Vol. 14(1): 179-192, January 2026

DOI: https://doi.org/10.23960/jsl.v14i1.1295

Jurnal Sylva Lestari

P-ISSN: 2339-0913
E-ISSN: 2549-5747

Journal homepage: https://sylvalestari.fp.unila.ac.id

*Full Length Research Article*

# Predicting Lumber Recovery of *Pinus patula* using Forest Inventory Variables: Model Development and Validation across Circular Sawmills

Gabriel Wilbald Mangi [1,2,*], Fortunatus Bulabo Makonda[2], Ernest William Mauya[2]

[1] Department of Ecosystems and Conservation, Sokoine University of Agriculture, Morogoro, Tanzania
[2] Department of Forest Engineering and Wood Sciences, Sokoine University of Agriculture, Morogoro, Tanzania
[*] Corresponding author. E-mail address: gabriel.mangi@sua.ac.tz

ARTICLE HISTORY:

*Received: 17 September 2025*
*Peer review completed: 2 December 2025*
*Received in revised form: 24 December 2025*
*Accepted: 30 January 2026*

KEYWORDS:

*Cross-validation*
*Regression models*
*Sawmill efficiency*
*Tree characteristics*
*Wood processing*



## ABSTRACT

The forestry sector plays a major role in Tanzania's timber economy, yet lumber recovery estimates in sawmills still rely on manual, log-based measurements that limit accuracy and operational planning. This study developed predictive models to estimate lumber recovery of *Pinus patula* using forest inventory variables: diameter at breast height (DBH), tree height, and taper. Data were obtained from 80 trees, yielding 254 logs and 2,364 boards, processed at the Laimet 120 and Slidetec Tommi Laine (STL) circular sawmills. Regression models (logarithmic, log-linear, polynomial, and power) were fitted and evaluated using Akaike information criterion, coefficient of determination, root mean square error, mean absolute error, coefficient significance, and K-fold cross-validation. Model performance showed that all equations explained more than 73% of the variation in lumber recovery, with polynomial models providing the highest accuracy, lowest error values, and most stable cross-validated estimates. Predictor importance differed by sawmill: DBH and height were most influential for Laimet 120, while taper improved predictions for STL due to greater variation in stem form. These results demonstrate that forest inventory data can be used to reliably estimate lumber recovery. The developed equations provide sawmills and forest managers with a practical tool for planning log allocation, enhancing efficiency, and minimizing processing waste.

## 1. Introduction

The forestry sector is vital to forest-rich countries such as Russia, Brazil, Canada, the United States, and China, as well as others with significant forest cover, including Australia, the Democratic Republic of the Congo, Indonesia, and Peru (Rascón-Solano et al. 2023; Sushko and Efimova 2025). Globally, forested land covers approximately 4.06 billion ha (Nesha et al. 2021). In Africa, the forested land is 645 million ha (Djenontin et al. 2021). In Tanzania, about 48.1 million ha of land are covered by forests and woodlands, representing roughly 7.5% of Africa's total forested area (Njana et al. 2021). These forests and woodlands serve as habitats for wildlife, support ecosystems, and provide water catchments, biodiversity, timber, and non-timber products.

Forests also generate significant government revenue through trade in forest products, particularly through processing industries. Globally, the forest processing industry contributes approximately USD 1.5 trillion annually to the economy (Alderman 2022). In Tanzania, the forest processing industry is dominated by timber processing, which accounts for approximately 4.26% of gross value added (Temu et al. 2024). To balance forest conservation with demand for forest products, plantation forests have emerged as a sustainable solution.

Tanzania has promoted plantation forestry since the 19th century to support the timber industry while conserving biodiversity and natural resources, and to enhance forest value, foreign exchange earnings, land productivity, environmental quality, and employment opportunities (Andrew 2022). The country has approximately 554,500 ha of plantation forests and small-scale woodlots dominated by species such as *P. patula*, *Eucalyptus* spp., *Cupressus lusitanica*, and *Tectona grandis* (Beleko et al. 2021). These plantations supply raw materials to key industries, including construction, furniture, pulp and paper, and packaging, and when well managed, they ensure a sustainable flow of timber to both local and international markets (Babune et al. 2021; Marnaek et al. 2024; Wong et al. 2019). Timber industries, particularly sawmills, convert logs into value-added products like lumber. In Tanzania, sawmills have been central to economic development since independence (Silver 2019), followed by pulp and paper, plywood, pole treatment plants, and furniture industries (Lolila et al. 2021). Sawmills process roundwood into planks and dimensional lumber, with lumber recovery representing the usable wood output after processing (Rawat et al. 2023).

Despite the long history of timber studies in Tanzania, little progress has been made in developing systems that use forest inventory variables, such as DBH, height, and taper, to quantify lumber recovery. Studies in Canada (Li et al. 2016) have shown that forest inventory data, such as DBH, height, and taper, can be effectively used to estimate lumber yield and improve wood resource utilization. However, similar approaches have not been applied in Tanzania. Current methods rely on log characteristics, such as length and diameter (Lolila et al. 2021; Mangi et al. 2025; Manina and Rayment 2025; Mauya et al. 2024), which limit the efficiency of sawmilling operations and the accuracy of decision-making tools.

Accurate estimation of lumber recovery is very important for understanding production efficiency and waste generation in sawmills. Without a simple and rapid system, lumber recovery is determined manually by measuring plank dimensions (Kolapo et al. 2020; Worku et al. 2022), which is labor-intensive and less precise. Model equations that use forest inventory data offer a promising approach to estimate lumber recovery efficiently. This study aimed to develop model equations to estimate lumber recovery from *P. patula* using forest inventory data. The specific objectives were: 1) To develop model equations for predicting lumber recovery using forest inventory variables, and 2) To evaluate and confirm the practical applicability of the developed model equations.

## 2. Materials and Methods

### *2.1. Study Area*

The study was conducted at Sokoine University of Agriculture Training Forest (SUATF) in Arusha Region (**Fig. 1**). SUATF lies between latitudes 3°15'–3°18' S and longitudes 36°41'–36°42' E, at 1,740 to 2,320 masl. It receives two rainy seasons per year: the long rains from March to May and the short rains from November to December. The mean annual temperature of the area ranges from 18°C to 23°C, during the morning and afternoon, respectively (Silayo et al. 2014). SUATF is planted with various tree species, dominated by *P. patula*, which is the major source of logs for sawmilling.

