

Correlation Analysis of Pyrolysis Yield Using a Linear Regression Model

Yi Wang ^{1,#}, Sisi Fang ^{1,#}, Bo Xu ^{2,#}, Boqian Li ^{3,#}

¹ Nanchang Institute of Technology, School of Electronics and Information, Nanchang, 330044, China

² East China Jiaotong University, School of Vocational Education, Nanchang, 330013, China

³ Jiangxi University of Technology, School of Artificial Intelligence, Nanchang, 330098, China

#These authors contributed equally

Abstract: The paper focuses on the impact of pyrolysis combination and mixture ratio on the yield of various pyrolysis products, such as tar, water, coke residue, and syngas, in the field of catalytic reaction analysis for pyrolysis product generation. Initially, data preprocessing was carried out, outlier detection and missing value processing were respectively performed, and a model was established to explain the relationship between mixing ratio and yield. Descriptive statistical analysis was employed to gain an initial understanding of the overall data situation. Subsequently, statistical indices such as mean, range, and standard deviation of pyrolysis products for each combination were quantified to unveil the extent of influence of different mixing ratios on yield. The correlation analysis and linear regression model were then used to establish the relationship model between mixture ratio and yield, with further explanation of the correlation and presentation of the functional expression. Finally, mathematical formulas and graphs were utilized to analyze the linear trend between product and mixing ratio under various pyrolysis combinations. This study offers robust data analysis and modeling support for comprehending the impact of different mixing ratios on pyrolysis product yield.

Keywords: Linear Regression Model; Pyrolysis Yield; Relevancy.

1. Introduction

With the gradual improvement of human life quality, the search for renewable energy materials has become the focus of the current world. Xinjiang is a region with abundant cotton stalk production, and its cotton planting area accounts for more than 80% of the country's total planting area. [1] However, as a renewable energy source, cellulose, lignin and other biomass contained in cotton stalk have attracted much attention, and biomass as an energy source has gradually become a trend.[2]Sun Zhiao[3] found that cotton stalk had better comprehensive combustion performance and lower ignition temperature and burnout temperature. Liu Simeng[4] found that hydrothermal oxidation treatment could produce higher fixed carbon content after pyrolysis of cotton stalk. Zhao Jiaying[5]Zhao Jiaying found in his research on the quality improvement of desulfurization ash and biomass pyrolysis products that desulfurization ash promoted the formation of small and medium molecular compounds during pyrolysis, thus increasing the yield of pyrolysis water and pyrolysis gas, while significantly decreasing the yield of pyrolysis oil. In order to further step into the field of the relationship between pyrolysis combination and pyrolysis products, this paper summarizes the influence of changing the mixture ratio of pyrolysis combination on the yield of various pyrolysis products such as tar, water, coke residue and syngas based on the research experience of the former. According to the data analysis provided by <https://www.nmmcm.org.cn/index/>, Firstly, descriptive statistical analysis was carried out on the average and standard difference data of each pyrolysis combination, and scatter plot was drawn to observe the influence trend of different mixing ratios of pyrolysis combinations on pyrolysis products and whether there was any significant influence. Then correlation analysis

and linear regression model were established, under the condition that different mixing ratios were taken as independent variables and the yield of pyrolysis products was taken as dependent variables. The relationship between the mixture ratio and the yield of pyrolysis products and the yield of pyrolysis products under different mixture ratio were quantitatively analyzed, and the corresponding influencing factors and rules were obtained.

2. Establishment and Solution based on Linear Regression Model

2.1. Data Pre-Processing

According to the data, first of all, the data are preliminarily preprocessed to check whether there are abnormal values or missing values in the data respectively. It can be clearly known from the three tables in the data that all provided are the yields of decomposition products of the pyrolysis combination, and the unit is 100%. Therefore, it can be seen that the sum of products under different mixing ratios should be 100.As can be seen from Table 1, there are no missing values in the sample data.

Based on observation, outliers are selected, that is, data that is different from the sample data. Two decimal places are retained for all data, and some data are integers or only one decimal place. Outliers account for a small proportion of the total sample data, so in order to maintain data consistency, they are selected to change the outliers. The specific data after the change is shown in the data marked in red in Table 2 and 3.

