BIODIVERSITAS
Volume 27, Number 1, January 2026
Pages: 9

ISSN: 1412-033X
E-ISSN: 2085-4722
https://doi.org/10.13057/biodiv/d270119

Generalized additive models vs. traditional models for teak biomass estimation in Northern Thailand

PONTHEP MEUNPONG[1], CHAKRIT NA TAKUATHUNG[2], LADDAWAN RIANTHAKOOL[2], JIRAWAT YINGDEE[2], THARNRAT KAEWGRAJANG[3], SUPASIT SRIARKARIN[4], NARINTHORN JUMWONG[5], PATTAMA SANGVISITPIROM[5], NARONGCHAI CHONLAPAP[6], THEERAPONG CHUMSANGSRI[2,♥]
[1]Department of Silviculture, Faculty of Forestry, Kasetsart University. 50 Phahonyothin Rd, Bangkok 10900, Thailand
[2]Department of Forest Engineering, Faculty of Forestry, Kasetsart University. 50 Phahonyothin Rd, Bangkok 10900, Thailand.
Tel.: +66-2-579-0170, ♥email: ffortpc@ku.ac.th
[3]Department of Forest Biology, Faculty of Forestry, Kasetsart University. 50 Phahonyothin Rd, Bangkok 10900, Thailand
[4]Department of Forest Management, Faculty of Forestry, Kasetsart University. 50 Phahonyothin Rd, Bangkok 10900, Thailand
[5]Forestry Research Center, Faculty of Forestry, Kasetsart University. 50 Phahonyothin Rd, Bangkok 10900, Thailand
[6]Forest Industry Organization. Rajadamnern Nok Avenue, Bangkok 10200, Thailand

Manuscript received: 1 November 2025. Revision accepted: 19 January 2026.

Abstract. *Meunpong P, Takuathung CN, Rianthakool L, Yingdee J, Kaewgrajang T, Sriarkarin S, Jumwong N, Sangvisitpirom P, Chonlapap N, Chumsangsri T. 2026. Generalized additive models vs. traditional models for teak biomass estimation in Northern Thailand. Biodiversitas 27 (1): d270119. https://doi.org/10.13057/biodiv/d270119.* Accurate estimation of aboveground biomass is essential for carbon accounting and sustainable management of teak (*Tectona grandis*) plantations, yet traditional allometric equations may not capture nonlinear variation in stem form and crown biomass allocation. We developed and evaluated generalized additive models (GAMs) incorporating internal stem geometry (mid-height and mid-diameter) using destructive sampling of 30 trees (n = 30) from managed teak plantations in northern Thailand. Model performance was assessed using leave-one-out cross-validation (LOOCV) and compared with a conventional log-linear DBH-only allometric model. The GAM substantially outperformed the traditional allometric model, reducing prediction error by approximately 44% (RMSE = 28.47 kg vs. 51.28 kg) and increasing predictive accuracy ($R^2 = 0.975$ vs. 0.918). GAM smoothers revealed interpretable nonlinear effects for DBH and mid-height, while mid-diameter was penalized toward linearity. Residual diagnostics indicated adequate model fit, and sensitivity analysis showed that GAM performance was robust to changes in smoother complexity. Branch biomass exhibited lower deviance explained but moderate LOOCV accuracy. Leaf biomass showed low predictability due to inherently high biological variability. Monte Carlo simulations propagated tree-level uncertainty to plot-level biomass. The GAM produced substantially narrower 95% confidence intervals (0.91-1.85 t ha^{-1}) than the allometric model (1.76-4.14 t ha^{-1}), indicating improved stability for operational carbon assessments. These results demonstrate that flexible modeling approaches incorporating internal stem geometry can significantly enhance both accuracy and precision of teak biomass estimation. Generalized additive models provide a robust alternative to traditional allometries, particularly where nonlinear stem structure influences biomass allocation, and offer clear advantages for plantation-scale carbon accounting and sustainable forest management.

Keywords: Allometric equation, biomass estimation, generalized additive model (GAM), Teak plantation, *Tectona grandis*

INTRODUCTION

Teak (*Tectona grandis* L.f.) is one of the most commercially important and ecologically significant tropical hardwood species, prized worldwide for its exceptional timber quality, durability, and resistance to decay (Fuseini et al. 2025). Native to Southeast Asia, it has been widely planted across tropical regions globally, making it a cornerstone of high-value forestry, plantation economics, and sustainable timber production (Stewart et al. 2021; Kollert et al. 2024). Beyond economic importance, mature teak stands contribute to habitat structure, microclimate regulation, and biodiversity support, particularly for understory plants, epiphytes, insects, and bird communities that utilize the multi-layered canopy architecture (Harikrishnan et al. 2012; Woods et al. 2019). In Thailand, teak plantations play an essential role in national forestry programs, rural livelihoods, and carbon sequestration initiatives. Therefore, accurate estimation of teak biomass is crucial for understanding its carbon sequestration potential, optimizing forest management, and participating in climate mitigation programs such as REDD$^+$ (Kenzo et al. 2020).

The foundational approach to biomass estimation in forestry has been the use of allometric equations, which relate tree dimensions, most commonly diameter at breast height (DBH), total tree height (H), and wood density (ρ), due to their simplicity and minimal data requirements (Chave et al. 2014; Vorster et al. 2020). In teak, allometric equations have been developed across multiple regions, including Thailand (Ounban et al. 2015; Kenzo et al. 2020), India (Buvaneswaran et al. 2006), Costa Rica (Cordero and Kanninen 2003), and Ecuador (Aguilar et al. 2019). Although widely adopted, these traditional equations often assume fixed parametric relationships and may not adequately capture nonlinearities in tree growth, variation in crown structure, and site-specific differences in stem form. In recent years, the forestry and ecological modeling community has increasingly recognized the limitations of rigid allometric equations, particularly for biomass components such as branches and leaves, which exhibit high structural variability (Henry et al. 2011; Picard et al. 2015). Crown architecture in teak can vary substantially with topography, stand density, silvicultural treatments, and competition for light (Cordero and Kanninen 2003). These factors influence biomass allocation patterns and introduce nonlinear relationships that traditional power-law allometries may not represent accurately. Furthermore, destructive sampling, the basis for allometric equation development, is costly and labor-intensive, and is typically conducted on small sample sizes, underscoring the need for flexible modeling approaches that can generalize across diverse plantation conditions. Generalized additive models (GAMs) offer a flexible, semi-parametric alternative for biomass estimation. GAMs use smoothing functions to model nonlinear relationships between biomass and predictor variables, without requiring a predetermined functional form (Wood 2017). Their capacity to handle complex, nonlinear effects makes them especially suitable for representing tree allometry, taper variation, and crown-related biomass. Incorporating internal stem geometry, such as mid-height and mid-diameter, may further improve predictive performance by capturing variation in bole shape and tapering, attributes particularly relevant for teak, which exhibits diverse stem forms across management regimes (Choochuen et al. 2021; Salekin et al. 2021; Pokhrel et al. 2024).