Two circular sawmills, Laimet 120 and Slidetec TL, were purposively selected for this study. Laimet 120 was 25 years old and used a steel circular saw (4.9 mm kerf, 100 cm blade diameter, log capacity 80 cm × 1.5–11.5 m, table speed 110 m/min), while Slidetec TL was 1 year old and operated with a carbide saw blade (3.1 mm kerf, 120 cm blade diameter, log capacity 65 cm × 1.5–11.5 m, table speed 110 m/min).

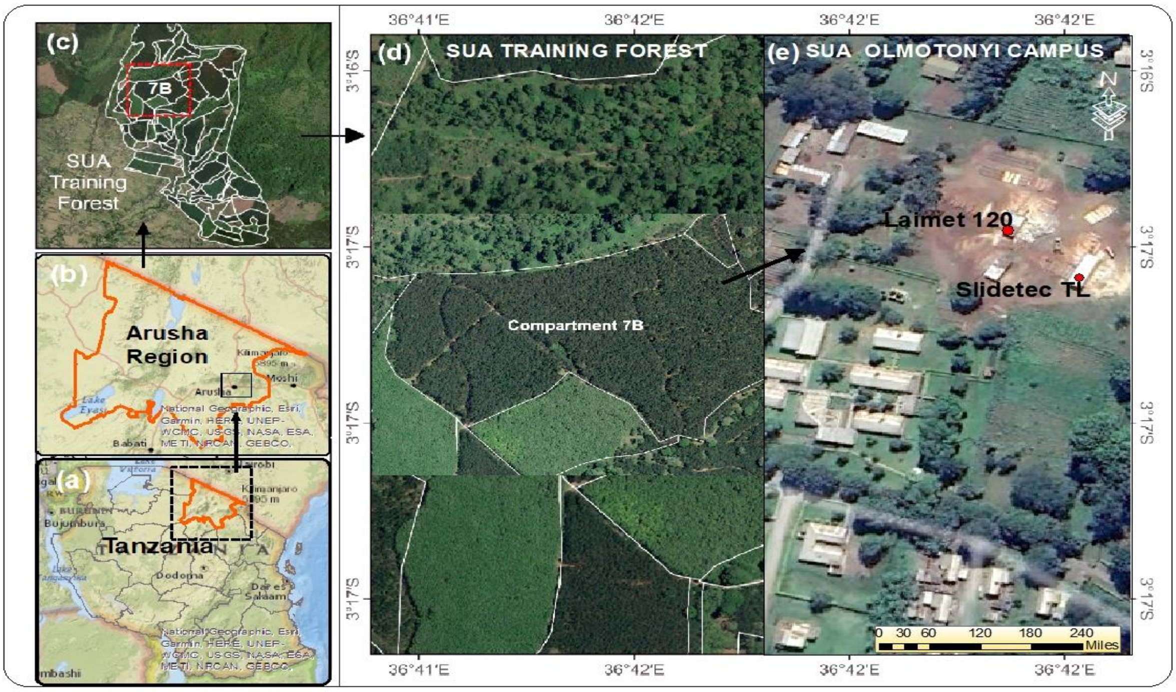


**Fig. 1.** Location map of the study area.

## *2.2. Sampling and Data Collection*

Eighty *P. patula* trees (DBH > 10 cm) were randomly selected from Compartment 7B (**Fig. 1**). Following the recommendation of Younis and Hassan (2019), who indicated that a sample size of 40 trees is sufficient for reliable estimation of stem and branch volumes and for testing model adequacy, 40 trees were allocated to each sawmill, giving a total sample of 80 trees for comparative analysis. Ten rectangular plots (30 m × 30 m) were randomly established, and all trees within each plot were numbered. From each plot, eight trees were then selected for measurement using a random number generator. DBH and total height were measured using a Haglof Mantax Blue 80 cm caliper and a Haglof Vertex 4 hypsometer, and trees were marked for identification. After felling, logs were bucked; diameters (bottom, mid, top) were measured with a caliper, and lengths were measured with a tape measure, from which taper was derived. In total, 254 logs (127 per sawmill) were processed. Log lengths ranged 5.3–6.6 m for Laimet 120 and 4.0–5.8 m for Slidetec TL, with corresponding processed volumes of 49.8 $m^3$ and 42.3 $m^3$, yielding 22.59 $m^3$ and 22.27 $m^3$ of lumber, respectively. Lumber from Laimet 120 measured 3.3–6.6 m in length, 1.4–5.3 cm in thickness, and 4–27.1 cm in width, while that from Slidetec TL ranged 2.57–5.8 m, 1.0–7.5 cm, and 3.6–30 cm. Final dimensions were verified using a digital caliper and tape measure. DBH, total height, and taper, standard inventory variables that are simple to measure, were used as predictors for model development.

### 2.3. Statistical Modeling

Data collected from tree and log measurements were entered into the appropriate form for statistical analysis. The processed data included log volume, taper and the volume of lumber recovered.

#### 2.3.1. Log volume

Log volume was computed using Newton's formula with the following Equation 1 (Manina and Rayment 2025).

$$L_v = \pi \frac{(D_1{}^2 + 4Dm^2 + D_2{}^2)}{240000} \times L \quad (1)$$

where $L_v$ is log volume in cubic metres (m³), $D_1$ is bottom diameter in centimetres, $D_m$ is mid diameter in centimetres, $D_2$ is top diameter in centimetres, and $L$ is log length in metres.

#### 2.3.2. Lumber volume

The lumber volume was computed for individual planks from each of the logs sawn using the following Equation 2 (Kolapo et al. 2020).

$$V_l = \sum n \; x \; l \; x \; w \quad (2)$$

where $V_l$ is lumber volume in cubic metres, $n$ is the number of lumber sawn, $l$ is the length of lumber in metres, $w$ is the width of lumber in metres, and $t$ is the thickness of lumber in metres.

