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

DOI: 10.29298/rmcf.v17i96.1646

Artículo de investigación

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

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

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

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.

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.

Introducción

La estimación confiable de atributos forestales como el volumen de madera y la biomasa aérea es fundamental para el manejo sostenible, la contabilidad de carbono y el monitoreo de ecosistemas (Hu & Sun, 2022). La precisión en estas estimaciones permite evaluar existencias de carbono y productividad, así como diseñar estrategias orientadas a la mitigación del cambio climático (Zadbagher et al., 2024).

Tradicionalmente, en los inventarios forestales se utiliza muestreo probabilístico sistemático con inicio aleatorio y estimadores basados en el diseño, como el Estimador *Horvitz-Thompson* (*HTE*, por sus siglas en inglés) (Horvitz & Thompson, 1952), que garantiza ausencia de sesgo bajo probabilidades de inclusión apropiadas. Sin embargo, en diseños sistemáticos o cuasi-sistemáticos, comunes en inventarios operativos como el programa FIA de Estados Unidos de América, la estimación de la varianza resulta intratable o inestable, por lo que con frecuencia se usan aproximaciones bajo muestreo aleatorio simple.

Aunque el Estimador *Horvitz-Thompson* presenta propiedades teóricas favorables, puede exhibir alta varianza en poblaciones heterogéneas (Ståhl et al., 2016). Para mitigar esta limitación, la estimación asistida por modelos incorpora información auxiliar disponible para toda la población, combinando predicciones con residuos ponderados (McConville et al., 2020), lo que produce estimadores insesgados respecto al diseño y potencialmente más eficientes cuando la variable auxiliar es predictiva (McRoberts et al., 2006).

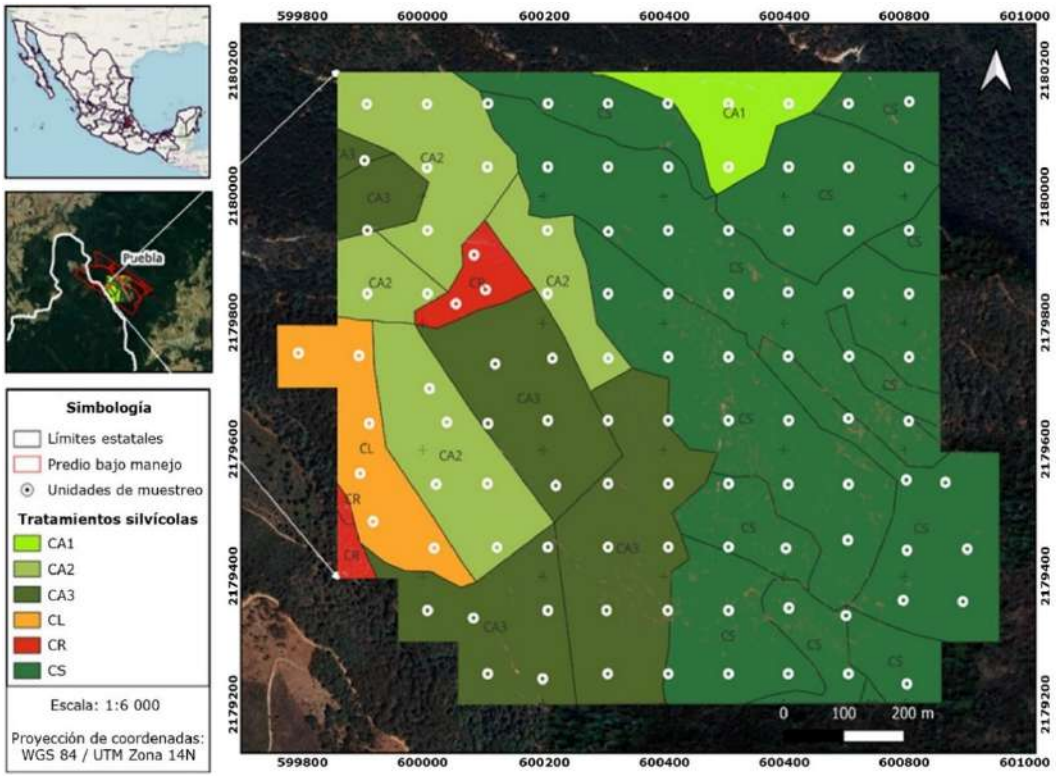
En este contexto, la disponibilidad de imágenes de *Sentinel-2* ha impulsado la estimación de variables forestales a gran escala (McRoberts et al., 2006). Índices de vegetación, en particular del borde rojo e infrarrojo cercano muestran alta correlación con atributos estructurales, lo cual permite el uso de modelos como regresión lineal, modelos aditivos y técnicas de aprendizaje automático (Khan et al., 2024).

No obstante, en México, particularmente en predios forestales bajo manejo, existe limitada evidencia sobre cómo la elección del modelo afecta la eficiencia del Estimador Asistido por Modelos (*MAE*, por sus siglas en inglés) desde una perspectiva de diseño, en especial cuando se dispone de una sola variable auxiliar. Por ello, en el presente estudio se evaluó el desempeño de distintos modelos en la estimación de volumen y biomasa bajo un enfoque *MAE*, con énfasis en la reducción de varianza, respecto al Estimador *Horvitz-Thompson* en un contexto de muestreo cuasi-sistemático.

