Revista Mexicana de Ciencias Forestales vol. 17 (96)
Proyecto Estratégico Forestal (2026)

DOI: 10.29298/rmcf.v17i96.1646
Research article

Model-assisted inference for mean estimation of forest volume and biomass in Mexico

Inferencia asistida por modelos para la estimación media de volumen y biomasa forestal en México

Efraín Velasco-Bautista[1], Martin Enrique Romero-Sánchez[2]*, Alma Delia Ortiz-Reyes[1], Jesús Valentín Gutiérrez-García[1]

Fecha de recepción/Reception date: 23 de febrero de 2026.
Fecha de aceptación/Acceptance date: 14 de mayo de 2026.

[1]Centro Nacional de Investigación Disciplinaria en Conservación y Mejoramiento de Ecosistemas Forestales. Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias. México.
[2]Campo Experimental Valle de México. Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias. México.

*Autor para correspondencia; correo-e: romero.martin@inifap.gob.mx
*Corresponding author; e-mail: romero.martin@inifap.gob.mx

Abstract

Accurate estimation of timber volume and above-ground biomass in terms of variance reduction is essential for sustainable forest management and for the estimation of above-ground carbon. This study evaluated the performance of the Model-assisted estimator (*MAE*) across different functional forms, given a single continuous auxiliary variable derived from Sentinel-2A, and a model that regards the treatment as a categorical variable, compared to the Horvitz–Thompson estimator (*HTE*) in its simple form. Eight models (linear, generalized, and nonlinear) were analyzed using a quasi-systematic field sampling scheme in a managed temperate forest in *Puebla*, Mexico. The Green Normalized Difference Vegetation Index (*GNDVI*) was selected as an auxiliary variable using regularization methods (LASSO and Elastic Net, with cross-validation as the selection criterion). The population means estimated using *MAE* were consistent across models and comparable to the Horvitz–Thompson estimator. Significant differences in relative efficiency were observed in both volume and above-ground biomass estimations when analyzing the variance of *MAE* relative to the *HTE*. For harvestable volume, the model using *GNDVI* and silvicultural management achieved a 37.65 % reduction in variance; for above-ground biomass, the reduction was 30.21 %. The findings show that model-assisted estimation significantly improves accuracy without compromising unbiasedness.

Keywords: Horvitz-Thompson estimator, quasi-systematic sampling, Sentinel-2, ecosystem services, variance, forest volume and biomass.

Resumen

La estimación precisa, en términos de reducción de la varianza, del volumen maderable y la biomasa aérea es fundamental para el manejo forestal sostenible y la estimación de carbono aéreo. Este estudio evaluó el desempeño del Estimador asistido por modelos (*MAE*, por sus siglas en inglés) al considerar diferentes formas funcionales cuando se dispone de una única variable auxiliar continua derivada de *Sentinel-2A* y un modelo que considera el tratamiento como variable categórica respecto al Estimador *Horvitz-Thompson* (*HTE*, por sus siglas en inglés) en su versión simple. Se analizaron ocho modelos (lineal, generalizados y no lineales) bajo un esquema de muestreo en campo cuasi-sistemático en un bosque templado con manejo forestal de Puebla, México. El Índice de Vegetación de la Diferencia Normalizada Verde (*GNDVI*, por sus siglas en inglés) fue seleccionado como variable auxiliar mediante procedimientos de regularización (*LASSO* y *Elastic Net*, criterio de selección validación cruzada). Las medias poblacionales estimadas mediante *MAE* fueron consistentes entre modelos y comparables con el Estimador de *Horvitz-Thompson*. Se observaron diferencias importantes en eficiencia relativa tanto en la estimación de volumen, como de biomasa aérea cuando se analiza la varianza de *MAE* respecto a *HTE*. Para volumen maderable, el modelo con *GNDVI* y el tratamiento silvícola logró la reducción de varianza de 37.65 %; en tanto que en la biomasa aérea fue de 30.21 %. Los hallazgos demuestran que la estimación asistida por modelos incrementa significativamente la precisión sin comprometer el insesgamiento.

Palabras clave: Estimador *Horvitz-Thompson*, muestreo cuasi-sistemático, *Sentinel-2*, servicios ecosistémicos, varianza, volumen y biomasa forestal.

Introduction

Reliable estimates of forest attributes such as timber volume and above-ground biomass are essential for sustainable management, carbon accounting, and ecosystem monitoring (Hu & Sun, 2022). The accuracy of these estimates enables assessment of carbon stocks and productivity, as well as the development of strategies to mitigate climate change (Zadbagher et al., 2024).

Traditionally, forest inventories have used systematic probability sampling with a random start and design-based estimators, such as the Horvitz-Thompson estimator (*HTE*) (Horvitz & Thompson, 1952), which ensures unbiasedness under appropriate inclusion probabilities. However, variance estimation proves intractable or unstable in systematic or quasi-systematic designs that are common in operational inventories such as the USDA' FIA program; therefore, approximations based on simple random sampling are frequently used.

Although in theory the Horvitz-Thompson estimator has favorable properties, it can exhibit high variance in heterogeneous populations (Ståhl et al., 2016). To mitigate this limitation, model-assisted estimation incorporates auxiliary information available for the entire population, combining predictions with weighted residuals (McConville et al., 2020), whereby design-unbiased and potentially more efficient estimators can be obtained when the auxiliary variable is predictive (McRoberts et al., 2006).