#### 2.3.3. Taper

The taper for each log was computed as the ratio of the diameter difference and log length using the following Equation 3 (Worku et al. 2022).

$$Taper = \frac{D_1 - D_2}{L} \quad (3)$$

where $D_1$ is the bottom diameter in centimetres, $D_2$ is the top diameter in centimetres, and $L$ is the log length in metres.

### 2.4. Model Development

Model development involved selecting predictor variables, choosing model forms, estimating parameters, and validating performance. Previous studies have shown that polynomial, power, exponential, multiple linear regression, double-log, semi-logarithmic, and combined-variable models are effective for predicting lumber recovery using simple inventory variables (Eguakun and Nwanko 2016; Vibrans et al. 2015). In this study, DBH, tree height, and taper were used as key predictors, and logarithmic (log-log), log-linear, polynomial, and power functions were applied to develop predictive equations. The Equations 4–7 were of the following form.

#### 2.4.1. Logarithmic (log-log) model

$$log(R) = \beta_0 + \beta_1 \, log(X) + \epsilon \quad (4)$$

#### 2.4.2. Log-linear model

$$R = \beta_0 + \beta_1 X + \beta_2 \, log(X) + \epsilon \quad (5)$$

*2.4.3. Polynomial model*

$$R = \beta_0 + \beta_1 X + \beta_2 X_2 + \beta_3 X_2 + \cdots + \beta_n X_n + \epsilon \quad (6)$$

*2.4.4. Power models*

$$R = \beta_0 X^{\beta 1} + \epsilon \quad (7)$$

where $R$ is the lumber recovery (dependent variable), $X$, $X_1$, and $X_2$ are the independent variables such as DBH, tree height, and taper, $\beta_0$, $\beta_1$, $\beta_2$, and $\beta_3$ are the coefficients, n is the maximum degree of the polynomial terms, and $\epsilon$ is the error term.

*2.5 Performance Evaluation of the Models*

The models were evaluated to select the best for estimating lumber recovery. The following are the criteria used to evaluate the model using Equations 8–11.

*2.5.1. Akaike information criterion (AIC)*

AIC was computed to aid in model selection (Pham 2019) as:

$$AIC = 2k - 2\log(L) \quad (8)$$

where $k$ is the number of parameters in the model, and $L$ is the likelihood of the model given the data.

*2.5.2. Coefficient of determination ($R^2$)*

$R^2$ was computed to aid in model selection (Pham 2019) as:

$$R^2 = 1 - [\frac{RSS}{TSS}] \quad (9)$$

where $R^2$ is the coefficient of determination, $RSS$ is the residual sum of squares, and $TSS$ is the total sum of squares.

*2.5.3. Root mean square error (RMSE)*

RMSE was computed to aid in model selection (Hodson 2022) as:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y} - y_i)}{n}} \quad (10)$$

where $y$ is the observed lumber recovery, $\hat{y}_i$ is Predicted lumber recovery and $n$ is the number of observations.

*2.5.4. Mean absolute error (MAE)*

MAE was computed to aid in model selection (Hodson 2022) as:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} | yi - \hat{y}i | \quad (11)$$

where $n$ is the number of observations, $yi$ is the observed value (actual value) for the i-th observation, $\hat{y}i$ is the predicted value for the i-th observation, and $|yi - \hat{y}i|$ is the absolute value of the difference between the observed and predicted values.

### *2.6. Statistical Analysis*

The best model for estimating lumber recovery was selected based on the highest $R^2$, lowest AIC, RMSE, and MAE values, and statistically significant regression coefficients. Model performance was further validated using K-fold cross-validation. Parameters of the arithmetic models were estimated using the least squares method, while their significance and the agreement between observed and predicted lumber recovery were tested using the Mann–Whitney U test ($\alpha = 0.05$) for logs processed by the two sawmills. Both descriptive and inferential statistical analyses were conducted, and all analyses were performed in R (version 4.3.3).

## 3. Results and Discussion

### *3.1. Characteristics of Trees and Lumber Measured*

**Table 1** indicates clear differences in log characteristics between the two sawmills. Laimet 120 processed logs of greater diameter and height, whereas Slidetec TL handled smaller logs, thereby affecting resultant log volume, conversion yield, and output piece counts. Accordingly, the observed variation is primarily attributable to heterogeneity in the input material. However, machine-specific parameters, such as kerf width, degree of automation and mechanization, and feed system stability, also modulated recovery efficiency.

**Table 1.** Descriptive statistics of 80 sample trees collected from *P. patula*

| Machine type | Variable | Mean | Max. | Min. | SD | SE |
|---|---|---|---|---|---|---|
| Laimet 120 | DBH (cm) | 36.52 | 51.1 | 24.7 | 6.184 | 0.977 |
| | Tree height (m) | 29.185 | 35.5 | 22.2 | 2.325 | 0.367 |
| | Log length (m) | 5.563 | 5.8 | 5.3 | 0.1006 | 0.0159 |
| | Taper (cm/m) | 0.911 | 1.368 | 0.304 | 0.255 | 0.04 |
| | Log volume ($m^3$) | 1.237 | 2.912 | 0.45 | 0.61 | 0.096 |
| | Lumber volume ($m^3$) | 0.564 | 1.722 | 0.195 | 0.3 | 0.047 |
| | Lumber pieces | 28 | 66 | 8 | 12.027 | 1.901 |
| Slidetec TL | DBH (cm) | 33.802 | 49 | 21.5 | 6.619 | 1.046 |
| | Tree height (m) | 25.482 | 31.3 | 11.5 | 4.6 | 0.727 |
| | Log length (m) | 5.453 | 5.7 | 4.59 | 0.198 | 0.031 |
| | Taper (cm/m) | 1.013 | 1.684 | 0.54 | 0.245 | 0.038 |
| | Log volume ($m^3$) | 1.059 | 2.798 | 0.175 | 0.638 | 0.1 |
| | Lumber volume ($m^3$) | 0.556 | 1.516 | 0.055 | 0.361 | 0.057 |
| | Lumber pieces | 29.2 | 76 | 3 | 17.793 | 2.813 |

Notes: Max.= maximum, Min. = minimum, SD = standard deviation, and SE = standard error.

