2025 | 16·20
GIJÓN | JUNIO

9º CONGRESO **FORESTAL** ESPAÑOL

9CFE-1444

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

Organiza

Sociedad Española de Ciencias Forestales
SECF

Valorización de parcelas del IFN fuera de España: estimación de biomasa aérea mediante la integración de datos locales y teledetección en el Bosque Nacional de Bastard, Pau (Francia)

NAVARRO FERNÁNDEZ, J.A. (1), ESTEBAN CAVA, J. (1), FIORE, J. (2), FERRY, J.N. (2) y TOMÉ MORÁN, J.L. (1)

1. Agresta Sociedad Cooperativa
2. TotalEnergies S.E.

Resumen

Disponer de datos precisos de biomasa forestal es crucial debido a su papel en la crisis climática y los mercados de carbono. Los altos costos de los inventarios suelen ser una limitación, pero integrar datos externos al área de interés puede reducirlos, utilizando métodos robustos y precisos que consideren la incertidumbre asociada al uso de datos dispares y la complejidad de los bosques. Este estudio analiza el uso de diferentes datos de sensores remotos como LiDAR aerotransportado, Landsat, Sentinel-2 o GEDI, para estimar la biomasa aérea mediante el ajuste de modelos kNN y Random Forest. Se utilizaron dos conjuntos de datos: una muestra local reducida y otra externa del IFN español. La calidad de modelos se evaluó mediante el error estándar (SE) de las estimaciones poblacionales obtenidas por *bootstrapping*. Usando datos locales reducidos como muestra de entrenamiento, los mejores resultados se obtuvieron con modelos Random Forest (SE=4,22%), mientras que kNN fue más consistente en la integración de datos externos con parcelas locales (SE=3,76%). Los resultados demuestran la posibilidad de integrar datos externos cuando los locales son limitados. No obstante, es importante tener un conocimiento exhaustivo de los métodos de inferencia aplicados para generar estimaciones robustas.

Palabras clave

Model-based, Inventario Forestal Nacional, incertidumbre, Sentinel, LiDAR.

1. Introducción

Los datos LiDAR aerotransportados (ALS, por sus siglas en inglés Airborne Laser Scanner) han demostrado ser valiosos para caracterizar la estructura de los bosques, incluidas las copas de varios estratos. La capacidad de penetración en el dosel permite al ALS capturar datos no sólo de la parte superior de éste, sino también del interior del propio dosel, proporcionando una visión más completa de la estructura del bosque y de la distribución de la biomasa aérea (AGB) (White et al., 2013). Gracias a esta característica, los datos ALS suelen conducir a estimaciones más precisas de la biomasa aérea (Holopainen et al., 2015). Sin embargo, la idea de disponer de datos ALS gratuitos en todo el mundo es probablemente poco realista, ya que su coste de adquisición puede resultar prohibitivo para aplicaciones a gran escala (Luther et al., 2019; Shang et al., 2019). Dadas estas limitaciones, es necesario explorar alternativas para mitigar los costes de adquisición de datos remotos y desarrollar inventarios forestales basados en

tecnologías más accesibles.

Los Inventarios Forestales Nacionales (IFN) se llevan a cabo para recopilar datos estadísticos sobre los bosques a escala nacional e internacional. Debido al extenso número de parcelas y variables que se analizan, representan la principal fuente de información forestal que proporciona información estandarizada y exhaustiva sobre las características estructurales, la composición y los volúmenes de biomasa forestal, lo que las convierte en un componente esencial para la elaboración de políticas relacionadas con los bosques (Alberdi, 2015; Fridman et al., 2014; Tomppo et al., 2010). Sin embargo, un desafío importante radica en la transferencia de datos del IFN, originalmente diseñados para un contexto específico, a aplicaciones locales en áreas con características ecológicas o de manejo distintas.

El uso de datos del IFN en lugares diferentes a los de su origen genera interrogantes sobre su representatividad y aplicabilidad. Las variaciones en la topografía, la composición forestal, las condiciones climáticas y las prácticas de manejo pueden introducir sesgos en las estimaciones que se basan en estos datos. Por tanto, es esencial estudiar métodos para adaptar y complementar los datos del IFN con datos locales con el fin de garantizar predicciones precisas, robustas e insesgadas. De esta forma, desarrollar modelos transferibles a partir de datos IFN que puedan aplicarse en entornos locales o incluso en otros territorios podría resultar muy relevante para los propietarios y gestores forestales, ya que disminuye la necesidad de recopilar datos de campo (Fekety et al., 2018; Tompalski et al., 2019). Sin embargo, la transferibilidad directa de los modelos no está asegurada, ya que, incluso aunque se utilice el mismo tipo de sensor y se estudien las mismas especies y su estructura, las variaciones en la configuración del vuelo o en los parámetros del sensor influyen en las características finales de los datos (Navarro Fernández, 2019).

Por otra parte, las imágenes de satélite y las tecnologías de teledetección ofrecen alternativas rentables para estimar la AGB en los bosques. Las imágenes de satélite de media resolución, como Landsat o Sentinel son opciones gratuitas y con una alta resolución temporal para estudiar la AGB, pero pueden carecer de la precisión de las mediciones de campo o de las técnicas de teledetección de alta resolución como el ALS. Sin embargo, algunos estudios han demostrado que la integración de diferentes sensores puede mejorar la precisión de las estimaciones (Campbell et al., 2021; Heiskanen et al., 2019).

La combinación de la información obtenida de inventarios forestales y los datos de sensores remotos, requiere de técnicas de modelización y métodos de inferencia robustos que permitan integrar diferentes resoluciones espaciales y temporales para generar estimaciones precisas y operativas. Por otra parte, muchas zonas se sitúan en áreas remotas y de difícil acceso, donde recopilar datos de inventarios forestales resulta costoso y difícil, por lo que resulta fundamental seleccionar enfoques de modelización que permitan usar de manera eficiente datos disponibles como los de los IFN y proporcionen estimaciones precisas sin necesidad de realizar un muestreo extensivo en campo.

En este estudio se han probado algoritmos de aprendizaje automático como k-Nearest Neighbors (kNN) y Random Forest (RF). Se han elegido estos dos tipos de modelo debido

a que kNN permite preservar la variabilidad natural de los datos y adaptarse a patrones locales, mientras que RF, aunque es robusto al ruido y las interacciones complejas entre variables predictoras y puede manejar grandes conjuntos de datos con interacciones variables complejas, puede ser sensible a pequeños cambios en los datos de entrenamiento y propenso al sobreajuste (Boateng et al., 2020; Li et al., 2011). Además. en este estudio, se han explorado combinaciones de datos del IFN español, de parcelas locales y métricas derivadas de diferentes sensores remotos como ALS, Landsat, Sentinel-2 y GEDI, evaluando cómo los diferentes métodos y bases de datos afectan la precisión de las estimaciones de biomasa.