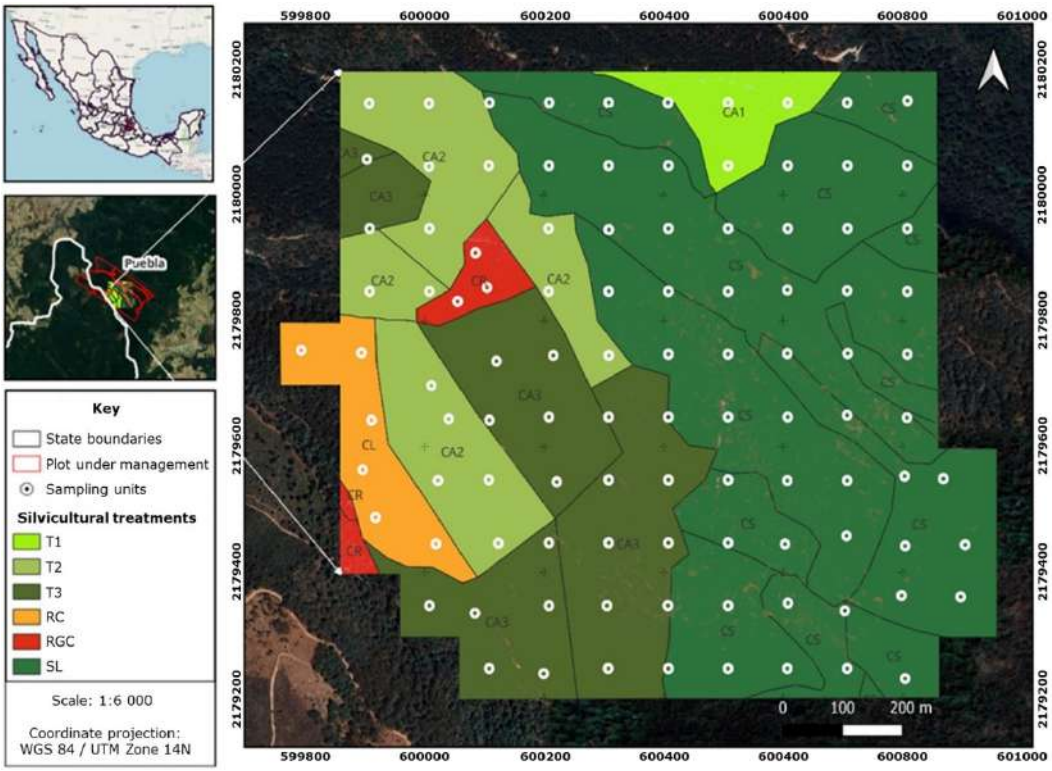
Within this context, the availability of Sentinel-2 imagery has advanced the estimation of forest variables on a large scale (McRoberts et al., 2006). Vegetation indexes, particularly those in the red and near-infrared bands, show a high correlation with structural attributes, which allows for the use of models such as linear regression, additive models, and machine learning techniques (Khan et al., 2024).

However, in Mexico, particularly in managed forest areas, there is limited evidence on how model choice affects the efficiency of the Model-assisted estimator (*MAE*) from a design perspective, especially when only a single auxiliary variable is available. Therefore, this study evaluates the performance of different models in estimating volume and biomass using an *MAE* approach, with an emphasis on variance reduction, compared to the Horvitz–Thompson estimator in a quasi-systematic sampling context.

Materials and Methods

Study area

The study was conducted in a managed forest in the *Emiliano Zapata ejido*, *Chignahuapan* municipality, state of *Puebla*, Mexico. This forest is characterized by the presence of pine trees (*Pinus ayacahuite* Ehrenb. *ex* Schltdl., *Pinus teocote* Schied. *ex* Schltdl. & Cham., *Pinus greggii* Engelm. *ex* Parl., *Pinus pseudostrobus* Lindl.), firs (*Abies religiosa* (Kunth) Schltdl. & Cham.), oaks (*Quercus rugosa* Née, *Quercus crassifolia* Bonpl., *Quercus laurina* Bonpl.), and, to a lesser extent, other broad-leaved trees (*Alnus acuminata* Kunth, *Arbutus xalapensis* Kunth). Field data were collected in late 2023 on a forest property managed with the Silvicultural Development Method (SDM) and the Mexican Method for the Management of Irregular Forests (MMOBI). The SDM included the following silvicultural treatments: thinning (T), release cutting (RC), and regeneration cutting (RGC), carried out in 2016, 2014, and 2014, respectively. The MMOBI study examined the selective logging (SL) treatment implemented in 2016. The total sampled area covered approximately 102 ha (Figure 1).



CA1 = T1; CA2 = T2; CA3 = T3; CL = RC; CR = RGC; CS = SL. Each circle represents a circular sampling unit measuring 1 000 m^2.

Figure 1. Quasi-systematic sampling design in the study area.

A total of 102 circular sampling units were established using a quasi-systematic sampling design, with a theoretical spacing of 100 m between plots (Figure 1). Each sampling unit covered 1 000 m^2. Based on the methodology defined for the National Forest and Soil Inventory (Comisión Nacional Forestal [Conafor], 2017), all trees within each plot with a diameter at breast height (*DBH*) above 7.5 cm were measured, using a 95-cm model Mantax Blue Haglöf® aluminum caliper and a model 283D Forestry Suppliers® diameter tape; total height (m) was measured with a model PM-5a Suunto® clinometer. In addition, the species of the measured individuals was recorded.

Field estimation of the timber volume and biomass

The timber volume and above-ground biomass of each tree were calculated using species-specific allometric equations previously developed for the study region. Based on the species identified in the field, the equations were taken from the following sources: Avendaño-Hernández et al. (2009), Soriano-Luna et al. (2015), Díaz-Ríos et al. (2016), Arias-Téllez and García-Martínez (2017) and Correa-Díaz et al., (2025).

The volume and total biomass per plot were obtained by aggregating the individual tree estimates for all the species found in each sampling unit. Based on the totals for each sampling unit, the *HTE* and *MAE* estimators were applied to obtain estimates of the average timber volume and average above-ground biomass per sampling unit (1 000 m^2 circular plot).

Supplementary data derived from Sentinel-2

The auxiliary data were obtained from Sentinel-2A multispectral images (October-November 2023) retrieved via the Google Earth Engine. Scenes with a cloud cover <5 % were selected, and cloud and shadow masking were applied (Khan et al., 2024).

To reduce temporal variability, a mean composite was created from the filtered dataset. Five widely used vegetation indices were calculated for this composite: *NDVI* (Normalized Difference Vegetation Index), *GNDVI* (Green Normalized Difference Vegetation Index), *EVI* (Enhanced Vegetation Index), *MSAVI* (Modified Soil Adjusted Vegetation Index), and *SIPI* (Structure-Insensitive Pigment Index) (Besic et al., 2025). The average values per plot, obtained through spatial interpolation, were used as auxiliary variables in the models.

Auxiliary variable selection

Exploratory analyses were conducted to assess the relationship between forest attributes and spectral indexes using scatter plots and histograms. Due to the slight nonlinearity and strong asymmetry in volume and biomass, Spearman's rank correlation coefficients were applied to assess monotonic relationships (Dhiman & Kumar, 2025).

The high correlation among the indices indicated multicollinearity; therefore, regularization techniques (LASSO and Elastic Net, using cross-validation for selection) were applied to select a parsimonious auxiliary variable (Valbuena et al., 2017). Both methods consistently identified the *GNDVI* as the most informative predictor.

Comparative analysis of models

The analysis was based on the simplified version of the Horvitz-Thompson estimator (*HTE*), which uses simple random sampling estimators, even though the data were collected in a quasi-systematic manner (Frank & Monleon, 2021). For comparative purposes with respect to the *HTE*, the Model-assisted estimator (Ståhl et al., 2016) was utilized, initially considering two generalized linear models, one generalized additive model, and four explicitly nonlinear models in its prediction term (Table 1).

Table 1. Models used (*MAE*) in the comparative analysis.

