Scholars Journal of Physics, Mathematics and Statistics
OPEN ACCESS
Abbreviated Key Title: Sch J Phys Math Stat
ISSN 2393-8056 (Print) | ISSN 2393-8064 (Online)
Journal homepage: https://saspublishers.com

Modelling of Aerial Biomass and Its Physicochemical Properties Using Artificial Intelligence and Response Surface Methodology

Kupolusi, Joseph A[1*], Adedeji, Shalom Odunayo[1]

[1]Department of Statistics, Federal University of Technology Akure, Nigeria

DOI: https://doi.org/10.36347/sjpms.2025.v12i03.001 | **Received:** 05.04.2024 | **Accepted:** 08.05.2024 | **Published:** 05.03.2025

***Corresponding author:** Kupolusi, Joseph A
Department of Statistics, Federal University of Technology Akure, Nigeria

Abstract **Original Research Article**

Aerial biomass and its associated physicochemical properties are pivotal components in understanding ecosystem dynamics, with far-reaching implications for environmental and agricultural management. This study delves into the dynamics of aerial biomass and its correlation with physicochemical properties, utilizing a dataset derived from research on Spartina alterniflora in the Cape Fear Estuary of North Carolina. Machine learning (ML) models, including Response Surface Methodology (RSM), the Quadratic Model, Artificial Neural Networks (ANN), and the Adaptive Neuro-Fuzzy Inference System (ANFIS), are employed to evaluate the relationships between these properties and aerial biomass. The RSM model outperformed other models with a remarkable Mean Squared Error (MSE) of 0.0579 and a high R-squared value of 0.9518, emphasizing its efficiency in estimating biomass. The quadratic model follows closely, with an MSE of 0.1778 and an R-squared value of 74.61%, providing valuable insights into biomass variation. Furthermore, Particle Swarm Optimization (PSO) is applied to optimize the models. The results highlight RSM-PSO as the most efficient technique, with a PSO value of 0.9872 and an ML R-squared of 0.9518, underscoring the robustness of the RSM model when combined with PSO for predicting aerial biomass. The findings emphasize the significance of pH and potassium content in biomass estimation and recommend the RSM model, particularly when coupled with PSO, for efficient biomass prediction. These insights have critical implications for environmental and agricultural management and may serve as a valuable tool for ecosystem optimization.

Keywords: Aerial Biomas, Artificial Intelligence, Response Surface Methodology, Particle Swarm Optimization, Artificial Neural Network, Adaptive Neuro-Fuzzy Inference System.

1.0 INTRODUCTION

Aerial biomass, consisting of above-ground plant components such as leaves, stems, branches, and reproductive structures, plays a pivotal role in shaping ecosystems, supporting agriculture, and contributing to the renewable energy sector (Petráš *et al*., 2021). The physicochemical properties of aerial biomass, encompassing attributes such as elemental composition, calorific value, moisture content, and nutrient concentrations, exert a profound influence on various ecological, agricultural, and technological processes (Aal *et al*., 2023). Aerial biomass serves as a key indicator of ecosystem health and functionality. It contributes to the terrestrial carbon cycle by capturing atmospheric carbon dioxide through photosynthesis, thus mitigating the impacts of climate change (Onyeaka *et al*., 2021).

In agriculture, the study of aerial biomass and its physicochemical properties is critical for optimizing crop yield and ensuring long-term sustainability (Talaviya *et al*., 2020). Aerial biomass contributes to crop architecture, light interception, and photosynthetic efficiency (Nhamo *et al*., 2020). Accurate modeling and analysis of aerial biomass attributes can guide agricultural practices, allowing farmers to make informed decisions about planting densities, irrigation strategies, and nutrient management (Nhamo *et al*., 2020). The utilization of aerial biomass as a feedstock for bioenergy production offers a promising avenue for achieving renewable energy goals (Jekayinfa *et al*., 2020). The physicochemical properties of biomass influence its suitability for conversion processes such as pyrolysis and fermentation, which yield biofuels like bioethanol and biogas (Ahorsu *et al*., 2018). Understanding the relationships between aerial biomass attributes and energy conversion efficiency is essential for optimizing bioenergy production (Batista *et al.*, 2023)

Despite the significance of aerial biomass and its physicochemical properties, the complexities inherent in their interrelationships pose challenges that warrant thorough investigation (P. Ralevic *et al*., 2010). Existing approaches to studying aerial biomass often rely on conventional methodologies that may be limited in scope, time-consuming, and prone to subjective biases. Furthermore, the intricate web of factors influencing aerial biomass composition and its physicochemical attributes demands a comprehensive analytical framework that can capture and optimize these multifaceted interactions. While Artificial Intelligent (AI) and Response Surface Methodology (RSM) have individually demonstrated remarkable potential in various domains, their integrated application to tackle the intricate dynamics of aerial biomass and its physicochemical properties remains relatively unexplored (Zhang *et al*., 2020).