2. Objetivos

Los objetivos de este estudio son 1) evaluar cómo afecta la combinación de datos del IFN y datos locales, junto con métricas de sensores remotos (ALS, Landsat, Sentinel-2 y GEDI), a la precisión de las estimaciones de AGB utilizando diferentes tipos de modelo; y 2) analizar la capacidad de los datos del IFN para ser adaptados a contextos locales y su impacto en la incertidumbre de las estimaciones.

3. Metodología

Área de estudio

La zona de estudio se sitúa al norte de Pau (Francia), concretamente en el Forêt de Bastard, con una superficie aproximada de 300 hectáreas. Se trata de un bosque formado por rodales gestionados en los que las especies forestales dominantes son: *Quercus robur*, *Castanea sativa* y *Quercus rubra*.

Datos de campo

Se obtuvieron datos de dos muestreos diferentes. Por un lado, se realizó una selección de parcelas del IFN4 del norte de España de masas más o menos puras de las principales especies presentes en el área de estudio (*Castanea sativa*, N=45; *Pinus sylvestris*, N=276; *Quercus petraea*, N=50; *Quercus robur*, N=42; *Quercus rubra*, N=11).

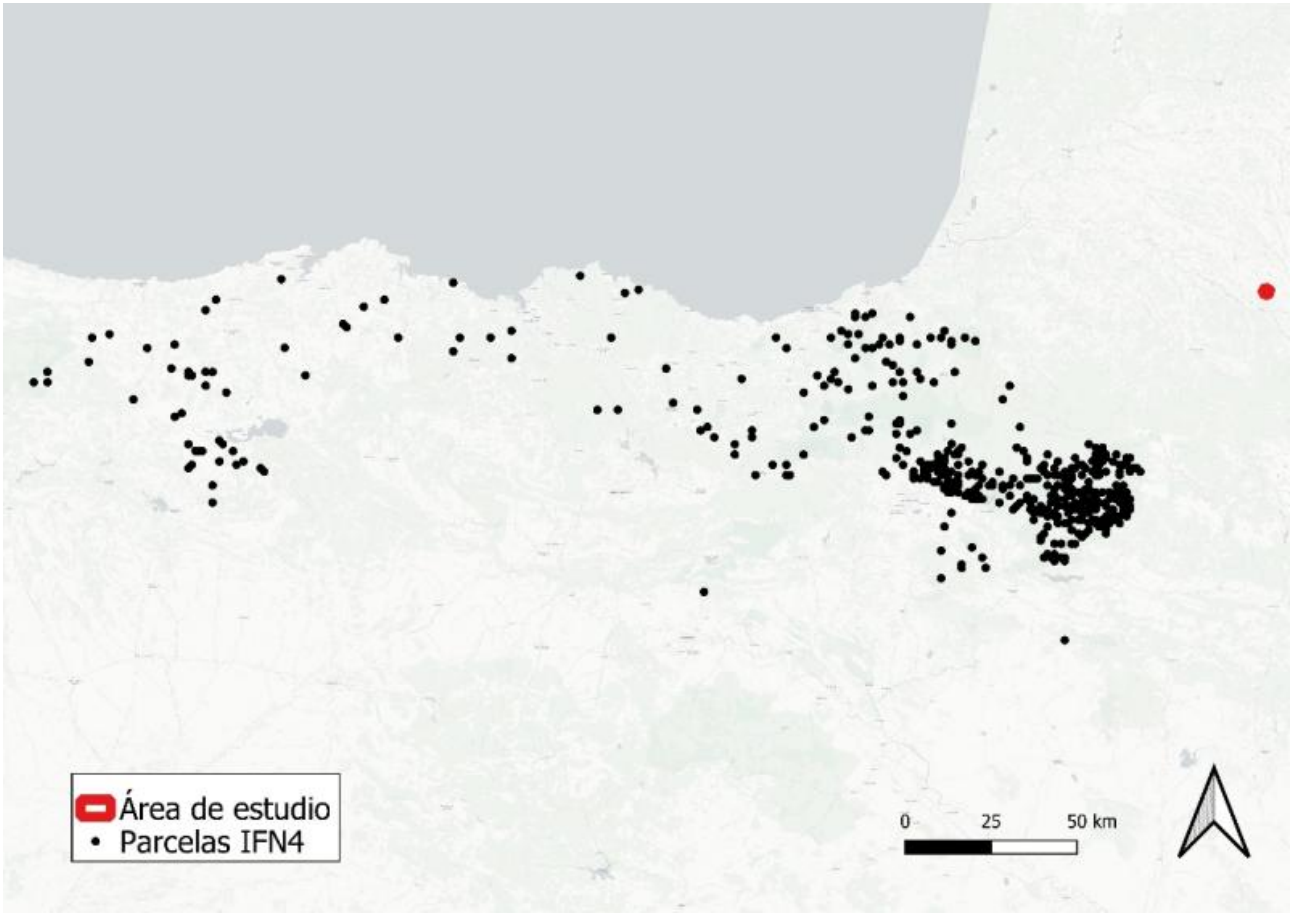

Figura 1. Localización de las parcelas del IFN4 utilizadas y del área de estudio.

La AGB de las parcelas de muestreo se estimó aplicando factores de expansión de biomasa (BEF) desarrollados por el Centro de Investigación Ecológica y Aplicaciones Forestales (CREAF) y utilizados en el Sistema Español de Inventario y Proyecciones de Emisiones de Gases de Efecto Invernadero a la Atmósfera (SEI).

Tabla 1. Métricas descriptivas de la AGB (t/ha) en las parcelas de muestreo seleccionadas del IFN4.

