2025 | 16·20
GIJÓN | JUNIO
9º CONGRESO FORESTAL ESPAÑOL

9CFE-1500

Actas del Noveno Congreso Forestal Español
Edita: **Sociedad Española de Ciencias Forestales. 2025**.
ISBN: **978-84-941695-7-1**

Organiza

Uso de datos LiDAR para generar inputs espacialmente explícitos de modelos de crecimiento de árbol individual.

MAURO GUTIERREZ F. (1), CANDEL PEREZ D. (2), GOMEZ ROUX M. (1), MANZANERA DE LA VEGA J. (3), GONZALEZ MESQUIDA J. (1), GONZÁLEZ-GARCÍA C. (3).
(1) iuFOR, EiFAB, Universidad de Valladolid, 42004 Soria, España
(2) Sección de Desarrollo Territorial Sostenible, NASUVINSA, Pamplona, España.
(3) Universidad Politécnica de Madrid. ETS de Ingeniería de Montes, Forestal y del Medio Natural

Resumen
La gestión forestal necesita de modelos que permitan anticipar la evolución de los bosques y los efectos de tratamientos alternativos que puedan plantearse en una determinada masa forestal. En la actualidad, existen en España modelos de crecimiento de árbol individual sensibles al clima que permiten satisfacer esta necesidad. Sin embargo, estas herramientas están infrautilizadas ya que no hay datos espacialmente explícitos que puedan usarse como inputs en estos modelos. En este trabajo hemos desarrollado modelos de imputación basados en datos LiDAR que permiten obtener datos espacialmente explícitos compatibles con los principales modelos de crecimiento de árbol individual disponibles para *Pinus sylvestris* en las provincias de Soria, Burgos y Palencia. Los modelos se han generado atendiendo a una priorización de seis variables respuesta realizada por un conjunto de profesionales del sector. Los modelos generados permiten obtener de forma simultánea predicciones para esas seis variables de interés, alcanzado coeficientes de determinación para la altura dominante y el volumen de 0.89 y 0.82 respectivamente. Además, los modelos de imputación seleccionados permiten reconstruir distribuciones diamétricas y listas de árboles que pueden emplearse para hacer proyecciones de crecimiento o para obtener inputs para simuladores de incendios.

Palabras clave
LiDAR, imputación, kNN, modelos de crecimiento de árbol individual, lista de árboles.

1. Introducción

La gestión forestal necesita de modelos que permitan anticipar la evolución de los bosques bajo un contexto de cambio global, su susceptibilidad a incendios forestales o los efectos de posibles tratamientos selvícolas alternativos que puedan plantearse en una determinada masa forestal. En la actualidad, existen en España modelos de crecimiento de árbol individual sensibles al clima integrados en simuladores como Medfate (DE CÁCERES et al., 2022) o SIMANFOR (BRAVO et al., 2012) y herramientas de predicción del comportamiento del fuego, que permiten satisfacer esta necesidad. Sin embargo, estas herramientas están infrautilizadas ya que salvo en Cataluña (PUKKALA et al., 2024), no hay datos espacialmente explícitos que puedan usarse como inputs en estos modelos.
Los inputs de este tipo de modelo reciben el nombre de tree-lists, listas de árboles, tabla de árboles o archivos de cohortes, y se caracterizan por su alto nivel de desagregación. Una lista de árboles es una tabla de longitud variable donde cada registro o fila representa un árbol prototipo con un determinado diámetro, altura y especie, que aparece en el territorio con una determinada frecuencia. La frecuencia se proporciona en un campo o columna de la tabla a través de un factor de expansión, que representa el número veces que se espera que arboles similares al prototipo aparezcan en una unidad de superficie. Desde un punto de vista del

usuario, una lista de árboles es análoga a la tabla de mediciones de un conjunto de parcelas medidas en un determinado rodal o área de estudio. Por tanto, empleando métodos bien conocidos, permiten calcular una gran cantidad de parámetros de interés forestal como áreas basimétricas o volúmenes totales, alturas dominantes u otros parámetros más desagregados como distribuciones diamétricas con anchos de clase arbitrarios.