In this context, the fusion of artificial intelligence (AI) and advanced statistical methodologies like response surface methodology (RSM) has emerged as a transformative avenue for unraveling the complexities of natural systems (Zhang & Wu, 2021). This research seeks to delve into the synergistic application of AI and RSM, focusing on the modeling and optimization of aerial biomass and its physicochemical properties. By employing cutting-edge AI algorithms, such as machine learning, alongside RSM's established optimization techniques, this research intends to construct accurate predictive models and identify optimal conditions for aerial biomass production and quality enhancement.

Table 1.1: Summary of Machine Learning Application studies

Area of Application	**Related Reference(s)**	**ML method**	**Location**	**Input**	**Prediction Type**
Health	(Dev *et al.*, 2022)	Principal Component Analysis	EHR dataset	correlation, stepwise analysis and machine learning algorithms	Value
Stroke type mortality	(Mainali *et al*., 2021)	Random forest		Mortality, non-invasive variables, stroke type	Classification
Vessel Operation	(Zhao *et al*., 2018)	Clustering	Vessel trade routes across the globe	Locations of vessels	Action
Asset Maintenance	(Oneto *et al.*, 2015)	Regularized kernel least square, SVM	Simulated data	Health status of equipment's	Value
Fraud detection	(Muhammad, 2021)	LR, SVM, RF, classification tree	AMAZON data		Classification

Adapted from: Barua and Zou (2021)

2.0. RESEARCH AND METHODS

The data reported in this study is from secondary sources obtained from research on factors affecting Spartina alterniflora growth and dieback in the Cape Fear Estuary of North Carolina (Linthurst, 1980). The data obtained was analyzed using the various machine learning algorithms presented below.

2.1 Data Preprocessing

The preprocessing of the dataset, which involves looking for missing values and incorrect inputs, is the first step in the research technique. The proper resolution of these problems may entail calculating missing values and removing incorrect entries.

2.2 Response Surface Methodology

Response Surface Methodology (RSM) is a statistical and mathematical approach to modeling the relationship between multiple input variables (predictors) and a single output variable (response) in order to optimize the response. RSM is commonly used in the design of experiments to identify the optimal levels of input variables that result in the maximum (or minimum) response. RSM involves fitting a regression model to the experimental data, typically using second-degree polynomials, and using statistical methods to identify the critical factors that affect the response. The fitted model can then be used to make predictions, find the optimal input values, and understand the nature of the relationship between the inputs and responses. Response Surface Methodology (RSM) uses a mathematical model to describe the relationship between the input variables and the response. The most common model used in RSM is a second-degree polynomial regression model. The general form of the model is given by (1)

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_{jj} x_j + \sum_{j=1}^{k} \beta_j x^2{}_j + \sum_{i<}^{k} \sum_{j=2}^{k} \beta_{ij} x_i x_j + \varepsilon_i \quad (1)$$

Where:
Y is the response variable.
$X_1, X_2, \cdots, X_k$ are the input variables factor
$\beta_{11}, \beta_{22}, \ldots, \beta_{ij}$ are the regression coefficients are the interaction terms,
and ε is the residual error.

The regression coefficients can be estimated using statistical techniques such as least-squares regression or maximum likelihood estimation. Once the model is fit, it can be used to make predictions, optimize the response, and understand the nature of the relationship between the inputs and response. For example, the model can be used to identify the levels of the input variables that result in the maximum or minimum response or to determine the optimal combination of input variables that results in the desired response. RSM also provides a measure of the goodness of fit of the model and can be used to perform hypothesis tests to determine the significance of the input variables and interaction terms in explaining the variation in the response.

The application of Response Surface Methodology (RSM) to model and optimize the growth and dieback of Spartina alterniflora involves considering aerial biomass, aeration, nitrogen, potassium, pH, and salinity as factors that affect the growth. The goal of the economic model is to identify the optimal levels of these factors to maximize the growth of Spartina alterniflora while minimizing dieback. The RSM model would use statistical techniques to analyze the data collected on the factors and their effect on the growth and dieback of the plant and to determine the optimal levels of these factors. The results of the RSM model could then be used to make informed decisions about the management of the growth and dieback of Spartina alterniflora and to optimize the economic benefits associated with its growth.

2.3 Adaptive Neuro-Fuzzy Inference System (ANFIS)

ANFIS is a powerful hybrid intelligent system that merges the strengths of fuzzy logic and neural networks. It is designed to model and understand complex relationships between input and output variables. The key idea behind ANFIS is to use a fuzzy inference system to capture the linguistic and symbolic aspects of human knowledge while employing a neural network to adaptively adjust the parameters of the fuzzy system based on data. The initial step involves transforming crisp, numerical input data into fuzzy sets using predefined membership functions. These membership functions capture the degree of belongingness of input data to different linguistic labels or categories, such as "low," "medium," or "high." A rule base is constructed, comprising a set of if-then rules that relate the fuzzy input variables to the fuzzy output variables. These rules can be determined through expert knowledge, data-driven methods, or a combination of both. Each rule consists of an antecedent (the 'if' part) and a consequent (the 'then' part). ANFIS employs the rules and the degree of membership of input data in each linguistic label to make fuzzy inferences. This process calculates the strength of each rule's contribution to the overall output. The outputs of individual rules are aggregated into a single fuzzy set. Common aggregation methods include the maximum (max) or summation (sum) of rule strengths. The final step is to convert the aggregated fuzzy set into a crisp, numerical output value. This numerical output is interpretable and serves as the model's prediction or decision.