Estrato	Mínimo	Máximo	Media	Deviación estándar
Castanea sativa	36,54	276,22	141,92	60,43
Pinus sylvestris	7,24	520,09	138,27	92,00
Quercus petraea	41,57	724,86	231,02	155,49
Quercus robur	7,03	257,87	112,34	64,18
Quercus rubra	16,73	362,73	151,25	104,57

Por otra parte, en octubre de 2023 se obtuvieron datos de campo de 39 parcelas de muestreo circulares con un radio de 11,3m con precisión horizontal submétrica (0,4 m), tratando de abarcar la mayor variabilidad posible en cuanto a formaciones forestales. En estas parcelas se registró el diámetro a la altura del pecho (*d*) y la especie de todos los árboles con un *d* >7,5 cm, y se midió la altura de cuatro árboles seleccionados al azar por parcela. Las alturas de los árboles restantes de cada parcela se predijeron utilizando un modelo generalizado altura-diámetro basado en las alturas de los árboles medidas en todas las parcelas de campo. El volumen del tronco se estimó para cada árbol utilizando los modelos alométricos desarrollados por (Deleuze et al., 2014) para diferentes especies de árboles en Francia. Para calcular la AGB a partir del volumen del tronco, se utilizaron los factores de expansión de volumen generados por (Longuetaud et al., 2013) y los valores de densidad de la madera de (Zanne et al., 2009):

$$V_{tot} = VEF \cdot V_{tronco}$$

$$AGB = V_{tot} \cdot \rho$$

Donde V_{tot} es el volumen total del árbol (m^3), V_{tronco} es el volumen del tronco (m^3), *VEF* es el factor de expansión de volumen (adimensional), AGB es la biomasa aérea (toneladas de materia seca) y es la densidad de la madera (g/cm^3).

Figura 2. Localización de las parcelas de muestreo en el área de estudio.

Tabla 2. Métricas descriptivas de la AGB (t/ha) en las parcelas de muestreo dentro del área de estudio.

Estrato	Área (ha)	N parcelas	Mínimo	Máximo	Media	Desviación estándar
C. sativa	50,32	7	77.97	319.26	200.05	79.71
P. sylvestris	23,32	5	113.49	160.98	129.21	18.69
Q. robur	169,68	17	16.92	433.44	249.01	108.96
Q. rubra	34,29	5	118.32	404.89	281.55	112.86
Q. petraea	14,31	5	2.48	50.44	30.94	19.24
Otras especies	19,52	-	-	-	-	-
Total	311,44	39	2.48	433.44	203,41	117,47

Finalmente, a partir de las dos muestras (IFN4 y local), se crearon dos bases de datos independientes. La primera incluyó una muestra balanceada de parcelas del IFN4, utilizando un máximo de 60 parcelas por especie con el fin de que la mayor proporción de alguna especie no afecte a los resultados de los modelos. La segunda muestra se generó combinando las parcelas del IFN4 y las parcelas locales en una única base de datos, asegurando que todas las parcelas locales estuvieran representadas en la muestra combinada. Para garantizar una representación equilibrada de todas las especies en la muestra, se seleccionaron 52 parcelas por especie en total, excepto para el caso de *Q. robur* y *Q. rubra*, que se agruparon en un único conjunto de 52 parcelas. Además, en ambas bases de datos se eliminaron las parcelas de roble del IFN4 con una densidad mayor a 1.000 pies/ha para conseguir muestra más representativa de las características locales.

Datos remotos

1. ALS PNOA: Para las parcelas del IFN se usó la cobertura LiDAR del PNOA más cercana a la fecha de campo de las parcelas del IFN de cada provincia, con densidades que oscilan entre 0,5 y 2 puntos/m^2. Para el área de estudio se utilizaron datos ALS del IGN francés correspondientes a la campaña LIDAR HD France con una densidad media de 34,21 puntos/m2. El procesado de los datos ALS se llevó a cabo utilizando software propietario desarrollado por Agresta S. Coop. y el paquete LidR (Roussel et al., 2020). El flujo de trabajo es el siguiente:

    1. Generación de modelos digitales de elevación (MDE).

    2. Normalización de los retornos láser clasificados como vegetación.

    3. Cálculo de estadísticos de la estructura de la vegetación.

    4. Generación de capas y bases de datos georreferenciadas con todas los estadísticos ALS con una resolución de 20 m.

2. Landsat: Se utilizaron imágenes Landsat-5 y Landsat-7 de los meses de mayo a octubre de los mismos años del IFN-4 para obtener el valor de las bandas y calcular diferentes índices de vegetación en las parcelas: índice de vegetación de diferencia normalizada (*ndvi*), índice de quemado normalizado (*nbr*), índice de humedad diferencial normalizado (*ndmi*), e índice de vegetación mejorado (*evi*). También se creó un compuesto anual de verano para el área de estudio a partir de imágenes de los mismos meses del año 2023, el cual se resampleó a la resolución de 20 m de los estadísticos ALS usando el método del vecino más próximo.

3. Sentinel-2: Debido a la fecha de lanzamiento del sensor posterior a las parcelas del IFN-4, sólo se utilizaron datos de Sentinel-2 para el área de estudio. La fecha de las imágenes, el método de resampleado a 20 m y los índices generados fueron los mismos que para el caso de Landsat. Para garantizar la coherencia espectral, se evitó utilizar las bandas espectrales de la región del *Vegetation Red Edge*.

4. Sentinel-1: Se utilizaron datos de primavera y verano de 2023 resampleados a la resolución de los estadísticos ALS para generar un modelo de altura de vegetación a partir de métricas de GEDI. Los productos de Sentinel-1 se preprocesaron con Sentinel-1 Toolbox siguiendo estos pasos: (i) eliminación de ruido térmico, (ii) calibración radiométrica, (iii) corrección del terreno basada en el modelo de elevación digital SRTM (resolución de 3 segundos) y (iv) conversión a decibelios (dB) mediante escala logarítmica (10*log10(x)).

5. GEDI: Los datos de GEDI se utilizaron para calibrar un modelo predictivo de alturas de vegetación basado en variables de Sentinel-1 y Sentinel-2, empleando la métrica *RH95* (altura relativa al 95% de energía). Para mitigar el error de geolocalización de las huellas de GEDI, se excluyeron las huellas cuyo *RH95* difería en ±2 m del percentil 95 de altura ALS (*zq95*), quedando 218 muestras. Para cada huella, se extrajeron y promediaron los valores de Sentinel-1 y Sentinel-2. Se ajustó un modelo del tipo *Support Vector Regression* (SVR) utilizando un núcleo de función de base radial (RBF) y una función de pérdida ε-insensible de Vapnik. El modelo se optimizó ajustando los parámetros clave mediante un método de búsqueda en malla (grid search) y se evaluó su calidad mediante validación cruzada con 10 particiones (10-fold cross-validation). La metodología permitió generar un mapa de alturas de vegetación a 20 m de resolución. Los resultados de la validación cruzada del modelo arrojaron un R^2 de 0,55, un RMSE de 4,25 m (21,35%) y un sesgo de -0,36 m (-1,68%).

