Problem Chosen	2023	Team Control Number
B	ShuWei Cup Summary Sheet	2023090819411

The catalytic reaction of cotton stalk pyrolysis

Summary

For question 1: Firstly, the data were pre-processed, and the relationship between different mixing ratios and pyrolysis product yields in the three pyrolysis combinations was analyzed descriptively using the combination graph (column + line). After linear observation, the Pearson correlation coefficient analysis was used to analyze the correlation between the mixing ratios in the three pyrolysis combinations and their pyrolysis product yields. By using Lasso regression model to investigate the effect, it is found that DFA plays an important role as a catalyst in promoting the pyrolysis of CS\CE and LG.

For question 2: Pearson correlation coefficient analysis was used to analyze the correlation between different mixing ratios in the three pyrolysis combinations and the yields of pyrolysis gases in each group. We found that DFA promoted the thermal decomposition of CS and increased the yield of CS pyrolysis gas. Meanwhile, DFA also had different catalytic effects on the generation of pyrolysis gas from CE and LG components. Next, the area map and line chart are further visualized.

For question 3: First, normality test was performed on the data. The yields of pyrolysis products Tar yield and Char yield failed the normality test. The difference analysis was conducted using paired sample Wilcoxon symbolic rank test, and it was found that there were significant differences in the yields of pyrolysis products CE and LG. Water yield and Syngas yield were tested by normality test, and differences were analyzed by paired sample T test. It was found that different biomass had no significant influence on Water yield, but had significant influence on Syngas yield. Different gas yields under different biomass were all normally distributed. Two-factor analysis of variance found that different biomass had significant effects on gas yields.

For question 4: the analytical diagram is used to visualize the thermal decomposition process of cotton straw. Next, the kinetic model is used to analyze the reaction mechanism model of desulfurization ash, and the overall catalytic effect of desulfurization ash is quantified. Next, the regression relationship between the mixture ratio, pyrolysis products and gas is analyzed by multiple linear regression.

For question 5: Based on the characteristics of small samples, the grey prediction model is used to predict the output. The stage ratio, posterior difference ratio C value and average relative error are all tested with good accuracy, and the specific data of tar, water, coke residue and syngas in the next seven stages are predicted. Combined with the standardized and non-standardized model formulas of Lasso model proposed above, the data of gray prediction is brought in. After testing, it can be seen that the corresponding mixing ratio in the later stages is more reasonable, and the combination of the two different methods makes the prediction more reliable.

Keyword: Pearson correlation coefficient, Lasso regression, normal test, paired sample T test, multiple linear regression, grey prediction

Content

1. Introduction

1.1 Background

As the basis for human survival and development, energy is an important guarantee for the prosperity and development of a country. With the rapid development of economy, the demand for energy is increasing day by day, so renewable energy has received unprecedented attention. As a kind of renewable energy, biomass as a zero-carbon emission raw material, the total CO2 emission in the whole life cycle is zero, and has developed into a relatively mature state. It is the only sustainable and renewable carbon-based resource[3].

Agricultural waste, cotton stalks, is one of them. Cotton stalk contains a lot of cellulose and lignin, which can produce a variety of renewable energy through pyrolysis. However, the energy efficiency of pyrolysis process is affected by many factors such as catalyst and temperature[4].

For question 1: For each of the pyrolysis combinations in Annex I, analyze the relationship between the yield of pyrolysis products (tar, water, coke residue, syngas) and the mixture ratio of the corresponding pyrolysis combinations, and indicate whether desulfurization ash as a catalyst plays an important role in promoting the pyrolysis of cotton stalk, cellulose and lignin.

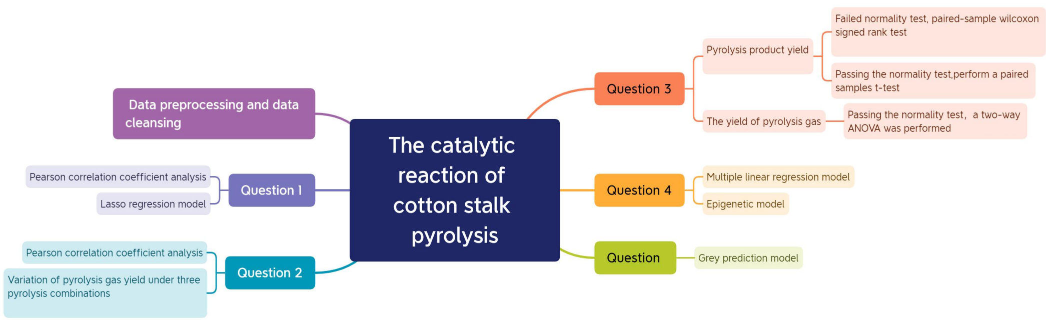
For question 2: According to Annex II, for each of the three pyrolysis combinations, the influence of the mixture ratio of the pyrolysis combinations on the pyrolysis gas yield of each group is discussed and illustrated by making corresponding images.

For question 3: Is there a significant difference in the yield of CE and LG pyrolysis products and the yield of pyrolysis gas components under the same proportion of desulfurization ash catalysis? Please state the reason.

For problem 4: the catalytic reaction mechanism model of desulfurization ash of CE and LG model compounds was established, and the reaction kinetics model was established for analysis.

For problem 5: using mathematical models or artificial intelligence learning methods, the yield or quantity of pyrolysis products is predicted under limited data conditions.

1.2 Work



2. Problem analysis

2.1 Analysis of question one

Firstly, descriptive analysis was conducted on the relationship between different mixing ratios and pyrolysis product yields in the three pyrolysis combinations. Then, Pearson correlation coefficient analysis was used to analyze the correlation between the mixing ratios and pyrolysis product yields in the three pyrolysis combinations. The Lasso regression model showed that DFA played an important role in promoting the pyrolysis of CS\CE and LG as a catalyst.

2.2 Analysis of question two

Pearson correlation coefficient analysis was used to analyze the correlation between the yields of pyrolysis gases in different mixtures of the three pyrolysis combinations, and it was found that DFA promoted the thermal decomposition of CS and increased the yield of CS pyrolysis gas, and also had different catalytic effects on the process of CE and LG components generating pyrolysis gas.

2.3 Analysis of question three

Firstly, normality test was conducted on the data. The yields of Tar yield and Char yield of pyrolysis products failed the normality test. The difference analysis was conducted using paired sample Wilcoxon symbolic rank test, and it was found that there were significant differences in the yields of CE and LG pyrolysis products. Water yield and Syngas yield were tested by normality test, and differences were analyzed by paired sample T test. It was found that different biomass had no significant influence on Water yield, but had significant influence on Syngas yield. Different gas yields under different biomass were all normally distributed. Two-factor analysis of variance found that different biomass had significant effects on gas yields.

2.4 Analysis of question four

First, the thermal decomposition process of cotton stalk was visualized using the analytical diagram; then, the reaction mechanism model of desulfurization ash was analyzed using the performance kinetic model to quantify the overall catalytic effect of desulfurization ash; then, the regression relationship between the mixture ratio, pyrolysis products and gas was analyzed using multiple linear regression. The model fit was good and the formation mechanism was quantified.

2.5 Analysis of question five

Based on the characteristics of small samples, we first collate the data and use the gray prediction model to forecast the output. The stage ratio, the posterior difference ratio C value and the average relative error have all passed the test with good accuracy, and the specific data of tar, water, coke residue and syngas etc. in the next seven stages have been predicted[6]. Combined with the standardized and non-standardized model formulas of Lasso model proposed above, the data of gray prediction is brought in. After testing, it can be seen that the corresponding mixing ratio in the later stages is more reasonable, and the combination of the two different methods makes the prediction more reliable.

