

Modelos no lineales y algoritmos de aprendizaje automático para modelar relaciones altura diámetro en *Pinus lawsonii* Roezl

*Nonlinear models and machine learning algorithms for modeling height-diameter relationships in *Pinus lawsonii* Roezl*

Montoya-Jiménez J. C.¹, Ruiz-González C. G.^{2*}, García-Martínez R.¹, Hernández-Soto F. N.¹, Estrada-Venegas E. G.², Equihua-Martínez A.²

 Montoya-Jiménez J. C.
juan.mj@vbravo.tecnm.mx

 Ruiz-González C. G.
ceci_forestal@hotmail.com

 García-Martínez R.
rene.gm@vbravo.tecnm.mx

 Hernández-Soto F. N.
div.forestal@vbravo.tecnm.mx

 Estrada-Venegas E. G.
edith_ev@yahoo.com.mx

 Equihua-Martínez A.
equihuaa@colpos.mx

¹Tecnológico Nacional de México-TES Valle de Bravo, División de Ingeniería Forestal. Carretera Federal Monumento-Valle de Bravo, km 30. Ejido San Antonio de la Laguna. Valle de Bravo. Estado de México.

²Colegio de Postgraduados, Entomología y Acarología, Carretera México-Texcoco Km. 36.5, Montecillo, Texcoco 56230, Estado de México.

Revista Iberoamericana de Bioeconomía y Cambio Climático
Universidad Nacional Autónoma de Nicaragua, León, Nicaragua
ISSN-e: 2410-7980
Periodicidad: Semestral
vol.10, núm.19, 2024
conrado.quirroz@ev.unanleon.edu.ni

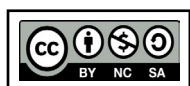
Recepción: 15 Agosto, 2024
Aprobación: 02 Diciembre, 2024

URL: https://revistas.unanleon.edu.ni/index.php/REB_ICAMCLI/article/view/1025

DOI: <https://doi.org/10.5377/ribcc.v10i19.19639>

Autor de correspondencia: ceci_forestal@hotmail.com

Copyright © 2024 Rev. iberoam. bioecon. cambio clim. Universidad Nacional Autónoma de Nicaragua León (UNAN-León), Área de Conocimiento de ciencias agrarias y veterinarias/ Área Específica de Agroecología y agronegocios /Centro de Investigación Ciencias Agrarias y Veterinarias. Dirección Académica. Departamento de Investigación. Unidad de publicaciones y eventos científicos.



Esta obra está bajo una licencia internacional
[Creative Commons AtribuciónNoComercialCompartirigual 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Resumen

Antecedentes: La medición de características del arbolado, particularmente la altura, en los inventarios forestales es una tarea costosa y requiere tiempo, especialmente en extensas superficies. Por esta razón, se han encaminado esfuerzos para generar modelos matemáticos para estimar la altura a partir del diámetro de los pinos. **Objetivo:** El objetivo de esta investigación fue comparar modelos estadísticos y algoritmos de aprendizaje automático para estimar la altura total en función del diámetro normal para árboles de *Pinus lawsonii*. **Metodología:** El estudio se realizó en el bosque natural del TESVB, para ello se midió la altura y diámetro de 295 árboles de diferente categoría diamétrica. A partir de los datos se ajustaron 15 modelos no lineales. Adicionalmente se probaron cuatro algoritmos de aprendizaje automático. El análisis se realizó con el software R. **Resultados:** Con los modelos no lineales, se observó que la mayoría de los modelos explicaron el 66 % (R²) de la varianza de la altura en función del diámetro normal, sin embargo, los modelos evaluados incumplieron los supuestos de normalidad y homocedasticidad. En el caso de los algoritmos de aprendizaje automático el porcentaje de la varianza explicada en la altura total a partir del diámetro normal, para los datos de entrenamiento osciló entre 61 y 84 %. **Conclusión:** Considerando el incumplimiento de los supuestos se recomienda bosques aleatorios para predecir altura total, además que obtuvieron predicciones aceptables y con congruencia biológica considerando la altura medida en campo de los árboles de *Pinus lawsonii*.

Palabras claves: Software R, coeficiente de determinación, Akaike, bayesiano, error.

Abstract

Background: Measuring tree characteristics, particularly height, in forest inventories is a costly and time-consuming task, especially in large areas. For this reason, efforts have been directed to generate mathematical models to estimate height from the diameter of pine trees. **Aim:** the objective of this research was to compare statistical models and machine learning algorithms to estimate total height based on normal diameter for *Pinus lawsonii* trees. **Methodology:** The study was carried out in the TESVB forest property, for which the height and diameter of 295 trees of different diameter categories were measured. Fifteen non-linear models were fitted from the data. Additionally, four machine learning algorithms were tested. The analysis was carried out with the R software. **Result:** With the non-linear models, it was observed that most of the models explained 66% (R²) of the variance in height based on normal diameter, however, all of them failed to meet the assumptions of normality and homoscedasticity. In the case of machine learning algorithms, the percentage of variance explained in total height from normal diameter for the training data ranged between 61 and 84 %. **Conclusion:** Considering the failure to meet assumptions, random forests are recommended to predict total height, and acceptable predictions with biological consistency were obtained considering the height measured in the field of *Pinus lawsonii* trees.

Keywords: R software, coefficient of determination, Akaike, Bayesian, error. Aphis spp.

Introducción

Los estudios relacionados con la biometría forestal dependen en gran medida de mediciones precisas en las dimensiones de los árboles. En los inventarios forestales donde el objetivo principal es cuantificar la cantidad de madera existente, las características principales que se miden en los árboles son el diámetro y la altura, a partir de estas mediciones es posible realizar cuantificaciones de volumen, área basal y biomasa aérea (Li et al., 2020; Özçelik et al., 2013; Şahin, 2024). Sin embargo, medir la altura de los árboles en campo es una tarea costosa y requiere mucho tiempo, especialmente en grandes superficies. Por lo tanto, los modelos que predicen la altura de los árboles a partir del diámetro normal (1.30 m), se han convertido en herramientas esenciales en las ciencias forestales (Castillo-Gallegos et al., 2018; Guerra-De la Cruz et al., 2018; Hernandez-Ramos et al., 2020; López-Villegas et al., 2017; Mayrinck et al., 2019).

En la literatura científica la regresión lineal y no lineal son técnicas estadísticas útiles y ampliamente aplicadas para modelar la relación altura - diámetro (Guerra-De la Cruz et al., 2018), las estimaciones del valor de los coeficientes de estas técnicas se obtienen generalmente mediante el método de mínimos cuadrados ordinarios (MCO), aunque la estimación de los parámetros del modelo a menudo se ve obstaculizada por la falta en el cumplimiento de los supuestos del método, que de acuerdo con Zdenek & Drápela (2016), es muy común.