La gran flexibilidad de las listas de árboles hace que su cartografía sea interesante, no solo por la posibilidad de usarlas como inputs en simuladores de dinámica forestal, sino también porque permiten generar mapas de parámetros derivados sin tener que desarrollar nuevos modelos cada vez que se quiere cartografiar una nueva variable. La generación de mapas de listas de árboles se ha realizado tradicionalmente empleando métodos de imputación de k vecinos más próximos (kNN). Éstas técnicas de aprendizaje automático se comenzaron a aplicar en problemas de cartografía de variables forestales por teledetección en Escandinavia y Norteamérica en la década de los 90 y de los 2000(MALTAMO & KANGAS, 1998; TEMESGEN et al., 2003; TOMPPO & KATILA, 1991). La generalización del uso de datos LiDAR en inventarios forestales reavivó el interés en estas técnicas ya que la información proporcionada por los sensores LiDAR permite que las listas de árboles predichas mediante kNN tengan un alto nivel de precisión y realismo (MAURO et al., 2019; PUKKALA et al., 2024).

Aunque actualmente no hay en España ninguna cartografía de listas de árboles, salvo en Cataluña; nos encontramos en un país con una información de teledetección y de campo de gran calidad. El Plan Nacional de Ortofotografía Aérea (PNOA) proporciona datos LiDAR multitemporales con cobertura nacional que puede complementarse con bases de datos globales de imágenes de satélite como las proporcionadas por Landsat, Sentinel 1 y Sentinel 2. Por otro lado, el Inventario Forestal Nacional IFN, recopila datos en decenas de miles de parcelas permanentes geolocalizas con precisión submétrica y distribuidas de forma “cuasi” sistemática por el territorio nacional. Estas piezas pueden combinarse para desarrollar modelos de imputación kNN que permitan generar mapas de listas de árboles para alimentar simuladores de dinámica facilitando así su uso.

Finalmente, un problema derivado del alto nivel de desagregación de las listas de árboles es que es necesario considerar múltiples respuestas de forma simultánea. La forma en que un conjunto de respuestas se pondera al seleccionar modelos de imputación para generar listas de árboles, puede tener un gran impacto en los resultados finales. Para solventar este problema, MAURO et al., (2019) propusieron interactuar con usuarios finales para obtener una ponderación apropiada de un las posibles respuestas que podrían considerarse al seleccionar modelos de predicción de listas de árboles. Este enfoque implica involucrar a los usuarios finales en la propia selección de modelos y se adecua a los métodos desarrollados por (MEDDENS et al., 2022) para facilitar el uso de productos generados por teledetección.

2. Objetivos

El objetivo del presente estudio es comparar distintos métodos de selección de modelos de imputación kNN para generar listas de árboles para masas de *Pinus sylvestris* en las provincias de Soria, Burgos y Palencia, empleando como información auxiliar datos LiDAR del PNOA. Entre los métodos comparados se analiza como incorporar las preferencias y necesidades de información de profesionales del sector en el proceso de selección de modelos de imputación.

3. Metodología

El área de estudio está constituida por el conjunto de polígonos del mapa forestal

de España (MFE) 1:25000 para Castilla y León (actualización 31/12/2021) de las provincias de Soria, Burgos y Palencia donde el Pino silvestre tiene más de un 60% de ocupación. En total, 9380 teselas que representan aproximadamente 262000 hectáreas lo que implica un tamaño medio de 28 hectáreas por tesela. La elevación varía entre 450 y 2006 m sobre el nivel del mar (Figura 1).

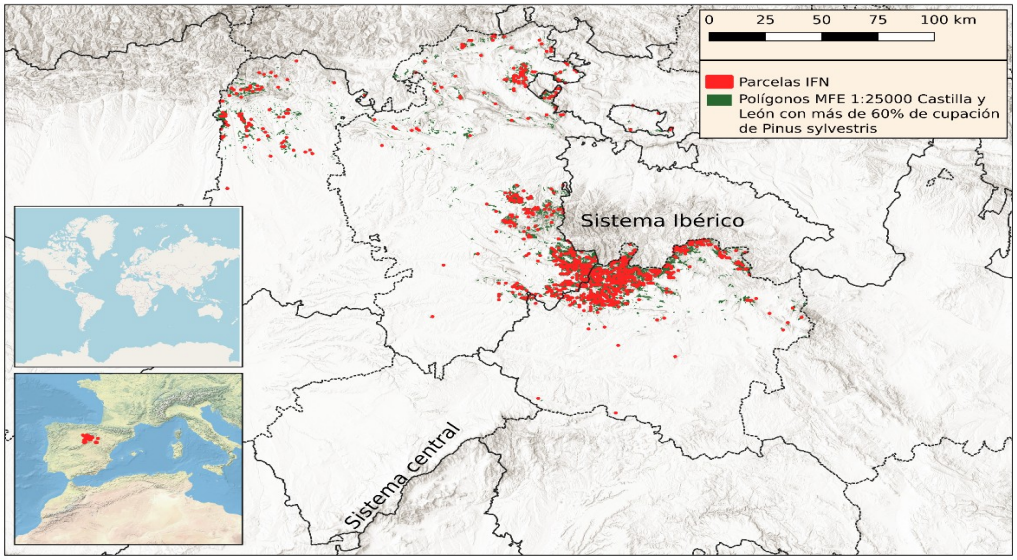

Figura 1. Área de estudio. Polígonos del mapa forestal con más del 60% de ocupación de *Pinus sylvestris* (verde) Parcelas IFN en dichos polígonos (rojo).