Table 1. Total yield of some products

Yield of Decomposition Products from DFA/CS Pyrolysis wt.%(daf)								
DFA/CS	0/100	20/100	30/100	40/100	50/100	60/100	80/100	100/100
Tar yield	19.46	15.43	14.14	13.89	13.21	12.84	12.57	12.13
Water yield	26.84	28.11	28.23	28.62	29.01	30.07	30.68	31.02
Char yield	29.21	29.29	29.34	29.14	29.33	29.47	29.64	29.87
Syngas yield	24.49	27.17	28.29	28.35	28.45	27.62	27.11	26.98
total output	100	100	14.14	13.89	13.21	12.84	12.57	12.13

Table 2. Yield of Decomposition Products from DA/CS Pyrolysis wt.%(daf)

Yield of DA/CS pyrolysis decomposition products wt.%(daf)								
DFA/CS	0/100	20/100	30/100	40/100	50/100	60/100	80/100	100/100
Tar yield	19.46	15.43	14.14	13.89	13.21	12.84	12.57	12.13
Water yield	26.84	28.11	28.23	28.62	29.01	30.07	30.68	31.02
Char yield	29.21	29.29	29.34	29.14	29.33	29.47	29.64	29.87
Syngas yield	24.49	27.17	28.29	28.35	28.45	27.62	27.11	26.98

Table 3. Yield of Gaseous Components from DFA/LG Pyrolysis(mL/g,daf)

Gas component yield from DFA/LG pyrolysis (mL/g, daf)					
DFA/LG	H2	CO	CO2	CH4	C2H6
20/100	0.67	15.23	24.35	27.67	0.08
40/100	0.72	15.76	37.91	33.25	0.06
50/100	0.78	17.31	39.26	33.86	0.09
80/100	0.81	19.08	41.23	30.26	0.07
100/100	0.79	20.94	43.56	29.87	0.05

After the above outliers and missing values are determined and changed, descriptive statistics are carried out on the three groups of data to understand the general data distribution.

2.2. Model Establishment

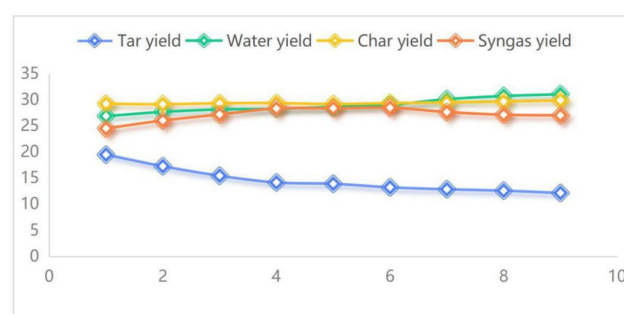
Table 4. Corresponds to the mean value, range and standard deviation of pyrolysis products

Pyrolytic combination	products	average value	Spannweite	standard deviation
DFA/CS	Coke yield	29.37889	0.76	0.246396
DFA/CE	Tar production	42.0825	10.86	3.76388
DFA/LG	water yield	20.435	8.39	2.53883

The mean value, range and standard deviation of corresponding pyrolysis products were calculated respectively, and some of the results were shown in Table 4. By calculating the average value of pyrolysis products of each combination, the average yield level can be roughly understood. If the range of pyrolysis products of each group is large, such as the range value of tar production in the table is 7.33, 10.86 and 9.87 respectively, which are all too large, it indicates that tar is significantly affected under different mixing ratios of pyrolysis combinations. Then, the dispersion degree of a data set is reflected according to the standard deviation data. In the yield table of DFA/CE pyrolysis decomposition products, the standard deviation values of water production and tar production are too large, indicating that the dispersion degree is poor and the yield deviation under different mixing ratios is large, while in the DA/CS pyrolysis decomposition products table, the standard

deviation of coke production is only 0.246. It shows that the product is basically unaffected by different mixing ratios.

Through descriptive statistics of the data, the products susceptible to different mixing ratios can be roughly understood. In order to more intuitively state their relationships, scatter plots of different pyrolysis combinations are drawn, as shown in FIG. 1, 2a and 2b.

**Figure 1.** Yield of Decomposition Products from DFA/CS Pyrolysis

It can be seen from FIG. 1 that desulfurization ash has a significant effect on tar yield, while other pyrolysis products all show a slight upward trend, and the overall stability is relatively stable, indicating that desulfurization ash plays an insignificant role in the pyrolysis reaction of cotton stalk. Furthermore, desulfurization ash/cellulose pyrolysis plays a relatively significant role in promoting the pyrolysis reaction, and the yield of decomposition products can be intuitively obtained. The tar yield increased with the increase of mixing ratio, indicating that the water yield of desulfurization ash in cellulose was slightly affected by the different mixing ratio.

