Research on the Impact of n-Hexane Insoluble Substances on Tar Yield Based on Pearson Correlation and Regression Analysis

Yenian Zhao *

School of Mathematics and Statistics, Lanzhou University, Lanzhou, China, 730107

* Corresponding Author Email: 3235218684@qq.com

Abstract. With increasing environmental pressures and energy demands, pyrolysis technology for biomass and waste has garnered attention. This study examines the impact of n-hexane insoluble substances on pyrolysis products, particularly tar yield, through Pearson correlation and multiple regression analyses. During data pre-processing, null values for n-hexane insoluble substances were set to zero, and outliers in product yields were removed to ensure data integrity and accuracy. Results indicate a significant positive correlation between n-hexane insoluble substances and tar yield, while correlations with water and coke slag yields are minimal. Linear and multiple regression models were employed to predict the yields of tar, water, and coke slag, revealing that n-hexane insoluble substances have a statistically significant effect on tar yield. Multiple regression analysis further shows that the production of various products is strongly influenced by both the content of n-hexane insoluble substances and their interaction with other components, especially evident in the interaction effect on coke slag yield. These findings provide new insights into the mechanism by which n-hexane insoluble substances affect product distribution during pyrolysis and offer theoretical support for optimizing the pyrolysis process and enhancing tar yield.

Keywords: n-hexane insoluble substances, tar yield, Pearson correlation, multiple regression analysis.

1. Introduction

In recent years, the high energy consumption and environmental concerns associated with fossil fuels have intensified interest in pyrolysis technology for biomass and waste conversion [1, 2]. Pyrolysis, as an efficient biomass transformation technique, converts complex biomass and waste materials into energy carriers and chemical precursors [3, 4]. However, the composition and yield of pyrolysis products are often influenced by various factors, including feedstock characteristics, temperature, pressure, and solvent properties [5, 6]. As a significant component in pyrolysis feedstock, n-hexane insoluble substances may impact the yields of products like tar and coke through complex mechanisms [7]. To enable effective yield prediction and optimization, quantitatively evaluating the influence of n-hexane insoluble substances on pyrolysis product yields is of great importance.

Pearson’s correlation coefficient, widely used for measuring linear relationships, aids in identifying the linear correlation between n-hexane insoluble substances and tar yield [8]. In addition, regression analysis [9] not only explores the associations among variables but also predicts trends in the yield of products like tar [10]. Therefore, this study employs Pearson correlation analysis and multiple regression modeling to quantify the impact of n-hexane insoluble substances on tar yield, further elucidating their potential influence mechanism in the pyrolysis process and providing a theoretical basis for optimizing subsequent pyrolysis processes.

2. Data Preprocessing and Correlation Analysis of Pyrolysis Products

The data for this study is sourced from http://www.nmmcm.org.cn/. In the pre-processing phase, null values for n-hexane insoluble substances in Annex 1 are treated as zero. Additionally, outliers present in individual pyrolysis product yields in Annex 1 are identified and removed.

Perason's correlation coefficient is widely used as a measure of the strength of the linear correlation between two variables, with values ranging from -1 to 1. Positive values indicate a positive correlation, negative values indicate a negative correlation, and values close to 0 indicate that there is little or no linear correlation between the two variables. The formula is as follows:

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X_i})(Y_i - \overline{Y_i})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X_i})^2} * \sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y_i})^2}} \tag{1}$$

The Perason correlation coefficients are shown in the table below:

Table 1. Person correlation results.

	Hexane insoluble substance (INS)g	Tar yield	productivity of fish	Coke slag yield	Hexane soluble yield
Hexane insoluble substance (INS)g	1 (0.000***)	0.719 (0.000***)	-0.145 (0.292)	-0.074 (0.591)	0.238 (0.080*)
Tar yield	0.719 (0.000***)	1 (0.000***)	-0.286 (0.034**)	0.036 (0.792)	0.241 (0.077*)
productivity of fish	-0.145 (0.292)	-0.286 (0.034**)	1 (0.000***)	-0.895 (0.000***)	-0.113 (0.411)
Coke slag yield	-0.074 (0.591)	0.036 (0.792)	-0.895 (0.000***)	1 (0.000***)	-0.05 (0.717)
Hexane soluble yield	0.238 (0.080*)	0.241 (0.077*)	-0.113 (0.411)	-0.05 (0.717)	1 (0.000***)

Note: ***, **, * represent 1 per cent, 5 per cent and 10 per cent significance levels, respectively.

