F:/数维杯/图片/heatmap.png', dpi=300, bbox_inches='tight')
plt.show()
```

```
6]: import pandas as pd
import matplotlib.pyplot as plt
data1 = pd.read_excel('F:/数维杯/DACS.xlsx', sheet_name='Sheet1')
data2 = pd.read_excel('F:/数维杯/DFACE.xlsx', sheet_name='Sheet1')
data3 = pd.read_excel('F:/数维杯/DFALG.xlsx', sheet_name='Sheet1')
```

```
1]: import pandas as pd
import matplotlib.pyplot as plt

# 读取数据
data1 = pd.read_excel('F:/数维杯/DACS.xlsx', sheet_name='Sheet1')
data2 = pd.read_excel('F:/数维杯/DFACE.xlsx', sheet_name='Sheet1')
data3 = pd.read_excel('F:/数维杯/DFALG.xlsx', sheet_name='Sheet1')

# 创建包含三个子图的图形
fig, axes = plt.subplots(nrows=1, ncols=3, figsize=(15, 5))

# 定义离群点的样式
flierprops = dict(marker='o', markerfacecolor='red', markersize=4, linestyle='none')

# 在每个子图中绘制箱线图，并设置英文标签
bp1 = data1.boxplot(column=['焦油产量', '水产量', '炭收率', '合成气收率'], ax=axes[0], grid=False, flierprops=flierprops)
axes[0].set_xticklabels(['TY', 'WY', 'CY', 'SGY'])
axes[0].set_title('DFA/CS')

bp2 = data2.boxplot(column=['焦油产量', '水产量', '炭收率', '合成气收率'], ax=axes[1], grid=False, flierprops=flierprops)
axes[1].set_xticklabels(['TY', 'WY', 'CY', 'SGY'])
axes[1].set_title('DFA/CE')

bp3 = data3.boxplot(column=['焦油产量', '水产量', '炭收率', '合成气收率'], ax=axes[2], grid=False, flierprops=flierprops)
axes[2].set_xticklabels(['TY', 'WY', 'CY', 'SGY'])
axes[2].set_title('DFA/LG')

# 显示图形
plt.show()
```

```
: import numpy as np

# 读取数据
data1 = pd.read_excel('F:/数维杯/DACS.xlsx', sheet_name='Sheet1')
data2 = pd.read_excel('F:/数维杯/DFACE.xlsx', sheet_name='Sheet1')
data3 = pd.read_excel('F:/数维杯/DFALG.xlsx', sheet_name='Sheet1')

# 定义检测异常值的函数
def detect_outliers(data, threshold=3):
    mean = np.mean(data)
    std = np.std(data)
    upper_threshold = mean + threshold * std
    lower_threshold = mean - threshold * std
    outliers = data[(data > upper_threshold) | (data < lower_threshold)]
    return outliers

# 检测异常值
outliers1 = detect_outliers(data1)
outliers2 = detect_outliers(data2)
outliers3 = detect_outliers(data3)

# 输出异常值
print("DFA/CS的异常值：")
print(outliers1)
print("DFA/CE的异常值：")
print(outliers2)
print("DFA/LG的异常值：")
print(outliers3)
```

```
: import pandas as pd
import matplotlib.pyplot as plt
data1 = pd.read_excel('F:/数维杯/DACS.xlsx', sheet_name='Sheet1')
data2 = pd.read_excel('F:/数维杯/DFACE.xlsx', sheet_name='Sheet1')
data3 = pd.read_excel('F:/数维杯/DFALG.xlsx', sheet_name='Sheet1')
```

```
: import pandas as pd
from scipy.stats import entropy

# 读取数据
data1 = pd.read_excel('F:/数维杯/DACS.xlsx', sheet_name='Sheet1')

# 提取需要计算权重的列
selected_columns = ['焦油产量', '水产量', '炭收率', '合成气收率']
data_selected = data1[selected_columns]

# 计算各列数据的熵值
entropies = [entropy(column) for column in data_selected.T]

# 计算权重
weights = [entropy / sum(entropies) for entropy in entropies]

# 打印各列数据的权重
for i, column in enumerate(selected_columns):
    print(f"权重 {column}: {weights[i]}")
```

```
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
from scipy.stats import entropy