The pyrolysis reaction of desulfurization ash/lignin showed that the output of tar, water, coke and syngas were not affected by different mixing ratios. The above has preliminarily analyzed whether desulfurization ash promotes the pyrolysis of cotton stalks, cellulose, and lignin, further establish correlation analysis, using Pearson correlation coefficient to illustrate the mixing ratio relationship the specific modeling



Figure 2a. Yield of DFA/CE pyrolysis decomposition products

process is as follows:

In numbers that are not all zero $\lambda_1, \lambda_2, \dots, \lambda_k$ make:

$$\lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_k X_k + \mu = 0 \quad (1)$$

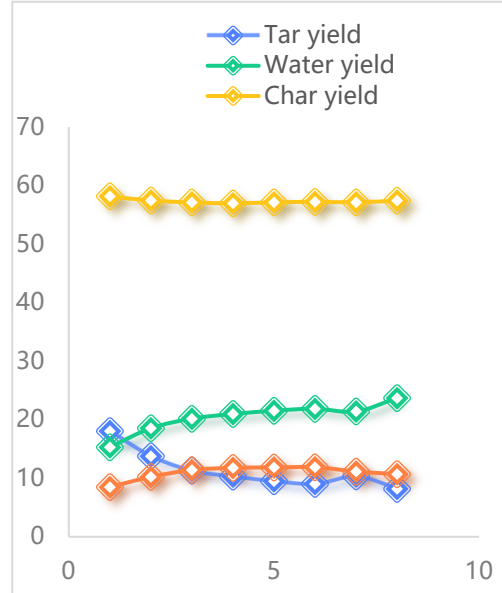


Figure 2b. Yield of DFA/LG pyrolysis decomposition products

established, of which μ is a random error term. When $\mu=0$, there is complete collinearity between explanatory variables, otherwise, it is called incomplete collinearity. According to the size of the semi partial correlation coefficient, explanatory variables that have no significant impact on the dependent variable can be selected. Consider the following two linear regression models

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (2)$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{i-1} X_{i-1} + \beta_{i+1} X_{i+1} + \dots + \beta_i X_i + \mu \quad (3)$$

The judgment coefficients are R^2 and R respectively. R^2 , the relationship between the two is:

$$sr_{1Y}^2 = R^2 - R_1^2 \quad (4)$$

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E((X-\mu_X)(Y-\mu_Y))}{\sigma_X \sigma_Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)} \sqrt{E(Y^2) - E^2(Y)}} \quad (5)$$

The correlation analysis model can be obtained from the above formula. Since the data we are studying is a continuous variable, we use Pearson correlation coefficient to complete the correlation test, due to the fluctuation of X and Y, the covariance value cannot fully demonstrate the correlation between the two variables. Therefore, we need to standardize the covariance to obtain the Pearson correlation coefficient. The specific steps are as follows:

$$\rho_{X,Y} = \frac{N \sum XY - \sum X \sum Y}{\sqrt{N \sum X^2 - (\sum X)^2} \sqrt{N \sum Y^2 - (\sum Y)^2}} \quad (6)$$

In addition, in order to obtain the relationship between the yield of pyrolysis products and the mixture ratio of pyrolysis combinations more accurately, a linear regression model was constructed to solve the function expression of the yield of

pyrolysis products under each pyrolysis combination, the Pearson correlation coefficients are 0.969, 0.979, 0.905, 0.949, 0.968, 0.968 and 0.959 respectively. The lowest is more than 0.9, so as to set the following multiple linear regression model:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m + \varepsilon \quad (7)$$

Where the independent variables x_1, x_2, \dots, x_n is the predictor, y is the dependent variable, ε is the random variable, and β is the parameter to be estimated. β can be calculated by least square method. The calculation formula is as follows:

$$Q(\beta) = \sum_{i=1}^n (\delta_i^2 (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_m x_{im}))^2 = (Y - X\beta)^T (Y - X\beta) \quad (8)$$

Finding β minimize $Q(\beta)$ gives a least squares estimate of β

Linear regression equations are usually verified by model fitting degree R^2 , analysis of variance (F test) and T test. R is the correlation coefficient between the predicted value and the observed value, which can represent the interpretation degree of the model to the observed data, and it is generally required to reach more than 0.85. The expression for R is as follows:

$$R = \frac{\sum (y_i - Y)(Y_i - Y)}{\sqrt{(\sum (y_i - Y)^2)(\sum (Y_i - Y)^2)}} \quad (9)$$