Despite the potential of GAMs and other flexible models, their application in teak biomass modeling remains limited, and few studies have rigorously compared them against traditional allometric approaches using robust validation frameworks such as leave-one-out cross-validation (LOOCV). Furthermore, little is known about how these modeling choices affect plot-level biomass and uncertainty propagation, both of which are critical for carbon stock assessments and sustainable forest management. This gap is particularly relevant in northern Thailand, where teak (*T. grandis*) plantations occur across diverse topographic and management conditions that influence stem form, crown architecture, and biomass distribution. Therefore, we evaluate (i) tree-level predictive performance of GAMs relative to a DBH-only allometric model, and (ii) how these differences propagate to plot-level biomass uncertainty using Monte Carlo simulation. Using detailed destructive sampling and leave-one-out cross-validation, we assess model performance, robustness, and implications for carbon accounting and sustainable forest management in teak plantations. Incorporating internal stem geometry into a flexible generalized additive model yields more accurate and stable estimates of teak aboveground biomass than a DBH-only allometric equation.

MATERIALS AND METHODS

Study area and sampling design

Tree biomass data were collected within 17-18 year old teak (*T. grandis*) plantations spaced at 4 × 4 m in Lampang Province, northern Thailand (18.26655° N, 99.75982° E). Elevation within the study area ranges from 300 to 700 m above sea level (masl), with long-term average temperatures of 14-36°C. Annual precipitation averages 1,055 mm, with most rainfall occurring from May to October (Electricity Generating Authority of Thailand 2024). The Lampang teak plantations are representative of managed teak systems in northern Thailand in terms of stand age, spacing, silvicultural practices, and environmental conditions, while recognizing that the single-site design limits broader geographic generalization. Twelve permanent 40 × 40 m plots were established across the plantation to capture variation in slope position (upper, middle, lower) and stand structure (thinned and unthinned), and served as the full inventory frame for stand-level measurements. All trees within each plot were measured for diameter at breast height (DBH) using a diameter tape, and total height was measured using a hypsometer (Haglöf Vertex 5). From these plots, 30 trees were selected for destructive sampling and stratified into three DBH classes (<10 cm, 10-15 cm, >15 cm) to ensure adequate representation across the whole diameter range present in the stand. The sampling target was 8-12 trees per class, resulting in 12 small trees (<10 cm), 11 medium trees (10-15 cm), and seven large trees (>15 cm). Destructive sampling was conducted in five plots (T1F1, T1F2, T1F3, T2F2, and T2S3) for operational reasons, resulting in a total destructively sampled area of 0.8 ha (5 × 0.16 ha).

Field measurements and biomass determination

After felling, each stem was cut into 1 m segments from base to tip. For each segment, upper and lower over-bark diameters (D_1 and D_2 cm) were measured perpendicular to the stem axis with a diameter tape (0.1 cm), and segment boundaries were recorded with a measuring tape (0.01 m). Mid-height was the midpoint between boundaries, and mid-diameter was the mean of D_1 and D_2. Total height was measured along the stem axis on the felled tree. DBH and segment diameters were measured by the same trained observer using calibrated instruments under a standardized protocol; repeatability was not formally quantified but was considered a minor limitation relative to biological variability. Each tree was separated into stem, branches, and leaves. Stem bolts (1 m) were weighed fresh and subsampled for oven drying at 65°C to constant mass; branches and leaves were similarly processed to estimate moisture content and dry biomass following IPCC (2006). Biomass was reported as tree-level dry mass (kg) and plot-level aboveground biomass (t ha^{-1}), with total aboveground biomass computed as the sum of stem, branch, and leaf dry biomass. Modeling variables included DBH, mid-diameter, mid-height, and dry mass of each component.

Allometric model and generalized additive models (GAMs)

A traditional log-linear allometric model was developed following widely used formulations in tropical biomass research (Chave et al. 2014). We focused on a DBH-only allometry because it represents the operational baseline model used in teak plantations. Preliminary analyses indicated that including total tree height in the allometric model resulted in only marginal improvements in predictive performance, and height was therefore excluded to retain a parsimonious DBH-only baseline. The baseline model used in this study was:

$$\log(B) = \beta_0 + \beta_1 \log(DBH)$$

Where B is the biomass component (total, stem, branch, or leaf), models were fitted using the lm() function in R. Predictions were back-transformed using the exponential function and corrected for retransformation bias using Duan's smearing estimator (Duan 1983), ensuring unbiased estimates on the original biomass scale.

In addition to the allometric model, we fitted generalized additive models (GAMs) using log-transformed biomass as the response variable. Predictors included DBH, mid-height, and mid-diameter, allowing the models to capture internal stem geometry and within-tree taper variation, thereby improving the flexibility of biomass estimation. The GAMs were calculated by using:

$$\log(B) \sim s(DBH) + s(\text{mid-height}) + s(\text{mid-diameter})$$

Where *s()* denotes a thin-plate regression spline smoother.