Materiales y Métodos

Área de estudio

El estudio se realizó en un bosque bajo manejo ubicado en el ejido Emiliano Zapata, municipio Chignahuapan, Puebla, México; el cual se caracteriza por la presencia de pinos (*Pinus patula* Schiede ex Schjtdl *et* Cham., *Pinus ayacahuite* Ehrenb. *ex* Schltdl., *Pinus teocote* Schied. *ex* Schltdl. & Cham., *Pinus greggii* Engelm. *ex* Parl., *Pinus pseudostrobus* Lindl.), oyamel (*Abies religiosa* (Kunth) Schltdl. & Cham.), encinos (*Quercus rugosa* Née, *Quercus crassifolia* Bonpl., *Quercus laurina* Bonpl.) y otras latifoliadas (*Alnus acuminata* Kunth, *Arbutus xalapensis* Kunth) en menor proporción. Los datos de campo se recolectaron a finales de 2023 en un área forestal con manejo mediante el Método de Desarrollo Silvícola (MDS) y el Método Mexicano de Ordenación de Bosques Irregulares (MMOBI). El MDS incluyó los tratamientos silvícolas: cortas de aclareo (CA), corta de liberación (CL) y corta de regeneración (CR) intervenidos en el 2016, 2014 y 2014, respectivamente; en el MMOBI se consideró la corta de selección (CS) implementada en el 2016. El área total muestreada abarcó aproximadamente 102 ha (Figura 1).



Cada círculo representa una unidad de muestreo circular de 1 000 m^2.

Figura 1. Diseño de muestreo cuasi-sistemático en el área de estudio.

Se establecieron 102 unidades de muestreo circulares, siguiendo un diseño de muestreo cuasi-sistemático, con una separación teórica de 100 m entre parcelas (Figura 1). Cada unidad de muestreo fue de 1 000 m^2. Con base en la metodología definida para el Inventario Nacional Forestal y de Suelos (Comisión Nacional Forestal [Conafor], 2017), dentro de cada parcela se midieron todos los árboles con un diámetro a la altura del pecho (*DAP*) mayor a 7.5 cm, con forcípula de aluminio 95 cm marca *Haglöf*® modelo Mantax Blue y cinta diamétrica *Forestry Suppliers*® modelo 283D; la altura total (m) se midió con clinométro *Suunto*® modelo PM-5a. Además, se registró la especie correspondiente a los individuos medidos.

Estimación de volumen maderable y biomasa en campo

El volumen de madera y la biomasa aérea a nivel de árbol se calcularon con ecuaciones alométricas específicas por especie, previamente desarrolladas para la región de estudio. Con base en las especies identificadas en campo, las ecuaciones consultadas fueron: Avendaño-Hernández et al. (2009), Soriano-Luna et al. (2015), Díaz-Ríos et al. (2016), Arias-Téllez y García-Martínez (2017) y Correa-Díaz et al., (2025).

El volumen y la biomasa total por parcela se obtuvieron mediante la agregación de las predicciones individuales de los árboles de todas las especies encontradas en cada unidad de muestreo. A partir de los totales en cada unidad de muestreo, se aplicaron los estimadores *HTE* y *MAE*, para obtener la estimación del volumen medio de madera y la biomasa aérea media por unidad de muestreo (parcela circular de 1 000 m^2).

Información auxiliar derivada de *Sentinel-2*

La información auxiliar se obtuvo a partir de imágenes multiespectrales de *Sentinel-2A* (octubre-noviembre de 2023) obtenidas en *Google Earth Engine*. Se seleccionaron escenas con cobertura nubosa <5 % y se aplicó enmascaramiento de nubes y sombras (Khan et al., 2024).

Para reducir la variabilidad temporal, se generó un compuesto mediano a partir de la colección filtrada. Sobre este se calcularon cinco índices de vegetación ampliamente utilizados: *NDVI* (Índice de Vegetación de Diferencias Normalizadas), *GNDVI* (Índice Verde de Vegetación de Diferencias Normalizadas), *EVI* (Índice de Vegetación Mejorado), *MSAVI* (Índice de Vegetación Ajustado al Suelo Modificado) y *SIPI* (Índice

de Pigmentación Insensible a la Estructura) (Besic et al., 2025). Los valores medios por parcela, obtenidos a partir de intersección espacial, se utilizaron como variables auxiliares en los modelos.

Selección de la variable auxiliar

Se realizaron análisis exploratorios para evaluar la relación entre los atributos forestales y los índices espectrales mediante diagramas de dispersión e histogramas. Debido a la ligera no linealidad y fuerte asimetría en volumen y biomasa, se aplicaron coeficientes de correlación de intervalos de *Spearman* para evaluar asociaciones monotónicas (Dhiman & Kumar, 2025).

