Predictive regression model for crop biomass in clayey soil amended with bamboo biochar

Karolina Villagra-Mendoza[1*], María José Gómez-Astorga[1], Federico Masís-Meléndez[2]

(1.Agricultural Engineering, CETIA Centro de Investigación y Extensión en Tecnología e Ingeniería Agrícola, Instituto Tecnológico de Costa Rica, 159-7050, Cartago, Costa Rica;

2. Chemistry, CEQIATEC, Centro de Investigación y de Servicios Químicos y Microbiológicos, Instituto Tecnológico de Costa Rica, 159-7050, Cartago, Costa Rica)

Abstract: By employing multiple linear regression (MLR) and random forest (RF) algorithms, the intricate interactions of clay soil amended with bamboo biochar were explored, aiming to introduce innovative strategies in precision crop management. The performance of two machine learning-based statistical models were assessed to predict maize biomass following a single application of biochar to clay soil at doses of 25 and 50 tons per hectare, respectively, either alone or in combination with vermicompost, across two cropping cycles. A total of 34 soil input variables, comprising chemical, physical, and hydraulic soil parameters, including pore size distribution, infiltration rate, and hydraulic conductivity, and one target crop biomass, were subjected to correlation analysis. The process of refining model parameters involved assessing feature importance and redundancy to optimize selection. Three models were chosen based on a dataset encompassing the 100th, 50th, and 75th percentiles. Among them, the RF model using the 50th percentile data demonstrated the best fit explaining 60% of the variability of the target crop biomass variable while yielding the smallest root mean square error (RMSE). Notably, all RF models identified potassium (K), phosphorus (P), and the magnesium-to-potassium (Mg/K) ratio as the most influential soil properties for biomass prediction. The RF models represent a valuable tool for predicting crop biomass yield in clay soils amended with biochar.

Keywords: machine learning, bamboo biochar, crop biomass, Maize, random forest, multiple linear regression

Citation: Villagra-Mendoza, K., M. J. Gómez-Astorga, F. Masís-Meléndez. 2025. Predictive regression model for crop biomass in clayey soil amended with bamboo biochar. Agricultural Engineering International: CIGR Journal, 27(1):24-34.

1 Introduction

The diverse interactions between soil and biochar not only affect soil properties but crop productivity. Biochar amendments yield varied outcomes on soil´s chemical, physical, and hydraulic properties, dependent on numerous factors, comprising the feedstock type, biochar composition, thermal production process, application rate, soil texture, and climate conditions. Singh et al. (2022) reviewed different chemical soil variations in response to biochar feedstock. The soil pH showed a notable increase with biochar derived from manure feedstock and applied to coarse-textured soil, while electrical conductivity was observed when biochar derived from herbaceous feedstocks, prepared below 500°C, was applied to fine-textured soils. Likewise,

Received date: 2024-02-24 **Accepted date:** 2024-07-24

***Corresponding author: Karolina Villagra-Mendoza,** Associate Professor**.** Agricultural Engineering, Instituto Tecnológico de Costa Rica. Email: kvillagra@tec.ac.cr. Tel: +506-2550-2876, Fax: +506-2550-2271**.**

the cation exchange capacity exhibited a more positive impact when biochar derived from manure or wood feedstocks was applied to coarse or fine-textured soils (although negatively impacting medium textures), especially at application rates exceeding 40 tons ha^{-1}.

Changes in physical soil properties have been also reported by several authors. Biochar derived from wood, produced between 350°C-400°C and applied at a rate of 30 tons ha^{-1} to a fallow loess soil, reduced bulk density up to 15% through a decrease in total porosity. Consequently, this led to a reduction in soil thermal conductivity and diffusivity (Usowicz et al., 2016). Similarly, incorporating 25% (v/v) peanut hulls biochar, produced at 500°C, into a loamy sand soil resulted in an 18% reduction in bulk density, a 36% decrease in saturated hydraulic conductivity, and an increased in volumetric water content (Githinji, 2014). Moreover, the application of willow biochar pyrolyzed at 320°C to a heavy clay soil resulted in an increased in soil porosity and plant-available water by up to 32% (Rasa et al., 2018). In general, positive effects on bulk density and porosity are observed regardless of the feedstock composition. However, these impacts are particularly pronounced in coarse-textured soils, at pyrolysis temperatures surpassing 500°C, and with application rates surpassing 40 tons ha^{-1} (Singh et al., 2022; Nkoh et al., 2021).

Positive effects in crop yield have been observed more in coarse-grained soils compared to fine-textured soils (Singh et al., 2022). However, high carbon sequestration potential is evident when biochar with a low H/C ratio is applied to fine-textures soils (Nkoh et al., 2021). Application rates exceeding 40 tons ha^{-1}, pyrolysis temperatures below 500°C, and biochar derived from herbaceous and lignocellulosic waste feedstocks contribute significantly to crop productivity (Singh et al., 2022). Moreover, in tropical regions, a potential 25% increase in crop yield may be observed compared to temperate zones due to biochar´s beneficial effect on acidic soils (Nkoh et al., 2021).