2.3. Model Solving and Analysis

Pearson correlation coefficient was established to represent the relationship between the yield of pyrolysis products (tar, water, coke slag, syngas) and the mixing ratio of

corresponding pyrolysis combinations, and the correlation diagram between desulfurization ash/cotton stalk, desulfurization ash/cellulose, desulfurization ash/lignin pyrolysis combinations and their pyrolysis products under different mixing ratios was made respectively. Figure 3 below takes desulfurization ash/cotton stalk pyrolysis combination as an example.

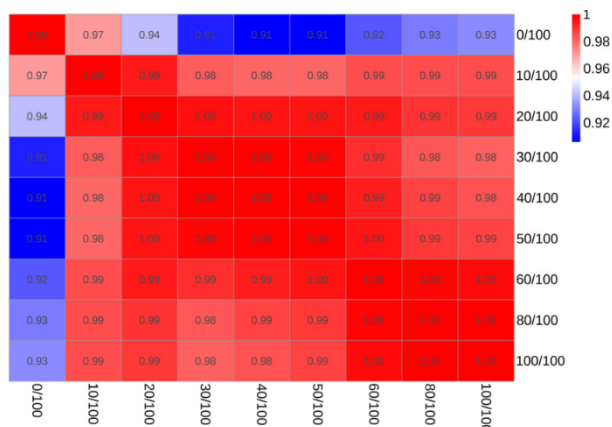


Figure 3. Desulfurization ash/cotton stalk (DFA/CS) correlation diagram

It can be intuitively seen from the above figure that, except for the poor correlation between the pyrolysis product tar and the mixture ratio of the pyrolysis combination, the overall correlation is high, and the Pearson coefficient obtained by solving it is 0.983. Then the corresponding function expression was solved by the linear regression model, and the linear diagram of the mixture ratio of pyrolysis products and pyrolysis combinations was obtained. Part of the linear diagram was shown in FIG. 4a and 4b. The specific solution expression is shown in Table 5 below.

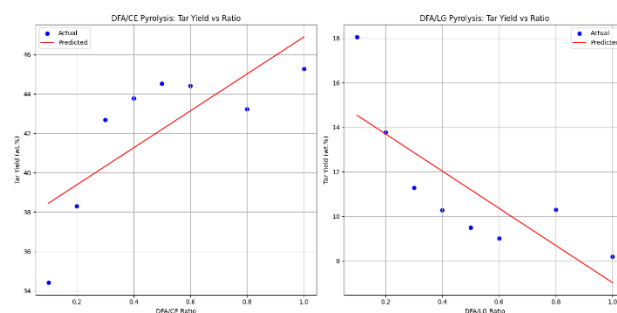


Figure 4a. Linear graph of tar production and different mixing ratios of desulfurization ash/cellulose

Figure 4b. Linear graph of tar yield and different mixing ratios of desulfurized ash/cotton stalk

Through linear regression analysis, we get the following results:

For DFA/CE (cellulose) pyrolysis products, the linear regression model coefficient of the relationship between tar yield and desulfurization fly ash mixture ratio is 9.38, and the intercept is 37.51. This indicates that the tar yield tends to increase with the increase of desulfurization fly ash mixture ratio.

For DFA/LG (lignin) pyrolysis products, the linear regression model coefficient of the relationship between tar yield and desulfurization fly ash mixture ratio is -8.36, and the intercept is 15.38. This indicates that the tar yield tends to decrease with the increase of desulfurization fly ash mixture ratio.

In both graphs, we can see the actual data points (blue) and the lines predicted by the model (red). In the case of DFA/CE, the model predicts an increase in tar yield as the proportion of desulfurized fly ash increases. In the case of DFA/LG, the model predicts that the tar yield will decrease as the proportion of desulfurized fly ash increases.

These results show that desulfurized fly ash has different catalytic effects on the pyrolysis of cellulose and lignin. In particular, desulfurized fly ash seems to promote the pyrolysis of cellulose to produce tar, while inhibiting the pyrolysis of lignin to produce tar.

