CHEMICAL ENGINEERING TRANSACTIONS

VOL. 113, 2024

A publication of
AIDIC
The Italian Association
of Chemical Engineering
Online at www.cetjournal.it

Guest Editors: Jeng Shiun Lim, Bohong Wang, Guo Ren Mong, Petar S. Varbanov
Copyright © 2024, AIDIC Servizi S.r.l.
ISBN 979-12-81206-14-4; **ISSN** 2283-9216

DOI: 10.3303/CET24113093

Regression Models for Predicting Physicochemical Properties of Biochar

Chiaw Hui Chiew[a], Li Yee Lim[a], Pei Ying Ong[a,*], Chunjie Li[b], Yee Van Fan[c]

[a]Faculty of Chemical and Energy Engineering, Universiti Teknologi Malaysia, 81310, Johor, Malaysia
[b]School of Environmental Science and Engineering, Shanghai Jiao Tong University, 200240, Shanghai, China
[c]Environmental Change Institute (ECI), University of Oxford, Oxford, OX1 3QY, United Kingdom
o.peiying@utm.my

This study evaluates the effectiveness of various regression models in predicting the physicochemical properties of biochar, essential for sustainable agriculture and environmental remediation. A comprehensive review of recent literature compared the predictive accuracies of linear, non-linear regression (NLR), quadratic, and multiple linear regression (MLR) models. Findings highlight that MLR models perform exceptionally well, with R^2 values exceeding 0.92, particularly in predicting complex interactions like cation exchange capacity (CEC) and electrical conductivity (EC). NLR models also demonstrated strong performance, achieving high median R^2 values, especially in predicting High Heating Value (HHV), with R^2 values up to 0.9802. Pyrolysis Temperature (PT) was identified as a frequent and significant predictor for properties such as EC and nitrogen content. However, properties like CEC and Specific Surface Area (SSA) presented challenges due to inconsistencies between high R^2 and higher Root Mean Square Error (RMSE) values, indicating underlying variability. Municipal Solid Waste (MSW) biochar was the most challenging to predict due to its heterogeneous composition. This study advocates for integrating MLR with non-linear techniques to develop hybrid models, enhancing predictive accuracy and practical usability, and optimizing biochar utilization in agriculture and environmental remediation.

1. Introduction

Biochar, derived from biomass pyrolysis, has garnered attention for its applications in soil improvement and environmental remediation (García et al., 2021). Incorporating biochar into agriculture enhances soil fertility, boosting nutrient availability and crop yields. It also aids in preventing nutrient leaching and conserving water by retaining moisture in the topsoil, contributing to enhanced crop output and improved water and nutrient efficiency (Mylavarapu et al., 2013). Chen et al. (2023) identified regression modelling as the most effective method for predicting the physicochemical properties of biochar among all machine learning techniques. However, there has been no comprehensive review on which specific type of regression modelling is best suited for biochar's physicochemical properties prediction. This presents a significant research gap, particularly in understanding the comparative efficacy of different regression models for predicting biochar's physicochemical properties. This study aims to bridge the knowledge gap by evaluating the performance of linear, non-linear regression (NLR), quadratic, and multiple linear regression (MLR) models in predicting biochar's properties. The analysis will focus on their predictive accuracy, as measured by R^2 values and Root Mean Square Error (RMSE), and their practical application in biochar production. The ultimate goal is to identify the most effective models that can be tailored to the unique characteristics of biochar and its diverse applications. Future research will likely explore integrating traditional regression models with advanced machine learning algorithms to develop hybrid models. These models aim to enhance predictive accuracy and practical usability, supporting more effective biochar utilization in agriculture and environmental remediation.

2. Review method

The data was collected from research articles published between 2016 and 2022, spanning various journals related to biochar production and utilization. The search terms included keywords such as "biochar,"

Paper Received: 19 April 2024; Revised: 17 September 2024; Accepted: 23 October 2024
Please cite this article as: Chiew C.H., Lim L.Y., Ong P.Y., Li C., Fan Y.V., 2024, Regression Models for Predicting Physicochemical Properties of Biochar, Chemical Engineering Transactions, 113, 553-558 DOI:10.3303/CET24113093

"physicochemical properties," "regression modelling". Relevant articles underwent screening, and data of predictors, predicted variables, model types, biomass content, R^2 values and RMSE were extracted.

3. Regression models for predicting physicochemical properties of biochar

In evaluating regression models for predicting the physicochemical properties of biochar, Linear, quadratic, NLR and MLR models were analyzed. Table 1 provides an overview of the proposed regression models, highlighting their performance metrics and applicability in predicting biochar's physicochemical properties.

