Science of the Total Environment 998 (2025) 180252

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

The potential of decision tree application in threshold analysis of hazardous volatile organic compound release from biochar: Implications for environmental risk assessment

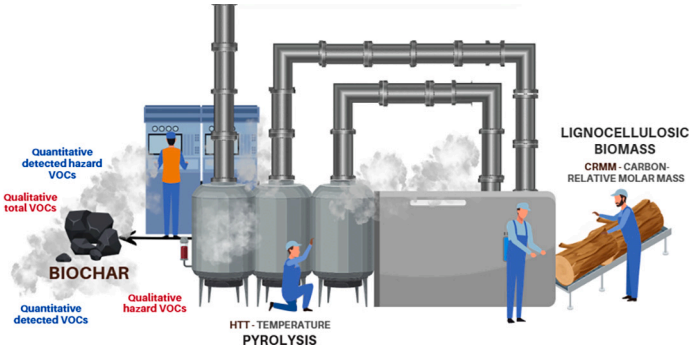
Ewa Syguła [a], Kamila Piasecka [a], Jacek Łyczko [b], Andrzej Białowiec [a,*]

[a] Department of Applied Bioeconomy, Wroclaw University of Environmental and Life Sciences, Wroclaw 51-630, Poland
[b] Faculty of Biotechnology and Food Science, Wroclaw University of Environmental and Life Sciences, 50-375 Wroclaw, Poland

HIGHLIGHTS

- CRMM and HTT parameters identified as hazardous VOC formation indicators
- Linear models insufficient to assess VOC threat levels from biochar
- Decision tree regression captured thresholds in hazardous VOC release.
- Novel use of decision trees to model biochar safety and environmental risk

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Toxic compounds
Biomass low pyrolysis
Biochar
CRMM
Environmental risk assessment
Machine learning

ABSTRACT

The release of hazardous volatile organic compounds (HVOCs) from biochar poses a potential threat to both human health and the environment. This study investigates how low pyrolysis temperature (HTT) and the chemical characteristics of lignocellulosic biomass, expressed as the carbon-relative molar mass (CRMM) index ($g \cdot mol^{-1}$), influence the release of hazardous volatile compounds from biochar. A total of 204 biochar samples were produced from pure biomass components: lignin (L), cellulose (C), hemicellulose (H) and their mixtures, and analyzed using stepwise multiple regression, agglomerative hierarchical clustering, principal component analysis (PCA), and regression decision trees. The analysis included both qualitative and quantitative assessments of HVOCs. Qualitative results indicated the presence of HVOCs, but quantitative patterns were nonlinear and inconsistent with expected trends. Decision tree analysis showed that the highest release occurred either at very high CRMM values (>34 g mol^{-1}) or at low-temperature conditions (200–325 °C) with low CRMM. The lowest release was associated with temperatures 350–425 °C and elevated CRMM. The results emphasize the need to consider both process and substrate parameters when assessing release risks, providing a new approach to evaluating the safety of biochar in the context of environmental and human health protection. To our knowledge, this is the first study to apply decision tree regression to model HVOC release from biochar based on CRMM and pyrolysis conditions.

* Corresponding author.
E-mail addresses: jacek.lyczko@upwr.edu.pl (J. Łyczko), andrzej.bialowiec@upwr.edu.pl (A. Białowiec).

https://doi.org/10.1016/j.scitotenv.2025.180252
Received 27 May 2025; Received in revised form 10 July 2025; Accepted 8 August 2025
Available online 14 August 2025

1. Introduction

1.1. The role of VOCs in biochar and their sources

Biochar has been extensively studied in recent years, not only for its beneficial properties but also for the presence of potential impurities such as VOCs (Zheng et al., 2018). Research to date has focused on analyzing the properties of biochar due to variable substrate and process conditions. This topic has been extensively studied by researchers worldwide, and biochar has gained popularity over the years, focusing on its valuable properties (Wu et al., 2019). In recent years, more and more research papers have been reported to analyze biochar impurities (Zheng et al., 2018). More than a decade ago, Spokas et al. found that approximately 20 % of studies on biochar indicated a negative effect after its application in the environment (Spokas et al., 2011). Possible contaminants that could cause such negative effects are organic pollutants such as dioxins, polyaromatic hydrocarbons (PAHs) (Sobol et al., 2023), or volatile organic compounds (VOCs), which are gaining importance. The difference between these pollutants in biochar is that dioxins and PAHs are captured in global biochar quality systems, which underscores the risks of biochar application to the environment, while VOCs have still not been standardized (Zheng et al., 2018). The high VOC content of biochar is highly phytotoxic (Kim et al., 2021). VOCs are mobile compounds, meaning that they can be easily leached into the environment or absorbed by living organisms (Han et al., 2022).

1.2. Impact of pyrolysis conditions on VOCs presence in biochar

Recondensation of gases and liquids during pyrolysis has recently been identified as a key source of VOCs in biochar (Buss et al., 2015). If the biochar production process is not properly controlled, the VOC content can exceed 100 $\mu g \cdot g^{-1}$, which is associated with their phytotoxicity (e.g., complete inhibition of cress germination) (Han et al., 2022). These reactions can lead to the decomposition of primary products into more complex chemical compounds, as well as their polymerization or cracking. Pyrolysis triggers various reaction pathways and the release of volatile compounds (Liu et al., 2019). The rate and efficiency of this process depend on retention time and temperature. A longer retention time promotes greater carbonization and energy densification, resulting in lower volatile content (Supriyanto et al., 2020). The rate of temperature increase is one of the fundamental parameters in pyrolysis, determining the type and efficiency of the process. Higher heating rates generate greater amounts of volatile and liquid products. Both fast and low pyrolysis produce volatiles, though in different proportions. In fast pyrolysis, low-molecular-weight compounds such as acetic acid, furfural, acetone, phenol, and methanol are commonly released due to the rapid breakdown of biomass polymers (Demirbas, 2006). In contrast, low pyrolysis tends to favor the formation of heavier compounds like long-chain hydrocarbons, aromatics (e.g., toluene, benzene), and polycyclic aromatic hydrocarbons (PAHs), resulting from secondary reactions occurring during prolonged residence time (Ben and Ragauskas, 2013). Regrouping reactions that create a more stable matrix occur primarily during low pyrolysis, whereas during fast pyrolysis, multiple chemical bonds are broken simultaneously before rearrangement can take place. It is preferable for volatiles formed during the process to remain in the chamber for as short a duration as possible to obtain a high-quality product (Buss and Mašek, 2014). Biomass is primarily composed of lignin, cellulose, and hemicellulose, which degrade differently and release distinct VOCs.

1.2.1. Lignin

Lignin decomposes and transforms structurally under thermal conditions. Its main decomposition occurs between 200 and 450 °C, peaking at 360–400 °C. Volatile compounds form due to unstable propyl chains and bonds between monomers and methoxyl groups. Thermal processing restructures the lignin carbon skeleton into polycyclic aromatics. The volatile gases released by these reactions are mostly small, non-condensable gases. During the thermal degradation of lignin, compounds such as guaiacol, syringol, vanillin, phenol, and catechol are commonly released (Kibet et al., 2012). In an inert atmosphere, the benzene rings that are the product of the thermal process become very stable, leading to a tendency for their concentration to increase in the lignin residue during the reaction. It is worth noting that most of the aromatic rings that form carbon already exist in the lignin, which makes it easier to understand the phenomena occurring during the process (Balat, 2008).

1.2.2. Cellulose

Cellulose undergoes thermal decomposition in specific temperature ranges. Its main decomposition occurs between 300 and 390 °C, peaking at 330–370 °C (Syguła et al., 2024). This range yields mainly condensed organics, often exceeding 70 % by weight. Above 400 °C, the cellulose residue becomes increasingly aromatic. These reactions release VOCs such as aldehydes, ketones, alcohols, and acids (Ponomarev and Ershov, 2014). These compounds are released during each step of cellulose conversion, and their composition and number can change as the temperature increases and the reaction progresses. As temperature increases, chemical reactions affect the composition of cellulose conversion products and their chemical properties. At lower temperatures, depolymerization processes result in the release of light volatile compounds, which affect the rate of combustion and gas production (Usino et al., 2020).

1.2.3. Hemicellulose

Hemicellulose, despite the complexity of its composition, undergoes the main conversion in the temperature range from 200 to 350 °C. Weight loss reaches a maximum at around 310 °C, with an apparent peak at around 260 °C. Conversion of xylan, one of the main components of hemicellulose, shows two peaks around 230 and 290 °C, suggesting a displacement of the maximum relative to glucomannan by about 20–30 °C. During the conversion of hemicellulose, volatile organic compounds are released, including a variety of substances such as aldehydes, ketones, alcohols, carboxylic acids, and other organic compounds. Thermal reactions lead to the regrouping of residues, especially in processes above 350 °C (Syguła et al., 2024).

1.3. Impact of biomass composition on VOCs formation on biochar surface

Many studies on biochar focus on plant species and basic traits like moisture, organics, or ash. This may lead to misleading comparisons when identical feedstocks are grown under varying conditions (Alburquerque et al., 2014). The simplest method to parameterize substrate properties is to determine the content of structural components that play a key role in the subsequent formation of VOCs. Syguła et al. presented a parameter, such as carbon relative molar mass (CRMM), unifying all biomass parameters such as moisture content, volatiles, ash content, C, H, N, S, and O content (Syguła et al., 2025a). CRMM correlates with structural components critical to VOC formation in biochar (Syguła et al., 2024). Inorganic components catalyze biomass decomposition, carbon formation, and volatile yield during pyrolysis (Zhou et al., 2020). Zhang et al. showed that biochar's inorganic content affects VOC adsorption mechanisms. Depending on the inorganic concentration, biochars showed different adsorption behavior. High inorganic biochars favored physical adsorption, while lower inorganic content promoted stronger chemical adsorption (Zhang et al., 2017a). In

contrast, the organic fraction of biomass accounts for up to 90 % of its composition, depending on morphological structure, origin, and environmental conditions (Tursi, 2019).