Table 5. The function expression of each pyrolysis combination and pyrolysis product

Function name	Function expression	Tendency
Tar yield to DFA/CS mix ratio	$y = -0.8245x + 18.669$	Decline
Water yield to DFA/CS mix ratio	$y = 0.509x + 26.368$	Rise
Coke yield and DFA/CS mixture ratio	$y = 0.0763x + 28.996$	Rise
Synthesis gas production to DFA/CS mixture ratio	$y = 0.2372x + 25.981$	Rise and Fall
Tar production to DFA/CE mix ratio	$y = 1.2688x + 36.373$	Rise
Water yield to DFA/CE mixture ratio	$y = -1.0188x + 23.847$	Decline
Coke yield and DFA/CE mixture ratio	$y = 0.1983x + 23.173$	Plateau after rise
Synthesis gas production to DFA/CE mixture ratio	$y = -0.4483x + 16.608$	Down and up
Tar production to DFA/LG mix ratio	$y = -1.1195x + 16.338$	Dcline
Water yield to DFA/LG mix ratio	$y = 0.9374x + 16.217$	Rise
Coke yield and DFA/LG mixture ratio	$y = -0.0746x + 57.672$	Plateau rise
Synthesis gas production and DFA/LG mixture ratio	$y = 0.2568x + 9.7732$	Plateau after rise

The relationship between the yield of pyrolysis products and the mixing ratio of the corresponding pyrolysis combination can be clearly understood from Table 5, which further explains the results of correlation analysis and improves the uncertainty of correlation analysis. In the end, the tar yield in desulfurization ash/cotton stalk, tar yield in desulfurization ash/cellulose and water yield has a significant impact on the mixing ratio of pyrolysis combination.

3. Conclusion

"Carbon peaking" and "carbon neutrality" were first included in the government work report during the 2021 Two Sessions, The vigorous development of new energy and the search for more renewable energy have become important tasks for China to achieve sustainable development strategy

In this context, the impact of pyrolysis combination mixing ratio on the yield of various pyrolysis products is proposed. Using descriptive statistics such as mean and standard deviation combined with drawing scatter plots of pyrolysis products, we observed that the combination of desulfurization ash and cotton straw pyrolysis has a significant inhibitory effect on tar yield, The combination of desulfurization ash and cellulose pyrolysis has a significant promoting effect on tar. Next, conduct a deeper correlation analysis to obtain a correlation heatmap, It was found that there is a high correlation between tar production and changes in catalyst mixing ratio, the Pearson coefficient is 0.983. Further establish a linear regression model to calculate the functional expressions of each pyrolysis combination and pyrolysis product yield, Draw the conclusion that Water yield to DFA/CS mix ratio, Coke yield and DFA/CS mixture ratio, Tar production to DFA/CE mix ratio, Water yield to DFA/LG mix ratio the mixing ratio of the four pyrolysis combinations shows an upward trend, The tar yield in desulfurization ash/cotton straw, tar yield in desulfurization ash/cellulose, and water yield have a significant impact on the final mixing ratio with pyrolysis.

References

- [1] Yang Jifan, Zhang Shouyu, Cao Zhongyao, et al. The effect of hydrothermal oxidation pretreatment on the physicochemical properties of cotton stem shaped particles [J] Chemical Progress, 2021. DOI: 10.16085/j.issn.1000-6613.
- [2] Liu Jiajia, Huang Fujiang. Current Status and Problems of Cotton Production in Xinjiang and Countermeasures 46 [J] Cotton Science, 2022, 44 (5): 15-19.
- [3] Sun Zhiao, Jin Baosheng, Zhang Mingyao, Liu Renping, Zhang Yong, Research on Fluidization Characteristics of Cotton Stalks and Their Combustion Behavior in Thermogravimetric Analysis, School of Energy and Environment, Southeast University, Nanjing, Jiangsu 210096.
- [4] Liu Simeng, Zhang Shouyu, Sun Mengyuan, Xu Jiaqing, Hu Nan, Wu Yuxin, Zhou Yi, Yang Jifan, Lang Sen. The influence of three components of cotton straw on its pyrolysis behavior during hydrothermal/hydrothermal oxidation treatment. Renewable Energy Volume 41, Issue 10, October 2023. [A], 1671-5292 (2023) 10-1279-08.
- [5] Zhao Jiaxing, Experimental study on the improvement of biomass pyrolysis gas-phase products by desulfurization ash, Liaoning Metallurgical Design and Research Institute, Liaoning University of Science and Technology.
- [6] Pan Jiahua The target direction for addressing climate change during the 14th Five Year Plan period [J]. Yuejiang Academic Journal, 2020, 12 (1): 20-32120.