# 读取数据
data1 = pd.read_excel('F:/数维杯/DFALG.xlsx', sheet_name='Sheet1')
# 计算平均值
avg_tar_yield = data1['焦油产量'].mean()
avg_water_yield = data1['水产量'].mean()
avg_char_yield = data1['炭收率'].mean()
avg_syngas_yield = data1['合成气收率'].mean()

# 计算标准差
std_tar_yield = data1['焦油产量'].std()
std_water_yield = data1['水产量'].std()
std_char_yield = data1['炭收率'].std()
std_syngas_yield = data1['合成气收率'].std()

# 计算最大值
max_tar_yield = data1['焦油产量'].max()
max_water_yield = data1['水产量'].max()
max_char_yield = data1['炭收率'].max()
max_syngas_yield = data1['合成气收率'].max()

# 计算最小值
min_tar_yield = data1['焦油产量'].min()
min_water_yield = data1['水产量'].min()
min_char_yield = data1['炭收率'].min()
min_syngas_yield = data1['合成气收率'].min()

# 计算峰度
kurt_tar_yield = data1['焦油产量'].kurtosis()
kurt_water_yield = data1['水产量'].kurtosis()
kurt_char_yield = data1['炭收率'].kurtosis()
kurt_syngas_yield = data1['合成气收率'].kurtosis()

# 计算偏度
skew_tar_yield = data1['焦油产量'].skew()
skew_water_yield = data1['水产量'].skew()
skew_char_yield = data1['炭收率'].skew()
skew_syngas_yield = data1['合成气收率'].skew()
```

```
# 计算变异系数
cv_tar_yield = std_tar_yield / avg_tar_yield * 100
cv_water_yield = std_water_yield / avg_water_yield * 100
cv_char_yield = std_char_yield / avg_char_yield * 100
cv_syngas_yield = std_syngas_yield / avg_syngas_yield * 100
# 创建包含统计结果的DataFrame
stats_data = pd.DataFrame({
    '指标': ['焦油产量', '水产量', '炭收率', '合成气收率'],
    '平均值': [avg_tar_yield, avg_water_yield, avg_char_yield, avg_syngas_yield],
    '标准差': [std_tar_yield, std_water_yield, std_char_yield, std_syngas_yield],
    '最大值': [max_tar_yield, max_water_yield, max_char_yield, max_syngas_yield],
    '最小值': [min_tar_yield, min_water_yield, min_char_yield, min_syngas_yield],
    '峰度': [kurt_tar_yield, kurt_water_yield, kurt_char_yield, kurt_syngas_yield],
    '偏度': [skew_tar_yield, skew_water_yield, skew_char_yield, skew_syngas_yield],
    '变异系数': [cv_tar_yield, cv_water_yield, cv_char_yield, cv_syngas_yield]
})

# 保存DataFrame到Excel文件
stats_data.to_excel('F:/数维杯/DFALG统计结果.xlsx', index=False)
```

```python
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from scipy.optimize import curve_fit

LABELS = {
    'H2': 'H2',
    'CO': 'CO',
    'CO2': 'CO2',
    'CH4': 'CH4',
    'C2H6': 'C2H6',
    'C3H8': 'C3H8',
    'C3H6': 'C3H6',
    'C2H4': 'C2H4',
    'C4H10': 'C4H10'
}

# 读取数据
data = pd.read_excel('F:/数维杯/DFACS-1.xlsx')

# 定义二次函数模型
def quadratic_function(x, a, b, c):
    return a * x**2 + b * x + c

# 提取DFA/CS作为x轴数据
x = data['DFA/CS']

# 提取其他列作为y轴数据
y_columns = ['H2', 'CO', 'CO2', 'CH4', 'C2H6', 'C3H8', 'C3H6', 'C2H4', 'C4H10']
# 设置子图布局
fig, axs = plt.subplots(3, 3, figsize=(15, 12))