Table 1: Regression models in predicting physicochemical properties of biochar

Model	Predictor	Predicted variables	Biomass Category	N	R^2	RMSE	Reference
Linear	Weight	CEC	Animal-Derived Residues, Forest Residues, Industrial Residues, Agricultural Residues	20	0.97	5.90	(Lago et al., 2021)
Linear	C	CEC	Animal-Derived Residues, Forest Residues, Industrial Residues, Agricultural Residues	20	0.96	15.52	(Lago et al., 2021)
Linear	PT	EC	Agricultural Residues	21	0.66	1.30	(Morais et al., 2021)
Linear	PT	EC	Animal-Derived Residues	21	0.67	0.20	(Morais et al., 2021)
Linear	Ash	HHV	Animal-Derived Residues, Forest Residues, Agricultural Residues, MSW	52	0.92	1.99	(Chen et al., 2022)
Linear	C, H, O, N, VM, FC, Ash	HHV	Municipal Solid Waste (MSW)	67	0.63	2.50	(Mari Selvam and Balasubramanian, 2023)
Linear	PT	N	Agricultural Residues	21	0.89	0.80	(Morais et al., 2021)
Linear	PT	N	Animal-Derived Residues	21	0.92	1.60	(Morais et al., 2021)
Linear	PT	N loss	Agricultural Residues	21	0.80	5.60	(Morais et al., 2021)
Linear	PT	N loss	Animal-Derived Residues	21	0.96	2.70	(Morais et al., 2021)
Linear	N/A	Yield	Agricultural Residues	77	0.82	N/A	(Narde and Remya, 2022)
Quadratic	H/C, O/C	Aromaticity	Animal-Derived Residues, Forest Residues, Agricultural Residues	98	0.89	0.09	(Cao et al., 2021)
Quadratic	PT	EC	Agricultural Residues	21	0.93	0.60	(Morais et al., 2021)
Quadratic	PT	EC	Animal-Derived Residues	21	0.67	0.20	(Morais et al., 2021)
Quadratic	PT	N	Agricultural Residues	21	0.99	0.20	(Morais et al., 2021)
Quadratic	PT	N	Animal-Derived Residues	21	0.94	1.30	(Morais et al., 2021)
Quadratic	PT	N loss	Agricultural Residues	21	0.98	1.60	(Morais et al., 2021)
Quadratic	PT	N loss	Animal-Derived Residues	21	0.96	2.60	(Morais et al., 2021)
Quadratic	PT	pH	Agricultural Residues	22	0.60	N/A	(Rafiq et al., 2016)
NLR	PT, H/C	HHV	Agricultural Residues	18	0.98	N/A	(Ortiz et al., 2020)
NLR	C,H	HHV	Animal-Derived Residues, Forest Residues, Agricultural Residues	1566	0.89	N/A	(Yaka et al., 2022)
NLR	PT, H/C	O/C	Agricultural Residues	18	0.99	N/A	(Ortiz et al., 2020)
NLR	PT, H/C	pH	Agricultural Residues	18	0.96	N/A	(Ortiz et al., 2020)
NLR	VM,AC,PT	Yield	Agricultural Residues	112	0.89	6.25	(Narde and Remya, 2022)
MLR	PT, H/C	EC	Agricultural Residues	18	0.92	N/A	(Ortiz et al., 2020)
MLR	PT, H/C	H/C	Agricultural Residues	18	0.99	N/A	(Ortiz et al., 2020)
MLR	C,N,O	HHV	Animal-Derived Residues, Forest Residues, Agricultural Residues, MSW	52	0.93	1.84	(Chen et al., 2022)
MLR	C,H	HHV	Animal-Derived Residues, Forest Residues, Agricultural Residues, MSW	1566	0.88	1.79	(Yaka et al., 2022)
MLR	C, H, N, S, MC, Ash, PT	SSA	Animal-Derived Residues, Forest Residues, Industrial Residues, Agricultural Residues, MSW	292	0.29	10.87	(Hai et al., 2023)
MLR	C, H, N, S, MC, Ash, PT	Yield	Animal-Derived Residues, Forest Residues, Industrial Residues, Agricultural Residues, MSW	292	0.82	4.76	(Hai et al., 2023)
MLR	PT, H/C	Yield	Agricultural Residues	18	0.96	N/A	(Ortiz et al., 2020)

The analysis of Figure 1, which depicts the distribution of R² and RMSE values for various predicted variables using different regression models, reveals several key insights into the models' predictive performance for biochar properties. Most reported R² values are above 0.6, indicating that the regression models generally provide a good fit for predicting the physicochemical properties of biochar. However, there are notable inconsistencies between high R² values and RMSE values, particularly for certain variables. HHV predictions are comparatively easier for regression models to capture, with the highest R² of 0.98 achieved using NLR by Ortiz et al. (2020). The RMSE values for HHV range from 1.80 to 2.5, indicating good predictive accuracy, but with variability suggesting that simpler linear models may not adequately capture the complexity of HHV.