The intricate interplay of factors influencing the interaction between biochar and soil properties, coupled with the imperative to optimize agronomic practices and water resources for sustainable crop management, poses a significant challenge. The onset of climate threats presents an additional factor, particularly amidst the escalating climate challenges, further complicating task of understanding the soil-plant complex system for efficient crop management. In recent years, machine learning (ML) has been used to implement predictive models, in an accurate, easy, and fast way, with the ability to learn from a complex, large, and multidimensional dataset, by identifying patterns between input and output variables (Chen et al., 2023; Hai et al., 2023). ML consists of finding a solution without being explicitly programmed to find a specific output (Fashoto et al., 2021). ML algorithms are categorized into three forms, namely supervised, semi-supervised, and unsupervised learning (Chen et al., 2023). Supervised ML uses a trained labeled dataset to predict future interested outputs. Unsupervised ML does not necessitate a labeled dataset as the algorithms infer data relationships autonomously. Semi-supervised ML is the combination of both, the unsupervised and supervised ML methods. It uses both, labeled and unlabeled, datasets to generate an appropriate function (Fashoto et al., 2021). Random forest (RF) is a supervised ML classification and prediction tool, and because of its simplicity, it has been employed in multiple agriculture and environmental-related fields, such as the prediction of efficiency on pollutant removal (Lopez et al., 2021; Asl et al., 2013), land risk assessment and remediation (Wu et al., 2013; Nighojkar et al., 2023), biochar yield (Zhu et al., 2019), biochar effect on crop yield (Xu et al., 2024), and carbon storage (Azzi et al., 2024). Multiple linear regression (MLR) analysis is a

statistical technique and a ML tool that requires several measured-input variables to predict the outcome of one response variable, and due to its simplicity, it has been used to predict crop yield (Shastry et al., 2017).

In Costa Rica, to some extent, approximately 61% of the territory faces issues related to degraded agricultural lands, which are predominantly characterized by their fine-textured soils. Despite its tropical location, the region is experiencing a rise in the frequency and duration of droughts. As a result, implementing sustainable agricultural practices, such as biochar amendments, may offer a solution to mitigate the effects of prolonged dry periods and enhance crop resilience. However, further studies are needed to understand the complex interactions within the biochar-soil-plant relationship and its effects on soil properties and crop production. In this study, we utilized MLR and RF modeling techniques to predict the yield of above ground maize biomass in fine-textured soil amended with biochar. This involved examining the correlation among the physicochemical, hydraulic, and hydrological properties of the soil.

2 Materials and methods

2.1 Experimental design for data collection

The data collection took place between 2020 and 2022 in a greenhouse located at the Experimental Field of the School of Agricultural Engineering at the Costa Rican Institute of Technology in Cartago, Costa Rica.

The study involved a one-time application of biochar at doses of 2.5% and 5% (dry mass percent), equivalent to 25 and 50 tons per hectare, respectively, co-applied with vermicompost by adding 30 g to each plant at sowing over two growing cycles. Six treatments, were established in triplicate plots of 2 m^2 each, following a completely randomized block design. The experimental design consisted of the following treatments: BC0 (control, only soil), BC0CO (soil + vermicompost), BC2.5 (soil + biochar 2.5%), BC2.5CO (soil + biochar 2.5% + vermicompost), BC5 (soil + biochar 5%), BC5CO (soil + biochar 5% + vermicompost). In total, 18 plots were arranged in an array of 3 blocks and 6 treatments. The maize variety DK7500 was selected as the test crop. Two maize cycles were studied: from November 2020 to March 2021 (Cycle I), and February to June 2022 (Cycle II), both corresponding to the dry season. Various physicochemical and hydraulic soil parameters were assessed in the study area, including the soil texture, pH, electrical conductivity, effective cation exchange capacity, concentrations of Ca^{2+}, Mg^{2+} and K^{+}, phosphorous content, total carbon and nitrogen content, water content at field capacity and permanent wilting point, available water content, air capacity, pore size distribution, soil water storage capacity index, air capacity index, hydraulic conductivity, and infiltration. Furthermore, aboveground biomass was measured.

2.2 Data analysis

The influencing factors of maize above-ground biomass were categorized into three groups: 1) the physicochemical soil parameters, 2) hydraulic soil parameters, and 3) hydrological soil parameters. A total of 34 input variables were obtained from two cropping cycles. The output variable was defined as the fresh above-ground biomass.

Linear dependence between any two variables was assessed using Pearson correlation (Equation 1). In cases where a high correlation was observed between two input variables, one variable was omitted from the regression analysis:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\sum_{i=1}^{n}(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (1)$$

where $\bar{x}$ and $\bar{y}$ were the mean values of variables *x* and *y*.