3. Symbol Description

Table 1 Symbol Description

Serial number	Symbol	Implication
1	Yi	i mixing ratio
2	Xi	Pyrolysis product i of cotton stalk
3	Xi - 1	The I-th product of cellulose
4	Xi - 2	The i-th product of lignin
5	λ	Model parameter
6	Zi	Cotton stalk i pyrolysis gas
7	Zi - 1	The i-th gas of cellulose
8	Zi - 2	The i-th gas of lignin
9	α	Standard deviation
10	σ	intercept

4. Fundamental assumptions

(1) The influence of temperature and heating rate on the pyrolysis of cotton stalk is constant.

(2) Under the same mixing ratio of desulfurization ash, the yield of CE and LG under high temperature decomposition will not affect each other.

(3) The relationship between pyrolysis products, pyrolysis gas and mixing ratio is quantifiable and can be described by mixing ratio.

(4) The generation of products and gases and the pyrolysis reaction are independent of each other, and no secondary reaction will occur.

(5) The model of catalytic effect and reaction mechanism can be quantitatively described.

(6) External factors affecting the pyrolysis process, such as air pressure, are not considered.

(7) Errors in experimental data can be ignored in the model.

5. Model establishment and solution

5.1 The establishment and solution of problem 1

5.1.1 Pearson correlation coefficient

The Pearson correlation coefficient between two variables is defined as the quotient of covariance and standard deviation between two variables:

$$\mathrm{\rho X,Y} = \frac{\mathrm{cov(X,Y)}}{\mathrm{\sigma X \sigma Y}} = \frac{\mathrm{E[(X-\mu X)(Y-\mu Y)]}}{\mathrm{\sigma X \sigma Y}},$$

The above defines the overall correlation coefficient, which is often represented by the Greek lowercase letter ρ. Pearson correlation coefficient can be obtained by estimating the covariance and standard deviation of the sample, commonly represented by the lowercase letter r:

$$r = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}}.$$

r can also be estimated by the mean value of the standard scores of sample points (X_i, Y_i), and an expression equivalent to the above formula can be obtained:

$$\mathrm{r} = \frac{1}{\mathrm{n}-1}\sum_{\mathrm{i-1}}^{\mathrm{n}}\left(\frac{\mathrm{X_i}-\overline{\mathrm{X}}}{\mathrm{\sigma X}}\right)\left(\frac{\mathrm{Y_i}-\overline{\mathrm{Y}}}{\mathrm{\sigma Y}}\right)$$

Where $\frac{X_i-\bar{X}}{\sigma X}$, $\bar{X}$ and σX are the standard score, sample mean and sample standard deviation of X_i sample respectively.

5.1.2 Lasso model

Lasso regression is an alternative to the least square method of compression estimation, full name is minimum absolute selection and contraction operator. Assume that the data $\{X_i, Y_i\}, X_i = \{x_{i1}, \ldots, x_{im}\}$ and Y_i are the explanatory and corresponding variables corresponding to the i th observation, respectively. Consider a linear regression model:

$$Y_i = \alpha_i + \sum_{j=1}^{m} \beta_j x_{ij} + \varepsilon_i \,, \varepsilon_i \sim N(0, \sigma^2)$$

In the usual regression structure, assuming that the observations are independent of each other, or independent if the observations are given, that is, Y_i is independent about X_i conditions, while assuming that x_{ij} is the most standardized, that is $\frac{1}{n}\sum_j x_{ij} = 0, \frac{1}{n}\sum_j x_{ij}^2 = 1$,Lasso estimates:

$$(\hat{\alpha}, \hat{\beta}) = arg \min_{\beta} \left\{ \sum_i (Y_i - \alpha_i - \sum_j \beta_j x_{ij})^2 \right\}$$

$$s.t. \sum_j |\beta_i| \le t$$

Here t>=0, which is the harmonic parameter. Now for all t, we have an estimate of α, $\hat{\alpha}=\bar{y}$. Without loss of generality, $\bar{y}$=0 is assumed, so α is omitted, and the control of the harmonic p2arameter t makes the regression coefficient smaller overall, if

$$t^0 = \sum_j |\beta_j| \, , t \le t^0$$

Some of the regression coefficients will shrink to zero, and some of the coefficients will even be equal to zero.

5.1.3 Solution of model

We first preprocessed the data and found that the sum of the output of the pyrolysis products was 100, abnormal values were excluded and there was no missing value in the table. Then we cleaned the data and found that the relative error of the parallel experiment was about 5%, and the next analysis could be carried out within the allowed error range.

Data preprocessing:

Let's figure out that the production does add up to 100. Exclude outliers. There are no missing values in the table.

Data cleaning:

The relative error of parallel test is about 5%.

As shown in Figure 1, it can be seen from the bar chart that in the combination of DFA/CS, the tar yield decreases significantly with the increase of the proportion of desulfurization ash, which indicates that DFA plays a promoting role in the process of CS pyrolysis and reduces the tar generation. With the increase of DFA proportion, Char yield gradually decreased from the first proportion to the second, Water yield increased to the first, Syngas yield increased and then gradually decreased. The line in Figure 1 shows the growth rate changes of pyrolysis products under different DFA ratios[7]. We find that the growth rate of Tai yield has maintained a negative growth, showing a continuous downward trend, while Water yield and Char yield have maintained a low growth. With the increase of the proportion of DFA, Syngas yield initially continued to decline to negative growth and then increased, and the overall production was unstable.

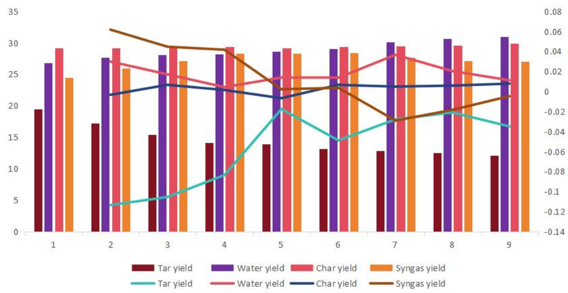


Figure 1 Yield and growth rate of DA/CS pyrolysis decomposition products.wt.%(daf)

In the combination of DFA/CE, with the increase of the ratio of DFA to CE, the tar yield has a significant impact. Tar production rose first, then declined slightly, but then increased again. The results showed that the increase of DFA ratio could increase the tar yield. The proportion of Water yield decreased first and then increased, and its proportion growth showed negative growth first, positive growth at the mixed ratio of 60/100-80/100, and then negative growth.

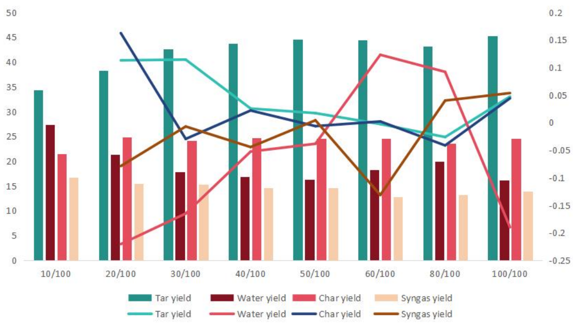


Figure 2 The yield and growth rate of DFA/CE pyrolysis decomposition products.wt.%(daf)

In the DFA/LG combination, with the increase of DFA, the tar yield showed a downward trend, while the Water yield showed an upward trend, and the yield of the other two products did not change significantly. This indicates that the increase of DFA effectively promotes the pyrolysis process of LG. The changes of pyrolysis product yields under different mixing ratios of DFA/LG pyrolysis combination are shown in figure 3. The column chart shows the proportion of pyrolysis products under different mixing ratios. The line chart shows the changes of growth rates of the yields of four pyrolysis products.

