Problem Chosen	2023	Team Control Number
B	ShuWei Cup	2023091119429
	Summary Sheet	

Abstract

Developing sustainable sources of renewable energy has become imperative in light of increasing energy demands and environmental concerns. Pyrolysis of agricultural residues, such as cotton stalks which are abundantly generated as waste from cotton cultivation, possess high cellulose and lignin content, making them suitable for conversion into renewable biofuels and biochemicals. Hence, it offers a promising avenue for the production of biofuels and bioproducts (Chen et al., 2011). However, the optimization of cotton stalk pyrolysis necessitates a comprehensive scientific understanding of the underlying mechanisms, product distribution, and catalyst influence. Although cotton stalk pyrolysis has been extensively studied, many reaction pathways and mechanisms are still unclear, highlighting the need for further research to provide scientific insights into the development of sustainable and efficient pyrolysis processes (Yang et al., 2020).

To this end, we aim to examine the 5 given questions related to cotton stalk pyrolysis. First, using the data in Annex I, we show that the addition of desulfurization ash (DFA) is a critical catalyst that effectively enhances pyrolysis reactions. Second, by mapping the figures using data in Annex II, we find that the pyrolysis gases of lignin are mainly composed of [CO], [CO_2]and [CH_4] with lower contents of H_2 and [C_2H_6]. While DFA has the potential to improve pyrolysis gas flammability and calorific value, it also reduces CO_2 emissions. Besides, it not only endorses the gasification reaction of cellulose and increased [H_2] generation but also inhibits cellulose cracking and decreases [CO], [CO_2], and hydrocarbon generation. However, DFA has a minimal effect on the pyrolysis reaction of lignin. To test the third question, we use a variety of normality tests (such as the Kolmogorov-Smirnov and the Shapiro-Wilk) as well as a matched samples t-test and a Wilcoxon analysis. We find no significant differences in the yields of cotton stalks pyrolyzed with and without catalysts, as well as the yields of pyrolysis gas components. In the subsequent question, we discover that [CO_2], [C_2H_6], and [CO] have a significant effect on the reaction kinetics equation when it comes from DFA/CE Pyrolysis, whereas [H_2] and [CO] affect the kinetics equation of [DFA/LG] Pyrolysis. Based on papers by Lamba, R., Prins, R., Jakobsen, H.A. Zabala S et al., Hai et al., we developed a kinetics equation. We employed a Trilayered neural network to predict the value of Annex-I-Table-1 in question 5, and find a favorable fit with a high R-square value. However, our methods and data entail several sources of lack inconsistencies that may affect our numerical results.

Finally, we have attached a memo in which we briefly report our findings after reviewing three papers on the economic impact of bio-renewable energy from cotton stalks. We hope that our research outcomes can also serve as a catalyst for socioeconomic development.

Keywords: Data plotting analysis, Normality test, Matched samples t-test, Wilcoxon analysis of paired samples test, Correlation analysis, Regression, Stepwise regression, Trilayered neural network.

Content

1. Introduction

1.1 Background

Rapidly growing global demand for renewable energy has led to increased attention on biomass energy as a mature and sustainable energy source. Among various biomass resources, cotton stalks, an abundant agricultural waste, have emerged as a significant biomass feedstock due to their rich biomass components, such as cellulose and lignin. Pyrolysis, a thermochemical conversion process, can effectively convert cotton stalks into various forms of renewable energy. However, the quality and yield of pyrolysis products are influenced by several factors, including pyrolysis temperature and the use of catalysts. Therefore, comprehensive research into the mechanism and properties of cotton stalk pyrolysis products, as well as the impact of catalysts during pyrolysis, holds significant importance for the efficient utilization and sustainable development of cotton stalks.

In the context of this study, a specific chemical engineering laboratory adopted the model compound method to establish pyrolysis combinations, involving the utilization of desulfurization ash alongside cotton stalks and model compounds. By subjecting these pyrolysis combinations to pyrolysis at different mixing ratios, the catalytic mechanism and effects of desulfurization ash on the pyrolytic conversion of cotton stalks were investigated. During the selection of model compounds, careful consideration was given to factors such as controllability, reaction stability, and their relevance to cotton stalk pyrolysis. Consequently, CE (Cellooligosaccharide) and LG (Lignin), which serve as representative components of cellulose and lignin in cotton stalks respectively, were chosen as model compounds. This approach enables a more refined analysis of the targeted catalytic effects of desulfurization ash on different biomass components.

The experimental outcomes, presented in Appendix 1 and Appendix 2, were derived from fixed-bed pyrolysis experiments conducted across a range of mixing ratios, including 10/100, 20/100, 30/100, 40/100, 50/100, 60/100, 80/100, and 100/100. The selection of these mixing ratios was based on the requirement that, under these experimental conditions, the relative error among parallel experiments is approximately 5%. Choosing mixing ratios that are too small, such as 5/100 or 7/100, would introduce a significant relative error, adversely impacting the accuracy of the experimental results and impeding the exploration and optimization processes. To overcome this limitation and provide valuable insights for the distribution of pyrolysis products, it is envisioned that the integration of mathematical modeling and AI learning can enable the prediction of pyrolysis product yields and distributions using limited experimental data. Such a predictive approach holds immense potential to expedite experimental optimization while consistently guiding the understanding of

trends in pyrolysis product distributions.

1.2 Problem Restatement

i. Per the findings outlined in Appendix I, a thorough analysis was conducted on the yield of pyrolysis products (tar, water, coke residue, syngas) in relation to the mixing ratios of the corresponding pyrolysis combinations. The objective was to determine the significant role, if any, played by desulfurization ash when utilized as a catalyst in facilitating the pyrolysis of cotton stalks, cellulose, and lignin.