The selected input variables were normalized to a range between 0 and 1. This was achieved by subtracting the minimum value from each individual value and dividing it by the difference between the

maximum and minimum values, as shown in Equation 2:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (2)$$

where x' was the normalized input value, x was the single value, x_{min} and x_{max} were the minimum and maximum values.

2.3 Modelling techniques and performance assessment

Two modeling techniques, MLR and RF were implemented to predict biomass production. The difference between these techniques lies in their approach: MLR utilizes two or more independent variables to predict the outcome of a dependent variable (Equation 3), aiming to minimize the difference between the observed and predicted values, while assuming linearity and independence among predictors (Hai et al., 2023). While the second technique employs trained data to construct and combine predictions randomly from multiple individual decision trees, RF has the added capability to accommodate non-linear relationships and interactions between predictors. Furthermore, RF can offer measures of feature importance, aiding in the identification of the most significant predictors within the model (Hai et al., 2023):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_i x_i \quad (3)$$

In both techniques, the database was randomly divided prior to model construction, with 80% of the original data allocated for training and the remaining 20% reserved for evaluating the model performance. Three predictive models were assessed based on the 100th percentile of the selected independent variables (utilizing all the input parameters), the 50th percentile (incorporating half of the input parameters) and the 75th percentile (employing 25% of the input parameter). Hence, depending on the selected percentile, three RF models were developed, RF100, RF50, RF75; and three MLR models were constructed MLR100, MLR50, and MLR75. The variance inflation factor (VIF) was used to assess multicollinearity among independent variables.

The performance of the developed models was evaluated using the goodness of fit (R^2), root mean squared error (RMSE), and the percent bias (PBIAS) (Zhu et al., 2019). The calculation formulas of R^2, RMSE, and PBIAS, are provided in Equations 4, 5, and 6 respectively:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i^{obs} - Y_i^{pred})^2}{\sum_{i=1}^{n}(Y_i^{obs} - Y^{obs})^2} \quad (4)$$

$$PBIAS = \left(\frac{\sum_{i=1}^{n}(Y^{obs} - Y^{pred})}{\sum_{i=1}^{n} Y^{pred}}\right) \times 100 \quad (5)$$

$$PBIAS = \left(\frac{\sum_{i=1}^{n}(Y^{obs} - Y^{pred})}{\sum_{i=1}^{n}(Y^{pred})}\right) \quad (6)$$

Where, Y_i^{obs} and Y_i^{pred} were the observed and predicted *i*th values, respectively, and $\overline{Y^{\exp}}$ denoted the mean experimental value. The R^2 coefficient was utilized to assess the correlation trend between observed and predicted data, with values ranging from 0 to 1, a value closer to 1 indicated a better-fit model. According to Hai et al. (2023), R^2 values exceeding 0.75 indicate a considerable predictive model, while typically R^2 > 0.5 is considered acceptable (Golmohammadi et al., 2014). *RMSE* helps to identify significant errors in the data (Hai et al., 2023), and lower *RMSE* values indicating higher model accuracy (Arriola-Valverde et al., 2020). *RMSE* values less than half the standard deviation of the observed values is considered indicative of a well-performing predictive model (Golmohammadi et al., 2014). *PBIAS* was utilized to assess whether the predicted data tended to be larger or smaller than their observed counterparts, with an optimal *PBIAS* value of 0.0. Negative values indicated underestimation bias, while positive values indicated overestimation of the predicted values. Low-magnitude values suggested an accurate model simulation (Arriola-Valverde et al., 2020).

Additionally, the Akaike information criterion (*AIC*) (Equation 7) was employed as a model selection criterion among the three MLR models evaluated (100th, 50th and 75th percentile), to

determine the best the best model with the minimum input parameter requirement. The optimal MLR model was selected based on the minimum *AIC* and the lowest calculated error (Mohammed et al., 2015).

$$AIC=(2k)-2ln(L) \quad (7)$$

Where, k is the number of estimated parameters, n the number of observed measurements, and L the value of the likelihood.

3 Results and discussion

3.1 Input parameter correlation

According to the Pearson correlation coefficient matrix obtained (Figure 1), four groups of variables with high correlation were identified. There were two groups associated with soil chemical properties, one group related to soil physical properties, and another group highly correlated with aboveground maize biomass. Within the two groups related to chemical properties, Group 1 comprised, Acidity, Nitrogen (N), Calcium (Ca), Magnesium (Mg), Manganese (Mn), effective cation exchange capacity (ECEC) and electrical conductivity (EC), while Group 2 included pH, acid saturation (AS), Iron (Fe) and the Calcium to Magnesium (Ca/Mg) ratio. The group related to physical soil properties encompassed total porosity (TP), macroporosity (Macro%), air capacity (AC), and air capacity index (ACI), whereas the group associated with crop biomass correlated with Phosphorus (P), Potassium (K), Carbon-to-Nitrogen ratio (CN) and fraction of biochar (BC).