1.4. Effect of ambient conditions on VOCs release

However, as indicated in Section 1.3, opinions on the effects of using biochar in the broader industry, environment, or agriculture vary widely despite many years of research on the material. Beyond biomass and process parameters, environmental conditions also influence biochar behavior (Kazemi Shariat Panahi et al., 2020). VOCs are compounds with high vapor pressure and moderate boiling points (50–250 °C), easily transitioning to a gas phase. VOCs originate from both anthropogenic and biogenic sources, the latter including plants, animals, and microorganisms (Zheng et al., 2018). Most of the VOCs in the atmosphere come from natural releases, called biogenic, emitted by plants, animals, or microorganisms. Anthropogenic sources emit about 142 million tons of VOCs annually, mainly through fuel use, solvent use, biofuels, and biomass burning (Evuti, 2013). Their widespread presence highlights the importance of environmental factors in VOC dynamics. Ambient temperature, humidity, pressure, concentration of VOCs, and contact time will all affect the interactions of biochar with the environment (Xiang et al., 2020). As Zhu et al. point out, high ambient temperature promotes the mobility of VOCs, which affects the lower sorption capacity of biochar. High humidity can compete with polar VOCs for sorption sites on hydrophilic biochar (Zhu et al., 2017). Recent studies focus on how these parameters affect VOC adsorption by biochar. Adsorption also depends on biochar pore size, VOC molecular size, and polarity. Thus, these factors likely influence VOCs release from biochar as well (Sadegh et al., 2024).

1.5. Standardization of VOC analysis in biochar — challenges and developments

Recently, the European Union has given direction to the use of biochar as a carbon sequestration product to move toward a closed-loop economy and sustainable development (Štrubelj, 2022). Biochar certification schemes include the European Biochar Certificate (EBC), the International Biochar Initiative (IBI), and the British Biochar Foundation (BBF). These systems are voluntary and lack legal enforceability in EU countries (Han et al., 2023). In addition, each country individually sets standards and rules for the use of biochar. Standardizing VOC analysis in biochar remains a key challenge alongside defining quality criteria. Comparability requires harmonized analytical standards. VOCs in biochar have been studied using varied instruments and storage protocols. All of these factors can affect the final result of an experiment (Dutta et al., 2017).

Laboratories analyzing biochar often adopt standards for soils, waste, or fuels to analyze this material, which indicates the need for standardization. Validated methods not tailored to biochar may yield misleading results. In response, the ISO/TC 238 technical committee began developing standards and guidelines for biochar. The goal of this work is to develop standards and guidelines for the production, properties, methods of analysis, and application of biochar. Ongoing work by ISO/TC 238 aims to establish dedicated standards to support safe biochar applications (Dutta et al., 2017; Bachmann et al., 2016). Additionally, while several regulatory frameworks define permissible exposure limits for specific VOCs in the air (OSHA, EPA), none currently establish quantitative limits for VOC content in solid biochar materials ($mg{\cdot}kg^{-1}$). This regulatory gap limits the ability to assess environmental or occupational risk based on VOC concentration in biochar. Hence, our study provides valuable insights into potential release levels of hazardous VOCs, which could inform future development of quality criteria and regulatory standards for biochar use in industrial and environmental applications. In parallel, there is also a need to develop predictive models, both quantitative and qualitative, that would support the optimization of biochar production by identifying the pyrolysis conditions and feedstock characteristics under which the release of hazardous HVOCs may occur. This article addresses this gap by proposing a modeling approach aimed at predicting such risks and supporting safer biochar applications.

1.6. Predictive modeling as a tool for risk assessment

Standardized methods for assessing VOC release from biochar remain lacking, which limits both regulatory action and practical risk assessment. This underscores the need for accurate and interpretable predictive tools (Dutta et al., 2017). Although nonlinear models are increasingly applied in environmental science, biochar modeling is still dominated by linear regression and deterministic approaches, with decision tree models being rarely used despite their interpretability and threshold-detection capabilities (Zhao et al., 2022). Moreover, nonlinear models can underperform when applied to environmental data outside their training domain, underscoring the need for interpretable and robust methods (Hsieh, 2023). Despite their proven value in threshold detection and pattern recognition, decision trees remain underutilized in the field. Machine learning (ML) techniques are increasingly applied in biochar research to improve production efficiency and characterization. Wang et al. (2024) suggest that ML can reduce the time and labor demands of experimental trials, offering practical advantages for process optimization (Wang et al., 2024). Oral et al. emphasize ML's ability to detect hidden patterns and make accurate predictions in biochar utilization (Oral et al., 2024). Among machine learning methods, decision tree regression offers a unique balance between nonlinearity and interpretability, making it particularly suitable for identifying release thresholds of HVOCs, something that traditional nonlinear models in our study failed to capture effectively.

To our knowledge, this is the first study to use decision tree regression to model the quantitative release of HVOCs from biochar based on both pyrolysis conditions and biomass composition. This approach not only captures non-linear relationships but also identifies interpretable thresholds necessary for practical risk assessment. Unlike more complex models such as random forests or support vector machines, decision trees offer a transparent structure that supports clear communication of risk categories.

1.7. Bibliometric analysis of VOC and biochar research and methodological novelty

According to the renowned international bibliographic databases Scopus and Web of Science (WoS), biochar has appeared around 46k times in scientific publications over the last 25 years. It is the subject of research by scientists from many fields, which confirms its usefulness and the positive effect of its use, especially in environmental sciences. Volatile organic compounds are also the subject of research by scientists, appearing 65 thousand times in scientific publications over the last 25 years. Research on VOCs and biochar has been confirmed by 373 (WoS) and 407 (Scopus) scientific papers. This indicates a niche that is beginning to gain scientific importance. Among these papers, scientists are trying in various ways to discover the significance of the occurrence of these substances in biochar or other derivatives. Section 1.5 indicates the lack of standardization in the presence of volatile organic compounds (VOCs) in biochar and the ongoing work of the relevant technical committee on a document that is to define guidelines for its safe use. Although methods such as linear modeling, multiple regression, cluster analysis, principal component analysis (PCA), and decision tree models are well established in environmental data analysis, their combined application to the study of volatile organic compounds (VOCs) in biochar remains extremely limited. To explore this gap, a bibliometric analysis was performed using Web of Science data covering the period from 2000 to 2025. All applied analysis methods were listed separately for biochar, VOCs, and their combination. While these methods, when

used separately with biochar or VOCs, resulted in a larger number of publications, when used together (VOCs and biochar), only 5 records were identified: one each for multiple regression, principal component analysis (PCA), and cluster analysis, and two for linear models. Importantly, no results were reported regarding the use of decision trees in the analysis of VOCs in biochar (Fig. 1).

While methods like regression and PCA are widely used in general environmental research, their implementation in VOC-focused biochar studies remains scarce. To the best of our knowledge, this study is the first to apply decision tree modeling to explore VOC release patterns in biochar, marking a novel contribution to both methodological practice and environmental risk assessment.

2. Materials and methods

2.1. Materials

The study used 12 chemically prepared mix samples based on the main structural components of lignocellulosic biomass: lignin (alkali lignin, CAS 8068-05-1), cellulose (microcrystalline cellulose, CAS 9004-34-6), and hemicellulose (xylan, CAS 9014-63-5). The components used in the study were purchased from certified suppliers providing product safety data sheets, including Sigma-Aldrich (St. Louis, MO, USA). Nine samples were mixtures reflecting the typical weight proportions of these components in natural wood biomass, while the remaining three consisted of pure substances (L, C, or H). All samples were prepared in powdered form, with a standardized mass of 50 g.

The Ferre diagram (Fig. 2, A) shows the structural composition of the materials included in the lignocellulosic biomass based on the literature (Tursi, 2019; Komilis and Ham, 2003; Zhang et al., 2023; Cai et al., 2017; Nawirska and Kwaśniewska, 2004; Kataki et al., 2015; Orejuela-Escobar et al., 2021; Menon and Rao, 2012; Taylor et al., 2019). The highest concentration of these materials is indicated, highlighting the most common proportions of structural components. The diagram Ferre (Fig. 2, B) shows the content of structural components in the research materials and confirms the validity of choosing these materials as a representative lignocellulosic biomass (Fig. 2). For data validation, one real biomass material was selected — pine wood and hemp straw, which is also indicated in the figure presenting the research materials.

To quantitatively represent the composition of the mixtures and introduce an independent variable describing the structural component contribution, the CRMM index (Carbon-Relative Ash-Free Molar Mass) was used. This index was developed by the authors in a previous study (Syguła et al., 2025a) and is calculated based on CHNSO elemental analysis. It enables direct comparison of materials regardless of specific component labels and serves as a robust input for statistical and modeling analyses. The full calculation procedure, theoretical rationale, and example values (21–36 $g \cdot mol^{-1}$) are described in detail in Syguła et al., where the CRMM index was first introduced. In the validation biomass, pine wood had a CRMM value of 25.84 $g \cdot mol^{-1}$, while hemp straw had a CRMM of 27.18 $g \cdot mol^{-1}$ (Syguła et al., 2025a).

2.2. Methods

2.2.1. Biochar production

Biochar was produced in sealed metal containers (57 × 57 × 82 mm) with an ~5 mm vent hole in the lid to equalize internal pressure. Twelve chemical mixtures and three types of lignocellulosic biomass were weighed using an analytical balance (±0.0001 g). The samples were dried at 105 °C for 24 h in a laboratory oven (WAMED KBC-65 W, Warsaw, Poland), then placed in a muffle furnace (SNOL 8.1/1100, Utena, Lithuania). Heating began 5 min after initiating the inert gas flow. To ensure an oxygen-free atmosphere during pyrolysis, the reactor chamber was purged with carbon dioxide at a flow rate of 10 $dm^3 \cdot h^{-1}$,

