Data Analysis of Biomass Isotopes Release During Thermal Conversion of Biomass

Member: Dingning Liu, Min Kim, Yutian Yang, Ziqin Wang
Client: Dr. Peter Thy, project scientist, Department of Earth and Planetary Sciences

Introduction

Biomass is plant-based material that can be used as fuel to produce heat or electricity. Examples are wood and wood residues, energy crops, agricultural residues, and waste from industry, farms, and households.[1] Since biomass fuels are primarily composed of carbon, hydrogen, and oxygen, the main products from burning biomass are carbon dioxide and water. Flame temperatures can exceed 2000°C, depending on the heating value and moisture content of the fuel, the amount of air used to burn the fuel, and the construction of the furnace. The utilization of renewable biomass for energy generation has distinct strategic, economic, and environmental advantages [2]. To enable a direct analysis of metal components being released from biomass as a function of temperature, the connection of Differential Thermal Analyzer (DTA) to an Inductively Coupled Plasma Mass Spectrometer (ICPMS) is performed. In the DTA-ICPMS procedure, elements such as carbon and calcium come off earlier while quite a few elements come off at the last stage with very high temperatures.

Scientific background and goals

The scientific challenge of the project: “Direct Monitoring of Elemental Losses From Thermal Conversion of Biomass” is to find a method to identify the fundamental patterns in the dataset that are related to chemical reactions such as organic decomposition and oxide formation and breakdown as a function of increasing temperature. Specifically, the scientific goal is to investigate how metal releases are related to different chemical reactions during the combustion and heating process of organic fuels. The primary question of interest is to find the relationship between the release of isotopes and time, furthermore, we would also like to study the effect of other factors such as total mass of individual elements and their kinetic behavior.

The goal of our data analysis is to find a statistical model or approach that can identify fundamental patterns in the dataset related to thermochemical reactions as a function of temperature or time. The data analysis involves statistical methodologies: preprocessing of data and exploratory data analysis and application of random forest.

Limitation and recommendations

The most important limitation of this analysis lies in the sample size. Since the isotopes release patterns depend on time, temperature, samples (particle sizes, distribution, weights, etc.). One combustion process is not sufficient to obtain robust statistical conclusions. Collecting more samples (5-10 combustion process) would increase the accuracy of our analysis. Another limitation is the correlation between time and temperature. With a constant heating rate, the temperature is consistently increasing as a function of time. However, some isotopes may need a longer time for releasing at a certain temperature. Therefore, adjusting the heating rate so that temperatures can be unchanged for a period of time would make our results more reliable.

Previous work

Recent research on experimenting tracing metal release during wood pyrolysis has elucidated the role and behavior of many elements during thermal treatment of biomass [3]. Machine learning techniques have been widely used in the modeling of combustion, gasification, and pyrolysis processes. Ji and Deng proposed a chemical reaction neural network (CRNN) to interpret reaction pathways and quantify kinetic parameters simultaneously from the weight of the neural network [4]. Kardani et al. [5] compared Extreme Gradient Boosting method (XGBoost), Multilayer Perceptron Artificial Neural Network (MLPANN), and Support Vector Machine (SVM) in predicting the conversion of any feedstock based on the composition and operating conditions. Aghaaminiha et al. [6] showed that machine learning models are appropriate for interpolating and extrapolating the kinetics of the HTC process but deviated from experimental data under certain conditions.

Data collection

The dataset of the project was obtained from two sources: DTA and ICPMS. DTA is a thermal technique where the material under study and an inert reference are made to undergo identical thermal cycles while recording any temperature difference between sample and reference. ICPMS is a type of mass spectrometry that uses an inductively coupled plasma to ionize the sample. It atomizes the sample and creates atomic and small polyatomic ions, which are then detected by their specific mass. The two datasets are synchronized and down-sampled to ease the computation with MATLAB or Python. The down-sampled dataset has 870 observations with 67 variables. It includes Time (s), Setpoint Temperature (C), Sample Temperature (C), Sample Weight (mg), Residual Mass Fraction, Heat Flow (J/s), and released count of 55 isotopes.

Exploratory data analysis

As there are a lot of overlaps during the release process, we performed exploratory data analysis on a few selected isotopes (C12, Na23, Si28, Ca44, K39, S34, and Cl35) shown in figure 1, The plot shows that metals such as Na, Ca, and K have similar release patterns before 3500s. After 3500s, the release of K and Na increased dramatically, while the release counts of Ca dropped and then increased gradually. This analysis allows us to detect patterns of a few selected isotopes. Additionally, the organic material will decompose and some metals that are bound in the organic molecules will also be released together with C-H-O. We also noticed that all selected isotopes have the same peak as C12, which indicates that they are probably bound in the organic cellulose. By looking at log counts of release as a function of temperature, we noticed there are secondary phases of breakdown, particularly sulfates and carbonates. This motivates us to group related releases. Therefore, a method to classify isotopes is needed so that we are able to investigate their interactions in the isotopes release process.