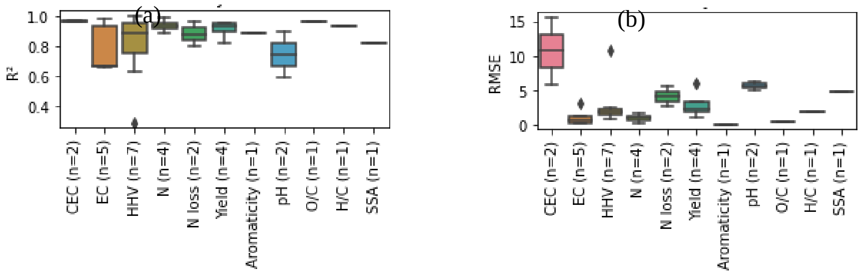

Figure 1: Distribution of (a) R2; (b) RMSE by Predicted Variables. n=number of studies

MLR models for predicting EC demonstrate high predictive accuracy, with the highest R² of 0.92 reported by Ortiz et al. (2020). The RMSE values for EC are relatively low, ranging from 0.2 to 1.3, indicating high accuracy, yet the variability in RMSE values highlights the significant influence of biomass type and pyrolysis conditions on the predictive performance.

Interestingly, while CEC shows high R² values, such as 0.97 reported by Lago et al. (2021) using a linear model, the RMSE values are not correspondingly low, ranging from 5.9 to 15.52. This discrepancy suggests that although the models explain a significant portion of the variance, there are underlying factors contributing to higher errors in prediction. One potential reason for this could be the significant variability in CEC values across different types of biochar. Gaskin et al. (2008) explained that CEC is influenced by the type of biomass and the temperature of biochar production. Manure-derived biochar tends to have higher CEC values compared to woody biochar, with CEC values such as 57.50 cmol/kg for algal biochar and 48.4 cmol/kg for poultry litter biochar at 500 °C, compared to 29.90 mol/kg for orange pomace-derived biochar (Tag et al., 2016). This inherent variability in CEC values can lead to higher RMSE despite high R² values, as the models must account for a wider range of values and conditions.

Yield predictions show high R² values, particularly in MLR models (up to 0.96), with RMSE values ranging from 4.76 to 6.25, indicating good accuracy but also highlighting the need for more precise models to capture the influence of volatile matter, ash content, and pyrolysis temperature on yield.

Nitrogen content and nitrogen loss predictions achieve high accuracy with quadratic and MLR models, reflected in high R² values (up to 0.99) and low RMSE (ranging from 0.2 to 5.6), indicating the models' effectiveness in capturing temperature-related impacts. The high R² values and low RMSE for nitrogen loss suggest that these models are well-suited for scenarios involving significant temperature effects on nitrogen dynamics.

Variables such as Aromaticity, O/C, H/C, and SSA that have a small number of studies (n =1) indicate that the results might not be broadly representative. These results are more susceptible to variability and may not generalize well to other contexts. These complex interactions and dependencies make it challenging for regression models to achieve consistent predictive accuracy across different studies and conditions.

Figure 2 illustrates the range of R² and RMSE values by model type and Figure 3 illustrates the sample size (N) vs R2 by model type, providing a comparative performance analysis of these models.