GAMs were fitted to log-transformed biomass using a Gaussian likelihood in the mgcv package (Wood 2017). In all models, log denotes the natural logarithm (ln). Predictor variables (DBH, mid-height, and mid-diameter) were entered on their original measurement scales and were not standardized, allowing direct interpretation of smoother shapes in relation to observed tree dimensions. Model summaries included effective degrees of freedom (edf), deviance explained (%), and F-statistics for each smooth term with significance of predictors ($p<0.05$). Predictions were back-transformed to the original biomass scale with bias correction applied where appropriate. A Gamma regression with log link was also tested, but failed to converge under LOOCV due to zero and near-zero biomass values, making the model unsuitable for this dataset.

Model validation: Leave-one-out cross-validation

Model performance was evaluated using leave-one-out cross-validation (LOOCV) at the tree level ($n = 30$). In each iteration, the model was refitted to 29 trees and used to predict the held-out individual. Because all models were fitted on the log-transformed scale, predictions were back-transformed and bias-corrected using Duan's smearing estimator, and all performance metrics were calculated on the original biomass scale. Predictive accuracy was quantified using RMSE, MAE, and R^2. GAMs were fitted using restricted maximum likelihood (REML) with a basis dimension of $k = 6$ for each smoother, and smoother complexity was evaluated using effective degrees of freedom (edf), where values near 1 indicate linearity and larger values reflect increasing nonlinearity. REML penalization automatically limited excessive smoothness, particularly in the mid-diameter term.

Diagnostic evaluation

Model assumptions and goodness of fit were assessed using standard diagnostic tools applied to the log-scale residuals. These included residuals versus fitted values to detect systematic trends, Q-Q plots to evaluate normality, scale-location plots to assess potential heteroscedasticity, and histograms to inspect residual symmetry. Heteroscedasticity was formally tested by regressing squared residuals against fitted values using an F-test. Although mild heteroscedasticity was detected, the strong predictive performance under leave-one-out cross-validation indicates that this variance structure does not compromise model reliability. Smoothing adequacy and basis-dimension choices were examined using the gam.check function in mgcv. All diagnostic evaluations were conducted in R using the mgcv and car packages.

Monte Carlo uncertainty propagation

To propagate tree-level prediction uncertainty to plot-level biomass, we used a residual-based Monte Carlo simulation for both the GAM and the DBH-only allometric model. For each of 5,000 iterations, log-scale residuals were resampled with replacement and added to the model linear predictor to generate simulated log-biomass values, which were then back-transformed using Duan's smearing estimator. Simulated tree biomass values were summed within each 40 × 40 m plot (0.16 ha) and converted to t ha^1. The 2.5th and 97.5th percentiles of the simulated distributions were used as empirical 95% confidence intervals, and interval widths were compared between models.

Sensitivity to smoother complexity

To evaluate sensitivity to smoother flexibility, GAMs were refitted with basis dimensions $k = 4$, 6, and 8. LOOCV metrics (RMSE, MAE, R^2) were compared across models to determine whether the choice of k influenced predictive stability or introduced unnecessary flexibility.

RESULTS AND DISCUSSION

Tree-level model performance (LOOCV)

Leave-one-out cross-validation (LOOCV) revealed apparent performance differences between the traditional DBH-only allometric model and the GAM (Table 1). The allometric model yielded RMSE = 51.28 kg, MAE = 29.39 kg, and $R^2 = 0.918$, whereas the GAM that incorporated DBH, mid-diameter, and mid-height achieved substantially higher predictive accuracy (RMSE = 28.47 kg, MAE = 16.62 kg, $R^2 = 0.975$). These results indicate that internal stem geometry reduced prediction error by approximately 44% relative to the DBH-only allometry. From component-level performance, stem biomass showed strong predictability ($R^2 = 0.955$), whereas branch biomass exhibited moderate accuracy ($R^2 = 0.883$), reflecting higher biological variability in crown-related components. Leaf biomass had very low predictability, with negative LOOCV R^2 (-10.76), indicating no predictive skill beyond using the mean and reflecting its small magnitude and inherently high biological variability. These patterns are consistent with the low deviance explained for branch and leaf models on the log scale and support the interpretation that crown-level biomass is more variable than stem biomass. This discrepancy arises because deviance explained is computed on the log scale and is strongly driven by trees with tiny branch or leaf masses, while LOOCV R^2 reflects predictive accuracy on the original scale in absolute biomass units. Despite their higher biological variability, branch and leaf components still exhibited meaningful predictive structure, although with greater unexplained variance than stem or total biomass. Bias correction affected LOOCV performance metrics differently across models. For the DBH-only allometric model, naïve back-transformation without correction underestimated prediction error, whereas applying Duan's smearing estimator increased RMSE and reduced R² to more realistic values. In contrast, GAM performance metrics changed negligibly after bias correction, reflecting lower residual variance and greater robustness to retransformation (Table 2). Because both GAM and allometric models were fitted under a common Gaussian likelihood, AIC values were computed on a consistent scale. However, AIC evaluates relative likelihood fit rather than predictive accuracy and can be sensitive to distributional assumptions. Accordingly, LOOCV metrics provide the primary basis for comparison in this study, and these consistently favored the GAM. Although AIC results showed the same model ranking, they were not used as the principal selection criterion (Tables 3 and 4).

Figure 1 shows the relationship between observed total biomass and LOOCV-predicted values for both the DBH-only allometric model and the GAM. Predictions from the GAM were more tightly clustered around the 1:1 line, indicating closer agreement with observed values across the biomass range. In contrast, the allometric model exhibited greater scatter and systematic under- and over-prediction at lower and higher biomass values, respectively. These visual patterns are consistent with the LOOCV accuracy metrics, where the GAM achieved a substantially lower RMSE (28.47 kg vs. 51.28 kg) and higher R² (0.975 vs. 0.918) than the allometric model. Together, these results confirm that the GAM provides more reliable predictions for individual trees and better captures the nonlinear biomass-stem geometry relationships present in the dataset. Branch and leaf biomass were more challenging to predict than stem or total biomass, a pattern that aligns with biological expectations. These components exhibit inherently greater variability because they are strongly influenced by crown architecture, competition, pruning history, and site-specific micro-environmental conditions that are not fully captured by DBH, mid-diameter, or mid-height (Ounban et al. 2015; Ruiz-Blandon et al. 2025). As a result, traditional power-law allometries tend to underperform for these components, and even flexible models such as GAMs explain only modest deviance on the log scale. GAMs achieved moderate LOOCV accuracy for branch biomass but very low accuracy for leaf biomass, consistent with their higher biological variability and the challenges of predicting small-mass components from trunk-based measurements. Future improvements will likely require predictors that more directly characterize crown structure, such as crown radius, crown length, or live-crown ratio, or remote sensing proxies from UAV photogrammetry or LiDAR, which have been shown to improve biomass estimation for crown-dominated components in other tropical forest systems (Soraya et al. 2025).