Observaciones de campo

Se seleccionaron las parcelas del inventario forestal nacional (IFN) incluidas en los polígonos seleccionados del MFE. En total, se contó con 795 parcelas con coordenadas GPS submétricas. Para cada una de las parcelas se calcularon el volumen total con corteza (VCC), el área basimétrica (G), la altura dominante (Ho), el número de pies por hectárea (N), la biomasa aérea total (B) y el diámetro cuadrático medio (Dg). En la Tabla 1 se muestra una síntesis de los valores de las variables respuestas en las parcelas.

Tabla 1. Síntesis de las parcelas empleadas para modelizar listas de árboles. VCC: Volumen con corteza por hectárea, G: el área basimétrica. Ho: altura dominante. N: número de pies por hectárea. B biomasa aérea total. DG diámetro cuadrático medio.

Variable	Mínimo	Media	Desviación típica	Máximo
VCC (m3/ha)	1.65	208.27	124.79	638.67
G (m^2/ha)	0.80	30.61	13.79	103.69
N(pies/ha)	5.09	671.28	505.95	4899.57
B (Mg/ha)	3.60	130.93	69.77	602.01
Ho (m)	3.10	15.89	4.47	31.01
Dg (cm)	8.39	26.89	8.51	59.38

Predictores LiDAR

Los datos LiDAR del PNOA2 fueron normalizados empleando la clasificación del suelo proporcionada por el PNOA utilizando el paquete lidR (ROUSSEL et al., 2020) de R. Una vez normalizados, se calcularon 60 métricas LiDAR descritas en HUDAK et al., (2020). Estas métricas incluyen, percentiles de la distribución de alturas de puntos LiDAR, porcentajes de cobertura para distintas alturas de corte y momentos

de la distribución de altura de puntos lidar (altura media, desviación estándar de la altura etc). Para cada parcela del IFN se obtuvieron las variables auxiliares anteriormente descritas. Las variables LiDAR en las parcelas se obtuvieron recortando las nubes de puntos del PNOA2 para un área circular de 16.93 m de radio de forma que la superficie de cómputo de métricas LiDAR, 900 m^2, fuese homogénea con la resolución espacial de Landsat, ya que en un futuro se pretenden desarrollar modelos de imputación que incluyan datos de series temporales de imágenes Landsat que tienen un tamaño de pixel de 900m^2.

Selección de modelos de imputación

Los modelos de kNN permiten obtener predicciones para un conjunto de parámetros arbitrario, es decir, todas las variables anteriormente descritas y otras adicionales pueden predecirse con un único modelo de imputación empleando los métodos descritos por PACKALÉN & MALTAMO, (2007) o MAURO et al., (2019). Sin embargo, la selección de modelos de kNN tradicionalmente se ha realizado minimizando algún tipo de métrica de bondad de ajuste que atiende a una única variable, o a lo sumo, a un conjunto de dos o tres variables respuesta que no son ponderadas por su posible importancia. En este estudio se siguió la propuesta de MAURO et al., (2019) y se empleó un conjunto de variables respuestas más amplio (seis en total) y se ponderaron las respuestas en base a la importancia que un conjunto de 45 profesionales del sector dieron dado a dichas variables empleando un cuestionario estandarizado. Con el fin de evaluar como cambian las predicciones de los métodos de imputación al emplear métodos de selección tradicionales basados en una única variable, se obtuvieron también modelos optimizados para cada una de las variables respuesta consideradas por separado. Finalmente, todos los modelos se compararon entre sí considerando cada una de las variables respuestas por separado.

Para la selección de variables predictoras y del número de vecinos usados en la imputación se empleó el método de eliminación sucesiva implementado en el paquete de R yaimpute (CROOKSTON & FINLEY, 2007). Como métrica de distancia para determinar los k vecinos más próximos se empleó la distancia random forest. Finalmente, el número de vecinos se permitió que variase entre 1 y 5 cinco. Cuando las variables se consideraron de forma individual, la métrica de error minimizada fue el error cuadrático medio (ECM) para la variable j, que puede expresarse como:

$$ECM_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{ij} - \hat{y}_{ij}\right)^2}$$

donde,

y_{ij} es el valor de dicha variable en la parcela i, e

$\hat{y}_{ij}$ es la predicción de dicha variable en la parcela i, y n es el número de parcelas.