Taper values were higher in logs sawn by Slidetec TL, reflecting greater stem-form variation. Since a higher taper often reduces conversion efficiency, this suggests that recovery from Slidetec TL could be more sensitive to log form than Laimet 120. Despite these differences, mean lumber volumes from both sawmills were comparable, although variation was higher for Slidetec TL, as indicated by the wider range of lumber piece counts. This greater variability points to the influence of log size distribution and stem form on sawing outcomes.

Accurate estimation of lumber recovery is therefore important for assessing sawmill performance and minimizing waste. Traditional approaches that rely on manual measurement are often labor-intensive, inconsistent, and prone to error. Tree-level variables such as DBH, total height, and stem taper have commonly been used as predictors in lumber recovery modeling

(Kankare 2015). Evidence from previous modeling efforts suggests that regression models incorporating DBH alone or in combination with height can provide reasonable estimates of lumber recovery across different species and processing conditions (Pieterse 2022). In several applications, polynomial and similar regression models have been used to link tree dimensions with recovery outcomes, and their predictive accuracy tends to improve when additional inventory variables are included rather than relying on a single variable alone (De Lima et al. 2018).

These results indicate that basic tree measurements may offer useful explanatory power for lumber recovery, although model performance can vary depending on species characteristics, sawmill technology, and operational settings (Ngobi et al. 2024). These findings indicate that simple, measurable tree attributes can reliably explain differences in lumber recovery under varying sawmill and operational contexts, supporting their continued use in recovery modeling where detailed processing data are unavailable. Additionally, the results show that differences in tree shape and size strongly affect sawn wood output.

### *3.2. Lumber Volume Recovery Prediction Models*

The modeling results presented in **Table 2** indicate that all tested regression models, logarithmic, log-linear, polynomial, and power, performed well in predicting lumber recovery from inventory data, with $R^2$ values exceeding 0.73.

**Table 2.** Results of models fitted for both Laimet 120 and Slidetec TL sawmills

| Model form | AIC | $R^2$ | RMSE | MAE |
|---|---|---|---|---|
| **Laimet 120** | | | | |
| (1) $R = 7.655 \times 10^{-6} D^{2.5462} H^{0.5731} T^{-0.1768}$ | 12.043 | 0.767 | 0.143 | 0.097 |
| (2) $R = 0.0349 \exp(0.0727D)$ | 11.580 | 0.745 | 0.151 | 0.107 |
| (3) $R = 2.145 \exp(-00525D) \exp(0.0012D^2)$ | -28.266 | 0.732 | 0.153 | 0.107 |
| (4) $R = -0.1078 + 1.6527 \text{ x } 10^{-5} D^2 H$ | -37.081 | 0.774 | 0.141 | 0.103 |
| (5) $R = 645.23 \exp(-0.0489D - 0.0002D^2 + 0.5512H + 0.0101H^2)$ | -39.321 | 0.816 | 0.127 | 0.100 |
| (6) $R = 28.28 \exp(-0.07971D) \exp(-0.0005D^2) \exp(-0.1612H) \exp(0.0053DH) \exp(-0.1345T)$ | -36.859 | 0.814 | 0.128 | 0.096 |
| (7) $R = 1.039 \exp(0.0578D) \exp(0.0014D^2) \exp(0.0289H) \exp(0.3132T) \exp(0.0124DT)$ | -29.143 | 0.774 | 0.141 | 0.103 |
| (8) $R = 3.7806 \times 10^{-5} D^{2.6457}$ | 10.931 | 0.749 | 0.155 | 0.109 |
| **Slidetec TL** | | | | |
| (1) $R = 1.658 \times 10^{-6} D^{3.3491} H^{0.3829} T^{-0.6485}$ | 28.509 | 0.824 | 0.304 | 0.249 |
| (2) $R = 0.0187 \exp(0.0935D)$ | 44.952 | 0.706 | 0.393 | 0.294 |
| (3) $R = -0.8 + 0.03D + 0.000000011D^2$ | -17.96 | 0.76 | 0.17 | 0.13 |
| (4) $R = -0.1017 + 2.1273 \exp(-5D^2H)$ | -16.940 | 0.741 | 0.181 | 0.135 |
| (5) $R = 0.481 \exp(-0.0160D) \exp(-0.00041D^2) \exp(-0.0135H) \exp(-0.00013H^2)$ | -15.200 | 0.767 | 0.172 | 0.130 |
| (6) $R = 1.116 \exp(-0.015D) \exp(-0.000D^2) \exp(-0.031H) \exp(-0.288DH) \exp(-0.001T)$ | -17.853 | 0.792 | 0.162 | 0.120 |
| (7) $R = 0.382 \exp(-0.034D) \exp(0.001D^2) \exp(0.001H) \exp(0.449T) \exp(-0.023DT)$ | -18.055 | 0.793 | 0.162 | 0.121 |
| (8) $R = 4.2721 \times 10^{6} D^{3.2975}$ | 39.600 | 0.743 | 0.195 | 0.132 |

Notes: R = lumber volume recovery, D = diameter at breast height, H = tree height, and T = tree taper.

This demonstrates that different functional forms can adequately describe the relationship between tree inventory variables and lumber recovery. Similar findings were reported by Eguakun and Nwankwo (2016), who showed that logarithmic, power, combined-variable, and polynomial models achieved acceptable prediction accuracy using different independent variables. Although all models performed well, polynomial models provided the most accurate predictions, as reflected by their higher $R^2$, lower AIC, RMSE, and MAE values, indicating stronger explanatory power. Borz et al. (2021) also reported that timber efficiency models are commonly evaluated using $R^2$, and the strength of $R^2$ guides model selection.

Stepwise variable selection further revealed differences in the importance of predictors between the two sawmills. For Laimet 120, DBH and tree height were sufficient to explain lumber recovery, reflecting the generally larger and taller logs processed at this mill. In contrast, for Slidetec TL, the inclusion of taper improved model performance, indicating greater sensitivity of recovery to stem form in smaller, more tapered logs. This highlights that the influence of tree attributes on lumber recovery depends on log size, stem form, and sawmill processing conditions, supporting observations that sawmill technology plays a key role in determining timber recovery (Ngobi et al. 2024).