GAM smoothers

The GAM smoothers (Table 5) showed moderate complexity, with effective degrees of freedom (edf) between 3.51 and 3.60 for DBH and mid-height, indicating nonlinear but well-penalized relationships. Mid-diameter displayed an edf = 1 and a non-significant p-value ($p = 0.7106$), meaning that the smoothing penalty shrank this term to near-linearity and minimized its influence on the model. This indicates that the GAM did not overfit the data and that the inclusion of mid-diameter was not detrimental because the REML procedure automatically penalized unnecessary curvature. DBH and mid-height remained highly significant predictors of biomass ($p<0.001$).

The structure of the GAM smoothers provides insight into how different stem attributes contribute to biomass prediction. DBH and mid-stem height exhibited moderate nonlinear effects, with effective degrees of freedom (3.51-3.6), confirming that these variables contain meaningful curvature that improves predictive accuracy. In contrast, mid-stem diameter displayed an edf =1 and was not statistically significant ($p = 0.7106$), indicating that the REML smoothing penalty shrank this term toward linearity and effectively minimized its influence on the model. This adaptive penalization is a key advantage of GAMs, and potentially useful predictors that can be included without risking overfitting, because unnecessary complexity is automatically penalized (Marra and Wood 2011; Wood 2017). The strong significance of DBH and mid-height ($p < 0.001$) underscores their dominant role in explaining variation in biomass, while the minimal contribution of mid-diameter suggests that stem form information is captured primarily through DBH and vertical stem position. Overall, the pattern of smoothers demonstrates that the GAM is both flexible and parsimonious, capturing essential nonlinear structure without introducing superfluous model complexity.

Table 1. Model validation using LOOCV (n = 30 trees)

Model	RMSE (kg)	MAE (kg)	R²
Allometry (log DBH)	51.28	29.39	0.918
GAM (DBH + mid geometry)	28.47	16.62	0.975

Table 2. LOOCV predictive performance for biomass components using GAMs

Component	RMSE (kg)	MAE (kg)	R²
Stem	27.66	17.89	0.955
Branch	18.93	12.60	0.883
Leaf	8.83	3.82	-10.76

Table 3. Effect of bias correction on LOOCV performance metrics for total aboveground biomass

Model	Bias correction	RMSE	MAE	R2
Allometric (DBH-only)	No (exp only)	47.989	27.742	0.929
Allometric (DBH-only)	Yes (Duan smearing)	51.277	29.386	0.918
GAM (DBH + mid geometry)	No (exp only)	25.739	16.230	0.979
GAM (DBH + mid geometry)	Yes (Duan smearing)	25.777	16.218	0.979

Table 4. Akaike Information Criterion (AIC) values for the DBH-only allometric model and the geometry-informed GAM fitted to log-transformed total aboveground biomass

Model	AIC
Allometric (DBH-only)	-13.00246
GAM (DBH + mid geometry)	-39.53434

Table 5. Estimated degrees of freedom (edf) and significance of GAM smoothers for total tree biomass

Smooth term	EDF	Ref.df	F-value	p-value
s(DBH)	3.51	4.10	13.65	0.000011
s(mid-diameter)	1.00	1.00	0.14	0.7106
s(mid-height)	3.60	4.26	8.00	0.000397

Note: Smooth terms were considered statistically significant at $p<0.05$

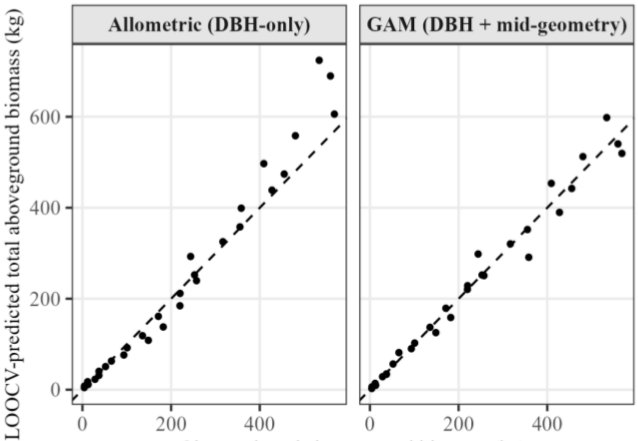


Figure 1. Observed vs LOOCV-predicted total aboveground biomass for the DBH-only allometric model and the GAM

Residual diagnostics

Model diagnostics indicated that the GAM provided an adequate representation of the relationship between tree attributes and total biomass. Figure 2 shows that the residuals were centered around zero with no visible structure, trend, or curvature across the fitted range, suggesting that the model captured the mean response appropriately. The Q-Q plot (Figure 3) demonstrates that the residuals closely followed the theoretical normal distribution across the central quantiles, with only mild deviations in the lower and upper tails. Such behavior is typical for biomass datasets modeled on the log scale, where smaller and larger trees often contribute slightly heavier-tailed residuals. Additional diagnostic plots, including the scale-location plot and the residual histogram (Figure 4), showed no severe violations of model assumptions. The scale-location plot indicated a modest decline in variance with increasing fitted values, an expected pattern for biomass data. The histogram confirmed that residuals were approximately symmetric and free of extreme outliers. A regression-based test for heteroscedasticity detected a significant relationship between squared residuals and fitted values ($F = 12.35$, $p = 0.0015$), but this effect was mild and did not produce problematic patterns in the diagnostic plots. Collectively, these results indicate that the GAM residuals were well-behaved and exhibited no patterns suggestive of misspecification. The strong cross-validation performance (LOOCV $R^2 = 0.975$) further supports the robustness of the model despite minor heteroscedasticity, which is typical of biomass datasets.