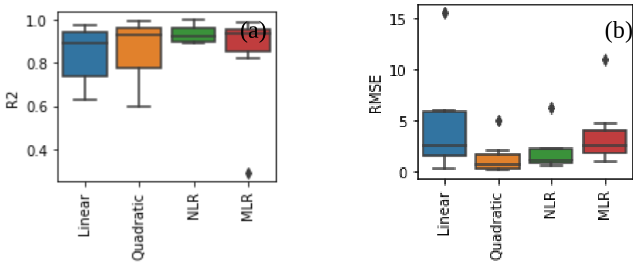

Figure 2: Range of (a) R2 and (b) RMSE by Model Type

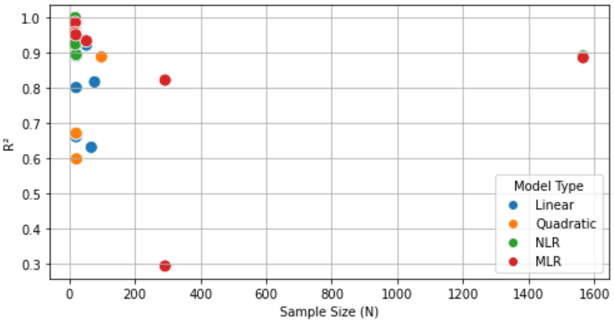

Figure 3: Sample Size (N) vs R^2 by Model Type

MLR demonstrates the best overall performance among the evaluated models. It has the highest median R² value, indicating superior effectiveness in explaining variance, and a narrow interquartile range (IQR), reflecting consistent performance. The presence of a dot below the whisker in Figure 2 indicates an outlier with a lower R² value. Figure 3 shows R² values decrease with an increase in sample size. Most studies reviewed used low sample sizes, ranging from 18 to 292, with only one study exceeding 1000. This trend of limited sample sizes may impact validity, making findings less generalizable and suitable for machine learning models that risk overfitting. High R² values might not accurately reflect the model's performance on new and unseen data. When MLR models encounter multicollinearity, different sample sizes (N), and outliers, the reliability, stability, and interpretability of the regression coefficients are compromised (Akinwande et al., 2015). Despite the minor outlier, MLR models maintain a robust fit to biochar data, with the lowest median RMSE values, underscoring MLR's potential in integrating multiple predictors for improved biochar production decisions.

NLR exhibits strong performance, with a high median R² value, marginally lower than that of MLR, and a narrow IQR indicating consistent model reliability. Despite their computational intensity, NLR models excel in accommodating complex patterns, rendering them suitable for capturing the intricate dynamics inherent in biochar-treated soils (Liu et al., 2018). The low median RMSE values highlight their predictive accuracy, although with slightly less consistency compared to MLR.

Quadratic Regression demonstrates moderate performance, characterized by a median R² value lower than both MLR and NLR. The larger IQR indicates higher variability in model effectiveness. Quadratic models, which are second-degree polynomials, are moderately effective in handling non-linear relationships but are sensitive to outliers, affecting their reliability (Henson and Friston, 2007). The moderate median RMSE values indicate higher prediction errors compared to MLR and NLR, underscoring their limited accuracy.

Linear Regression exhibits the lowest performance, marked by the lowest median R² value and the largest IQR, indicating significant variability and frequent poor performance. These models often fail to capture the complex interactions within biochar systems, resulting in the highest median RMSE values and the least accurate predictions. The presence of highly negatively correlated features can lead to inflated coefficients that may cancel each other out, further skewing results (Iqbal, 2020).

Figure 4 represents the range of R² and RMSE values by biomass category, highlighting the variability and performance of different regression models across various biomass types.

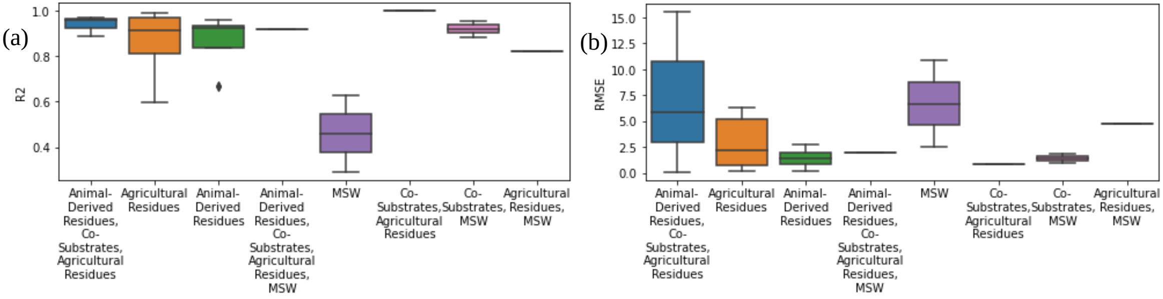

Figure 4: Range of (a) R2 and (b) RMSE by Biomass Category

Animal derived residues, co-substrates and agricultural residues categories with higher median R² values suggest models with better explanatory power for those specific biomass types, indicating that the chosen independent variables are highly effective in predicting the predicted variable for these categories. MSW exhibits a low median in R² and a high IQR in RMSE as the content of MSW is inherently challenging due to its heterogeneous composition, variable generation rates, inconsistent data collection methods, and seasonal variations (Chandra et al., 2021). These complexities lead to lower accuracy and limited predictive power for models dealing with MSW.