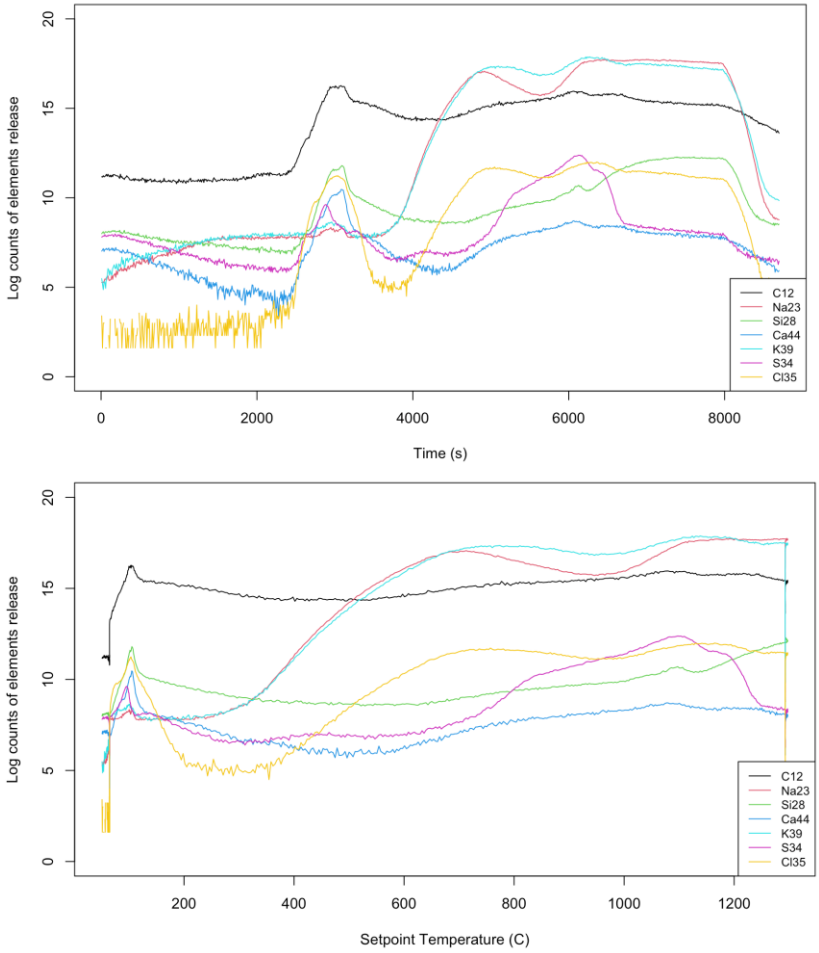

Figure 2. Summary Diagrams of Isotopes

Random forest modeling

Modeling

Random forest is a decision tree-based ensemble learning method that performs both classification and regression tasks. For classification, the output of the random forest is the class selected by most trees (the mode). For regression, the mean or average prediction of the individual trees is returned.

Random forest is usually treated as a supervised learning method; it uses a training dataset with known labels to train the model then performs prediction with the trained model on the test data. However, if a dissimilarity matrix (guideline metrics that include the independent variables like time, total mass, kinetic behavior, etc.) can be produced using Random Forest, we will be able to implement Random Forest on unsupervised learning, meaning that the model will assign labels to the dataset according to metrics in the dissimilarity matrix without being prespecified. Then the results will be similar to the clustering (another commonly used unsupervised learning method) results, but with a random forest model.

The goal of this data analysis is to identify fundamental patterns in the dataset related to chemical reactions as a function of time. The random forest can achieve this goal by clustering isotopes into different groups based on its clustering algorithm. For example, in a simple case, when time is less than 2000s, the total mass of the individual element is less than 20mg, isotopes whose release count is greater than 10000 will be clustered in the same group. After clustering isotopes, we can investigate each clustered group of isotopes that share similar properties. Furthermore, we could keep utilizing the potentially relevant predictors to improve the results.

Result

The first approach is to use random forest methods for classification. This is intended for calculating the dissimilarity matrix (or distance matrix, or proximity matrix). If the two isotopes are classed together, it means that the distance between them is very close, which implies that they should be clustered into the same group. Figure 2 shows that the isotopes are being clustered into four groups [9]. However, there are not enough samples, we can observe a large number of variations in Figure 3, in which case we are not able to draw robust statistical conclusions.