La alta correlación entre los índices evidenció multicolinealidad, por lo que se aplicaron técnicas de regularización (*LASSO* y *Elastic Net*, Criterio de selección validación cruzada) para seleccionar una variable auxiliar parsimoniosa (Valbuena et al., 2017). Ambos métodos identificaron consistentemente el *GNDVI* como el predictor más informativo.

Análisis comparativo de modelos

En el análisis se utilizó como base el Estimador *Horvitz-Thompson* (*HTE*) en su versión simplificada, que consiste en usar los estimadores del muestreo aleatorio simple, aun cuando los datos se recolectaron de manera cuasi-sistemática (Frank & Monleon, 2021). Para fines comparativos, respecto al *HTE*, se utilizó el Estimador Asistido por

Modelos (*MAE*) (Ståhl et al., 2016) que en su término de predicción consideró en un inicio dos modelos lineales generalizados: un modelo aditivo generalizado y cuatro modelos explícitamente no lineales (Cuadro 1).

Cuadro 1. Modelos utilizados (*MAE*) en el análisis comparativo.

Tipo de modelo	Estructura del modelo
Lineal generalizado con distribución normal (LIM)	Distribución: $y_i \sim Normal(\eta_i, \sigma^2)$, $E(y_i)=\mu_i$, $var(y_i)=\sigma^2$ Predictor lineal: $\eta_i=\beta_0+\beta_1 x_i$ Función de enlace: $\eta_i=\mu_i$, enlace identidad
Lineal generalizado con distribución Gamma (GLM)	Distribución: $y_i \sim Gamma(a_i,\beta_i)$; $E(y_i)= a_i\beta_i=\mu_i$, $var(y_i)=a_i\beta_i^2=\frac{\mu_i^2}{a_i}$ Predictor lineal: $\eta_i=\beta_0+\beta_1 x_i$ Función de enlace: $\eta_i= log(\mu_i)$, enlace identidad
Aditivo generalizado con distribución Gamma (GAM)	Distribución: $y_i \sim Gamma(a_i,\beta_i)$; $E(y_i)= a_i\beta_i=\mu_i$, $var(y_i)=a_i\beta_i^2=\frac{\mu_i^2}{a_i}$ Predictor Lineal: $\eta_i=\beta_0+f_1(x_i)$ Función de enlace: $\eta_i= log(\mu_i)$, enlace identidad
Schumacher (NL1)	$y_i=e^{\beta_0+\beta_1/x_i}+\varepsilon_i$, $\varepsilon_i \sim N(0,\sigma^2)$
Exponencial (NL2)	$y_i=e^{\beta_0+\beta_1 x_i}+\varepsilon_i$, $\varepsilon_i \sim N(0,\sigma^2)$
Potencial (NL3)	$y_i=\beta_0 x_i^{\beta_1}+\varepsilon_i$, $\varepsilon_i \sim N(0,\sigma^2)$
Michaelis-Menten invertido (NL4)	$y_i=\frac{\beta_0}{\beta_1+x_i}+\varepsilon_i$, $\varepsilon_i \sim N(0,\sigma^2)$

$a_i..n$; $\beta_{i...n0}, \beta_1$ = Coeficientes de regresión; μ_i = Media poblacional; σ^2 = Varianza poblacional; ε_i = Error aleatorio; y = Volumen o biomasa; x = *GNDVI*; x y y están dados por unidad de muestreo.

Con base en la dispersión observada en las relaciones volumen-*GNDVI* y biomasa-*GNDVI*, se evaluaron siete modelos con el propósito de analizar, en comparación con el *HTE*, para la sensibilidad y el comportamiento de los estimadores asistidos por modelos con diferentes estructuras funcionales, se utilizó una única variable auxiliar continua: el *GNDVI*. En este contexto, *Y* representa la variable respuesta (volumen o

biomasa por unidad de muestreo), mientras que *X* corresponde a la variable predictora (*GNDVI*) (Cuadro 1).

Adicionalmente, se consideró un modelo generalizado (Ståhl et al., 2016) que incluyó además del *GNDVI*, al tratamiento silvícola (*CAT*) como variable categórica.

Evaluación de ajuste del modelo

Al tratarse de modelos lineales generalizados y modelos no lineales, en conjunto, como criterios de ajuste se utilizaron la Raíz del Cuadrado Medio del Error (*RCME*), el Error Absoluto Medio (*EAM*), el Error Residual Medio (*ERM*) y la pseudo-R^2, así como el Criterio de Información de *Akaike* (*AIC*) y el Criterio de Información *Bayesiano* (*BIC*) (González-Rosales & Ortiz-Paniagua, 2022).

Estas métricas no se utilizaron como criterio único de selección, sino como herramientas para caracterizar el comportamiento de los residuos, dado que en el marco de la estimación asistida por modelos, la varianza del estimador depende directamente de la variabilidad residual bajo el diseño muestral (Ståhl et al., 2016).