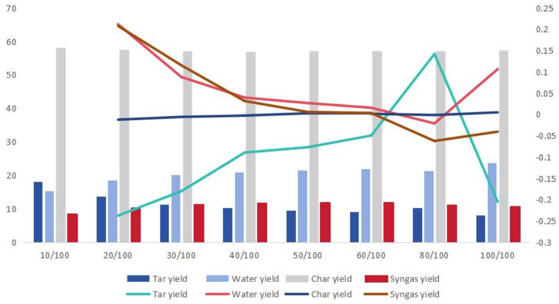


Figure 3 The yield and growth rate of DFA/LG pyrolysis decomposition products.wt.%(daf)

We first processed the data in Annex I, and then selected Pearson correlation coefficient to analyze the relationship between the yield of pyrolysis products and the mixing ratio of corresponding pyrolysis combinations. At this time, the significance P-values of DFA/CS and Tai yield, Water yield and Char yield were 0.002***, 0.000*** and 0.001***, respectively, which were all less than 0.05. It shows that DFA/CS is significantly correlated with Tai yield, Water yield and Char yield, and the correlation coefficient between DFA/CS and Tai yield is -0.883, that is, there is a negative correlation between them. The greater DFA/CS is, the less Tai yield is. The correlation coefficients between DFA/CS and Water yield and Char yield are 0.984 and 0.898 respectively, that is, they show a high positive correlation. The larger the DFA/CS is, the more Water yield and Char yield will be. The significance P-value of DFA/CE and Syngas yield was 0.265 and greater than 0.05, indicating that there was no significant correlation between the two, that is, the change of mixing ratio had no correlation with Syngas yield.

The significant P values of DFA/CE and Tai yield and Syngas yield were 0.029** and 0.020***, respectively, which were both less than 0.05, indicating that DFA/CE was significantly correlated with Tai yield and Syngas yield. The correlation coefficient between DFA/CE and Tai yield is 0.759, that is, there is a high positive correlation between them. The greater the DFA/CE, the more Tai yield. The correlation coefficient between DFA/CE and Syngas yield is -0.864, that is, there is a negative correlation between them. The greater the DFA/CE, the lower the Syngas yield. The P values of significance between DFA/CE and Water yield and Char yield were 0.073* and 0.293, respectively, which were greater than 0.05, indicating that DFA/CE was not significantly correlated with Water yield and Char yield. That is, changes in Water yield and Char yield are independent of changes in mixing ratio.

The significant P-values of DFA/LG with Tai yield and Water yield were 0.019** and 0.005***, respectively, which were both less than 0.05, indicating that there was a significant correlation between DFA/LG and Tai yield and Water yield. Moreover, the correlation coefficient between DFA/LG and Tai yield is -0.794, that is, there is a high negative correlation between them. When DFA/LG is larger, Tai yield is smaller. The correlation coefficient between DFA/LG and Water yield is 0.866, that is, they are highly correlated, and the greater the DFA/CE, the more Water yield. The significance P-values of DFA/CE with Char yield and Syngas yield were 0.334 and 0.290, respectively, which were greater than 0.05, indicating that DFA/CE was not significantly correlated with Char yield and Syngas yield. That is, changes in Char yield and Syngas yield are independent of changes in the mixing ratio.

Table 2 Phase relationship between corresponding pyrolysis combination and pyrolysis product yield

	DFA/CS	DFA/CE	DFA/LG
Tar yield	-0.883(0.002***)	0.759(0.029**)	-0.794(0.019**)
Water yield	0.984(0.000***)	-0.604(0.113)	0.866(0.005***)
Char yield	0.898(0.001***)	0.367(0.371)	-0.394(0.334)
Syngas yield	0.416(0.265)	-0.79(0.020**)	0.428(0.290)

We used Lasso regression model to analyze the data in the DFA/CS combination, and found that the parameters $\lambda=0.02$ and R^2 were 0.973 at this time, indicating that the model had a good fitting effect, and the regression results could explain the data, that is, desulfurization and acting as a catalyst played an important role in promoting the pyrolysis of mogan. The results of the regression model show that the standardization coefficient of Char yield and Syngas yield is 0, so these two pyrolysis products are deleted from the standardization formula. However, the non-standardization coefficient of Char yield in the non-standardization formula is 0.322, while the coefficient of Syngas yield is still 0. Considering that the change of DFA proportion in Pearson correlation coefficient analysis is significantly correlated with Tar yield, Water yield and Char yield, but has no significant effect on Syngas yield, the non-standardized formula of the model is chosen here to explain the relationship between DFA/CS and pyrolysis products.

In the DFA/CE combination, λ value determined by Lasso regression model is 0.04 and R^2 is 0.66. The model fits well and can be used to explain the relationship between DFA and Tai yield, Water yield, Char yield and Syngas yield. The results show that DFA plays an important role in promoting the pyrolysis of CE as a catalyst.

In the DFA/LG combination, λ value and R^2 value of Lasso regression model are 0.263 and 0.858, which can be used to explain the relationship between DFA and Tai yield, Water yield, Char yield and Syngas yield in this pyrolysis combination. The results show that DFA plays an important role in promoting the pyrolysis of LG as a catalyst.

5.2 The establishment and solution of problem 2

5.2.1 Pearson correlation coefficient

Same as 5.1.1

5.2.2 Solution of model

Pearson correlation coefficient was used to analyze the correlation between the three pyrolysis combinations and the changes in gas yield. The significant P-values of

DFA/CS and H_2, CO_2 and C_3H_8 gas yields were 0.000***, 0.021** and 0.002***, respectively, less than 0.05. The results showed that the yield changes of the three gases, H_2, CO_2 and C_3H_8, were correlated with the increase of the proportion of DFA. The correlation coefficients of H_2 and C_3H_8 were all greater than 0.95, showing a high positive correlation. The higher the proportion of DFA, the higher the yield of the two gases. The correlation coefficient between CO_2 yield and DFA is -0.879, which indicates that they have a negative correlation. With the increase of DFA, CO_2 yield decreases. The P-values of the yield significance of DFA and the six gases CO, CH_4, C_2H_6, C_3H_6, C_2H_4 and C_4H_{10} are all greater than 0.05, as shown in Table several, indicating that the increase of DFA ratio has no significant correlation with the yield of these six gases, that is, the change of DFA ratio has little effect on the change of the yield of these six gases.

In the DFA/CE combination, only H_2, CO, CO_2, CH_4 and C_2H_6 are produced, among which the significance P-values of DFA with CO, CO_2 and CH_4 are 0.011**, 0.034** and 0.042**, respectively, indicating that DFA is significantly correlated with CO, CO_2 and CH_4. Their correlation coefficients are all less than -0.89, which means that they have a negative correlation, that is, the greater the proportion of DFA, the lower the yield of CO, CO_2 and CH_4. However, the significance P-value of DFA with H_2 and C_2H_6 is greater than 0.05, which indicates that the increase of DFA proportion has no significant correlation with the change of H_2 and C_2H_6 yield, that is, the change of DFA proportion has little effect on the yield of these two gases.

In the DFA/LG combination, the significance P value of DFA and CO is only 0.002***, indicating that there is a significant correlation between the two, and their correlation coefficient is 0.987, which indicates that the CO yield increases with the increase of DFA proportion. The significance P-values of DFA with H_2, CO_2, CH_4 and C_2H_6 are all higher than 0.05, as shown in Table 3, which means that DFA has no significant correlation with the changes in the yield of these four gases, and has little impact on their yield.

Table 3 Pearson phase relation table of pyrolysis gas yield under three pyrolysis combinations

	DFA/CS	DFA/CE	DFA/LG
H_2	0.988(0.000***)	0.819(0.090*)	0.86(0.062*)
CO	-0.272(0.602)	-0.955(0.011**)	0.987(0.002***)
CO_2	-0.879(0.021**)	-0.907(0.034**)	0.848(0.070*)
CH_4	0.568(0.240)	-0.892(0.042**)	-0.002(0.998)
C_2H_6	0.687(0.132)	-0.871(0.054*)	-0.594(0.291)
C_3H_8	0.961(0.002***)		
C_3H_6	0.312(0.547)		
C_2H_4	0.362(0.481)		
C_4H_{10}	-0.506(0.306)		

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

The accumulation area diagram of nine gas yields in different DFA/CS combinations is shown in Figure 4. We found that with the increase of DFA ratio, the pyrolysis yield of CS gradually increased, from 120.24mL/g,daf, 0/100 to

153.2mL/g,daf, 100/100, which was evident in the pyrolysis process of DFA/CS. DFA has obvious catalytic effect on the formation of main components in pyrolysis gas.