Model type	Model structure
Generalized linear model (LIM) with normal distribution	Distribution: $y_i \sim Normal(\eta_i, \sigma^2)$, $E(y_i)=\mu_i$, $var(y_i)=\sigma^2$ Linear predictor: $\eta_i=\beta_0+\beta_1 x_i$ Liaison role: $\eta_i=\mu_i$, identity link
Generalized linear model (GLM) with Gamma distribution	Distribution: $y_i \sim Gamma(a_i, \beta_i)$; $E(y_i)= a_i\beta_i=\mu_i$, $var(y_i)=a_i\beta_i^2=\frac{\mu_i^2}{a_i}$ Linear predictor: $\eta_i=\beta_0+\beta_1 x_i$ Liaison role: $\eta_i= log(\mu_i)$, $log\, link$
Generalized additive model (GAM) with Gamma distribution	Distribution: $y_i \sim Gamma(a_i, \beta_i)$; $E(y_i)= a_i\beta_i=\mu_i$, $var(y_i)=a_i\beta_i^2=\frac{\mu_i^2}{a_i}$ Linear Predictor: $\eta_i=\beta_0+f_1(x_i)$ Liaison role: $\eta_i= log(\mu_i)$, log link
Schumacher (NL1)	$y_i=e^{\beta_0+\beta_1/x_i}+\varepsilon_i$, $\varepsilon_i \sim N(0,\sigma^2)$
Exponential (NL2)	$y_i=e^{\beta_0+\beta_1 x_i}+\varepsilon_i$, $\varepsilon_i \sim N(0,\sigma^2)$
Potential (NL3)	$y_i=\beta_0 x_i^{\beta_1}+\varepsilon_i$, $\varepsilon_i \sim N(0,\sigma^2)$
Inverse Michaelis-Menten (NL4)	$y_i=\frac{\beta_0}{\beta_1+x_i}+\varepsilon_i$, $\varepsilon_i \sim N(0,\sigma^2)$

$a_i..n$; $\beta_{i...n0}, \beta_1$ = Coefficients of regression; μ_i = Population mean; σ^2 = Population variance; ε_i = Random error; y = Volume or biomass; x = *GNDVI*; x and y are provided per sample unit.

Based on the observed dispersion in the volume-*GNDVI* and biomass-*GNDVI* ratios, seven models were evaluated to analyze, in comparison to the *HTE*, the sensitivity and behavior of model-assisted estimators with different functional structures; a single continuous auxiliary variable was used: the *GNDVI*. Within this context, *Y* represents the response variable (volume or biomass per sampling unit), while *X* represents the predictor variable (*GNDVI*) (Table 1).

In addition, a generalized model (Ståhl et al., 2016) was considered, including not only the *GNDVI* but also the silvicultural treatment (*CAT*) as a categorical variable.

Model fit assessment

Since the models included both generalized linear models and nonlinear models, the following fit criteria were utilized: Root mean square error (RMSE), Mean absolute error (*MAE*), Mean residual error (*MRE*), and pseudo-R^2, as well as the Akaike information criterion (*AIC*) and the Bayesian information criterion (*BIC*) (González-Rosales & Ortiz-Paniagua, 2022).

These metrics were not used as the sole selection criterion, but rather as tools to characterize the behavior of the residuals, given that, within the context of model-assisted estimation, the variance of the estimator depends directly on the residual variability under the sampling design (Ståhl et al., 2016).

Assessment of *MAE*'s performance regarding *HTE*

Using the *HTE* and *MAE* estimators, as mathematically defined by McConville et al. (2020), the means for volume and biomass per sampling unit were calculated, along with their respective variances. The *HTE* uses only field data, whereas the *MAE* estimates the population mean by combining the average of the model's predictions for all units (N=1 089) with an adjustment based on the average of the residuals. The performance of the *MAE* was evaluated against the *HTE* using the following criteria (Dettmann et al., 2022):

Relative efficiency (*RE*): $RE = \frac{\widehat{Var}(\hat{a}_{MAE})}{\widehat{Var}(\hat{a}_{HTE})}$ (1)

Variance reduction: (*VR*): $VR = (1 - RE)100$ (2)

Standard error (*SE*) of the estimator: $SE = \sqrt{\widehat{Var}(\hat{a})}$: (3)

Relative sampling error (*RSE*): $RSE = \frac{Z_{\propto/2}\, SE}{\hat{a}} \times 100$ (4)

Where:

$\hat{a}$ = Population parameter estimation

$Z_{\propto/2}$ = Critical value of the normal distribution for the confidence level.

$\widehat{Var}$ = Variance

Variable selection and exploratory analysis were performed using SAS version 9.4 (SAS Institute Inc., 2024). Model fitting, as well as the calculation of model-based estimators and their variances, were performed in R version 4.5.0 (R Core Team, 2025), using the mgcv and nls2 packages.

Results

The inventory, comprising 102 sample plots, revealed a sample mean of 31.18 m^3 for timber volume and 15 711 kg for above-ground biomass (*AGB*), both per sample plot (green vertical line in Figure 2). Both variables exhibited positive asymmetry (asymmetry coefficients of 0.78 for volume and 0.82 for biomass). The Gamma distribution, which belongs to the exponential family, was found to converge rapidly and provide the best theoretical fit for timber volume (*AIC*=819.85 and *BIC*=825.10); it also proved to be an ideal candidate for biomass (*AIC*=2 056 and *BIC*=2 061) (Figure 2).