2.4 Artificial Neural Network (ANN)

An artificial neural network (ANN) is a computational model that draws inspiration from the structure and functionality of the human brain. ANNs are composed of interconnected nodes, known as neurons, that process information and transmit it to other neurons. These neurons are organized into layers, typically including an input layer, one or more hidden layers, and an output layer. Neurons within a layer are interconnected, and these connections are associated with weights. Each neuron employs an activation function, such as the sigmoid or ReLU, to determine its output based on the weighted sum of its inputs. The introduction of non-linearity through activation functions allows ANNs to model complex relationships.

ANNs can be trained to learn from data through a process called training. During training, the network adjusts the weights of its connections to minimize a predefined error or loss function, typically using the backpropagation algorithm. This enables ANNs to learn complex relationships between input and output variables, making them capable of tasks such as pattern recognition, prediction, and classification. One of the distinguishing features of ANNs is their ability to generalize from the training data, meaning they can make predictions on new, unseen data that was not part of the training set. This generalization is crucial for the practical applications of ANNs in various fields. Artificial neural networks have found applications in diverse domains, including image and speech recognition, natural language processing, recommendation systems, financial forecasting, autonomous vehicles, and biological and medical applications. They continue to be a rapidly evolving field with ongoing research and development, leading to improved network architectures, training techniques, and expanded applications across various industries.

2.5 Criteria for Comparison

Statistical measures including the correlation coefficient (R), coefficient of determination (R2), adjusted R^2, mean square error (MSE), and average absolute deviation (AAD) were employed to evaluate the developed models' predictive efficacy.

2.5.1 Mean Square Error Estimator (MSE)

If we have a scalar parameter to be estimated and the statistic is used as an estimator, then the mean square error is given as:

$$MSE(\theta,\beta) = E((\theta-\beta)^2,\beta) \quad (2)$$
$$= var(\theta,\beta) + (b\beta)^2 \quad (3)$$
$$MSE = E(\theta-\beta)^2 \quad (4)$$
$$= E(\theta - E(\theta) + E(\theta) - \beta)^2 \quad (5)$$
$$MSE = Var(\theta) + [b(\beta)]^2 \quad (6)$$

Meaning that mean square error is the sum of variance estimator and squared of bias estimator and is efficient if smaller value is obtained compare to other estimators. In the case of unbiased estimators, it is just the ratio of their variance and the one with smaller variance will be more efficient if among all the unbiased estimators of $\hat{\beta}$ is the one with the smallest variance then it been called the most efficient estimator of, that is for two unbiased estimators $\tilde{\beta}_1$ and $\tilde{\beta}_2$ for the parameter with variance $var(\tilde{\beta}_1)$ and $var(\tilde{\beta}_2)$ respectively, the efficiency of $\tilde{\beta}_1$ relative $\tilde{\beta}_2$ is defined by;

$$e(\hat{\beta}_1,\hat{\beta}_2) = \frac{V(\hat{\beta}_2)}{V(\hat{\beta}_1)} \quad (7)$$

$$MSE = \frac{\sum_{i-1}^{n}(y_i-\hat{y}_i)^2}{n-p} = \frac{\sum_{i=1}^{n} e_i^2}{n-p} \quad (8)$$

2.5.2 Coefficient of Determination (R^2)

The coefficient of determination is the square of the Pearson's correlation coefficient between X and Y. It is expressed mathematically as;

$$r^2 = \frac{SSR}{SST} \quad (9)$$

$$r^2 = 1 - \frac{SSE}{SST} \quad (10)$$

$$R^2 = \left(\frac{n\Sigma xy - \Sigma x\Sigma y}{\sqrt{n\sum x^2-(\sum x)^2}\sqrt{n\sum y^2-(\sum y)^2}}\right)^2 \quad (11)$$

For the general linear model

$$Y = X\beta + \varepsilon \quad (12)$$

The coefficient of determination (R^2) is defined as follows;

$$R^2 = \frac{\hat{\beta}'X'Y - n\bar{y}'}{Y'Y - n\bar{Y}'} \quad (13)$$

Where;
SSR = Residual sum of squares
SST = Total sum of square
x = the explanatory variable(s)
y = response variable(s)
n = number of observations

2.6 Process Parameter Optimization

Process parameter optimization is a systematic approach used to fine-tune and improve the parameters of a given process with the goal of enhancing its performance, efficiency, or quality. In this process, the first step is to identify the parameters that significantly affect the outcome, which could include factors like temperature, pressure, or flow rate. An appropriate experimental design, such as factorial designs or response surface methodology, is chosen to systematically vary these parameters and collect data on the process's performance. These models can range from simple linear equations to more complex polynomial equations or machine learning models like neural networks. After the model is established, optimization techniques are used to find the best combination of parameter values that either maximizes or minimizes the response variable. Common optimization algorithms include gradient descent, genetic algorithms, or particle swarm optimization.

