Exploring nonlinear models for predicting Diameter-Height and Diameter-Volume relationships in *Shorea robusta* trees: A study in the forests of Putalaibazar Municipality, Syangja

Sagar Budhathoki¹, Sanchita Budhathoki², Lochana Adhikari³, Sahayog Chhetri¹, Padam Prakash Jaishi⁴

**ABSTRACT**

This study aims to determine the most suitable models among a set of five candidate models to describe the relationships between diameter height and diameter volume for individual *Shorea robusta* trees within the forests of Putalaibazar Municipality, Syangja. The methodology involved measuring the diameter at breast height and total height of 137 individual trees and calculating their respective volumes. The normality of these variables was assessed by a non-parametric Kolmogorov-Smirnov test (p ≤ 0.05), which revealed non-normality and only five nonlinear models were employed to fit the height-diameter and volume-diameter relationship by a transformation of variables. The study estimated model parameters, including intercepts and regression coefficients, and assessed model performance using fit statistics such as the adjusted coefficient of determination (adj. R²) and root mean square error (RMSE). Statistical significance of parameters was determined through parametric t-tests for regression parameters, with all parameters found to be statistically significant (p ≤ 0.05). The selection of the best-fitting model was based on models exhibiting the highest adjusted coefficient of determination, lowest root mean square error, and lowest Akaike Information Criterion (AIC) value. Visual assessments, including histogram analyses, normal probability plot curves, and scattered plot diagrams, were also employed. Among the models tested, the Wykoff model (H(Height, m)=Bh(Breast Height, m)+exp(3.19+(-.9.203)/(D+1)), where D represents diameter at breast height in cm, demonstrated superior performance for characterizing the Diameter-Height relationship. For the Diameter-Total Volume relationship, the model V = (-0.049) + 0.001 * D² proved to be optimal. These selected models are recommended for predicting the height and volume of individual *Shorea robusta* trees. It is essential to note that these models are explicitly site-specific and should be applied exclusively to sites, sizes, and stand conditions congruent with those examined in this study.

**Keywords:** Model, best-fit, RMSE, adjusted R², etc.
1. INTRODUCTION

Background
Tree diameter distribution is one of the most important aspects to be considered by forest managers when making decisions about the management of forest stands as it provides a wide range of information, from timber assortments to carbon stock and even forest biodiversity. Two key measurements in forest inventories are the height of trees (H) and their diameter at breast height (D). Measurement of these variables is important in calculating things like how much space the trees take up, how much they weigh, how much carbon they store, and how likely they are to survive (Curtis, 1967). To effectively manage forests at the local, regional, or national level, it’s crucial to have a good understanding of the size and composition of the forest (Rahman et al., 2022). With this knowledge, forest managers can develop strategies that ensure the forest ecosystem continues to thrive.

In Nepal, the *Shorea robusta* tree is particularly valuable. It’s used for construction and making furniture, and it’s the primary source of firewood in the Terai region. *Shorea robusta* leaves are also important as food for animals and for making disposable plates (Jackson, 1994). The allometric equation is a valuable tool for establishing a connection between easily measurable morphometric variables, such as tree diameter (D), and the overall height and volume of a tree. Traditionally, models describing the relationship between tree height (H) and diameter (D) and between tree volume and diameter have been developed and applied primarily in pure, even-aged forest stands or plantations. In these models, diameter (D) serves as a predictor variable for tree height (H) in the H-D model (Huang et al., 2000). More recent studies have expanded the scope by considering additional stand attributes, including site quality, stand age, stand density, and dominant height, in mixed-effect H-D relationship models (Castaño-Santamaría et al., 2013).

Mixed-effect H-D relationship models incorporate both population-averaged parameters (fixed parameters common to the population) and subject-specific effects as random effects. This approach enhances accuracy compared to nonlinear models that rely on minimizing sums of squares. The distribution of tree diameters holds exceptional importance for forest managers when making decisions regarding forest stand management. It furnishes a wealth of information, ranging from the types of timber available to the estimation of carbon storage and the preservation of forest biodiversity (Pradip-Saud et al., 2016). Therefore, this research aims to develop best-fitted height-diameter and diameter volume models for the *Shorea robusta* forest of Syanga district of western Nepal. This developed model may be recommended to forest managers for predicting total heights and volume for *S. robusta* trees in western Nepal and reference for all Nepal. It is expected that the proposed model will be a useful tool for forest managers, forest users, and researchers.