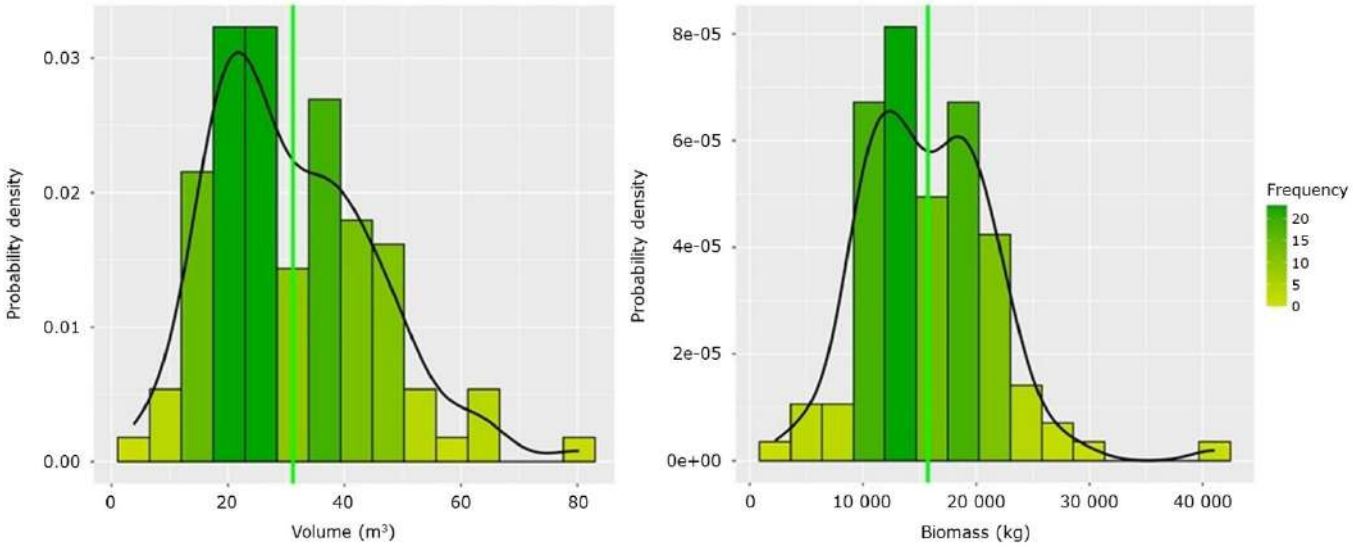


Figure 2. Distribution of the variables volume (m^3) and biomass (kg) per sampling unit.

Auxiliary variable selection

The correlation analysis revealed significant relationships between forest attributes and all Sentinel-2 spectral indices. The timber volume showed a stronger correlation with spectral data (*mean*=-0.51) than the biomass (*mean*=-0.42), suggesting that these variables may be useful for predicting these forest variables at the sampling unit level. The LASSO and Elastic Net regression procedures identified *GNDVI* as the auxiliary variable with the greatest predictive power, and thus the greatest impact for both parameters (the minimum cross-validation lambda for volume was 0.35 (*log lambda*=-1.06) for LASSO and 1.46 (*log lambda*=0.38) for Elastic Net. Meanwhile, for biomass, the value was 341.13 (*log lambda*=5.84) for Lasso and 682.26 (*log lambda*=6.25) for Elastic Net. In this regard, the Spearman correlation coefficients were -0.52 between volume and *GNDVI*, and -0.45 between biomass and *GNDVI*.

Comparative model fitting for timber volume and above-ground biomass

With the exception of pseudo-R^2, the seven models evaluated showed little difference in performance when used to predict timber volume (Table 2).

Table 2. Fitting parameters of the models evaluated for predicting harvestable volume.

Model		*RMSE*	*MAE*	*MRE*	Pseudo-R^2	*AIC*	*BIC*
LIM	$\hat{\beta}_0$=*232.75*** $\hat{\beta}_1$=*-291.18***	11.76	9.58	0.001	0.29	798.4	806.28
GLM	$\hat{\beta}_0$=*9.661*** $\hat{\beta}_1$=*-9.028***	11.59	9.41	0.047	0.31	789.9	797.86
GAM	$\hat{\beta}_0$=*3.412*** *s(GNDVI)=1.12***	11.58	9.41	0.092	0.31	790.0	798.21
NL1	$\hat{\beta}_0$=*-3.458*** $\hat{\beta}_1$=*4.746* **	11.52	9.34	0.039	0.31	794.1	802.02
NL2	$\hat{\beta}_0$=*10.421*** $\hat{\beta}_1$=*-10.136***	11.56	9.36	0.055	0.31	794.7	802.66
NL3	$\hat{\beta}_0$=*2.330** $\hat{\beta}_1$=*-6.942* **	11.54	9.35	0.047	0.31	794.4	802.32
NL4	$\hat{\beta}_0$=*3.304*** $\hat{\beta}_1$=*-0.580***	11.49	9.34	0.002	0.321	793.52	801.396

RMSE = Root mean square error; *MAE* = Mean absolute error; *MRE* = Mean residual error; *AIC* = Akaike information criterion; *BIC* = Bayesian information criterion. Models: LIM = Generalized linear model with normal distribution; GLM = Generalized linear model with Gamma distribution; GAM = Generalized additive model with Gamma distribution; NL1 = Schumacher; NL2 = Exponential; NL3 = Potential; NL4 = Inverse Michaelis-Menten. *Significant coefficients; **Highly significant coefficients.

Although there are certain similarities in the fitting criteria, the NL4 nonlinear (inverse Michaelis–Menten) model had the lowest value for *RMSE* (11.490) and the highest for pseudo-R^2 (0.321). Similarly, the GLM and GAM models, which incorporated the Gamma distribution, had the lowest *AIC* values (789.9 and 790.0, respectively), indicating a better fit compared to the linear model (LIM), which had the highest *AIC* (798.4). This behavior was also observed with the *BIC*.

Consequently, for the volume case, all models produced point estimates that were very similar to one another across the *GNDVI* spectrum and close to the observed mean obtained using the Horvitz–Thompson estimator (*HTE*), which was 31.18 m^3 (Figure 3).

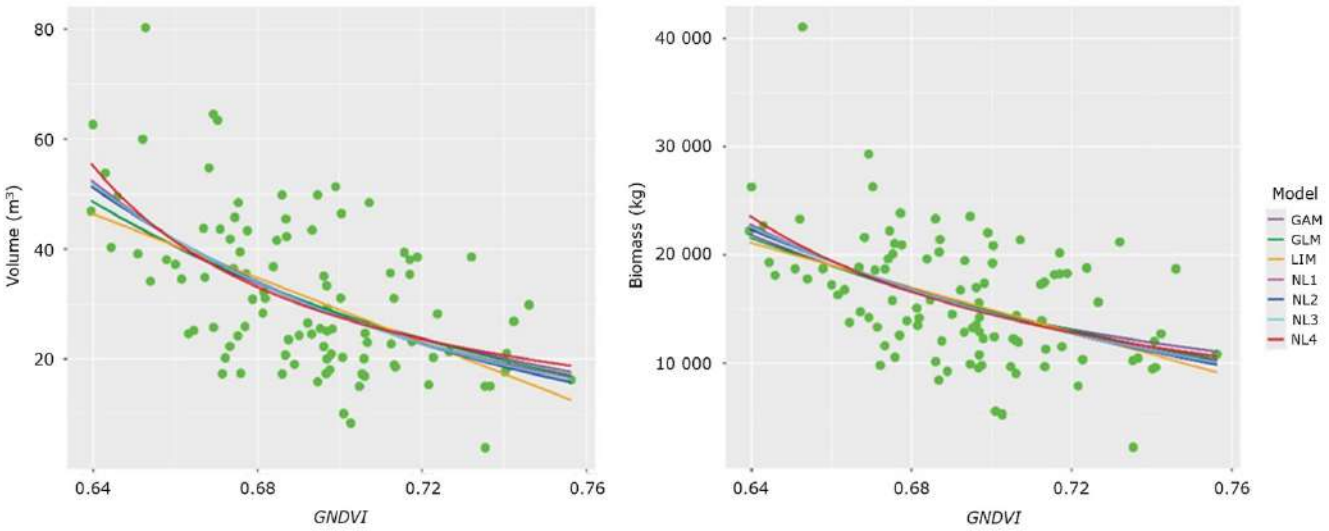


Figure 3. Volume distribution *versus GNDVI* (left) and above-ground biomass *versus GNDVI* (right), along with the curves from the evaluated models.