Model 3, which used DBH alone, performed well across both sawmills, demonstrating that DBH is a practical and reliable predictor of lumber recovery when inventory data are limited. This agrees with Magarik et al. (2020), who showed that accurately measured DBH can effectively estimate volume and other difficult-to-measure forest attributes. However, prediction accuracy improved when additional variables such as height and taper were included, particularly for Slidetec TL, which processed more variable stem forms. This trend is consistent with Li et al. (2016), who reported that model prediction accuracy generally increases with the number of independent variables. Finally, the strong model fit and evenly distributed residuals shown in **Fig. 2** and **Fig. 3** indicate homogeneity of variance and reliable predictions. This agrees with De Lima et al. (2018), who stated that small, evenly spread residuals around the regression line indicate constant variance and reliable predictions.

### *3.3. Model Validation*

The K-fold cross-validation results, an approach that minimizes overfitting and provides a robust estimate of predictive accuracy (Nti et al. 2021), showed that the models explained approximately 74.5% of the variation in lumber recovery, with low RMSE and MAE values and stable error estimates across folds. Consistent with these findings, the Mann–Whitney U tests (**Table 3**) revealed no significant differences between observed and predicted values for either sawmill, confirming the strong reliability of the models.

Among the tested model forms, polynomial models demonstrated the best overall performance, with higher goodness-of-fit and more evenly distributed residuals, indicating improved predictive accuracy and a more homogeneous variance (**Fig. 2** and **3**). These findings suggest that polynomial relationships are well-suited for describing the association between inventory variables and lumber recovery under the conditions examined in this study. This observation is consistent with De Lima et al. (2018), who noted that lower residual values and a regular distribution of data points around the regression line are key indicators of good model quality and reliable predictions. Additionally, Dos Santos et al. (2024) also noted that graphical analysis of residuals is an effective method for evaluating model precision.

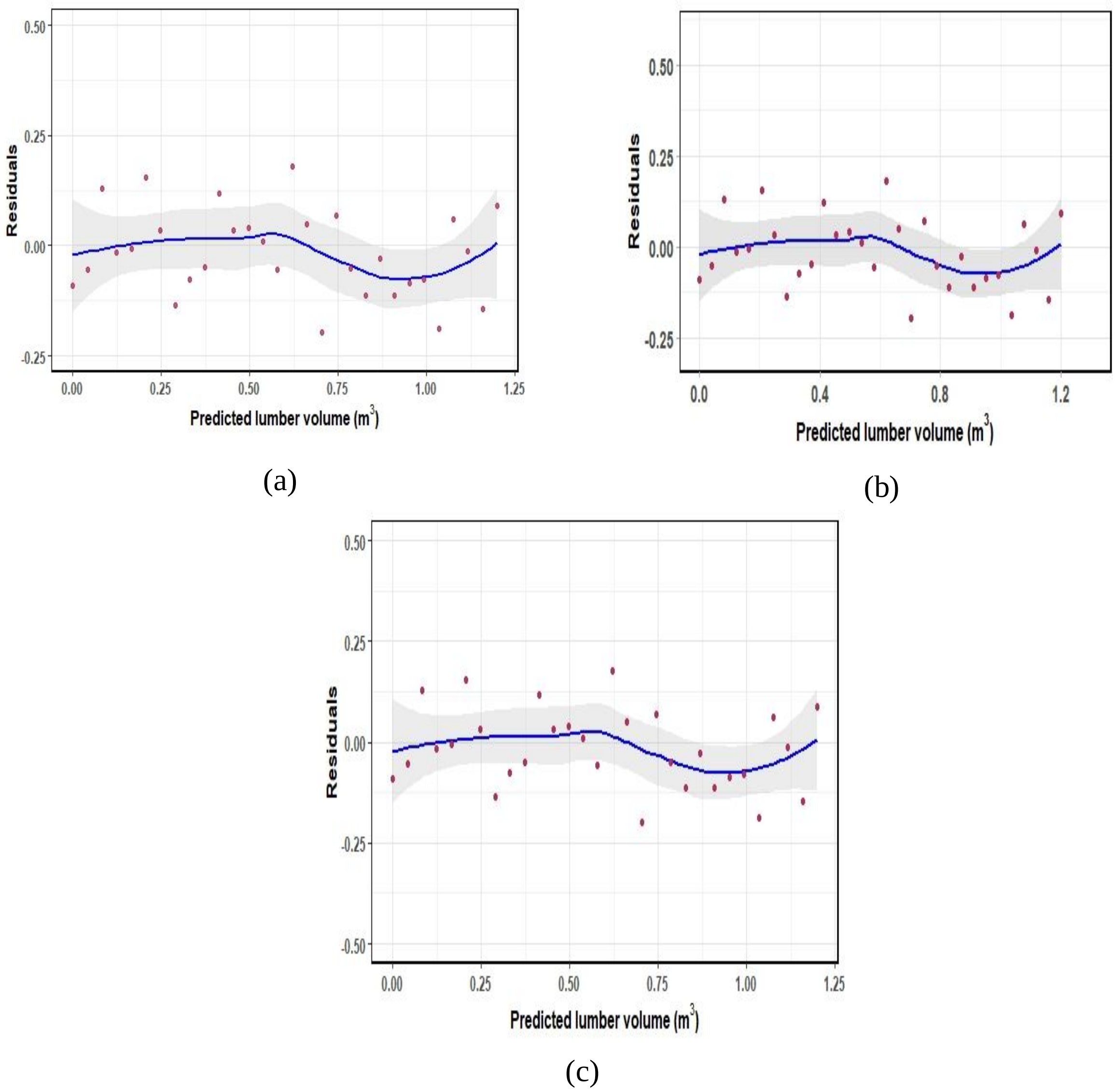


**Fig. 2.** Residuals for the Laimet 120 best model using (a) DBH, (b) DBH and height, and (c) DBH, height, and taper.

Stepwise selection revealed clear, sawmill-specific differences in predictor importance. For Laimet 120, including DBH and height increased the polynomial model's $R^2$ from 0.73 (DBH only) to 0.82, while the addition of taper reduced it slightly to 0.81. In contrast, for Slidetec TL, model performance improved from an $R^2$ of 0.76 (DBH only) to 0.79 when DBH, height, and taper were included, indicating that taper contributed more to prediction accuracy under these processing conditions. These results demonstrate that the effect of tree attributes on lumber recovery depends on sawmill characteristics and log form. Notably, Model 3, which uses DBH alone, showed consistent performance across both sawmills, highlighting its practical value where inventory data are limited. The findings demonstrate that polynomial models, when validated with cross-validation, offer reliable and practical tools for predicting lumber recovery. Consistent with De Lima et al. (2018), such models can enhance sawn timber estimation and support operational efficiency in different milling settings.