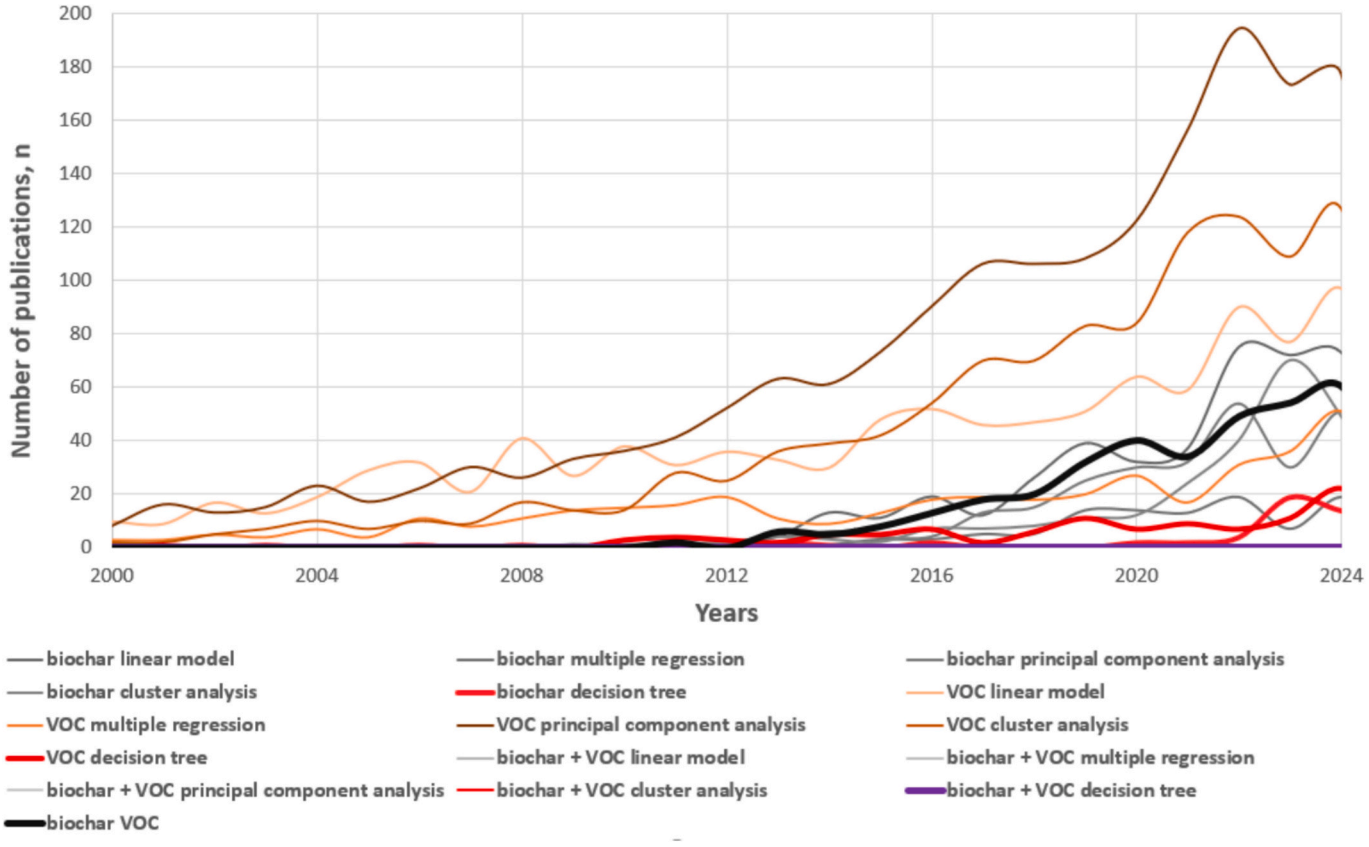


Fig. 1. Number of publications in Web of Science and Scopus (2000–2024) using selected statistical methods applied to biochar, VOC, and their combination.

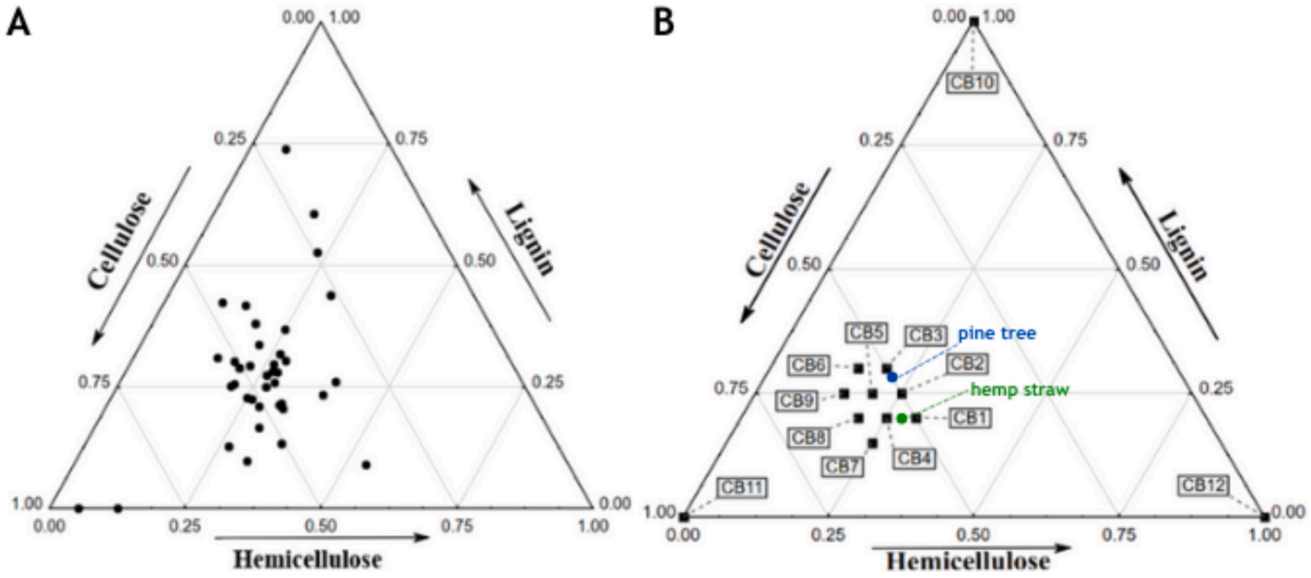


Fig. 2. Comparison of the structural components content (lignin, cellulose, hemicellulose) in lignocellulosic biomass (A) and mix samples (B) with validation sample.

with a heating rate of 50 °C·min^{-1}. After reaching the target temperature, the samples were held for 60 min before cooling. Once cooled, the resulting biochar samples were stored in hermetically sealed vacuum bags to preserve their properties. Biochar was produced at temperatures ranging from 200 °C to 600 °C, at 25 °C intervals (17 temperatures), resulting in a total of 204 biochar samples. For validation purposes, two real biomass types, pine wood and hemp straw, were selected. Biochar from these materials was produced under identical process conditions, but only within the 200–425 °C (9 temperatures) range. The upper limit was set due to the lack of detectable VOCs above 425 °C, as described in detail in Section 2.2.2. This resulted in 18 additional biochar samples.

2.2.2. GC–MS analysis of VOCs

The analysis of volatile organic compounds (VOCs) was performed using a gas chromatograph coupled with a mass spectrometer, Shimadzu GCMS-QP2020 Plus (Shimadzu, Kyoto, Japan), equipped with a ZB-5Msi capillary column (30 m × 0.25 mm × 0.25 μm; Phenomenex, Torrance, CA, USA).

Approximately 0.5 g (±0.01 g) of each powdered biochar sample was weighed into 20 ml glass vials in triplicate, under conditions that minimized external contamination. Each vial was supplemented with 0.5 ml of distilled water and a piece of filter paper (2 × 2 cm) impregnated with 20 μg of internal standard (caryophyllene, 1 mg·ml^{-1}, 3 μg·cm^{-2}). Vials were sealed and incubated for 10 min at 45 °C. VOCs were extracted from the headspace using solid-phase microextraction (SPME Arrow, DVB/C-WR/PDMS, 1.10 mm) for 30 min at the same temperature.

The analytes were desorbed in the GC injection port at 50 °C for 3 min. The oven temperature program was as follows: initial temperature of 50 °C (held for 2 min), ramped at 3 °C·min^{-1} to 180 °C, then at 20 °C·min^{-1} to 270 °C, and held at this final temperature for 5 min. Helium was used as the carrier gas at a constant flow rate of 1 ml·min^{-1}. The MS detector operated with a source and interface temperature of 250 °C, scanning in the 40–400 *m/z* range. The internal standard was used to enable quantitative analysis of volatile compounds.

The analysis followed a semi-quantitative approach based on an internal standard (caryophyllene). Although individual LOD and LOQ values were not determined, the method enabled consistent comparison of relative VOC concentrations across all samples. Final concentrations were calculated in μg·kg^{-1} using peak area ratios normalized to the internal standard. All biochar samples were analyzed within a week of production and stored in sealed polyethylene bags to preserve their properties. Samples of biochar produced at temperatures above 425 °C did not detect any volatile organic compounds, but analyses of all samples were performed to confirm this trend. To keep the results comparable, the extraction conditions were not altered to avoid artificially increasing the release of compounds under unrepresentative conditions. In total, 101 different VOCs were identified in the biochar samples produced from chemically defined biomass mixtures (mix samples). All identified compounds were subsequently checked against Annex I of Regulation (EC) No 1272/2008. Based on this comparison, 10 VOCs were classified as hazardous to human health or the environment and were selected for further quantitative analysis and modeling. For the validation biomass (validation samples — pine wood and hemp straw), a total of 33 different VOCs were identified, of which 6 were classified as hazardous. In this study, we focus on volatile organic compounds that are retained in the solid biochar matrix and can be released after production. All of the VOCs data are publicly available in Syguła et al. (2025b).

2.2.3. Statistical analysis

Statistical analysis was carried out on the results of 120 biochar samples (temperatures from 200 to 425 °C) in three replicates. The independent variables of the experiment were the temperature of the pyrolysis process (HTT) and the CRMM parameter to characterize the properties of the substrate. A matrix of 360 observations (n), described by HTT and CRMM values (independent variables X_1 and X_2) and qualitative and quantitative variables describing the release of VOCs from biochar (Y_i dependent variables), was used for the analysis. The dependent variables of the experiment were qualitative-detected VOCs (Y_1) and detected HVOCs (Y_2), given in units. Qualitative analysis provides information about the possibility of occurrence of the phenomenon, but not about its scale. The dependent variables, total VOCs (Y_3) and hazard total HVOCs (Y_4) in mg·kg^{-1} were used to determine the scale of the hazard. HVOCs were extracted from the stream of total VOCs based on Annex 1 of European Union regulation EC No. 1272/2008, in which hazardous substances for humans and the environment are indicated. This resulted in 10 HVOCs (Y_5 to Y_{15}) identified in the test samples (Table 1). For model validation, an additional 18 biochar samples (pine wood and hemp straw) were produced under identical process conditions and analyzed using the same procedure.

2.2.4. Modeling data

First, the linear dependence of the individual aggregated 4 variables

Table 1
Data modeling procedure for qualitative and quantitative variables for all biochar samples produced at temperatures ranging from 200 to 425 °C with varying CRMM values.