# 生成更多的x值
x_smooth = np.linspace(x.min(), x.max(), 100)
```

```python
# 遍历每个y轴数据进行拟合和绘图
for i, y_col in enumerate(y_columns):
    row = i // 3
    col = i % 3
    y = data[y_col]
    popt, pcov = curve_fit(quadratic_function, x, y)
    residuals = y - quadratic_function(x, *popt)
    ss_res = np.sum(residuals**2)
    ss_tot = np.sum((y - np.mean(y))**2)
    r_squared = 1 - (ss_res / ss_tot)
    if col == 0:
        axs[row, col].scatter(x, y, color='Maroon')
    elif col == 1:
        axs[row, col].scatter(x, y, color='Navy')
    elif col == 2:
        axs[row, col].scatter(x, y, color='Olive')
    axs[row, col].plot(x_smooth, quadratic_function(x_smooth, *popt), 'r-')
    equation = f'y = {round(popt[0], 3)}x^2 + {round(popt[1], 3)}x + {round(popt[2], 3)}'
    r_squared_text = f'R squared = {round(r_squared, 2)}'
    axs[row, col].text(0.05, 0.95, equation, transform=axs[row, col].transAxes, fontsize=12, va='top', ha='left')
    axs[row, col].text(0.05, 0.85, r_squared_text, transform=axs[row, col].transAxes, fontsize=12, va='top', ha='left')
    axs[row, col].set_title(LABELS[y_col])  # 设置子图标题

plt.tight_layout()
plt.savefig("F:/数维杯/2-1.png", dpi=400)
plt.show()
```

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from scipy.optimize import curve_fit

# 数据文件路径
file_paths = ['F:/数维杯/DFACE-1.xlsx']
sheet_name = 'Sheet1'

# 定义二次函数模型
def quadratic_function(x, a, b, c):
    return a * x**2 + b * x + c

for file_path in file_paths:
    # 读取数据
    data = pd.read_excel(file_path, sheet_name=sheet_name)

    # 提取DFA/CS作为x轴数据
    x = data['DFA/CE']

    # 提取其他列作为y轴数据
    y_columns = ['H2', 'CO', 'CO2', 'CH4', 'C2H6']
    # 设置子图布局
    fig = plt.figure(figsize=(15, 12))

    # 生成更多的x值
    x_smooth = np.linspace(x.min(), x.max(), 100)

    # 遍历每个y轴数据进行拟合和绘图
    for i, y_col in enumerate(y_columns):
        ax = fig.add_subplot(3, 3, i+1)
        y = data[y_col]
        popt, pcov = curve_fit(quadratic_function, x, y)
        residuals = y - quadratic_function(x, *popt)
        ss_res = np.sum(residuals**2)
        ss_tot = np.sum((y - np.mean(y))**2)
        r_squared = 1 - (ss_res / ss_tot)
        if i % 3 == 0:
            ax.scatter(x, y, color='Maroon')
        elif i % 3 == 1:
            ax.scatter(x, y, color='Navy')
        elif i % 3 == 2:
            ax.scatter(x, y, color='Olive')
        ax.plot(x_smooth, quadratic_function(x_smooth, *popt), 'r-')
        equation = f'y = {round(popt[0], 3)}x^2 + {round(popt[1], 3)}x + {round(popt[2], 3)}'
        r_squared_text = f'R squared = {round(r_squared, 2)}'
        ax.text(0.05, 0.95, equation, transform=ax.transAxes, fontsize=12, va='top', ha='left')
        ax.text(0.05, 0.85, r_squared_text, transform=ax.transAxes, fontsize=12, va='top', ha='left')
        ax.set_title(y_col)  # 设置子图标题

    plt.subplots_adjust(hspace=0.5, wspace=0.3)
```

```
# 删除空白图
if len(y_columns) < 9:
    for i in range(len(y_columns), len(fig.axes)):
        fig.delaxes(fig.axes[i])

# 显示图像
plt.savefig("F:/数维杯/2-2.png", dpi=400)
plt.show()
```

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from scipy.optimize import curve_fit

# 数据文件路径
file_paths = ['F:/数维杯/DFALG-1.xlsx']
sheet_name = 'Sheet1'

# 定义二次函数模型
def quadratic_function(x, a, b, c):
    return a * x**2 + b * x + c

for file_path in file_paths:
    # 读取数据
    data = pd.read_excel(file_path, sheet_name=sheet_name)