Para el conjunto de todas las variables ponderadas, la métrica de error que se minimizó fue la distancia cuadrática media de Mahalanobis ponderada entre observaciones y predicciones (DCMMP) (KRUSIŃSKA & LIEBHART, 1987). Los factores de ponderación usados fueron los pesos medios que los 45 profesionales entrevistados asignaron a cada variable. La DCMMP se expresa como

$$DCMMP = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i)^t \boldsymbol{W}\boldsymbol{\Sigma}^{-1}\boldsymbol{W}^t(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i)\right]}$$

donde
$W = diag(w_1, w_2, \ldots, w_6)$
es una matriz con los pesos medios asignados a cada una de las seis variables consideradas en la diagonal y ceros en el resto de posiciones ,
y_i

e

$\hat{y}_i$

representan, respectivamente, los vectores de con las observaciones y predicciones realizadas para cada una de las seis variables en la parcela i ,
$\sum^{-1}$
es la inversa de la matriz de varianzas y covarianzas de las seis variables consideradas, y t indica transposición de vectores o matrices (Ecuación 2).Esta medida toma valores menores de uno cuando el modelo permite reducir la incertidumbre en la predicción de la variable o conjunto de variables consideradas por debajo del nivel de incertidumbre que se tendría si no existiese el modelo y sólo se tuvieran datos de campo. Finalmente, tanto para el conjunto de variables ponderadas como para las variables individuales se seleccionó el modelo con menor valor de ECM o DCMMP para comprobaciones posteriores.

Predicción de variables consideradas en la selección de modelos y distribuciones diamétricas.

Finalmente, el modelo seleccionado al considerar todas las variables con sus factores de ponderación fue empleado para obtener predicciones de las seis variables consideradas en un monte dominado por *Pinus sylvestris*. El monte seleccionado fue el monte 172 del catálogo de montes de utilidad pública, Pinar Grande.

4. Resultados

Los resultados de la ponderación de variables respuestas efectuados por los 45 profesionales forestales indicó que el número de pies por hectárea y la altura dominante tenían una importancia ligeramente superior al resto, a su vez el diámetro cuadrático medio y la biomasa total por hectárea fueron las variables que menos peso recibieron (Tabla 2).

Tabla 2. Importancia media escalada al intervalo 0-1 del que un conjunto variables respuestas consideradas según la evaluación que 45 profesionales del sector forestal han proporcionado a dichas variables. VCC: Volumen con corteza por hectárea, G: el área basimétrica. Ho: altura dominante. N: número de pies por hectárea. B biomasa aérea total. Dg diámetro cuadrático medio.

Variable	Peso proporcionado por usuarios a cada variable
VCC (m3/ha)	0.17
G (m2/ha)	0.17
N(pies/ha)	0.19
B (Mg/ha)	0.16
Ho (m)	0.18
Dg (cm)	0.13

El número de vecinos de los modelos seleccionados varió sustancialmente al seleccionar modelos kNN específicos para las distintas variables respuestas. Para las variables consideradas de forma aislada el valor de k osciló entre 1 y 5, sin mostrar patrones claros, siendo el valor óptimo de k igual a: 1 para modelos que

optimizan la predicción de Dg y G, 3 para el modelo de B, 4 para el modelo de N y 5 para los modelos que optimizan la predicción de VCC y Ho. Cuando todas las variables se consideraron de forma conjunta y ponderando cada variable por la importancia proporcionada por los profesionales entrevistados, el valor óptimo de k fue de 2. En todos los casos, los valores de DMMP fueron menores que 1 lo que indica que todos los modelos permiten hacer predicciones en las que se gana información con respecto a usar solo información de campo (Figura 2).

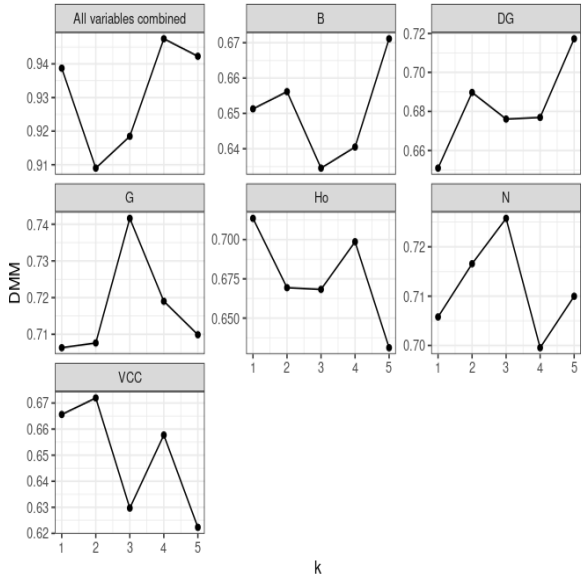

Figura 2. Distancia media de Mahalanobis entre observaciones y predicciones para los mejores modelos de imputación con uno a cinco vecinos. VCC: Volumen con corteza por hectárea, G: el área basimétrica. Ho: altura dominante. N: número de pies por hectárea. B biomasa aérea total. Dg diámetro cuadrático medio.