Dependent variables Y_i		Independent variables X_p	Type of analysis	Method
Variables aggregated based on the occurrence of hazardous chemical compounds in the study	Y_1 — detected VOC, n Y_2 — hazard VOC n Y_3 — total VOC $mg \cdot kg^{-1}$ Y_4 — hazard VOC $mg \cdot kg^{-1}$	X_1 — HTT X_2 — CRMM	Quantitative and qualitative analysis	Multiple regression (stepwise)
Hazardous compounds contributing to Y_2 and Y_4 and considered in Y_1 and Y_3	Y_5–…Y_{15}		Qualitative analysis	PCA Cluster analysis Decision trees

(Y_1–Y_4) on HTT and CRMM was verified. For this purpose, one of the multivariate regression methods known as stepwise regression was used. This method involves the gradual addition or removal of variables from the model, taking into account their statistical significance. Because there are two independent variables in the study in equation, backward stepwise regression was used. Stepwise regression is based on classical linear regression. The initial model was a regression function including all dependent variables. In the possibility of determining which variable has a greater impact on the dependent variable, the variables were standardized.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \in$$

Y_i — standardized dependent variable, selected for the modelβ_0 — standardized intercept,$\beta_1 - \beta_p$ — standardized regression coefficients, $X_1 - X_p$ — standardized independent variables,$p = 1, 2, \ldots$ k. k corresponds to the number of independent variables included in the model.

To assess the significance of the independent variables, a *t*-test was used by comparing the estimated regression coefficients with their standard errors. The standard error was calculated based on the residual sum of squares and the degrees of freedom. Residuals were also evaluated graphically using residual plots and Q–Q plots to assess model assumptions and potential deviations from normality. If the *p*-value of a given variable is greater than the significance threshold (set at $p = 0.05$), the variable is removed from the model. Only statistically significant variables were retained in the model, and standardized regression coefficients β were calculated to compare the relative impact of independent variables on the dependent variable Y. Model quality was assessed using the coefficient of determination (R^2) and the root mean square error (RMSE). Higher R^2 values indicate better model fit, while lower RMSE values reflect smaller deviations between predicted and actual values.

Subsequent analysis focused on the quantitative assessment of selected HVOCs. To identify variables that significantly influence specific compound groups, Principal Component Analysis (PCA) was applied. This technique reduces the dimensionality of the dataset by transforming the original variables into a set of uncorrelated principal components while preserving most of the total variance. PCA enables the identification of variables that contribute most strongly to the formation of individual components. Each principal component corresponds to an eigenvalue of the data covariance matrix, reflecting the amount of variance explained. The total variance is equal to the sum of all eigenvalues. Due to differences in units among the variables, the data were standardized before analysis. The number of retained components was determined using the Kaiser criterion (eigenvalues > 1) and a scree plot, with the point of inflection indicating the optimal number of components to retain. Components selected before this point were considered significant and further interpreted.

Cluster analysis was conducted for selected hazardous volatile organic compounds (HVOCs) to identify groups of compounds with similar properties. This exploratory technique classifies multivariate objects into clusters based on similarity, without prior assumptions about their structure. Objects are iteratively grouped using a chosen distance metric (Euclidean), forming increasingly larger clusters until a complete hierarchical structure is obtained. The results were visualized using a dendrogram, illustrating the sequence of cluster mergers.

To verify the influence of selected independent variables, such as pyrolysis HTT and the CRMM index, a regression model based on the Decision Tree Regression algorithm was applied. This method is particularly useful for analyzing nonlinear data and is more robust to outliers compared to traditional linear models (e.g., regression, correlation, PCA). Additionally, it can handle both quantitative and qualitative data without the need for variable standardization. The decision tree is composed of decision nodes, branches, and leaves that classify the data into predicted ranges of the dependent variable. The model is trained by iteratively splitting the dataset at points that maximize the difference between subsets, often using criteria such as minimizing the root mean square error. In the decision tree model, each node includes the mean (Mu) and the variance (Var) of the target variable, allowing for basic interpretation of prediction uncertainty and variability within individual splits. In this study, the model enabled the identification of variables with the greatest influence on the content of selected HVOCs released from biochar. Extreme HVOC concentration values were handled effectively, which grouped them into distinct terminal nodes based on CRMM and HTT thresholds. This allowed for localized variance to be captured and interpreted without data transformation or outlier removal.

3. Results and discussion

3.1. Stepwise multiple regression

This study aimed to model the effect of pyrolysis temperature and biomass CRMM on VOC release from biochar. HVOCs were identified based on Annex I of Regulation (EC) No. 1272/2008. This document defines the hazard classes of individual substances, ensuring a high level of protection for human health and the environment. GC–MS analysis was conducted for 204 biochar samples in triplicate. No VOCs were detected in biochar produced above 425 °C. The non-detectable levels of VOCs observed in biochars above 425 °C can be attributed to advanced carbonization processes and structural transformations. At these elevated temperatures, most volatile precursors are thermally degraded or restructured into more stable aromatic carbon forms. Additionally, progressive pore closure and carbon matrix densification may reduce both the formation and subsequent release of VOCs, as surface functionalities and reactive sites are significantly diminished (Wang et al., 2024; Zhang et al., 2017b). Therefore, 120 biochar samples (in triplicate) were selected for model development. For each sample, an individual VOC profile was determined, and the results were summed based on the total number of detected VOCs and the number of identified HVOCs, representing a qualitative assessment of compound release. Stepwise multiple regression analysis was used to evaluate the influence of temperature and CRMM on the qualitative and quantitative release of VOCs from biochar. For model validation, 9 biochar samples from pine wood and 9 from hemp straw were produced (200–425 °C) and analyzed using the same procedure as for the mix samples.

In the case of the qualitative analysis, the model showed a good fit, with $R^2 = 0.87$ for detected VOCs and $R^2 = 0.61$ for detected HVOCs

(Table 2). Lower RMSE values indicate higher predictive accuracy of the model. The negative β coefficients for temperature suggest that higher pyrolysis temperatures lead to a reduction in both detected VOCs and detected HVOCs. Similarly, the β coefficients for CRMM indicate that higher CRMM values of biomass are associated with lower levels of detected VOCs and HVOCs. The stronger effect observed for detected HVOCs compared to total detected VOCs suggests that biomass with higher CRMM may result in fewer compounds classified as hazardous. Wing linear or inverse linear relationships between pyrolysis temperature and total VOC release. For example, Spokas et al. demonstrated that higher pyrolysis temperatures significantly reduce the concentration of adsorbed VOCs on biochar surfaces (Spokas et al., 2011). Similarly, Tripathi et al. reported that VOC release decreases with increasing temperature, following a relatively linear pattern (Tripathi et al., 2016). Previous studies, such as Hernowo et al., have proposed detailed pseudo-kinetic mathematical models to predict the yield of specific volatile compounds formed during biomass pyrolysis. Their approach focused on simulating the formation of 45 gaseous and liquid compounds based on temperature-dependent activation energies and biomass type, offering a valuable tool for process design. However, their models rely on compound-specific kinetic constants, which are difficult to obtain and limit their general applicability (Hernowo et al., 2022). Ghidotti et al. examined the relationship between total VOC signal intensity and biochar properties (H/C and VM), presenting linear correlations. Although they did not build predictive models, their results confirmed that higher carbonization corresponds to lower VOC content. However, their approach was qualitative, based on normalized peak areas rather than absolute VOC concentrations (Ghidotti et al., 2017).

The quantitative analysis revealed different trends compared to the qualitative analysis. The model fit for total VOCs was $R^2 = 0.11$, and for HVOCs $R^2 = 0.27$, indicating weak predictive performance (Table 2). The high RMSE values suggest a considerable prediction error, particularly for total VOCs. Therefore, the β coefficients were not interpreted in the context of the quantitative analysis. This is an important observation, as the study focuses on assessing the potential hazards associated with the production and use of biochar.

To test the robustness of the models, validation was performed using an independent data set based on two real biomasses (pine wood and hemp straw). For the qualitative analysis, the results confirmed high consistency: R^2 values remained high (0.72 for detected VOCs and 0.66 for detected HVOCs) and the direction of the relationship (positive or negative) between predictors and response variables was preserved. In contrast, quantitative models showed increased R^2 values (up to 0.50 for total VOCs and 0.39 for HVOCs), but RMSE values increased significantly (from 261 to 45,418 for total VOCs and from 141 to 2337 for HVOCs), indicating poor prediction accuracy (Table 3). These results confirm the limited usefulness of linear regression for modeling release rates and justify the use of machine learning approaches in the next steps of the analysis.

Qualitative analysis provides information about the possibility of an occurrence, but not about its magnitude. This suggests that there may be situations in which high pyrolysis temperature or a high CRMM value does not guarantee the safety of biochar production or application. Annex I of the European Union Regulation EC No. 1272/2008 provides specific threshold values ($mg \cdot kg^{-1}$) for each substance, indicating their toxic effects on humans or the environment. Therefore, to verify the influence of process conditions and biomass properties on the potential negative impact of biochar on human health and the environment, the focus was shifted to quantitative analysis. For this purpose, alternative data analysis methods based on variability, similarity, and clustering were applied, as discussed in the following sections.

3.2. Principal component analysis PCA

To further investigate the potential negative impact of biochar on human health and the environment, principal component analysis (PCA) was conducted. The analysis focused on identified specific HVOCs: furfural, ethylbenzene, 4-heptanone, styrene, cyclohexanone, limonene, benzaldehyde, acetophenone, mequinol, and nonanoic acid. The sum of these compounds represents the HVOCs value ($mg \cdot kg^{-1}$). These substances are listed in Annex I of European Union Regulation EC No. 1272/2008 as hazardous to humans or the environment. To avoid statistical multicollinearity and ensure the independence of predictive variables, the analysis included only specific HVOCs, without combining them with total VOCs or overall HVOCs values. This approach avoids redundancy, as total VOCs inherently include HVOCs, which could otherwise lead to misleading statistical interpretations.

The PCA results allowed the identification of variables that contributed most to the variability in the dataset and clarified their interrelationships. PCA is commonly used in VOC studies as a method for providing clear insights into release patterns and the relationships between individual compounds (Delgado-Rodríguez et al., 2012; Geng et al., 2009). The first four principal components (PC1–PC4) explained 78.1 % of total variance, effectively capturing the VOC release patterns. The eigenvalue analysis revealed that PC1–PC4 had comparable values and gradually decreasing contributions to the total variance (Table 4). The absence of a dominant principal component suggests that each plays a meaningful role in describing the variability of VOC release.

To better visualize the relationships between variables, the principal components were presented in two separate plots: the first illustrating the relationship between PC1 and PC2, and the second between PC3 and PC4 (Fig. 2). The first two components (PC1 and PC2) jointly explain 47.87 % of the total variance, indicating that the majority of information about the variability in the data is represented in the first plot. PC3 and PC4 add another 30.24 % of explained variance, suggesting that further differences in HVOC release are more dispersed and require detailed interpretation.