    # 提取DFA/CS作为x轴数据
    x = data['DFA/LG']

    # 提取其他列作为y轴数据
    y_columns = ['H2', 'CO', 'CO2', 'CH4', 'C2H6']
    # 设置子图布局
    fig = plt.figure(figsize=(15, 12))

    # 生成更多的x值
    x_smooth = np.linspace(x.min(), x.max(), 100)

    # 遍历每个y轴数据进行拟合和绘图
    for i, y_col in enumerate(y_columns):
        ax = fig.add_subplot(3, 3, i+1)
        y = data[y_col]
        popt, pcov = curve_fit(quadratic_function, x, y)
        residuals = y - quadratic_function(x, *popt)
        ss_res = np.sum(residuals**2)
        ss_tot = np.sum((y - np.mean(y))**2)
        r_squared = 1 - (ss_res / ss_tot)
        if i % 3 == 0:
            ax.scatter(x, y, color='Maroon')
        elif i % 3 == 1:
            ax.scatter(x, y, color='Navy')
        elif i % 3 == 2:
            ax.scatter(x, y, color='Olive')
```

```
        equation = f'y = {round(popt[0], 3)}x^2 + {round(popt[1], 3)}x + {round(popt[2], 3)}'
        r_squared_text = f'R squared = {round(r_squared, 2)}'
        ax.text(0.05, 0.95, equation, transform=ax.transAxes, fontsize=12, va='top', ha='left')
        ax.text(0.05, 0.85, r_squared_text, transform=ax.transAxes, fontsize=12, va='top', ha='left')
        ax.set_title(y_col)  # 设置子图标题

    plt.subplots_adjust(hspace=0.5, wspace=0.3)

# 删除空白图
if len(y_columns) < 9:
    for i in range(len(y_columns), len(fig.axes)):
        fig.delaxes(fig.axes[i])

# 显示图像
plt.savefig("F:/数维杯/12-3.png", dpi=400)
plt.show()
```

```
import pandas as pd

# 读取数据
data1 = pd.read_excel('F:/数维杯/DFACS-1.xlsx', sheet_name='Sheet1')
data2 = pd.read_excel('F:/数维杯/DFACE-1.xlsx', sheet_name='Sheet1')
data3 = pd.read_excel('F:/数维杯/DFALG-1.xlsx', sheet_name='Sheet1')

# 数据1描述性统计
data1_description = data1.describe()
data1_kurtosis = data1.kurtosis()
data1_skewness = data1.skew()
data1_coefficient_of_variation = data1.std() / data1.mean()

# 数据2描述性统计
data2_description = data2.describe()
data2_kurtosis = data2.kurtosis()
data2_skewness = data2.skew()
data2_coefficient_of_variation = data2.std() / data2.mean()

# 数据3描述性统计
data3_description = data3.describe()
data3_kurtosis = data3.kurtosis()
data3_skewness = data3.skew()
data3_coefficient_of_variation = data3.std() / data3.mean()

# 创建 DataFrame
data1_description_df = pd.DataFrame(data1_description)
data1_kurtosis_df = pd.DataFrame(data1_kurtosis)
data1_skewness_df = pd.DataFrame(data1_skewness)
data1_coefficient_of_variation_df = pd.DataFrame(data1_coefficient_of_variation)

data2_description_df = pd.DataFrame(data2_description)
data2_kurtosis_df = pd.DataFrame(data2_kurtosis)
data2_skewness_df = pd.DataFrame(data2_skewness)
data2_coefficient_of_variation_df = pd.DataFrame(data2_coefficient_of_variation)

data3_description_df = pd.DataFrame(data3_description)
data3_kurtosis_df = pd.DataFrame(data3_kurtosis)
data3_skewness_df = pd.DataFrame(data3_skewness)
data3_coefficient_of_variation_df = pd.DataFrame(data3_coefficient_of_variation)

# 保存为Excel文件
with pd.ExcelWriter('F:/数维杯/描述性统计结果.xlsx') as writer:
    data1_description_df.to_excel(writer, sheet_name='Data1 描述统计')
    data1_kurtosis_df.to_excel(writer, sheet_name='Data1 峰度')
    data1_skewness_df.to_excel(writer, sheet_name='Data1 偏度')
    data1_coefficient_of_variation_df.to_excel(writer, sheet_name='Data1 变异系数')
```