Cuando no se cumplen con los supuestos de normalidad, independencia y homocedasticidad, los modelos de regresión lineal y no lineal tienen limitaciones significativas. Para resolver lo expuesto anteriormente se recurren a enfoques paramétricos y no paramétricos espaciales (Kangas & Haara, 2012), modelos de efectos mixtos (Zdenek & Drápela, 2016), uso de estimadores por mínimos cuadrados generalizados y con funciones de varianza combinadas (Quiñonez-Barraza & García-Espinoza, 2018). Varios autores (Lima et al., 2022; Nunes-Miranda et al., 2022; Özçelik et al., 2013 y Şahin, 2024), emplean modelos de aprendizaje automático; estos enfoques no requieren suposiciones sobre la forma de la función de ajuste, distribuciones no gaussianas, multicolinealidad, valores atípicos y ruido en los datos (Che et al., 2018).

Se han realizado estudios para analizar la relación altura-diámetro en las siguientes especies: *Pinus teocote*, *P. montezumae*, *P. chiapensis*, *P. arizonica*, *P. ayacahuite*, *P. durangensis*, *P. herrerae*, *P. lumholtzii*, *P. douglasiana* y *P. pseudostrobus* (Castillo-Gallegos et al., 2018; Corral Rivas et al., 2019; Flores-Morales et al., 2019; Guerra-De la Cruz et al., 2018) y otras especies que no son coníferas (García Cuevas et al., 2017; Lima et al., 2022), sin embargo, se han realizado muy pocos esfuerzos para estudiar dicha relación en *Pinus lawsonii* (Mayrinck et al., 2019), por lo que el presente estudio se realizó con el objetivo de analizar 15 modelos estadísticos y cuatro algoritmos de aprendizaje automático y seleccionar los que mejor estimen la altura total en función del diámetro normal para árboles de *Pinus lawsonii* del predio forestal del Tecnológico de Estudios Superiores de Valle de Bravo (TESVB). Aunque el presente estudio no se contempló cuantificar volumen biomasa o carbono las estimaciones de altura total se pueden utilizar para hacer inferencias sobre dichas variables.

Materiales y métodos

Área de estudio

El estudio se realizó en el bosque natural del Tecnológico de Estudios Superiores de Valle de Bravo (TESVB) ubicado en el municipio de Donato Guerra al suroeste del Estado de México (Figura 1), los árboles tienen una edad promedio de 75 años. El área de estudio presenta vegetación de bosque de pino-encino, tiene una altitud de 1,800 m, con clima del tipo (A)C(w"1)(wi)g templado semicálido, con precipitaciones que van 1,100 a 1,300 mm, la temperatura media anual va de 12 a 22 °C (INEGI, 2010).

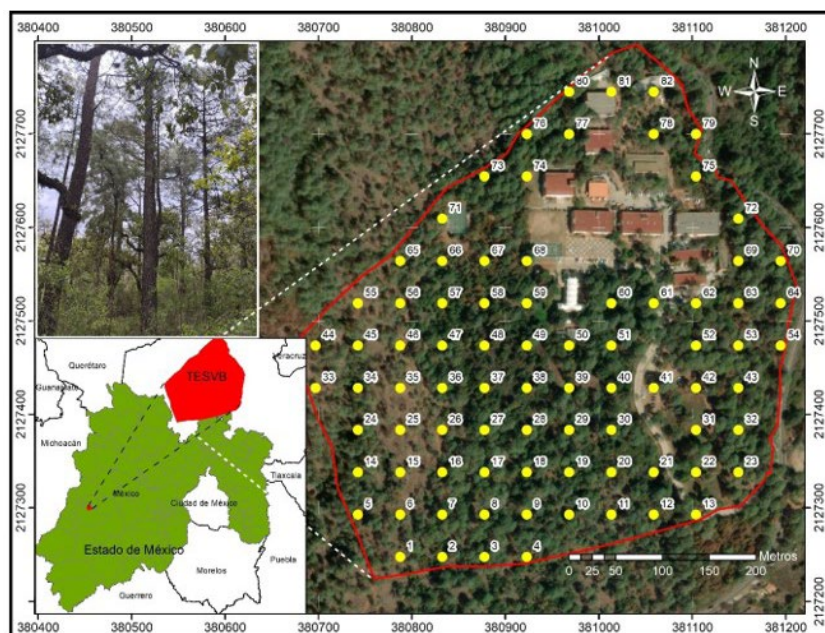


Figura 1. Ubicación del área de estudio en el Tecnológico de Estudios Superiores de Valle de Bravo (TESVB), Estado de México.

Obtención de datos dasométricos

A partir de un muestreo sistemático se establecieron 82 sitios circulares de 500 m² en toda el área de estudio. En cada sitio se midió la altura total y el diámetro normal (a 1.30 m) de los árboles, estas variables se midieron con una pistola Haga ® y una forcípula Haglof respectivamente, en total se midieron 295 individuos. Los árboles muestreados reflejaron la variación en las clases de diámetro y altura existentes en el bosque del TESVB (Tabla 1). Los 295 árboles medidos se dividieron en dos grupos: datos de entrenamiento (modelo) (208 árboles, 70 % del total de datos) y datos de prueba (87 árboles, el 30 % restante). La [Tabla 1](#) muestra estadísticas descriptivas para los conjuntos de datos utilizados en el presente estudio.

Tabla 1. Estadísticas descriptivas de los conjuntos de datos global, de ajuste y validación. (STD): desviación estándar

Datos	N	Variables	Promedio	Máximo	Mínimo	Sd
Global	295	dap (cm)	33.38	75.00	7.00	15.31
	295	h (m)	16.61	30.00	4.00	6.00
Entrenamiento	208	dap (cm)	33.50	75.00	7.70	15.00
	208	h (m)	16.36	30.00	4.00	5.93
Prueba	87	dap (cm)	33.08	75.00	7.00	16.11
	87	h (m)	17.21	30.00	4.00	6.16

El presente estudio se realizó en un bosque natural que cuenta con árboles de diferentes edades que van desde los 15 años a los 105 años. La [Figura 2](#) denota la distribución de las mediciones de altura y diámetro realizadas.