Evaluación del desempeño de *MAE* respecto a *HTE*

Mediante los estimadores *HTE* y *MAE*, definidos matemáticamente por McConville et al. (2020), se calcularon las medias de volumen y biomasa por unidad de muestreo, junto con sus respectivas varianzas. El *HTE* utiliza únicamente datos de campo, mientras que el *MAE* estima la media poblacional combinando la media de las

predicciones del modelo para todas las unidades (*N*=1 089) con un ajuste basado en el promedio de los residuales. El desempeño del *MAE* se evaluó respecto al *HTE* mediante los siguientes criterios (Dettmann et al., 2022):

Eficiencia relativa (*ER*): $ER = \frac{\widehat{Var}(\hat{a}_{MAE})}{\widehat{Var}(\hat{a}_{HTE})}$ (1)

Reducción de la varianza (*RV*): $RV = (1 - ER)100$ (2)

Error estándar (*EE*) del estimador: $EE = \sqrt{\widehat{Var}(\hat{a})}$ (3)

Error relativo de muestreo (*ERM*): $ERM = \frac{Z_{\propto/2}\, EE}{\hat{a}} \times 100$ (4)

Donde:

$\hat{a}$ = Estimación del parámetro poblacional

$Z_{\propto/2}$ = Valor crítico de la distribución normal para el nivel de confianza

$\widehat{Var}$ = Varianza

La selección de variables y el análisis exploratorio se realizaron en *SAS* versión 9.4 (SAS Institute Inc., 2024). El ajuste de modelos, así como el cálculo de estimadores asistidos por modelos y sus varianzas, se realizaron en *R* versión 4.5.0 (R Core Team, 2025), con los paquetes *mgcv* y *nls2*.

Resultados

El inventario con 102 unidades de muestreo reveló una media muestral de 31.18 m^3 para el volumen de madera y 15 711 kg para la biomasa aérea (*AGB*), ambos por unidad de muestreo (línea vertical verde de la Figura 2). Las dos variables presentaron asimetría positiva (coeficientes de asimetría de 0.78 para el volumen y 0.82 para la biomasa). Se identificó que la distribución Gamma, perteneciente a la familia exponencial, es de rápida convergencia y representa el mejor ajuste teórico para el volumen de madera (*AIC*=819.85 y *BIC*=825.10) y resulta ser una candidata idónea para la biomasa (*AIC*=2 056 y *BIC*=2 061) (Figura 2).

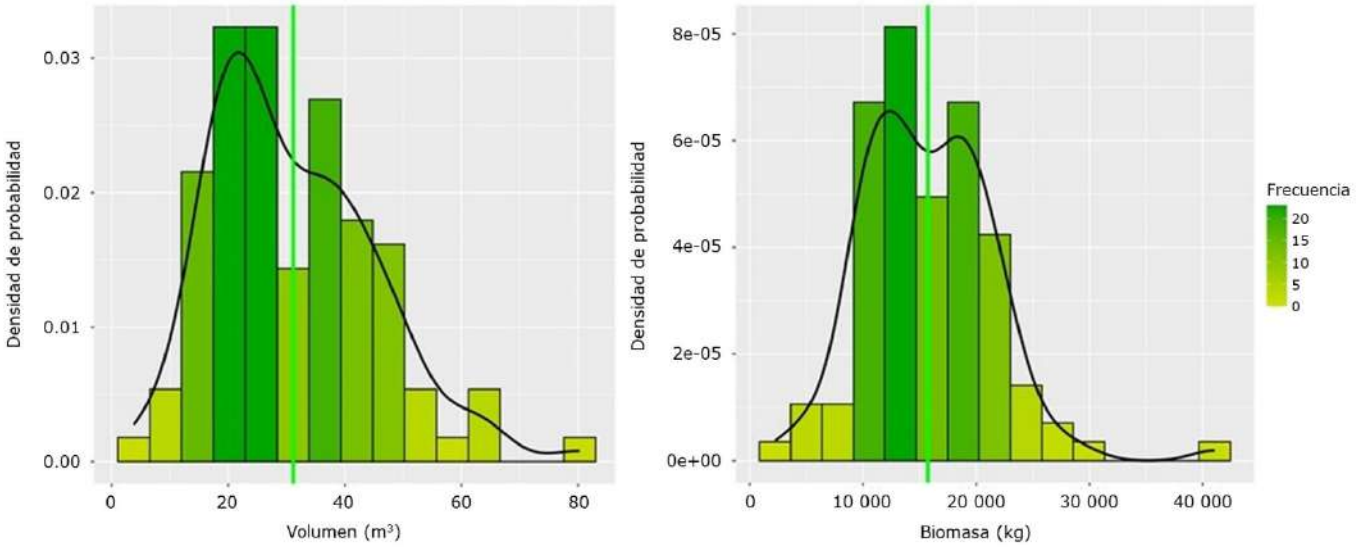


Figura 2. Distribución de las variables volumen (m^3) y biomasa (kg) por unidad de muestreo.