```
    data1_coefficient_of_variation_df.to_excel(writer, sheet_name='Data1 变异系数')
```


```

    data2_description_df.to_excel(writer, sheet_name='Data2 描述统计')
    data2_kurtosis_df.to_excel(writer, sheet_name='Data2 峰度')
    data2_skewness_df.to_excel(writer, sheet_name='Data2 偏度')
    data2_coefficient_of_variation_df.to_excel(writer, sheet_name='Data2 变异系数')

    data3_description_df.to_excel(writer, sheet_name='Data3 描述统计')
    data3_kurtosis_df.to_excel(writer, sheet_name='Data3 峰度')
    data3_skewness_df.to_excel(writer, sheet_name='Data3 偏度')
    data3_coefficient_of_variation_df.to_excel(writer, sheet_name='Data3 变异系数')
```

```
import pandas as pd
import matplotlib.pyplot as plt

# 读取数据
data1 = pd.read_excel('F:/数维杯/DFACS-1.xlsx', sheet_name='Sheet1')
data2 = pd.read_excel('F:/数维杯/DFACE-1.xlsx', sheet_name='Sheet1')
data3 = pd.read_excel('F:/数维杯/DFALG-1.xlsx', sheet_name='Sheet1')

# 创建包含三个子图的图形
fig, axes = plt.subplots(nrows=3, figsize=(10, 15))

# 定义离群点的样式
flierprops = dict(marker='o', markerfacecolor='red', markersize=4, linestyle='none')

# 在每个子图中绘制箱线图，并设置英文标签
bp1 = data1.iloc[:, 1:].boxplot(ax=axes[0], grid=False, flierprops=flierprops)
axes[0].set_xticklabels(data1.columns[1:])
axes[0].set_title('DFA/CS')

bp2 = data2.iloc[:, 1:].boxplot(ax=axes[1], grid=False, flierprops=flierprops)
axes[1].set_xticklabels(data2.columns[1:])
axes[1].set_title('DFA/CE')

bp3 = data3.iloc[:, 1:].boxplot(ax=axes[2], grid=False, flierprops=flierprops)
axes[2].set_xticklabels(data3.columns[1:])
axes[2].set_title('DFA/LG')

# 调整子图之间的间距
plt.tight_layout()

# 保存图形为PNG文件
plt.savefig('F:/数维杯/boxplot.png')

# 显示图形
plt.show()
```

```python
# 第一个雷达图
# 变量名称
variables1 = ['Tar ', 'Water ', 'Char ', 'Syngas ']
# 变量值
values1 = [4.426, 2.64, 22.897, 1.634]
# 计算角度
angles1 = np.linspace(0, 2 * np.pi, len(variables1), endpoint=False).tolist()
angles1 += angles1[:1]
# 数据闭合
values1 += values1[:1]
# 设置整个图的尺寸为 10x5
plt.figure(figsize=(10, 5))
# 创建第一个子图
ax1 = plt.subplot(1, 2, 1, polar=True)
# 绘制第一个雷达图
ax1.plot(angles1, values1, color='pink', linewidth=1, linestyle='solid', label='Value')
ax1.fill(angles1, values1, color='lightgreen', alpha=0.3)
ax1.set_xticks(angles1[:-1])
ax1.set_xticklabels(variables1)
ax1.set_ylim(0, 30)  # 将范围设为0-30
ax1.set_yticks(np.arange(0, 30, 5))  # 设置刻度线的位置
ax1.set_yticklabels(['0', '5', '10','15','20','25'])  # 设置刻度线的标签
ax1.grid(True)