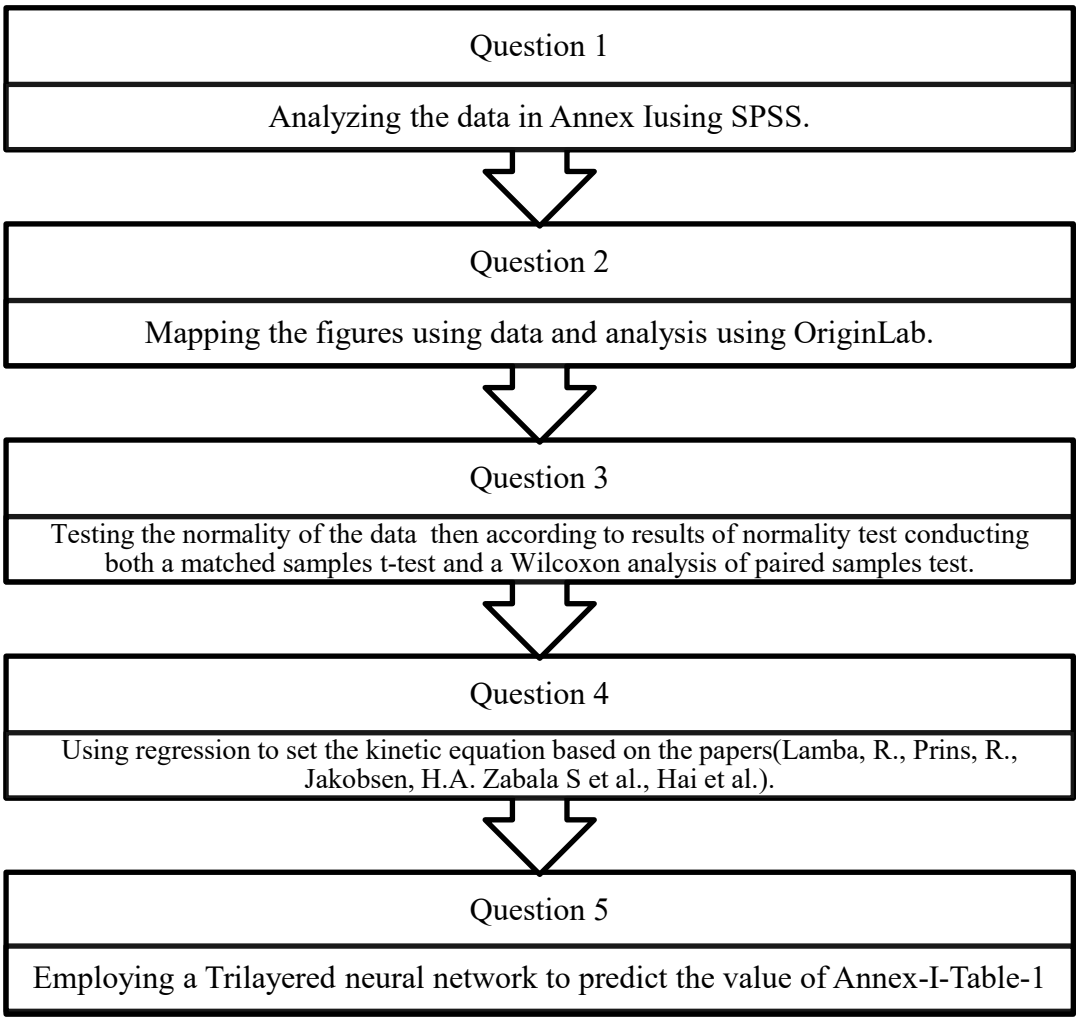
ii. In accordance with the information presented in Appendix II, a comprehensive examination was undertaken to assess the impact of mixing ratios in the pyrolysis combinations on the yield of each group of pyrolysis gas. To assist with understanding and clarification, corresponding images were generated and utilized for explanation.

iii. To ascertain whether there are notable differences in the yields of products generated from the pyrolysis of CE and LG, as well as the yields of the components of the pyrolysis gas, under the catalytic influence of an equal proportion of desulfurized ash, an in-depth investigation was conducted. The results found and reasons behind any observed dissimilarities are provided in the analysis.

iv. The establishment of a catalytic reaction mechanism model of desulfurized ash for model compounds such as CE and LG, alongside the modeling of reaction kinetics for analysis purposes, were rigorously pursued. These efforts aimed to offer insights into the underlying processes and mechanisms involved in the catalytic reactions.

v. Leveraging mathematical models and AI learning methods, predictions were generated on the yields or quantities of the pyrolysis products under conditions of limited data availability. By combining available data with predictive modeling techniques, valuable estimates were obtained to aid in understanding and forecasting pyrolysis product outcomes.

2. Problem analysis



2.1 Question 1

Based on the data in Annex I, for the first question, we used OriginLab for mapping, and based on the images we performed descriptive analysis

For each pyrolysis combination listed in Annex I, I analyzed the relationship between the yields of pyrolysis products (namely, tar, water, coke residue, and syngas) and the corresponding mixing ratios. In the DFA/CS combination, the tar yield showed a downward trend with the increase in DFA/CS, and the rate of decrease diminished over time. Conversely, the water yield increased as the DFA/CS ratio increased, while the char yield also increased but not to the same extent as the water yield. The syngas yield displayed an initial increase followed by a decrease with the increase of DFA/CS, and the inflection point was determined at DFA/CS=0.5.

As for the DFA/CE combination, the tar yield showed a significant increasing trend with the increase in DFA/CE, and this trend was more pronounced before DFA/CE=0.5. The water yield conversely decreased with the increasing DFA/CE ratio, but there was an increase observed between DFA/CE=0.6-0.8. Char yield

increased from DFA/CE=0.1 to 0.2, but then remained steady at around 24, despite minor fluctuations. Syngas yield, on the other hand, initially decreased with the increasing DFA/CE ratio from 0.1 to 0.2 and then increased from DFA/CE=0.1 to 0.2, exhibiting an overall downward trend.

Lastly, for the DFA/LG combination, the tar yield displayed a decreasing trend with the increasing DFA/LG ratio, albeit with a slight increase at DFA/LG=0.6-0.8. The water yield, on the other hand, increased as the DFA/LG ratio increased, while the char yield remained mostly constant. The syngas yield initially decreased from DFA/LG=0.1 to 0.2 and then exhibited an increase followed by another decrease. The inflection point was around DFA/LG=0.5.

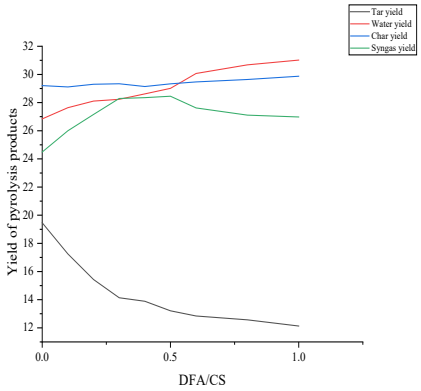


Figure 1 Yield of pyrolysis products--DFA/CS

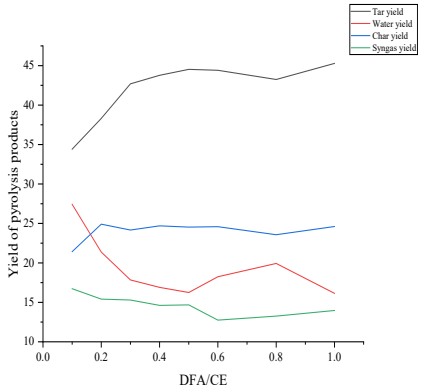


Figure 2 Yield of pyrolysis products--DFA/CE

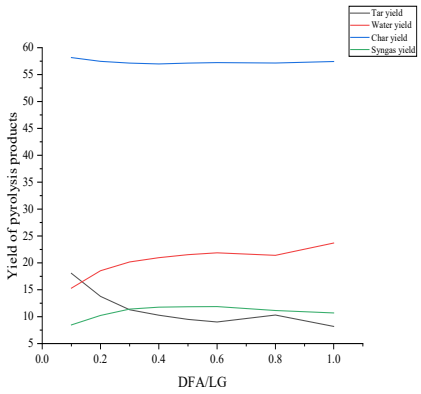


Figure 3 Yield of pyrolysis products--DFA/LG

Based on the data provided, desulfurization ash (DFA) plays a significant role as a catalyst in facilitating the pyrolysis of cotton stalks, cellulose, and lignin.