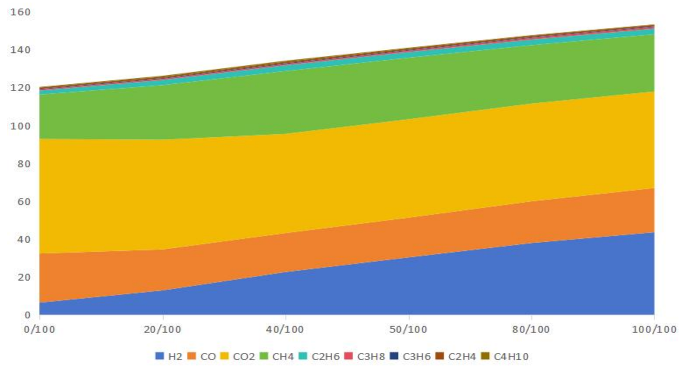


Figure 4 Yield of Gaseous Components from DFA/CS Pyrolysis (mL/g,daf)

The yield changes of nine gases under different mixing ratios of DFA/CS are shown in figure 5. It can be seen from FIG. 1 that DFA significantly promotes the generation of H_2 during CS pyrolysis. When the mixing ratio of DFA/CS is 50/100 and 100/100, the yield of H_2 is 30.35 mL/g and 43.62 mL/g, respectively. The H_2 yield is 4.68 times and 6.73 times that of 6.48m L/g during separate pyrolysis of cotton stalk, respectively. With the increase of DFA/CS mixture ratio, the production of CO and CO_2 is inhibited, and the yield of CO and CO_2 decreases continuously. When the mixture ratio of desulfurization ash/cotton stalk is 100/100, the CO and CO_2 yields are 89.93% and 84.29% of the CO_2 and CO yields of cotton stalk pyrolysis alone. When the mixture ratio of desulfurization ash/cotton stalk is 50/100, the yields of CO and CO_2 are 20.95 mL/g and 51.92 mL/g, respectively, which are 14.04% and 19.21% lower than the yields of corresponding products during the pyrolysis of cotton stalk alone.

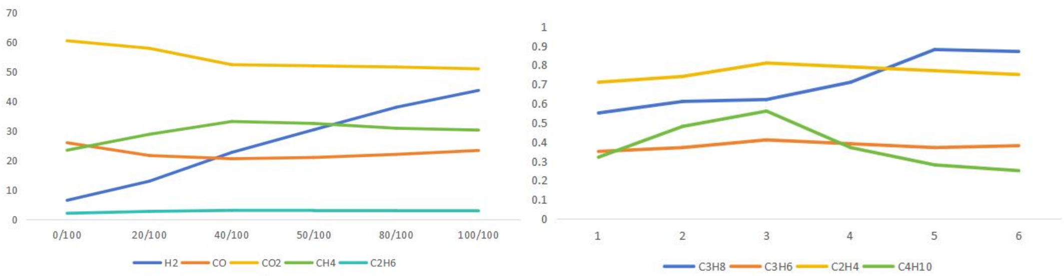


Figure 5 Yield of Gaseous Components from DFA/CS Pyrolysis (mL/g,daf)

As shown in Figure 6, the yield accumulation area diagram of five gases with different mixing ratios in different DFA/CE combinations shows that the total yield of five gases increases and then decreases, indicating that different mixing ratios effectively affect the gas yield, among which the yield of H_2 changes most obviously. The mixture ratio of 20/100-80/100 showed a continuous increase and decreased at higher mixing ratios.

As shown in Figure1, the visualization diagram of DFA/CE combination under different mixing ratios shows, we can find that the change of H_2 is the most obvious. When the H_2 yield increases, the yield of the other four gases shows a downward trend, and when the H_2 yield decreases, the yield of the other four gases increases.

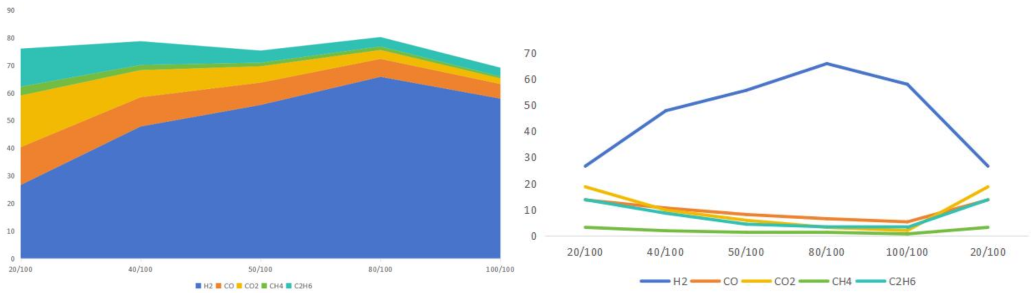


Figure 6 Visualization of DFA/CE yield of five gases at different mixing ratios

In the DFA/LG combination, as shown in Figure 7, we found that the total gas yield showed an upward trend under the combination, indicating that with the increase of DFA ratio, the gas yield also increased.

The visual diagram of the changes in the yield of five gases under different mixing ratios of DFA/LG combination is shown in Figure 7, where the left ordinate represents the yield percentage of the three gases CO, CO2 and CH4, and the right ordinate represents the yield percentage of H2 and CH6. It can be found that the yield of H2 and CH6 gases in this pyrolysis combination is relatively small, and the yield of CO, CO2 and CH4 generally increases first and then decreases with the increase of DFA proportion.

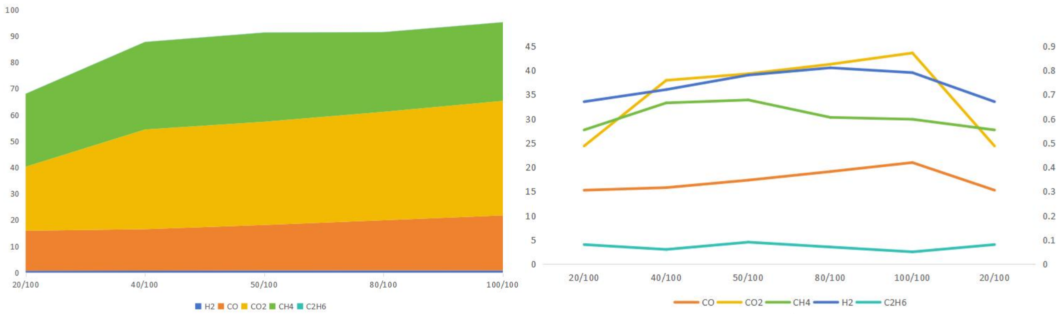


Figure 7 Yield of Gaseous Components from DFA/LG Pyrolysis (mL/g,daf)

5.3 The establishment and solution of problem 3

5.3.1 Paired sample T test

The paired sample T test can be regarded as an extension of the single sample T test, but the object of the test is changed from a circle of independent samples from the normal distribution to the difference of the observed values of the two-group paired samples. The basic steps are as follows:

Step1: Assume

Null hypothesis: The difference d of the paired sample means 0

Alternative hypothesis: The difference d between the two paired samples is not 0 on average

Step2: Sampling distribution type

The difference d between the two pairs is approximately normally distributed, so it meets the T-distribution condition.

Step3: Test direction

The alternative hypothesis is that the difference d between the two paired samples is different, whether greater or less, so the two-tail test must be used.

Step4: Solve the T value

Based on the current null hypothesis and the current sample, the T-value is calculated, representing the probability of drawing the current sample and the more extreme sample by the probability of the T-value and the occurrence of a value more extreme than the value.