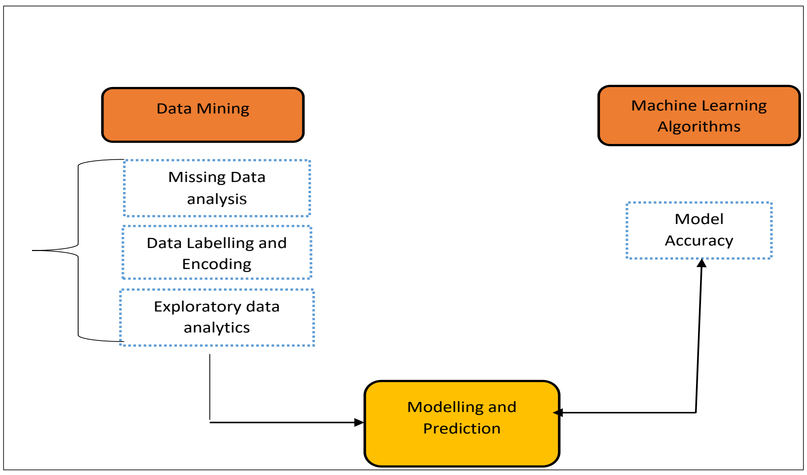

Figure 2.1: Proposed System's block diagram
Adapted from: Tazin *et al.*, (2021)

3.0 RESULTS AND DISCUSSION

3.1 Descriptive Statistics of the Variables

Table 3.1 provides a comprehensive summary of key statistics for variables related to aerial biomass and its physicochemical properties. Focusing on the most significant values, the average biomass (Bio) is found to be 1000.8 units, with a considerable standard deviation of 660.08, implying a notable variability in biomass values. Ranging from 236 to 2436 units, the distribution illustrates a diverse spectrum, and its skewness of 0.60 indicates a moderate rightward skewness, suggesting the presence of higher biomass values. Moving to salinity (Sal), the mean concentration is 30.27, exhibiting moderate variability with a standard deviation of 3.72. Ranging from 24 to 38, the dataset encompasses a moderate span of salt concentrations. The skewness of 0.31 suggests a mild rightward skewness, indicating a slight tail to the right in the distribution of salinity. Analyzing acidity and alkalinity (pH), the average pH level is 4.6, denoting an acidic to neutral environment. With a standard deviation of 1.25, the dataset displays variability, and its pH values range from 3.2 to 7.45. The skewness of 0.90 indicates a moderate rightward skewness, hinting at the presence of higher pH values and a tail to the right. Regarding potassium content (K), the mean is 797.62, indicating a typical amount. The dataset exhibits variability with a standard deviation of 297.6, and potassium values range from 350.73 to 1441.67. A mild rightward skewness of 0.48 suggests a slight tail to the right in the distribution of potassium content. Considering sodium content (Na), the average is 16596.73, displaying considerable variability with a standard deviation of 6882.48. Ranging from 7886.5 to 35185.5, sodium content spans a broad spectrum of values. The skewness of 0.82 indicates a moderate rightward skewness, signifying a tail to the right in the distribution of sodium levels. Lastly, examining zinc content (Zn), the mean is 17.88, and the dataset shows variability with a standard deviation of 8.28. Ranging from 0.21 to 31.29, the distribution illustrates variability in zinc content. The skewness of -0.68 suggests a moderate leftward skewness, indicating a tail to the left in the distribution of zinc content.

Table 3.1: Summary Statistics

Variables	Mean	SD	Min	Max	Skewness
Bio	1000.8	660.08	236	2436	0.60
Sal	30.27	3.72	24	38	0.31
PH	4.6	1.25	3.2	7.45	0.90
K	797.62	297.6	350.73	1441.67	0.48
Na	16596.73	6882.48	7886.5	35185.5	0.82
Zn	17.88	8.28	0.21	31.29	-0.68

3.2 Application of Machine Learning Algorithms

Table 3.2 shows the performance comparison of the models. The result demonstrates that of the various models formulated, the RSM model stands out with an impressive MSE of 0.0579 and a high R-squared value of 0.9518. This was followed by the quadratic model with a MSE value of 0.1778 and an R square value of 74.61%. The ANN model and ANFIS model did not perform absolutely well, with R square values of 27.36% and 14.56%, respectively. Hence, the response surface model (RSM) is chosen to be the best model for estimating the effect of physicochemical properties on aerial biomass since it performed much better than other models.