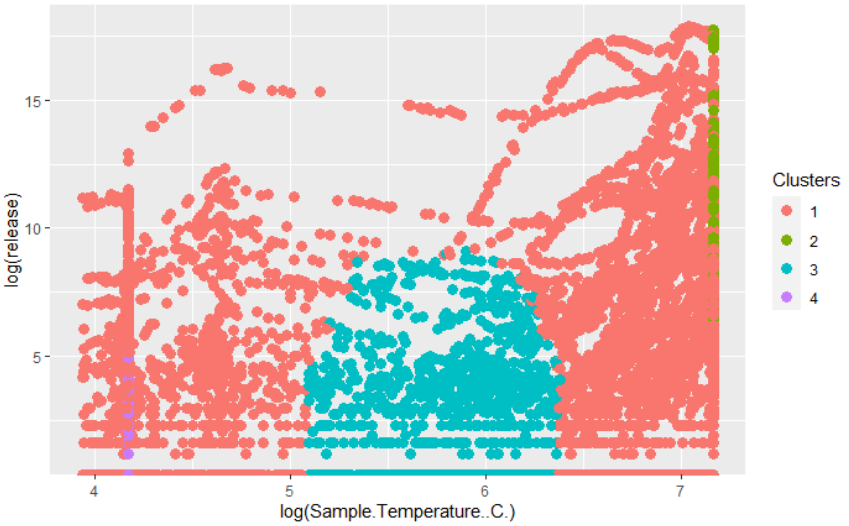

Figure 2. Cluster Assigned for Isotopes with Temperature vs. Release

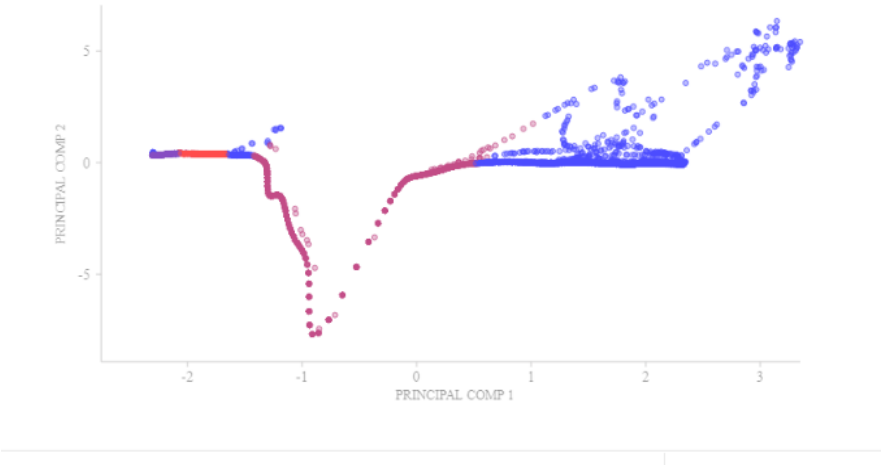

Figure 3. Cluster Assigned for Isotopes with PCA

A potential method to solve this issue is to calculate the weighted average of the clusters assigned for each isotope, then compare the weighted average clusters to determine which isotopes are being released together during the combustion, but we cannot assume the conclusion is statistically significant.

Ruptures

Change Point Detection

Another approach is to treat this dataset as a time-series dataset [8]. We only consider the release of isotopes and time in this case. We treat the release and time of each isotope separately and use the Ruptures package in Python to detect the change point of the release of each isotope.

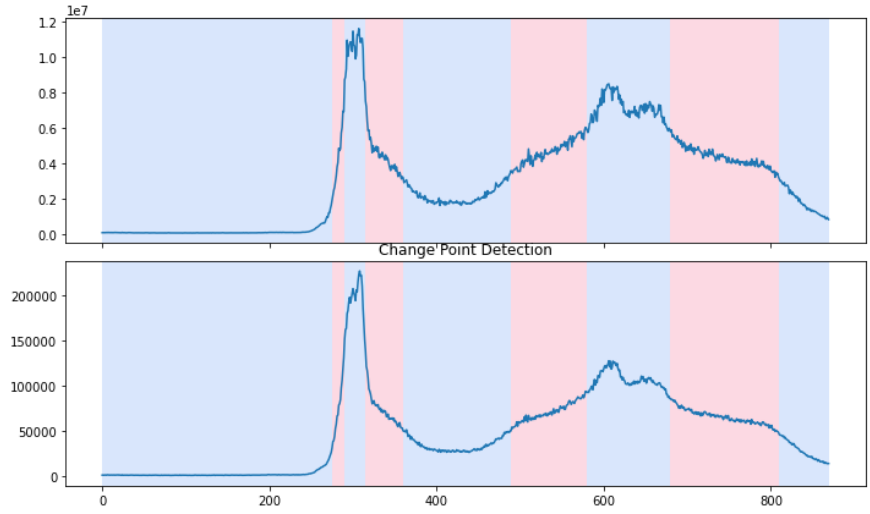

Figure 4. Example of isotopes that have similar change points

From Figure 4 we can see that C12 and C13 have a very similar distribution of release and thus they have the same change points, which suggests that C12 and C13 are released together during the combustion, so we should group C12 and C13 together [7]. By using this technique, we can compare the isotopes pairwise to conclude the pattern of isotopes released together under different times (or temperatures).