Limitations of the study
The relationship between a tree’s diameter, height, and volume is subject to variation depending on various environmental factors, site quality, stand density, stand age, competition, and silvicultural treatments applied, among others (Forrester, 2017). Numerous studies have indicated that in dense forests, trees tend to grow taller compared to less dense forests, assuming other factors remain constant. Conversely, trees in dense forest environments tend to have smaller diameters, primarily due to heightened competition (López-Sanchez et al., 2003; Calama and Montero, 2004). Additionally, it’s worth noting that the limited duration of data collection resulted in the analysis of data from a relatively small sample of 137 trees.

2. METHODOLOGY

Study Site
The research takes place in the Shorea robusta forest of Putalibazar municipality, Syanja, Nepal. *Shorea robusta* (Sal), *Syzygium cumini* (Jamun), *Adina cardifolia* (Karma), *Lagerstomia parviflora* (Botdhayero), *Terminalia belirica* (Barro), *Castanopsis indica* (Katus), and others are among the principal tree species with slopes ranging from 0 to 48 degrees (Figure 1). The research area is located between 83.8470 and 83.960 and 28.070 and 28.1190.
Sampling Design and Data Collection
Forest management plans were used to determine the size variation of the individual species (S. robusta) within the research region. To account for a wider range of altitudes, aspects, slopes, stand origins, stand densities, stand age and size classes, and stand treatments, the selection of stands was subjective rather than using a complex sampling technique in which the selected individuals would have adequately represented the entire population (Adinugroho and Sidiyasa, 2006; Diéguez-Aranda et al., 2006; Castedo-Dorado et al., 2006). Wolf trees, malformed, top broken, inhibited, leaning, or unhealthy trees were not measured (Sharma, 2009). One hundred and thirty-seven (137) trees were measured. The diameter at breast height (DBH) and the trees’ total height were measured using diameter tape and vertex-IV, respectively.

Data Analysis
Diameter-Height Relationship
The normality of these variables was assessed using a non-parametric Kolmogorov-Smirnov test (p ≤ 0.05), which revealed non-normality. Only five nonlinear models (Table 1) were employed to fit the height-diameter relationship by transformation of variables. All of these models have a small number of parameters and are theoretically sound. Thus, they’re frequently employed to describe different tree and stand characteristics.

<table>
<thead>
<tr>
<th>Model</th>
<th>Equation (Linear and Nonlinear)</th>
<th>Model Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>H = Bh + a + bD + Ei</td>
<td>General linear</td>
</tr>
<tr>
<td>M2</td>
<td>H = Bh + exp(a/(b+D)+1) + Ei</td>
<td>Wykoff</td>
</tr>
<tr>
<td>M3</td>
<td>H = Bh + aDb + Ei</td>
<td>Lundqvist/Korf</td>
</tr>
<tr>
<td>M4</td>
<td>H = Bh + a exp(-b/d) + Ei</td>
<td>Ratkowsky</td>
</tr>
<tr>
<td>M5</td>
<td>H = Bh + D2/(b+D)2 + Ei</td>
<td>Naslund</td>
</tr>
</tbody>
</table>
Where, H= Total height(m); D= DBH= Diameter at breast height(cm), Bh= Breast height; a,b= regression coefficients(parameters); exp= Exponential; Ei= Unexplained error (random) and distributed residuals (error).

**Diameter- Total Volume Relationship**

The normality of these variables was assessed by a non-parametric Kolmogorov-Smirnov test (p ≤ 0.05), which revealed non-normality and only five nonlinear models (Table 1) were employed to fit the volume-diameter relationship by transformation of variables. All of these models have a small number of parameters and are mathematically sound, so they’re commonly used to model various tree and stand characteristics.