En la Figura 3, se presenta un subconjunto de las comparaciones que se pueden realizar entre los modelos seleccionados. El encabezado de cada panel indica la variable respuesta que se está considerando y debajo la variable o conjunto de variables que se han tenido en cuenta al hacer la selección del modelo de imputación. Cuando se comparan las predicciones de Dg, VCC y Ho obtenidas con los modelos que optimizaban el error para dichas respuestas, con las predicciones generadas por modelos alternativos en los que se optimiza una variable diferente, se puede observar que efectivamente hay un efecto de "*trade-off*" y que si una variable se optimiza es a expensas de empeorar la predicción de otras. Por otro lado, también se puede observar que el modelo obtenido para el conjunto de variables ponderadas nunca supone una gran pérdida con respecto al modelo obtenido minimizando el error para una única variable. En el caso del volumen, el modelo seleccionado considerando el conjunto de respuestas ponderadas es el mejor. Esto es un artefacto debido a que los métodos de selección de variables empleados no son exhaustivos (Figura 3).

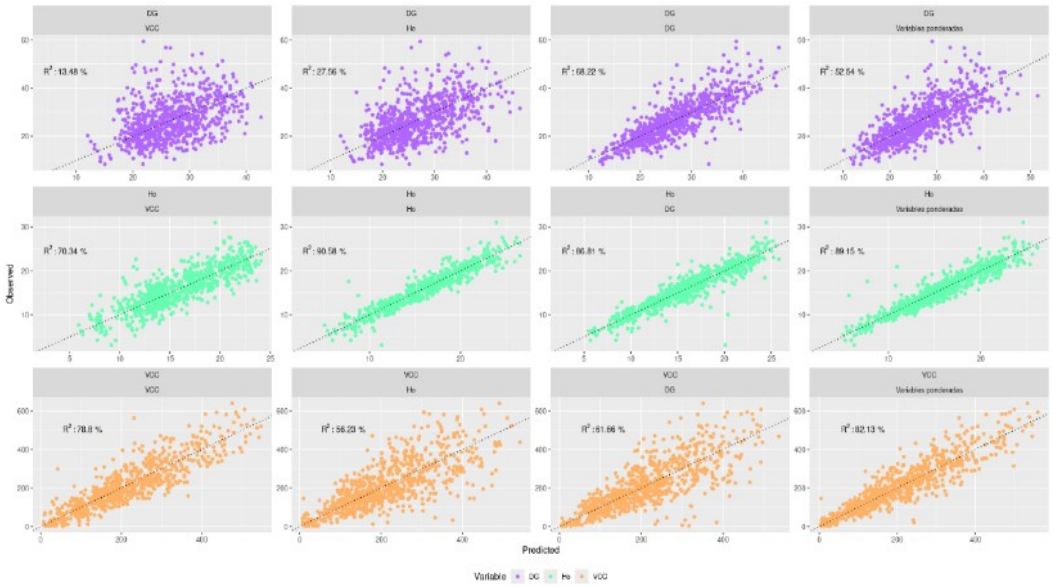

Figura 3. Diagramas observación predicción para volumen con corteza por hectárea VCC (naranja), altura dominante, Ho (verde) y diámetro cuadrático medio, Dg (morado) para los modelos de imputación obtenidos minimizando: los errores para VCC (primera columna), Ho (segunda columna), Dg (tercera columna) y minimizando la distancia de Mahalanobis media entre observaciones y predicciones ponderando todas las variables analizadas por la importancia dada por 45 profesionales del sector forestal (cuarta columna).

Al aplicar en el monte de Pinar Grande, el modelo obtenido considerando todas las variables de forma ponderada, podemos observar que para todas las variables se pueden diferenciar las unidades de gestión de dicho monte y que las predicciones en estas unidades son coherentes entre sí. Las zonas con menores valores de volumen y área basimétrica se corresponden con zonas con una menor altura dominante como cabría esperar. La densidad de pies por hectárea, N, es la variable que muestra una menor señal de las seis analizadas (Figura 4).