# 第二个雷达图
# 变量名称
variables2 = ['H2', 'CO', 'CO2', 'CH4', 'C2H6']
# 变量值
values2 = [3.358, 1.564, 2.07, 9.471, 1.457]
# 计算角度
angles2 = np.linspace(0, 2 * np.pi, len(variables2), endpoint=False).tolist()
angles2 += angles2[:1]
# 数据闭合
values2 += values2[:1]
# 创建第二个子图
ax2 = plt.subplot(1, 2, 2, polar=True)
# 绘制第二个雷达图
ax2.plot(angles2, values2, color='skyblue', linewidth=1, linestyle='solid', label='Value')
ax2.fill(angles2, values2, color='lightgreen', alpha=0.25)
ax2.set_xticks(angles2[:-1])
ax2.set_xticklabels(variables2)
ax2.set_ylim(0, 12)  # 将范围设为0-12
ax2.set_yticks(np.arange(0, 12, 2))  # 设置刻度线的位置
ax2.set_yticklabels(['0', '2','4','6','8','10'])  # 设置刻度线的标签
ax2.grid(True)
```

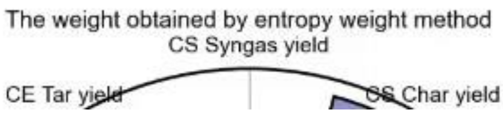
```python
# 第二个雷达图
# 变量名称
variables2 = ['H2', 'CO', 'CO2', 'CH4', 'C2H6']
# 变量值
values2 = [3.358, 1.564, 2.07, 9.471, 1.457]
# 计算角度
angles2 = np.linspace(0, 2 * np.pi, len(variables2), endpoint=False).tolist()
angles2 += angles2[:1]
# 数据闭合
values2 += values2[:1]
# 创建第二个子图
ax2 = plt.subplot(1, 2, 2, polar=True)
# 绘制第二个雷达图
ax2.plot(angles2, values2, color='skyblue', linewidth=1, linestyle='solid', label='Value')
ax2.fill(angles2, values2, color='lightgreen', alpha=0.25)
ax2.set_xticks(angles2[:-1])
ax2.set_xticklabels(variables2)
ax2.set_ylim(0, 12)  # 将范围设为0-12
ax2.set_yticks(np.arange(0, 12, 2))  # 设置刻度线的位置
ax2.set_yticklabels(['0', '2','4','6','8','10'])  # 设置刻度线的标签
ax2.grid(True)
# 添加标题