Selección de variables auxiliares

El análisis de correlación mostró relaciones significativas entre los atributos forestales y todos los índices espectrales de *Sentinel-2*. El volumen de madera presentó una correlación más fuerte con los datos espectrales (*promedio*=-0.51) que la biomasa (*promedio*=-0.42), lo cual indica que pueden ser útiles para la predicción de estas variables forestales a nivel de unidad de muestreo. Los procedimientos de regresión *LASSO* y *Elastic Net* identificaron al *GNDVI* como la variable auxiliar con mayor poder predictivo, y por tanto de mayor impacto para ambos parámetros (lambda mínima de validación cruzada para volumen fue de 0.35 (*log lambda*=-1.06) correspondiente a *LASSO* y de 1.46 (*log lambda*=0.38) para *Elastic Net*. Mientras que para biomasa fue de 341.13 (*log lambda*=5.84) correspondiente a *LASSO* y de 682.26 (*log lambda*=6.25) para *Elastic Net*. En este sentido, la correlación de *Spearman* entre volumen y *GNDVI* fue de -0.52 y, entre biomasa y *GNDVI* de -0.45.

Ajuste comparativo de modelos para volumen de madera y biomasa aérea

Excepto para la pseudo-R^2, los siete modelos evaluados mostraron desempeños poco diferenciados en el ajuste para la predicción del volumen de madera (Cuadro 2).

Cuadro 2. Valores de ajuste de los modelos evaluados para predecir volumen maderable.

Modelo		*RCME*	*EAM*	*ERM*	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

RCME = Raíz del cuadrado medio del error; *EAM* = Error absoluto medio; *ERM* = Error residual medio; *AIC* = Criterio de información de *Akaike*; *BIC* = Criterio de información *Bayesiano*. Modelos: LIM = Lineal generalizado con distribución normal; GLM = Lineal generalizado con distribución Gamma; GAM = Aditivo generalizado con distribución Gamma; NL1 = *Schumacher*; NL2 = Exponencial; NL3 = Potencial; NL4 = *Michaelis-Menten* invertido. *Coeficientes significativos; **Coeficientes altamente significativos.

Aun cuando se reflejan ciertas similitudes en los criterios de ajuste, el modelo no lineal NL4 (*Michaelis–Menten* invertido) presentó relativamente la menor *RCME* (11.490) y el mayor pseudo-R^2 (0.321). Asimismo, los modelos GLM y GAM, que incorporaron la distribución Gamma, registraron los valores de *AIC* más bajos (789.9 y 790.0, respectivamente), lo que indica un mejor ajuste en comparación con el modelo lineal (LIM), que presentó el *AIC* más alto (798.4). Este comportamiento también se observó en el *BIC*.

Por lo anterior, para el caso de volumen todos los modelos produjeron predicciones puntuales muy similares entre sí a lo largo del espectro de *GNDVI* y cercanas a la media observada que se obtuvo mediante el Estimador *Horvitz-Thompson* (*HTE*) (Figura 3), la cual fue de 31.18 m^3.

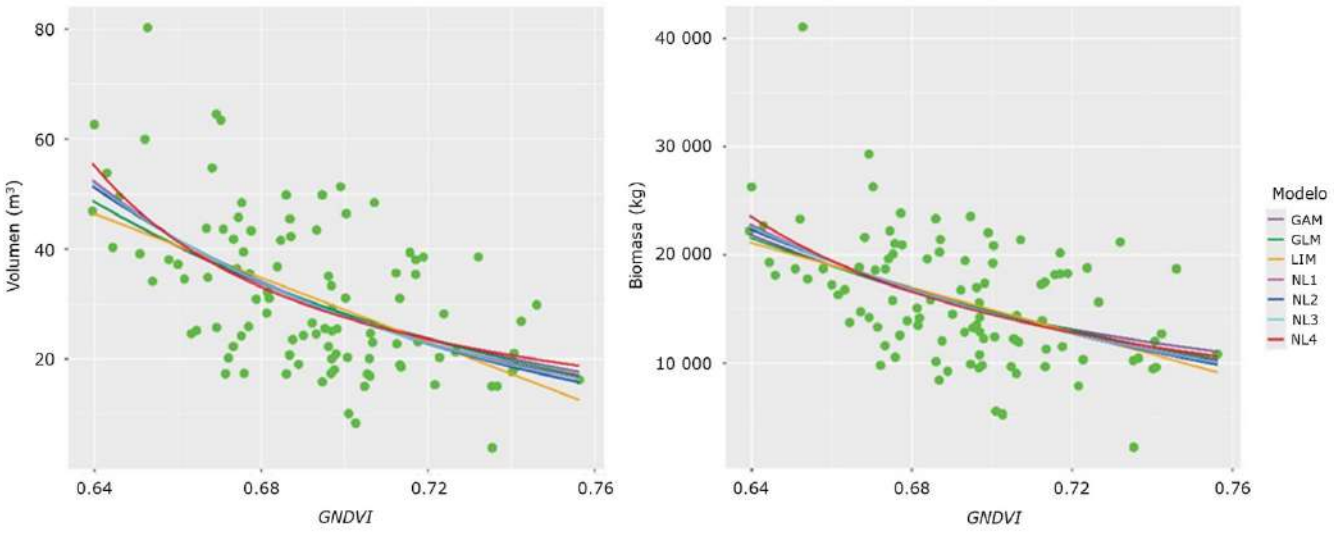


Figura 3. Dispersión del volumen *versus GNDVI* (izquierda) y biomasa aérea *versus GNDVI* (derecha) y curvas de los modelos evaluados.