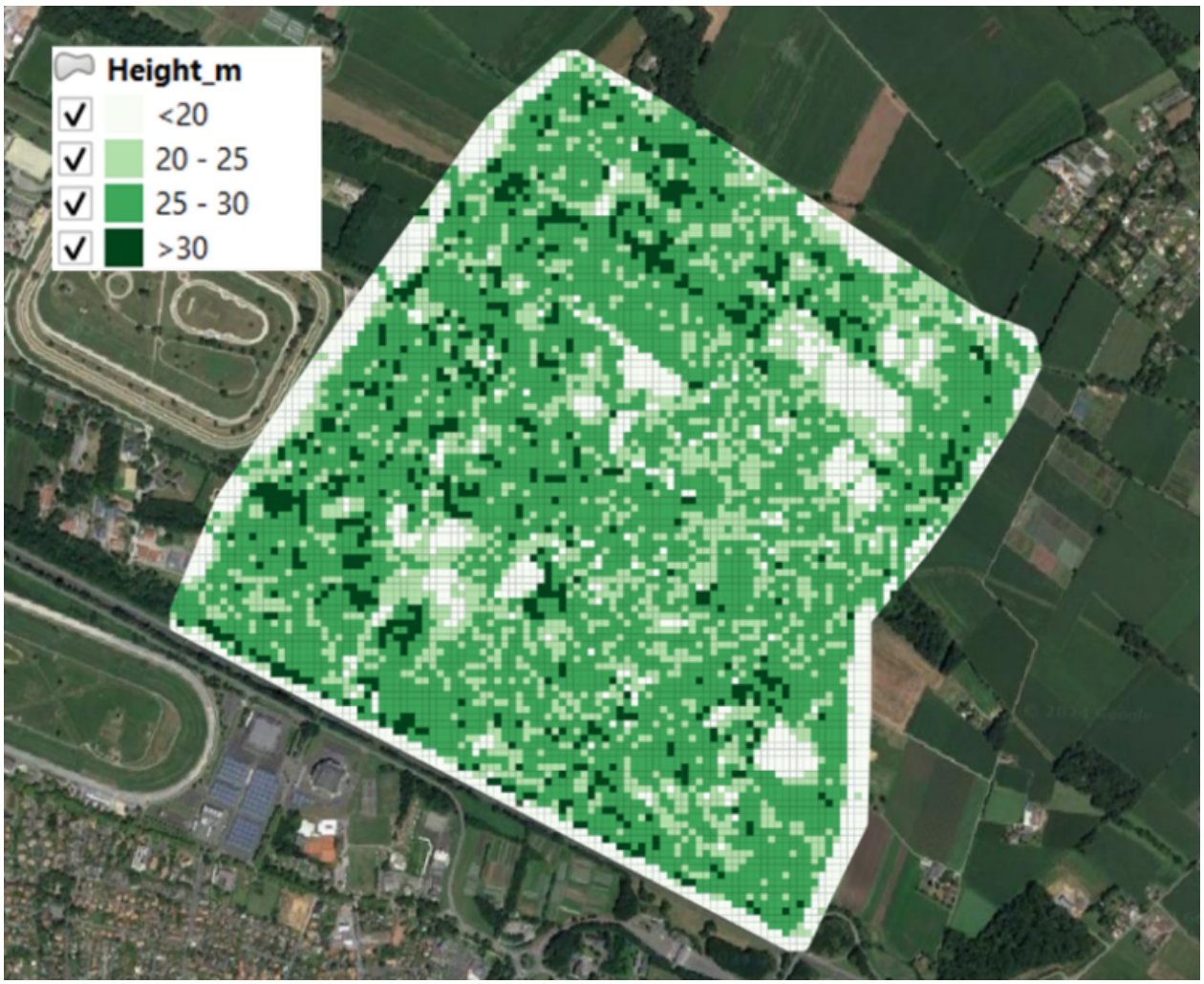

Figura 3. Mapa de alturas de vegetación modeladas a partir de GEDI, Sentinel-1 y Sentinel-2.

Modelos de AGB

Se ajustó una batería de modelos utilizando diferentes combinaciones de datos de entrada y técnicas de aprendizaje automático. Así, una parte de los modelos ajustados fueron calibrados con datos de campo del área de estudio, es decir, con un set reducido de parcelas locales. Mientras, el resto de modelos utilizaron como datos de entrenamiento un set de parcelas que combinaba las parcelas locales con una muestra de parcelas del IFN4 para evaluar si una base de datos más amplia que incluya información externa puede mejorar los resultados de la estimación de AGB y cómo responden las diferentes técnicas de modelización a datos heterogéneos procedentes de distintas fuentes con patrones o tendencias variables. No se ajustaron modelos diferentes para cada especie presente en la muestra de entrenamiento, sino que los modelos utilizaron la muestra completa para predecir la AGB de toda el área de estudio.

Ya que el uso de datos muy dispares puede afectar a la sensibilidad y el desempeño de los modelos, se consideró esencial analizar la capacidad de diversas técnicas de aprendizaje automático como RF y kNN en el manejo de los diferentes datos de entrada.

Las diferentes combinaciones de datos utilizadas para construir los modelos quedan recogidas en la tabla 3:

Tabla 3. Tipos de modelo y fuentes de datos utilizadas.

Modelo	Muestra	LiDAR	Landsat	*zq95*	*RH95*
Random Forest	IFN4 + Local	?	?	?	-
kNN	IFN4 + Local	?	?	?	-
Random Forest	Local	?	?	?	?
kNN	Local	?	?	?	?

Se eligió usar modelos RF por ser una de las técnicas más comunes en la estimación de variables forestales a partir de datos remotos (Ahmed et al., 2015; Esteban et al., 2019; Gleason & Im, 2012). Los modelos RF se ajustaron utilizando el paquete *randomForest* (Liaw & Wiener, 2002). Aunque se trata de un método estadístico que funciona bien con un gran número de variables predictoras, incluso con cierta autocorrelación, para mejorar la robustez de los modelos y evitar el sobreajuste, se utilizó la metodología de selección optimizada de variables predictoras desarrollada en el paquete *VSURF* (Genuer et al., 2015). Por su parte, el método kNN fue seleccionado por su capacidad de adaptarse a bases de datos heterogéneas y su sensibilidad a la estructura de los datos, lo que lo hace adecuado para abordar los desafíos asociados con la integración de datos de diferentes fuentes (Ali et al., 2019; Rani, 2017). Sin embargo, la elección de k y de la métrica de la distancia puede influir significativamente en la capacidad del modelo para captar las relaciones entre las parcelas del proyecto y del IFN4, garantizando una integración eficaz de los datos y unos resultados de regresión coherentes para las variables de interés. Así, se ajustaron distintos modelos kNN con diferentes valores de *k* y métricas de distancia, como triangular, euclídea, Mahalanobis, Triweight o Epanechnikov, seleccionando el mejor modelo para cada combinación de datos de entrada.