Step5: Calculate the sample size and solve the P-value of degrees of freedom

If the difference between the two paired samples x1i and x2i is di=xi1-xi2 independent and from normal distribution, then whether the population expected value of di is 0 can be obtained using the following statistics:

$$t = \frac{\bar{d} - \mu 0}{sd/\sqrt{n}}$$

Where i=1...n, $\bar{d}=\frac{\sum_{i=1}^{n} d_i}{n}$ is the mean of the difference between the paired samples, $s_d=\sqrt{\frac{\sum_{i=1}^{n}(d_i-\bar{d})^2}{n-1}}$ is the standard deviation of the difference between the paired samples, and n is the number of paired samples. This statistic t follows a t distribution with n-1 degrees of freedom under the null hypothesis: μ=μ0 is true.

Step6: Solve the mean confidence interval

Lower limit of confidence interval = difference between two mean values -t value * standard error

Upper limit of confidence interval = difference between two mean values +t value * standard error

5.3.2 Two-factor variance factor

To test whether H_0 is true, we need to determine the statistics of the test row factors and column factors respectively. As with the method of constructing statistics by one-way ANOVA, it is also necessary to decompose the total sum of squares. The total sum of squares is the sum of the squared errors of all sample observations $x_{ij}(i = 1,2,\ldots,k; j=1,2,\ldots,r)$ and the total sample mean $\bar{\bar{x}}$, denoted as SST, i.e. :

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{r}(x_{ij} - \bar{\bar{x}})^2 = \sum_{i=1}^{k}\sum_{j=1}^{r}(x_{i*} - \bar{\bar{x}})^2 + \sum_{i=1}^{k}\sum_{j=1}^{r}(x_{*j} - \bar{\bar{x}})^2 + \sum_{i=1}^{k}\sum_{j=1}^{r}(x_{ij} - x_{i*} - x_{*j} - \bar{\bar{x}})^2$$

The first term on the right hand side of the decomposed equation is the sum of the squares of the errors generated by the row factors. Let's call it SSR, or $\sum_{i=1}^{k}\sum_{j=1}^{r}(x_{i*} - \bar{\bar{x}})^2$.

The second term is the sum of the squares of the errors resulting fro m the factors listed, denoted as SSC, $\sum_{i=1}^{k}\sum_{j=1}^{r}(x_{*j}-\bar{\bar{x}})^2$.

The third item is the sum of the squares of errors generated by the remaining factors other than the row factors and column factors, which is called the sum of squares of random errors and is denoted by SSE, that is, $\sum_{i=1}^{k}\sum_{j=1}^{r}(x_{ij}-x_{i*}-x_{*j}-\bar{\bar{x}})^2$.

Relation of sum of squares: SST = SSR + SSC +SSE

On the basis of the above sum of squares of error, the mean square is calculated, that is, the sum of squares is divided by the corresponding degree of freedom, which is the mean square[9]. The degrees of freedom corresponding to the sum of squares of error are:

The total sum of squares SST is kr-1.

The square error of the row factor and the degree of freedom of SSR are k-1.

The squares of error of the column factors and the degrees of freedom of SSC are r-1;

The random error squares and SSE have degrees of freedom (k-1)(r-1).

To construct the test statistic, the following mean squares need to be calculated:

Mean square of row factors, denoted as MSR:$MSR=\frac{SSR}{k-1}$

The mean square of the column factors, denoted MSC:$MSC=\frac{SSC}{r-1}$

The mean square of the random error term, denoted as MSE:

$$MSE=\frac{SSE}{(k-1)(r-1)}$$

To test whether the influence of the row factors on the dependent variables is significant, the following statistics are used:

$$F_R=\frac{MSR}{MSE}\sim F(k-1,(k-1)(r-1))$$

To test whether the influence of column factors on dependent variables is significant, the following statistics are used:

$$F_C=\frac{MSC}{MSE}\sim F(r-1,(k-1)(r-1))$$

5.3.3 Solution of model

We first conducted a normal test on the yield of CE and LG pyrolysis products under the same mixing ratio. Wilcoxon signed rank test was used to pair the yields of four pyrolysis products Tar yield, Water yield, Char yield and Syngas yield of CE and LG biomass. It was found that the significance P-values of Tar yield and Char yield were both 0.012** and less than 0.05, showing significance at the level, so the data did not meet the normal distribution and did not pass the normal test. The model could co nduct difference analysis on the two groups of data. At this time, Cohen's d values of the yields of the two products are 8.804 and 39.195, respect ively, which are greater than 0.8, indicating that there are significant differ ences in the yields of Tar yield and Char yield of CE and LG pyrolysis pr oducts under the catalysis of the same proportion of desulfurization ash. While the significance P-values of Water yield and Syngas yield are 0.084* and 0.335, respectively, greater than 0.05, showing no significance at the level. Therefore, the data meet the normal distribution, and independent s ample T test is required for difference analysis. Under this model, the sig nificance P value of Water yield is 0.607 and greater than 0.05, and Cohe n's d value is 0.19, which indicates that there is no significant difference i n Water yield in different biomass components. The significance P value o f Syngas yield is 0.002*** less than 0.05, which indicates that there are si gnificant differences in the yield of Syngas yield in different biomass comp onents. Cohen's d value of 1.634 is greater than 0.8, which indicates that the Syngas yield varies greatly in different biomass components.

Table 4 Paired samples Wilcoxon signed rank test

Variable name	Mean value	Standard deviation	Skewness	Kurtosis	S-W test	P	Cohen's d
Tar yield.1pairing Tar yield.2	30.783	6.954	-1.578	2.022	0.818(0.045**)	0.012**	8.804
Char yield.1 pairing Char yield.2	-33.271	1.453	-2.467	6.362	0.649(0.001***)	0.012**	39.195

Table 5 Paired sample T test results

Variable name	Mean value	Standard deviation	Skewness	Kurtosis	S-W test	P	Cohen's d
Water yield.1 pairing Water yield.2	-1.173	6.165	1.661	3.055	0.845(0.084*)	0.607	0.19
Syngas yield.1 pairing Syngas yield.2	3.661	2.241	1.256	2.168	0.907(0.335)	0.002***	1.634

Shapiro-Wilk test method was used to test the normality of the pyrolysis gas yields of CE and LG biomass. The sample size was less than 5000, and t he test results were shown in Table several. It was found that the significant P -values were all greater than 0.05, indicating that the pyrolysis gas yields met the normal distribution[10]. Therefore, we can use two-factor ANOVA to analyze the difference of pyrolytic gases.

Table 6 Overall description result

Variable name	Sample size	Median	Mean value	Standard deviation	Skewness	Kurtosis	S-W test	The K-S test
H2.1	5	55.6	50.74	14.942	-1.273	1.887	0.909(0.461)	0.288(0.907)
CO.1	5	8.1	8.84	3.364	-0.686	-0.634	0.954(0.767)	0.187(0.981)
CO2.1	5	5.9	7.92	6.73	1.283	1.3	0.889(0.354)	0.218(0.929)
CH4.1	5	1.3	1.68	0.92	1.202	1.686	0.904(0.435	0.255(0.827)
C2H6.1	5	4.4	6.68	4.54	1.241	0.453	0.836(0.153)	0.292(0.694)
H2.2	5	0.78	0.754	0.058	-0.85	-0.888	0.907(0.453)	0.274(0.764)
CO.2	5	17.305	17.661	2.365	-0.524	-1.329	0.943(0.685)	0.189(0.978)
CO2.3	5	39.26	37.262	7.526	-1.791	3.548	0.813(0.102)	0.334(0.529)
CH4.4	5	30.26	30.982	2.557	-0.064	-1.65	0.929(0.591)	0.212(0.940)
C2H6.5	5	0.07	0.07	0.016	0	-1.2	0.987(0.967)	0.136(1.000)

In the two-factor analysis of variance, we take CE and LG as the first la yer classification, and the five mixed ratios of 20/100, 40/100, 50/100, 80/100 and 100/100 as the second level classification. At this time, the P values of H 2, CO, CO2, CH4 and C2H6 groups were 0.002***, 0.025**, 0.010***, 0.000 *** and 0.031** *, respectively, which were all less than 0.05, showing signifi cant significance at the level, indicating that different biomass had significant e ffects on the yields of the five pyrolysis gases. In other words, under the catal ytic action of the same proportion of desulfurization ash, there is a significant difference in the yield of CE and LG pyrolysis gas.