The following different models were tested using SPSS,

\[ V = a + b \times D \] ...............................(M1)
\[ \ln V = a + b \times \ln D \]...........................(M2)
\[ V = a + b \times \ln D \]............................(M3)
\[ \ln V = a + b \times D \]............................(M4)
\[ V = a + b \times D^2 \] ...............................(M5)

Parameter estimation and model evaluation: In this study, the two most commonly used modeling approaches were also used. The first step is to fit the candidate models, and the second step is to evaluate the models that have been fitted. In the first step, regression analysis was used to fit candidate models M1 to M5.

The second step, i.e., the evaluation of the fitted models, was carried out using the following criteria:

- **Adjusted coefficient of determination (R² adj):** With adjustments to the number of parameters, p, and the number of non-missing observations, n, it indicates the fraction of total variance explained by the model. It is estimated as follows:

\[ R^2_{adj} = 1 - \left( 1 - R^2 \right) \frac{(n-1)}{(n / p)} \]

- **Significance of the parameter values:** Parameter estimates should be significantly different from zero (p ≤ 0.05).

- **Homogeneity of the residuals:** Plotting of the residuals from the model overpredicted values or independent variables should show a random, constant variance pattern around a residual value of zero (Clutter et al., 1983).

- **Distribution of residuals:** Histograms of residuals were plotted to display the distribution (normal or abnormal) patterns of the residuals.

- **Root Mean Squared Error (RMSE):** It determines the accuracy of model predictions, and it is considered one of the most important model evaluation criteria. RMSE was calculated using the following formula:

\[ \text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n / p}} \]

Where Yi and Ŷi are the observed and predicted values, respectively; n is the total number of observations used to fit the model; and p is the number of parameters.

**3. RESULTS AND DISCUSSION**

The percentage occupancy of the sampled number of species in every 10 cm DBH class interval in ascending order was found to be virtually in contrast order of 23, 49, 43, 16, 4, and 2, respectively, as shown in (Table 2). In every average 10 cm DBH starting from 5 -15 cm to 55 - 65 cm intervals, the average height performance of this species is in the ratio of 1: 1.36: 1.6: 1.62 :1.72: 1.78: 1.62 respectively, which is not linear means that it is the indication of biological logic for the change of height with respect to DBH. Height growth rate (difference in height in relation to DBH) grew up to a specific DBH limit in the early stages, but it decreased with growing DBH in the later stages.
Because a tree needs more strength to withstand external pressures such as wind, diameter development should be faster than height growth when the tree grows to larger and taller sizes, the thickening of its bole should be faster than height growth (Khanna and Chaturvedi, 1994; Cato et al., 2006; Sharma, 2009). This species’ average DBH and height ratio was discovered to be 1.54:1, and its average DBH and volume ratio was determined to be 46.38:1. The standard errors of DBH and height are 6.22 and 4.77, respectively, while the standard errors of DBH and total volume are percentages of the estimates of respective variables for this species (Kalle, 2001). We found that the average DBH, height, and volume of all species in that forest were 25.42 cm, 16.411 m, and 0.54 m$^3$, respectively, with standard errors of DBH, height, and volume of 0.93, 0.32, and 0.044

Diameter- Height Relationship

The following table shows the intercept and regression coefficient parameters with estimated values and fit statistics with adjusted coefficient of determination (adj. R$^2$) and root mean square error (RMSE) values for the candidate models M1 and M5.

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter with estimates</th>
<th>Fit Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>M1</td>
<td>10.441</td>
<td>0.235</td>
</tr>
<tr>
<td>M2</td>
<td>3.19</td>
<td>-9.203</td>
</tr>
<tr>
<td>M3</td>
<td>4.075</td>
<td>1.544</td>
</tr>
<tr>
<td>M4</td>
<td>0.9818</td>
<td>0.6245</td>
</tr>
<tr>
<td>M5</td>
<td>23.594</td>
<td>-8.058</td>
</tr>
</tbody>
</table>

The estimated values of all parameters were found to be statistically significant (p ≤ 0.05) using the parametric t-test for regression parameters, as shown in (Table 3). Except for model M4, four of the models had a root mean square error (RMSE) of less than two. As a result, M4 was eliminated from further study due to poor fit statistics, particularly RMSE. Models M1 and M3 were eliminated because their adj. R$^2$ and RMSE were lower than those of M2 and M5. The best-fit models for the height-diameter connection were M2 and M5, which had higher adj. R$^2$, lowest RMSE, and lower AIC.