The key evidence supporting this is:

i. For cotton stalk pyrolysis, increasing the DFA ratio leads to decreased yields of tar and char, but increased yields of syngas and water. This demonstrates that DFA promotes gasification reactions that break down tar and char into syngas and water.

ii. For cellulose (CE) pyrolysis, adding DFA decreases tar and char yields while

increasing syngas yield. This shows DFA breaks down cellulose into syngas, though the effect is less pronounced compared to cotton stalks.

iii. For lignin (LG) pyrolysis, DFA addition decreases tar yield and slightly increases syngas yield. However, the catalytic effect is weaker compared to cellulose and cotton stalks.

iv. The trends of changing product yields with increasing DFA are consistent across the different biomass types. This confirms DFA's role as a catalyst in facilitating pyrolysis reactions.

v. The catalytic activity is strongest for heterogeneous, complex cotton stalks and weakest for the more homogeneous lignin.

Here we provide more details,

i. For DFA/CS pyrolysis, the tar yield decreases as the DFA content increases, indicating that DFA catalyzes the cracking of tar into syngas. The water yield increases slightly with the DFA content, suggesting that DFA promotes the dehydration of biomass. The char yield remains relatively stable, implying that DFA has little effect on the char formation. The syngas yield increases significantly with the DFA content, especially for the hydrogen content, showing that DFA enhances the gasification of biomass and the reforming of tar and water. Therefore, DFA plays a significant role in facilitating the pyrolysis of cotton stalks and improving the quality of pyrolysis gas.

ii. For DFA/CE pyrolysis, the tar yield increases sharply as the DFA content increases, indicating that DFA inhibits the cracking of tar into syngas. The water yield decreases drastically with the DFA content, suggesting that DFA suppresses the dehydration of cellulose. The char yield increases slightly with the DFA content, implying that DFA promotes the char formation. The syngas yield decreases slightly with the DFA content, especially for the hydrogen content, showing that DFA reduces the gasification of cellulose and the reforming of tar and water. Therefore, DFA plays a negative role in facilitating the pyrolysis of cellulose and degrading the quality of pyrolysis gas.

iii. For DFA/LG pyrolysis, the tar yield decreases significantly as the DFA content increases, indicating that DFA catalyzes the cracking of tar into syngas. The water yield increases moderately with the DFA content, suggesting that DFA promotes the dehydration of lignin. The char yield remains relatively stable, implying that DFA has little effect on the char formation. The syngas yield increases slightly with the DFA content, especially for the hydrogen content, showing that DFA enhances the gasification of lignin and the reforming of tar and water. Therefore, DFA plays a positive role in facilitating the pyrolysis of lignin and improving the quality of pyrolysis gas.

iv. Changes in pyrolysis product yields with DFA addition provide clear evidence that desulfurization ash serves as an effective catalyst to promote pyrolysis reactions, especially gasification reactions that generate syngas. DFA displays measurable catalytic activity in converting lignocellulosic biomass feedstocks like cotton stalks, cellulose, and lignin into useful syngas and other valuable

products.

2.2 Question 2

We use OriginLab to draw the figures, here are what we find,

i. DFA/CS pyrolysis combination: As can be seen from Table 1, with the increase of DFA content, the yields of H_2, CH_4, C_2H_6, C_3H_8, C_3H_6, C_2H_4 and C_4H_{10} all show an increasing trend, while the yields of CO and CO_2 show a decreasing trend. This indicates that the addition of DFA is conducive to improving the flammability and calorific value of pyrolysis gas, while reducing the emission of CO_2. Shown in Figure 4.

ii. DFA/CE pyrolysis combination: As can be seen from Table 2, with the increase of DFA content, the yield of H_2 significantly increases, while the yield of CO, CO_2, CH_4 and C_2H_6 significantly decreases. This indicates that the addition of DFA can promote the gasification reaction of cellulose and produce more H_2, while inhibit the cracking reaction of cellulose and produce less CO, CO_2 and hydrocarbons. Shown in Figure 4.

iii. DFA/LG pyrolysis combination: As can be seen from Table 3, with the increase of DFA content, the yields of H_2, CO and CO_2 slightly increase, while the yields of CH_4 and C_2H_6 slightly decrease. This indicates that the addition of DFA has little effect on the pyrolysis reaction of lignin. The pyrolysis gases of lignin are mainly composed of CO, CO_2 and CH_4, while the contents of H_2 and C_2H_6 are low. Shown in Figure 6.

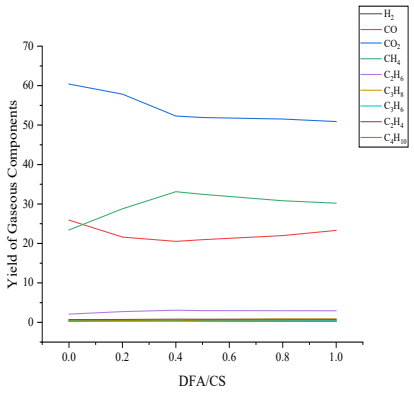


Figure 4 Yield of Gaseous Components--DFA/CS

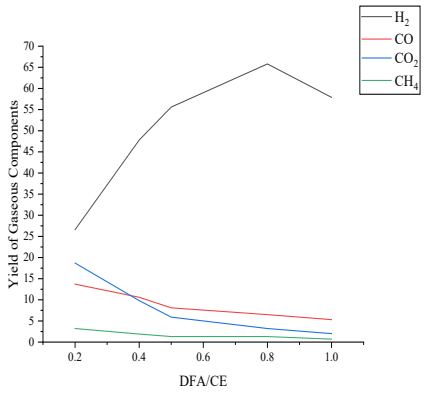


Figure 5 Yield of Gaseous Components--DFA/CE

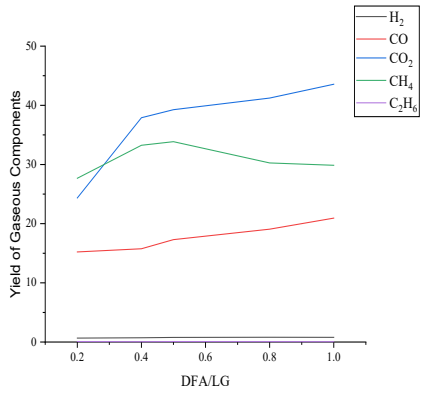


Figure 6 Yield of Gaseous Components--DFA/LG

2.3 Question 3

We first test the normality of the data in the four tables of Annex I and Annex II under the catalytic action of the same proportion of desulfurization ash. If both sets of data are normally distributed, we can use the T-test or analysis of variance (ANOVA). If the data does not follow a normal distribution, then non-parametric tests can be used, such as the Mann-Whitney U test or the Wilcoxon signed rank test. The results are as follows,