Table 3.2: Performance of the Developed Machine Learning Models

Models	MSE	R squared	AAD
RSM	0.0579	0.9518	0.2821
ANN	0.5785	0.2736	0.7059
ANFIS	0.6219	0.14565	0.6412
Quadratic Model	0.1778	0.7461	0.3208

3.3 Model Estimation

The regression model for the optimized aerial biomass and psychochemical properties is presented in Table 3.3. The intercept coefficient of 1309.57 is the expected value of the dependent variable when all of the independent variables are equal to zero. The p-value of 0.000295 indicates that this intercept value is statistically significant, meaning that we can reject the null hypothesis that the true intercept is zero. The estimated coefficients for the independent variables suggest that PH and Na have a positive effect on the dependent variable, while Sal, K, and Zn have a negative effect. The estimated coefficient for PH is positive but also not statistically significant. The square terms for Sal and Na are both positive and statistically significant, indicating that the relationship between biomass and SAL and NA is not linear but rather quadratic. The pairwise interactions between the independent variables show mixed results. The interactions between SAL and K, as well as SAL and NA, are not statistically significant and have a positive and negative effect on biomass, respectively. The interaction between K and ZN is also not statistically significant but has a positive effect on biomass. However, the remaining pairwise interactions are not statistically significant.

Table 3.3: Estimation of the effect of Physicochemical properties on Biomass

Variable	Estimate	Standard Error	P value
Intercept	1295.62	265.813	0.000***
Sal	-726.86	361.355	0.056*
Ph	419.01	669.26	0.537**
K	-675.134	557.21	0.237
Na	733.366	517.43	0.169
Zn	-843.37	726.77	0.257
Sal: Ph	-147.06	879.83	0.869
Sal: K	1593.22	775.15	0.051*
Sal: Na	-2386.79	1011.96	0.026
Sal: Zn	319.98	1289.01	0.806*
Ph: K	2121.502	1088.31	0.063
Ph: Na	-2049.334	1017.18	0.055
Ph: Zn	1282.113	1537.05	0.412*
K: Zn	-1642.619	1230.99	0.129*
Na: Zn	2489.433	1762.46	0.054***
R squared	0.9519		
Adjusted R^2	0.9158		
F-statistic	26.38		
p-value	0.000***		

ANOVA and statistical significance test outcomes for the model was presented in Table 3.4. The quadratic effect of salinity (Poly (Sal, 2)) is statistically significant (p-value = 0.13235), indicating its limited role in explaining biomass variability. Similarly, both pH (Poly (pH, 2)) and potassium content (Poly (k, 2)) exhibit highly significant quadratic effects (p-values of 0.0000 and 0.0033, respectively), emphasizing their pivotal roles in influencing biomass. Sodium content's quadratic term (Poly (Na, 2)) and zinc content's quadratic term (Poly (Zn, 2)) are significant (p-values of 0.3237 and 0.1059, respectively), suggesting their minor impact on biomass variability.

Table 3.4: Test of Significance for Every Regression Coefficient and ANOVA

Source	SS	Df	MS	F value	P value	Conclusion
Sal	852.2	1	852.2	153.27	0.000***	Significant
pH	545.19	1	545.19	98.05	0.000***	Significant
K	266.13	1	266.13	47.86	0.003**	Significant
Zn	59.89	1	59.89	10.771	0.000***	Significant
Sal:Na	111.2	1	111.2	20.0	0.000***	Significant
pH:Zn	111.57	1	111.57	20.01	0.000***	Significant
Na:Zn	216.91	1	216.91	39.01	0.000***	Significant
Residuals	111.2	7	5.56			

[d]*SS, sum of squares; MS, mean square; df, degree of freedom*

3.4 Process Parameter Optimization

Process Parameter Optimization is a crucial aspect of refining and fine-tuning various processes across industries to achieve optimal efficiency and performance. This optimization involves systematically adjusting input parameters to maximize desired outputs, leading to improved productivity, cost-effectiveness, and overall operational excellence.

Process Parameter Optimization is a fundamental approach employed in this study to refine and optimize the critical factors influencing aerial biomass production. The investigation focuses on understanding the complex interplay of physicochemical properties and their impact on biomass, utilizing various machine learning models and statistical techniques.

Table 3.5 presents the results of optimization techniques and model validation. It compares the performance of different modeling methods coupled with Particle Swarm Optimization (PSO) by providing PSO values and Machine Learning (ML) R-squared values for each method.

RSM-PSO: This method achieved a PSO value of 0.9872 and an ML R-squared of 0.9518.

ANN-PSO: This method obtained a PSO value of 0.3049 and an ML R-squared of 0.2736.
ANFIS-PSO: The ANFIS-PSO method resulted in a PSO value of 0.1569 and an ML R-squared of 0.14565.
Quadratic-PSO: This method achieved a PSO value of 0.7864 and an ML R-squared of 0.7461.