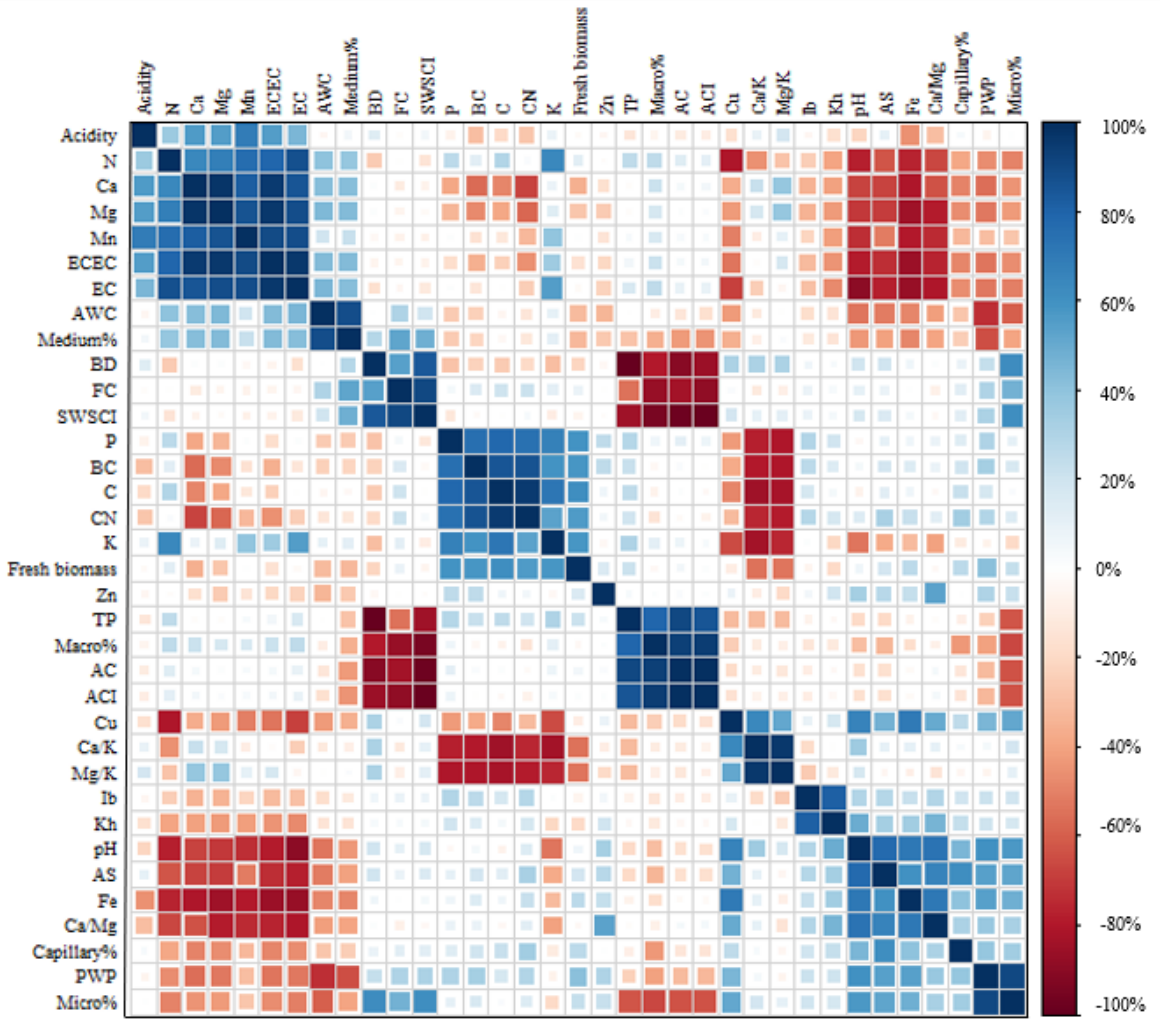

Figure 1 Pearson correlation matrix between any two variables

3.2 Model assessment

The relative importance of the input variables in the model prediction, for the 100th, 50th and 75th percentiles are shown in Figure 2. Features were ranked by their respective importance based on the RF (blue bars) and the MLR (red dots) models. In all

RF models the top three features corresponds to Magnesium to Potassium (Mg/K) ratio, Phosphorus (P) and Potassium (K). Contrarily, the most significant features in the MLR models vary depending on the selected percentile. For MLR100, the top three input variables are Available Water Content (AWC), Manganese (Mn) and pH. In MLR50, the variables P, K and Mg/K ratio (like RF models) hold greater importance, whereas in MLR75, Calcium (Ca), K, and permanent wilting point (PWP) are among the top features. The collective significance of input parameters P, K, and Mg/K ratio (a fertility indicator), aligns with the findings of the review from Jindo et al. (2020). This review suggests that biochar enhances the plant-available P and K, thus presenting a viable option for sustainable agriculture in the long term.