In the prediction for above-ground biomass, the results were consistent with the trends observed for volume; *i. e.*, the fitted values showed little variation. In this case, the relatively highest *AIC* and *BIC* values were observed for GLM and GAM (Table 3).

Table 3. Fitting parameters of the models evaluated for predicting above-ground biomass.

	Model	*RMSE*	*MAE*	*MRE*	Pseudo-R^2	*AIC*	*BIC*
LIM	$\hat{\beta}_0$=*86 838*** $\hat{\beta}_1$=*-102 748***	5 063.57	4 024.07	0.000	0.21	2 035.54	2 043.42
GLM	$\hat{\beta}_0$=*13.999*** $\hat{\beta}_1$=*-6.284***	5 022.83	3 984.25	11.39	0.22	2 036.00	2 043.88
GAM	$\hat{\beta}_0$=*9.648*** $s(GNDVI)$=*1.316***	5 000.37	3 962.27	31.76	0.23	2 036.02	2 044.73
NL1	$\hat{\beta}_0$=*4.824*** $\hat{\beta}_1$ =*3.333***	4 999.07	3 984.82	10.59	0.23	2 032.93	2 040.80
NL2	$\hat{\beta}_0$=*14.513*** $\hat{\beta}_1$=*-7.032* ***	5 014.41	3 992.71	13.51	0.22	2 033.55	2 041.43
NL3	$\hat{\beta}_0$=*2 589.40*** $\hat{\beta}_1$=*-4.847***	5 006.31	3 988.85	12.07	0.23	2 033.22	2 041.10
NL4	$\hat{\beta}_0$=*2 257.397*** $\hat{\beta}_1$=*-0.544***	4 984.17	3 976.36	11.82	0.23	2 032.32	2 040.19

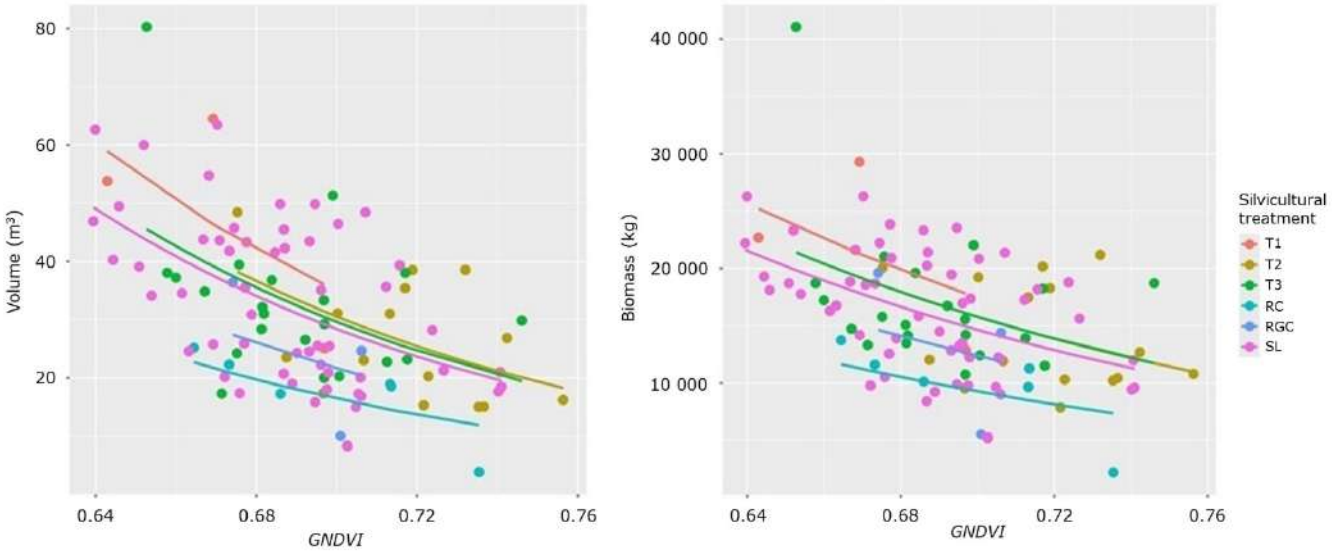
RMSE = Root mean square error; *MAE* = Mean absolute error; *MRE* = Mean residual error; *AIC* = Akaike information criterion; *BIC* = Bayesian information criterion. Models: LIM = Generalized linear model with normal distribution; GLM = Generalized linear model with Gamma distribution; GAM = Generalized additive model with Gamma distribution; NL1 = Schumacher; NL2 = Exponential; NL3 = Potential; NL4 = Inverse Michaelis-Menten. **Highly significant coefficients.

In this case, the NL4 model performed best (*RMSE*=4 984.17), with *AIC* and *BIC* values similar to those of the other models. The linear and generalized models exhibited lower variability (Coefficient of variation [*CV*]≈16 %) than the nonlinear models (*CV*≈18 %), with differences at the extremes.

Given the similar performance in *MAE*, the silvicultural treatment was included in the GLM along with *GNDVI*, due to its high internal variability (*CV* up to 55.7 %). This

improved the fit (*RMSE*=11.011, pseudo-R^2=0.377, *AIC*=786.361). Both predictors were significant (p<0.05), resulting in treatment-specific models based on *GNDVI*.

The above-ground biomass was modeled using a GLM with silvicultural treatment and *GNDVI* as predictors, given the high variability among treatments (*CV* up to 54 %). The model improved the fit (*RMSE*=4 770.92; pseudo-R^2=0.30; *AIC*=2 032.5), and both predictors were significant (p<0.05). The model includes treatment-specific indicator variables, allowing for differentiated estimates based on the *GNDVI* (Figure 4).



T1 = Thinning 1; T2 = Thinning 2; T3 = Thinning 3; RC = Release cutting; RGC = Regeneration cutting; SL = Selective logging.

Figure 4. Volume (left) and biomass (right) estimated using a GLM that includes *GNDVI* and silvicultural treatment.