# 将最外面的圆线加粗为2
ax1.spines['polar'].set_linewidth(1.5)
# 将最外面的圆线加粗为2
ax2.spines['polar'].set_linewidth(1.5)

# 调整子图之间的间距
plt.subplots_adjust(wspace=0.5)
# 保存图片
plt.savefig('F:/数维杯/radar_plot.png')
# 显示雷达图
plt.show()
```

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import rcParams

LABELS = {
    'DFA/CS': 'DFA/CS',
    'DFA/CE': 'DFA/CE',
    'DFA/LG': 'DFA/LG',
    '焦油产量': 'Tar Yield',
    '水产量': 'Water Yield',
    '炭收率': 'Char Yield',
    '合成气收率': 'Syngas Yield',
}

rcParams['font.family'] = 'Arial'

data1 = pd.read_excel("D:/O/shuwei/1/DACS.xlsx", sheet_name='Sheet1')
data2 = pd.read_excel("D:/O/shuwei/1/DFACE.xlsx", sheet_name='Sheet1')
data3 = pd.read_excel("D:/O/shuwei/1/DFALG.xlsx", sheet_name='Sheet1')

fig, axs = plt.subplots(4, 3, figsize=(15, 12))

datasets = [data1, data2, data3]

for i, data in enumerate(datasets):
    x = data['DFA/CS'] if i == 0 else data['DFA/CE'] if i == 1 else data['DFA/LG']
    y1 = data['焦油产量']
    y2 = data['水产量']
    y3 = data['炭收率']
    y4 = data['合成气收率']

    for j, ax in enumerate(axs[:, i]):
        ax.scatter(x, [y1, y2, y3, y4][j], color=['Maroon', 'Navy', 'Silver', 'Olive'][j])

        # 使用拟合方程生成更多的 x 值
        x_smooth = np.linspace(x.min(), x.max(), 100)

        coef = np.polyfit(x, [y1, y2, y3, y4][j], 2)
        poly1d_fn = np.poly1d(coef)

        ax.plot(x_smooth, poly1d_fn(x_smooth), color="red", linewidth=2.5)

        ax.set_xticks(x)

        equation_text = f'y = {coef[0]:.2f}x^2 + {coef[1]:.2f}x + {coef[2]:.2f}'
        ax.text(0.05, 0.91, equation_text, ha='left', va='top', transform=ax.transAxes, fontsize=12)

        residuals = np.array([y - poly1d_fn(x_val) for x_val, y in zip(x, [y1, y2, y3, y4][j])])
```

```
        residuals = np.array([y - poly1d_fn(x_val) for x_val, y in zip(x, [y1, y2, y3, y4][j])])
        ss_res = np.sum(residuals**2)
        ss_tot = np.sum((np.mean([y1, y2, y3, y4][j]) - [y1, y2, y3, y4][j])**2)
        r_squared = 1 - (ss_res / ss_tot)
        r_squared_text = f'R_squared: {r_squared:.2f}'
        ax.text(0.335, 0.83, r_squared_text, ha='right', va='top', transform=ax.transAxes, fontsize=

        # 只给第一列子图添加y轴标题
        if i == 0:
            axs[j, i].set_ylabel([LABELS['焦油产量'], LABELS['水产量'], LABELS['炭收率'], LABELS['合

    # 设置x轴标题
    axs[3, i].set_xlabel(LABELS['DFA/CS'], fontsize=12) if i == 0 else axs[3, i].set_xlabel(LABELS['

axs[0, 0].set_xticklabels(["0", '0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8'])
axs[1, 0].set_xticklabels(["0", '0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8'])
axs[2, 0].set_xticklabels(["0", '0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8'])
axs[3, 0].set_xticklabels(["0", '0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8'])

axs[0, 1].set_xticklabels(['0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8'])
axs[1, 1].set_xticklabels(['0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8'])
axs[2, 1].set_xticklabels(['0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8'])
axs[3, 1].set_xticklabels(['0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8'])

axs[0, 2].set_xticklabels(['0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8'])
axs[1, 2].set_xticklabels(['0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8'])
axs[2, 2].set_xticklabels(['0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8'])
axs[3, 2].set_xticklabels(['0.1', '0.2', '0.3', '0.4', '0.5', '0.6', '0.7', '0.8'])

for i in range(4):
    for j in range(3):
        axs[i, j].spines['top'].set_linewidth(2.5)
        axs[i, j].spines['right'].set_linewidth(2.5)
        axs[i, j].spines['bottom'].set_linewidth(2.5)
        axs[i, j].spines['left'].set_linewidth(2.5)

for i in range(4):
    for j in range(3):
        axs[i, j].tick_params(axis='x', labelsize=14)
        axs[i, j].tick_params(axis='y', labelsize=14)

plt.tight_layout()
plt.savefig("D:/0/shuwei/1/1.png", dpi=400)
plt.show()
```

```python
# 设置中文字体
plt.rcParams['font.sans-serif'] = ['Arial Unicode MS']

# 解决负号'-'显示为方块的问题
plt.rcParams['axes.unicode_minus'] = False

# 创建包含数据的 DataFrame
data = {
    "DFA/CS": ["0/100", "20/100", "40/100", "50/100", "80/100", "100/100"],
    "H2": [6.48, 12.94, 22.65, 30.35, 37.93, 43.62],
    "CO": [25.93, 21.61, 20.53, 20.95, 22, 23.32],
    "CO2": [60.4, 57.84, 52.31, 51.92, 51.53, 50.91],
    "CH4": [23.41, 28.79, 33.11, 32.45, 30.83, 30.21],
    "C2H6": [2.09, 2.73, 3.08, 2.99, 2.96, 2.94],
    "C3H8": [0.55, 0.61, 0.62, 0.71, 0.88, 0.87],
    "C3H6": [0.35, 0.37, 0.41, 0.39, 0.37, 0.38],
    "C2H4": [0.71, 0.74, 0.81, 0.79, 0.77, 0.75],
    "C4H10": [0.32, 0.48, 0.56, 0.37, 0.28, 0.25]
}

df = pd.DataFrame(data)

# 设置DFA/CS作为索引
df.set_index('DFA/CS', inplace=True)

# 可视化堆积柱状图
df.plot(kind='bar', stacked=True, zorder=3, color=[
    (42/255, 68/255, 88/255),
    (51/255, 100/255, 133/255),
    (62/255, 134/255, 181/255),
    (231/255, 229/255, 223/255),
    (153/255, 184/255, 107/255),
    (104/255, 138/255, 76/255),
    (115/255, 107/255, 157/255),
    (183/255, 131/255, 175/255),
    (245/255, 166/255, 115/255),
    (252/255, 219/255, 114/255)
], edgecolor='black', linewidth=1.5)

plt.title('Yield of Decomposition Products from DFA/CS Pyrolysis wt.%(daf)', fontsize=16, pad=20)
plt.legend(bbox_to_anchor=(1, 1))
plt.grid(True, axis='y', linestyle='--', alpha=0.6)

plt.tick_params(axis='x', rotation=0, labelsize=12)

plt.tick_params(axis='y', rotation=0, labelsize=12)
plt.ylim(0, 160)

plt.savefig("D:/0/shuwei/4/3.png", dpi=400, bbox_inches="tight")
plt.tight_layout()
```