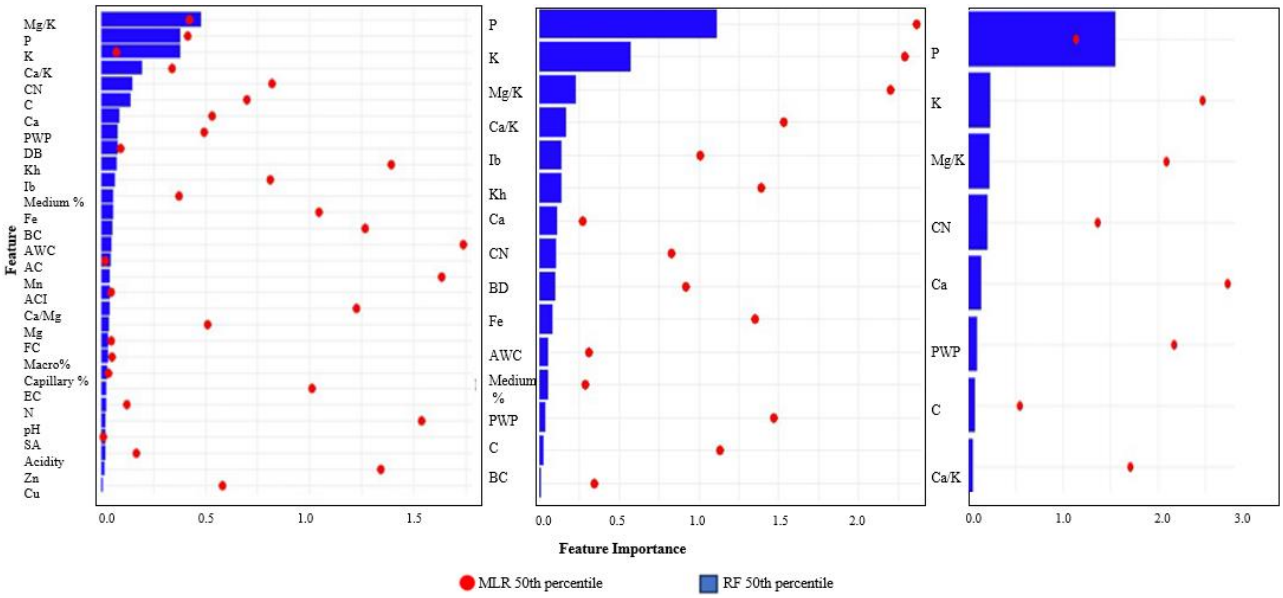

Figure 2 Important features obtained from the prediction models, random forest (blue bars), and multiple linear regression (red dots) with input parameters based on the 100th, 50th, and 75th percentile

Table 1 Coefficients of the RF and MLR model parameters at the 100th, 50th, and 75th percentiles

Parameters	RF100	RF50	RF75	MLR100	MLR50	MLR75
Intercept				-2.1240	-0.1469	0.2587
BC		0.0570		-0.5498	-0.0662	
pH	0.0225			1.6911		
Acidez	0.0195			-0.0812		
Ca		0.0885		-1.5449	-0.1236	-0.6623
Mg	0.0366			-2.1039		
K		0.3807		-0.0905	0.4403	0.5129
AS	0.0202			-0.0041		
P		0.3820		-0.1286	0.4248	0.2106
Zn	0.0159			0.5124		
Cu	0.0073			-0.2267		
Fe		0.0578		0.8558	0.3924	
Mn	0.0439			1.0911		
EC	0.0262			2.6439		
C		0.1431		-1.3677	-0.4168	0.2016
N	0.0241			0.1562		
CN	0.1492		0.1492	1.3674	0.3750	-0.5899
Ca/Mg	0.0420			-1.4646		
Ca/K	0.1976		0.1976	1.2579	-0.9271	-0.8810

Parameters	RF100	RF50	RF75	MLR100	MLR50	MLR75
Mg/K		0.4799		-1.6530	1.0687	1.0133
Ib		0.0668		-0.1377	-0.1628	
Kh		0.0776		-0.2917	-0.2289	
BD		0.0786		0.8136	-0.1291	
AC	0.0451			0.2336		
AWC		0.0513		0.6510	-0.0690	
FC	0.0356			-0.3523		
PWP		0.0798		1.1744	0.2588	0.3036
Macro%	0.0350			0.6887		
Capillary%	0.0284			0.1512		
Medium%		0.0603		0.9277	0.0745	
ACI	0.0432			0.5699		

