

GreenWest: inteligencia artificial para la predicción de créditos de carbono en proyectos de (re)forestación en España

Correspondiente al grupo B1 (GreenWest) de las líneas de investigación de DemIA

Maidar Araceli Urbón Jiménez^{*1}, Jaime Gabriel Vegas¹, Ana de Luis Reboredo¹, Belén Pérez Lancho¹ y Ana-Belén Gil-González¹

¹Grupo B1, Equipo de investigación BISITE, Universidad de Salamanca, Facultad de Ciencias, Salamanca, Castilla y León, España,

{murbon001, JaimeGabrielVegas, adeluis, lancho, abg}@usal.es

30 de octubre de 2025

Resumen

Este trabajo presenta **GreenWest**, un modelo de inteligencia artificial diseñado para predecir la cantidad de carbono capturado en proyectos de forestación y reforestación en España. El modelo se entrena con datos multifuente: registros del **Inventario Forestal Nacional (IFN3–IFN4, MITECO)** [?], variables climáticas derivadas de **Copernicus/ERA5-Land** [?] e índices espectrales procedentes de **imágenes Landsat** (Collection 2, Level 2, USGS) [?]. Estos datos se integran en una base de datos relacional jerárquica descrita en un trabajo complementario [?], que organiza la información por parcela, especie y clase diamétrica, manteniendo trazabilidad y coherencia estructural entre inventarios.

El modelo desarrollado responde a la pregunta: *Dado un cultivo forestal con características concretas de vegetación, clima y terreno, ¿cuánto CO₂ contendrá pasados unos años?* Esta capacidad predictiva permite su integración en marcos de optimización forestal, abordando cuestiones como la selección de especies o la asignación óptima de terrenos para maximizar la fijación de carbono.

Se evaluaron múltiples enfoques de aprendizaje supervisado, destacando **CatBoost** [?] como el modelo con mejor rendimiento ($R^2 > 0,80$, RMSE<15), con alta capacidad de generalización temporal mediante validación cruzada por grupos. Los resultados demuestran el potencial del enfoque para estimar la absorción futura de CO₂ y optimizar decisiones de gestión forestal sostenible, contribuyendo a la transición hacia una economía baja en emisiones [? ?].

Palabras clave: créditos de carbono, inteligencia artificial, forestación, reforestación, modelado predictivo, cambio climático.

^{*} Autora de correspondencia: murbon001@usal.es

1. INTRODUCCIÓN

El cambio climático es uno de los mayores desafíos globales y su manifestación más directa es el aumento de las concentraciones atmosféricas de dióxido de carbono (CO_2), con impactos sobre criosfera, extremos climáticos y ecosistemas [?]. Los bosques actúan como sumideros naturales al fijar CO_2 en biomasa vía fotosíntesis, por lo que su gestión resulta clave para la mitigación.

A lo largo de las últimas décadas, instrumentos internacionales como la *Convención Marco de las Naciones Unidas sobre el Cambio Climático (CMNUCC)* y el *Protocolo de Kioto* [?] han establecido los marcos regulatorios para reducir las emisiones de gases de efecto invernadero mediante mecanismos basados en el mercado. En este contexto surgen los *créditos de carbono*, unidades que representan la cantidad de dióxido de carbono (CO_2) —habitualmente una tonelada— que ha sido capturada o cuya emisión ha sido evitada a través de proyectos certificados de mitigación.

Entre las actividades elegibles, la forestación y reforestación destacan por su capacidad de actuar como sumideros naturales de carbono, fijando CO_2 en la biomasa y el suelo. No obstante, para que estas actuaciones puedan generar créditos de carbono válidos, deben cumplir una serie de criterios técnicos y legales definidos en la normativa internacional y nacional vigente:

- **Intervención humana directa:** Los árboles deben provenir de actividades de intervención humana, como la plantación, siembra o fomento de semilleros naturales. Esto significa que los cultivos forestales naturales no son elegibles para la contabilización de carbono.
- **Período mínimo de 30 años:** Para que un proyecto sea válido, debe garantizarse que los árboles permanezcan en el terreno durante un período mínimo de tiempo, generalmente 30 años, lo que excluye la absorción de carbono de cultivos estacionales, cuyo carbono es liberado nuevamente al ser cosechados.
- **Superficie mínima de 1 hectárea:** El proyecto debe abarcar al menos 1 hectárea de terreno para ser considerado.
- **Fracción mínima de cabida cubierta del 20 %:** Para que un área sea considerada como bosque, debe cubrir al menos el 20 % del área con especies arbóreas.
- **Altura mínima de los árboles maduros de 3 metros:** Los árboles deben alcanzar una altura mínima de 3 metros en su madurez, aunque no es necesario que alcancen esta altura al inicio de la plantación.

Este trabajo presenta **GreenWest**, un modelo de inteligencia artificial para estimar la cantidad de carbono que capturará un cultivo forestal en España a partir de variables de vegetación, clima y terreno. Este enfoque innovador tiene el potencial de transformar la gestión de proyectos de forestación y reforestación, optimizando las prácticas de plantación y maximizando la cantidad de carbono que se puede capturar en estos ecosistemas.

La pregunta operativa es: *dadas las características iniciales de una plantación, ¿cuánto CO_2 contendrá tras t años?* Para responderla, se integran datos del **Inventario Forestal Nacional** (IFN3–IFN4, MITECO) [?], reanálisis **ERA5-Land** [?] e **índices espectrales Landsat** (Collection 2, L2) [?] en una base de datos relacional jerárquica descrita en un trabajo complementario [?].

Este modelo no solo mejorará la comprensión del comportamiento de los sumideros de carbono, sino que también proporcionará herramientas útiles para la toma de decisiones estratégicas tanto en el ámbito empresarial como en el ambiental. De esta forma, el proyecto *GreenWest* contribuye a la transición hacia una economía baja en carbono, alineándose con los objetivos globales de sostenibilidad establecidos en el marco de la CMNUCC y el *Protocolo de Kioto*, y promoviendo la creación de un mercado de créditos de carbono más eficiente y accesible para los actores económicos involucrados en la gestión de los recursos naturales.