With the above Table 1, it can be seen that n-hexane insoluble matter has a significant effect on tar yield, while it has little effect on the yield of other pyrolysis products.

3. Surface Weathering Analysis and Data Processing

Linear regression analysis is used to predict the value of one variable based on the value of another variable. The variable to be predicted is called the dependent variable. The variable used to predict the value of another variable is called the independent variable.

This form of analysis estimates the coefficients of a linear equation involving one or more independent variables that best predict the value of the dependent variable. Linear regression fits a line or surface that minimises the difference between the predicted value and the actual output. There are simple linear regression calculators that use the "least squares" method to find the best fit line for a set of paired data. The value of X (the dependent variable) can then be estimated from Y (the independent variable).

The solution results demonstrate the relationships between tar yield, water yield, and coke slag yield with n-hexane insoluble matter. The fitted equations are as follows:

Tar yield vs. n-hexane insoluble matter: y=0.083+0.116×INS (g), with P=0.0001

Water yield vs. n-hexane insoluble matter: y=0.125−0.027×INS (g), with P=0.292

Coke slag yield vs. n-hexane insoluble matter: y=0.666−0.026×INS (g), with P=0.591

The significance coefficient P values from each product’s fit indicate that n-hexane insoluble matter has a statistically significant impact on tar yield.

Multiple regression analysis, or simply multiple regression, is a method used to quantitatively represent the linear dependence between a dependent variable and multiple independent variables through regression equations. The core idea of regression analysis is that, while no strict deterministic functional relationship may exist between the independent and dependent variables, one can attempt to identify a mathematical expression that best captures the relationship between them. This approach allows for the measurement of the impact of each independent variable on the dependent variable.
The multiple regression formula is derived as follows:

$$f(x) = k^T * x + b * 1 \tag{2}$$

Expand the formula:

$$y=k_1*x_1+k_2*x_2+\ldots\ldots+k_m*x_m \tag{3}$$

It may be useful to let:

$$X=(x,1)=[\ x_1,\ x_2,\ \ldots\ldots,\ x_m,1] \tag{4}$$

$$\hat{k}=(k,b)=[\ k_1,\ k_2,\ldots\ldots k_m] \tag{5}$$

This gives $y=\hat{k}^T\ X$.

Further, introducing the mean square error J $(\hat{k})$.

$$J(\hat{k}\ )=(y-X\hat{k})^T\ (y-X\hat{k}\ ) \tag{6}$$

The mean square error function takes the derivative of $\hat{k}$ derivative of the mean square error function to 0.

$$\frac{\partial}{\partial \hat{k}}\ E(\hat{k}\ )=0 \tag{7}$$

Rolled out:

$$\hat{k}\ =(X^TX)^{-1}X^Ty \tag{8}$$

Getting the regression model:

$$f(\hat{y_i}\ )=\hat{x_i}(X^TX)^{-1}X^Ty \tag{9}$$

Combine the mixing ratios of various biomass and coal-based materials with n-hexane insoluble content and the yield rates of each pyrolysis product. Use MATLAB to perform multiple regression analysis and create scatter plots along with residual plots.