RF and MLR model coefficients are shown in Table 1. Common input parameters such as Ca, K, P, C, Mg/K, and PWP were observed for all RF and MLR models. Additionally, in the RF models, the percentage of medium-sized pores was a common parameter, while in the MLR models, the CN and the Ca/K ratio were also frequently present. Overall, the models acknowledged the significance of soil parameters such as liming, carbon content, fertility, water holding capacity in the dry zone (up to a matric potential of -1500 kPa), and nutrient availability such as P and K. The RF50 and MLR50 models considered soil chemical parameters and input variables related soil capacity and intensity, whereas the RF75 and MLR75 predominantly included soil chemical parameters. The importance of the soil parameters has been highlighted in various studies, as evidenced by the impact of biochar on soil nutrient availability (Al-Wabel et al., 2018), its contribution to carbon sequestration (Semida et al., 2019; Lorenz and Lal, 2014), alterations in hydraulic conductivity and infiltration (Villagra-Mendoza and Horn, 2019; Githinji, 2014), modifications in soil water retention capacity (Rasa et al., 2018), and changes in pore size distribution, ultimately influencing crop productivity (Vijay et al., 2021).

The scatter plots in Figure 3 depict the observed and predicted aboveground maize biomass for models based on the 100th, 50th and 75th percentiles. Based on the visual analysis of the observed and predicted crop biomass, the overall performance of the RF models seems reasonably accurate. However, the MLR100 model fails to deliver acceptable results at low biomass outputs, while this issue is also observed with lower significance for MLR50 and MLR75.

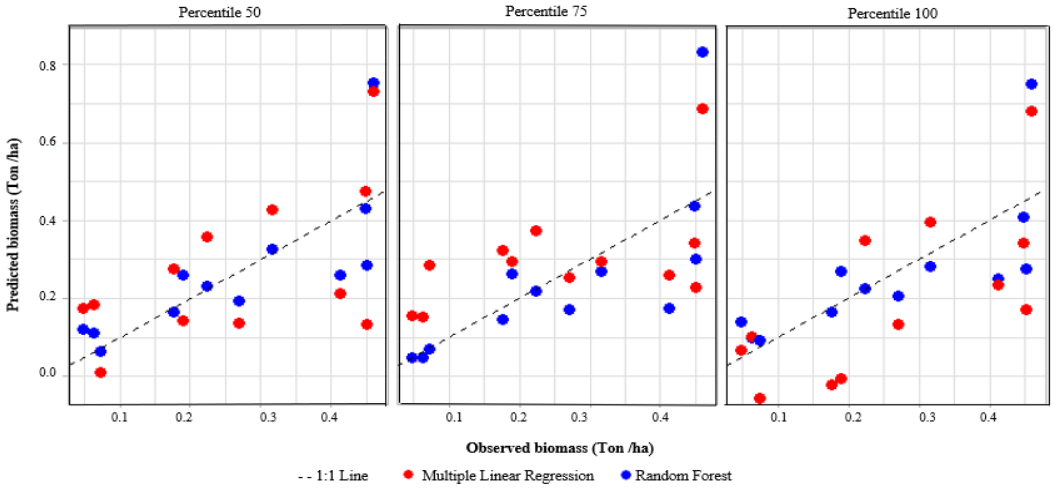

Figure 3 Performance of crop biomass predictive random (RF) forest and multiple linear regression (MLR) models based on the 100th, 50th, and 75th percentiles

The performance of the models in relation to the simulated crop biomass was further validated using statistical criteria, as summarized in Table 2. Improved model performances were accepted if R^2 approached unity, RMSE approached zero and was less than half the standard deviation of the observed biomass, and if PBIAS exhibited small relative values, as suggested by Arriola-Valverde et al. (2020). Additionally, the selection of model criteria encompassed the AIC criterion specifically for the MLR models. Based on the R^2 value, the RF50 model exhibited a satisfactory fit, explaining 60% of the observed variability in the aboveground biomass of the crop (Golmohammadi et al., 2014). The RF50 model exhibited the smallest *RMSE*, which remained significantly lower than half the standard deviation of the observed values (*RMSE*=0.1140). Comparing RMSE among the models, MLR75 also presented a small value, but R^2 was insufficient to deem it a satisfactory fit. *PBIAS* was close to zero across all RF models, indicating minor over-estimation for RF100 and RF50, and an under-estimation for RF75. In contrast, the MLR models showed slightly larger PBIAS values, consistently under-estimating the predicted biomass. Despite the AIC selection criterion favoring MLR100 as the optimal model, the error data did not robustly support the preference of a MLR model over a RF model. Considering the minimal data variability and the most accurate predicted trend, the RF50 model emerged as the best-fitting model.