Table 1 Yield of Decomposition Products from DFA/CE Pyrolysis wt.%(daf)

Name	N	Mean	SD	Skewness	kurtosis	Kolmogorov Smirnov test p	Shapiro Wilk test p
R1	4	25.000	7.654	0.355	-1.115	0.000***	0.929
R2	4	25.000	9.701	1.043	1.563	0.000***	0.681
R3	4	25.000	12.369	1.499	2.081	0.000***	0.263
R4	4	25.000	13.243	1.428	1.705	0.000***	0.280
R5	4	25.000	13.720	1.466	1.805	0.000***	0.222
R6	4	25.000	13.815	1.306	1.741	0.000***	0.477
R7	4	25.000	12.888	1.345	2.250	0.000***	0.445
R8	4	25.000	14.278	1.444	1.748	0.000***	0.258
*p<0.05 **p<0.01 ***p<0.001							

The Kolmogorov-Smirnov normality test for all variables (R1, R2, R3, R4, R5, R6, R7, R8) results in a non-normal distribution, while the Shapiro-Wilk normality test results in a normal distribution. However, considering the actual situation of chemical production and the amount of data given is too small, we believe that these data are in line with normal distribution.

Table 2 Yield of Decomposition Products from DFA/LG Pyrolysis wt.%(daf)

Name	N	Mean	SD	Skewness	kurtosis	Kolmogorov Smirnov test p	Shapiro Wilk test p
R1	4	25.000	22.478	1.804	3.428	0.000***	0.086

R2	4	25.000	21.906	1.858	3.514	0.000***	0.056
R3	4	25.000	21.820	1.795	3.201	0.000***	0.048*
R4	4	25.000	21.838	1.737	2.971	0.000***	0.086
R5	4	25.000	22.051	1.688	2.784	0.000***	0.120
R6	4	25.000	22.182	1.656	2.665	0.000***	0.144
R7	4	25.000	22.020	1.707	2.833	0.000***	0.086
R8	4	25.000	22.663	1.514	2.017	0.000***	0.199
*p<0.05 **p<0.01 ***p<0.001							

The Kolmogorov-Smirnov normality test for all variables except R3 (R1, R2, R4, R5, R6, R7, R8) results in a non-normal distribution, while the Shapiro-Wilk normality test results in a normal distribution. R3 is non-normal distribution.

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

Name	N	Mean	SD	Skewness	kurtosis	Kolmogorov Smirnov test p	Shapiro Wilk test p
R1	5	15.200	8.521	-0.141	0.960	0.000***	0.857
R2	5	15.740	18.249	2.029	4.372	0.000***	0.016*
R3	5	15.060	22.797	2.170	4.770	0.000***	0.003**
R4	5	16.040	27.879	2.211	4.910	0.000***	0.001**
R5	5	13.820	24.700	2.210	4.906	0.001**	0.001**
*p<0.05 **p<0.01 ***p<0.001							

The Kolmogorov-Smirnov normality test for all variables except R1 (R2, R3, R4, R5) results in a non-normal distribution, R1 is a normal distribution.

Table 4 Yield of Gaseous Components from DFA/LG Pyrolysis (mL/g,daf)

Name	N	Mean	SD	Skewness	kurtosis	Kolmogorov Smirnov test p	Shapiro Wilk test p
R1	5	13.600	12.905	-0.117	-2.866	0.030*	0.260
R2	5	17.540	17.701	0.167	-2.825	0.030*	0.264
R3	5	18.259	18.173	0.119	-2.773	0.030*	0.292
R4	5	18.289	18.081	0.169	-2.140	0.030*	0.458
R5	5	19.041	18.814	0.193	-1.852	0.030*	0.481
*p<0.05 **p<0.01 ***p<0.001							

The Kolmogorov-Smirnov normality test for all variables (R1, R2, R3, R4, R5) results in a non-normal distribution, while the Shapiro-Wilk normality test results in a normal distribution. However, considering the actual situation of chemical production and the amount of data given is too small, we believe that these data are in line with normal distribution.

Considering that there are both data conforming to the normal distribution and data disconfirming to the normal distribution, we adopted matched samples t-test and

Wilcoxon analysis of paired samples.

Table 5 Data from Annex I- Pyrolysis Product Yields of Three Pyrolysis Combinations

Results of matched samples t- test							
Pairing number	Term	Mean	SD	Mean difference	t	p	Effect Value
Pair1	R1	25.000	7.654	0.000	0.000	1.000	0.000
	R1'	25.000	22.478				
Pair2	R2	25.000	9.701	0.000	0.000	1.000	0.000
	R2'	25.000	21.906				
Pair3	R3	25.000	12.369	0.000	0.000	1.000	0.000
	R3'	25.000	21.820				
Pair4	R4	25.000	13.243	0.000	0.000	1.000	0.000
	R4'	25.000	21.838				
Pair5	R5	25.000	13.720	0.000	0.000	1.000	0.000
	R5'	25.000	22.051				
Pair6	R6	25.000	13.815	0.000	0.000	1.000	0.000
	R6'	25.000	22.182				
Pair7	R7	25.000	12.888	0.000	0.000	1.000	0.000
	R7'	25.000	22.020				
Pair8	R8	25.000	14.278	0.000	0.000	1.000	0.000
	R8'	25.000	22.663				
*p<0.05 **p<0.01 ***p<0.001							

There was no significant difference between the two groups (p>0.05).

Table 6 Data from Annex I- Pyrolysis Product Yields of Three Pyrolysis Combinations

Results of Wilcoxon analysis of paired samples					
Name	Paired median (P25, P75)		Median M difference (pair 1 - pair 2)	W	p
	Pair 1	Pair 2			
R1 Pairing R1'	24.425(13.593,29.17)	16.68(13.593,28.087)	7.745	6.000	0.875
R2 Pairing R2'	23.14(12.885,28.26)	16.155(12.885,28.27)	6.985	6.000	0.875
R3 Pairing R3'	21.005(11.38,28.8)	15.79(11.38,29.41)	5.215	5.000	1.000
R4 Pairing R4'	20.8(11.397,29.47)	16.37(11.397,29.972)	4.43	5.000	1.000
R5 Pairing R5'	20.395(11.252,29.538)	16.685(11.252,30.433)	3.71	5.000	1.000
R6 Pairing R6'	21.42(11.165,29.545)	16.875(11.165,30.71)	4.545	5.000	1.000
R7 Pairing R7'	21.75(10.93,28.488)	16.275(10.93,30.345)	5.475	5.000	1.000
R8 Pairing R8'	20.375(10.065,29.777)	17.19(10.065,32.125)	3.185	5.000	1.000
*p<0.05 **p<0.01 ***p<0.001					

There was no significant difference between the two groups (p>0.05).