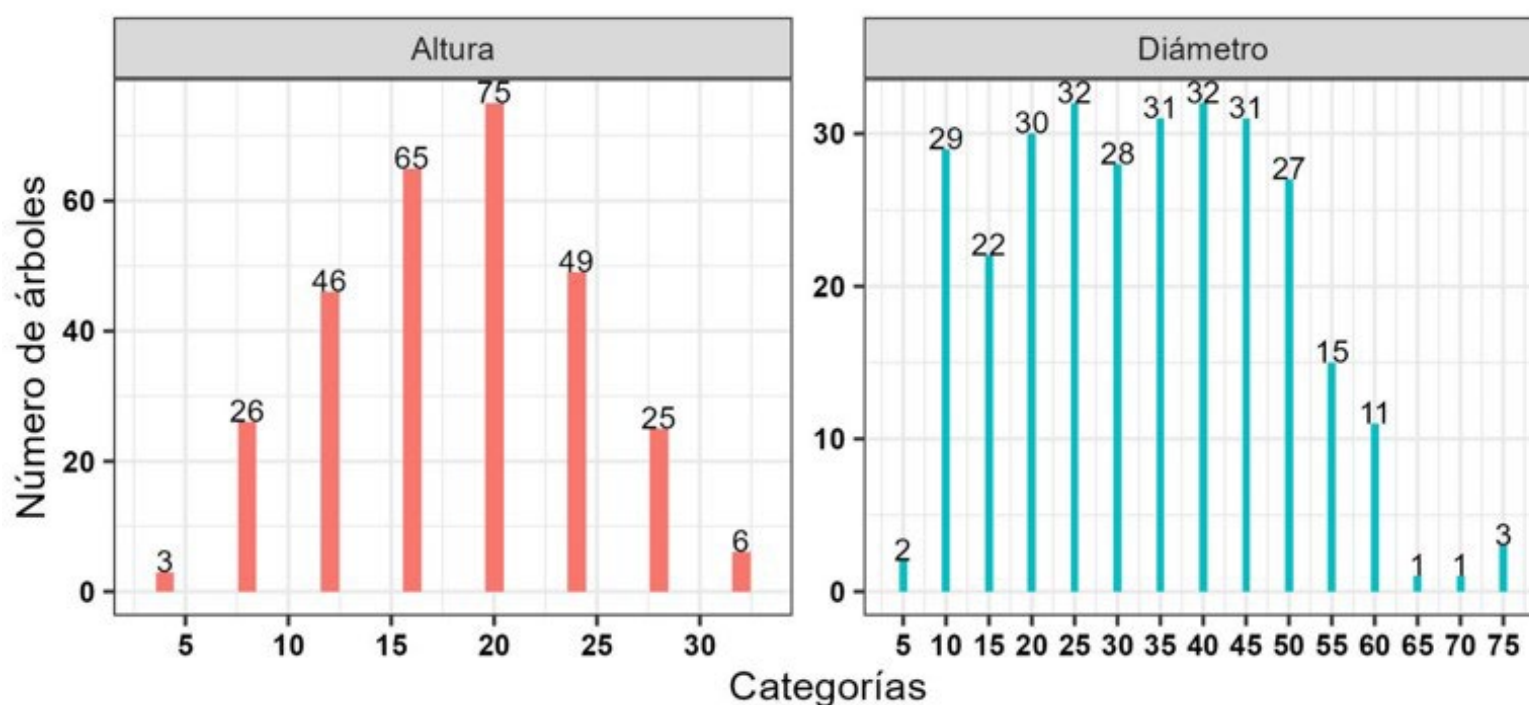


Figura 2. Distribución de categorías de medición de altura y diámetro para ajustar los modelos y algoritmos de aprendizaje automático en *Pinus lawsoni*.

Modelos, algoritmos y precisión

En el presente estudio se ajustaron 14 modelos no lineales y un modelo cuadrático los cuales son ampliamente utilizados en estudios sobre relaciones alométricas (Tabla 2). La estructura de los modelos se tomaron del estudio de Guerra-De la Cruz et al. (2018). Los modelos se ajustaron con el conjunto de datos global a través del software R (R Core Team, 2024) utilizando la librería gsnls (Chau, 2024).

Tabla 2. Distribución de categorías de medición de altura y diámetro para ajustar los modelos y algoritmos de aprendizaje automático en *Pinus lawsoni*.

Modelo	Ecuación	Referencia
MC	$h = \beta_0 [1 - e^{-\beta_1 d}]^{\beta_2}$	(Clutter et al., 1983)
MT1	$h = \beta_0 [1 - e^{-\beta_1 d}]^{\beta_2}$	(Temesgen et al., 2014)
MT2	$h = e^{\left(\beta_0 + \frac{\beta_1}{d + \beta_2}\right)}$	
MC1	$h = \beta_0 e^{(-\beta_1 d^{-\beta_2})}$	(Castillo-Gallegos et al., 2018)
MC2	$h = \frac{d^{\beta_0}}{\beta_1 + \beta_2 d^{\beta_0}}$	
MC3	$h = \frac{\beta_0 d}{\beta_1 + d}$	
MC4	$h = \left(\beta_0 + \frac{\beta_1}{d}\right)^{-\beta_2}$	
MC5	$h = \beta_0 \left(1 + \frac{1}{d}\right)^{-\beta_1}$	
MC6	$h = \beta_0 \left(-\ln\left(\frac{1}{d}\right)\right)^{\beta_1}$	
MC7	$h = \left(\beta_0 + \frac{\beta_1}{d}\right)^{-\beta_2}$	
MS	$h = \beta_0 d^{(\beta_1 + \beta_2 d)}$	(Sharma, 2009)
MG1	$h = \beta_0 + \beta_1 d^{\beta_2}$	(García Cuevas et al., 2017)
MG2	$h = \beta_0 + \beta_1 d + \beta_2 d^2$	
MG3	$h = \beta_0 e^{\beta_1 \left(\frac{1}{d}\right)}$	
MG4	$h = e^{\beta_0 + \beta_1 \ln(d)}$	

donde: h = Altura total; d = Diámetro normal; β β = Parámetros del modelo. Tabla modificada del trabajo de (Guerra-De la Cruz et al., 2018).

Los algoritmos de aprendizaje automático que se utilizaron en el presente estudio fueron bosques aleatorios (RF por sus siglas en inglés) ([Breiman, 2001](#)), máquinas de soporte vectorial (SVM por sus siglas en inglés) ([Betancourt, 2005](#)), potenciación de gradiente extremo (XGBoost por sus siglas en inglés) ([Darmatasia & Arymurthy, 2017](#)), y redes neuronales artificiales (ANN por sus siglas en inglés) ([Krogh, 2008](#)). Estos algoritmos de aprendizaje automático son de utilidad debido a que abordan problemas de clasificación (datos categóricos) y regresión (datos continuos) especialmente cuando no se cumple alguno de los supuestos de los modelos paramétricos. En la biometría forestal se han utilizado exitosamente para predecir altura ([Zdenek & Drápela, 2016](#)), volumen ([Şahin, 2024](#)), biomasa aérea ([Li et al., 2020](#)).