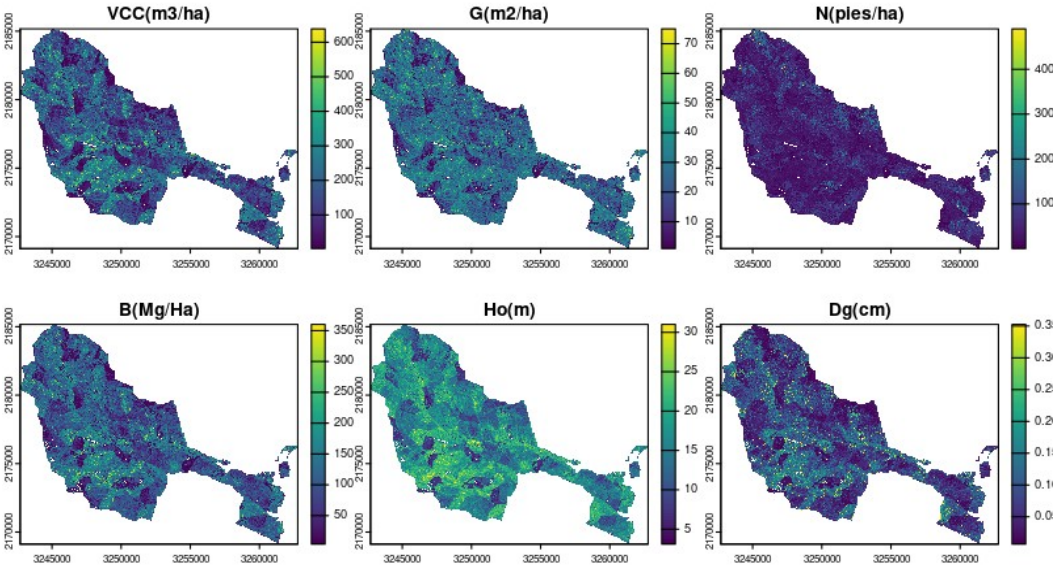

Figura 4. Predicciones con el modelo de imputación que pondera las variables respuesta VCC, G, B, B, Ho y Dg, de acuerdo con la importancia proporcionada a dichas variables por un conjunto de 45 profesionales forestales

entrevistados. VCC: Volumen con corteza por hectárea, G: el área basimétrica. Ho: altura dominante. N: número de pies por hectárea. B biomasa aérea total. Dg diámetro cuadrático medio.

Como ejemplo, en la Figura 5 se muestran funciones de densidad para el diámetro normal construidas a partir de las listas de árboles generadas por el modelo seleccionado para el conjunto de las seis variables. Las distribuciones de diámetros han sido suavizadas utilizando un kernel gaussiano en cada rodal. Se muestran un rodal en con regenerado (GIS-ID 18013), un rodal en estado de latizal (GIS-ID 18018) y un rodal en estado de fustal (GIS-ID 1808). Estas funciones de densidad se transforman en distribuciones diamétricas que proporcionen número de pies por hectárea multiplicando por el valor de la curva por el valor medio de N en el rodal.