In the PC1 and PC2 plots, clear groupings of variables are visible depending on their associations with the principal components (Fig. 3). These compounds are likely associated with lower pyrolysis temperatures, suggesting that milder thermal conditions favor their retention in biochar. Although previous studies report that styrene formation

Table 2
Results of stepwise multiple regression for individual parameters describing the qualitative and quantitative release of VOCs from biochar.

Analysis	Parameter	Sampels	β		R^2	RMSE
			Temperature, °C	CRMM, $g \cdot mol^{-1}$		
Qualitative	Detected VOCs, n	Mix samples	−0.909	−0.202	0.87	4.43
		Validation	−0.847	−0.010	0.72	3.02
	Detected *HVOCs, n	Mix samples	−0.713	−0.317	0.61	0.93
		Validation	−0.813	−0.026	0.66	0.84
Quantitative	Total, VOCs, $mg \cdot kg^{-1}$	Mix samples	−0.305	0.132	0.11	261
		Validation	0.7097	0.028	0.50	45,418
	*HVOCs, $mg \cdot kg^{-1}$	Mix samples	−0.352	0.379	0.27	141
		Validation	0.635	0.095	0.39	2337

* Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures.

Table 3
Descriptive statistics of VOC-related dependent variables for mix and validation samples of biochar.

Samples	Mix samples				Validation samples			
Analysis	Qualitative		Quantitative		Qualitative		Quantitative	
Dependent variables	Detected VOC, n	Detected * HVOCs, n	Total, VOCs, $mg \cdot kg^{-1}$	*HVOCs, $mg \cdot kg^{-1}$	Detected VOC, n	Detected * HVOCs, n	Total, VOCs, $mg \cdot kg^{-1}$	*HVOCs, $mg \cdot kg^{-1}$
n	360	360	360	360	54	54	54	54
Mean	27.3	4.29	145	74.9	26.3	2.59	35,774	1617
Standard deviation	12.2	1.49	277	165	5.74	1.46	65,126	3076
Minimum	4	1	0	0	20	1	75.6	6.56
Maximum	43	6	2783	1438	33	4	322,803	15,521
Skewness	0.00551	−0.255	5.09	5.25	−0.187	−0.152	2.48	2.82
Standard error of skewness	0.129	0.129	0.129	0.129	0.325	0.325	0.325	0.325
25th percentile (Q1)	16	3	6.89	1.14	20	1	171	20.8
50th percentile (Median/Q2)	25	4	69.6	28.5	30	3	695	332
75th percentile (Q3)	41	6	168	83.3	31	4	34,287	1472

* Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures.

Table 4
Summary of principal component analysis (PCA).

Principal component	Eigenvalues	% variance	% cumulative
1	2.72	27.2	27.2
2	2.06	20.6	47.8
3	1.72	17.2	65.1
4	1.30	13.0	78.1

increases with pyrolysis temperature, this study observed a negative correlation between temperature and styrene concentration in biochar. This can be explained by the fact that at higher temperatures, styrene is more likely to volatilize and leave the solid matrix, whereas at lower temperatures, it may remain adsorbed or trapped in the biochar (Crawford and Lungu, 2011). Mequinol was previously identified in the tar and bio-oil fractions obtained from the pyrolysis of *Abies alba* wood, with slightly increasing concentrations at higher temperatures (López et al., 2013). In the present study, however, mequinol showed no clear correlation with HTT or CRMM, suggesting that its presence in biochar may depend more on specific precursor availability or release kinetics during pyrolysis rather than bulk process parameters. Their grouping suggests a shared precursor or mechanism.

Furfural and nonanoic acid showed a clear positive association with the HTT vector, suggesting a positive correlation as their release increases with rising pyrolysis temperature (Fig. 2). This finding aligns with the literature. Furfural is a known decomposition product of hemicellulose at elevated temperatures, and fatty acids and their derivatives are typically released more intensively during advanced stages of biomass degradation (Rasmussen et al., 2014). Temperature had the strongest overall influence, while biomass composition had a more moderate effect. CRMM had a moderate influence, especially on furfural, 4-heptanone, and nonanoic acid. This suggests that biomass composition, while secondary to temperature, still contributes meaningfully to their release (Fig. 3). Their proximity to CRMM in the PC3–PC4 space suggests a partial influence of substrate properties on their release. 4-heptanone is located higher and more to the left, clearly separated from the rest, suggesting that its release may be driven by a distinct, less dominant mechanism.

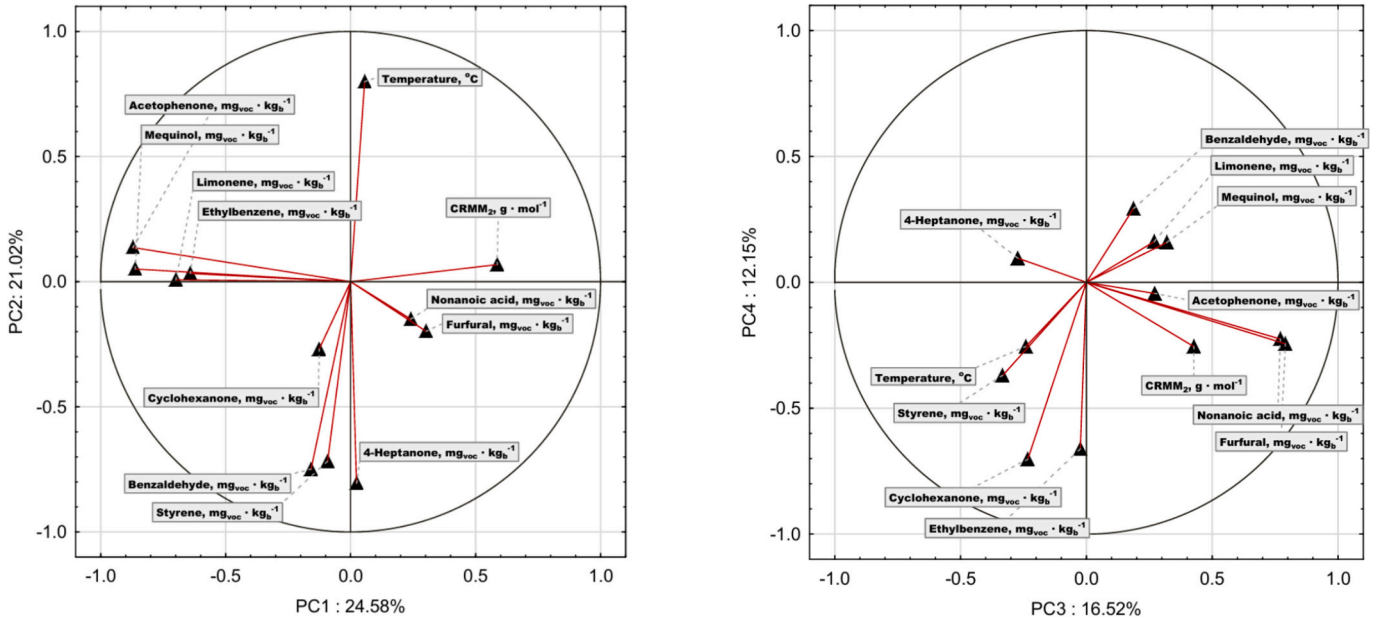


Fig. 3. Principal component analysis (PCA) biplot showing the loadings of HVOCs, pyrolysis temperature (HTT), and CRMM on the first and second (PC1–PC2), third and fourth (PC3–PC4) principal components.

A comparison of both plots shows that variables with high values on PC1 and PC2 (styrene, 4-heptanone, acetophenone) also play an important role in PC3 and PC4, but under different dependency patterns. This indicates that each principal component contributes meaningful information to the analysis of HVOCs, although their influence is distributed across multiple axes. This distribution of variance suggests that the release of HVOCs from biochar is not dominated by a single variable or group of compounds but is influenced by multiple factors highlighting the importance of considering all four components in the interpretation of results.

While PCA was used to explore structural patterns and groupings among HVOCs and biomass characteristics, it did not allow for identifying specific process conditions that govern the occurrence and concentration of hazardous compounds in biochar. Therefore, decision tree regression was employed to detect threshold-based relationships between parameters and VOC release. These methods serve complementary roles PCA supports an exploratory understanding of multivariate variance, whereas decision trees provide predictive insight critical for risk assessment and process optimization.

3.3. Agglomerative hierarchical clustering

To group individual HVOCs based on their co-occurrence patterns in biochar samples, agglomerative hierarchical cluster analysis was performed. This method involved creating a similarity matrix of the classified compounds and iteratively merging the most similar objects into clusters.

Acetophenone and mequinol exhibited the most similar release profiles (Fig. 4). Nonanoic acid was subsequently added to this VOC cluster, showing a comparable release pattern across biochar samples. Although nonanoic acid is derived from lipid degradation, its clustering with lignin-derived compounds may result from similar thermal formation ranges and release profiles (Ha et al., 2020). Moderate similarity was observed for cyclohexanone and limonene. In contrast, styrene, 4-heptanone, furfural, benzaldehyde, and ethylbenzene indicating were grouped at higher linkage distances, indicating more distinct release behaviors likely driven by different formation mechanisms. In the work of Kibet et al., benzene, toluene, and styrene are explicitly identified as key products of lignin pyrolysis. Additionally, the occurrence of benzaldehyde and ethylbenzene is confirmed, while the presence of other compounds such as acetophenone and mequinol may be inferred based on the prevalence of aromatic ketones and phenolic derivatives (Alburquerque et al., 2014). The dispersion and distances between clusters suggest that HVOC release from biochar is a complex phenomenon influenced by multiple factors, including pyrolysis temperature and the chemical composition of the biomass.