Los algoritmos se entrenaron y evaluaron utilizando los dos conjuntos de datos descritos en la Tabla 1 (entrenamiento y prueba), a través del software R ([R Core Team, 2024](#)) utilizando las librerías: randomForest ([Liaw & Wiener, 2002](#)), e1071 ([Meyer et al., 2023](#)), xgboost ([Chen & Guestrin, 2016](#)) y h2o (H2O.ai., 2022).

Para evaluar y comparar la bondad de ajuste de los modelos no lineales y el cuadrático se examinaron seis criterios estadísticos: suma de cuadrados del error (SCE), cuadrado medio del error (CME), raíz del cuadrado medio del error (RCME) coeficiente de determinación (R^2), criterio de información de Akaike corregido (AICc) y el criterio de información Bayesiano (BIC). Finalmente, a cada uno de los criterios estadísticos se les asignó una calificación del 1 al 15 (1 = mejor ajuste y 15 = menor ajuste) las calificaciones obtenidas en los criterios se sumaron para cada uno de los modelos y se obtuvo una calificación final (CF), el modelo con menor puntuación en la calificación final representó el mejor ajuste. Para los algoritmos de aprendizaje automático se consideraron los primeros cuatro estadísticos que se utilizaron en los modelos no lineales y con base en estos se asignó la CF siguiendo el procedimiento anteriormente explicado. Además de evaluar la bondad de ajuste, se realizó la exploración gráfica de los residuales y el cumplimiento de los supuestos de normalidad y homogeneidad de varianza a través de la prueba de Shapiro-Wilks y Levene respectivamente.

Resultados

Con base en el coeficiente de determinación (R^2) de los modelos analizados, se observó que la mayoría explicaron el 66 % de la varianza de la altura en función del diámetro normal de *Pinus lawsonii*, excepto por los modelos MC5, MG3 y MG4 los cuales explicaron menor porcentaje de varianza (65 %) ([Tabla 3](#)).

Tabla 3. Estadísticos de bondad de ajuste de los modelos analizados para modelar la altura de *Pinus lawsonii*

Mod	SCE	CME	RCME	R2	AICc	BIC	CF
MC	3510.7 (3)	12.02 (3)	3.46 (3)	0.66 (3)	1575.84 (4)	1590.51 (5)	21
MT1	3510.5 (2)	12.02 (2)	3.46 (2)	0.66 (2)	1575.83 (3)	1590.49 (4)	15
MT2	3515.5 (5)	12.03 (6)	3.46 (6)	0.66 (5)	1576.25 (6)	1590.91 (7)	35
MC1	3524.9 (9)	12.07 (10)	3.47 (10)	0.66 (10)	1577.04 (11)	1591.71 (11)	61
MC2	3515.8 (6)	12.04 (7)	3.46 (7)	0.66 (7)	1576.28 (7)	1590.94 (8)	42
MC3	3526.8 (10)	12.03 (5)	3.46 (5)	0.66 (6)	1575.16 (1)	1586.18 (1)	28
MC4	3515.8 (7)	12.04 (8)	3.46 (8)	0.66 (8)	1576.28 (8)	1590.94 (9)	48
MC5	3614.7 (13)	12.33 (13)	3.51 (13)	0.65 (13)	1582.42 (13)	1593.44 (13)	78
MC6	3548.1 (12)	12.10 (11)	3.47 (11)	0.66 (11)	1576.93 (10)	1587.95 (2)	57
MC7	3517.8 (8)	12.04 (9)	3.47 (9)	0.66 (9)	1576.44 (9)	1591.11 (10)	54
MS	3507.4 (1)	12.01 (1)	3.46 (1)	0.66 (1)	1575.57 (2)	1590.23 (3)	9
MG1	3537.6 (11)	12.11 (12)	3.48 (12)	0.66 (12)	1578.1 (12)	1592.76 (12)	71
MG2	3511.6 (4)	12.02 (4)	3.46 (4)	0.66 (4)	1575.92 (5)	1590.58 (6)	27
MG3	3633.9 (14)	12.40 (14)	3.52 (14)	0.65 (14)	1583.98 (14)	1595 (14)	84
MG4	3654.8 (15)	12.47 (15)	3.53 (15)	0.65 (15)	1585.68 (15)	1596.7 (15)	90

donde: Mod = Modelo; SCE = Suma de cuadrados del error; CME = Cuadrado medio del error; RCME = Raíz del cuadrado medio del error; R^2 = Coeficiente de determinación; AICc = Criterio de información de Akaike corregido; BIC = criterio de información bayesiano; CF = Calificación final.

De acuerdo con la interpretación de la calificación final, que aquellos modelos con menores valores son los que mostraron los mejores valores de todos los criterios de ajuste, se dilucidó que los mejores modelos fueron MS y MT1 (9 y 15 en la calificación final). Es importante señalar que los modelos MG4, MG3, MC5, MG1, MC1 obtuvieron menor ajuste debido a que sus calificaciones finales fueron las más altas (entre 60 y 90), lo cual indicó que, en la SCE, el CME, la RCME, el AICc y el BIC obtuvieron valores altos y en el R^2 obtuvieron valores bajos. Los ocho modelos restantes obtuvieron un ajuste medio (Tabla 3).

Adicional a la evaluación estadística de los modelos, para la selección de los mejores modelos se consideró la significancia de los estimadores de los modelos evaluados, a partir de estos resultados se observó que la mayoría de los modelos resultaron significativos sus estimadores, excepto por los modelos MC7, MG1 y MG2 en los que uno o sus tres estimadores resultaron no significativos (Tabla 4).

Tabla 4. Valores y significancia de los coeficientes de los modelos analizados para *Pinus lawsonii*

Mod	b0	Pr(> t)	b1	Pr(> t)	b2	Pr(> t)
MC	26.7752	<.0001	0.033536	0.0005	1.038306	<.0001
MT1	26.57666	<.0001	0.029748	<.0001	1.030421	<.0001
MT2	3.560681	<.0001	28.38669	<.0001	8.01972	0.0316
MC1	50.67122	0.0067	6.37025	<.0001	0.513278	0.0018
MC2	1.171673	<.0001	1.614935	0.0332	0.029873	<.0001
MC3	39.94324	<.0001	42.47135	<.0001		
MC4	33.44772	<.0001	0.018465	0.0026	1.172748	<.0001
MC5	29.47461	<.0001	16.46817	<.0001		
MC6	1.501589	<.0001	-1.95477	<.0001		
MC7	0.065138	0.3588	1.562929	0.0139	1.312784	0.0071
MS	0.996878	0.0002	0.889054	<.0001	-0.00204	0.0009
MG1	-45.2961	0.4352	37.21877	0.4844	0.149597	0.2993
MG2	1.65259	0.0689	0.637072	<.0001	-0.00468	<.0001
MG 3	29.09546	<.0001	-15.7366	<.0001		
MG 4	0.799668	<.0001	0.581772	<.0001		

donde: b_0 , b_1 y b_2 = Coeficientes de regresión; $Pr>[t]$ = Valor de Probabilidad de t.