Table 7 Data from Annex II-Pyrolysis Gas Yields of Three Pyrolysis Combinations

Results of matched samples t- test						
Pairing number	Term	Mean	SD	Mean difference	t	p
Pair1	R1	15.200	8.521	1.600	0.186	0.862
	R1'	13.600	12.905			
Pair2	R2	15.740	18.249	-1.800	-0.126	0.906
	R2'	17.540	17.701			
Pair3	R3	15.060	22.797	-3.199	-0.198	0.853
	R3'	18.259	18.173			
Pair4	R4	16.040	27.879	-2.249	-0.123	0.908
	R4'	18.289	18.081			
Pair5	R5	13.820	24.700	-5.221	-0.303	0.777
	R5'	19.041	18.814			
*p<0.05 **p<0.01 ***p<0.001						

There was no significant difference between the two groups (p>0.05).

Table 8 Data from Annex II-Pyrolysis Gas Yields of Three Pyrolysis Combinations

Results of Wilcoxon analysis of paired samples					
Name	Paired median (P25, P75)		Median M difference (pair 1 - pair 2)	W	p
	Pair 1	Pair 2			
R1 Pairing R1'	13.8(0.67,18.7)	15.23(0.67,24.35)	-1.43	8.000	1.000
R2 Pairing R2'	9.8(0.72,10.6)	15.76(0.72,33.25)	-5.96	8.000	1.000
R3 Pairing R3'	5.9(0.78,8.1)	17.305(0.78,33.86)	-11.405	9.000	0.812
R4 Pairing R4'	3.4(0.81,6.5)	19.075(0.81,30.26)	-15.675	9.000	0.812
R5 Pairing R5'	3.2(0.79,5.3)	20.935(0.79,29.87)	-17.735	9.000	0.812
*p<0.05 **p<0.01 ***p<0.001					

There was no significant difference between the two groups.

According to matched samples t-test and Wilcoxon analysis of paired samples, Under the catalytic action of the same proportion of desulfurized ash, there is no significant difference in the yields of the products generated from the pyrolysis of CE and LG, as well as the yields of the components of the pyrolysis gas.

2.4 Question 4

Here are some models that can be used to establish a catalytic reaction mechanism,

i. Eley–Rideal (ER) model: This model describes a mechanism in which a chemical reaction is considered to take place when one reactant from the bulk liquid phase collides with another reactant already adsorbed on the catalyst.[12 13 14]

ii. Pseudo-homogeneous (PH) model: In this model, fluid and solid phases are considered as a single pseudo-phase and the balances are imposed for only one phase. Heat and mass transport coefficients inside the bed are calculated by expressions which account for the simultaneous presence of two phases.[12 15 16 17]

iii. Langmuir–Hinshelwood (LH) model: In this model, two molecules adsorb on neighboring sites and the adsorbed molecules undergo a bimolecular reaction[13 19]

Here, we try different ways to establish a catalytic reaction mechanism model of desulfurized ash for model compounds according to papers[12 13 14 15 16 17].

Set

$$v=k[H_2]^{\alpha}[CH_4]^{\beta}[C_2H_6]^{\gamma}[CO]^{\lambda}[CO_2]^{\theta}$$

Then,

$$\ln v = \ln k + \alpha \ln[H_2]+\beta\ln[CH_4]+\gamma\ln[C_2H_6]+\theta \ln[CO_2]+\lambda\ln[CO]$$

Then we use linear regression to model the kinetics of the catalytic reaction, we use STATA to run a multivariate linear regression.

General model set,

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon$$

i. y: dependent variable, random variable;

ii. $x_1, \ldots, x_p$: independent variable, non-random;

iii. ε: random error term, random variable;

iv. β_0: intercept term;

v. β_j: slope term corresponding to x_j;

vi. The model shows that the dependent variable y is approximately equal to the linear combination of the independent variables.

vii. $E\varepsilon = 0, \mathrm{Var}(\varepsilon) = \sigma^2$

For n sets of observations,

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, 2, \ldots, n$$

For the random error term, $\varepsilon_i, i = 1, 2, \ldots, n$, assume,

i. Expectation is zero and follows normal distribution.

ii. Homogeneity of variance: variance is independent of the value of the independent variable.

iii. Independent of each other.

In short, $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are independent of each other, subject to distribution

$\mathrm{N}(0, \sigma^2)$.

Data format is like,

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_p & y \\ x_{11} & x_{12} & \cdots & x_{1p} & y_1 \\ x_{21} & x_{22} & \cdots & x_{2p} & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} & y_n \end{pmatrix}$$

According to the model,

$$Ey = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p.$$

The unknown parameters are regression coefficients $\beta_0, \beta_1, \ldots, \beta_p, \sigma^2$, and the error term is also unknown. The least squares method is still used to estimate the model, and the coefficients are estimated $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p$ and error variance estimate $\hat{\sigma}^2$. Plug the coefficient estimates into the model and write the estimated multiple linear regression equation as follows:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_p x_p.$$

For the group i observations, the values of the independent variables are substituted into the estimated regression equation to obtain the fitting values

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip}.$$

$e_i = y_i - \hat{y}_i$ is called the residual. Regression parameter estimation, using the minimum residual as the goal, make

$$Q = \sum_{i=1}^{n} \left[y_i - \left(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} \right) \right]^2$$

Take the minimum$Q$, $(\beta_0, \beta_1, \ldots, \beta_p)$ as a parameter estimation, called least squares estimation, denoted by $(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p)$. The minimum value obtained is SSE, and the estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n-p-1}\mathrm{SSE} = \frac{1}{n-p-1}\sum_{i=1}^{n} e_i^2.$$

Here we assume that the higher the catalyst content, the faster the chemical reaction, which means，

$$\mathrm{v} \propto \mathrm{DFA/CE(LG)}$$

Therefore,

$$\mathrm{v} = \mathrm{k[DFA/CE(LG)]}$$

k is the adjustment factor.

Therefore,

$$\mathrm{lnv} = \mathrm{lnk} + \mathrm{ln[DFA/CE(LG)]}$$

Then set DFA/CE and DFA/LG as dependent variables, and H_2, CO, CO_2, CH_4, C_2H_6 as independent variables. Because the sample data is small and there is overfitting, we choose the regression variables by calculating the correlation. Dataset is from Annex II. [X] means X gas yields of three pyrolysis combinations.

First we calculate Tab 2 Yield of Gaseous Components from DFA/CE Pyrolysis (mL/g,daf)

Table 9 Matrix of correlations of Tab 2 Yield of Gaseous Components from DFA/CE Pyrolysis (mL/g,daf)

Variables	(1)	(2)	(3)	(4)	(5)	(6)
(1) ln[DFA/CE]	1.000					
(2) ln[H_2]	0.915	1.000				
(3) ln[CO]	-0.985	-0.860	1.000			
(4) ln[CO_2]	-0.989	-0.858	0.999	1.000		
(5) ln[CH_4]	-0.951	-0.827	0.967	0.963	1.000	
(6) ln[$C2H_6$]	-0.963	-0.922	0.972	0.964	0.926	1.000

We adopted the table above and the stepwise regression method, finally selected CO, CO_2, C_2H_6 as the independent variable.