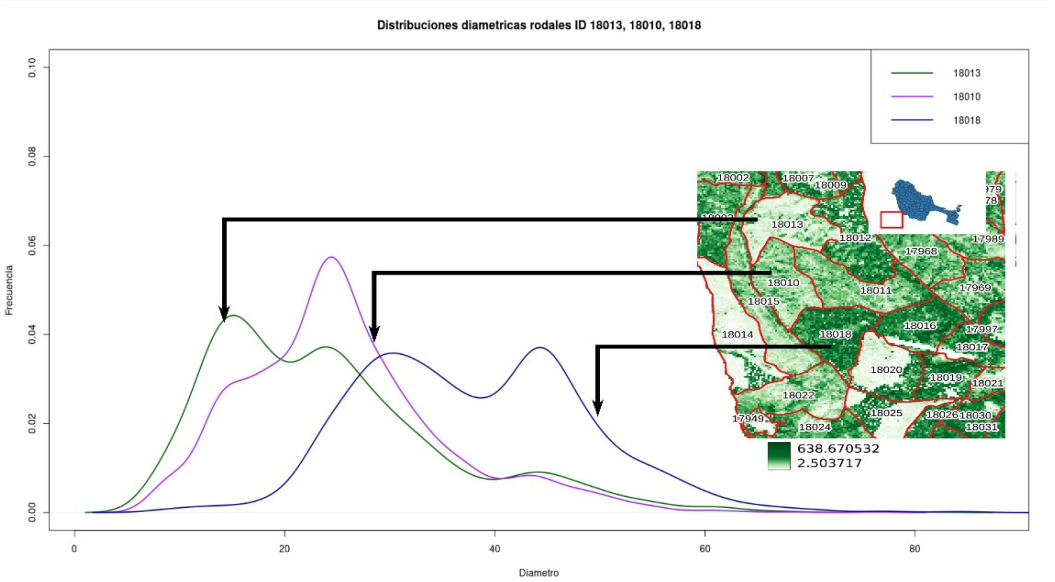

Figura 5. Funciones de densidad para el diámetro normal para un rodal de regenerando (18013, verde), un rodal de latizal (18010, morado) y un rodal en estado de fustal (18018, azul). La integral de estas distribuciones por el número de pies del rodal proporciona la distribución diamétrica.

5. Discusión

El presente estudio analiza distintos métodos para generar listas de árboles con las que alimentar modelos de simulación de dinámica forestal. Una de las principales aportaciones de este trabajo es que el proceso de selección de modelos de imputación kNN está basado en las necesidades expresadas por distintos profesionales del sector forestal. Es decir, implica a potenciales usuarios de los mapas generados en la propia selección de modelos de predicción. Actualmente, se es consciente de que existe cierta desconexión entre los productores de cartografía de variables de interés forestal mediante teledetección y los usuarios finales de dichas cartografías. Dicha desconexión resulta en una baja tasa de adopción de dichas cartografías por parte de profesionales del sector. Para resolver este problema, se ha propuesto aumentar la interacción con usuarios finales de dichas cartografías para conocer mejor sus necesidades y adaptar de acuerdo a estas los mapas generados (MEDDENS et al., 2022).

En el presente trabajo se ha adoptado la filosofía de trabajo propuesta por MEDDENS et al., (2022) y se ha comprobado que al integrar múltiples variables en el proceso de selección de modelos de imputación kNN, ponderándolas de acuerdo

con la importancia que los usuarios les han dado, es posible obtener modelos de predicción de gran calidad. Al ponderar las variables respuesta por la importancia que les han dado los profesionales entrevistados, se llega a modelos de imputación que proporcionan predicciones muy cercanas a las que se obtienen enfocando la selección de modelos a una única respuesta. Aunque es inevitable que el modelo kNN obtenido al considerar todas las variables pierda cierto nivel de precisión, las perdidas observadas han sido menores y los outputs obtenidos han mostrado un alto nivel de coherencia para múltiples atributos.

Uno de los problemas detectados en el presente estudio es que los métodos de selección de variables por eliminación sucesiva que tradicionalmente se han empleado en problemas de imputación exploran un conjunto de modelos alternativos sumamente reducido, (60x59)/2, en nuestro caso. Este número es ínfimo comparado con el número de modelos alternativos que pueden realizarse con 60 predictores LiDAR que es 2^{60}-1. Esto claramente indica que es necesario recurrir a técnicas heurísticas que mejoren los procesos de selección de variables para métodos de imputación. En este sentido, trabajos preliminares con algoritmos genéticos han mostrado resultados prometedores y trabajos existentes han demostrado que métodos de búsqueda basadas “simulated annealing” (PUKKALA et al., 2024) también permiten hacer una búsqueda eficiente de modelos de imputación.

En este trabajo se han analizado variables medias o agregadas. En general las técnicas de imputación son consistentes cuando se consideran variables aditivas, pero no todas las variables de interés para gestores forestales son de este tipo. Futuros estudios deben analizar que ocurre al emplear técnicas de imputación para otro tipo de variables como índices de biodiversidad. Finalmente, estos mapas de listas de árboles son los primeros mapas de este tipo desarrollados en Castilla y León y se espera que faciliten el uso de herramientas de simulación de dinámica forestal en la región.

6. Conclusiones

La principal conclusión de este trabajo es que considerar de forma conjunta múltiples variables respuestas, ponderándolas por la importancia que tienen para usuarios del sector forestal, permite generar listas de árboles realistas que predicen con gran nivel de precisión múltiples atributos de interés forestal y cuyos outputs pueden integrarse en modelo de crecimiento de árbol individual independientes de la distancia.

7. Agradecimientos

Los autores agradecen la financiación recibida por parte del Ministerio de Ciencia e Innovación para el desarrollo del proyecto TLM-PROJECT (PID2022-140104OA-I00). Francisco Mauro recibió una ayuda María Zambrano financiada por la Universidad de Valladolid. David Candel-Pérez recibió una ayuda postdoctoral (CONTPO-2021-105) financiada por la Universidad de Valladolid. Manuel Gomez-Roux disfruta de una ayuda predoctoral FPI (CONTFPI-2023-91) asociada al proyecto TLM-PROJECT.