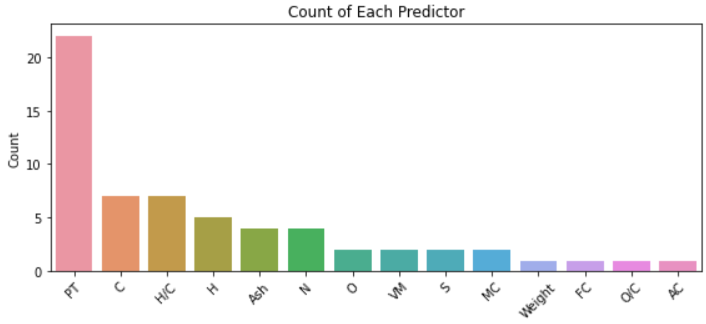

Figure 5: Predictor counts

Figure 5 shows the frequency of predictors that applied in predicting the biochar properties. PT stands out with a higher predictor count, exhibiting strong explanatory power in predicting EC and N properties. For EC prediction, PT demonstrated high significance across both MLR and quadratic models, with R^2 values ranging from 0.92 to 0.93. PT showed high significance across both linear and quadratic models, with R^2 values ranging from 0.89 to 0.99 for N related properties.

4. Conclusions

This research highlights the complexities involved in modelling biochar properties and demonstrates that no single model uniformly outperforms others across all scenarios. While linear models often fall short in capturing the dynamic interactions within biochar systems, non-linear and polynomial models provide greater flexibility in dealing with complex, variable-dependent reactions. MLR emerges as a particularly robust method for operational applications, allowing for the integration of multiple variables that influence biochar's effectiveness in soil improvement. However, its reliance on assumptions about linear relationships may limit its accuracy, highlighting the necessity for integrating MLR with non-linear approaches to fully capture the multifaceted nature of biochar interactions. Future studies should focus on developing hybrid models that merge the predictive power of machine learning algorithms with the interpretability of traditional regression methods. By adopting a strategic approach that combines the strengths of different modelling techniques, practitioners can significantly enhance the precision and applicability of biochar utilization, contributing to more effective environmental management and agricultural practices.

Nomenclature

AC – Active Carbon
C – Carbon
CEC – Cation Exchange Capacity
EC – Electrical Conductivity
HHV – High Heating Value
IQR – Interquartile range
ML – Machine Learning
MLR – Multiple Linear Regression
MSW – Multiple Solid Waste
MC - Moisture Content
N – Sample Size
NLR – Non-linear Regression
n – Number of studies
PT – Pyrolysis Temperature
R^2 – Coefficient of Determination
RMSE – Root Mean Square Error
SSA – Specific Surface Area
VM – Volatile Matter

Acknowledgments

This research was funded by the Malaysian Ministry of Higher Education under the Fundamental Research Grant Scheme (FRGS), grant no. FRGS/1/2019/STG05/UTM/02/9. The authors acknowledge the funding and support provided by the EU project Sustainable Process Integration Laboratory (SPIL), funded as project No. CZ.02.1.01/0.0/0.0/15_003/0000456, through the Czech Republic Operational Programme Research, Development, and Education, Priority 1: Strengthening Capacity for Quality Research, under a collaboration agreement with Universiti Teknologi Malaysia (UTM), Malaysia.

References

Akinwande M.O., Dikko H.G., Samson A., 2015, Variance inflation factor: As a condition for the inclusion of suppressor variable(s) in regression analysis, Open Journal of Statistics, 5(7), 754-767.

Cao H., Milan Y.J., Mood S.H., Ayiania M., Zhang S., Gong X., Lora E.E.S., Yuan Q., Garcia-Perez M., 2021, A novel elemental composition based prediction model for biochar aromaticity derived from machine learning, Artificial Intelligence in Agriculture, 5, 133-141.

Chandra, Sunayana,Kumar, Sunil,Kumar, Rakesh., 2021, Forecasting of municipal solid waste generation using non-linear autoregressive (NAR) neural models, Waste Management, 121. 206-214.

Chen D., Yin L., Wang H., He P., 2014, Pyrolysis technologies for municipal solid waste: a review, Waste Management, 34(12), 2466-2486.