Therefore, the equation is,

$$\ln v = \ln k + \gamma \ln[\mathrm{C_2H_6}] + \theta \ln[\mathrm{CO_2}] + \lambda \ln[\mathrm{CO}]$$

Results are as follows,

Table 10 Linear regression of Tab 2 Yield of Gaseous Components from DFA/CE Pyrolysis (mL/g,daf)

ln[DFA/CE]	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
ln[CO]	5.417	3.937	1.38	.4	-44.609	55.443	
ln[CO_2]	-2.604	1.468	-1.77	.327	-21.262	16.055	
ln[$C2H_6$]	-.596	.456	-1.31	.415	-6.387	5.194	
Constant	-6.546	5.219	-1.25	.428	-72.856	59.764	

Mean dependent var	-0.688	SD dependent var	0.631
R-squared	0.993	Number of obs	5
F-test	46.468	Prob > F	0.107
Akaike crit. (AIC)	-8.260	Bayesian crit. (BIC)	-9.822

*** p<.01, ** p<.05, * p<.1

The R-squared value of the model is 0.9929, indicating that the regression model can explain approximately 99.29% of the variance in the target variable ln[DFA/CE]. This indicates a high level of fit. The regression model has a high level of fit.

Therefore,

$$\ln v = -6.546 - 0.596\ln[C_2H_6] - 2.604\ln[CO_2] + 5.417\ln[CO]$$

Same we can get the result of Tab 3 Yield of Gaseous Components from DFA/LG Pyrolysis (mL/g,daf)

Table 11 Matrix of correlations of Tab 3 Yield of Gaseous Components from DFA/LG Pyrolysis (mL/g,daf)

Variables	(1)	(2)	(3)	(4)	(5)	(6)
(1) ln[DFA/CE]	1.000					
(2) ln[H_2]	0.935	1.000				
(3) ln[CO]	0.946	0.856	1.000			
(4) ln[CO_2]	0.924	0.899	0.761	1.000		
(5) ln[CH_4]	0.246	0.373	-0.028	0.590	1.000	
(6) ln[$C2H_6$]	-0.547	-0.221	-0.532	-0.467	0.108	1.000

Table 12 Linear regression of Tab 3 Yield of Gaseous Components from DFA/LG Pyrolysis (mL/g,daf)

lln[DFA/LG]	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
ln[H_2]	3.782	2.381	1.59	.253	-6.462	14.026	
ln[CO]	2.594	1.41	1.84	.207	-3.471	8.659	
Constant	-7.042	4.632	-1.52	.268	-26.971	12.887	
Mean dependent var	-0.688		SD dependent var	0.631			
R-squared	0.953		Number of obs	5			

F-test	20.372	Prob > F	0.047
Akaike crit. (AIC)	-0.848	Bayesian crit. (BIC)	-2.019
*** p<.01, ** p<.05, * p<.1			

Therefore,

$$\ln v = -7.041946 + 3.782118 \ln[H_2] + 2.594206 \ln[CO]$$

2.5 Question 5

We employed a trilayered neural network to predict the value of Annex I- Pyrolysis Product Yields of Three Pyrolysis Combinations Tab 1 Yield of Decomposition Products from DA/CS Pyrolysis wt.%(daf). In this study, DA/CS was considered as the dependent variable, while tar yield was treated as the independent variable. The training MATLAB code has been provided in the appendix, and the MATLAB executable program file is included in the supporting materials.

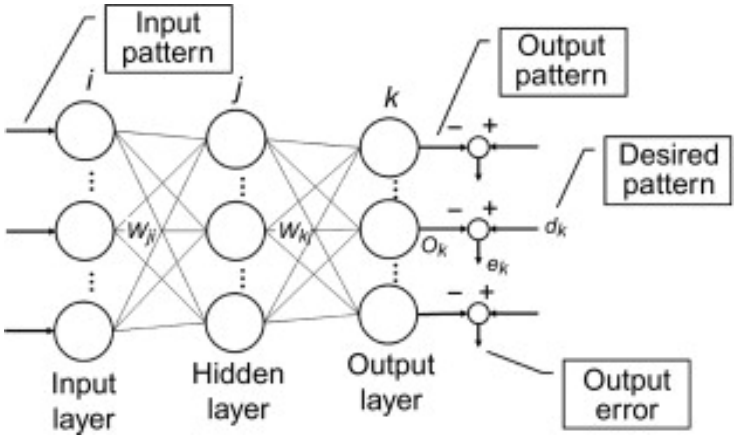


Figure 7 Three-Layered Neural Network

i. Set the initial values of $w_{ji},\ w_{kj},\ \theta_j,\ \theta_k, and\ \eta > 0$.

ii. Specify the desired values of the output $d_k, k = 1,2,\ldots,k$

iii. corresponding to the input data $x_i, i = 1,2,\ldots,I$ in the input layer. Step 3: Calculate the outputs of the neurons in the hidden layer and output layer by

$$\text{net}_j = \sum_{i=1}^{l} w_{ji}\, x_i - \theta_j, O_j = f\text{net}_j, \quad fx = \frac{1}{1+\text{e}^{-x}}$$

$$\text{net}_k = \sum_{j=1}^{J} w_{ji}\, O_i - \theta_k, \quad O_k = f\text{net}_k$$

iv. Calculate the error e_k and generalized errors δ_k, δ_j by

$$e_k = d_k - O_k, \delta_k = e_k O_k 1 - O_k$$

$$\delta_k = \sum_{k=1}^{K} \delta_k \, w_{kj} O_j 1 - O_j$$

v. If e_k is sufficiently small for all k, END and otherwise.

$$\Delta w_{kj} = \eta O_j \delta_k, w_{kj} \Leftarrow w_{kj} + \Delta w_{kj}$$

$$\Delta w_{ji} = \eta O_i \delta_j, w_{ji} \Leftarrow w_{ji} + \Delta w_{ji}$$

vi. Go to Step 3.

Here is the Training Results:

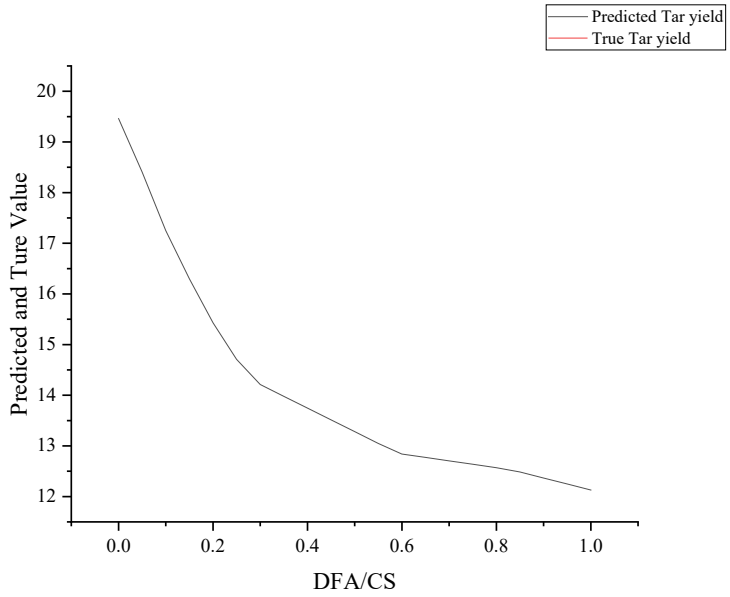
Metric	**Value**
RMSE (Validation)	0.37064
R-Squared (Validation)	0.98
MSE (Validation)	0.13738
MAE (Validation)	0.28966
Prediction speed	-440 obs/sec
Training time	2.461 sec
Model size (Compact)	~8 kB