The yield of high-temperature decomposition gas components met the nor mality test, so two-factor ANOVA was used to classify CE and LG as the first layer, and the five mixing ratios of 20/100, 40/100, 50/100, 80/100 and 100/1 00 as the second level, and the differences of each gas were discussed.

Table 7 Results of two-factor ANOVA F test

Item	H2	CO	CO2	CH4	C2H6
Intercept	0.002***	0.000***	0.002***	0.000***	0.029**
Group	0.002***	0.025**	0.010***	0.000***	0.031**
Mixing ratio	0.494	0.99	0.999	0.709	0.499

5.4 The establishment and solution of problem 4

5.4.1 Linear regression

Linear regression is a statistical analysis method that uses regression analy sis in mathematical statistics to determine the interdependent quantitative relatio nship between two or more variables, which is widely used. It is expressed as $y = w'x+e$, where e is a normal distribution where the error follows a mean of 0.

In regression analysis, only one independent variable and one dependent v ariable are included, and the relationship between them can be approximated b y a straight line. This regression analysis is called unitary linear regression ana lysis. If the regression analysis includes two or more independent variables and there is a linear relationship between the dependent variable and the independ ent variable, it is called multiple linear regression analysis.

In general, linear regression can be solved by least square method or grad ient descent method, which can be calculated for y=bx+a line. Take the least s quare method as an example. In general, there is more than one factor affectin g y[11]. Let's say x1, x2,... xk, k factors, can usually be considered as the follo wing linear relationship

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon$$

For y and x1, x2... xk makes n independent observations at the same time to obtain n sets of observations (xt1, xt2,... xtk), t=1,2,... ,n (n>k+1), they sat isfy the relation:

$$y = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_k x_{tk} + \varepsilon_t$$

Among them, $\varepsilon_1 \ldots \varepsilon_n$ is uncorrelated and is a random variable with the sa me distribution as ε. In order to express the above formula as a matrix, let:

$$Y = X\beta + \varepsilon$$

The solution of β is obtained using the least square method

$$\widehat{\beta} = (X^T X)^{-1} X^T Y$$

Where $(X^T X)^{-1} X^T Y$ is called the pseudo-inverse of X

Slope b calculation method

Method 1: Use σ

$$u(b) = \frac{\sigma}{\sqrt{\sum (x - \bar{x})^2}}$$

5.4.2 Solution of model

CS contains two kinds of biomass, CE and LG. Under the catalytic action of DFA, these two kinds of biomass decompose and gasification under heat to release volatile gases. The following figure is the simulation diagram of the thermal decomposition process of CE and LG.

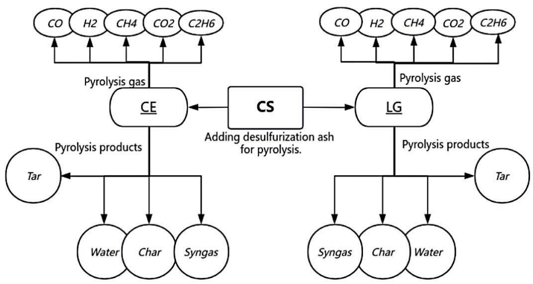


Figure 8 The most basic single step pyrolysis process for CE and LG

The process of biomass pyrolysis: biomass → volatile matter + solid carb on.

The biomass sample with initial mass m_0 will undergo pyrolysis reaction under heating condition. At a certain time t and mass m, the decomposition rat e can be expressed as follows:

$$d\alpha/dt = kf(\alpha)$$

$$\alpha = \frac{(m_0 - m)}{(m_0 - m_\infty)}$$

$$\mathrm{k} = \mathrm{Aexp}(-\mathrm{E}/\mathrm{RT})$$

In the above formula:

A: pre-factor (1/min);

K:Arrhenius rate constant;

E: reaction activation energy (KJ/mol);

R=8.315J/(mol·K), gas constant;

T: temperature (K);

t: Reaction time (s).

Multiple linear regression models were used to analyze the pyrolysis produ ct yields catalyzed by different mixing ratios of DFA and biomass, and the ana lysis results were shown in Table 8. In this case, the R-square values were all greater than 0.75, indicating that the model had a good fit and could be used as a reaction mechanism model for desulfurization ash catalysis.

Table 8 Three model summaries and parameter estimates

Independent variable: mixing ratio									
Dependent variable		Model summary					Parameter estimates		
		R square	F	Degree of freedom 1	Degrees of freedom 2	significance	constant	b1	b2
CE		0.988	113.774	3	4	0.000	26.635	87.414	-138.746
		0.993	201.185	3	4	0.000	37.519	-120.313	217.487
		0.778	4.682	3	4	0.085	18.859	36.950	-69.679
		0.879	9.722	3	4	0.026	16.987	-4.050	-9.061
LG	Tar yield	0.992	171.042	3	4	0.000	24.184	-72.184	114.796
	Water yield	0.995	250.451	3	4	0.000	10.862	52.992	-86.350
	Char yield	0.912	13.861	3	4	0.014	58.938	-9.906	15.949
	Syngas yield	0.996	303.565	3	4	0.000	6.016	29.098	-44.395

Figure 9 shows the fitting diagram of the multiple linear regression model under the pyrolysis combination of DFA/CE:

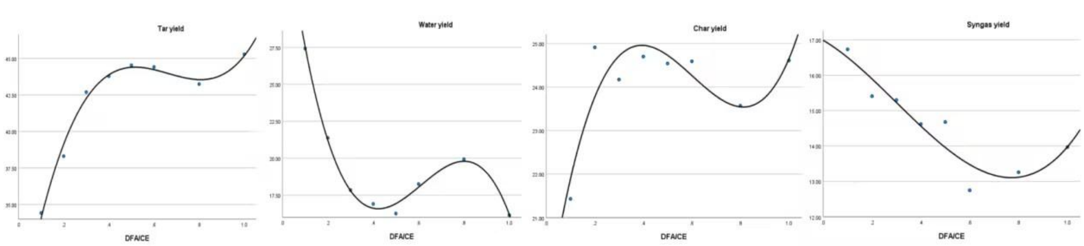


Figure 9 Multiple linear fitting of CE

Figure 10 shows the fitting diagram of multiple linear regression model un der the pyrolysis combination of DFA/LG.