Los valores observados de altura total de los árboles de *Pinus lawsonii* se compararon gráficamente con los valores estimados por los modelos MT1 y MS y se observó que ambos modelos predicen de forma similar en el intervalo de 7.5 a 65 cm de diámetro normal, es a partir de los 70 cm que se nota una ligera diferencia entre los modelos, sin embargo, ambos modelos mostraron una solución válida para la estimación de la altura de *Pinus lawsonii* (Figura 3).

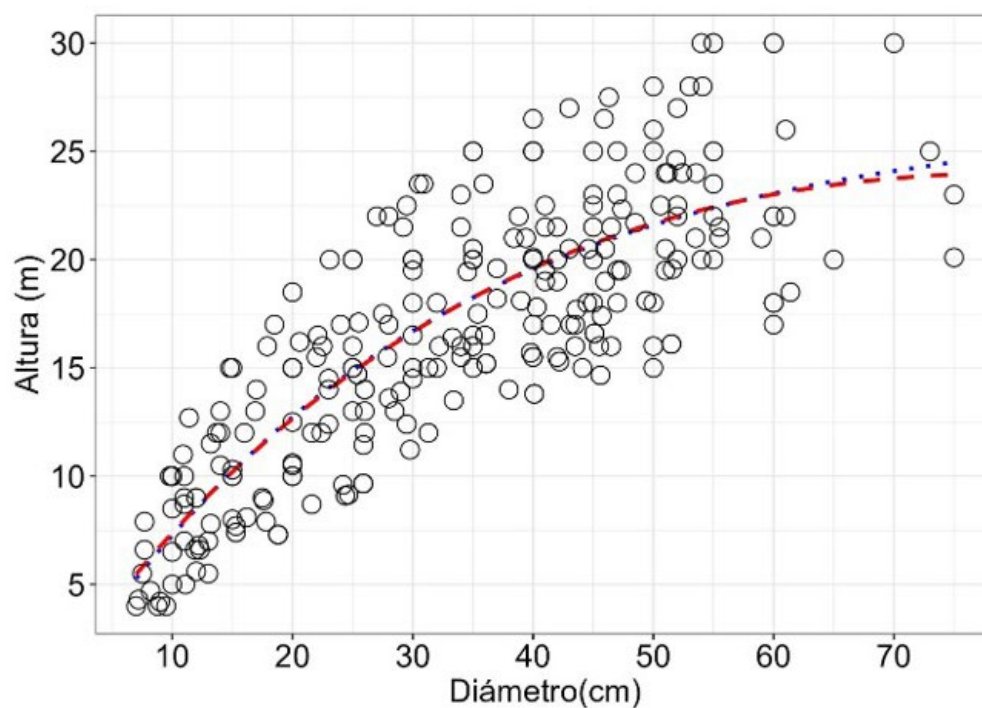


Figura 3. Valores observados y predichos de los modelos MT1 (línea azul) y MS (línea roja) ajustados para predecir altura en *Pinus lawsonii*

A pesar de que los modelos se ajustaron a los datos y que la mayoría de estos resultaron significativos sus estimadores, en todos los casos se incumplieron los supuestos de normalidad y homocedasticidad en los residuos. De acuerdo con la prueba de Shapiro-Wilks aplicada sobre los residuos de los modelos MT1 y MS se observó que ambos incumplieron con el supuesto ($W = 0.97442$, $p\text{-value} = 4.046e-05$ y $W = 0.97448$, $p\text{-value} = 4.135e-05$), el mismo resultado se obtuvo en la prueba de Levene para los dos modelos ($Pr(>F) = 0.04115$ y $Pr(>F) = 0.04115$), este resultado se confirmó a través de la exploración gráfica de los residuos (Figura 4).

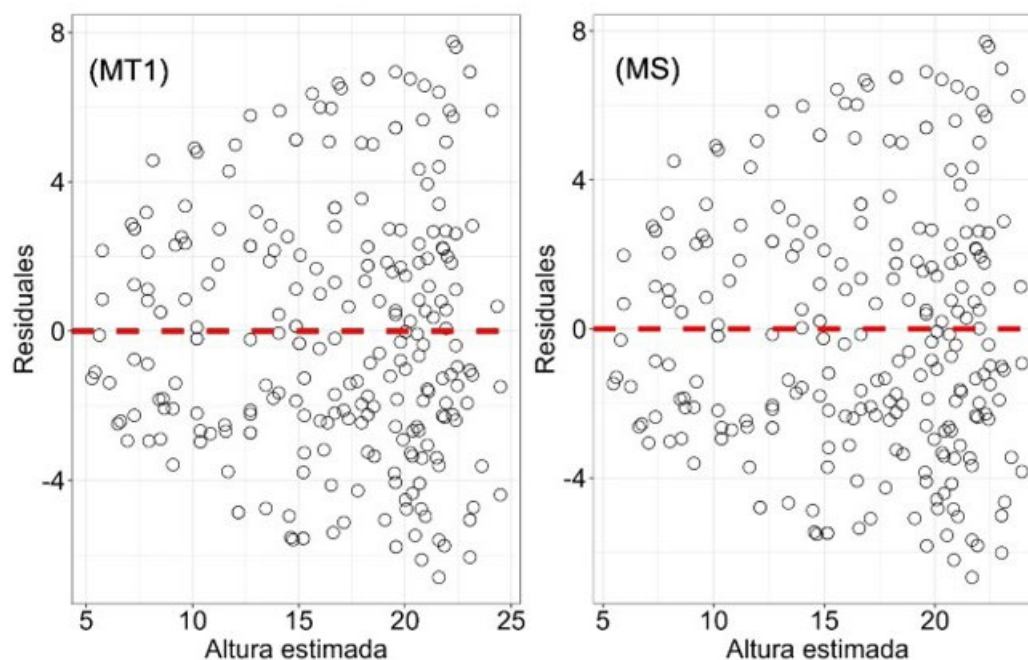


Figura 4. Gráfica de residuales contra altura estimada de los modelos de altura-diámetro ajustados para *Pinus lawsonii*