Contribución. GreenWest (i) modela la captura futura de carbono a resolución parcela–especie–clase diamétrica, (ii) demuestra *generalización temporal* mediante validación cruzada por grupos de periodo, y (iii) habilita su uso en marcos de optimización (selección de especie y emplazamiento) manteniendo compatibilidad con los requisitos de elegibilidad de créditos [?]. Estos resultados facilitan la planificación de proyectos de (re)forestación y la toma de decisiones para una gestión forestal alineada con una economía baja en carbono.

2. OBJETIVOS Y JUSTIFICACIÓN

El presente estudio tiene como objetivo principal desarrollar un modelo de inteligencia artificial capaz de predecir con precisión la capacidad de absorción de dióxido de carbono (CO_2) en cultivos forestales españoles. Este modelo se basa en variables que describen la especie arbórea, las características del terreno y las condiciones climáticas. A partir de este objetivo general se derivan varias metas específicas, que en conjunto justifican la relevancia y aplicabilidad del proyecto.

Objetivos específicos

- **Desarrollar un modelo predictivo robusto:** Construir un modelo de aprendizaje automático que estime la cantidad de CO_2 que será capturado a lo largo del tiempo por un cultivo forestal, a partir de datos como especie, tipo de suelo, clase diamétrica, clima y otras variables relevantes.
- **Optimizar la captura de carbono:** Utilizar el modelo para identificar combinaciones óptimas de especies y terrenos que maximicen la fijación de carbono, contribuyendo a la planificación eficiente de proyectos de forestación y reforestación.
- **Asegurar la compatibilidad con las normativas internacionales:** Garantizar que las predicciones y salidas del modelo sean compatibles con los marcos normativos definidos por la *Convención Marco de las Naciones Unidas sobre el Cambio Climático* (CMNUCC) y el *Protocolo de Kioto*, cumpliendo así los criterios necesarios para la validación de créditos de carbono.
- **Analizar los factores determinantes del desarrollo forestal:** Estudiar la influencia de variables climáticas (como la temperatura y la precipitación) y edáficas (como el tipo de suelo o la pendiente) sobre el crecimiento forestal y su capacidad de capturar carbono.
- **Apoyar la toma de decisiones ambientales y empresariales:** Proporcionar una herramienta práctica y validada que permita a técnicos, gestores y empresas seleccionar las especies más adecuadas y planificar actuaciones de forestación con la mayor eficiencia posible en términos de secuestro de carbono.

Justificación

La necesidad de contar con herramientas predictivas para estimar la captura de CO_2 se ha intensificado ante el crecimiento del mercado voluntario de créditos de carbono, y las obligaciones adquiridas en el marco de la CMNUCC y el Protocolo de Kioto. Según estos acuerdos, cada país debe reportar sus emisiones y absorciones de gases de efecto invernadero, y puede utilizar actividades de forestación y reforestación como mecanismos de compensación.

Para que estos proyectos sean elegibles, deben cumplir criterios específicos, los cuales hacen imprescindible disponer de modelos que no solo estimen el carbono actual, sino que sean capaces de prever su evolución a futuro con base en condiciones iniciales y variables predictoras.

Este trabajo busca cubrir ese vacío mediante el uso de inteligencia artificial aplicada a datos reales y multifuente. Integrar su manejo dentro del sistema de créditos de car-

bono puede representar una importante oportunidad para la economía local y para la mitigación del cambio climático.

3. REVISIÓN DE LA LITERATURA

El secuestro de carbono en ecosistemas forestales ha cobrado una importancia creciente en la literatura científica, impulsada tanto por los compromisos internacionales en materia de cambio climático [?] como por el auge de los mercados de créditos de carbono. Esto ha motivado el desarrollo de modelos orientados a cuantificar la biomasa forestal y estimar el contenido de carbono, aprovechando avances recientes en sensores remotos y técnicas de inteligencia artificial (IA).

Una de las estrategias más consolidadas es la estimación del carbono almacenado en un momento dado a partir de datos de teledetección. Goetz et al. (2009) [?] revisan el uso de imágenes satelitales (MODIS, Landsat) en modelos empíricos de biomasa aérea, destacando su eficacia a escala regional en zonas boreales. Este tipo de estimaciones suele realizarse mediante regresiones lineales o algoritmos de mínimos cuadrados generalizados, con coeficientes de determinación (R^2) típicamente entre 0.6 y 0.8 según la resolución de entrada y la heterogeneidad del ecosistema.

La aplicación de aprendizaje profundo ha permitido mejorar sustancialmente la precisión y resolución espacial de estas estimaciones. Por ejemplo, Zhang et al. (2022) [?] integran imágenes Sentinel-2 con redes neuronales convolucionales, alcanzando un R^2 de 0.84 para estimar el carbono en bosques subtropicales. Del mismo modo, Yang et al. (2023) [?] desarrollan el modelo *ForestCarbonAI*, entrenado con datos multiespectrales y LIDAR, con el que generan mapas de carbono forestal de alta resolución (10 m), reportando errores medios absolutos (MAE) inferiores a 3.5 tC/ha en zonas templadas. Otros trabajos recientes, como Reiersen et al. (2022) [?] o Dong et al. (2023) [?], también demuestran la eficacia del deep learning para estimaciones estáticas, aunque se centran en contextos tropicales y no consideran el componente temporal.

Frente a estos enfoques descriptivos, algunas iniciativas han intentado proyectar la evolución del carbono a futuro. En el ámbito nacional, el Ministerio para la Transición Ecológica (MITECO) ha implementado herramientas como la calculadora ex ante de absorciones [?], que permite obtener estimaciones simplificadas del carbono que puede fijarse en una plantación forestal en función de la especie y la zona agroclimática. No obstante, este instrumento se basa en coeficientes tabulados y no incorpora variables edafoclimáticas reales ni técnicas de modelización basadas en datos, lo que limita su precisión y capacidad de adaptación a contextos específicos.

En este escenario, el presente trabajo propone una metodología innovadora centrada en la predicción dinámica de carbono a largo plazo. A diferencia de los modelos anteriores, que estiman el carbono ya almacenado, este estudio se enfoca en anticipar cuánto carbono capturará un cultivo forestal en un horizonte temporal concreto. Para ello, se estudian diversos modelos de aprendizaje supervisado entrenados con datos históricos del Inventario Forestal Nacional (IFN2, IFN3 e IFN4), variables climáticas de Copernicus, características edáficas y métricas espectrales derivadas de imágenes Landsat [? ? ?]. Los detalles sobre la arquitectura del modelo, las variables utilizadas, los algoritmos implementados y las métricas de evaluación se desarrollan en la siguiente sección.