En la predicción de biomasa aérea, los comportamientos fueron consistentes con las tendencias observadas en volumen; es decir, valores de ajuste poco diferenciados. En este caso, los *AIC* y *BIC* relativamente mayores correspondieron a GLM y GAM (Cuadro 3).

Cuadro 3. Valores de ajuste de los modelos evaluados para predecir biomasa aérea.

	Modelo	*RCME*	*EAM*	*ERM*	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

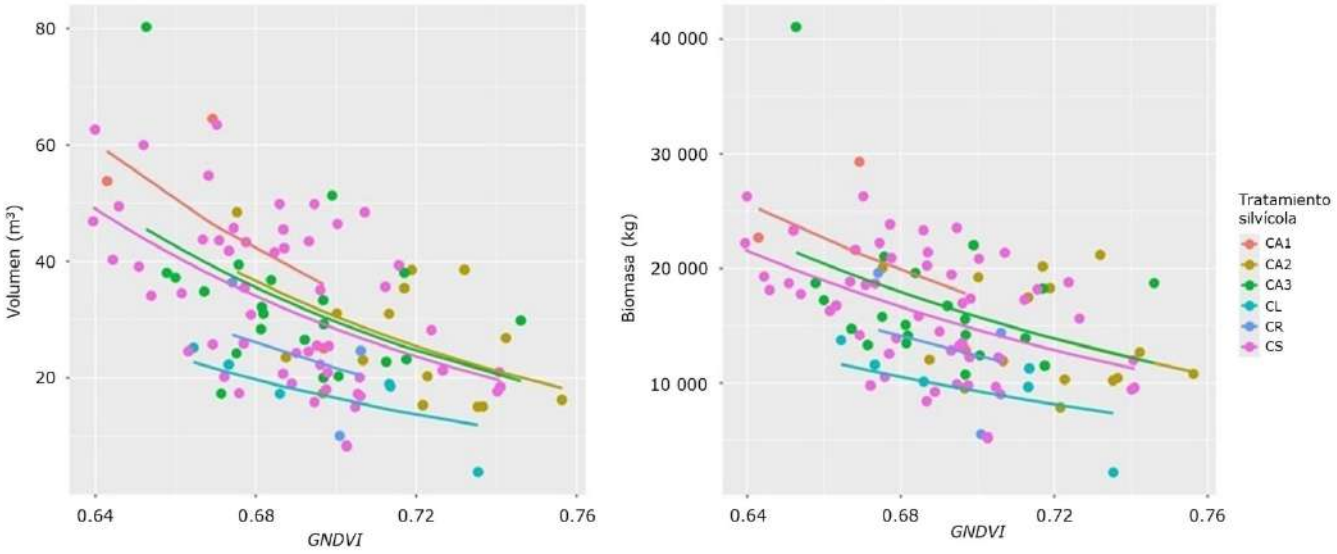
RCME = Raíz del cuadrado medio del error; *EAM* = Error absoluto medio; *ERM* = Error residual medio; *AIC* = Criterio de información de *Akaike*; *BIC* = Criterio de información *Bayesiano*. Modelos: LIM = Lineal generalizado con distribución normal; GLM = Lineal generalizado con distribución Gamma; GAM = Aditivo generalizado con distribución Gamma; NL1 = *Schumacher*; NL2 = Exponencial; NL3 = Potencial; NL4 = *Michaelis-Menten* invertido. **Coeficientes altamente significativos.

En este caso el modelo NL4 presentó el mejor desempeño (*RCME*=4 984.17), con *AIC* y *BIC* similares a los demás. Los modelos lineales y generalizados tuvieron menor variabilidad (Coeficiente de variación [*CV*]≈16 %) que los no lineales (*CV*≈18 %), con diferencias en los extremos.

Dado el comportamiento similar en *MAE*, se incorporó el tratamiento silvícola al GLM junto con *GNDVI*, debido a su alta variabilidad interna (*CV* hasta 55.7 %). Esto mejoró el ajuste (*RCME*=11.011, pseudo-R^2=0.377, *AIC*=786.361). Ambos

predictores fueron significativos ($p<0.05$), generando modelos diferenciados por tratamiento en función de *GNDVI*.

La biomasa aérea se modeló con un GLM usando tratamiento silvícola y *GNDVI*, dada la alta variabilidad entre tratamientos (*CV* hasta 54 %). El modelo mejoró el ajuste (*RCME*=4 770.92; pseudo-R^2=0.30; *AIC*=2 032.5), ambos predictores resultaron significativos ($p<0.05$). El modelo incluyó variables indicadoras por tratamiento, permitiendo estimaciones diferenciadas en función del *GNDVI* (Figura 4).



CA1 = Corta de aclareo 1; CA2 = Corta de aclareo 2; CA3 = Corta de aclareo 3; CL = Corta de liberación; CR = Corta de regeneración; CS = Corta de selección.