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np

# 读取excel文件
data = pd.read_excel("D:/0/shuwei/1/熵权法.xlsx")
labels = np.array(data['yield'])
weights = np.array(data['rate'])
theta = np.linspace(0, 2*np.pi, len(labels), endpoint=False)
radii = np.sqrt(weights / max(weights))

fig = plt.figure(figsize=(8, 8))

# 设置颜色列表
colors = [ (224/255, 210/255, 229/255), (191/255, 192/255, 222/255), (143/255, 152/255, 201/255), (9

# 绘制南丁格尔玫瑰图
ax = plt.subplot(111, polar=True)

# 调整极坐标网格线样式
ax.xaxis.grid(linewidth=1, color='black', alpha=0.5)  # 设置径向网格线样式
ax.yaxis.grid(linewidth=1, color='black', alpha=0.5, linestyle='--')  # 设置角度网格线样式

# 创建每个扇区的条形图，并设置颜色和边框
for i in range(len(labels)):
    color = colors[i % len(colors)]
    edgecolor = 'black'  # 设置边框颜色
    linewidth = 2  # 设置边框宽度
    bar = ax.bar(theta[i], radii[i], width=(2 * np.pi) / len(labels), bottom=0.0, color=color, linew

# 设置角度刻度和刻度标签，并加大字体
plt.xticks(theta, labels, fontsize=14)
plt.yticks([])  # 不显示半径刻度

# 加粗极径线
ax.spines['polar'].set_linewidth(2)

# 添加标题
plt.title('The weight obtained by entropy weight method', fontsize=15)
plt.savefig("D:/0/shuwei/1/5.png", dpi=400)
plt.show()
```



```python
import seaborn as sns

# 读取excel数据
data1 = pd.read_excel("D:/0/shuwei/2/2.xlsx", sheet_name='Sheet1')

# 提取 'DFA/CS' 和 'sum' 列数据
DFA_CS_data = data1['DFA/CE']
sum_data = data1['sum']

# 创建画布和坐标轴
fig, ax1 = plt.subplots()

# 定义颜色
colors = ["#edea9b",'#e6e7c5', '#b6b295', '#6d755e', '#5b6543', "#48513e"]

# 绘制柱状图
bars = ax1.bar(range(len(DFA_CS_data)), sum_data, width=0.6)
for i in range(len(bars)):
    bars[i].set_color(colors[i % len(colors)])

ax1.set_xlabel('DFA/CE', fontsize=14)

# 设置横坐标刻度和标签
ax1.set_xticks(range(len(DFA_CS_data)))
ax1.set_xticklabels(DFA_CS_data, rotation=0, fontsize=12)

# 设置纵坐标刻度字体大小
ax1.tick_params(axis='y', labelsize=12)

# 创建另一个坐标轴
ax2 = ax1.twinx()

# 绘制散点图
ax2.scatter(list(range(len(DFA_CS_data))), sum_data, color='#80A6E2')

# 添加线性回归线和置信区间
sns.regplot(x=list(range(len(DFA_CS_data))), y=sum_data, color="black", scatter=False, ci=95, ax=ax2

# 设置右边y轴范围
ax2.set_ylim(60, 90)

# 为每个柱状图添加黑色边框
for bar in bars:
    bar.set_edgecolor('black')
    bar.set_linewidth(1.2)
plt.savefig("D:/0/shuwei/3/2.png", dpi=400)
# 显示图形
plt.show()
```

Appendix（附录）