4. METODOLOGÍA

Esta sección describe el procedimiento seguido para entrenar y validar los modelos predictivos desarrollados. La metodología se fundamenta en la identificación de los factores que condicionan el crecimiento forestal y, por tanto, la capacidad de captura de carbono en un periodo determinado. Para ello, se ha trabajado sobre una base de datos específicamente estructurada para facilitar su uso en tareas de modelización ecológica y análisis predictivo.

El carbono absorbido por los árboles se acumula en su biomasa a lo largo del tiempo. La cantidad almacenada depende del tamaño de los individuos, el cual está influenciado por múltiples variables ambientales y estructurales. Así, la estimación del carbono capturado entre dos momentos se realiza midiendo el crecimiento de los árboles y aplicando factores de conversión, considerando que aproximadamente el 50% de la biomasa seca corresponde a carbono [?].

La base de datos empleada recoge características forestales, climáticas y espectrales a nivel de parcela, especie y clase diamétrica. Esta estructura relacional permite capturar la evolución del bosque entre inventarios y alimentar modelos capaces de predecir el contenido futuro de carbono a partir de observaciones pasadas. El conjunto de variables integradas satisface los principales requisitos ecológicos y estadísticos del problema, garantizando una representación suficiente del entorno forestal.

El crecimiento de las especies arbóreas está condicionado por factores edáficos, climáticos, topográficos y de competencia intraespecífica. Las condiciones meteorológicas, como la temperatura o la precipitación, afectan a la fotosíntesis y la disponibilidad hídrica. La orientación, la pendiente y la altitud modifican la radiación recibida y la capacidad de retención de agua. A su vez, la densidad de árboles por superficie determina el nivel de competencia por los recursos, lo cual varía según la especie y su tolerancia ecológica [? ?].

A partir de estos principios, la base de datos recoge un conjunto diverso de variables que incluyen:

- **Características del terreno:** textura, tipo de suelo, orientación, pendiente y elevación, que condicionan el entorno físico del crecimiento.
- **Estructura y composición forestal:** cobertura total y arbórea, ocupación específica por especie, forma y estado de la masa, y distribución espacial del arbolado.
- **Variables dendrométricas:** número de pies, área basimétrica y volumen por clase diamétrica, como indicadores directos de biomasa y densidad forestal.
- **Condiciones climáticas:** medias y extremos de temperatura y precipitación por estación, obtenidas a partir de registros satelitales y reanálisis.
- **Índices de vegetación:** métricas como NDVI, NDII, GNDVI y EVI, que capturan el vigor, humedad y densidad de la cubierta vegetal (véase Sección ??).

4.1. Objetivo y alcance

El objetivo principal de este trabajo es estimar el **carbono total** que una parcela forestal será capaz de capturar en un horizonte temporal de aproximadamente 20 a 30 años. Este valor constituye un indicador clave del potencial de mitigación climática

de los ecosistemas forestales y resulta esencial para la planificación de proyectos de compensación de emisiones.

El modelo se entrena para predecir dos variables de respuesta relacionadas con la captura de carbono: *c* y *carbono_bruto*. La primera (*c*) representa el carbono total contenido en la biomasa aérea y subterránea por especie, parcela e inventario, expresado en toneladas de carbono por hectárea (tC/ha). Su origen se encuentra en las variables del Inventario Forestal Nacional (IFN), donde se reportan estimaciones del contenido de carbono aéreo (*ca*), radical (*cr*) y total (*c*). Sin embargo, la documentación del IFN no detalla el método de cálculo empleado, por lo que se asume que siguen las directrices de la *Guía del MITECO para la estimación de absorciones de CO₂* [?]. Estas variables presentan vacíos notables en el IFN4 y ausencia completa en los inventarios previos (IFN2 e IFN3), por lo que al crear la base de datos se recurrió a un modelo *Random Forest Regressor* para imputar los valores faltantes a partir de variables dendrométricas observadas (Especie, CD, VSC, NPies, ABas, IAVC, VCC y VLE). Aunque este enfoque mostró un rendimiento satisfactorio ($R^2_{test} > 0,90$), sus estimaciones derivan de un modelo estadístico y no de mediciones directas.

La variable *carbono_bruto* (tC) se desarrolló como alternativa trazable y físicamente interpretable, calculada directamente a partir de variables medidas en campo: número de pies (*npies*), altura media (*ht*), tipo de especie (*clase_especie*) y clase diamétrica (*cd_id*). Su estimación sigue un modelo alométrico adaptado de [?] y las directrices del IPCC [?]. La biomasa aérea individual se obtiene mediante ecuaciones específicas por tipo de especie:

$$\text{Biomasa}_{\text{coníferas}} = 0,05 \times (CD^2 \times Ht)^{0,9}, \quad \text{Biomasa}_{\text{frondosas}} = 0,067 \times (CD^2 \times Ht)^{0,9}.$$

A partir de esta biomasa se calcula el carbono por árbol aplicando el factor de conversión recomendado por el IPCC:

$$\text{Carbono}_{\text{árbol}} = \frac{\text{Biomasa} \times 0,5}{1000}.$$

Finalmente, el *carbono_bruto* total se obtiene multiplicando el carbono por árbol por el número de pies de cada clase diamétrica:

$$\text{Carbono_bruto} = \text{Carbono}_{\text{árbol}} \times \text{npies}.$$

El resultado se expresa en toneladas de carbono (t) por parcela y especie, sin normalizar por superficie, lo que facilita la trazabilidad y la comparación entre inventarios. Esta aproximación evita dependencias con factores de expansión específicos del IFN y proporciona una medida más transparente y replicable del carbono capturado a partir de datos observados. En coherencia con los criterios de proyectos de reforestación y forestación, las entradas correspondientes a brinzales o plantones se consideran con valor de carbono cero, dado que el carbono de las fases tempranas de desarrollo no se contabiliza como carbono capturado en los registros oficiales.