Table 2 Statistical metrics including R^2, RMSE, PBIAS and the AIC, for the predictive RF and MLR models using the 100th, 50th and 75th percentiles of the dataset

Model	R^2	RMSE	PBIAS	AIC
RF 100	0.5405	0.0138	0.8870	-
RF 50	0.6113	0.0119	1.9573	-
RF75	0.5393	0.0205	-4.3323	-
MLR 100	0.4799	0.0255	-47.3647	-56.70
MLR 50	0.3018	0.0268	-47.6685	-43.73
MLR75	0.3239	0.0191	-35.3835	-33.30

4 Conclusions

The observed maize aboveground biomass in a clayey soil amended with biochar served as the basis for evaluating the performance of MLR and a RF model. The models were constructed using a three-level selection of input parameters at the 100th, 50th, and 75th percentiles. Across all RF models, the top three input parameters consistently centered around nutrient availability (P and K) and fertility (Mg/K). The MLR models varying degrees of feature importance depending on input parameters (percentile) employed in each model. In addition to the top three parameters in RF models, features such as Ca, C and PWP were utilized across all models to predict biomass. This underscores the significance of including the effects of liming and water content at very negative matric potentials into the prediction models. Following the evaluation of the models using the statistical metrics including R^2, *RMSE*, *PBIA*S, and the selection criterion *AIC*, the RF50 model delivered the most favorable statistical performance.

Acknowledgments

This study was funded by the Vice-rector of Research and Extension of the Costa Rican Institute of Technology. The authors would like to thank Fernando Watson for the support provided on the statistical analysis.

References

Al-Wabel, M. I., Q. Hussain, A. R. A. Usman, M. Ahmad, A. Abduljabbar, A. S. Sallam, and Y. S. Ok. 2018. Impact of biochar properties on soil conditions and agricultural sustainability: A Review. *Land Degradation and Development*, 29(7): 2124–2161.

Arriola-Valverde, S., L. C. Villalobos-Avellán, K. Villagra-Mendoza, and R. Rimolo-Donadio. 2020. Erosion quantification in runoff agriculture plots by multitemporal high-resolution uas digital photogrammetry. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13: 6326-6336.

Asl, S. M. H., M. Ahmadi, M. Ghiasvand, A. Tardast, and R. Katal. 2013. Artificial neural network (ANN) approach for modeling of Cr (VI) adsorption from aqueous solution by zeolite prepared from raw fly ash (ZFA). *Journal of Industrial and Engineering Chemistry*, 19(3): 1044–1055.

Azzi, E. S., H. Li, H. Cederlund, E. Karltun, and C. Sundberg. 2024. Modelling biochar long-term carbon storage in soil with harmonized analysis of decomposition data. *Geoderma*, 441: 116761.

Chen, M. W., M. S. Chang, Y. Mao, S. Hu, and C. C. Kung. 2023. Machine learning in the evaluation and prediction models of biochar application: A review. *Science Progress*, 106(1): 1-22.

Fashoto, S. G., E. Mbunge, G. Ogunleye, and J. Van Den Burg. 2021. Implementation of machine learning for predicting maize crop yields using multiple linear regression and backward. *Malaysian Journal of Computing*, 6(1): 679–697.

Githinji, L. 2014. Effect of biochar application rate on soil physical and hydraulic properties of a sandy loam. *Archives of Agronomy and Soil Science*, 60(4): 457–470.

Golmohammadi, G., S. Prasher, A. Madani, and R. Rudra. 2014. Evaluating three hydrological distributed watershed models: MIKE-SHE, APEX, SWAT. *Hydrology*, 1(1): 20–39.

Hai, A., G. Bharath, M. F. A. Patah, W. M. A. W. Daud, K. Rambabu, P. Show, and F. Banat. 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.

Jindo, K., Y. Audette, F. S. Higashikawa, C. A. Silva, K. Akashi, G. Mastrolonardo, M. A. Sánchez-Monedero, and C. Mondini. 2020. Role of biochar in promoting circular economy in the agriculture sector. Part 1: A review of the biochar roles in soil N, P and K Cycles. *Chemical and Biological Technologies in Agriculture*, 7(1): 15.

Lopez, A. M., A. Wells, and S. Fendorf. 2021. Soil and aquifer properties combine as predictors of groundwater uranium concentrations within the Central Valley, California. *Environmental Science and Technology*, 55(1): 352–361.

Lorenz, K., and R. Lal. 2014. Biochar application to soil for climate change mitigation by soil organic carbon sequestration. *Journal of Plant Nutrition and Soil Science*, 177(5): 651–670.

Mohammed, E. A., Ch. Naugler, and B. H. Far. 2015. Emerging business intelligence framework for a clinical laboratory through big data analytics. In *Emerging Trends in Computational Biology, Bioinformatics, and Systems Biology*, eds Q. N. Tran, and H. Arabnia, 577–602. Massachusetts, United States, Elsevier.