Figura 4. Volumen (izquierda) y biomasa (derecha) estimados con GLM que incluye *GNDVI* y tratamiento silvícola.

Eficiencia y ganancias de precisión (*MAE vs. HTE*)

En conjunto, el *MAE* superó al *HTE* en todas las estructuras de modelo evaluadas. La *ER* promedio, al considerar los siete modelos de *MAE*, respecto a *HTE*, fue de 0.6897; es decir, en promedio, independiente del modelo evaluado, la varianza de *MAE* fue de 68.97 % la varianza de *HTE*, lo cual condujo a una reducción promedio de la varianza del volumen por *MAE*, respecto a *HTE*, de 31.03 %. El modelo NL4 mostró una reducción de varianza 32.17 %, ligeramente mayor que el resto de los modelos, lo que indicó una disminución del error relativo de muestreo de 6.93 %. Las medias puntuales de volumen para *MAE* y *HTE* fuerron prácticamente las mismas, 31 m^3 por parcela que equivale a 310 m^3 por ha (Cuadro 4).

Cuadro 4. Estimaciones de volumen maderable por unidad de muestreo y valores de la varianza por modelo evaluado.

Estimador	Modelo	Media	Varianza	*ER*	*RV* (%)	*EE*	*ERM* (%)
HTE	Testigo	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

ER = Eficiencia relativa; *RV* = Reducción de la varianza; *EE* = Error estándar; *ERM* = Error relativo de muestreo; *HTE* = Estimador *Horvitz-Thompson*; *MAE* = Estimador asistido por modelos.

Para el caso de la biomasa aérea, la eficiencia relativa de *MAE*, en relación con *HTE*, fue de 0.7706, lo que originó una reducción promedio de varianza de 22.94 %, al

considerar los siete modelos de trabajo en conjunto. El modelo NL4 alcanzó una ganancia de precisión de 23.84 %, ligeramente mayor que el resto de los modelos, lo que resultó en un error relativo de muestreo de 5.95 % (Cuadro 5).

Cuadro 5. Estimaciones de biomasa por unidad de muestreo y valores de la varianza por modelo evaluado.

Estimador	Modelo	Media	Varianza	*ER*	*RV* (%)	*EE*	*ERM* (%)
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

ER = Eficiencia relativa; *RV* = Reducción de la varianza; *EE* = Error estándar; *ERM* = Error relativo de muestreo; *HTE* = Estimador *Horvitz-Thompson*; *MAE* = Estimador asistido por modelos.

Aunque las estimaciones promedio por parcela (1 000 m^2) fueron similares (15 711.14 kg para *HTE* y 15 560.87 kg para *MAE*), la precisión del *MAE* varió sustancialmente entre los siete modelos evaluados. En el enfoque asistido por modelos, la reducción de la varianza depende de la capacidad explicativa del modelo; en ausencia de poder predictivo, el *MAE* converge en eficiencia al *HTE*.

El modelo generalizado que incorpora *GNDVI* y tratamiento silvícola presentó la mayor eficiencia relativa (62.4 % para volumen y 69.8 % para biomasa), lo que equivale a reducciones de la varianza de 37.65 % y 30.21 %, respectivamente, así como errores de muestreo de 6.6 % y 5.7 % (Cuadro 6). El error estándar confirmó su superioridad bajo las condiciones evaluadas. Aunque los modelos basados únicamente en *GNDVI* mostraron desempeños similares, la inclusión conjunta de variables cuantitativas y

cualitativas produjo mejoras sustanciales respecto al estimador basado exclusivamente en datos de campo.

Cuadro 6. Eficiencia relativa y reducción de varianza correspondientes al modelo GLM que considera *GNDVI* y tratamiento silvícola (*CAT*).

Variable	**Media**	**Varianza**	***ER***	***RV* (%)**	***EE***	***ERM* (%)**
Volumen	30.950	1.088	0.624	37.649	1.043	6.606
Biomasa	15 647.160	204 255.187	0.698	30.213	451.946	5.662

ER = Eficiencia relativa; *RV* = Reducción de la varianza; *EE* = Error estándar; *ERM* = Error relativo de muestreo.

Por otra parte, una eficiencia relativa de 0.624 indica que utilizando el estimador asistido por el modelo GLM-*CAT*, sería suficiente una muestra de aproximadamente 64 unidades de muestreo (102×0.624≈64) para obtener una varianza en la estimación del volumen maderable equivalente a la alcanzada con una muestra de 102 unidades bajo el estimador basado en muestreo aleatorio simple. De manera análoga, en la estimación de biomasa aérea, una eficiencia relativa de 0.698 sugiere que el estimador *MAE* asistido mediante el modelo GLM-*CAT* requiere únicamente cerca de 71 unidades de muestreo (102×0.698≈71) para alcanzar la misma varianza obtenida bajo un esquema de muestreo aleatorio simple con 102 unidades.