The GAM substantially outperformed the DBH-only allometric model, consistent with previous studies showing that flexible regression methods better capture nonlinear biomass-structure relationships than parametric allometries (Zianis et al. 2005; Henry et al. 2011; Goodman et al. 2014). This improvement indicates that internal stem attributes, particularly mid-stem height, provide meaningful information beyond DBH alone. The GAM smoothers revealed moderate nonlinearity in DBH and mid-stem height effects, while mid-stem diameter was penalized toward linearity, demonstrating appropriate smoothing through REML (Wood 2017) and no evidence of overfitting. Diagnostic plots supported model adequacy, residuals exhibited no structural patterns, Q-Q plots showed acceptable normality, and only mild heteroscedasticity was observed, an expected feature of log-transformed biomass data (Picard et al. 2015). This variance pattern is common in biomass datasets and does not invalidate model predictions given the strong cross-validation performance. Given the strong LOOCV performance, this variance structure does not compromise predictive reliability. These findings highlight the potential of flexible modeling approaches to improve biomass estimation in plantation forestry. Internal stem measurements require more effort than DBH alone, but analogous structural information can increasingly be derived from terrestrial and airborne LiDAR, enabling integration of GAM-based models into operational inventories (Dassot et al. 2011; Soraya et al. 2025). For applications requiring rapid biomass assessments, height-inclusive allometries may serve as a practical compromise between accuracy and measurement cost (Chave et al. 2014).

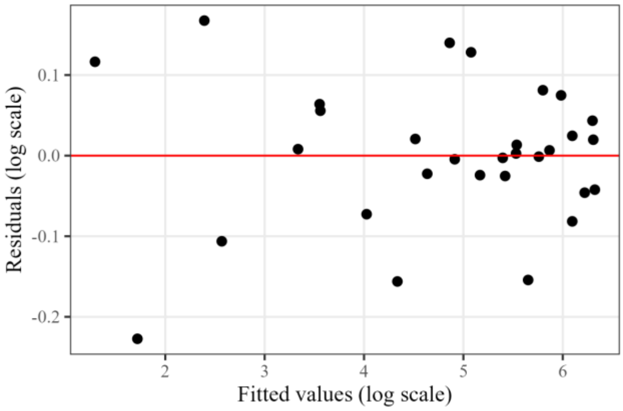


Figure 2. Residuals vs fitted values for the GAM predicting total tree biomass on the log scale

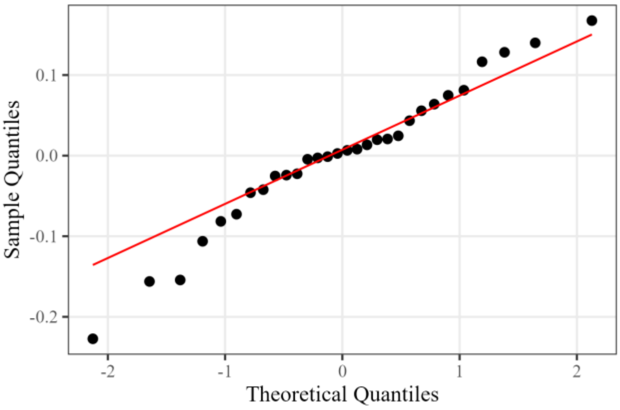
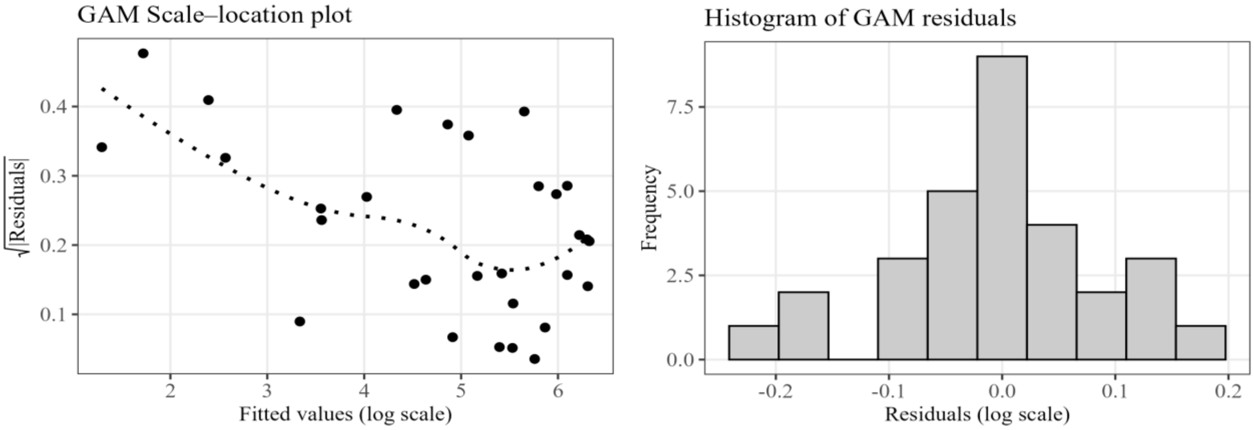


Figure 3. Quantile-quantile (Q-Q) plot of GAM residuals on the log scale



A

B

Figure 4. Additional diagnostic plots: A. The scale-location plot, and B. The residual histogram

Bias-correction effects

Because models were fitted on the log scale, we evaluated the impact of back-transformation. Without correction, the GAM produced biased, systematically low predictions and an artificially small training RMSE of 15.87 kg. After applying Duan's smearing estimator, RMSE recovered to the correct LOOCV value of 28.47 kg. This demonstrates the necessity of bias correction and ensures accurate comparison of model performance. A Gamma regression with log link was fitted for completeness, but LOOCV failed due to numerical divergence (infinite RMSE), likely caused by zero or very small dry-weight biomass components, which violate the distributional assumptions of the Gamma model.cv (Gamma model evaluated under cross-validation). This further supports the suitability of the log-GAM with smearing correction and confirms that bias correction was essential for valid model comparisons.