3.4. Decision tree

To investigate the nonlinear relationship between pyrolysis conditions and hazardous volatile organic compound (HVOC) release, a regression decision tree model was used. The goal was to identify thresholds and interactions between CRMM substrate characteristics and process temperature (HTT) concerning the total release of HVOCs. Fig. 5 presents the resulting regression tree. In the figure, Mu represents the mean VOC release for each terminal node, and Var denotes the variance within that node. High variance (Var) indicates greater dispersion of VOC release, reflecting heterogeneity within the group. The previously used stepwise regression model did not accurately reflect these relationships, showing a poor overall fit. This is due to the high variability of the data, i.e. the variance values. This suggests that HVOC release is governed by complex, nonlinear mechanisms that require a model capable of detecting threshold behavior and interaction variables — such as the regression tree method used here. The first and most important division is based on CRMM value. Samples with CRMM > 34 g·mol^{-1} form a distinct group with the highest average release (Mu =

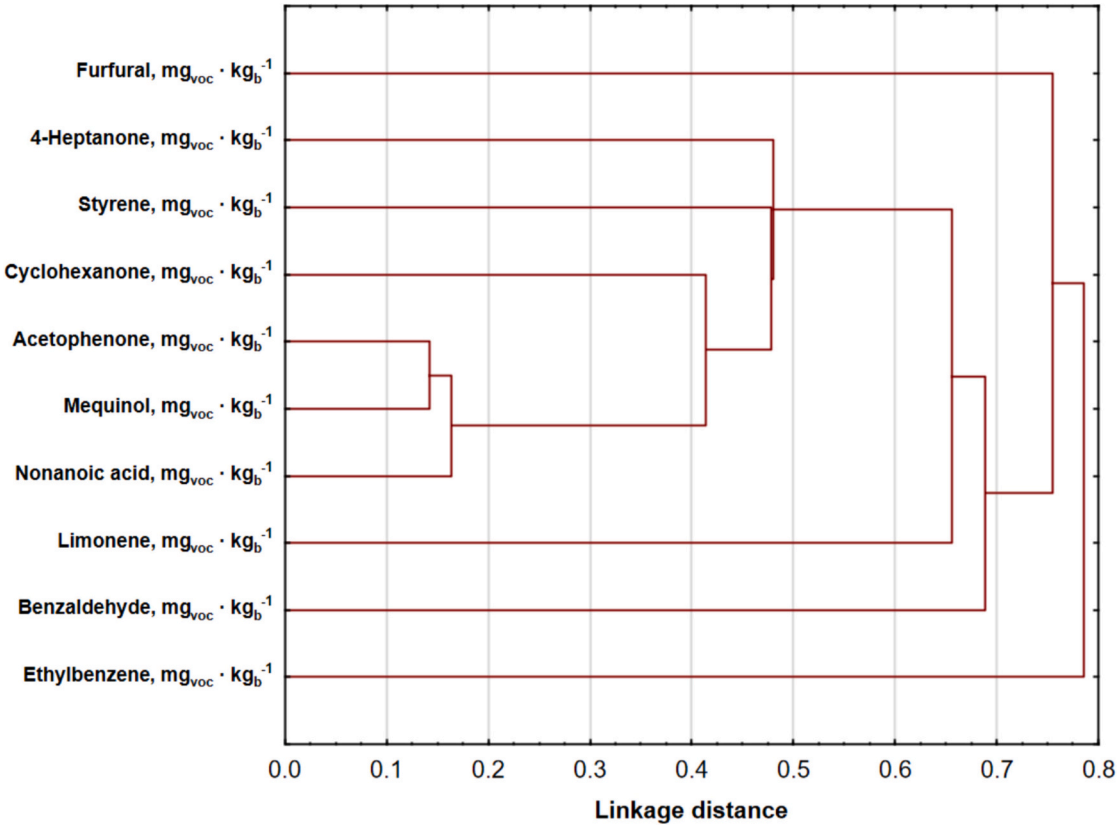


Fig. 4. Dendrogram of HVOCs released from biochar based on their co-occurrence patterns in biochar samples.

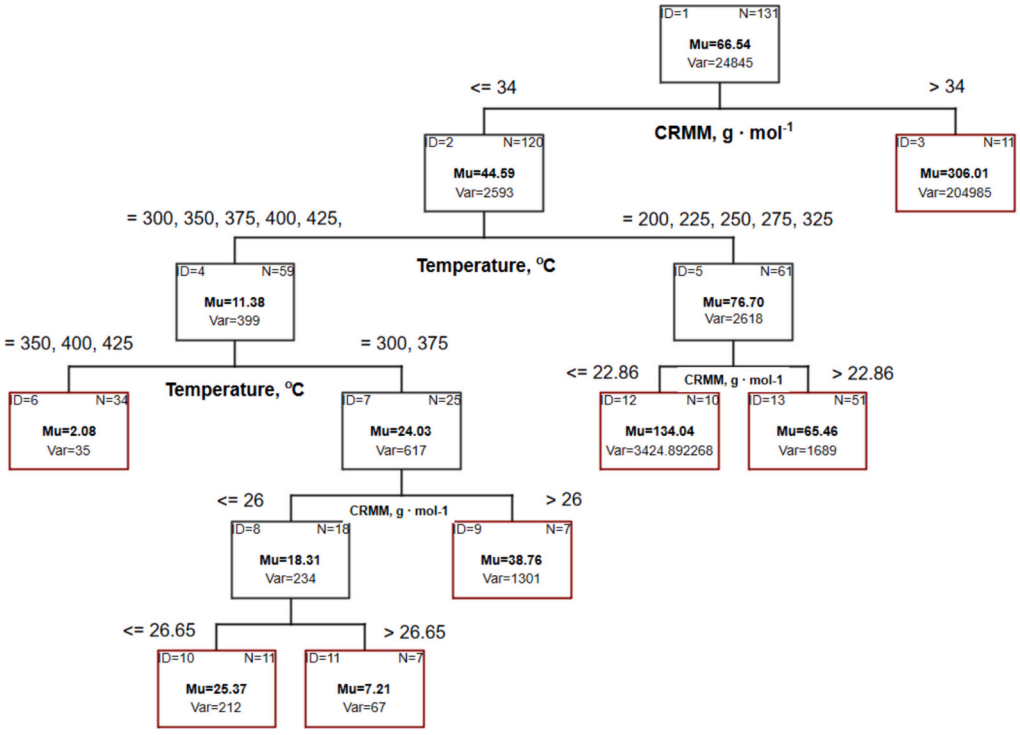


Fig. 5. Regression decision tree modeling the influence of CRMM and temperature on HVOC release for mix samples.

306.01 mg·kg^{-1}). This suggests that high CRMM values are strongly associated with excessive and unstable VOC release. Among samples with CRMM ≤ 34, the model further differentiates groups based on temperature. Higher temperatures (300, 350, 375, 400, and 425 °C) result in lower release (Mu = 11.38 mg·kg^{-1}). Lower temperatures (200–275 and 325 °C) correspond to higher release (Mu = 76.70 mg·kg^{-1}). The observations were divided almost evenly. Another branching of the high temperature group reveals that the highest temperatures 350–425 °C, lead to minimal HVOC release (Mu = 2.08 mg·kg^{-1}), while at 300–375 °C, release increases slightly (Mu = 24.03 mg·kg^{-1}). In this group, the CRMM value again modulates HVOC release: lower CRMM values lead to Mu = 25.37 mg·kg^{-1}, while CRMM > 26.65 reduces release to Mu = 7.21 mg·kg^{-1}. However, the lower temperature group (200–275 °C) also showed a strong dependence on CRMM. The highest average release (Mu = 134.04 mg·kg^{-1}) occurs when CRMM ≤ 22.86. Slightly reduced release is observed when CRMM exceeds this value (Mu = 65.46 mg·kg^{-1}).

In summary, the regression tree shows that CRMM is the main predictor of HVOC release, while HTT acts as a modifying factor. High release is associated with low pyrolysis temperatures and higher CRMM, while the lowest release occurs at high temperatures combined with low CRMM values. The tree structure effectively organizes the data into interpretable subgroups with different levels of release risk. Most of the end nodes show low internal variability, confirming the reliability of the partitioning. The results of the regression decision tree demonstrate that hazardous VOC release from biochar can be effectively predicted based on process parameters and substrate composition. Although this method has previously been applied to predict the physicochemical properties of biochar, such as specific surface area or ash content (Kandpal et al., 2024; Nguyen et al., 2024), it has not been used in the context of VOC release. In other studies, decision tree algorithms were applied to estimate VOC concentrations in air samples, but not to release from solid materials like biochar (Chen et al., 2017).

Both decision tree models for the test mix (Fig. 5) and validation samples (Fig. 6) showed a consistent pattern of relationships between the amount of HVOCs, pyrolysis temperature and CRMM index value. In both cases, temperature proved to be the dominant factor a marked decrease in HVOCs was observed when the temperature increased above 300–350 °C. CRMM is an additional differentiating factor, but its importance was less than that of temperature, which may be due to the limited range of this variable in the validation set for pine wood the CRMM value was 25.84 g·mol^{-1}, and for hemp straw 27.18 g·mol^{-1} compared to a much wider range (20–36 g·mol^{-1}) in the mix samples. The concordance of the breakdown structure and mean values (Mu) at the final nodes confirms the correctness of the model's generalization to real biomasses and the reliability of the prediction based on process parameters.

The risk of overfitting in decision tree models depends on the balance between the number of predictors and the dataset size. A greater number of predictors is generally suitable for large datasets, where complex patterns can be learned without compromising generalizability. In this study, we used a medium-sized dataset and deliberately limited the model to two key predictors (temperature and CRMM) to maintain interpretability and avoid excessive complexity. However, even in such cases, decision trees may still capture local patterns that do not generalize, especially when threshold effects vary across the data range. The current model does not include cross-validation, which should be addressed in future studies to improve robustness.

4. Conclusions and future studies

The results of the study demonstrated that the release of volatile organic compounds (VOCs) from biochar is influenced by both pyrolysis temperature and biomass properties, particularly the CRMM index. The CRMM (carbon-relative molar mass, g·mol^{-1}) parameter enables a simplified, yet representative quantification of biomass composition by aggregating five elemental variables into a single descriptor. This allowed for the linearization of biomass characteristics and their effective incorporation into modeling frameworks. Stepwise regression analysis revealed that higher pyrolysis temperatures and higher CRMM values are associated with a lower number of detected VOCs and HVOCs. However, these parameters only indicate the possibility of VOCs release,