Table 13 Original Data

DFA/CS	0	0.1	0.2	0.3	0.4	0.5	0.6	0.8	1
Tar yield	19.46	17.25	15.43	14.14	13.89	13.21	12.84	12.57	12.13

Table 14 Predicted Results

DFA/CS	Predicted Tar yield	True Tar yield	**DFA/CS**	Predicted Tar yield	True Tar yield	**DFA/CS**	Predicted Tar yield	True Tar yield
0	19.46	19.46	**0.35**	13.9792	-	**0.7**	12.705	-
0.05	18.4063	-	**0.4**	13.7467	13.89	**0.75**	12.6375	-
0.1	17.25	17.25	**0.45**	13.5142	-	**0.8**	12.57	12.57
0.15	16.2978	-	**0.5**	13.2817	13.21	**0.85**	12.485	-
0.2	15.43	15.43	**0.55**	13.0492	-	**0.9**	12.3667	-
0.25	14.7091	-	**0.6**	12.84	12.84	**0.95**	12.2483	-
0.3	14.2117	14.14	**0.65**	12.7725	-	**1**	12.13	12.13



3. Symbol and Assumptions

3. 1 Symbol Description

Variable name		Description
Difference Analysis for Annex I		
DFA/CE	DFA/LG	Rate(DFA/CE or DFA/LG)
R1	R1’	0.1
R2	R2’	0.2
R3	R3’	0.3
R4	R4’	0.4
R5	R5’	0.5
R6	R6’	0.6
R7	R7’	0.8
R8	R8’	1
Difference Analysis for Annex II		
DFA/CE	DFA/LG	Rate(DFA/CE or DFA/LG)
R1	R1’	0.2
R2	R2’	0.4
R3	R3’	0.5
R4	R4’	0.8
R5	R5’	1.0
Description of other variables		

k	Adjustment factor between reaction rate and catalyst ratio
ln	Natural logarithm
[X]	X gas yields of three pyrolysis combinations.

3.2 Fundamental assumptions

i. One of the reactants is adsorbed on the surface while the other reactant approaches from the gas phase.
ii. There is no interaction between the adsorbed reactant and the gas-phase reactant.
iii. The reaction occurs only when the gas-phase reactant collides with the adsorbed reactant on the surface.
iv. The reaction occurs in a well-mixed, single-phase system.
v. The surface is treated as a homogeneous site where reactants can freely interact with each other.
vi. The reaction rate is determined by the concentration of the surface species.
vii. Both reactants are absorbed on the surface and interact with each other.
viii. The adsorbed reactants form an intermediate complex before the reaction takes place.
ix. The reaction rate is determined by the concentration of the adsorbed species.
x. The higher the catalyst content, the faster the chemical reaction speed.

4. Sensitivity Analysis

According to the data in the correlation table, we tried to regression ln[DFA/CE] ln[CO] ln[CO_2] for Tab 2 Yield of Decomposition Products from DFA/CE Pyrolysis wt.%(daf). We found that the overall explanatory power of the model was strong, and R-squared was 0.9807, indicating that about 98.07% of the variance could be explained by the model. The adjusted R-squared is 0.9613. This shows the overall stability of the model.

ln[DFA/CE]	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
ln[CO]	1.669	3.147	0.53	.649	-11.873	15.21	
ln[CO_2]	-1.417	1.345	-1.05	.403	-7.205	4.371	
Constant	-1.723	4.305	-0.40	.728	-20.245	16.798	

Mean dependent var	-0.688	SD dependent var	0.631
R-squared	0.981	Number of obs	5
F-test	50.746	Prob > F	0.019
Akaike crit. (AIC)	-5.269	Bayesian crit. (BIC)	-6.441

*** p<.01, ** p<.05, * p<.1

Additionally, we do the same process with Tab 3 Yield of Decomposition Products from DFA/LG Pyrolysis wt.%(daf), we regress ln[DFA/LG] ln[CO] ln[H_2] ln[CO_2]

ln[DFA/LG]	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
ln[CO]	2.665	.745	3.58	.174	-6.804	12.134	
ln[H_2]	.359	1.866	0.19	.879	-23.355	24.072	
ln[CO_2]	1.243	.5	2.48	.244	-5.115	7.6	
Constant	-12.69	3.341	-3.80	.164	-55.137	29.758	

Mean dependent var	-0.688	SD dependent var	0.631
R-squared	0.993	Number of obs	5
F-test	50.719	Prob > F	0.103
Akaike crit. (AIC)	-8.694	Bayesian crit. (BIC)	-10.257

*** p<.01, ** p<.05, * p<.1

The overall explanatory power of the model is strong, R-squared is 0.9935, indicating that about 99.35% of the variance can be explained by the model. The adjusted R-squared is 0.9739. This shows the overall stability of the model.

5. Strengths and Weakness

5.1 Advantages

i. The establishment of the mathematical model is based on a comprehensive review of relevant literature, incorporating recognized research findings from previous studies.
ii. Correlation analysis and stepwise regression are employed to identify the optimal fitting equation, ensuring a robust and accurate model.
iii. The utilization of the gradient descent algorithm in machine learning prediction enables the identification of the most suitable parameters, thereby improving the prediction accuracy.

5.2 Disadvantages

i. The limited size of the dataset may introduce potential confounding factors and impact the regression analysis of the variables.
ii. The significance of the relationships between variables during the regression process might be uncertain. Further investigation and analysis of a larger dataset are necessary to enhance the accuracy of the assessment. The F statistic of the model is 46.47, with a probability value (Prob > F) of 0.1073. The relatively large probability value suggests that there may be room for improvement in the fit of the model.
iii. The coefficients (Coefficient) of the independent variables determine their impact on the target variable. However, for ln[CO], ln[CO_2], and ln[C_2H_6], their coefficients have relatively large p-values, indicating that their effects on the target variable may not be significant. Additionally, the confidence intervals should also be considered to accurately evaluate their effects.
iv. The coefficient of the intercept (Constant, _cons) is -6.54625, and its p-value is also relatively large, suggesting that the intercept may not have a significant impact on the target variable.