Discusión

Se estableció un procedimiento para identificar la forma funcional que minimiza los residuos y maximiza la eficiencia en términos de reducción de varianza del inventario de madera y biomasa. Un hallazgo central fue la estabilidad de las estimaciones de la media poblacional por unidad de muestreo entre los modelos LIM, GLM, GAM, NL1 a NL4 y GLM-*CAT*, lo que confirma que los estimadores asistidos por modelos mantienen la falta de sesgo y consistencia respecto al diseño, con independencia de la especificación del modelo (Ståhl et al., 2016).

La integración de datos procedentes de multifuentes mejora la estimación del almacenamiento de carbono, especialmente en regiones complejas (Matiza et al., 2023), mientras que técnicas avanzadas como redes neuronales o transformadores permiten capturar dependencias espaciales y espectrales, reduciendo errores en atributos estructurales (Tanase et al., 2025). No obstante, la robustez observada en este estudio refuerza la aplicabilidad del *MAE* en contextos operativos, en los que la especificación exacta del modelo no está garantizada y la precisión depende más de la calidad de los datos que del modelo en sí (Ameztegui et al., 2022).

El *GNDVI* demostró ser una variable auxiliar eficaz, dada su alta sensibilidad al contenido de clorofila (Gitelson et al., 1996). Los valores observados de 0.64 a 0.76 son consistentes con condiciones de vegetación saludable y comparables con estudios que han utilizado este índice para detectar condiciones de estrés (Zhang et al., 2025).

En términos de eficiencia, el modelo GLM-CAT (*GNDVI* más tratamiento silvícola) presentó la mayor reducción de varianza para volumen (37.65 %), mientras que los modelos no lineales, como *Schumacher* (NL1) y el modelo potencial (NL3), no mostraron mejoras sustanciales. Para biomasa aérea, la reducción máxima también se obtuvo con GLM-*CAT*, aunque en menor magnitud (30.21 %), lo que sugiere una relación más débil con el *GNDVI*. Este resultado confirma que las ganancias en

eficiencia dependen más de la fuerza de la relación entre variables auxiliares que de la complejidad del modelo (Ståhl et al., 2016).

Aunque se identificó que una distribución Gamma mejoró los criterios *AIC* y *BIC*, desde la perspectiva basada en el diseño, la especificación de la distribución influye en la precisión (reducción de varianza), pero no en las estimaciones puntuales. Las reducciones de varianza observadas de 30 % a 37 % son relevantes considerando una sola variable auxiliar óptica. Sin embargo, son menores que las citadas en estudios que utilizan datos *LiDAR*, los cuales proporcionan mayor capacidad explicativa (McRoberts et al., 2006). Por ejemplo, McRoberts et al. (2013) registraron reducciones superiores a 80 % con datos *LiDAR*, mientras que Breidenbach y Astrup (2012) determinaron eficiencias relativas entre 0.35 y 0.87. Asimismo, Zhao et al. (2024) documentaron eficiencias relativas de 0.73 en estimaciones anuales de biomasa con datos *Sentinel-2*.

Los valores del pseudo-R^2 de 0.38 para volumen y 0.30 para biomasa son consistentes con estudios previos en bosques templados que usan índices espectrales como predictores (Hernández-Ramos et al., 2020; López-Serrano et al., 2021). En conjunto, los hallazgos confirman que la estimación asistida por modelos mejora la eficiencia sin comprometer la validez basada en el diseño. Futuras investigaciones deberán evaluar la estabilidad en muestras finitas y explorar la integración de datos multifuente y técnicas avanzadas, en línea con las tendencias actuales del monitoreo forestal (Besic et al., 2025; Ståhl et al., 2016).

Conclusiones

El análisis comparativo confirma que el Estimador Asistido por Modelos (*MAE*) es superior al Estimador *Horvitz-Thompson* (*HTE*) con independencia de la especificación del modelo por utilizar. No obstante, la evidencia sugiere que la incorporación de una distribución Gamma (mediante GLM o GAM) o el uso de formas no lineales específicas (NL4) proporciona las mayores ganancias en precisión, porque se reduce la varianza cuando se usa un índice espectral como el *GNDVI* y el tratamiento silvícola. Estos modelos disminuyen eficazmente el error de muestreo cercano a 6.9 %, lo que conforma una alternativa costo-efectiva frente a inventarios que se basan solo en trabajo de campo intensivo.

Agradecimientos

Se agradece al personal técnico de la Consultoría de Fomento Ambiental y Desarrollo Social de Comunidades Forestales S. C. por el apoyo en la toma de datos dasométricos y al INIFAP por haber financiado el proyecto "Manejo integrado de recursos forestales para la sustentabilidad de los servicios ecosistémicos ante el cambio climático".

Conflicto de intereses

Los autores declaran no tener conflicto de intereses.

Contribución por autor

Efraín Velasco-Bautista y Martin Enrique Romero-Sánchez: conceptualización y supervisión del estudio; Efraín Velasco-Bautista, Alma Delia Ortiz-Reyes y Jesús Valentín Gutiérrez-García: trabajo de campo y análisis de datos; Efraín Velasco-Bautista: diseño metodológico y análisis estadístico. Todos los autores contribuyeron a la redacción, la revisión crítica y la aprobación de la versión final del manuscrito.

Referencias

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.