La bondad de ajuste de los modelos ajustados se evaluó a nivel de parcela mediante validación cruzada (10 particiones repetidas 10 veces), estimando el sesgo y error cuadrático medio (RMSE) absolutos y relativos (Parresol, 1999) a partir de los residuos de la validación.

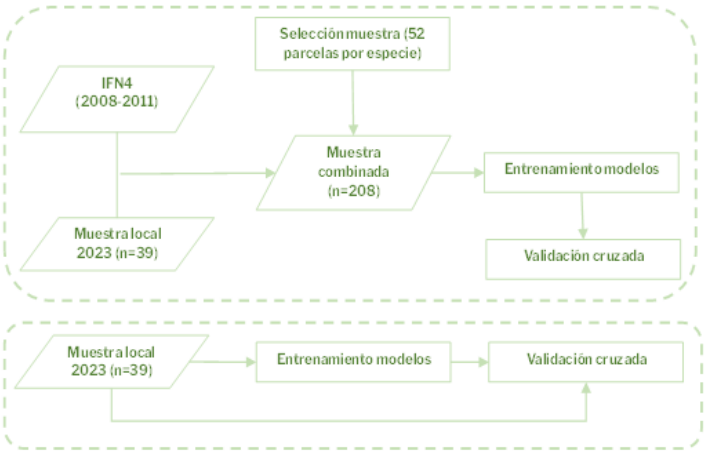

Figura 4. Diagrama esquemático de los métodos de modelización de AGB seguidos.

Incertidumbre de la estimación de AGB

Para estimar la incertidumbre de las predicciones de los modelos ajustados se siguieron los siguientes pasos:

- Selección de muestras bootstrap mediante un remuestreo a partir de la muestra original basado en la técnica de bootstrap de pares.
- Ajuste de nuevos modelos predictivos de AGB con las muestras Bootstrap.
- Cálculo de AGB de la población para cada réplica Bootstrap (10.000 iteraciones para cada conjunto de datos).
- Estimar el error estándar del bootstrap (SE) de la estimación de AGB poblacional.

El estimador basado en modelos de la media poblacional, $??_{??}$, se basa en predicciones de unidades poblacionales, ŷi (predicción de AGB para cada uno de los N píxeles del área de estudio).

$$\hat{\mu}_{MB} = \frac{1}{N}\sum_{i=1}^{N}\hat{y}_i$$

La media poblacional bootstrap (??$_{????}$) se calculó como:

$$\hat{\mu}_{boot} = \frac{1}{n_{boot}} \sum_{b=1}^{n_{boot}} \hat{\mu}_b$$

Donde es la media poblacional estimada de AGB y n_{boot} es el número de iteraciones bootstrapping.

La estimación bootstrap del sesgo se calcula como:

$$\widehat{Bias}_{boot} = \hat{\mu}_{boot} - \hat{\mu}_{MB}$$

El error estándar (SE) se calculó como:

$$SE(\hat{\mu}_{boot}) = \sqrt{\frac{1}{n_{boot} - 1} \sum_{b=1}^{n_{boot}} (\hat{\mu}_b - \hat{\mu}_{boot})^2}$$

4. Resultados

Los resultados de la validación cruzada muestran que los modelos basados en datos ALS del IGN francés tuvieron errores más bajos que los que usaron otros datos remotos. Los modelos basados únicamente en datos ópticos derivados de Landsat obtuvieron los peores resultados. Sin embargo, la inclusión del *zq95* del LIDAR francés mejoró significativamente la precisión del modelo, confirmando la importancia de la altura de la vegetación en la estimación de la AGB. También es posible observar que la adición de una variable de altura como la métrica *RH95* de GEDI a los datos ópticos (Sentinel-2) mejora significativamente la precisión de los modelos en términos de RMSE

Aunque todos los modelos ajustados únicamente datos de parcelas locales con RF tuvieron mejores resultados que los kNN, no existen diferencias importantes ellos. Sin embargo, cabe destacar el aumento significativo del sesgo en el modelo kNN frente al modelo RF cuando los datos de altura derivados de GEDI se incluyen a las variables de Sentinel-2.

Tabla 4. Resultados de los modelos ajustados.

Modelo	Muestra	ALS		Landsat	
Modelo	**Muestra**	**Sesgo (%)**	**RMSE (%)**	**Sesgo (%)**	**RMSE (%)**
kNN	IFN4 + PAU	-1,76	37,81	-0,80	53,64
RF	IFN4 + PAU	-0,81	54,55	0,23	60,74
kNN	PAU	-0,86	30,88	-0,47	52,85
RF	PAU	1,39	29,43	-1,29	50,67
Modelo	**Muestra**	**ALS**		**Landsat**	
Modelo	**Muestra**	**Sesgo (%)**	**RMSE (%)**	**Sesgo (%)**	**RMSE (%)**
kNN	IFN4 + PAU	-0,75	37,26	-	-
RF	IFN4 + PAU	0,34	62,35	-	-
kNN	PAU	-0,69	34,86	7,65	39,71
RF	PAU	-0,49	32,90	0,76	35,69

Tabla 5. Incertidumbre en la estimación de los valores poblacionales de AGB.

Modelo	Muestra	Sensor	SE (%)	Sesgo (%)	Media (t/ha)
kNN	IFN4 + PAU	LiDAR	3,76	-0,14	235.96
kNN	IFN4 + PAU	Landsat	5,58	-0,20	203.68
kNN	IFN4 + PAU	Landsat + LIDAR (*zq95*)	4,59	-0,70	230.87
RF	IFN4 + PAU	LiDAR	4,73	3,65	215.84
RF	IFN4 + PAU	Landsat	5,73	7,88	194.37
RF	IFN4 + PAU	Landsat + LIDAR (*zq95*)	3,81	2,38	225.74
kNN	PAU	LiDAR	4,86	-0,53	246.85
kNN	PAU	Landsat	8,56	0,36	201.03
kNN	PAU	Landsat + LIDAR (*zq95*)	6,20	-0,36	237.52
RF	PAU	LiDAR	4,22	-0,24	236.28
RF	PAU	Landsat	7,52	-0,63	205.28
RF	PAU	Landsat + LIDAR (*zq95*)	4,78	0,88	235.44
kNN	PAU	S2 + GEDI (*RH95*)	6,23	0,39	239.87
RF	PAU	S2 + GEDI (*RH95*)	7,17	4,06	227.60

La Tabla 5 muestra los resultados del análisis de la incertidumbre en la estimación de AGB con diferentes modelos, combinaciones de datos y sensores remotos. El mejor desempeño se observa en los modelos kNN que combinan datos de la muestra local con datos del IFN4. En general, estos modelos produjeron estimaciones más precisas e insesgadas que RF, destacando su capacidad para integrar datos dispares de manera más eficaz que RF en este contexto.