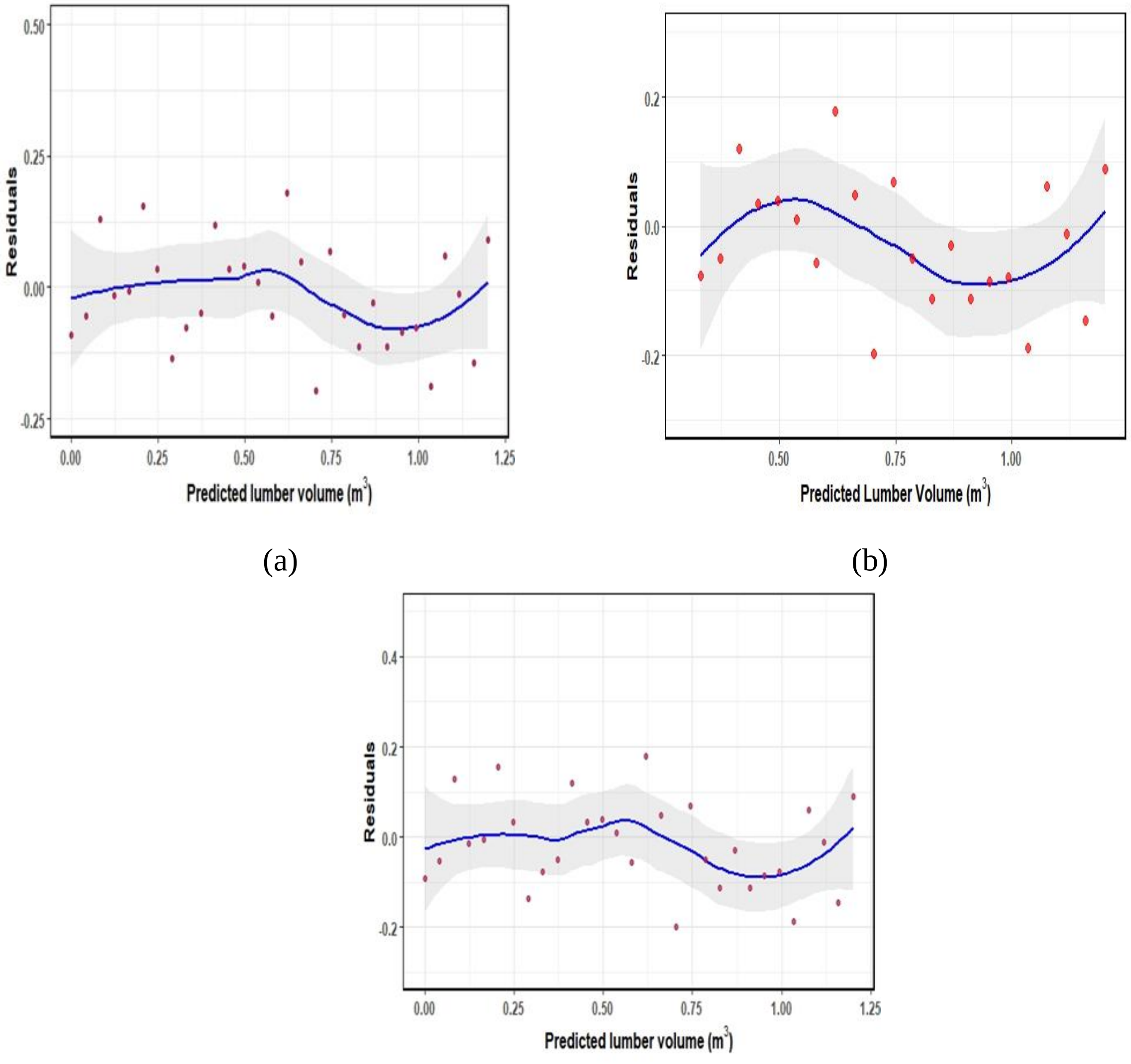


**Fig. 3.** Residuals for the Slidetec TL best model using (a) DBH, (b) DBH and height, and (c) DBH, height, and taper.

**Table 3.** Mann-Whitney U test summary for the fitness of the selected models

| Model No. | Mean ($m^3$) | Max. ($m^3$) | Min. ($m^3$) | SD | SE | P-value |
|---|---|---|---|---|---|---|
| **Laimet 120** | | | | | | |
| Model 3 | 0.5647 | 1.3554 | 0.2308 | 0.2575 | 0.0407 | 0.6755 |
| Model 4 | 0.5647 | 1.4242 | 0.1160 | 0.2647 | 0.0418 | 0.8821 |
| Model 5 | 0.5647 | 1.6084 | 0.1538 | 0.2718 | 0.0429 | 0.958 |
| Model 6 | 0.5647 | 1.6125 | 0.1577 | 0.2715 | 0.0429 | 0.837 |
| Model 7 | 0.5647 | 1.4785 | 0.0976 | 0.2643 | 0.0418 | 0.8211 |
| **Slidetec TL** | | | | | | |
| Model 3 | 0.5569 | 1.3079 | 0.0032 | 0.3151 | 0.0498 | 0.7545 |
| Model 4 | 0.5569 | 1.3744 | 0.0162 | 0.3112 | 0.0492 | 0.6912 |
| Model 5 | 0.5578 | 1.3570 | 0.0457 | 0.3138 | 0.0496 | 0.7472 |
| Model 6 | 0.5569 | 1.3462 | 0.0132 | 0.3219 | 0.0509 | 0.8146 |
| Model 7 | 0.5569 | 1.3531 | 0.0081 | 0.3221 | 0.0509 | 0.852 |

Notes: Max = Maximum, Min = Minimum, SD = Standard deviation and SE = Standard error.