Sensitivity analysis of smoother basis dimension (k)

Sensitivity analysis showed that GAM performance was largely insensitive to the choice of smoother basis dimension. Models fitted with k = 4, 6, and 8 produced nearly identical LOOCV metrics (Table 5), indicating stable predictive performance and no evidence of overfitting across reasonable levels of smoother complexity. The close agreement among k-values demonstrates that the GAM is not overly sensitive to basis dimension and that the default configuration (k = 6) provides adequate flexibility without overfitting.

The GAM sensitivity analysis showed that varying the smoothing basis dimension (k = 4, 6, 8) produced nearly identical LOOCV performance (RMSE = 28.17-28.47 kg; $R^2 = 0.9749$-0.9754), indicating results were not driven by smoother complexity. This stability likely reflects REML penalization effectively limiting overfitting in a moderate sample (Wood 2017), consistent with previous ecological and forestry studies where GAMs are robust to reasonable changes in k (Marra and Wood 2011). The weak dependence on k also suggests the nonlinear patterns captured are meaningful rather than artifacts, supporting the GAM as a reliable model of biomass-stem geometry relationships in this study.

Plot-level biomass estimates from GAM and allometric models

Monte Carlo uncertainty propagation further highlighted the advantages of the GAM over the DBH-only allometric equation. Across the five sampled 40 × 40 m plots, the GAM produced substantially narrower 95% confidence intervals for plot-level aboveground biomass (0.91-1.85 t ha^{-1}) compared to the allometric model (1.76-4.14 t ha^{-1}), despite producing similar mean estimates within plots (e.g., 13.16 vs. 12.67 t ha^{-1} in T1F1 and 8.34 vs. 8.89 t ha^{-1} in T1F2) (Figure 5). Although spatial replication was limited, the observed reduction in uncertainty reflects methodological improvements of the GAM framework rather than site-specific effects. The most significant reductions in uncertainty occurred in T1F1 and T2F2, where the allometric CI widths (3.09 and 4.14 t ha^{-1}, respectively) were more than twice those of the GAM (1.55 and 1.62 t ha^{-1}).

Plot-level aboveground biomass estimates and associated 95% confidence intervals derived from Monte Carlo simulation are summarized in Table 6. These results demonstrate that the GAM not only improves predictive accuracy at the tree level (LOOCV R^2 = 0.975) but also enhances precision when uncertainty is propagated to the operational scale of forest plots. The reduction in uncertainty likely arises from the GAM's ability to capture nonlinear relationships between biomass and stem geometry (Wood 2017), as well as its penalized smoothing framework, which minimizes overfitting even with modest sample sizes. Similar findings have been reported in studies showing that flexible models incorporating stem form or crown metrics yield more stable stand-level biomass estimates than traditional allometries that rely solely on DBH (Chave et al. 2014; Goodman et al. 2014; Picard et al. 2015). As remote sensing technologies increasingly provide detailed structural attributes such as height profiles and diameter distributions, GAM-based approaches may offer more accurate and reliable pathways for scaling biomass estimates in plantation systems and tropical forestry contexts. Propagation of prediction uncertainty from tree to plot scale revealed that GAM-based estimates consistently exhibited tighter distributions and reduced upper-tail variance. The allometric model produced broader uncertainty bands due to its lower accuracy at the tree level. These findings demonstrate the operational advantages of GAMs for plot-level biomass estimation. Overall, our findings demonstrate that flexible regression approaches leveraging stem geometry offer significant improvements in accuracy and precision for biomass estimation in plantation forestry.

Accurately modeling aboveground biomass in tropical species such as teak is essential for forest management and carbon accounting. This study provides one of the first direct evaluations of generalized additive models (GAMs) against traditional DBH-only allometric equations for teak. While conventional log-linear models remain widely used due to their simplicity (Ledo et al. 2016; Vorster et al. 2020), their fixed form limits their ability to predict crown-related components, which are strongly influenced by architecture, competition, and silvicultural treatment (Ounban et al. 2015; Ruiz-Blandon et al. 2025).

Table 5. Sensitivity of GAM performance to basis dimension (k)

Model	RMSE (kg)	MAE (kg)	R^2
GAM (k = 4)	28.17	16.79	0.97539
GAM (k = 6)	28.47	16.62	0.97488
GAM (k = 8)	28.46	16.69	0.97488

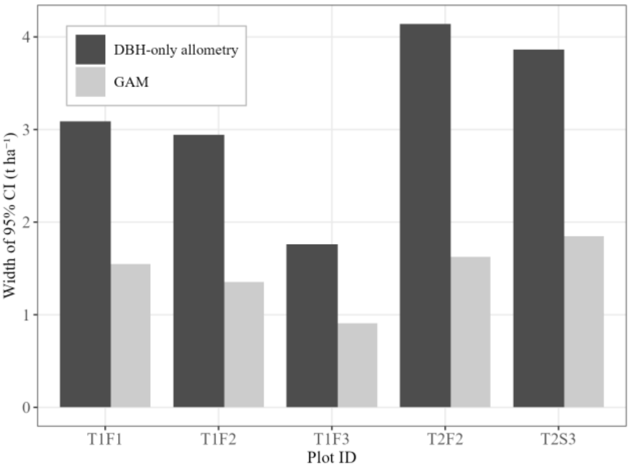


Figure 5. Width of 95% confidence intervals (t ha^{-1}) for plot-level aboveground biomass from Monte Carlo simulation (5,000 iterations)

Table 6. Plot-level aboveground biomass estimates (t ha^{-1}) and 95% confidence intervals derived from residual-based Monte Carlo simulation for the GAM and DBH-only allometric model

Plot ID	GAM mean	GAM low95	GAM high95	GAM width	Allo mean	Allo low95	Allo high95	Allo width
T1F1	13.16	12.38	13.93	1.55	12.67	11.12	14.2	3.09
T1F2	8.34	7.65	9.01	1.35	8.89	7.44	10.38	2.94
T1F3	4.41	3.94	4.84	0.91	4.27	3.37	5.13	1.76
T2F2	6.97	6.11	7.73	1.62	8.71	6.66	10.8	4.14
T2S3	9.34	8.39	10.24	1.85	10.18	8.22	12.08	3.86

GAMs, by contrast, flexibly model nonlinear relationships using smoothing functions (Wood 2017), producing substantially lower prediction error across biomass components. Improvements were most pronounced for branch and leaf biomass, which are components that simple power-law equations often fail to represent, supporting evidence from other forest systems that semi-parametric models better capture canopy complexity (Mauro et al. 2020). Incorporating mid-diameter and mid-height further enhanced model performance by capturing stem taper and internal geometry, key traits distinguishing teak trees with similar DBH but different structure (Salekin et al. 2021; Pokhrel et al. 2024). These results highlight the value of internal geometry for biomass estimation in species with variable form and crown dynamics (Cordero and Kanninen 2003).