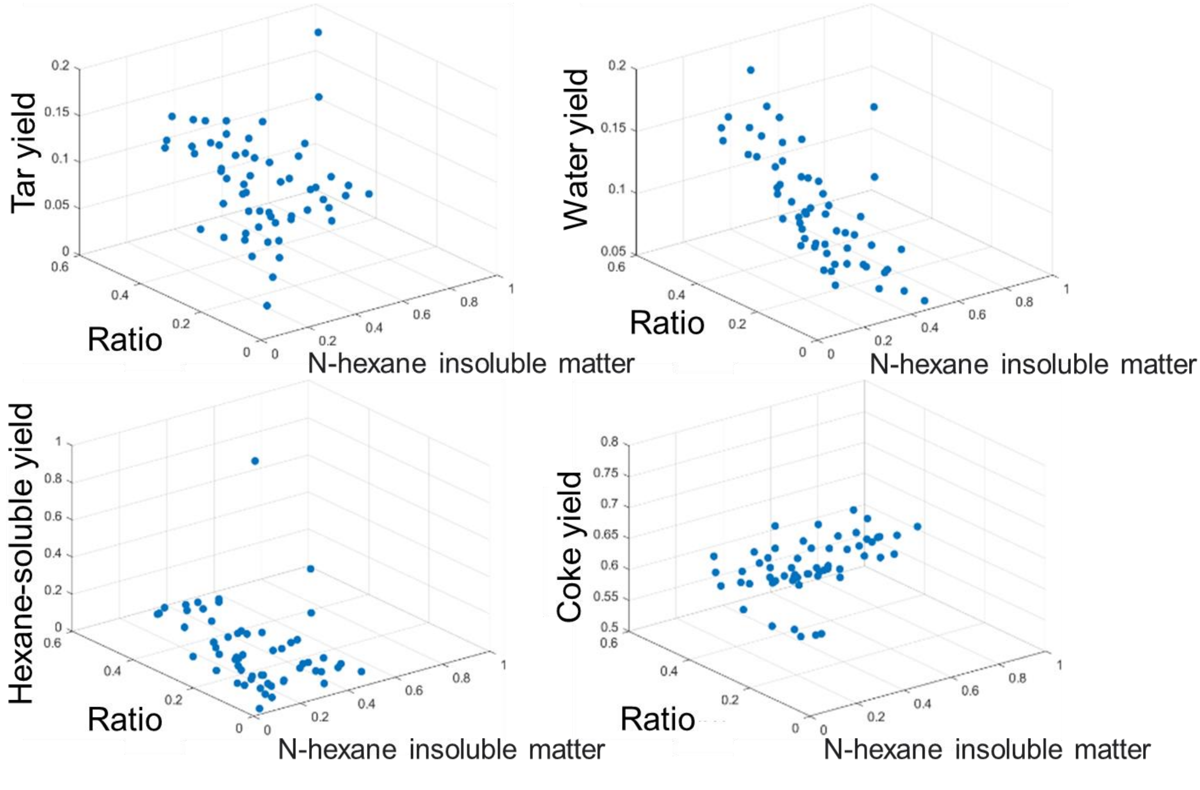

Figure 1. Multiple linear regression scatter plot.

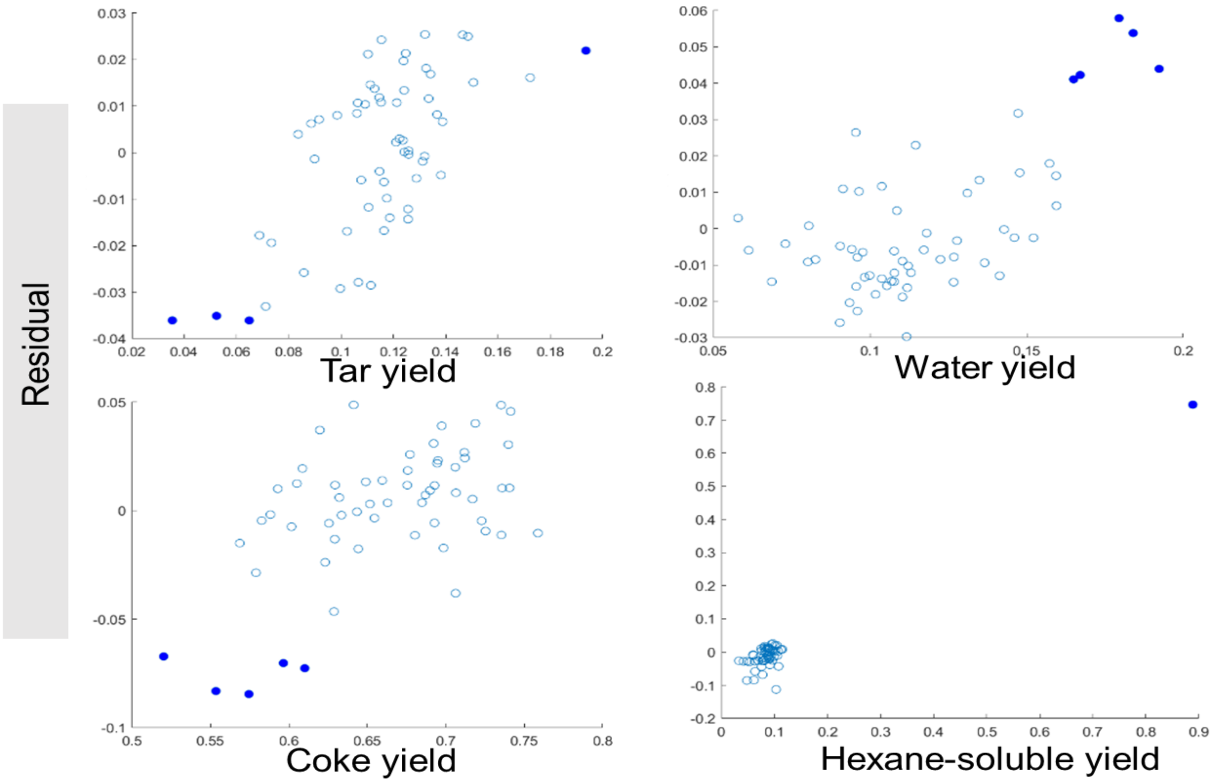

Figure 2. Multiple linear regression residual plot.