Efficiency and accuracy gains (*MAE* vs. *HTE*)

Overall, the *MAE* outperformed the *HTE* in all of the model structures evaluated. The average *RE* value for *MAE* relative to *HTE*, considering all seven models, was 0.6897; that is, on average, regardless of the evaluated model, the variance of *MAE* was 68.97 % of the variance of *HTE*, resulting in an average reduction in volume variance for *MAE* relative to *HTE* of 31.03 %. The NL4 model showed a 32.17 % reduction in variance, slightly higher than the other models, indicating a 6.93 % decrease in the relative sampling error. The average volume figures for *MAE* and *HTE* are virtually identical, at 31 m^3 per plot, which is equivalent to 310 m^3 per hectare (Table 4).

Table 4. Estimates of harvestable timber volume per sampling unit and variance values for each evaluated model.

Estimator	Model	Mean	Variance	*RE*	*VR* (%)	*SE*	*RSE* (%)
HTE	Control	31.18	1.74	1		1.321	8.30
MAE	LIM	30.74	1.24	0.714	28.64	1.115	7.10
MAE	GLM	30.78	1.20	0.691	30.91	1.098	6.99
MAE	GAM	30.80	1.20	0.690	31.03	1.097	6.98
MAE	NL1	30.74	1.19	0.683	31.66	1.092	6.96
MAE	NL2	30.74	1.19	0.687	31.29	1.095	6.98
MAE	NL3	30.74	1.19	0.685	31.54	1.093	6.97
MAE	NL4	30.75	1.18	0.678	32.16	1.088	6.93

RE = Relative efficiency; *VR* = Variance reduction; *SE* = Standard error; *RSE* = Relative sampling error. *HTE* = Horvitz-Thompson estimator; *MAE* = Model-assisted estimator.

In the case of above-ground biomass, the relative efficiency of *MAE* compared to *HTE* was 0.7706, resulting in an average reduction in variance of 22.94 % when the seven models were considered together. The NL4 model achieved an accuracy gain of 23.84 %, slightly higher than the other models, resulting in a relative sampling error of 5.95 % (Table 5).

Table 5. Biomass estimates per sampling unit and variance values for each evaluated model.

Estimator	Model	Mean	Variance	*RE*	*VR* (%)	*SE*	*RSE* (%)
HTE	-	15 711.14	292 983.16	1		541.002	6.749
MAE	LIM	15 557.05	230 081.39	0.786	21.389	479.668	6.043
MAE	GLM	15 570.64	226 393.25	0.774	22.649	475.808	5.989
MAE	GAM	15 579.67	224 365.16	0.767	23.342	473.672	5.959
MAE	NL1	15 555.26	224 256.23	0.766	23.379	473.557	5.967
MAE	NL2	15 554.68	225 633.55	0.771	22.909	475.009	5.985
MAE	NL3	15 554.96	224 905.47	0.768	23.157	474.242	5.980
MAE	NL4	15 553.80	222 920.90	0.762	23.835	472.145	5.950

RE = Relative efficiency; *RV* = Variance reduction; *SE* = Standard error; *RSE* = Relative sampling error. *HTE* = Horvitz-Thompson estimator; *MAE* = Model-assisted estimator.

Although the average estimates per plot (1 000 m^2) were similar (15 711.14 kg for *HTE* and 15 560.87 kg for *MAE*), the accuracy of the *MAE* varied substantially among the seven evaluated models. In the model-assisted approach, variance reduction depends on the model's explanatory power; in the absence of predictive power, the *MAE* converges to the *HTE* in efficiency.

The generalized model incorporating *GNDVI* and silvicultural treatments showed the highest relative efficiency (62.4 % for volume and 69.8 % for biomass), which corresponds to reductions in variance of 37.65 % and 30.21 %, and sampling errors of 6.6 % and 5.7 %, respectively (Table 6). The standard error confirmed its superiority under the conditions evaluated. Although models based solely on *GNDVI* showed similar performance, the combined inclusion of quantitative and qualitative variables resulted in substantial improvements over the estimator based exclusively on field data.

Table 6. Relative efficiency and variance reduction for the GLM model that accounts for *GNDVI* and silvicultural treatment (*CAT*).

Variable	Mean	Variance	*RE*	*VR* (%)	*SE*	*RSE* (%)
Volume	30.950	1.088	0.624	37.649	1.043	6.606
Biomass	15 647.160	204 255.187	0.698	30.213	451.946	5.662

RE = Relative efficiency; *VR* = Variance reduction; *SE* = Standard error; *RSE* = Relative sampling error.

Furthermore, a relative efficiency of 0.624 indicates that, using the GLM-*CAT* Model-assisted estimator, a sample of approximately 64 units (102×0.624≈64) would be enough to obtain a variance in the harvestable volume estimate equivalent to that achieved with a sample of 102 units under the estimator based on simple random sampling. In a similar way, when estimating above-ground biomass, a relative efficiency of 0.698 suggests that the *MAE* estimator assisted by the GLM-*CAT* model requires only approximately 71 sampling units (102×0.698≈71) to achieve the same variance under a simple random sampling scheme with 102 units.

Discussion

A procedure was established to identify the operational approach that minimizes waste and maximizes efficiency in terms of reducing variance in the timber and biomass inventory. A key finding was the stability of the population mean estimates per sampling unit across the LIM, GLM, GAM, NL1 through NL4, and GLM-*CAT* models, confirming that model-assisted estimators remain unbiased and consistent with respect to the design, regardless of the model specification (Ståhl et al., 2016).

The integration of data from multiple sources improves the carbon storage estimates, particularly in complex regions (Matiza et al., 2023), while advanced techniques such as neural networks or transformers enable the capture of spatial and spectral dependencies, thereby reducing errors in structural attributes (Tanase et al., 2025). Nevertheless, the robustness observed in this study reinforces the applicability of the *MAE* in operational contexts, where the exact specification of the model is not guaranteed and accuracy depends more on data quality than on the model itself (Ameztegui et al., 2022).

The *GNDVI* proved to be an effective auxiliary variable, given its high sensitivity to chlorophyll content (Gitelson et al., 1996). The observed values, ranging from 0.64 to 0.76, are consistent with healthy vegetation conditions and comparable to those reported in studies that have used this index to detect stress conditions (Zhang et al., 2025).

In terms of efficiency, the GLM-*CAT* model (*GNDVI* plus silvicultural treatment) showed the greatest reduction in variance for volume (37.65 %), while nonlinear models, such as Schumacher's (NL1) and the potential model (NL3), did not show substantial improvements. For above-ground biomass, the maximum reduction, though less significant (30.21 %), was also achieved using GLM-*CAT*, suggesting a weaker relationship with the *GNDVI*. This result confirms that efficiency gains depend more on the strength of the relationship between the auxiliary variables than on the complexity of the model (Ståhl et al., 2016).