## 4. Conclusions

This study developed and validated lumber recovery models for *P. patula* using three readily available forest inventory measurements: DBH, total tree height, and stem taper, which represent tree size, vertical form, and stem shape variability, respectively. For Laimet 120, DBH ranged from 24.7–51.1 cm, height from 22.2–35.5 m, and taper from 0.304–1.368 cm/m, while for Slidetec TL, DBH ranged from 21.5–49.9 cm, height from 11.5–31.3 m, and taper from 0.504–1.684 cm/m. Polynomial models combining these variables performed best, explaining over 79% of the variation in recovery for both Laimet 120 and Slidetec TL sawmills. The Mann-Whitney U test confirmed no significant differences between observed and predicted values, underscoring the robustness of the models. By linking inventory variables to lumber recovery, this study offers a practical tool for sawmills and forest managers to forecast yields, optimize log allocation, and reduce waste. The models also provide a scientific basis for policy interventions to enhance value addition and sustainably utilize forest resources. Future research should expand model development across other species and ecological zones and explore integration with advanced machine learning techniques to enhance broader applicability.

**Acknowledgments**

Gratitude is extended to the management of Sokoine University of Agriculture for providing the funding for this research.

**Author Contributions**

G.M.: Conceptualization, Methodology, Data Curation, Formal Analysis, Writing – Original Draft Preparation, Writing – Review and Editing; E.W.M.: Conceptualization, Methodology, Data Curation, Formal Analysis, Writing – Original Draft Preparation, Writing – Review, Editing and supervision; F.B.M.: Conceptualization, Methodology, Data Curation, Formal Analysis, Writing – Original Draft Preparation, Writing – Review, Editing and supervision.

**Conflict of Interest**

The authors have no conflict of interest.

**Declaration of Generative AI and AI-Assisted Technologies in the Manuscript Preparation**

During the preparation of this work, the authors used QuillBot and Grammarly for grammar, spelling, and language refinement to enhance clarity and readability. All content was carefully reviewed and edited by the authors, who take full responsibility for the final version of the publication.

## References

Alderman, D. 2022. *United States Forest Products Annual Market Review and Prospects, 2015–2021*. Department of Agriculture, Forest Service, Forest Products Laboratory, U.S.

Andrew, S. M. 2022. Drivers, Trends, and Management of Forest Plantation Fires in Tanzania. *Trees, Forests and People* 10(1): 100355. DOI: 10.1016/j.tfp.2022.100355

Babune, G. J., Mwalimu, T., and Memorial, N. 2021. The Impact of Wood Processing Factories on Community Livelihoods in Dodoma City, Tanzania. *Journal of Co-operative and Business Studies (JCBS)* 5(2): 95–105. DOI: 10.13140/rg.2.2.21407.69287

Beleko, S., Malimbwi, R. E., Mugasha, W. A., Balama, C., and Mpiri, A. 2021. Pine Resin Productivity at Sao Hill Forest Plantation, Southern Tanzania. *Tanzania Journal of Forestry and Nature Conservation* 90(1): 82–95.

Borz, S. A., Oghnoum, M., Marcu, M. V., Lorincz, A., and Proto, A. R. 2021. Performance of Small-Scale Sawmilling Operations : A Case Study on Time Consumption, Productivity and Main Ergonomics for a Manually Driven Bandsaw. *Forests* 12(6): 810. DOI: 10.3390/f12060810

De Lima, R. B., Ferreira, R. L. C., Da Silva, J. A. A., Guedes, M. C., De Oliveira, C. P., Da Silva, D. A. S., and Silva, R. M. D. A. 2018. Lumber Volume Modeling of Amazon Brazilian Species. *Journal of Sustainable Forestry* 38(3): 262–274. DOI: 10.1080/10549811.2018.1546596

Djenontin, I. N., Zulu, L. C., and Etongo, D. 2021. Ultimately, What is Forest Landscape Restoration in Practice? Embodiments in Sub-Saharan Africa and Implications for Future Design. *Environmental Management* 68(3): 619–641. DOI: 10.1007/s00267-020-01360-y

Dos Santos, A. S., Amaro, M. A., Ribeiro, Í. F. N., Fernandes, A. C., and De Carvalho, C. A. 2024. Adjustment of Thinning Equations for Forest Management Area in the Western Amazon. *Jurnal Sylva Lestari 12*(1): 76–85. DOI: 10.23960/jsl.v12i1.844

Eguakun, F. S., and Nwankwo, P. 2016. Predicting Wood Conversion Efficiency in Illoabuchi Sawmill, Port Harcourt, Nigeria. *New York Science Journal* 9(11): 20–26.

Hodson, T. O. 2022. Root-Mean-Square Error (RMSE) or Mean Absolute Error (MAE): When to Use Them or Not. *Geoscientific Model Development* 15(14): 5481–5487. DOI: 10.5194/gmd-15-5481-2022

Kankare, V. 2015. *The Prediction of Single-Tree Biomass, Logging Recoveries and Quality Attributes with Laser Scanning Techniques*. [Dissertation]. University of Helsinki, Finland.

Kolapo, M. A., Nafiu, A. K., and Kolapo, S. R. 2020. Assessment of Lumber Recovery Efficiency and Waste Generated in Some Selected Wood Processing Industry in South Western Nigeria. *FULafia Journal of Science and Technology* 6(2): 1–5.

Li, C., Huang, S., Barclay, H., and Sidders, D. 2016. Modeling Lumber Yield of White Spruce in Alberta, Canada : A Comparative Approach. *Journal of Forest Research* 21(6): 271–279. DOI: 10.1007/s10310-016-0544-3

Lolila, N. J., Mchelu, H. A., Mauya, E. W., and Madundo, S. D. 2021. Lumber Recovery and Production Rates of Small-Scale Mobile Sawmilling Industries in Northern Tanzania. *Tanzania Journal of Forestry and Nature Conservation* 90(3): 74–83.