6. Memo

We also read some non-technical literatures on cotton stalk biomass for bioenergy and economic studies during the modeling process. We hope our research will contribute to the economy and society.

i. Tatsiopoulos, I.P. and Tolis, A.J. described a model, which simulates the cotton biomass supply chain.[23]
ii. Pandirwar, A.P. et al. estimated the production cost of ethanol from cotton stalks and compares it with other lignocellulosic feedstocks and fossil fuels.[24]
iii. Uyan, M. et al. analyzed the energy balance, greenhouse gas emissions, and economic performance of the proposed biorefinery system and compares it with the conventional cotton production system.[25]

References

1. Yang Songlin (Ed.), Liberal Arts Mathematics, Soochow University Press, 2015.01, p. 107
2. Shaou-Gang Miaou; Jin-Syan Chou. Fundamentals of probability and statistics. Gao Li Books. 2012: 147.
3. Hu Y. Normal distribution [J]. Business Weekly, 2009 (24): 94-94.
4. Anderson T W, Anderson T W, Anderson T W, et al. An introduction to

multivariate statistical analysis[M]. New York: Wiley, 1958.
5. Zhao Zhengsong, Pan Dengdeng. Analysis of taxi industry operation law based on two independent samples and paired samples T-test [J]. Transportation Construction and Management, 2013 (08): 86-87.
6. Dual-population (σ_1~2,σ_2~2 unknown,n≤30) paired sample T-test and independent sample T-test based on SPSS. Reading and Writing (Journal of Education and Teaching), 2016
7. Qi Yongzhong. The impact of two RRR cuts on interest rate risk premium in China in 2014: An empirical analysis based on two paired sample T-test [J]. China Collective Economy, 2016 (10): 57-59.
8. Han Shuguang, Wu Jing, Chen Qiong. Case evaluation of fresh e-commerce logistics service based on paired sample T-test [J]. Journal of Zhejiang University of Science and Technology (Social Science Edition), 2016, 36(03): 246-251.
9. Wu Xizhi and Zhao Bojuan. Nonparametric Statistics: China Statistics Press, 2011
10. Qin Shan, Wang Xiaoyin. Application of rank sum test in feed science research [J]. Feed Wide Angle, 2003(21):30-32. (in Chinese)
11. Committee of the Dictionary of Mathematics. Mathematics Ci Hai (1-6). Volume 6 [M]. Shanxi Education Press, 2002.
12. Lamba, R. Reaction kinetic models for heterogeneous solid catalyzed esterification of decanoic acid with ethanol. Biomass Conv. Bioref. 13, 12157–12166 (2023). https://doi.org/10.1007/s13399-022-02413-4
13. Wikipedia contributors. "Reactions on surfaces." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 1 Dec. 2022. Web. 16 Nov. 2023.
14. Prins, R. Eley–Rideal, the Other Mechanism. Top Catal 61, 714–721 (2018). https://doi.org/10.1007/s11244-018-0948-8
15. https://core.ac.uk/download/pdf/213394653.pdf
16. Jakobsen, H.A. (2014). Packed Bed Reactors. In: Chemical Reactor Modeling. Springer, Cham. https://doi.org/10.1007/978-3-319-05092-8_11
17. Zabala S, Reyero I, Campo I, Arzamendi G, Gandía LM. Pseudo-Homogeneous and Heterogeneous Kinetic Models of the NaOH-Catalyzed Methanolysis Reaction for Biodiesel Production. Energies. 2021; 14(14):4192. https://doi.org/10.3390/en14144192
18. De Falco, M. (2011). Membrane Reactors Modeling. In: De De Falco, M., Marrelli, L., Iaquaniello, G. (eds) Membrane Reactors for Hydrogen Production Processes. Springer, London. https://doi.org/10.1007/978-0-85729-151-6_4
19. Hai, Tran & Nguyen, Dinh & Do, Phuong & Tran, Uyen. (2023). Kinetics of photocatalytic degradation of organic compounds: a mini-review and new

approach. RSC Advances. 13. 16915-16925. https://doi.org/10.1039/D3RA01970E
20. Jolliffe, I. T. (2002). Principal component analysis. Wiley Online Library1
21. Abdi, H., & Williams, L. J. (2010). Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2(4), 433-4592
22. StataCorp. (2017). Stata Statistical Software: Release 15. StataCorp LLC
23. Tatsiopoulos, I.P. and Tolis, A.J. (2003). Economic aspects of the cotton-stalk biomass logistics and comparison of supply chain methods. Biomass and Bioenergy, 24(3), 199-214. doi: 10.1016/s0961-9534(02)00115-0. Available at: https://eurekamag.com/research/003/722/003722081.php
24. Pandirwar, A.P., Khadatkar, A., Mehta, C.R. et al. Technological Advancement in Harvesting of Cotton Stalks to Establish Sustainable Raw Material Supply Chain for Industrial Applications: a Review. Bioenerg. Res. 16, 741–760 (2023). https://doi.org/10.1007/s12155-022-10520-3
25. Uyan, M., Alptekin, F.M., Bastabak, B. et al. Combined biofuel production from cotton stalk and seed with a biorefinery approach. Biomass Conv. Bioref. 10, 393–400 (2020). https://doi.org/10.1007/s13399-019-00427-z

Appendix

```
function [trainedModel, validationRMSE] = trainRegressionModel(trainingData,
responseData)

inputTable = array2table(trainingData', 'VariableNames', {'row_1'});

predictorNames = {'row_1'};
predictors = inputTable(:, predictorNames);
response = responseData;
isCategoricalPredictor = [false];

regressionNeuralNetwork = fitrnet(...
    predictors, ...
    response, ...
    'LayerSizes', [10 10 10], ...
    'Activations', 'relu', ...
    'Lambda', 0, ...
    'IterationLimit', 1000, ...
    'Standardize', true);

predictorExtractionFcn = @(x) array2table(x', 'VariableNames', predictorNames);
neuralNetworkPredictFcn = @(x) predict(regressionNeuralNetwork, x);
trainedModel.predictFcn = @(x)
neuralNetworkPredictFcn(predictorExtractionFcn(x));
```

```
trainedModel.RegressionNeuralNetwork = regressionNeuralNetwork;
trainedModel.About = 'This struct is a trained model exported from Regression Learner R2023b.';
trainedModel.HowToPredict = sprintf('To make predictions on a new predictor row matrix, X, use: \n    yfit = c.predictFcn(X) \nreplacing "c" with the name of the variable that is this struct, e.g. "trainedModel". \n \nX must contain exactly 1 rows because this model was trained using 1 predictors. \nX must contain only predictor rows in exactly the same order and format as your training \ndata. Do not include the response row or any rows you did not import into the app. \n \nFor more information, see <a href="matlab:helpview(fullfile(docroot, "stats", "stats.map"), "appregression_exportmodeltoworkspace")">How to predict using an exported model</a>.');

inputTable = array2table(trainingData', 'VariableNames', {'row_1'});

predictorNames = {'row_1'};
predictors = inputTable(:, predictorNames);
response = responseData;
isCategoricalPredictor = [false];

partitionedModel = crossval(trainedModel.RegressionNeuralNetwork, 'KFold', 5);

validationPredictions = kfoldPredict(partitionedModel);

validationRMSE = sqrt(kfoldLoss(partitionedModel, 'LossFun', 'mse'));
```