Although a Gamma distribution was found to yield better *AIC* and *BIC* values, from a design-based perspective, the choice of distribution affects precision (reduction in variance) but not the exact estimates. The observed reductions in variance, ranging from 30 % to 37 %, are significant when considering a single auxiliary optical variable. However, they are lower than those reported in studies using LiDAR data, which provide greater explanatory power (McRoberts et al., 2006). For example, McRoberts et al. (2013) described reductions of more than 80 % using LiDAR data, while Breidenbach and Astrup (2012) found relative efficiencies ranging from 0.35 to 0.87. Similarly, Zhao et al. (2024) reported relative efficiencies of 0.73 in annual biomass estimates using Sentinel-2 data.

The pseudo-R^2 values of 0.38 for volume and 0.30 for biomass are consistent with previous studies in temperate forests using spectral indices as predictors (Hernández-Ramos et al., 2020; López-Serrano et al., 2021). Overall, the findings confirm that model-assisted estimation improves efficiency without compromising design-based validity. Future research should assess stability in finite samples and explore the integration of multi-source data and advanced techniques, in line with current trends in forest monitoring (Besic et al., 2025; Ståhl et al., 2016).

Conclusions

The comparative analysis confirms that the Model-assisted estimator (*MAE*) outperforms the Horvitz-Thompson estimator (*HTE*) regardless of the model specification used. However, evidence suggests that incorporating a Gamma distribution (using GLM or GAM) or employing specific nonlinear forms (NL4) yields the greatest gains in accuracy because it reduces variance when using a spectral index such as *GNDVI* and silvicultural treatment. These models effectively reduce the sampling error to approximately 6.9 %, which renders them a cost-effective alternative to inventories that rely solely on intensive fieldwork.

Acknowledgments

The authors would like to thank the technical staff of the *Consultoría de Fomento Ambiental y Desarrollo Social de Comunidades Forestales S. C.* (Consultancy on Environmental Promotion and Social Development for Forest Communities) for their assistance in collecting mensuration data, and the *Instituto Nacional de Investigaciones Forestales*, *Agrícolas y Pecuarias*, *INIFAP* (National Institute for

Research on Forest, Agriculture and Livestock) for funding the project "Integrated management of forest resources for the sustainability of ecosystem services in the face of climate change".

Conflict of interest

The authors declare that they have no conflicts of interest.

Contribution by author

Efraín Velasco-Bautista and Martin Enrique Romero-Sánchez: study design and supervision; Efraín Velasco-Bautista, Alma Delia Ortiz-Reyes, and Jesús Valentín Gutiérrez-García: fieldwork and data analysis; Efraín Velasco-Bautista: methodological design and statistical analysis. All authors contributed to the drafting, critical revision, and approval of the final version of the manuscript.

References

Ameztegui, A., Rodrigues, M., & Granda, V. (2022). Uncertainty of biomass stocks in Spanish forests: a comprehensive comparison of allometric equations. *European Journal of Forest Research*, *141*, 395-407. https://doi.org/10.1007/s10342-022-01444-w

Arias-Téllez, A., & García-Martínez, R. (2017). Almacén de carbono en plantaciones de *Pinus patula* y *Pinus ayacahuite* en San Miguel Tenextepec, Amanalco, Estado de México. En V. J. C. Vinay, V. A. Esqueda E., O. H. Tosquy V., A. Ríos U., M. V. Vázquez H. & C. Perdomo M. (Comps.), *Avances en investigación agrícola, pecuaria, forestal, acuícola, pesquera, desarrollo rural, transferencia de tecnología, biotecnología, ambiente, recursos naturales y cambio climático 2017* (pp. 1057-1065). Instituto

Nacional de Investigaciones Forestales, Agrícolas y Pecuarias. https://www.researchgate.net/publication/322448012_Almacen_de_carbono_en_plantaciones_de_Pinus_patula_y_Pinus_ayacahuite_en_San_Miguel_Tenextepec_Amanalco_Estado_de_Mexico

Avendaño-Hernández, D. M., Acosta-Mireles, M., Carrillo-Anzures, F., & Etchevers-Barra, J. D. (2009). Estimación de biomasa y carbono en un bosque de *Abies religiosa*. *Revista Fitotecnia Mexicana*, *32*(3), 233-238. https://revfitotecnia.mx/index.php/RFM/article/view/756

Besic, N., Picard, N., Vega, C., Bontemps, J.-D., Hertzog, L., Renaud, J.-P., Fogel, F., Schwartz, M., Pellissier-Tanon, A., Destouet, G., Mortier, F., Planells-Rodriguez, M., & Ciais, P. (2025). Remote-sensing-based forest canopy height mapping: some models are useful, but might they provide us with even more insights when combined? *Geoscientific Model Development*, *18*(2), 337-362. https://doi.org/10.5194/gmd-18-337-2025

Breidenbach, J., & Astrup, R. (2012). Small area estimation of forest attributes in the Norwegian National Forest Inventory. *European Journal of Forest Research*, *131*, 1255-1267. https://doi.org/10.1007/s10342-012-0596-7

Comisión Nacional Forestal. (2017). *Inventario Nacional Forestal y de Suelos. Procedimientos de muestreo. Versión 19.0*. Comisión Nacional Forestal. https://www.conafor.gob.mx/apoyos/docs/externos/2022/DocumentosMetodologicos/2019/ANEXO_Procedimientos_de_muestreo_2019.pdf

Correa-Díaz, A., Villanueva-Díaz, J., Gutiérrez-García, J. V., Velasco-Bautista, E., Moreno-Sánchez, F., & Zamora-Morales, B. P. (2025). Efecto del clima y el manejo forestal en el crecimiento radial de un bosque de coníferas en Puebla, México. *Madera y Bosques*, *31*, Artículo e312717. https://doi.org/10.21829/myb.2025.312717

Dettmann, G. T., MacFarlane, D. W., Radtke, P. J., Weiskittel, A. R., Affleck, D. L. R., Poudel, K. P., & Westfall, J. (2022). Testing a generalized leaf mass estimation method for diverse tree species and climates of the continental United States. *Ecological Applications*, *32*(7), Article e2646. https://doi.org/10.1002/eap.2646

Dhiman, V., & Kumar, A. (2025). Species exhibiting positive association demonstrate high above-ground biomass accumulation in the subtropical Himalayan forest

ecosystem, India. *Applied Ecology and Environmental Research*, *23*(1), 1433-1452. https://doi.org/10.15666/aeer/2301_14331452

Díaz-Ríos, M. de J., Vázquez-Alarcón, A., Uribe-Gómez, M., Sánchez-Vélez, A., Lara-Bueno, A., & Cruz-León, A. (2016). Ecuaciones alométricas para estimar biomasa y carbono en aile obtenidas mediante un método no destructivo. *Revista Mexicana de Ciencias Agrícolas*, (Pub. Esp. 16)*,* 3235-3249. https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-09342016001203235