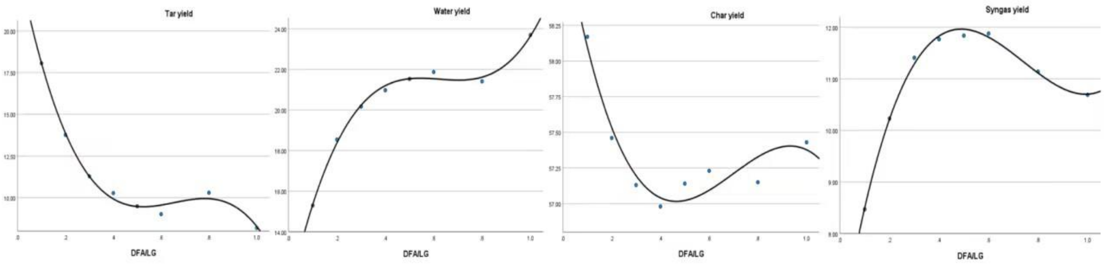


Figure 10 Multiple linear fitting of LG

The regression formula for the yield of four pyrolysis products in the DFA/CE combination under the multiple linear model is as follows:

$$y|\text{Tar yield} = 26.635 + 87.414 * x + (-138.746 * x * x) + 69.878 * x * x * x$$

$$y|\text{Water yield} = 37.519 + (-120.313 * x) + 217.487 * x * x + (-118.497 * x * x * x)$$

$$y|\text{Char yield} = 18.859 + 36.950 * x + (-69.679 * x * x) + 38.517 * x * x * x$$

$$y|\text{Syngas yield} = 16.987 + (-4.050 * x) + (-9.061 * x * x) + 10.102 * x * x * x$$

The regression formula for the yield of four pyrolysis products in the DFA/LG combination under the multiple linear model is as follows:

$$y|\text{Tar yield} = 24.184 + (-72.184 * x) + 114.796 * x * x + (-58.518 * x * x * x)$$

$$y|\text{Water yield} = 10.862 + 52.992 * x + (-86.350 * x * x) + 46.138 * x * x * x$$

$$y|\text{Char yield} = 58.938 + (-9.906 * x) + 15.949 * x * x + (-7.604 * x * x * x)$$

$$y|\text{Syngas yield} = 6.0158 + 29.098 * x + (-44.395 * x * x) + 19.984 * x * x * x$$

Multiple linear regression models were used to analyze pyrolysis gas yields catalyzed by different mixing ratios of DFA and biomass, and the analysis results were shown in Table 9. In this case, R-square values were all greater than 0.95, indicating a good fit of the model, which could be used as a reaction mechanism model for desulfurization ash catalysis.

Table 9 Three model summaries and parameter estimates

Independent variable: mixing ratio									
Dependent variable		Model summary					Parameter estimates		
		R square	F	Degree of freedom 1	Degrees of freedom 2	significance	constant	b1	b2
CE	H2	1.000	16464.257	3	1	0.006	0.690	137.672	-30.198
	CO	0.983	19.673	3	1	0.164	19.510	-34.265	30.572
	CO2	0.996	87.100	3	1	0.079	35.462	-104.732	115.798
	CH4	0.986	22.997	3	1	0.152	6.524	-22.311	32.234
	C2H6	0.969	10.397	3	1	0.223	24.324	-62.920	59.419
LG	H2	0.965	9.181	3	1	0.237	0.610	0.232	0.363
	CO	0.982	17.707	3	1	0.173	14.680	0.757	8.891
	CO2	0.996	92.796	3	1	0.076	-3.578	187.822	-262.878
	CH4	0.998	138.207	3	1	.062	11.378	114.558	-183.711
	C2H6	0.580	0.460	3	1	0.763	0.106	-0.225	0.491

Under the multiple linear model, the regression formula of five pyrolysis gas yields in the DFA/CE combination is as follows:

$$y|H2 = 0.690 + 137.671x + (-30.198 * x * x + (-50.255 * x * x * x)$$

$$y|CO = 19.510 + (-34.265 * x) + 30.572 * x * x + (-10.459 * x * x * x)$$

$$y|CO2 = 35.462 + (-104.732 * x) + 115.798 * x * x + (-44.473 * x * x * x)$$
$$y|CH4 = 6.524 + (-22.311 * x) + 32.234 * x * x + (-15.731 * x * x * x)$$
$$y|C2H6 = 24.324 + -62.920 * x + 59.419 * x * x + (-17.517 * x * x * x)$$

Under the multiple linear model, the regression formula of five pyrolysis gas yields in the DFA/LG combination is as follows:

$$y|H2 = 0.610 + 0.232 * x + 0.363 * x * x + (-0.417 * x * x * x)$$
$$y|CO = 14.680 + 0.757 * x + 8.891 * x * x + (-3.435 * x * x * x)$$
$$y|CO2 = -3.578 + 187.822 * x + -262.878 * x * x + 122.253 * x * x * x$$
$$y|CH4 = 11.378 + 114.558 * x + (-183.711 * x * x) + 87.628 * x * x * x$$
$$y|C2H6 = 0.106 + (-0.225 * x) + 0.491 * x * x + (-0.323 * x * x * x)$$

5.5 The establishment and solution of problem 5

5.5.1 Grey prediction

The brief principle of A prediction model is to first use the accumulation technology to make the data have exponential rules, then establish a first-order differential equation and solve it, and then reduce and reduce the results, that is, the gray predicted value, so as to predict the future. Step 1: Before establi shing the gray prediction model, the feasibility of the modeling method must b e guaranteed, that is, the known original data need to be tested by level ratio. Let the initial non-negative data sequence be：

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(k))$$
$$x^{(r)} = (x^{(r)}(1), x^{(r)}(2), \dots, x^{(r)}(k))$$

Then the summation production is:

$$= x^{(r-1)}(1) + x^{(r-1)}(2) + \dots + x^{(r-1)}(k) = \sum_{i=1}^{k} x^{(r-1)}(i)$$
$$= x^{(r)}(k-1) + x^{(r-1)}(k)$$

For $x^{(r)}$, its inverse accumulation generates the formula:

$$a^{(1)}(x^{(r)}(k)) = a^{(0)}((x^{(r)}(k))) - a^{(0)}(x^{(r)}(k-1)) = x^{(r)}(k) - x^{(r)}(k-1) = x^{(r-1)}(k)$$

Perform first-order cumulative generation on sequence $x^{(0)}$, generate sequen ce,

namely$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i)(k = 1,2,\dots,10)$,

It is obtained that the prediction model of sequence $x^{(1)}$,$\mathbf{GM(1，1)}$is a dy namic model of gray differential equation with first order single variable:

$$x^{(0)}(k) + az^{(1)}(k) = b \quad (k = 1,2,\dots,10)$$

Where $z^{(1)}(k)$ is the generation of the adjacent mean of $x^{(1)}(k)$, and the change of the new sequence is $\frac{dx^{(1)}}{dt} + ax^{(1)} = B$

Apply the least square method to $x_i^{(1)}$ to determine the model parameters, and obtain:

$$\hat{a} = (a, b)^T = (B^T B)^{-1} \cdot B^T \cdot Y_K,$$

$$\text{Among } B=\begin{bmatrix} -\frac{1}{2}\ (x^{(1)}(1)+x^{(1)}(2), & 1 \\ -\frac{1}{2}\ (x^{(1)}(2)+x^{(1)}(3), & 1 \\ \cdots\cdots & \cdots\cdots \\ -\frac{1}{2}\ (x^{(1)}(k-1)+x^{(1)}(k), & 1 \end{bmatrix}$$

$$Y_k=\left[x^{(0)}(2),x^{(0)}(3),\ldots,x^{(0)}(k)\right]^T$$

5.5.2 Analysis of model

Because the sample size of pyrolysis product yield is small, we choose the grey prediction model to predict it.

Under the gray prediction model, the fitting data of the gray prediction model of the DFA/CS combination are shown in Table 10, wherein the stage ratios of the yields of four pyrolysis products are all within the interval (0.819, 1.221)), indicating that the data is suitable for the construction of the gray prediction model, and the posterior difference ratio is less than 0.35, indicating that the model has high accuracy. The average relative error is also less than 2.5%, and the model has a good fitting effect. In summary, the model has a good prediction effect on the pyrolysis product yield of DFA/CS combination. The stage ratios of the yields of the nine pyrolysis gases are all within the interval (0.751, 1.331), indicating that the gas yield data are suitable for the construction of gray prediction models, and the yield posterior difference ratios of H2, CO, CO2, CH4, C2H6, C3H8 and C4H10 are all less than 0.35, indicating high accuracy of the models. The posterior difference ratio of C3H6 and C2H4 yields is 0.519 and 0.5, respectively. The accuracy of the model is basically qualified, and the average relative error is less than 8%. The model has a good fitting effect and can be used for later prediction. The predicted results of pyrolysis product yield and pyrolysis gas yield under DFA/CS combination in the next seven periods are shown in Table 10.