Magarik, Y. A., Roman, L. A., and Henning, J. G. 2020. How Should We Measure the DBH of Multi-stemmed Urban Trees?. *Urban Forestry & Urban Greening* 47: 126481. DOI: 10.1016/j.ufug.2019.126481

Mangi, G., Makonda, F.B., and Mauya, E.W. 2025. Productivity and Conversion Efficiency of Circular Sawmills. A Case Study of Laimet 120 and Slidetec TL in Northern Tanzania. *Tanzania Journal of Forestry and Nature Conservation* 94(2): 12–24.

Manina, I. L. O., and Rayment, M. 2025. Estimation of Wood Residues from Small-Scale Sawmill Operations in Kawetire Forest Plantation in the Southern Highlands, Tanzania. *International Journal of Multidisciplinary Research* 7(2): 1–14. DOI: 10.2989/20702620.2024.2316913

Marnaek, R. H., Setiawan, Y., and Hermawan, R. 2024. Development of Allometric Model for Estimating Biomass and Carbon Storage of Hybrid Eucalyptus (*E. grandis* x *E. urophylla*) in Industrial Plantation Forests of North Sumatra Province. *Jurnal Sylva Lestari* 12(3): 948–961. DOI: 10.23960/jsl.v12i3.951

Mauya, E. W., Gesase, L., Mangi, G., Temba, G., Tesha, D., Madundo, S., and Jonas, J. N. 2024. Estimation of Lumber Recovery for Small-Scale Mobile Sawmills in Tanzania. *Southern Forests: A Journal of Forest Science* 86(1): 24–29. DOI: 10.2989/20702620.2024.2316913

Nesha, M. K., Herold, M., Duchelle, A. E., Martius, C., Branthomme, A., and Pekkarinen, A. 2021. An Assessment of Data Sources, Data Quality, and Changes in National Forest Monitoring Capacities in the Global Forest Resources Assessment 2005–2020. *Environmental Research Letters* 16(5): 054029. DOI: 10.1088/1748-9326/abd81b

Ngobi, J., Kambugu, R. K., Mugabi, P., and Banana, A. Y. 2024. Performance of Different Forest Plantation Sawmill Technologies in Uganda. *Research Square* 8(3): 1–14. DOI: 10.21203/rs.3.rs-4011950/v1

Njana, M. A., Mbilinyi, B., and Eliakimu, Z. 2021. The Role of Forests in the Mitigation of Global Climate Change: Empirical Evidence from Tanzania. *Environmental Challenges* 4: 10170. DOI: 10.1016/j.envc.2021.100170

Nti, I. K., Nyarko-Boateng, O., and Aning, J. 2021. Performance of Machine Learning Algorithms with different K values in K-Fold Cross-Validation. *International Journal of Information Technology and Computer Science* 13(6): 61–71. DOI: 10.5815/ijitcs.2021.06.05

Pham, H. 2019. A New Criterion for Model Selection. *Mathematics* 7(12): 1215. DOI: 10.3390/math7121215

Pieterse, J. L. 2022. *A Predictive Model for Precision Tree Measurements using Applied Machine Learning*. [Dissertation]. Stellenbosch University, South Africa.

Rascón-Solano, J., Aguirre-Calderón, O. A., Nájera-Luna, J. A., Olivas-García, J. M., Alanís-Rodríguez, E., Jiménez-Pérez, J., and Treviño-Garza, E. 2023. Comparative Analysis of Volumetric Yield of Sawn Timber from Two Sawmill Variants in Chihuahua. *Revista Chapingo Serie Ciencias Forestales y del Ambiente* 29(3): 17–29. DOI: 10.5154/r.rchscfa.2022.09.069

Rawat, Y. S., Eba, M., and Nebiyu, M. 2023. Lumber Recovery Rate of *Cupressus lusitanica* in Arsi Forest Enterprise, Ethiopia. *Sustainability* 15(2): 1046. DOI: 10.3390/su15021046

Silayo, D. S. A., Migunga, G. A., and Shemwetta, D. T. K. 2014. Modelling the Effect of Training and Job Interruptions on Tree Cutting Time for Chainsaw Operators in Plantation Forests, Tanzania. *Industrial Engineering Letters* 4(7): 1–16

Silver, M. S. 2019. *The Growth of the Manufacturing Industry in Tanzania*. Routledge, New York.

Sushko, O. P., and Efimova, M. V. 2025. Analysis of the Timber Industry Complex of Forest-rich Countries. *Izvestiya Vuzov Lesnoy Zhurnal* 1(2): 154–171. DOI: 10.37482/0536-1036-2025-2-154-171

Temu, B. J., Monela, G. C., Darr, D., Abdallah, J. M., and Pretzsch, J. 2024. Forest Sector Contribution to the National Economy: Example Wood Products Value Chains Originating from Iringa Region, Tanzania. *Forest Policy and Economics* 164(1): 103246. DOI: 10.1016/j.forpol.2024.103246

Vibrans, A. C., Moser, P., Oliveira, L. Z., and de Maçaneiro, J. P. 2015. Generic and Specific Stem Volume Models for Three Subtropical Forest Types in Southern Brazil. *Annals of Forest Science* 72(6): 865–874. DOI: 10.1007/s13595-015-0481-x

Wong, B. J. M., Boyou, M., and Wong, N. S. Y. 2019. *Tropical Tree Plantation and the Economic Value Chain*. CRC Press, U.S.

Worku, S., Belay, B., and Mariam, Y. G. 2022. Effect of Log Taper and Length on Lumber Volume Recovery of *Cupressus Lusitanica* in Injibara Sawmill, Awi Zone, Ethiopia : A Research Article. *Dmujids* 6(1). DOI: 10.20372/dmujids.1000

Younis, A. J., and Hassan, M. K. 2019. Assessing Volume of *Quercus aegilops* L. Trees in Duhok Governorate, Kurdistan Region of Iraq. *Journal of Duhok University* 22(1): 265–276.