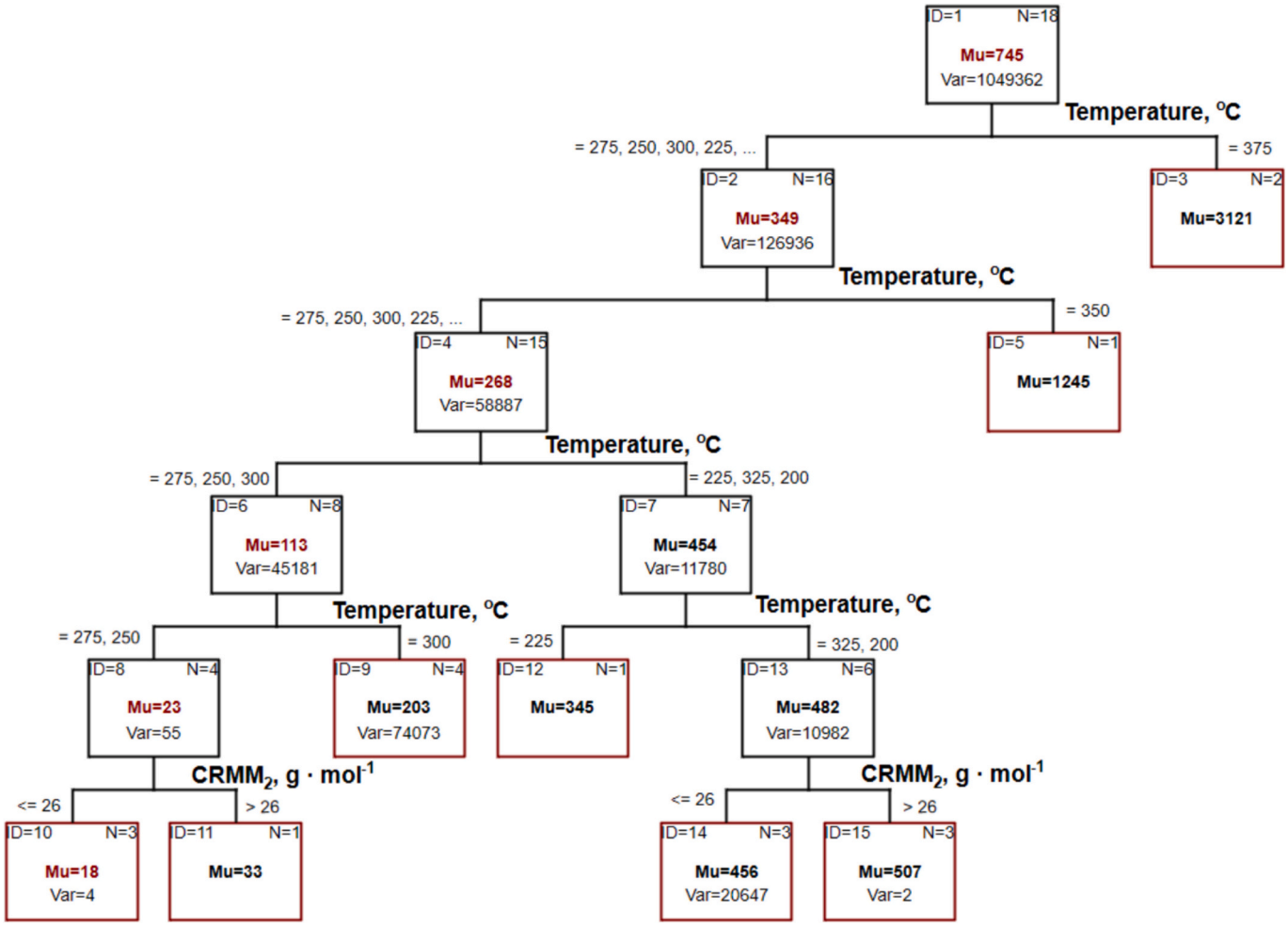


Fig. 6. Regression decision tree modeling the influence of CRMM and temperature on HVOC release for validation samples.

not their magnitude. Therefore, both qualitative and quantitative analyses were performed. The qualitative analysis highlighted the potential presence of hazardous compounds, while the quantitative analysis showed that linear models are insufficient for predicting release levels.

In contrast, regression trees captured nonlinear threshold effects determining VOC release levels, relevant to human and environmental health. The highest average releases, around 300 $mg \cdot kg^{-1}$, were observed in biochars with very high CRMM values (>34 $g \cdot mol^{-1}$), regardless of temperature, as well as at lower pyrolysis temperatures (200–325 °C) with low CRMM values around 134 $mg \cdot kg^{-1}$. The lowest releases were recorded at higher temperatures (350–425 °C), particularly when combined with moderate to high CRMM values. This is the first study to apply decision tree regression to model HVOC release from biochar based on CRMM and pyrolysis temperature. Existing studies have applied this method primarily to predict physicochemical properties (e.g., ash content) or estimate VOC concentrations in air. This approach would also the future sensitivity analysis or feature importance assessment would enhance the interpretability and reliability of the decision tree results.

This opens the possibility of applying these parameters to assess other test materials by determining their CRMM values. Future work should analyze a narrower HTT range using finer temperature intervals, but also other technological parameters like retention time, ramp-up temperature rate, or inert gas flow rate. Standardized VOC analysis methods are also needed to support regulatory quality standards for biochar. Although occupational and environmental limits exist for certain VOCs ($mg \cdot m^{-3}$ for inhalation), no such thresholds currently apply to VOCs in solid biochar. Since VOC emission depends on environmental conditions and material properties, the measured content should be viewed as a potential release indicator rather than a direct exposure metric. Overall, our results provide actionable guidance for biochar producers seeking to minimize the formation of hazardous compounds through informed process control. The study supports ongoing standardization efforts and underscores the need to develop regulatory benchmarks to ensure environmental and occupational safety in biochar production and use.

An important additional consideration is the potential for scaling and transferring the obtained findings into real-world industrial applications. While the results were derived under controlled laboratory conditions, the modeled threshold values of HTT and CRMM provide an initial framework for anticipating HVOC emissions from biochar. However, these findings require validation using biochar produced under industrial-scale conditions. Such validation would enhance the reliability of the proposed indicators and support the development of practical guidelines for optimizing biochar production, with due consideration to environmental and human health safety.

CRediT authorship contribution statement

Ewa Syguła: Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Methodology, Investigation, Data curation, Conceptualization. **Kamila Piasecka:** Validation, Software, Methodology, Data curation. **Jacek Łyczko:** Methodology, Investigation, Data curation. **Andrzej Białowiec:** Writing – review & editing,

Validation, Supervision, Project administration, Funding acquisition, Formal analysis, Conceptualization.

Funding

This research was funded in whole by the National Science Centre, Poland, Grant Number UMO-2019/35/O/ST8/03353. For the purpose of Open Access, the author has applied a CC-BY public copyright license to any Author Accepted Manuscript (AAM) version arising from this submission.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Andrzej Bialowiec reports financial support was provided by National Science Centre Poland. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The APC is financed by the Wroclaw University of Environmental and Life Sciences. This work was conducted at the Waste and Biomass Valorization Group (WBVG), and the Department of Applied Bioeconomy, Wroclaw University of Environmental and Life Sciences, Poland.

Data availability

The reference to the data repository is included in the manuscript.

References

Alburquerque, J.A., Calero, J.M., Barrón, V., Torrent, J., del Campillo, M.C., Gallardo, A., et al., 2014. Effects of biochars produced from different feedstocks on soil properties and sunflower growth. J. Plant Nutr. Soil Sci. 177 (1), 16–25.

Bachmann, H.J., Bucheli, T.D., Dieguez-Alonso, A., Fabbri, D., Knicker, H., Schmidt, H. P., et al., 2016. Toward the standardization of biochar analysis: the COST action TD1107 interlaboratory comparison. J. Agric. Food Chem. 64 (2), 513–527.

Balat, M., 2008. Mechanisms of thermochemical biomass conversion processes. Part 1: reactions of pyrolysis. Energy Sour. A Recov. Util. Environ. Eff. 30 (7), 620–635.

Ben, H., Ragauskas, A.J., 2013. Comparison for the compositions of fast and slow pyrolysis oils by NMR characterization. Bioresour. Technol. 147, 577–584. https://doi.org/10.1016/j.biortech.2013.07.151.

Buss, W., Mašek, O., 2014. Mobile organic compounds in biochar — a potential source of contamination — phytotoxic effects on cress seed (*Lepidium sativum*) germination. J. Environ. Manag. 137, 111–119.

Buss, W., Mašek, O., Graham, M., Wüst, D., 2015. Inherent organic compounds in biochar—their content, composition and potential toxic effects. J. Environ. Manag. 156, 150–157.

Cai, J., He, Y., Yu, X., Banks, S.W., Yang, Y., Zhang, X., et al., 2017. Review of physicochemical properties and analytical characterization of lignocellulosic biomass. Renew. Sust. Energ. Rev. 76, 309–322.

Chen, Z., Zheng, Y., Chen, K., Li, H., Jian, J., 2017. Concentration estimator of mixed VOC gases using sensor array with neural networks and decision tree learning. IEEE Sensors J. 17 (6), 1884–1892.

Crawford, S., Lungu, C.T., 2011. Influence of temperature on styrene emission from a vinyl ester resin thermoset composite material. Sci. Total. Environ. 409 (18), 3403–3408. Available from: https://doi.org/10.1016/j.scitotenv.2011.05.042.

Delgado-Rodríguez, M., Ruiz-Montoya, M., Giraldez, I., López, R., Madejón, E., Díaz, M. J., 2012. Use of electronic nose and GC-MS in detection and monitoring some VOC. Atmos. Environ. 51, 278–285. https://doi.org/10.1016/j.atmosenv.2012.01.006.

Dutta, T., Kwon, E., Bhattacharya, S.S., Jeon, B.H., Deep, A., Uchimiya, M., et al., 2017. Polycyclic aromatic hydrocarbons and volatile organic compounds in biochar and biochar-amended soil: a review. GCB Bioenergy 9 (6), 990–1004.

Evuti, A.M., 2013. A synopsis on biogenic and anthropogenic volatile organic compounds emissions: hazards and control a synopsis on biogenic and anthropogenic volatile organic compounds emissions: hazards and control. Int. J. Eng. Sci. 2 (5). Available from: https://www.researchgate.net/publication/236612089.

Demirbas, M.F., 2006. Current technologies for biomass conversion into chemicals and fuels. Energy Sour. A Recov. Util. Environ. Eff. 28 (13), 1181–1188.

Geng, F., Cai, C., Tie, X., Yu, Q., An, J., Peng, L., et al., 2009. Analysis of VOC emissions using PCA/APCS receptor model at city of Shanghai, China. J. Atmos. Chem. 62 (3), 229–247.

Ghidotti, M., Fabbri, D., Hornung, A., 2017. Profiles of volatile organic compounds in biochar: insights into process conditions and quality assessment. ACS Sustain. Chem. Eng. 5 (1), 510–517.

Ha, G.S., El-Dalatony, M.M., Kim, D.H., Salama, E.S., Kurade, M.B., Roh, H.S., et al., 2020. Biocomponent-based microalgal transformations into biofuels during the pretreatment and fermentation process. Bioresour. Technol. 302, 122809. https://doi.org/10.1016/j.biortech.2020.122809.

Han, H., Buss, W., Zheng, Y., Song, P., Khalid Rafiq, M., Liu, P., et al., 2022. Contaminants in biochar and suggested mitigation measures — a review. Chem. Eng. J. 429, 132287. https://doi.org/10.1016/j.cej.2021.132287.