Choosing the best model based solely on the value of adj. R$^2$ and RMSE is not a wise decision. As a result, a graphical analysis of residuals was also performed. A residual, which may be thought of as the difference between the data and the fit, is a measure of the variability not described by the regression model that is also supported by AIC. The residuals are the errors’ realized or observed values. As a result, any deviations from the errors’ underlying assumptions should show up in the residuals. The analysis of residuals is a useful tool for identifying many types of model flaws (Jayaraman, 2000). The histogram was used to determine if the residual distribution was nearly normal or aberrant, as well as which model had the best fit.
Figure 2 Histogram for residuals for M2 and M5.

Figure 2 shows that the residuals of both models M2 and M5 were determined to be within the usual normal Z value of 4 (99% confidence range) and that the residual histogram was roughly bell-shaped with symmetrical (normal) distributions. As a result, there was no significant heteroscedasticity issue with the promising model (Sharma, 2009). However, the model M2 revealed a more normal distribution of residuals than the model M5, indicating that the normal curve of M2 covers more area of the histogram rectangles. Model M2’s normal probability plot curves revealed a cluster of residuals pointing towards the equal distribution line, whereas M5’s did not (Figure 3). It is typical to evaluate the residuals to see if the data meets the assumptions required for the regression analysis.

Figure 3 Normal P-P plot of standardized residuals of M2 and M5.

The scatter plot resembles Figure 4, which indicates that the residuals can be contained in a horizontal band on both sides of the zero-mean value of residuals, and then there are no obvious model defects. This type of pattern was observed in both the model M2 and M5. Therefore, no decision can be made on these scatter plots. The scatter plot of residuals versus the corresponding predicted (fitted) value is useful for detecting several common types of model inadequacies (Jayaraman, 2000). From the available information, we considered the model M2 the best-fitted model among the available ones.
Figure 4 Scatter plot of predicted vs. residual heights.

**Diameter- Total Volume Relationship**

The following table shows the intercept and regression coefficient parameters with estimated values and fit statistics with adjusted coefficient of determination (adj. R2) and root mean square error (RMSE) values for the candidate models M1 through M5.

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter with estimates</th>
<th>Fit Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>M1</td>
<td>-0.613</td>
<td>0.046</td>
</tr>
<tr>
<td>M2</td>
<td>5.057</td>
<td>2.435</td>
</tr>
<tr>
<td>M3</td>
<td>-2.424</td>
<td>0.948</td>
</tr>
<tr>
<td>M4</td>
<td>10.234</td>
<td>0.097</td>
</tr>
<tr>
<td>M5</td>
<td>-0.049</td>
<td>0.001</td>
</tr>
</tbody>
</table>

From the parametric t-test for regression parameters, Table 4 shows that the estimated values of all parameters were statistically significant (p ≤ 0.05). The root mean square error (RMSE) was more than 1 in two of the models (M2 and M4). As a result, M2 and M4 were left out of the final analysis due to poor fit statistics, particularly RMSE. Models M3 and M4 were eliminated because they had a lower adj. R2 and a greater RMSE than M1 and M5. M1 and M5 were the best-fit models for the total volume-diameter relationship, with higher adj. R square, lowest RMSE, and lower AIC. In addition, the RMSE and AIC of model M5 are significantly lower than those of model M1, and the adjusted R2 of model M5 is higher than that of model M1, indicating that model M5 is better suited to the Diameter-Volume relationship.