Los estadísticos de bondad de ajuste (SCE, CME, RCME y R2) para el conjunto de datos de entrenamiento y prueba, que se obtuvieron para los algoritmos de aprendizaje automático utilizados en la predicción de la altura de *Pinus lawsonii*, denotaron que el porcentaje de la varianza explicada en la altura total a partir del diámetro normal, para los datos de entrenamiento osciló entre 61 y 84 % para SVM y RF. En los datos de prueba este porcentaje de varianza explicada disminuyó significativamente, el valor más bajo se obtuvo en NNA (0.57), por el contrario, bosques aleatorios obtuvo el valor más alto (0.61). Considerando el mismo estadístico (R2) el algoritmo XGBoost se encontró con un valor intermedio (Tabla 5)

Tabla 5. Estadísticos de bondad de ajuste de los algoritmos de aprendizaje automático analizados para modelar la altura de *Pinus lawsonii*

Alg	Datos	SCE	CME	RCME	R2	CF
RF	Entrenamiento	1188.19 (1)	5.71 (1)	2.39 (1)	0.84 (1)	4
	Prueba	1283.4 (1)	14.75 (1)	3.84 (1)	0.61 (1)	4
SVM	Entrenamiento	345.85 (3)	11.27 (3)	3.35 (3)	0.61 (4)	13
	Prueba	1327.27 (3)	15.25 (3)	3.90 (3)	0.59 (3)	12
XGBoost	Entrenamiento	1656.64 (2)	7.96 (2)	2.82 (2)	0.77 (2)	8
	Prueba	1293.30 (2)	14.86 (2)	3.85 (2)	0.60 (2)	8
NNA	Entrenamiento	2517.38 (4)	12.10 (4)	3.47 (4)	0.65 (3)	15
	Prueba	1382.14 (4)	15.88 (4)	3.98 (4)	0.57 (4)	16

donde: Alg = Algoritmo; RF = bosques aleatorios; SVM = máquinas de soporte vectorial; XGBoost = potenciación de gradiente extremo; NNA = Redes neuronales artificiales; SCE = Suma de cuadrados del error; CME = Cuadrado medio del error; RCME = Raíz del cuadrado medio del error; R^2 = Coeficiente de determinación; CF = Calificación final.

Considerando la calificación final obtenida en los cuatro algoritmos, se dilucidó que el ajuste general de los algoritmos evaluados en orden descendente fue el siguiente: RF, XGB, SVM y NNA. Lo anterior denotó que RF obtuvo valores bajos en la SCE, el CME, la RCME, el AICc y el BIC y valores altos en el R2 (Tabla 5), por lo que fue el algoritmo que se seleccionó como el más idóneo para hacer inferencias sobre la altura total de *Pinus lawsonii*.

Además de la evaluación estadística de los algoritmos, se graficaron los valores observados de altura de los datos de prueba y los valores estimados a través de los cuatro algoritmos, se observó que RF obtuvo predicciones bastante razonables y con congruencia biológica considerando la altura observada (Figura 5a), por el contrario, SVM mostró que con valores pequeños y altos en el diámetro normal sobrestima y subestima la altura total (Figura 5b), el algoritmo XGB reveló un buen ajuste, sin embargo, para diámetros grandes subestima la altura (Figura 5c), finalmente NNA mostró un bajo ajuste respecto a los datos observados (Figura 5d). (Figura 5).

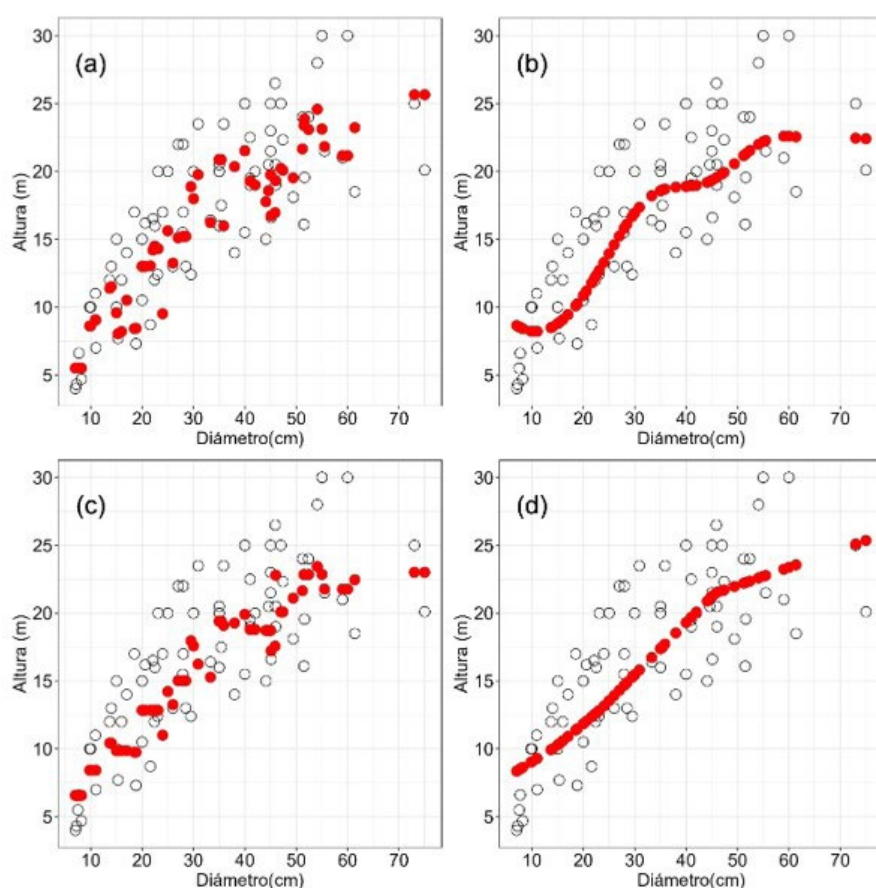


Figura 5. Gráfica de valores observados (círculos negros) y predichos (puntos rojos) con los datos de prueba del ajuste de los algoritmos: bosques aleatorios (a) máquinas de soporte vectorial (b), XGB (c) y redes neuronales artificiales (d) para predecir altura en *Pinus lawsonii*.

Discusión

La bondad de ajuste de los modelos medida a través de los estadísticos considerados (SCE, CME, RCME, R2, AICc, BIC, CF) fue baja. Con fines comparativos en el estudio de Guerra-De la Cruz et al. (2018) encontraron que el coeficiente determinación (R2) de los modelos analizados para *Pinus montezumae* y *P. teocote* osciló entre 0.89 y 0.91 y 0.80 y 0.87 respectivamente, en el presente estudio se obtuvieron valores entre 0.65 y 0.66 los cuáles fueron más bajos. Este mismo comportamiento se repitió respecto a los resultados que reportaron López-Villegas et al. (2017) para *P. pseudostrobus* (0.73-0.88), *P. patula* (0.67-0.94), *P. oaxacana* (0.69-0.89) y *P. douglasiana* (0.67-0.87).