Para que un proyecto forestal sea elegible dentro de los programas de *créditos de carbono*, debe cumplir una serie de requisitos técnicos establecidos en los marcos regulatorios internacionales [? ?]. Estos criterios aseguran la validez ambiental y la permanencia del secuestro de carbono, y entre los principales se encuentran:

- **Intervención humana directa:** el incremento del carbono debe ser resultado de una acción planificada (reforestación, restauración o manejo forestal sostenible).
- **Permanencia mínima de 30 años:** el carbono capturado debe mantenerse almacenado durante al menos tres décadas.
- **Superficie mínima de 1 hectárea:** el proyecto debe tener una extensión suficiente para generar beneficios climáticos verificables.
- **Fracción mínima de cabida cubierta del 20 %:** el área arbolada debe superar este umbral para ser considerada masa forestal.
- **Altura mínima de los árboles de 3 metros en la madurez:** requisito que garantiza la representatividad estructural del bosque.

El objetivo principal de este trabajo es estimar el **carbono total** que una parcela forestal será capaz de capturar en un horizonte temporal de aproximadamente 20–30 años. El conjunto de datos incluye dos variables de carbono: una en *toneladas totales* y otra en *toneladas por hectárea*, lo que permite analizar el contenido de carbono en términos absolutos y normalizados por superficie.

4.2. Origen y estructura de los datos

La base de datos empleada en este trabajo integra información forestal, climática y espectral estructurada en torno a la parcela como unidad básica. Cada parcela se describe mediante sus coordenadas geográficas, características edáficas y su evolución a través de distintos inventarios (IFN2, IFN3, IFN4).

Los datos forestales incluyen información por especie y clase diamétrica, como número de pies, volumen con y sin corteza, área basimétrica, carbono aéreo, radical y total. Estos valores permiten caracterizar con precisión la estructura y crecimiento de la vegetación.

A cada parcela se asocian también estadísticas climáticas agregadas por estación e inventario: temperaturas (superficie, aire y subsuelo) y precipitaciones, resumidas mediante métricas como media, máxima, mínima y desviación típica.

Finalmente, se incorporan índices espectrales derivados de imágenes satelitales (NDVI, EVI, NDII, GNDVI), que permiten cuantificar propiedades biofísicas de la vegetación:

- **NDVI (Normalized Difference Vegetation Index):** estima la actividad fotosintética.
- **EVI (Enhanced Vegetation Index):** mejora la sensibilidad en zonas densamente vegetadas.
- **NDII (Normalized Difference Infrared Index):** refleja el contenido hídrico de la vegetación.
- **GNDVI (Green NDVI):** variante del NDVI basada en la banda verde, sensible al clorofila.

Estructura de la base de datos

Estos datos se organizan en las siguientes entidades troncales:

- **parcelas:** contiene la información básica de localización y características edáficas de cada parcela.
- **parcela_inventario:** describe el estado de cada parcela en un inventario determinado (*parcela_id*, *inventario_id*), incluyendo atributos edáficos y de contexto (p. ej., *nivel1_id*, *textura_id*).
- **parcela_inventario_especie:** detalla la presencia y condición de cada especie dentro de una parcela e inventario, incorporando descriptores de masa y tratamientos silvícolas.
- **parcela_inventario_especie_cd:** describe las poblaciones arbóreas por parcela, especie y *clase diamétrica* (*cd_id*): n.º de pies (*npies*), área basimétrica (*abas*), volúmenes (*vcc*, *vsc*, *vle*), incrementos (*iavc*) y carbono (*ca*, *cr*).
- **parcela_especie_arbol:** caracteriza los pies mayores identificados por parcela y especie en el inventario cuarto. Recoge las características particulares de cada pie como altura (*ht*), diámetros (*dn1* y *dn2*), ubicación respecto del centro de la parcela (*rumbo*, *distancia*), volumen (*vcc*, *vsc*, *vle*), incremento (*iavc*) y carbono (*ca*, *cr*).
- **parcela_inventario_estacion:** almacena agregados climático-biofísicos por estación (*estacion_id*) en la misma granularidad parcela-inventario, incluyendo variables como precipitación (*PR*) y temperatura (*T2M*, *SKT*, *STL**), junto a índices de vegetación (*NDVI*, *EVI*, *NDII*, *GNDVI*).
- **especies y grupos:** recogen la información taxonómica y su clasificación jerárquica, estableciendo la relación entre especies individuales y grupos funcionales.

Cada variable categórica posee una tabla de catálogo propia (*cat_*), donde se definen los valores posibles y sus descripciones. Por ejemplo, *cat_textura*, *cat_nivel1*, *cat_tratmasa* o *cat_origen*. Todas siguen un patrón uniforme: la clave primaria es el identificador de la variable (*<variable>_id*), y las tablas troncales referencian este mismo campo como clave foránea. Además la base de datos incluye una tabla llamada *meta_variaciones* que recoge los metadatos.

La Figura ?? muestra el esquema general de las tablas troncales y sus principales relaciones. Este diagrama resume la estructura interna de la base de datos y su jerarquía de dependencias.