Los resultados del análisis de incertidumbre de las estimaciones confirman, al igual que la validación cruzada de los modelos, la superioridad de ALS frente a otros sensores en la estimación de AGB. La sola inclusión de un solo estadístico ALS como *zq95* en modelos con variables ópticas derivadas de Landsat redujo significativamente el SE, en comparación con aquellos que solo usan variables de Landsat. La inclusión de una variable de altura como *RH95* también redujo la incertidumbre de las estimaciones frente a modelos basados únicamente en datos ópticos, permitiendo además estimaciones significativamente mayores que los modelos basados en Landsat y del mismo orden de magnitud que las que producen los modelos basados en datos ALS. Esto resalta la importancia de las métricas de altura para mejorar la precisión de las estimaciones poblacionales y evitar la posible saturación de la señal en sensores ópticos.

Por otro lado, se observa que RF realizó estimaciones sistemáticamente más sesgadas que kNN en los modelos que combinan datos locales con datos externos del IFN4. Sin embargo, RF tuvo mejores resultados que kNN con muestras reducidas, es decir,

trabajando únicamente con una muestra pequeña local de parcelas. Sin embargo, el modelo basado en Sentinel-2 y GEDI presentó un sesgo notablemente mayor que el resto, mientras que kNN logró un modelo insesgado con menor SE (6,23% frente a 7,17% de RF).

5. Discusión

Los resultados de este estudio confirman la viabilidad e importancia de integrar datos de inventarios forestales nacionales (IFN4) con mediciones locales y datos de sensores remotos para estimar la biomasa aérea en situaciones con recursos limitados o donde la recopilación de datos de campo sea difícil. Esta integración permite la utilización de datos preexistentes en lugares con condiciones ecológicas o de gestión forestal distintas de aquellos donde fueron recopilados.

La validación de los modelos ajustados confirma la capacidad superior del ALS para caracterizar la estructura de la vegetación (en este caso AGB) frente a los sensores pasivos como Landsat, incluso cuando se utilizan datos de baja resolución LiDAR PNOA (Domingo et al., 2018; Gobakken & Næsset, 2008; Maltamo et al., 2006). Los resultados de la validación cruzada de los modelos que usan datos ALS están dentro del rango encontrado por otros autores. Domingo et al. (2018) obtuvieron valores de RMSE relativo en modelos basados en datos ALS para la estimación de AGB en bosque de *Pinus halepensis* entre 19,21% y 24,93%. Por su parte, el modelo de (Fekety et al., 2015) para estimar carbono en la AGB tuvo un RMSE del 52% en bosques templados mixtos de EE.UU.

Los modelos que utilizaron únicamente datos ópticos tuvieron, como era de esperar, resultados sistemáticamente peores que los modelos que incluyeron una variable de altura. Aunque los resultados de estos modelos se pueden considerar satisfactorios, tendieron a subestimar la AGB en los valores más altos. Esto coincide con lo observado en otros estudios que predicen variables forestales con datos ópticos (Astola et al., 2019; Jiang et al., 2021, 2022). Así, la inclusión de variables directamente relacionadas con la altura de la vegetación como zq95 se muestra fundamental para mejorar la precisión de los modelos, incluso cuando se combinaron con datos ópticos de menor resolución como los provenientes de Landsat, reafirmando la importancia de los datos tridimensionales en la caracterización de la estructura de los bosques y en la captura de patrones que no son fácilmente detectables mediante datos 2D. En este sentido, el *RH95* derivado de GEDI destaca como una solución prometedora para áreas donde no es posible acceder a datos LiDAR de alta resolución. Aunque los resultados de este estudio hacen notar que el uso de un modelo de *RH95* puede disminuir el efecto de saturación, Morin et al. (2023) observaron la persistencia de este efecto en modelos que combinan datos de diversos sensores satelitales y GEDI.

Los algoritmos de aprendizaje automático utilizados (RF y kNN) mostraron comportamientos diferenciados según las combinaciones de datos y el tamaño de las muestras. Los modelos kNN presentaron una mejor capacidad para tratar con datos heterogéneos y combinar de manera efectiva datos externos del IFN4 con los datos de las parcelas recogidos en el área de estudio, produciendo estimaciones con menor sesgo y menor incertidumbre en comparación con RF. Sin embargo, cuando se utilizó una muestra

local reducida (n=39), RF demostró ser más robusto y menos sensible a las limitaciones de los datos de entrenamiento, lo que lo convierte en una herramienta adecuada para entornos con datos limitados pero coherentes. Estos resultados coinciden con lo observado por Domingo et al. (2017), quienes hallaron RMSE de 16,7% para RF y 23,2% en kNN con una muestra de ajuste de 49 parcelas.

Un aspecto relevante de este estudio es la cuantificación de la incertidumbre mediante bootstrap. A la vista de los resultados, la evaluación de la incertidumbre se muestra fundamental para garantizar la fiabilidad de las estimaciones, permitido enriquecer cuantificación de recursos forestales (Condés & McRoberts, 2017; Fortin et al., 2018). En este sentido, se puede observar que modelos con resultados aparentemente peores en la fase de validación, pueden producir estimaciones más precisas y robustas cuando se evalúa la incertidumbre. Los resultados de este análisis de la incertidumbre permiten observar que las mejores estimaciones de AGB se obtuvieron con la aplicación de modelos kNN que combinan información externa del IFN4 con datos locales, poniendo en valor los datos de programas de IFN incluso en contextos diferentes a los que fueron tomados y la capacidad de kNN para tratar con datos heterogéneos de forma armónica y robusta. Esto demuestra que es posible reducir los costos de los inventarios mediante la integración de datos remotos y de inventarios preexistentes sin comprometer la calidad de las estimaciones, facilitando el monitoreo en áreas remotas o con escasos recursos para medir parcelas de campo.

Los resultados ponen de manifiesto que las técnicas empleadas, como el uso de bootstrap son esenciales para identificar y cuantificar las limitaciones de los modelos ajustados. Esto lo convierte en una herramienta esencial en el contexto de programas como REDD, los inventarios nacionales de gases de efecto invernadero (GEI) o proyectos de los mercados de carbono, donde la evaluación de la incertidumbre es necesaria para garantizar la fiabilidad de las estimaciones con márgenes de error conocidos y transparentes.