Table 10 Grey prediction model fitting data and prediction results summary table

Type	Name	Stage ratio	A posteriori difference over C value	Mean relative error	Backward prediction phase 1	Backward prediction phase 2	Backward prediction phase 3	Backward prediction phase 4	Backward prediction phase 5	Backward prediction phase 6	Backward prediction phase 7
Decom position	Tar yield	(0.819，1.221)	0.038	2.447%	11.15	10.627	10.128	9.653	9.2	8.768	8.356
	Water yield	(0.819，1.221)	0.027	0.646%	31.522	32.075	32.638	33.21	33.793	34.386	34.989
	Char yield	(0.819，1.221)	0.224	0.298%	29.81	29.902	29.995	30.087	30.18	30.274	30.367
	Syngas yield	(0.819，1.221)	0.376	2.214%	27.742	27.797	27.852	27.908	27.963	28.019	28.075
Gas	H2	(0.751，1.331)	0.013	5.822%	54.922	65.626	77.488	90.633	105.201	121.345	139.236

group	CO	（0.751，1.331）	0.112	2.384%	23.225	23.769	24.324	24.893	25.476	26.072	26.682
	CO2	（0.751，1.331）	0.124	1.867%	48.569	47.219	45.907	44.631	43.39	42.184	41.012
	CH4	（0.751，1.331）	0.197	3.719%	31.241	31.295	31.35	31.404	31.459	31.514	31.569
	C2H6	（0.751，1.331）	0.086	2.509%	3.029	3.059	3.09	3.121	3.152	3.184	3.216
	C3H8	（0.751，1.331）	0.08	3.862%	1.001	1.113	1.237	1.374	1.528	1.698	1.887
	C3H6	（0.751，1.331）	0.519	2.587%	0.378	0.376	0.374	0.372	0.37	0.369	0.367
	C2H4	（0.751，1.331）	0.5	2.334%	0.766	0.764	0.762	0.76	0.758	0.757	0.755
	C4H10	（0.751，1.331）	0.204	7.914%	0.181	0.121	0.063	0.008	-0.044	-0.093	-0.139

Testing of Lasso model:

The predicted data based on the gray prediction of decomposition products are brought into the Lasso formula proposed above as independent variables. The standardized formula of the mixed ratio model is y=-5.039-0.013 × Tar yield+0.196 × Water yield and the non-standardized formula of the model is as follows: y=-10.544-0.011 × Tar yield+0.167 × Water yield+0.215 × Char yield. Similarly, based on the predicted data of the gaseous group, the standardized formula of the mixture ratio model is: y=-0.398+0.029 × H2+0.051 × CO-0.003 × CO2-0.004 × CH4+0.17 × C2H6+0.36 × C3H8-2.257 × C3H6-1.448 × C2H4+1.281 × C4H10. Non-standardized formula of the model: y=-0.39+0.029 × H2+0.051 × CO-0.002 × CO2-0.004 × CH4+0.169 × C2H6+0.356 × C3H8-2.233 × C3H6-1.461 × C2H4+1.276 × C4H10.

Table 11 Lasso model test results

Type	Model formula	Backward prediction phase 1	Backward prediction phase 2	Backward prediction phase 3	Backward prediction phase 4	Backward prediction phase 5
Decomposition product	Standardization	0.994362	1.109549	1.226384	1.344671	1.464828
	Non-standardization	1.006674	1.124558	1.244063	1.364592	1.486931
Gas group	Standardization	1.253379	1.571343	1.930384	2.335765	2.849345
	Non-standardization	1.301124	1.617538	1.975008	2.378787	2.890497

The mixing ratio data corresponding to the following five phases can be obtained, and the mixing ratio of the topic ends at 100/100 (that is, 1). It can be seen that the mixing ratio data of the latter is more reasonable. Therefore, the gray prediction model has good applicability, which can also verify the rationality of the previous prediction of pyrolysis products and pyrolysis gas.

6.Strengths and Weakness

(1) The advantage of Pearson correlation coefficient model is that it is fast and easy to calculate, and it can calculate correlation coefficient and p-value at the same time. In addition, the Pearson correlation coefficient is suitable for

the analysis of linear correlation and can measure the strength and direction of the linear relationship between two variables[12]. However, the disadvantage of Pearson correlation coefficient is that it is more demanding on the data, needs to meet the premise of linear correlation, and is sensitive to outliers and outliers.

(2) Lasso model has strong selection ability, can automatically select important features, improve the model's interpretation ability and generalization ability, can deal with multicollinearity problems, easier to explain and understand. The disadvantage is that it is more sensitive to outliers, because it imposes constraints on all parameters, including the parameter where the outlier resides. In the case of very high feature dimension, lasso regression algorithm may choose too many features, and the regularization parameter λ needs to be carefully selected. When there is a strong correlation between features, the lasso regression algorithm tends to select one of the features while ignoring the other relevant features.

(3) Independent sample T test may lead to legacy effects or a small amount of noise. Independent T-tests do not help control for the effects of the environment, and changes in the environment may affect the output of the T-test. The T-test cannot be used for multiple comparisons because it causes type I errors.

(4) The advantage of linear regression is that the modeling speed is fast, it does not require very complex calculation, and it still runs very fast in the case of a large amount of data, and the understanding and interpretation of each variable can be given according to the coefficient. The disadvantage is that it can not fit nonlinear data well[13].

(5) Grey prediction model has wide application range, remarkable effect, high accuracy and simple calculation. However, grey prediction method also has some shortcomings: strict data requirements, lack of theoretical basis, lack of unified mathematical framework support.

References

[1] Hu J H. Study on pyrolysis/gasification characteristics and interaction between coal and biomass [D]. Huazhong University of Science and Technology,2018.
[2] Gao Y. Study on mechanism of hydrothermal degradation and product formation of biomass [D]. Huazhong University of Science and Technology,2013.
[3] ZHANG L Q. Experimental study on hierarchical pyrolysis of lignocellulosic biomass [D]. University of Science and Technology of China,2019. (in Chinese)
[4] Du Qi, Tang Chuyang, An Xinyuan et al. Study on catalytic extraction of biomass pyrolysis gas from desulfurization ash based on model compounds [J]. Renewable Energy, 2019,41(06):
[5] LIU Xuanzuo, Yao Zonglu, Zhao Lixin et al. Construction of biomass pyrolysis reaction rate model improved by integral method [J]. Transactions of the Chinese Society of Agricultural Engineering,2019,35(23):
[6] GAO R. Study on gasification process of two-stage gas bed coupled with pyrolysis and gasification [D]. East China University of Science and Technology,2020. (in Chinese)
[7] SHEN L. Study on kinetics and model of lignin pyrolysis reaction [D]. Xiamen University,2021. (in Chinese)
[8] He Fei, Wang Tao, Zhang Xue-Bing, et al. Study on reduction kinetics and reaction Model of molten Iron Catalyst H2-TPR [J]. Low Carbonization Science and Chemical Industry,
[9] Li Sen, Jing Dongfeng. Distributed activation energy model of biomass pyrolysis [J]. Agricultural Mechanization Research,2006,(08):
[10] He Fang, Yi Weiming, Sun Rongfeng et al. Analysis of pyrolysis reaction and pyrolysis kinetics of wheat and corn straw [J]. Transactions of the Chinese Society of Agricultural Engineering,2002,(04):
[11] SONG H. Establishment of molecular model of corn stalk and study of pyrolysis reaction kinetics [D]. Northeast Dianli University,2020.
[12] XU L. Experimental and theoretical research on model parameters of solid fuel pyrolysis [D]. University of Science and Technology of China,2017. (in Chinese)
[13] LI Xiaomin. Research on high temperature air combustion of straw and biomass [D]. University of Science and Technology of China,2012.