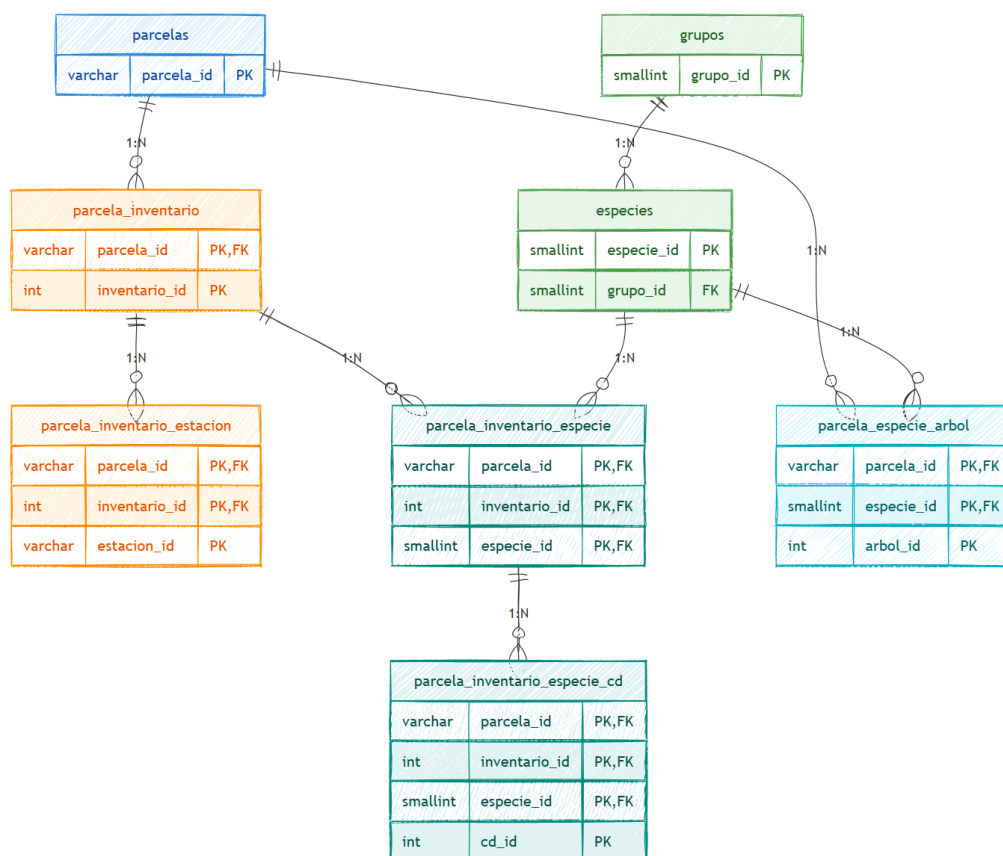


Figura 4.1: Esquema relacional de las tablas principales de la base de datos. Tabla extraída de [?], donde se pueden consultar más detalles sobre las variables.

Diccionario resumido de variables

Tabla 4.1: Resumen de variables principales por entidad. Tabla extraída de [?].

Variable	Descripción	Unidad	Tipo de dato
parcelas			
parcela_id	Identificador único de parcela (IFN).	–	Identificador
latitud, longitud	Coordenadas geográficas (WGS84).	°	Geográfico
coorx, coory	Coordenadas UTM; huso especifica zona.	m (UTM)	Geográfico
elevacion	Cota sobre el nivel del mar (NASADEM).	m	Númérico
pendiente	Inclinación del terreno.	°	Númérico
orientacion	Orientación del terreno (0–360).	°	Númérico
presencia_id	Presencia en IFN → cat_presencia.	–	Catégorico
tipsuelo1_id, tipsuelo2_id, tipsuelo3_id	Tipos de suelo → cat_tipsuelo*.	–	Catégorico
rocosidad_id	Rocosisdad → cat_rocosidad.	–	Catégorico
radio, superficie	Radio de parcela y superficie derivada.	m; ha	Númérico
parcela_inventario			
Continúa en la siguiente página			

Variable	Descripción	Unidad	Tipo de dato
parcela_id, inventario_id	Clave compuesta (parcela-inventario).	–	Identificador
ano	Año de apeo.	año	Numérico
nivel1_id, nivel2_id	Morfoestructura. → cat_nivel*.	–	Categórico
textura_id	Textura de suelo → cat_textura.	–	Categórico
merosiva_id	Manifestaciones erosivas → cat_merosiva.	–	Categórico
matorg_id	Materia orgánica → cat_matorg.	–	Categórico
modcomb_id	Modelo de combustible → cat_modcomb.	–	Categórico
disesp_id	Distribución espacial → cat_disesp.	–	Categórico
comesp_id	Composición específica → cat_comesp.	–	Categórico
fccarb, fcctot	Fracción de cabida cubierta (árboles).	%	Numérico
parcela_inventario_especie			
parcela_id, inventario_id, especie_id	Clave compuesta (parcela-inventario-especie).	–	Identificador
ocupa	Grado de ocupación de la especie.	(0–10)	Numérico
estado_id	Estado de desarrollo. → cat_estado.	–	Categórico
fpmasa_id	Forma principal de masa → cat_fpmasa.	–	Categórico
tratmasa_id	Tratamientos selvícolas → cat_tratmasa.	–	Categórico
orgmasa1_id	Origen de masa (IFN3/4)→ cat_orgmasa1.	–	Categórico
masa_id	Clasificación de masa → cat_masa.	–	Categórico
origen_id	Origen de la masa (IFN2) → cat_origen.	–	Categórico
parcela_inventario_especie_cd			
parcela_id, inventario_id, especie_id	Clave compuesta (parcela-inventario-especie- cd).	–	Identificador
cd_id	Clase diamétrica (CD) reglamento IFN.	cm	Numérico dis- creto
npies	Número de pies.	pies/ha	Numérico
abas	Área basimétrica.	m ² /ha	Numérico
vcc, vsc, vle	Volúmenes (con/sin corteza; leñas).	m ³ /ha	Numérico
iavc	Incremento anual del volumen con corteza.	m ³ /ha·año	Numérico
ca, cr	Carbono aéreo y radical.	t/ha	Numérico
ht	Altura media (modelo CatBoost).	m	Numérico
carbono_bruto	Carbono total estimado (alometrías).	t	Numérico
parcela_especie_arbol			
parcela_id, especie_id	Clave compuesta (parcela-especie-árbol).	–	Identificador
arbol_id	Identificador del árbol dentro de parcela y espe- cie.	–	Entero
rumbo	Rumbo desde el centro de la parcela al árbol.	grados cente- simales	Numérico
distancia	Distancia desde el centro de la parcela al árbol.	m	Numérico
cd	Clase diamétrica (reglamento IFN).	cm	Numérico dis- creto