Sin embargo, el intervalo de valores obtenidos en el R2 (0.65-0.66) fue superior respecto al intervalo de valores que obtuvieron Castillo-Gallegos et al. (2018) en el ajuste de modelos para la predicción de altura total en *Pinus chiapensis* (0.33 y 0.35). Algo similar a lo que encontraron Hernández-Ramos et al. (2020) en el que se ajustaron ecuaciones locales y generalizadas de altura-diámetro para *Pinus patula*, su intervalo de valores en el coeficiente determinación para las ecuaciones locales (0.59-0.62) fue inferior al presente estudio (0.65-0.66).

Este ajuste bajo que se obtuvo en los modelos se debe a la dispersión tan amplia de los datos, misma que pudiera explicarse por calidad del sitio, la densidad del rodal, competencia, incluso que la especie se haya desarrollado fuera del rango óptimo de condiciones ambientales en las que generalmente crece ([Castillo-Gallegos et al., 2018](#)). De acuerdo con [Corral Rivas et al. \(2019\)](#) el ajuste de las ecuaciones locales altura-diámetro se puede mejorar significativamente (6.6 %) utilizando ecuaciones generalizadas de altura-diámetro, lo cual representa una alternativa viable para incrementar el porcentaje de variabilidad explicada en la altura a partir del diámetro normal en especies forestales.

A pesar que el ajuste de los modelos no fue tan bajo comparado con otros estudios, a través del análisis de los residuales, se observó que todos los modelos incumplieron con el supuesto de homogeneidad de varianza, por lo que, se obtendrían estimaciones de parámetros con varianza no mínima e intervalos de predicción poco fiables ([Quiñonez-Barraza & García-Espinoza, 2018](#)). De acuerdo con los autores anteriormente citados utilizar funciones de varianza ayuda a modelar la heterocedasticidad.

Los modelos de aprendizaje automático utilizados en el presente estudio fueron de utilidad ya que los modelos no lineales incumplieron los supuestos, en estudios como el de [Lima et al. \(2022\)](#) utilizaron RF para modelar la altura de Eucalyptus; [Nunes-Miranda et al. \(2022\)](#) usaron RF para varias especies de árboles; [Özçelik et al. \(2013\)](#) y [Sahin, \(2024\)](#) utilizaron NNA para estimar la altura de Juniperus excelsa ambos estudios realizados en Turquía. Nuestros algoritmos ajustados obtuvieron menor ajuste en ambos conjuntos de datos (entrenamiento y prueba) respecto a los trabajos de [Nunes-Miranda et al., \(2022\)](#) y [Özçelik et al. \(2013\)](#). Obtener que el ajuste de los datos de validación sea inferior a los datos de entrenamiento es común ([Ercanli et al., 2022](#)), lo cual se puede mejorar a través de la búsqueda de los mejores parámetros del algoritmo (Afinación) ([Sahin, 2024](#)).

Los resultados de este estudio difieren respecto a los estudios [Che et al. \(2018\)](#) y [Kangas & Haara \(2012\)](#) en los que afirman que los algoritmos de aprendizaje automático pueden superar al método de regresión tradicional. Sin embargo, se visualizó a través de la gráfica de valores observados y predichos que las predicciones mostraron realismo biológico respecto al conjunto de datos de prueba en el siguiente orden: RF, XGBoost, NNA y SVM (mayor-menor realismo).

Conclusiones

Los enfoques de modelado de regresión no lineal, así como los algoritmos de aprendizaje automático son herramientas útiles para la predicción de la altura total de los árboles de *Pinus lawsonii*, sin embargo, en el presente estudio los algoritmos de aprendizaje automático representaron una solución viable, debido a que no se cumplieron los supuestos de la regresión no lineal, a partir de estos resultados se observó que los algoritmos de aprendizaje automático se ajustan mejor cuando se tienen suficientes datos para describir un fenómeno. Considerando los resultados actuales para futuras investigaciones se puede mejorar a través del uso de ecuaciones generalizadas. RF obtuvo la calificación final más alta, en consecuencia, se seleccionó como el más idóneo para hacer inferencias sobre la altura total de *Pinus lawsonii*, y utilizar la información en los inventarios forestales para realizar cuantificaciones de volumen, área basal y biomasa aérea o para cumplir con los objetivos que se quieran alcanzar en el ámbito forestal.

Fondos: Este estudio no fue financiado.

Cumplimiento de estándares éticos: No aplica

Contribuciones de autor: Los autores participaron en la Conceptualización, Metodología, Redacción, Borrador Original, Redacción, Revisión y Edición y Supervisión

Conflicto de intereses: Los autores declaran que no tienen conflictos de interés.

Disponibilidad de datos: Los conjuntos de datos analizados en el presente estudio pueden ser solicitados a través del correo electrónico del autor de correspondencia.