Figure 1 and 2 illustrate the relationship between n-hexane insoluble substances and various pyrolysis product yields, as well as the performance of regression models. There is a clear positive correlation between n-hexane insoluble matter and tar yield, showing that as the amount of insoluble matter increases, tar yield rises. However, the relationships with other products, such as water, coke, and hexane-soluble yields, are much weaker, indicating a minimal impact of insoluble substances on these products. The residual plots show that the regression model fits well for tar yield, with residuals close to zero, while the models for the other products have larger residuals, suggesting that the models do not explain these yields as effectively. From the fitted R-test values, it is clear that the interaction effect of scorched slag is the most pronounced, indicating that it plays a more significant role in influencing the overall outcomes compared to other variables. Overall, n-hexane insoluble substances significantly affect tar yield, but the effect on other products is weaker, with the interaction effect of scorched slag being the most notable.

4. Conclusion

This study provides a comprehensive analysis of the impact of n-hexane insoluble substances on pyrolysis product yields, with a particular focus on tar yield. Using Pearson correlation analysis, we found a significant positive correlation between n-hexane insoluble content and tar yield, suggesting that higher levels of insoluble substances promote tar production. Regression modeling further corroborated these findings, showing that n-hexane insoluble substances serve as a strong predictor of tar yield, while having a minimal effect on the yields of other pyrolysis products like water and coke.

Our analysis also reveals that the interaction effect of scorched slag is particularly significant, implying that this component may influence the pyrolysis outcomes more substantially than others. Residual plots validate that our regression model aligns well with the actual tar yield data, while showing limitations in predicting yields for water and coke. These findings provide valuable insights into optimizing pyrolysis processes by adjusting the concentration of n-hexane insoluble substances to maximize tar yield, thereby enhancing the efficiency and environmental sustainability of biomass and waste pyrolysis.

References

[1] WANG G, DAI Y, YANG H, et al. A review of recent advances in biomass pyrolysis [J]. Energy & fuels, 2020, 34(12): 15557-15578.

[2] JHA S, NANDA S, ACHARYA B, et al. A review of thermochemical conversion of waste biomass to biofuels [J]. Energies, 2022, 15(17): 6352.

[3] AMENAGHAWON A N, ANYALEWECHI C L, and OKIEIMEN C O, et al. Biomass pyrolysis technologies for value-added products: a state-of-the-art review [J]. Environment, development and sustainability, 2021: 1-55.

[4] ZHANG H, YANG K, TAO Y, et al. Biomass directional pyrolysis based on element economy to produce high-quality fuels, chemicals, carbon materials–a review[J]. Biotechnology Advances, 2023: 108262.

[5] AL-RUMAIHI A, SHAHBAZ M, MCKAY G, et al. A review of pyrolysis technologies and feedstock: A blending approach for plastic and biomass towards optimum biochar yield [J]. Renewable and Sustainable Energy Reviews, 2022, 167: 112715.

[6] TIWARI M, DIRBEBA M J, LEHMUSTO J, et al. Analytical and applied pyrolysis of challenging biomass feedstocks: Effect of pyrolysis conditions on product yield and composition [J]. Journal of Analytical and Applied Pyrolysis, 2024, 177: 106355.

[7] CHE Y, SHI K, CUI Z, et al. Conversion of low temperature coal tar into high value-added chemicals based on the coupling process of fast pyrolysis and catalytic cracking[J]. Energy, 2023, 264: 126169.

[8] FU T, TANG X, CAI Z, et al. Correlation research of phase angle variation and coating performance by means of Pearson's correlation coefficient [J]. Progress in Organic Coatings, 2020, 139: 105459.

[9] TYAGI K, RANE C, MANRY M. Regression analysis[M]//Artificial intelligence and machine learning for EDGE computing. Elsevier, 2022:53-63.

[10] SUREIMAN O, MANGERA C M. F-test of overall significance in regression analysis simplified [J]. Journal of the Practice of Cardiovascular Sciences, 2020, 6(2): 116-122.