Continúa en la siguiente página

Variable	Descripción	Unidad	Tipo de dato
ht	Altura total del árbol inventariado.	m	Numérico
dn1, dn2	Diámetros normales perpendiculares.	mm	Numérico
abas	Área basimétrica del pie medido.	m ²	Numérico
iavc	Incremento anual del volumen con corteza.	dm ³ /año	Numérico
vcc, vsc, vle	Volúmenes (con corteza, sin corteza, leñas).	dm ³	Numérico
ca, cr	Carbono aéreo y radical del árbol.	t	Numérico
parcela_inventario_estacion			
parcela_id, inventario_id, estacion_id	Clave compuesta (agregado estacional).	–	Identificador
PR_*	Estadísticos de precipitación (mean, max, min, std, sum).	mm/(m ² ·día), mm/m ²	Numérico
T2M_*, SKT_*	Aire 2 m y temperatura superficial (mean, max, min, std).	°C	Numérico
STL1_*–STL4_*	Temperatura del suelo por niveles (mean, max, min, std).	°C	Numérico
NDVI_*, EVI_*, NDII_*, GNDVI_*	Índices de vegetación (max, mean, median, min, std).	adimensional	Numérico
especies y grupos			
especie_id	Identificador de especie IFN.	–	Identificador
nombre, sinonimia	Denominación IFN y sinónimos.	–	Texto
tipo_especie	0 = conífera; 1 = frondosa.	–	Categorico
grupo_id	Grupo funcional → grupos.	–	Identificador
grupos.nombregrupo	Nombre del grupo.	–	Texto

Cardinalidad y completitud

El volumen de entradas por tabla es:

Tabla	Número de registros
parcelas	52,298
parcela_inventario	147,995
parcela_inventario_especie	417,119
parcela_inventario_especie_cd	1,191,070
parcela_especie_arbol	855,860
parcela_inventario_estacion	470,056
especies	195
grupos	33

4.3. Preparación y tratamiento de los datos

Para caracterizar el crecimiento de los árboles en los datos que se alimentan al modelo es necesario seleccionar dos inventarios distintos, permitiendo calcular la evolución de las variables forestales en el periodo transcurrido entre ambos. Dado que los inventarios IFN3 e IFN4 tienen una estructura similar y caracterizan mejor el terreno (más variables) la primera elección natural es emplear estos inventarios.

Filtrado de registros

Se descartan todas aquellas parcelas en las que el valor de carbono total (C o carbono_bruto) en la segunda inventariación es inferior a la primera. Estos casos suelen deberse a episodios de deforestación, incendios u otras perturbaciones, y no representan un crecimiento forestal neto.

El conjunto de datos empleado se restringe únicamente a las parcelas que presentan una *fccarb* (fracción de cabida cubierta arbórea) igual o superior al 20 %. Este umbral define la proporción mínima de superficie ocupada por copas de árboles respecto al área total de la parcela, y constituye una de las condiciones esenciales para considerar una superficie como terreno forestal. La exclusión de parcelas con *fccarb* inferior al 20 % permite asegurar que las estimaciones de carbono se realicen sobre masas forestales consolidadas, evitando sesgos asociados a áreas agrícolas o matorrales.

Tras aplicar los criterios de inclusión definidos —esto es, $fccarb > 20$ y la no reducción del carbono total entre inventarios ($c3 \geq c4$ o $carbono_bruto3 \geq carbono_bruto4$)—, se obtuvo la siguiente distribución de registros:

- **Caso C ($c3 \geq c4$):** de un total de 171 170 parcelas, 104 558 (61.1 %) cumplen únicamente la condición de $fccarb > 20$, mientras que 3 660 (2.1 %) cumplen solo la condición de no reducción del carbono total. El número de parcelas que satisfacen ambas condiciones simultáneamente —y, por tanto, se consideran válidas para el entrenamiento de modelos— asciende a 53 888 (31.5 % del total). En conjunto, el 94.7 % de las parcelas cumplen al menos una de las dos condiciones de inclusión.
- **Caso carbono_bruto ($carbono_bruto3 \geq carbono_bruto4$):** igualmente, sobre las mismas 171 170 parcelas iniciales, 123 774 (72.3 %) presentan una $fccarb > 20$, y 1 738 (1.0 %) muestran un incremento neto de carbono bruto sin cumplir la primera condición. Las parcelas que cumplen ambas simultáneamente suman 34 672 (20.3 %), lo que representa el conjunto final de datos empleados para entrenamiento en este caso. El 93.6 % de las parcelas cumplen al menos una de las condiciones de inclusión.

En la Tabla ?? se resume la relación entre ambas condiciones de filtrado en cada caso, evidenciando que la fracción de cabida cubierta es el criterio más restrictivo.

Tabla 4.2: Resumen del filtrado de parcelas según las condiciones de inclusión.

Caso	Total	F1 ($fccarb > 20$)	F2	F1 \wedge F2 (entrenamiento)
C ($c3 \geq c4$)	171 170	104 558	3 660	53 888
Carbono bruto ($cb3 \geq cb4$)	171 170	123 774	1 738	34 672

Cálculo y agregación de variables

Cada registro de entrada se genera a nivel de combinación parcela–especie, incorporando las variables correspondientes de la primera medición (IFN3 en la configuración empleada) y la variable objetivo (carbono) de la segunda medición (IFN4). Las variables de parcela y *parcela_inventario* se desdoblan para cada especie. Las entradas

de la tabla `parcela_inventario_especie_cd` se agrupan por parcela y especie y se comprimen en una única entrada creando un conjunto de variables para cada clase diamétrica.

La Tabla ?? resume las variables empleadas como entrada al modelo, integradas desde las distintas tablas que conforman la base de datos relacional.