Referencias bibliográficas

- Betancourt, G. A. (2005). Las máquinas de soporte vectorial (SVMs). *Scientia et Technica*, 27, 67–72.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1201/9780429469275-8>
- Castillo-Gallegos, E., Jarillo-Rodríguez, J., & Escobar-Hernández, R. (2018). Relación altura-diámetro en tres especies cultivadas en una plantación forestal comercial en el Este tropical de México. *Revista Chapingo, Serie Ciencias Forestales y Del Ambiente*, 24(1), 33–48. <https://doi.org/10.5154/r.rchscfa.2017.05.033>
- Chau, J. (2024). *gslns: GSL Multi-Start Nonlinear Least-Squares Fitting*. R package version 1.3.2, <<https://CRAN.R-project.org/package=gslns>>. (p.37). <https://doi.org/10.32614/CRAN.package.gslns>
- Che, S., Tan, X., Xiang, C., Sun, J., Hu, X., Zhang, X., Duan, A., & Zhang, J. (2018). Stand basal area modelling for Chinese fir plantations using an artificial neural network model. *Journal of Forestry Research*, 30(5), 1641–1649. <https://doi.org/10.1007/s11676-018-0711-9>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Clutter, J. L., Fortson, J. C., Pienaar, L. V., Brister, G. H., & Bailey, R. L. (1983). *Timber Management: A quantitative approach*. (Primera ed). ohn Wiley & Sons.
- Corral Rivas, S., Alma Mireya Silva, A., & Quiñonez-Barraza, G. (2019). A generalized nonlinear height-diameter model with mixed-effects for seven *Pinus* species in Durango, Mexico. *Revista Mexicana de Ciencias Forestales*, 10(53), 1–32. <https://doi.org/https://doi.org/10.29298/rmcf.v10i53.500>
- Darmatasia, & Arymurthy, A. M. (2017). Predicting the status of water pumps using data mining approach. 2016 International Workshop on Big Data and Information Security, IWBIS 2016, 57–63. <https://doi.org/10.1109/IWBIS.2016.7872890>
- Ercanli, İ., Senyurt, M., & Bolat, F. (2022). a Major Challenge To Machine Learning Models: Compatible Predictions With Biological Realism in Forestry: a Case Study of individual tree volume. *Proceeding of the “3rd International Conference on Environment and Forest Conservation (ICEFC)”*, 1–31.
- Flores-Morales, E. A., Aguirre-Calderón, O. A., Quiñonez-Barraza, G., González-Tagle, M. A., & Jiménez-Pérez, J. (2019). Normal diameter, height and volume estimation of *Pinus pseudostrobus* Lindl. From the diameter of the stump. *Revista Mexicana de Ciencias Forestales*, 10(55), 154–170. <https://doi.org/10.29298/rmcf.v10i55.547>
- García Cuevas, X. A., Hernández Ramos, J., Hernández Ramos, A., Quiñonez Barraza, G., Tamarit Urías, J. C., & García Espinoza, G. G. (2017). Predicción del diámetro normal, altura y volumen a partir del diámetro del tocón en especies tropicales. *Revista Mexicana de Ciencias Forestales*, 8(43), 89–116. <https://doi.org/10.29298/rmcf.v8i43.67>
- Guerra-De la Cruz, V., Islas-Gutiérrez, F., Flores-Ayala, E., Acosta-Mireles, M., Buendía-Rodríguez, E., Carrillo-Anzures, F., Tamarit-Urías, J. C., & Pineda-Ojeda, T. (2018). Modelos locales altura-diámetro para *Pinus montezumae* Lamb. y *Pinus teocote* Schiede ex Schltdl. en Nanacamilpa, Tlaxcala. *Revista Mexicana de Ciencias Forestales*, 10(51), 133–156. <https://doi.org/https://doi.org/10.29298/rmcf.v10i51.407>
- H2O.ai. (2022). *h2o: R Interface for H2O R package (version 3.42.0.2.)*. <https://github.com/h2oai/h2o-3>
- Hernández-Ramos, J., Aviles-Castillo, A., Jesus Garcia-Magana, J., Hernández-Ramos, A., García-Cuevas, X., & Flores-Lopez, C. (2020). Local and generalized height-diameter equations for *Pinus patula* Schl. et Cham. in Veracruz, Mexico. *Ecosistemas Y Recursos Agropecuarios*, 7(3), 1–11. <https://doi.org/10.19136/era.a7n3.2457>
- INEGI. (2010). *Compendio de información geográfica municipal 2010*. Instituto Nacional de Estadística y Geografía, 10. <http://mapserver.inegi.org.mx/mgn2k/>.
- Kangas, A., & Haara, A. (2012). Comparison of nonspatial and spatial approaches with parametric and nonparametric methods in prediction of tree height. *European Journal of Forest Research*, 131(6), 1771–1782. <https://doi.org/10.1007/s10342-012-0631-8>

- Krogh, A. (2008). What are artificial neural networks? *Nature Biotechnology*, 26(2), 195–197. <https://doi.org/10.1038/nbt1386>
- Li, Y., Li, M., Li, C., & Liu, Z. (2020). Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine learning algorithms. *Scientific Reports*, 10(1), 1–12. <https://doi.org/10.1038/s41598-020-67024-3>
- Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2(3), 18–22. <https://cran.r-project.org/doc/Rnews/>
- Lima, E. D. S., Souza, Z. M. De, Oliveira, S. R. D. M., Montanari, R., & Farhate, C. V. V. (2022). Random forest model to predict the height of Eucalyptus. *Engenharia Agrícola*, v.42, special issue, e20210153 <https://doi.org/10.1590/1809-4430-Eng.Agric.v42nepe20210153/2022>
- López-Villegas, M. F., Santiago-García, W., Quiñonez-Barraza, G., Suárez-Mota, M. E., Santiago-Juárez, W., & Santiago-García, E. (2017). Ecuaciones globales y locales de altura-diámetro de 12 especies de interés comercial en bosques manejados. *Revista Mexicana de Agroecosistemas*, 4(2), 113–126.
- Mayrincck, R. C., Roque, V. G. R., Filho, A. C. F., Filho, E. M., Arias-King, F., & Ribeiro, A. (2019). Height and volume functions for *Pinus lawsonii*, *Pinus leiophylla*, *Pinus oocarpa* and *Pinus pringlei* plantations in Guareí, São Paulo, Brazil. *Southern Forests*, 81(4), 325–334. <https://doi.org/10.2989/20702620.2019.1636196>
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & F., L. (2023). e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. *R News*, 1–67. <https://cran.r-project.org/package=e1071>
- Nunes-Miranda, E., Groenner-Barbosa, B. H., Godinho-Silva, S. H., Ussi-Monti, C. A., Phin-Tng, D. Y., & Rezende-Gomide, L. (2022). Variable selection for estimating individual tree height using genetic algorithm and random forest. *Forest Ecology and Management*, 504(119828), 1–13. <https://doi.org/10.1016/j.foreco.2021.119828>
- Özçelik, R., Diamantopoulou, M. J., Crecente-campo, F., & Eler, U. (2013). Estimating Crimean juniper tree height using nonlinear regression and artificial neural network models. *Forest Ecology and Management*, 306, 52–60. <https://doi.org/10.1016/j.foreco.2013.06.009>
- Quiñonez-Barraza, G., & García-Espinoza, G. G. (2018). ¿Cómo corregir la heterocedasticidad y autocorrelación de residuales en modelos de ahusamiento y crecimiento en altura? How to correct the heteroscedasticity and autocorrelation of residuals in taper and height growth models?. *Revista Mexicana de Ciencias Forestales*, 9(49), 28–59. <https://doi.org/https://doi.org/10.29298/rmcf.v9i49.151>
- R Core Team. (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Şahin, A. (2024). Analyzing regression models and multi-layer artificial neural network models for estimating taper and tree volume in Crimean pine forests. *IForest*, 17, 36–44. <https://doi.org/10.3832/ifor4449-017>
- Sharma, R. P. (2009). Modelling height-diameter relationship for Chir pine trees. *Banko Janakari*, 19(2), 3–9. <https://doi.org/DOI:10.3126/banko.v19i2.2978>
- Temesgen, H., Zhang, C. H., & Zhao, X. H. (2014). Modelling tree height-diameter relationships in multi-species and multi-layered forests: A large observational study from Northeast China. *Forest Ecology and Management*, 316, 78–89. <https://doi.org/10.1016/j.foreco.2013.07.035>
- Zdenek, A., & Drápela, K. (2016). Comparison of parametric and nonparametric methods for modeling height-diameter relationships. *IForest*, 10, 1–8. <https://doi.org/10.3832/ifor1928-009>