GAMs also improved operational relevance by producing narrower plot-level uncertainty intervals in Monte Carlo simulations, strengthening their suitability for carbon accounting and inventory applications. Although GAMs require more data and computational support than traditional allometries, their gains in predictive accuracy and uncertainty reduction demonstrate clear benefits. Mixed-effects, ensemble, and Bayesian hierarchical models can offer advantages in larger or more structured datasets (Pinheiro and Bates 2000; Breiman 2001; Gelman and Hill 2007). However, our dataset included only five plots with destructively sampled trees, providing insufficient replication to reliably estimate random effects or hierarchical variance components, and making ensemble approaches statistically unstable under cross-validation when sample size is small (Hastie et al. 2009). In contrast, GAMs provided strong predictive performance under LOOCV and allowed biologically interpretable nonlinear effects, consistent with the strengths of penalized spline models in ecological applications (Marra and Wood 2011; Wood 2017).

Future work with larger hierarchical datasets may explore NLME or Bayesian models to enable multilevel inference and propagate uncertainty across spatial scales. A practical pathway for forestry applications may involve using GAMs to identify key nonlinear relationships and structural predictors, then translating these insights into simplified parametric equations for field deployment. This balance between accuracy and usability is important for large-scale biomass estimation frameworks. However, our study is limited by its modest sample size (n = 30) and single-site focus. Future work should evaluate model transferability across stands, age classes, and management regimes, incorporate crown metrics or UAV-LiDAR predictors, and explore hierarchical or Bayesian GAMs for more comprehensive uncertainty propagation.

In conclusion, this study demonstrates that flexible modeling approaches substantially improve aboveground biomass estimation in teak (*T. grandis*) plantations compared with traditional DBH-only allometric equations. Using destructive sampling of 30 trees from managed plantations in northern Thailand, the generalized additive model (GAM) incorporating internal stem geometry markedly enhanced tree-level predictive performance. Under leave-one-out cross-validation, the GAM reduced prediction error by approximately 44%, lowering RMSE from 51.28 kg to 28.47 kg and increasing predictive accuracy from $R^2 = 0.918$ to $R^2 = 0.975$. Model smoothers revealed significant nonlinear effects of DBH and mid-stem height, while mid-stem diameter was automatically penalized toward linearity, indicating that the GAM captured essential structural information without overfitting. This study is based on a relatively small destructive sample size (n = 30) drawn from a single plantation site and age class, which may limit direct transferability of model parameters to other regions, stand ages, or management regimes. Although leave-one-out cross-validation provides a robust assessment of predictive performance, the analyses remain correlational and focused on stem-based predictors, which constrain the predictability of crown-dominated components such as leaves. Future research should evaluate the transferability of GAM-based biomass models across broader environmental gradients, stand structures, and age classes, and incorporate explicit crown metrics (e.g., crown length, crown radius, live-crown ratio) or structural data derived from UAV or LiDAR platforms. Expanding sample size and spatial replication would also enable hierarchical or Bayesian GAM frameworks to represent multilevel variability better and propagate uncertainty across tree, plot, and landscape scales.

ACKNOWLEDGEMENTS

The authors sincerely thank the Chief and all staff of Mae Jang Forest Plantation, Forest Industry Organization (FIO), for their invaluable support during the fieldwork. Their assistance and cooperation greatly contributed to the success of this study. This research was funded by the Thailand Science Research and Innovation and the Kasetsart University Research and Development Institute, KURDI, FF(KU)42.67.

REFERENCES

Aguilar FJ, Nemmaoui A, Peñalver A, Rivas JR, Aguilar MA. 2019. Developing allometric equations for teak plantations located in the coastal region of Ecuador from terrestrial laser scanning data. Forests 10 (12): 1050. https://doi.org/10.3390/f10121050.

Buvaneswaran C, George M, Pérez D, Markku K. 2006. Biomass of teak plantations in Tamil Nadu, India, and Costa Rica compared. J Trop For Sci 18 (3): 195-197.

Chave J, Réjou-Méchain M, Búrquez A, Chidumayo E, Colgan MS, Delitti WB, Duque A, Eid T, Fearnside PM, Goodman RC, et al. 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. Glob Chang Biol 20 (10): 3177-3190. https://doi.org/10.1111/gcb.12629.

Choochuen T, Suksavate W, Meunpong P. 2021. Development of a taper equation for teak (*Tectona grandis* L.f.) growing in eastern Thailand. Environ Nat Resour J 19 (3): 176-185. https://doi.org/10.32526/ennrj/19/2020183.

Cordero LP, Kanninen M. 2003. Aboveground biomass of *Tectona grandis* plantations in Costa Rica. J Trop For Sci 15 (1): 199-213.

Breiman L. 2001. Random forests. Machine Learning 45 (1): 5-32. https://doi.org/10.1023/A:1010933404324.

Dassot M, Constant T, Fournier M. 2011. The use of terrestrial LiDAR technology in forest science: Application fields, benefits, and challenges. Ann For Sci 68: 959-974. https://doi.org/10.1007/s13595-011-0102-2.

Duan N. 1983. Smearing estimate: A nonparametric retransformation method. Am Stat Assoc 78: 383. 605-610. https://doi.org/10.1080/01621459.1983.10478017.

Electricity Generating Authority of Thailand. 2024. Mae Moh Weather Information. www.watertele.egat.co.th/maemoh.