Resumen de Datos de Entrada del Modelo			
Variable	Tipo	Descripción	Anexo
parcela_id	varchar	Identificador <i>único</i> de parcela.	–
especie_id, tipo_especie, grupo_id	int (CF)	Especie (código), tipo de especie y grupo taxonómico.	Anexo ??
ocupa	int	Grado de ocupación/presencia de la especie en la parcela (0–10).	–
estado_id, fpmasa_id, tratmasa_id, orgmasa_1_id	int (CF)	Estado/fase de desarrollo, forma principal de masa, tratamiento de masa, organización de masa.	–
tipsuelo1_id, tipsuelo2_id, tipsuelo3_id	int (CF)	Tipos de suelo de la parcela (niveles jerárquicos).	Anexo ??
rocosidad_id, textura_id, matorg_id, modcomb_id, disesp_id, comesp_id, merosiva_id	int (CF)	Rociedad, textura, materia orgánica, modo de combustibilidad, distribución/composición específica, manifestaciones erosivas.	Anexo ??, ??, Anexo ??, ??, ??, ??
radio	float	Radio de la parcela (m).	–
orientacion, elevacion, pendiente	float	Orientación (grados), elevación (m s.n.m.), pendiente (%).	–
nivel1_id, nivel2_id	int (CF)	Niveles jerárquicos/estratos de inventario.	–
fccarb, fcctot	float	Fracción de cabida cubierta arbórea y total.	–
npies_{1,2,5,10,15, 20,25,30,35,40,45, 50,55,60,65,70}	float	Número de pies por clase diamétrica CD (cm). Cada campo corresponde a la CD indicada.	–
periodo	int	Años transcurridos entre inventarios considerados en el modelo.	–
evi_{stat}_{est}	float	Índice EVI por estación; stat ∈ {max, mean, median, min, std}, est ∈ {invierno, otoño, primavera, verano}.	–
gndvi_{stat}_{est}	float	Índice GNDVI por estación (misma convención de stat y est).	–
ndii_{stat}_{est}	float	Índice NDII por estación (misma convención).	–
ndvi_{stat}_{est}	float	Índice NDVI por estación (misma convención).	–
pr_{stat}_{est}	float	Precipitación: stat ∈ {max, mean, min, std, sum} por estación est.	–
skt_{stat}_{est}	float	Temperatura de superficie (skin temperature): stat ∈ {max, mean, min, std}.	–
stl1-4_{stat}_{est}	float	Temperatura de suelo por capa (1–4): stat ∈ {max, mean, min, std}.	–
t2m_{stat}_{est}	float	Temperatura del aire a 2 m: stat ∈ {max, mean, min, std}.	–

Tabla 4.3: Resumen de variables de entrada del modelo. Para variables estacionales se usa la notación `variable_{stat}_{est}`, con estadísticas `stat` y estaciones `est` en {invierno, otoño, primavera, verano}. Las variables `npies_{CD}` se repiten para cada clase diamétrica indicada.

Codificación y normalización

Las variables categóricas se codifican mediante *one-hot encoding*, generando variables binarias para cada clase. Las variables numéricas se escalan (normalización estándar o min-max, según el modelo) para asegurar que todas las magnitudes tengan el mismo orden de importancia durante el entrenamiento.

4.4. Limitación temporal y mitigación

De todos los criterios, la **permanencia de 30 años** es el único que impacta directamente en el entrenamiento: el intervalo IFN3–IFN4 es, en la mayoría de los casos, claramente inferior a dos décadas. La Figura ?? muestra la distribución de los años transcurridos entre inventarios.

El conteo de valores de la variable `periodo` confirma esta tendencia: los años más frecuentes son 15 - 17. En contraste, los valores superiores a 20 años son menos comunes, tal como se aprecia en el gráfico.

Para reducir el riesgo de sesgo temporal, se emplea validación cruzada por *grupos de periodo* y se evalúa la robustez del modelo ante cambios en el horizonte temporal.

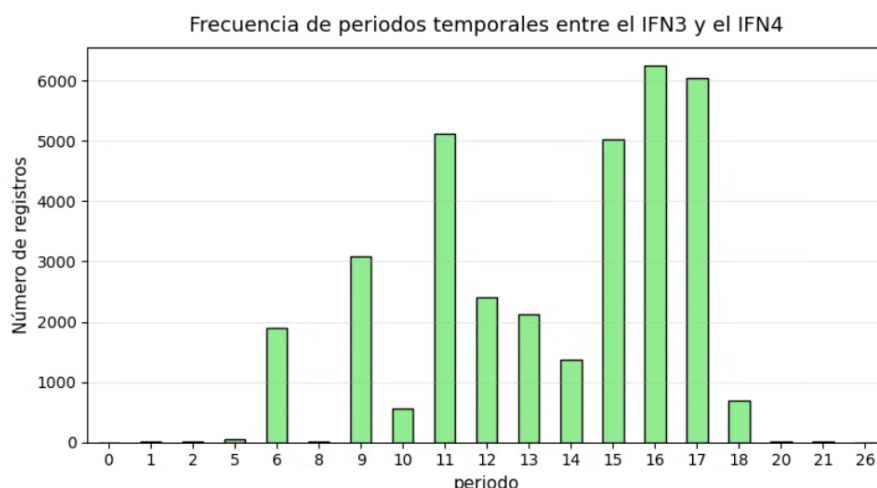


Figura 4.2: Distribución del número de años transcurridos entre la inventariación tercera y cuarta de cada parcela (considerando únicamente las parcelas inventariadas en ambas ocasiones). Se aprecia un intervalo modal entre 15 y 17 años, con valores superiores a 20 muy poco frecuentes.

4.5. Partición y validación

Para obtener una estimación imparcial del rendimiento y evitar *fugas de información* debidas a la correlación espacial dentro de cada parcela, la partición del conjunto de datos se realiza **por identificador de parcela** (`parcela_id`). Todas las observaciones asociadas a una misma parcela se asignan *íntegramente* a un único subconjunto, de modo que ninguna parcela aparece simultáneamente en entrenamiento y evaluación. Tras aplicar los filtros de preprocesamiento, se dispone de **XXX** registros correspondientes a **XXX** parcelas, que se dividen en un **80 %** para entrenamiento y un **20 %** para evaluación.

Validación interna y control de sesgo temporal. Sobre el subconjunto de entrenamiento (80 %) se aplica *validación cruzada por grupos* utilizando como agrupador los *años transcurridos entre inventarios* (p. ej., 15, 16, 17, ...). Esta estrategia comprueba la *estabilidad* del modelo frente a cambios en el horizonte temporal y reduce el riesgo de sobreajuste específico de un periodo. La selección de hiperparámetros se realiza exclusivamente dentro de esta validación interna; el conjunto de evaluación (20 %) permanece *sellado* para la prueba final.

Métricas de evaluación. El rendimiento se informa con dos medidas complementarias:

- **RMSE (Root Mean Squared Error):** raíz del error cuadrático medio entre valores observados y predichos; se expresa en las mismas unidades que la variable objetivo y penaliza con mayor peso los errores grandes. Valores más bajos indican mejor ajuste.
- **R^2 (coeficiente de determinación):** proporción de la varianza observada explicada por el modelo (idealmente en $[0, 1]$). Valores cercanos a 1 denotan alta capacidad explicativa; puede ser negativo si el modelo es peor que la predicción constante.