A major limitation of this method is that we are only considering time and release but ignoring other useful factors such as total mass and kinetic energy. Additionally, this method also omitted the information of release counts, but we could use logarithm transformation to keep the release on the same scale level.

Conclusion and discussion

In this project, we proposed clustering and change point detection to identify the fundamental release pattern of isotopes during combustion of biomass. By analyzing the data with random forest modeling, we can conclude that isotopes can be clustered to 4 groups with similar properties. The change point detection approach compares the change point of the release of each

isotope during the combustion pair-wisely. So, we can identify pairs of isotopes with the same change point. Both methods provided us with general ideas and results of clustering. However, this result is not solid due to the limitation of sample sizes. Future studies can address this issue by collecting larger sample sizes and more relevant factors involved will provide more accurate clustering groups. Additionally, clustering methods such as spectral clustering and K-medoids can be performed to investigate how isotopes release are related for different chemical reactions during the combustion and heating process of organic fuels.

Acknowledgement

We take this opportunity to express my sincere gratitude to Prof. Bala, Dr. Thy, Dr Jenkins and TA Ben and Tianke for their valuable guidance and kind help throughout.

Reference

1. https://en.wikipedia.org/wiki/Inductively_coupled_plasma_mass_spectrometry
2. https://www.sciencedirect.com/journal/fuel
3. P. Thy, G.H. Barfod, A.M. Cole, E.L. Brown, B.M. Jenkins, C.E. Lesher, Trace metal release during wood pyrolysis, Fuel, Volume 203, 2017, Pages 548-556, ISSN 0016-2361
4. Weiqi Ji and Sili Deng, The Journal of Physical Chemistry A 2021 125 (4), 1082-1092, DOI: 10.1021/acs.jpca.0c09316
5. Navid Kardani, Mojtaba Hedayati Marzbali, Kalpit Shah & Annan Zhou (2021) Machine learning prediction of the conversion of lignocellulosic biomass during hydrothermal carbonization, Biofuels, DOI: 10.1080/17597269.2021.1894780
6. Aghaaminiha, M., Mehrani, R., Reza, T. et al. Comparison of machine learning methodologies for predicting kinetics of hydrothermal carbonization of selective biomass. Biomass Conv. Bioref. (2021).
7. https://techrando.com/2019/08/14/a-brief-introduction-to-change-point-detection-using-python/
8. https://centre-borelli.github.io/ruptures-docs/
9. http://gradientdescending.com/unsupervised-random-forest-example/

Statement of Group Contribution

This statement is to recognize each group member's contribution to this data analysis. Each member contributed jointly on the completion of the following: forming statistical approach, making presentation, writing project proposal, initial and final report. More specifically, Dingning Liu and Ziqin Wang contributed to the initial statistical analysis or approach such as transformed additive model, linear, Poisson, and logistic model for the detection of fundamental patterns with release counts of isotopes. Corresponding exploratory data analysis is conducted by Dingning Liu. Yutian Yang contributed to the statistical methods used for this data analysis such as random forest modeling for clustering isotope into groups of each different characteristic and a times series model using ruptures package in Python for the detection of change points of isotopes. Dingning Liu and Min Kim contributed to administrative tasks such as scheduling and holding zoom meetings for group discussions and document organization using a shared Google Folder for documents created by all group members. Min Kim was responsible for the overall communication with the teaching team and clients via zoom meeting and emails. More specific contributions on writing of the proposal and final report are listed below:

- Dingning Liu:
    - Proposal: Introduction, Data Collection, and Initial Statistical Analysis
    - Report: Introduction, Previous Work, Data Collection, Limitation, EDA, and Conclusion (Final)
- Min Kim:
    - Proposal: Scientific Challenges and Goals, and Data Collection, Initial Approach, Statistical Approach, The Goal of Data Analysis
    - Report: Scientific Background and Goals, Data Collection, Introduction, Scientific Challenges and Goals, and Data Collection
- Yutian Yang:
    - Proposal: Random Forest Modeling
    - Report: Limitation, Random Forest Modeling, Change Detection Analysis (Ruptures), Results, and Conclusion (Final).
- Ziqin Wang:
    - Proposal: Initial Data Analysis (linear, logistic and Poisson)
    - Report: Conclusion (Draft).

I agree with the Statement of Group Contribution.

Name: Min Kim
Name: Dingning Liu
Name: Yutian Yang
Name: Ziqin Wang