Han, K.-H., Yun, S.-I., Kwak, J.-H., Lee, S.-I., 2023. A review on international carbon credit certification methodologies for biochar as a soil amendment. Kor. J. Soil Sci. Fertil. 56 (4), 572–594.

Hernowo, P., Rasrendra, C.B., Budhi, Y.W., Rizkiana, J., Irawan, A., Marno, S., et al., 2022. Volatile state mathematical models for predicting components in biomass pyrolysis products. J. Eng. Technol. Sci. 54 (1).

Hsieh, W.W., 2023. Improving predictions by nonlinear regression models from outlying input data. J. Environ. Inf. 41 (2), 75–87.

Kandpal, S., Tagade, A., Sawarkar, A.N., 2024. Critical insights into ensemble learning with decision trees for the prediction of biochar yield and higher heating value from pyrolysis of biomass. Bioresour. Technol. 411 (August), 131321. Available from: https://doi.org/10.1016/j.biortech.2024.131321.

Kataki, Rupam, Rsc, A., Mishra, M., Bordoloi, N., Saikia, R., Bhaskar, T., 2015. Feedstock Suitability for Thermochemical Processes.

Kazemi Shariat Panahi, H., Dehhaghi, M., Ok, Y.S., Nizami, A.S., Khoshnevisan, B., Mussatto, S.I., et al., 2020. A comprehensive review of engineered biochar: production, characteristics, and environmental applications. J. Clean. Prod. 270, 122462. https://doi.org/10.1016/j.jclepro.2020.122462.

Kibet, J., Khachatryan, L., Dellinger, B., 2012. Molecular products and radicals from pyrolysis of lignin. Environ. Sci. Technol. 46 (23), 12994–13001.

Kim, H. Bin, Kim, J.G., Kim, T., Alessi, D.S., Baek, K., 2021. Interaction of biochar stability and abiotic aging: influences of pyrolysis reaction medium and temperature. Chem. Eng. J. 411, 128441. https://doi.org/10.1016/j.cej.2021.128441.

Komilis, D.P., Ham, R.K., 2003. The effect of lignin and sugars to the aerobic decomposition of solid wastes. Waste Manag. 23 (5), 419–423.

Liu, CH, Chu, W, Li, H, Boyd, SA, Teppen, BJ, Mao, J, et al. Quantification and characterization of dissolved organic carbon from biochars. Geoderma. 2019;335 (April 2018):161–9. Available from: https://doi.org/10.1016/j.geoderma.2018.08.019.

López, F.A., Rodríguez, O., Urien, A., Lobato, B., Centeno, T.Á., Alguacil, F.J., 2013. Physico-chemical characteristics of the products derived from the thermolysis of waste *Abies alba* Mill. wood. J. Environ. Prot. 04 (01), 26–30.

Menon, V., Rao, M., 2012. Trends in bioconversion of lignocellulose: biofuels, platform chemicals & biorefinery concept. Prog. Energy Combust. Sci. 38 (4), 522–550. Available from: https://doi.org/10.1016/j.pecs.2012.02.002.

Nawirska, A., Kwaśniewska, M., 2004. Frakcje błonnikaw wytłokach z owoców. Acta Sci. Pol. Technol. Aliment. 3 (1), 13–20.

Nguyen, V.G., Sharma, P., Ağbulut, Ü., Le, H.S., Cao, D.N., Dzida, M., et al., 2024. Improving the prediction of biochar production from various biomass sources through the implementation of eXplainable machine learning approaches. Int. J. Green Energy 21 (12), 2771–2798. Available from: https://doi.org/10.1080/15435075.2024.2326076.

Oral, Burcu, Coşgun, Ahmet, Günay, M. Erdem, Yıldırım, R., 2024. Machine learning-based exploration of biochar for environmental management and remediation. J. Environ. Manag. 360.

Orejuela-Escobar, L.M., Landázuri, A.C., Goodell, B., 2021. Second generation biorefining in Ecuador: circular bioeconomy, zero waste technology, environment and sustainable development: the nexus. J. Bioresour. Bioprod. 6 (2), 83–107. https://doi.org/10.1016/j.jobab.2021.01.004.

Ponomarev, A.V., Ershov, B.G., 2014. Radiation-induced high-temperature conversion of cellulose. Molecules 19 (10), 16877–16908.

Rasmussen, H., Sørensen, H.R., Meyer, A.S., 2014. Formation of degradation compounds from lignocellulosic biomass in the biorefinery: sugar reaction mechanisms. Carbohydr. Res. 385, 45–57. Available from: https://doi.org/10.1016/j.carres.2013.08.029.

Sadegh, F., Sadegh, N., Wongniramaikul, W., Apiratikul, R., Choodum, A., 2024. Adsorption of volatile organic compounds on biochar: a review. Process Saf. Environ. Prot. November 2023 (182), 559–578. Available from: https://doi.org/10.1016/j.psep.2023.11.071.

Sobol, Ł., Dyjakon, A., Soukup, K., 2023. Dioxins and furans in biochars, hydrochars and torrefiicates produced by thermochemical conversion of biomass: a review. Environ. Chem. Lett. 21 (4), 2225–2249. Available from: https://doi.org/10.1007/s10311-023-01600-7.

Spokas, K.A., Novak, J.M., Stewart, C.E., Cantrell, K.B., Uchimiya, M., DuSaire, M.G., et al., 2011. Qualitative analysis of volatile organic compounds on biochar. Chemosphere 85 (5), 869–882. https://doi.org/10.1016/j.chemosphere.2011.06.108.

Štrubelj, L., 2022. Waste, fertilising product, or something else? EU regulation of biochar. J. Environ. Law 34 (3), 529–540.

Supriyanto, Usino, D.O., Ylitervo, P., Dou, J., Sipponen, M.H., Richards, T., 2020. Identifying the primary reactions and products of fast pyrolysis of alkali lignin. J. Anal. Appl. Pyrolysis 151 (June).

Syguła, E., Ciolkosz, D., Białowiec, A., 2024. The significance of structural components of lignocellulosic biomass on volatile organic compounds presence on biochar — a review. Wood Sci. Technol. 58 (3), 859–886.

Syguła, E., Łyczko, J., Białowiec, A., 2025a. Dataset on fuel properties and volatile organic compounds from chemically synthesized biomass components for modeling and predicting biomass properties in pyrolysis processes. Data Brief 111296. https://doi.org/10.1016/j.dib.2025.111296.

Syguła, E., Łyczko, J., Białowiec, A., 2025b. Dataset on fuel properties and volatile organic compounds from chemically synthesized biomass components for modeling and predicting biomass properties in pyrolysis processes. Data Brief 59, 10–16.

Taylor, M.J., Alabdrabalameer, H.A., Skoulou, V., 2019. Choosing physical, physicochemical and chemical methods of pre-treating lignocellulosic wastes to repurpose into solid fuels. Sustainability 11 (13).

Tripathi, M., Sahu, J.N., Ganesan, P., 2016. Effect of process parameters on production of biochar from biomass waste through pyrolysis: a review. Renew. Sust. Energ. Rev. 55, 467–481. https://doi.org/10.1016/j.rser.2015.10.122.

Tursi, A., 2019. A review on biomass: importance, chemistry, classification, and conversion. Biofuel Res. J. 6 (2), 962–979.

Usino, D.O., Supriyanto, Ylitervo, P., Pettersson, A., Richards, T., 2020. Influence of temperature and time on initial pyrolysis of cellulose and xylan. J. Anal. Appl. Pyrolysis 147 (November 2019), 104782. Available from: https://doi.org/10.1016/j.jaap.2020.104782.

Wang, Wei, Chang, Jo-Shu, Lee, Duu-Jong, 2024. Machine learning applications for biochar studies: a mini-review. Bioresour. Technol. 394.

Wu, P., Ata-Ul-Karim, S.T., Singh, B.P., Wang, H., Wu, T., Liu, C., et al., 2019. A scientometric review of biochar research in the past 20 years (1998–2018). Biochar 1 (1), 23–43. https://doi.org/10.1007/s42773-019-00002-9.

Xiang, W., Zhang, X., Chen, K., Fang, J., He, F., Hu, X., et al., 2020. Enhanced adsorption performance and governing mechanisms of ball-milled biochar for the removal of volatile organic compounds (VOCs). Chem. Eng. J. 385, 123842. https://doi.org/10.1016/j.cej.2019.123842.

Zhang, X., Gao, B., Zheng, Y., Hu, X., Creamer, A.E., Annable, M.D., et al., 2017a. Biochar for volatile organic compound (VOC) removal: sorption performance and governing mechanisms. Bioresour. Technol. 245 (September), 606–614. Available from: https://doi.org/10.1016/j.biortech.2017.09.025.

Zhang, X., Gao, B., Zheng, Y., Hu, X., Creamer, A.E., Annable, M.D., et al., 2017b. Biochar for volatile organic compound (VOC) removal: sorption performance and governing mechanisms. Bioresour. Technol. 245 (September), 606–614. https://doi.org/10.1016/j.biortech.2017.09.025.

Zhang, Y., Liang, Y., Li, S., Yuan, Y., Zhang, D., Wu, Y., et al., 2023. A review of biomass pyrolysis gas: forming mechanisms, influencing parameters, and product application upgrades. Fuel 347 (January), 128461. Available from: https://doi.org/10.1016/j.fuel.2023.128461.

Zhao, Y., Li, Y.L., Yang, F., 2022. A state-of-the-art review on modeling the biochar effect: guidelines for beginners. Sci. Total. Environ. 802, 149861. Available from: https://doi.org/10.1016/j.scitotenv.2021.149861.

Zheng, H., Liu, B., Liu, G., Cai, Z., Zhang, C., 2018. Potential toxic compounds in biochar: knowledge gaps between biochar research and safety. In: Biochar from Biomass and Waste: Fundamentals and Applications. Elsevier Inc., pp. 349–384. Available from: https://doi.org/10.1016/B978-0-12-811729-3.00019-4

Zhou, X., Moghaddam, T.B., Chen, M., Wu, S., Adhikari, S., 2020. Biochar removes volatile organic compounds generated from asphalt. Sci. Total. Environ. 745 (27), 141096. Available from: https://doi.org/10.1016/j.scitotenv.2020.141096.

Zhu, X., Chen, B., Zhu, L., Xing, B., 2017. Effects and mechanisms of biochar-microbe interactions in soil improvement and pollution remediation: a review. Environ. Pollut. 227, 98–115. Available from: https://doi.org/10.1016/j.envpol.2017.04.032.