Chen J., Ding L., Wang P., Zhang W., Li J., Mohamed B.A., Chen J., Leng S., Liu T., Leng L., Zhou W., 2022, The estimation of the higher heating value of biochar by data-driven modeling, Journal of Renewable Materials, 10(6), 1555-1574.

Chen M.-W., Chang M.-S., Mao Y., Hu S., Kung C.-C., 2023, Machine learning in the evaluation and prediction models of biochar application: A review, Science Progress, 106(1), 003685042211488.

García R., Gil M.V., Fanjul A., González A., Majada J., Rubiera F., 2021, Residual pyrolysis biochar as additive to enhance wood pellets quality, Applied Energy, 234, 1155-1164.

Gaskin, J. W., Steiner, C., Harris, K., Das, K. C., Bibens, B., 2008, Effect of low-temperature pyrolysis conditions on biochar for agricultural use. Trans Asabe, 51(6), 2061–2069.

Hai A., Bharath G., Patah M.F.A., Daud W.M.A.W., Rambabu K., Show P.L., Banat F., 2023, Machine learning models for the prediction of total yield and specific surface area of biochar derived from agricultural biomass by pyrolysis, Environmental Technology and Innovation, 30(103071), 103071.

Henson R., Friston K., 2007, Convolution models for fMRI, In: Statistical Parametric Mapping, Elsevier, 178-192.

Iqbal M.A., 2020, Application of Regression Techniques with their Advantages and Disadvantages, International Journal of Science and Research, 9(12), 117-120.

Lago B.C., Silva C.A., Melo L.C.A., Morais E.G., 2021, Predicting biochar cation exchange capacity using Fourier transform infrared spectroscopy combined with partial least square regression, Science of the Total Environment, 794(148762), 148762.

Leysieffer F., Moore D.S., 1999, The basic practice of statistics, American Mathematical Monthly, 106(2), 181.

Liu L., Tan Z., Zhang L., Huang Q., 2018, Influence of Pyrolysis Conditions on Nitrogen Speciation in a Biochar ‘Preparation-Application’ Process, Journal of Energy Institute, 91, 916-926.

Mari Selvam S., Balasubramanian P., 2023, Influence of biomass composition and microwave pyrolysis conditions on biochar yield and its properties: A machine learning approach, Bioenergy Research, 16(1), 138-150.

Morais E.G., Silva C.A., Gao S., Melo L.C.A., Lago B.C., Teodoro J.C., Guilherme L.R.G., 2021, Electrical Conductivity and Nitrogen Content in Biochar as Influenced by Pyrolysis Temperature, 2021 IEEE International Conference on RFID Technology and Applications (RFID-TA), IEEE, 203-208.

Mylavarapu R., Nair V., Morgan K., 2013, An Introduction to Biochars and Their Uses in Agriculture, EDIS, 2013(8).

Narde S.R., Remya N., 2022, Biochar production from agricultural biomass through microwave-assisted pyrolysis: predictive modelling and experimental validation of biochar yield, Environment, Development and Sustainability, 24(9), 11089-11102.

Ortiz L.R., Torres E., Zalazar D., Zhang H., Rodriguez R., Mazza G., 2020, Influence of pyrolysis temperature and bio-waste composition on biochar characteristics, Renewable Energy, 155, 837-847.

Rafiq M.K., Bachmann R.T., Rafiq M.T., Shang Z., Joseph S., Long R., 2016, Influence of pyrolysis temperature on physico-chemical properties of corn Stover (Zea mays L.) biochar and feasibility for carbon capture and energy balance, PLoS One, 11(6), e0156894.

Roberts J.J., Cassula A.M., Osvaldo Prado P., Dias R.A., Balestieri J.A.P., 2015, Assessment of dry residual biomass potential for use as alternative energy source in the party of General Pueyrredón, Argentina, Renewable and Sustainable Energy Reviews, 41, 568-583.

Tag, A. T., Duman, G., Ucar, S., Yanik, J., 2016, Effects of feedstock type and pyrolysis temperature on potential applications of biochar. J. Anal. Appl. Pyrolysis, 120, 200–206.

Yaka H., Insel M.A., Yucel O., Sadikoglu H., 2022, A comparison of machine learning algorithms for estimation of higher heating values of biomass and fossil fuels from ultimate analysis, Fuel, 320, 123971.

Zhao Y., Li Y., Yang F., 2022, A state-of-the-art review on modeling the biochar effect: Guidelines for beginners, Science of the Total Environment, 802,149861.