Frank, B., & Monleon, V. J. (2021). Comparison of variance estimators for systematic environmental sample surveys: Considerations for post-stratified estimation. *Forests*, *12*, Article 772. https://doi.org/10.3390/f12060772

Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*, *58*(3), 289-298. https://doi.org/10.1016/S0034-4257(96)00072-7

González-Rosales, A., & Ortiz-Paniagua, C. F. (2022). Superficie forestal afectada por incendios en México: Apuntes iniciales hacia un modelo de manejo preventivo. *Revista de Ciencias Ambientales*, *56*(1), 1-27. https://www.scielo.sa.cr/scielo.php?pid=S2215-38962022000100001&script=sci_abstract&tlng=es

Hernández-Ramos, J., García-Cuevas, X., Pérez-Miranda, R., González-Hernández, A., & Martínez-Ángel, L. (2020). Inventario y mapeo de variables forestales mediante sensores remotos en el estado de Quintana Roo, México. *Madera y Bosques*, *26*(1), Artículo e2611884. https://myb.ojs.inecol.mx/index.php/myb/article/view/e2611884

Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, *47*(260), 663-685. https://doi.org/10.1080/01621459.1952.10483446

Hu, Y., & Sun, Z. (2022). Assessing the capacities of different remote sensors in estimating forest stock volume based on high-precision sample plot positioning and the random forest method. *Nature Environment and Pollution Technology*, *21*(3), 1113-1123. https://doi.org/10.46488/nept.2022.v21i03.016

Khan, M. N., Tan, Y., Gul, A. A., Abbas, S., & Wang, J. (2024). Forest aboveground biomass estimation and inventory: evaluating remote sensing-based approaches. *Forests*, *15*, Article 1055. https://doi.org/10.3390/f15061055

López-Serrano, P. M., Vega-Nieva, D. J., Ramírez-Aldaba, H., García-Montiel, E., & Corral-Rivas, J. J. (2021). Estimación de parámetros forestales mediante datos de *Sentinel* 2A en Pueblo Nuevo, Durango. *Revista Mexicana de Ciencias Forestales*, *12*(68), 81-106. https://doi.org/10.29298/rmcf.v12i68.1075

Matiza, C., Mutanga, O., Peerbhay, K., Odindi, J., & Lottering, R. (2023). A systematic review of remote sensing and machine learning approaches for accurate carbon storage estimation in natural forests. *Southern Forests: A Journal of Forest Science*, *85*(3-4), 123-141. https://doi.org/10.2989/20702620.2023.2251946

McConville, K. S., Moisen, G. G., & Frescino, T. S. (2020). A tutorial on model-assisted estimation with application to forest inventory. *Forests*, *11*(2), Article 244. https://doi.org/10.3390/f11020244

McRoberts, R. E., Holden, G. R., Nelson, M. D., Liknes, G. C., & Gormanson, D. D. (2006). Using satellite imagery as ancillary data for increasing the precision of estimates for the Forest Inventory and Analysis program of the USDA Forest Service. *Canadian Journal of Forest Research*, *36*, 2968-2980. https://doi.org/10.1139/x05-222

McRoberts, R. E., Næsset, E., & Gobakken, T. (2013). Inference for lidar-assisted estimation of forest growing stock volume. *Remote Sensing of Environment*, *128*, 268-275. https://doi.org/10.1016/j.rse.2012.10.007

R Core Team. (2025). *R: A language and environment for statistical computing* (Version 4.5.0) [Computer software]. R Foundation for Statistical Computing. https://www.r-project.org/

SAS Institute Inc. (2024). *SAS/STAT® User's guide: The GLIMMIX Procedure 2024.04-2024.09**. SAS Institute Inc. https://documentation.sas.com/api/collections/pgmsascdc/v_052/docsets/statug/content/glimmix.pdf

Soriano-Luna, M. de los Á., Ángeles-Pérez, G., Martínez-Trinidad, T., Plascencia-Escalante, F. O., & Razo-Zárate, R. (2015). Estimación de biomasa aérea por

componente estructural en Zacualtipán, Hidalgo, México. *Agrociencia*, *49*(4), 423-438. https://www.agrociencia-colpos.org/index.php/agrociencia/article/view/1156

Ståhl, G., Saarela, S., Schnell, S., Holm, S., Breidenbach, J., Healey, S. P., Patterson, P. L., Magnussen, S., Næsset, E., McRoberts, R. E., & Grégoire, T. G. (2016). Use of models in large-area forest surveys: Comparing model-assisted, model-based and hybrid estimation. *Forest Ecosystems*, *3*, Article 5. https://doi.org/10.1186/s40663-016-0064-9

Tanase, M., Martini, J. P., Miranda, P., García-García, D., Wilke, V., Diez, J., Natal, S., & San Martín, D. (2025). Estimación de variables forestales a partir de sensores Lidar y ópticos e inteligencia artificial. *Cuadernos de Investigación Geográfica*, *51*(2), 121-132. https://doi.org/10.18172/cig.6767

Valbuena, R., Hernando, A., Manzanera, J. A., Görgens, E. B., Almeida, D. R. A., Mauro, F., García-Abril, A., & Coomes, D. A. (2017). Enhancing of accuracy assessment for forest above-ground biomass estimates obtained from remote sensing via hypothesis testing and overfitting evaluation. *Ecological Modelling*, *366*, 15-26. https://doi.org/10.1016/j.ecolmodel.2017.10.009

Zadbagher, E., Marangoz, A. M., & Becek, K. (2024). Estimation of above-ground biomass using machine learning approaches with InSAR and LiDAR data in tropical peat swamp forest of Brunei Darussalam. *iForest*, *17*(3), 172-179. https://doi.org/10.3832/ifor4434-017

Zhang, M., Zhu, J., Song, L., Qi, K., Zheng, X., Zhang, X., Ge, J., & Tian, H. (2025). How to characterize the decline of natural *Pinus sylvestris* var. *mongolica* forests on sandy land? *Global Ecology and Conservation*, *64*, Article e03955. https://doi.org/10.1016/j.gecco.2025.e03955

Zhao, A., Cheng, X., Cao, R., Huang, L., & Hou, Z. (2024). Continuous monitoring of forests in wetland ecosystems with remote sensing and probability sampling. *Remote Sensing*, *16*(18), Article 3508. https://doi.org/10.3390/rs16183508



Todos los textos publicados por la **Revista Mexicana de Ciencias Forestales –**sin excepción– se distribuyen amparados bajo la licencia *Creative Commons 4.0* Atribución-No Comercial (CC BY-NC 4.0 Internacional), que permite a terceros utilizar lo publicado siempre que mencionen la autoría del trabajo y a la primera publicación en esta revista.