Una de las limitaciones de este estudio fue el uso de modelos para diferentes especies en vez de modelos específicos para cada estrato del bosque, ya que el área de estudio está compuesta por rodales más o menos puros de diversas especies arbóreas. Otros estudios comparados en esta discusión utilizaron modelos para masas monoespecíficas (Domingo et al., 2018) o masas mixtas (Fekety et al., 2015), pero no un modelo único que funcione para rodales monoespecíficos de varias especies. Finalmente, este estudio destaca la necesidad de seguir desarrollando metodologías robustas y transferibles que permitan a aprovechar las herramientas de teledetección y los datos disponibles para facilitar el monitoreo de áreas remotas o de proyectos de carbono.

6. Conclusiones

Este estudio demuestra la viabilidad de integrar datos de inventarios forestales nacionales (IFN4) con mediciones locales y datos de sensores remotos para estimar la biomasa aérea (AGB) en áreas remotas o con recursos limitados donde los métodos tradicionales no son viables. Además, los resultados destacan que en los casos en que no se dispone de datos ALS, se pueden emplear eficazmente datos remotos alternativos, de libre acceso como Landsat, Sentinel-2 o GEDI, para estimar la AGB, obteniendo

resultados con niveles aceptables de incertidumbre (como indica un error estándar bootstrapped inferior al 10%). Sin embargo, es importante tener en cuenta el efecto de saturación encontrado en los datos ópticos, que llevó a subestimar la AGB. La combinación de los datos ópticos con una variable de altura (ya sea ALS francés o GEDI) mitigó en general este problema.

El enfoque seguido en este estudio facilita la transferencia de datos de IFN a contextos locales, lo que optimiza el monitoreo en áreas remotas o con recursos limitados para mediciones de campo, aunque la eficacia de este proceso depende en gran medida del tipo de modelo utilizado. Se observó que los modelos kNN superaron significativamente a RF cuando trabajaron con conjuntos de datos que combinaban datos locales con datos externos del IFN4, mientras que RF se mostró más efectivo en el manejo de muestras locales reducidas, pero produjo estimaciones sesgadas con la muestra combinada.

La cuantificación de la incertidumbre mediante bootstrap permite confiar más en las estimaciones, lo cual es clave para programas de monitoreo de gases de efecto invernadero (GEI). De cara a aprovechar el potencial del uso combinado de datos existentes y sensores remotos, se recomienda seguir desarrollando metodologías robustas y transferibles a otros ecosistemas forestales y localizaciones.

7. Agradecimientos

Los autores agradecen la colaboración de TOTAL en la financiación de los trabajos, y del Instituto Geográfico Nacional de España y Francia y del Inventario Forestal Nacional en la facilitación de acceso gratuito a los datos ALS y de inventario.

8. Bibliografía

AHMED, O. S., FRANKLIN, S. E., WULDER, M. A., & WHITE, J. C.; 2015. Characterizing stand-level forest canopy cover and height using Landsat time series, samples of airborne LiDAR, and the Random Forest algorithm. *ISPRS J Photogramm Remote Sens*, *101*, 89–101. https://doi.org/10.1016/j.isprsjprs.2014.11.007

ALBERDI, I.; 2015. *Metodología para la estimación de indicadores armonizados a partir de los inventarios forestales nacionales europeos con especial énfasis en la biodiversidad forestal*. Universidad Politécnica de Madrid.

ALI, N., NEAGU, D., & TRUNDLE, P.; 2019. Evaluation of k-nearest neighbour classifier performance for heterogeneous data sets. *SN Appl Sci*, *1*(12). https://doi.org/10.1007/s42452-019-1356-9

ASTOLA, H., HÄME, T., SIRRO, L., MOLINIER, M., & KILPI, J.; 2019. Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region. *Remote Sens Environ*, *223*. https://doi.org/10.1016/j.rse.2019.01.019

BOATENG, E. Y., OTOO, J., & ABAYE, D. A.; 2020. Basic Tenets of Classification Algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest and Neural Network: A Review. *JDAIP*, *08*(04). https://doi.org/10.4236/jdaip.2020.84020

CAMPBELL, M. J., DENNISON, P. E., KERR, K. L., BREWER, S. C., & ANDEREGG, W. R. L.; 2021. Scaled biomass estimation in woodland ecosystems: Testing the individual and combined capacities of satellite multispectral and lidar data. *Remote Sens Environ*, *262*. https://doi.org/10.1016/j.rse.2021.112511

CONDÉS, S., & MCROBERTS, R. E.; 2017. Updating national forest inventory estimates of growing stock volume using hybrid inference. *For Ecol Manage*, *400*, 48–57. https://doi.org/10.1016/j.foreco.2017.04.046

DELEUZE, C., MORNEAU, F., RENAUD, J.-P., RIVOIRE, M., SANTENOISE, P., LONGUETAUD, F., MOTHE, F., HERVÉ, J.-C., VALLET, P., & VIVIEN, Y.; 2014. Estimer le volume total d'un arbre, quelles que soient l'essence, la taille, la sylviculture, la station. *Rendez-Vous Techniques*, *44*.

DOMINGO, D., LAMELAS-GRACIA, M. T., MONTEALEGRE-GRACIA, A. L., & DE LA RIVA-FERNÁNDEZ, J.; 2017. Comparison of regression models to estimate biomass losses and CO2 emissions using low-density airborne laser scanning data in a burnt Aleppo pine forest. *EuJRS*, *50*(1), 384–396. https://doi.org/10.1080/22797254.2017.1336067

DOMINGO, D., LAMELAS, M. T., MONTEALEGRE, A. L., GARCÍA-MARTÍN, A., & DE LA RIVA, J.; 2018. Estimation of total biomass in Aleppo pine forest stands applying parametric and nonparametric methods to low-density airborne laser scanning data. *Forests*, *9*(4). https://doi.org/10.3390/f9040158

ESTEBAN, J., MCROBERTS, R., FERNÁNDEZ-LANDA, A., TOMÉ, J., & NÆSSET, E.; 2019. Estimating Forest Volume and Biomass and Their Changes Using Random Forests and Remotely Sensed Data. *Remote Sens.*, *11*(16), 1944. https://doi.org/10.3390/rs11161944

FEKETY, P. A., FALKOWSKI, M. J., & HUDAK, A. T.; 2015. Temporal transferability of LiDAR-based imputation of forest inventory attributes. *Can J For Res*, *45*, 422–435. https://doi.org/10.1139/cjfr-2014-0405

FEKETY, P. A., FALKOWSKI, M. J., HUDAK, A. T., JAIN, T. B., & EVANS, J. S.; 2018. Transferability of Lidar-derived Basal Area and Stem Density Models within a Northern Idaho Ecoregion. *Can J Remote Sens*, *44*(2), 131–143. https://doi.org/10.1080/07038992.2018.1461557