However, choosing the best model solely based on adj. R2 and RMSE is not a good decision. As a result, we ran a graphical analysis of the residuals. A residual, which can be thought of as the difference between the data and the fit, is a measure of the variability not explained by the regression model, and AIC also supports it. The realized or observed values of the errors are known as residuals. As a result, any deviations from the errors’ underlying assumptions should be reflected in the residuals. The use of residual analysis to investigate a variety of model flaws is a powerful tool (Jayaraman, 2000). The histogram was used to determine if the residual distribution was approximately normal or aberrant and which model best fit the data.
Figure 5 Histogram for residuals for M1 and M5.

Figure 5 illustrates that the residual of models M1 and M5 were found within standard normal Z value as 4 (99% confidence interval) and the residual histogram looked approximately bell-shaped with symmetrical (normal) distributions except for some outliers. Thus, this implied that there was no substantial heteroscedasticity problem with the promising model (Sharma, 2009). But comparatively, model M5 showed a more normal distribution of residuals than model M1, so the normal curve of M5 covers more area of the rectangles of the histogram. Model M5’s normal probability plot curves revealed a cluster of residuals pointing towards the equal distribution line, whereas M1’s did not (Figure 6). It is typical to evaluate the residuals to see if the data meets the assumptions required for the regression analysis.

Figure 6 Normal P-P plot of standardized residuals of M1 and M5.

If the residuals can be contained in a horizontal band on both sides of the zero-mean value of residuals, the scatter plot looks like Figure 7, indicating that there are no evident model flaws. Model M5 revealed this pattern. As a result, model M5 is rated the best fit among the available models. The scatter plot of residuals vs. the matching anticipated (fitted) value can be used to spot a variety of model flaws (Jayaraman, 2000). Based on the information supplied, we determined that the model M5 was the best-fitting model among the various models.
4. CONCLUSION
Among several modes tested in this study, \( H(\text{Height, m}) = B_h(\text{Breast Height, m}) + \exp(3.19 + (-9.203)/(D+1)) \) (Diameter at breast height (D) in cm), Wykoff model showed the best performance in terms of numerical significant parameters, fit statistics and graphical appearance for Diameter-Height relationship and \( V = (-0.049) + 0.001*D^2 \) for Diameter-Total Volume relationship. As a result, these models are suggested for predicting the height and volume of individual Shorea robusta trees. Because this model is expressly site-specific, it must be applied to sites, sizes, and stand conditions that are identical to those used in this study. Validation, verification, and re-calibration of the proposed model using additional data from a broader range of sites, sizes, and stand conditions of Shorea robusta trees is also suggested. More research is needed to build a strong association between diameter and height, as well as diameter and total volume, by incorporating a large number of data from a wide range of distributions.

Acknowledgment
After every completion, there are many faces and voices whose contribution made a difference towards the completion of work. This report is also not the result of the individual effort. Different persons contribute to the work from different levels. I owe my heartfelt gratitude to all of them. I would like to express my sincere gratitude to AFO Syangja, Ramesh Parajuli for his help, guidance, encouragement, and practical suggestions. My deepest debt of gratitude goes to Sabin Lamichhane for his hospitality and willingness to impart knowledge and experience during the field survey. I must thank my dear colleague, Ashish Pathak, who always supported me whenever needed and boosted me. I have to thank my batch as a whole for being so together and so helpful. My acknowledgment goes to my family for being inspired to undertake this research.

Author Contributions
Sagar Budhathoki: Conceptualization (lead); data curation (lead); formal analysis (lead); writing the original manuscript (lead); writing—review and editing (lead). Sanchita Budhathoki and Padam Prakash Jaishi: Data collection (equal); Formal analysis (equal). Lochana Adhikari: Data analysis, editing manuscript.

Informed consent
Not applicable.

Ethical approval
The ethical guidelines for plants & plant materials are followed in the study.

Conflicts of interests
The authors declare that there are no conflicts of interests.
Funding
The study has not received any external funding.

Data and materials availability
All data associated with this study are present in the paper.

REFERENCES AND NOTES
17. Rahman MM, Khan SA, Kabir ME. Status of timber productivity in exotic and indigenous tree plots in Sakhipur areas of Tangail district in Bangladesh. Discovery, 2022, 58(314), 125-133