Fuseini A, Addaney M, Akudugu JA, Nyarko BJ, Kang-Milung S. 2025. Understanding the teak timber value chain in the Bono Region of Ghana: Analysis and policy implications. For Econ Rev 7 (2): 177-196. https://doi.org/10.1108/FER-08-2023-0008.

Gelman A, Hill J. 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511790942.

Goodman RC, Phillips OL, Baker TR. 2014. The importance of crown dimensions to improve tropical tree biomass estimates. Ecol Appl 24 (4): 680-698. https://doi.org/10.1890/13-0070.1.

Hastie T, Tibshirani R, Friedman J. 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. Springer, New York. https://doi.org/10.1007/978-0-387-84858-7.

Harikrishnan S, Vasudevan K, Udhayan A, Mathur PK. 2012. Biodiversity values of abandoned teak, *Tectona grandis,* plantations in the Southern Western Ghats: Is there a need for management intervention?. Basic Appl Ecol 13 (2): 139-148. https://doi.org/10.1016/j.baae.2012.01.001.

Henry M, Picard N, Trotta C, Manlay R, Valentini R, Bernoux M, Saint André L. 2011. Estimating tree biomass of sub-Saharan African forests: A review of available allometric equations. Silva Fenn 45 (3B): 38. https://doi.org/10.14214/sf.38.

IPCC. 2006. Guidelines for National Greenhouse Gas Inventories: Vol 4 Agriculture, Forestry and Other Land Use. Intergovernmental Panel on Climate Change. Institute for global environmental strategies, Hayama, Kanagawa, Japan.

Kenzo T, Himmapan W, Yoneda R, Tedsorn N, Vacharangkura T, Hitsuma G, Noda I. 2020. General estimation models for above- and belowground biomass of teak (*Tectona grandis*) plantations in Thailand. For Ecol Manag 457: 117701. DOI: 10.1016/j.foreco.2019.117701.

Kollert W, Sandeep S, Sreelakshmy MP. 2024. Global Teak Resources and Market Assessment 2022. International Union of Forest Research Organizations (IUFRO), Vienna.

Ledo A, Cornulier T, Illian JB, Iida Y, Kassim AR, Burslem DFRP. 2016. Re-evaluation of individual diameter: Height allometric models to improve biomass estimation of tropical trees. Ecol Appl 26 (8): 2376-2382. DOI: 10.1002/eap.1450.

Marra G, Wood SN. 2011. Practical variable selection for generalized additive models. Comput Stat Data Anal 55 (7): 2372-2387. https://doi.org/10.1016/j.csda.2011.02.004.

Mauro F, Hudak AT, Fekety PA, Frank B, Temesgen H, Bell DM, Gregory MJ, McCarley TR. 2020. Regional modeling of forest fuels and structural attributes using airborne laser scanning data in Oregon. Remote Sens 13 (2): 261. https://doi.org/10.3390/rs13020261.

Ounban W, Puangchit L, Diloksumpun S. 2015. Development of general biomass allometric equations for *Tectona grandis* Linn.F. and *Eucalyptus camaldulensis* Dehnh. plantations in Thailand. Agric Nat Resour 50 (1): 48-53. https://doi.org/10.1016/j.anres.2015.08.001.

Picard N, Rutishauser E, Ploton P, Ngomanda A, Henry M. 2015. Should tree biomass allometry be restricted to power models? For Ecol Manag 353: 156-167. https://doi.org/10.1016/j.foreco.2015.05.035.

Pinheiro JC, Bates DM. 2000. Mixed-Effects Models in S and S-PLUS. Springer, New York. https://doi.org/10.1007/978-1-4419-0318-1.

Pokhrel NR, Subedi MR, Malego B. 2024. Fitting and evaluating taper functions to predict upper stem diameter of planted teak (*Tectona grandis* L.F.) in eastern and central regions of Nepal. Forests 16 (1): 77. https://doi.org/10.3390/f16010077.

Ruiz-Blandon BA, Hernández-Alvarez E, Martínez-Trinidad T, Amaringo-Cordova LP, Ucañay-Ayllon TM, Bernaola-Paucar RM, Hernández-Plascencia G, Orellana-Mendoza E. 2025. Growth, productivity, and biomass-carbon allometry in teak (*Tectona grandis*) plantations of western Mexico. Forests 16 (10): 1521. https://doi.org/10.3390/f16101521.

Salekin S, Catalán CH, Boczniewicz D, Phiri D, Morgenroth J, Meason DF, Mason EG. 2021. Global tree taper modelling: A review of applications, methods, functions, and their parameters. Forests 12 (7): 913. https://doi.org/10.3390/f12070913.

Soraya E, Umarhadi DA, Wardhana W, Senawi S, Ardiansyah F. 2025. Aerial imagery and UAV-LiDAR data fusion for quantifying aboveground carbon stock of teak plantation. Ecol Eng Environ Technol 26 (2): 178-192. ttps://doi.org/10.12912/27197050/197151.

Stewart HTL, Race DH, Rohadi D, Schmidt DM. 2021. Growth and profitability of smallholder sengon and teak plantations in the Pati district, Indonesia. For Policy Econ 130: 102539. https://doi.org/10.1016/j.forpol.2021.102539.

Vorster AG, Evangelista PH, Stovall AEL, Ex S. 2020. Variability and uncertainty in forest biomass estimates from the tree to landscape scale: The role of allometric equations. Carbon Balance Manag 15 (1): 8. https://doi.org/10.1186/s13021-020-00143-6.

Wood SN. 2017. Generalized Additive Models: An Introduction with R. Chapman and Hall. New York. https://doi.org/10.1201/9781315370279.

Woods CL, Nevins LM, Didier EJ. 2019. Structural heterogeneity of trees influences epiphyte distributions in a northern temperate rainforest. J Veg Sci 30 (6): 1134-1142. https://doi.org/10.1111/jvs.12797.

Zianis D, Muukkonen P, Mäkipää R, Mencuccini M. 2005. Biomass and Stem Volume Equations for Tree Species in Europe. Silva Fennica Monographs. http://www.metla.fi/silvafennica/full/smf/smf004.pdf