FORTIN, M., MANSO, R., & SCHNEIDER, R.; 2018. Parametric bootstrap estimators for hybrid inference in forest inventories. *Forestry*, *91*(3). https://doi.org/10.1093/forestry/cpx048

FRIDMAN, J., HOLM, S., NILSSON, M., NILSSON, P., RINGVALL, A. H., & STÅHL, G.; 2014. Adapting National Forest Inventories to changing requirements - The case of the Swedish National Forest Inventory at the turn of the 20th century. https://doi.org/10.14214/sf.1095

GENUER, R., POGGI, J.-M., & TULEAU-MALOT, C.; 2015. VSURF: An R Package for Variable Selection Using Random Forests. *The R Journal*, *7*(2), 19–33. https://hal.archives-ouvertes.fr/hal-01251924

GLEASON, C. J., & IM, J.; 2012. Forest biomass estimation from airborne LiDAR data using machine learning approaches. *Remote Sens Environ*, *125*, 80–91. https://doi.org/10.1016/j.rse.2012.07.006

GOBAKKEN, T., & NÆSSET, E.; 2008. Assessing effects of laser point density, ground sampling intensity, and field sample plot size on biophysical stand properties derived from airborne laser scanner data. *Can J For Res*. https://doi.org/10.1139/x07-219

HEISKANEN, J., ADHIKARI, H., PIIROINEN, R., PACKALEN, P., & PELLIKKA, P. K. E.; 2019. Do airborne laser scanning biomass prediction models benefit from Landsat time series, hyperspectral data or forest classification in tropical mosaic landscapes? *Int J Appl Earth Obs Geoinf*, *81*. https://doi.org/10.1016/j.jag.2019.05.017

HOLOPAINEN, M., VASTARANTA, M., KARJALAINEN, M., KARILA, K., KAASALAINEN, S., HONKAVAARA, E., & HYYPPÄ, J.; 2015. Forest inventory attribute estimation using airborne laser scanning , aerial stereo imagery , radargrammetry and interferometry – Finnish experiences of the 3D techniques. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, *2*(March), 63–69. https://doi.org/10.5194/isprsannals-II-3-W4-63-2015

JIANG, F., KUTIA, M., MA, K., CHEN, S., LONG, J., & SUN, H.; 2021. Estimating the aboveground biomass of coniferous forest in Northeast China using spectral variables, land surface temperature and soil moisture. *Sci. Total Environ.*, *785*. https://doi.org/10.1016/j.scitotenv.2021.147335

JIANG, F., DENG, M., TANG, J., FU, L., & SUN, H.; 2022. Integrating spaceborne LiDAR and Sentinel-2 images to estimate forest aboveground biomass in Northern China. *Carbon Balance Manag.*, *17*(1). https://doi.org/10.1186/s13021-022-00212-y

LI, S., HARNER, E. J., & ADJEROH, D. A.; 2011. Random KNN feature selection - a fast and stable alternative to Random Forests. *BMC Bioinformatics*, *12*(1). https://doi.org/10.1186/1471-2105-12-450

Liaw, a, & Wiener, M.; 2002). Classification and Regression by randomForest. *R News*, *2* (December), 18–22. https://doi.org/10.1177/154405910408300516

LONGUETAUD, F., SANTENOISE, P., MOTHE, F., SENGA KIESSÉ, T., RIVOIRE, M., SAINT-ANDRÉ, L., OGNOUABI, N., & DELEUZE, C.; 2013. Modeling volume expansion factors for temperate tree species in France. *For Ecol Manage*, *292*. https://doi.org/10.1016/j.foreco.2012.12.023

LUTHER, J. E., FOURNIER, R. A., VAN LIER, O. R., & BUJOLD, M.; 2019. Extending ALS-based mapping of forest attributes with medium resolution satellite and environmental data. *Remote Sens.*, *11*(9). https://doi.org/10.3390/rs11091092

MALTAMO, M., EERIKÄINEN, K., PACKALÉN, P., & HYYPPÄ, J.; 2006. Estimation of stem volume using laser scanning-based canopy height metrics. *Forestry*, *79*(2). https://doi.org/10.1093/forestry/cpl007

MORIN, D., PLANELLS, M., MERMOZ, S., & MOURET, F.; 2023. Estimation of forest height and biomass from open-access multi-sensor satellite imagery and GEDI Lidar data: high-resolution maps of metropolitan France. *ArXiv Preprint ArXiv:2310.14662*.

NAVARRO FERNÁNDEZ, J. A.; 2019. *Aplicaciones de la fotogrametría digital aérea en el inventario forestal* [E.T.S.I. Montes, Forestal y del Medio Natural (UPM)]. https://doi.org/10.20868/UPM.thesis.57694

PARRESOL, B. R.; 1999. Assessing tree and stand biomass: A review with examples and critical comparisons. *For. Sci.*, *45*(4). https://doi.org/10.1093/forestscience/45.4.573

RANI, P.; 2017. A Review of various KNN Techniques. *Int. J. Eng. Sci. Technol.*, *V*(VIII). https://doi.org/10.22214/ijraset.2017.8166

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

SHANG, C., TREITZ, P., CASPERSEN, J., & JONES, T.; 2019. Estimation of forest structural and compositional variables using ALS data and multi-seasonal satellite imagery. *Int J Appl Earth Obs Geoinf*, *78*. https://doi.org/10.1016/j.jag.2018.10.002

TOMPALSKI, P., WHITE, J. C., COOPS, N. C., & WULDER, M. A.; 2019. Demonstrating the transferability of forest inventory attribute models derived using airborne laser scanning data. *Remote Sens Environ*, *227*(April), 110–124. https://doi.org/10.1016/j.rse.2019.04.006

TOMPPO, E., GSCHWANTNER, T., LAWRENCE, M., & MCROBERTS, R. E.; 2010. National forest inventories: Pathways for common reporting. In *National Forest Inventories: Pathways for Common Reporting*. https://doi.org/10.1007/978-90-481-3233-1

WHITE, J. C., WULDER, M. A., VARHOLA, A., VASTARANTA, M., COOPS, N. C., COOK, B. D., PITT, G. D., & WOODS, M.; 2013. A best practices guide for generating forest inventory attributes from airborne laser scanning data using an area-based approach. *For. Chron.*, *89*, 722–723. https://doi.org/10.5558/tfc2013-132

ZANNE, A., LOPEZ-GONZALEZ, G., COOMES, D.A., ILIC, J., JANSEN, S., LEWIS, S.L., MILLER, R.B., SWENSON, N.G., WIEMANN, M.C., & CHAVE, J.; 2009. Data from: Towards a worldwide wood economics spectrum. *Dryad Digital Repository*. https://doi.org/doi:10.5061/dryad.234