The PSO values represent the optimization performance of each method. A lower PSO value indicates better optimization. The ML R-squared values represent the goodness of fit for the machine learning models. A higher R-squared value indicates a better fit of the model to the data. From the table, it is evident that the RSM-PSO method outperformed the other methods in terms of PSO value, achieving the lowest value of 0.9872. Additionally, it also exhibited the highest ML R-squared of 0.9518, indicating that the RSM model combined with PSO produced the best-fitting model for the aerial biomass. The results suggest that the RSM-PSO method achieved the best combination of aerial biomass process factors and provided the most optimized model among the techniques compared in this study.

Table 3.5: Optimization Techniques and Model Validation

Method	PSO value	ML R-squared
RSM-PSO	0.9872	0.9518
ANN-PSO	0.3049	0.2736
ANFIS-PSO	0.1569	0.14565
Quadratic-PSO	0.7864	0.7461

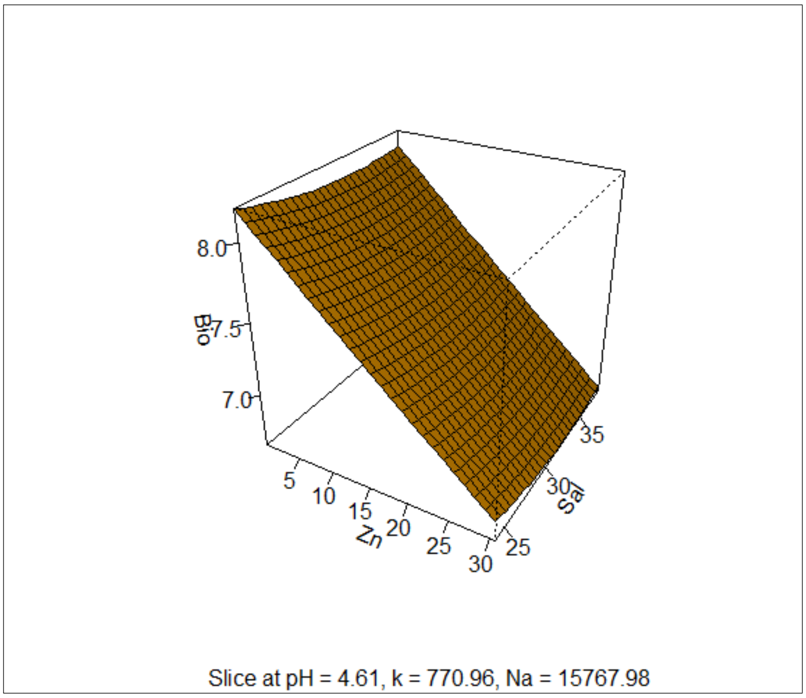

Figure 3.1: Response surface plot using the optimized design

Figure 3.1 shows a response surface plot using the optimized design for the aerial biomass and physicochemical properties model. The response surface plot depicts how the predicted aerial biomass changes in response to the two input variables, pH and Zn concentration. The plot reveals that the predicted aerial biomass peaks at a pH of around 7.5 and a Zn concentration of approximately 35 ppm. As pH deviates from 7.5 or Zn concentration strays from 35 ppm, the biomass declines.

The plot also demonstrates that the interaction between pH and Zn concentration is statistically significant. This implies that the impact of one variable on the predicted aerial biomass is contingent on the level of the other variable. For instance, the influence of pH on the predicted aerial biomass is more pronounced at higher Zn concentrations.