Nighojkar, A., Sh. Pandey, M. Naebe, B. Kandasubramanian, W. W. Soboyejo, A. Plappally, and X. Wang. 2023. Using machine learning to predict the efficiency of biochar in pesticide remediation. *Npj Sustainable Agriculture*, 1(1): 1.

Nkoh, J. N., M. Abdulaha-Al Baquy, S. Mia, R. Shi, M. A. Kamran, K. Mehmood, and R. Xu. 2021. A critical-systematic review of the interactions of biochar with soils and the observable outcomes. *Sustainability*, 13(24): 13726.

Rasa, K., J. Heikkinen, M. Hannula, K. Arstila, S. Kulju, and J. Hyväluoma. 2018. How and why does willow biochar increase a clay soil water retention capacity? *Biomass and Bioenergy*, 119: 346–353.

Semida, W. M., H. R. Beheiry, M. Sétamou, C. R. Simpson, T. A. Abd El-Mageed, M. M. Rady, and S. D. Nelson. 2019. Biochar implications for sustainable agriculture and environment: a review. *South African Journal of Botany*, 127: 333–347.

Shastry, A., H. A. Sanjay, and E. Bhanusree. 2017. Prediction of crop yield using regression techniques. *International Journal of Soft Computing*, 12(2): 96–102.

Singh, H., B. K. Northup, C. W. Rice, and P. V. V. Prasad. 2022. Biochar applications influence soil physical and chemical properties, microbial diversity, and crop productivity: A meta-analysis. *Biochar*, 4(1): 8.

Usowicz, B., J. Lipiec, M. Łukowski, W. Marczewski, and J. Usowicz. 2016. The effect of biochar application on thermal properties and albedo of loess soil under grassland and fallow. *Soil and Tillage Research*, 164: 45–51.

Vijay, V., S. Shreedhar, K. Adlak, S. Payyanad, V. Sreedharan, G. Gopi, T. S. van der Voort, P. Malarvizhi, S. Yi, J. Gebert, and P. B. Aravind. 2021. Review of large-scale biochar field-trials for soil amendment and the observed influences on crop yield variations. *Frontiers in Energy Research*, 9: 710766.

Villagra-Mendoza, K., and R. Horn. 2019. Changes in water infiltration after simulated wetting and drying periods in

two biochar amendments. *Soil Systems*, 3(4): 63.

Wu, G., C. Kechavarzi, X. Li, S. Wu, S. J. T. Pollard, H. Sui, and F. Coulon. 2013. Machine learning models for predicting PAHs bioavailability in compost amended soils. *Chemical Engineering Journal*, 223: 747–754.

Xu, X., T. Li, K. Cheng, Q. Yue, and G. Pan. 2024. Geographical differences in the effect of biochar on crop yield and greenhouse gas emissions – A global simulation based on a machine learning model. *Current Research in Environmental Sustainability*, 7: 100239.

Zhu, X., Y. Li, and X. Wang. 2019. Machine learning prediction of biochar yield and carbon contents in biochar based on biomass characteristics and pyrolysis conditions. *Bioresource Technology*, 288: 121527.

Acronyms

Acronym	Parameter	Unit
BC	Fraction of biochar	-
pH	Potential of hydrogen	-
Ca	Calcium	cmol(+) L^{-1}
Mg	Magnesium	cmol(+) L^{-1}
K	Potassium	cmol(+) L^{-1}
Acidity	Acidity	cmol(+) L^{-1}
ECEC	Effective cation exchange capacity	cmol(+) L^{-1}
AS	Acid saturation percentage	%
P	Phosphorous	mg L^{-1}
Zn	Zinc	mg L^{-1}
Cu	Copper	mg L^{-1}
Fe	Iron	mg L^{-1}
Mn	Manganese	mg L^{-1}
EC	Electrical conductivity	mS cm^{-1}
C	Total carbon	%
N	Total nitrogen	%
CN	Carbon-to-nitrogen (C/N) ratio	-
Ca/Mg	Calcium-to-magnesium ratio	-
Ca/K	Calcium-to-potassium ratio	-
Mg/K	Magnesium-to-potassium ratio	-
Ib	Basic infiltration	mm h^{-1}
Kh	Hydraulic conductivity	mm h^{-1}
BD	Soil bulk density	g cm^{-3}
AC	Soil air capacity	cm^3 cm^{-3}
AWC	Available water content	cm^3 cm^{-3}
FC	Field capacity	cm^3 cm^{-3}
PWP	Permanent wilting point	cm^3 cm^{-3}
TP	Total porosity	%
Macro%	Macropores percentage	%
Capillary%	Capillary pores percentage	%
Medium%	Medium pores percentage	%
Micro%	Micropores percentage	%
SWSCI	Soil water storage capacity index	-
ACI	Air capacity index	-