8. Bibliografía

BRAVO, F.; RODRIGUEZ, F.; ORDÓÑEZ, C.; 2012. A web-based application to simulate alternatives for sustainable forest management: SIMANFOR. *Forest Systems*, *21*(1), 4. https://doi.org/10.5424/fs/2112211-01953

CROOKSTON, N. L.; FINLEY, A. O.; 2007. yaImpute: An R Package for kNN Imputation. *Journal of Statistical Software*, *23*(10). http://www.jstatsoft.org/v23/i10

DE CÁCERES, M.; MOLOWNY-HORAS, R.; CABON, A.; MARTÍNEZ-VILALTA, J.;

MENCUCCINI, M.; GARCÍA-VALDÉS, R.; NADAL-SALA, D.; SABATÉ, S.; MARTIN-STPAUL, N.; MORIN, X.; BATLLORI, E.; AMÉZTEGUI, A.; 2022. MEDFATE 2.8.1: A trait-enabled model to simulate Mediterranean forest function and dynamics at regional scales. *Geoscientific Model Development Discussions*, *2022*, 1–52. https://doi.org/10.5194/gmd-2022-243

HUDAK, A. T.; FEKETY, P. A.; KANE, V. R.; KENNEDY, R. E.; FILIPPELLI, S. K.; FALKOWSKI, M. J.; TINKHAM, W. T.; SMITH, A. M. S.; CROOKSTON, N. L.; DOMKE, G. M.; CORRAO, M. V.; BRIGHT, B. C.; CHURCHILL, D. J.; GOULD, P. J.; MCGAUGHEY, R. J.; KANE, J. T.; DONG, J.; 2020. A carbon monitoring system for mapping regional, annual aboveground biomass across the northwestern USA. *Environmental Research Letters*, *15*(9), 095003. https://doi.org/10.1088/1748-9326/ab93f9

KRUSIŃSKA, E.; LIEBHART, J.; 1987. Objective evaluation of degree of illness with the weighted Mahalanobis distance. A study for patients suffering from chronic obturative lung disease. *Computers in Biology and Medicine*, *17*(5), 321–329. https://doi.org/10.1016/0010-4825(87)90021-7

MALTAMO, M.; KANGAS, A.; 1998. Methods based on k-nearest neighbor regression in the prediction of basal area diameter distribution. *Canadian Journal of Forest Research*, *28*(8), 1107–1115.

MAURO, F.; FRANK, B.; MONLEON, V. J.; TEMESGEN, H.; FORD, K. R.; 2019. Prediction of diameter distributions and tree-lists in southwestern Oregon using LiDAR and stand-level auxiliary information. *Canadian Journal of Forest Research*, *49*(7), 775–787.

MEDDENS, A. J.; STEEN-ADAMS, M. M.; HUDAK, A. T.; MAURO, F.; BYASSEE, P. M.; STRUNK, J.; 2022. Specifying geospatial data product characteristics for forest and fuel management applications. *Environmental Research Letters*, *17*(4), 045025.

PACKALÉN, P.; MALTAMO, M.; 2007. The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. *Remote Sensing of Environment*, *109*(3), 328–341.

PUKKALA, T.; AQUILUÉ, N.; JUST, A.; CORBERA, J.; TRASOBARES, A.; 2024. Developing kNN forest data imputation for Catalonia. *Journal of Forestry Research*, *35*(1), 80. https://doi.org/10.1007/s11676-024-01735-5

ROUSSEL, J.-R.; COOPS, N. C.; TOMPALSKI, P.; GOODBODY, T. R. H.; MEADOR, A. S.; BOURDON, J.-F.; BOISSIEU, F. de; ACHIM, A.; 2020. lidR: An R package for analysis of Airborne Laser Scanning (ALS) data. *Remote Sensing of Environment*, *251*, 112061. https://doi.org/10.1016/j.rse.2020.112061

TEMESGEN, B. H.; LEMAY, V. M.; FROESE, K. L.; MARSHALL, P. L.; 2003. Imputing tree-lists from aerial attributes for complex stands of south-eastern British Columbia. *Forest Ecology and Management*, *177*(1–3), 277–285. https://doi.org/10.1016/S0378-1127(02)00321-3

TOMPPO, E.; KATILA, M.; 1991. Satellite image-based national forest inventory of Finland. *International Archives of Photogrammetry and Remote Sensing*, *28*(7–1), 419–424.