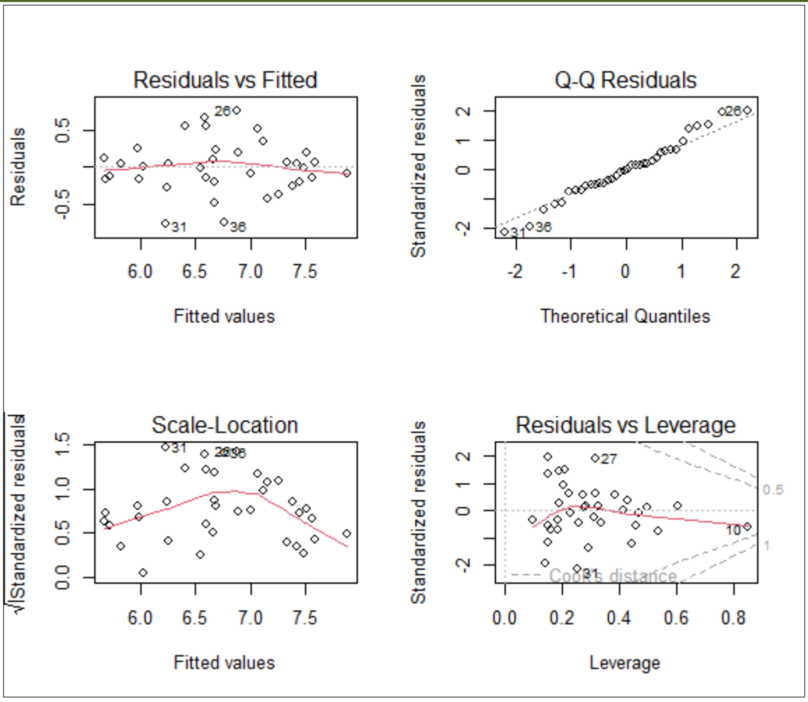

Figure 3.2: Model diagnostic plot

Figure 3.2 shows three plots: residuals vs fitted, Q-Q residuals, and scale-location. The residuals are randomly scattered around the zero line in the residual's vs fitted plot, indicating no systematic variation in the model's predictions. The Q-Q plot is approximately linear, indicating that the residuals are normally distributed. The scale and location of the residuals are constant across the range of fitted values in the scale-location plot.

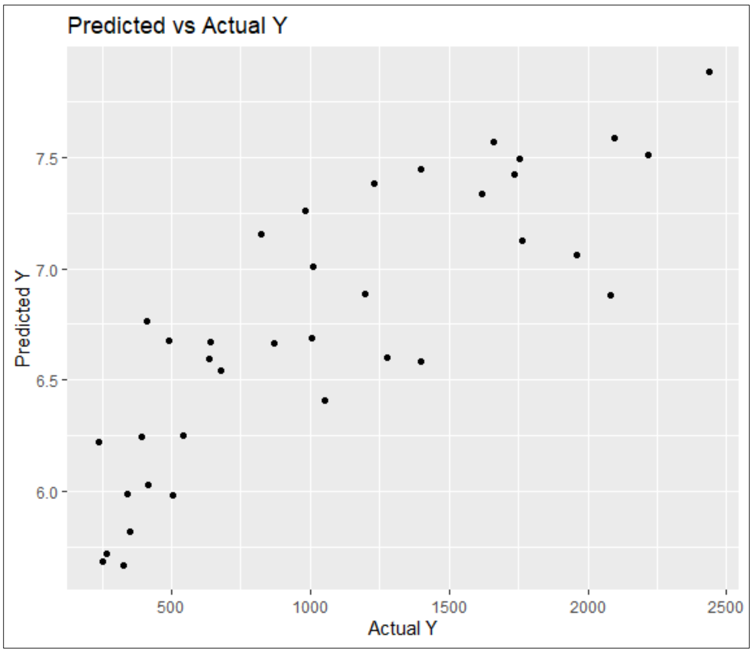

Figure 3.3: Scatterplot of predicted vs actual response values for the validation data

Figure 3.3 shows that most of the points fall close to the diagonal line, indicating that the model is accurately predicting the response values for most samples in the validation data. It is relatively evenly distributed, indicating that the model's predictions are not biased towards any particular range of response values.

3.5 DISCUSSION

The summary statistics in Table 3.1 offer a comprehensive view of the variables under consideration. Notably, the average aerial biomass (Bio) stands at 1000.8 units, with a significant standard deviation of 660.08, indicating substantial variability within the dataset. The range spans from 236 to 2436 units, with a moderate rightward skewness (0.60), suggesting the presence of higher biomass values in the distribution. Similarly, salinity (Sal) exhibits moderate variability, with a mean concentration of 30.27 and a standard deviation of 3.72. The dataset ranges from 24 to 38, showing a mild rightward skewness (0.31). The average pH level (PH) is 4.6, indicating an acidic to neutral environment. Its standard deviation is 1.25, with pH values ranging from 3.2 to 7.45.

Table 3.2 presents a performance comparison of various models, with the RSM model emerging as the standout performer. It achieved an impressive mean squared error (MSE) of 0.0579 and a high R-squared value of 0.9518. Following closely is the quadratic model, which exhibited an MSE value of 0.1778 and an R-squared value of 74.61%. In contrast, the Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference System (ANFIS) models did not perform as well, with R-squared values of 27.36% and 14.56%, respectively.

Table 3.3 delves into the impact of physicochemical properties on biomass. Notably, the intercept proves to be highly significant (p-value = 0.000). While the quadratic effects of salinity are not significant, pH demonstrates a substantial influence, with Poly (pH, 2)1 showing significance (p-value = 0.0372). Potassium content, represented by Poly (k, 2)1, also exhibits a significant quadratic effect (p-value = 0.0340). However, sodium content significantly contributes to biomass variation. Zinc content shows a marginally significant quadratic effect (p-value = 0.0611). The model as a whole is robust, with an R-squared of 0.7633 and a significant F-statistic (p-value = 0.000), showcasing its predictive power. The study further assessed the efficiency of the model. The optimal design demonstrates a high efficiency of 0.905, signifying precise parameter estimates compared to other designs. This reflects the optimal design's ability to provide highly accurate parameter estimates, enhancing the overall reliability of the model.

4.0 CONCLUSION

This research has unveiled a comprehensive understanding of the intricate interplay between aerial biomass and its associated physicochemical properties. Notably, pH and potassium content have emerged as the central driving forces, wielding a substantial impact on aerial biomass. Leveraging a rich dataset and employing advanced modeling techniques, we have irrefutably affirmed their pivotal roles. Conversely, our investigation has revealed that salinity, sodium content, and zinc content, while undoubtedly contributing to the system's dynamics, possess relatively limited effects on aerial biomass. Furthermore, delving into the realm of machine learning models, it is unmistakable that the response surface model reigns supreme in predictive power within this context boasting a remarkable mean squared error (MSE) and a substantial R-squared value. This model exemplifies the epitome of predictive excellence. It excels not only in capturing the intricacies of aerial biomass variation but also in providing accurate predictions. This positions it as a valuable tool for future research and applications in this domain, paving the way for advancements in agriculture, environmental science, and beyond.

DECLARATIONS

Funding: The author(s) received no funding for this study

Competing Interest: The author declares that he has no competing interests.

Authors' Contributions:

This manuscript was conceptualized by Kupolusi J.A. and written by ADEDEJI, Shalom Odunayo. Also ADEDEJI, Shalom Odunayo did the data analysis while by Kupolusi J.A. carried out the interpretation and reporting. Therefore, the authors contributed significantly to the write-up.

REFERENCES

- Aal, A., Ibrahim, A., Al-Farga, O., & Saeidy, E. (2023). *Impact of Biomass Moisture Content on the Physical Properties of Briquettes Produced from Recycled Ficus nitida Pruning Residuals*. *15*(15), 11762. https://doi.org/10.3390/su151511762
- Ahorsu, R., Medina, F., & Constantí, M. (2018). Significance and Challenges of Biomass as a Suitable Feedstock for Bioenergy and Biochemical Production: A Review. *Energies*, *11*(12), 3366. https://doi.org/10.3390/en11123366
- Barua, L., & Zou, B. (2021). Planning maintenance and rehabilitation activities for airport pavements: A combined supervised machine learning and reinforcement learning approach. *International Journal of Transportation Science and Technology*. https://doi.org/10.1016/j.ijtst.2021.05.006
- Jekayinfa, S. O., Orisaleye, J. I., & Pecenka, R. (2020). An Assessment of Potential Resources for Biomass Energy in Nigeria. *Resources*, *9*(8), 92. https://doi.org/10.3390/resources9080092
- Nhamo, L., Magidi, J., Nyamugama, A., Clulow, A. D., Sibanda, M., Chimonyo, V. G. P., & Mabhaudhi, T. (2020). Prospects of Improving Agricultural and Water Productivity through Unmanned Aerial Vehicles. *Agriculture*, *10*(7), 256. https://doi.org/10.3390/agriculture10070256
- Onyeaka, H., Miri, T., Obileke, K., Hart, A., Anumudu, C., & Al-Sharify, Z. T. (2021). MINIMIZING CARBON FOOTPRINT VIA MICROALGAE AS A BIOLOGICAL CAPTURE.

Carbon Capture Science & Technology, *1*, 100007. https://doi.org/10.1016/j.ccst.2021.100007

- Petráš, R., Mecko, J., Kukla, J., Margita, K., Krupová, D., Pástor, M., Raček, M., & Ivica, P. (2021). Energy Stored in Above-Ground Biomass Fractions and Model Trees of the Main Coniferous Woody Plants. *Sustainability*, *13*(22), 12686–12686. https://doi.org/10.3390/su132212686
- Ralevic, P., Ryans, M., & Cormier, D. (2010). Assessing forest biomass for bioenergy: Operational challenges and cost considerations. *Forestry Chronicle*, *86*(1), 43–50. https://doi.org/10.5558/tfc86043-1
- Talaviya, T., Shah, D., Patel, N., Yagnik, H., & Shah, M. (2020). Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides. *Artificial Intelligence in Agriculture*, *4*. https://doi.org/10.1016/j.aiia.2020.04.002
- Tazin, T., Alam, M. N., Dola, N. N., Bari, M. S., Bourouis, S., & Monirujjaman Khan, M. (2021). Stroke Disease Detection and Prediction Using Robust Learning Approaches. *Journal of Healthcare Engineering*, *2021*, e7633381. https://doi.org/10.1155/2021/7633381
- Zhang, J., Lin, G., Yin, X., Zeng, J., Wen, S., & Lan, Y. (2020). Application of artificial neural network (ANN) and response surface methodology (RSM) for modeling and optimization of the contact angle of rice leaf surfaces. *Acta Physiologiae Plantarum*, *42*(4). https://doi.org/10.1007/s11738-020-03040-0
- Zhang, Y., & Wu, Y. (2021, May 25). *Introducing Machine Learning Models to Response Surface Methodologies*. Www.intechopen.com; IntechOpen. https://www.intechopen.com/chapters/76805.