Protocolo de reporte. Para cada modelo se reportan: (i) el rendimiento medio y la dispersión en la validación cruzada por grupos (entrenamiento), y (ii) el desempeño final en el conjunto de evaluación independiente (20 %). Este protocolo garantiza comparabilidad entre modelos, control del sesgo espacial por parcela y verificación explícita de la robustez temporal.

4.6. Modelos evaluados

A continuación, se detalla el diseño general y las estrategias empleadas para la selección y optimización de modelos.

Entrenamiento y optimización

Se aplicó *RandomizedSearchCV*, una técnica de búsqueda aleatoria de hiperparámetros que evalúa distintas combinaciones utilizando validación cruzada. Este procedimiento permite optimizar el rendimiento de cada modelo sin incurrir en un coste computacional tan elevado como el de una búsqueda exhaustiva.

VALIDACIÓN CRUZADA POR GRUPOS. En algunas configuraciones probadas, se emplea la validación cruzada por grupos (*Group k-Fold Cross Validation*). Este método divide el área de estudio en k bloques, basados en una característica común de los datos, asegurando que los datos dentro de cada bloque estén relacionados. El modelo se entrena con $k - 1$ bloques y se valida con el bloque restante, repitiendo este proceso k veces.

Este enfoque es útil cuando los datos tienen agrupaciones naturales, como por ejemplo, diferentes parcelas o periodos de tiempo. Al mantener los datos relacionados en un mismo bloque, se evita la filtración de información entre los conjuntos de entrenamiento y validación, lo que permite una mejor evaluación de la capacidad de generalización del modelo. Así, se asegura que el modelo no se sobreajuste y sea robusto al ser

evaluado en contextos no vistos previamente.

Modelos ensemble

Para mejorar la precisión y robustez, se emplearon diversos métodos de *ensemble learning*, que combinan múltiples modelos base (*base learners*) para generar una predicción agregada. Esta estrategia se inspira en la teoría de la sabiduría colectiva: la combinación de estimaciones independientes tiende a superar a cualquier estimador individual.

TÉCNICAS UTILIZADAS:

- **Voting y Averaging:** combinan modelos ya entrenados mediante votación mayoritaria o promedio.
- **Bagging y Boosting:** construyen modelos desde cero y los combinan. Bagging reduce la varianza al entrenar modelos en subconjuntos aleatorios; Boosting mejora el sesgo al entrenar secuencialmente, corrigiendo errores anteriores.
- **Stacking:** combina modelos optimizados usando un metamodelo que aprende a integrar sus predicciones.

Boosting y aprendizaje gradual

El *boosting* se basa en el aprendizaje secuencial, donde cada nuevo modelo intenta corregir los errores residuales del anterior. Esta técnica permite construir modelos fuertes a partir de modelos débiles, alcanzando gran precisión. Sin embargo, requiere una cuidadosa configuración de hiperparámetros para evitar el sobreajuste.

Entre las implementaciones destacadas se incluyen:

- **XGBoost:** modelo GBM que optimiza rendimiento con gradientes de primer y segundo orden, regularización L1/L2, manejo automático de valores faltantes, y técnicas de generalización como *shrinkage* y *column subsampling*.
- **LightGBM:** algoritmo eficiente para grandes volúmenes de datos, con crecimiento *leaf-wise* y soporte nativo para variables categóricas.
- **AdaBoost:** ajusta modelos simples secuencialmente, enfocando el aprendizaje en observaciones mal clasificadas.
- **CatBoost:** especializado en variables categóricas y robusto frente a datos ruidosos, usando codificación por orden aleatorio.
- **Gradient Boosting Decision Trees (GBDT):** construye árboles secuenciales ajustados a residuos, optimizando mediante descenso por gradiente.

Bagging

El *bagging* (Bootstrap Aggregating) entrena múltiples modelos independientes sobre subconjuntos de datos generados por muestreo con reemplazo. Las predicciones se combinan por promedio o votación. Esta técnica reduce la varianza y mejora la estabilidad de modelos inestables.

- **Random Forest:** combina árboles de decisión (CART) con selección aleatoria de características en cada división. Es escalable, robusto a datos faltantes, y menos propenso al sobreajuste.

- **Bagged Decision Trees (BaggedDT):** genera árboles sin poda entrenados en muestras bootstrap. Promedia sus predicciones para reducir la varianza.

Otros modelos utilizados

Además de los métodos ensemble, se evaluaron modelos representativos de distintos paradigmas de aprendizaje supervisado:

- **K-Nearest Neighbors (KNN):** modelo basado en instancia que predice a partir de los vecinos más cercanos. Sensible a la escala y a *outliers*.
- **Multi-Layer Perceptron (MLP):** red neuronal con una o más capas ocultas, capaz de modelar relaciones no lineales complejas.
- **Support Vector Regression (SVR):** modelo de márgenes para regresión, con soporte para kernels no lineales.
- **SVM con kernel:** modelo poderoso para clasificación y regresión no lineal, aunque costoso y sensible a hiperparámetros.
- **Bayesian Neural Network:** enfoque probabilístico que estima incertidumbre en las predicciones. Incluye variantes como la *Bayesian Ridge Regression*.
- **Naive Bayes:** clasificador probabilístico rápido y simple, útil en texto y alta dimensionalidad. Se evaluaron variantes:
 - *Gaussian Naive Bayes:* para datos continuos.
 - *Multinomial Naive Bayes:* para conteos y texto.
 - *Bernoulli Naive Bayes:* para variables binarias.

Comparación y justificación de modelos

La evaluación de múltiples modelos responde a la necesidad de identificar no solo el de mejor rendimiento, sino también el más adecuado según la naturaleza del problema y los datos disponibles. Se compararon algoritmos lineales, no lineales, basados en vecinos, redes neuronales, modelos probabilísticos y diferentes técnicas de *ensemble*. Vemos un resumen de los modelos aplicados en la tabla ??.

Modelo	Tipo / Técnica	Características destacadas	Observaciones
Random Forest	Bagging (Árboles)	Uso de bootstrap, selección aleatoria de atributos, reducción de varianza	Robusto y escalable; menor interpretabilidad
Bagged Decision Trees (BaggedDT)	Bagging	Árboles sin poda, entrenados en paralelo sobre muestras con reemplazo	Preciso pero costoso computacionalmente
XGBoost	Boosting (GBM)	Regularización L1/L2, manejo de valores faltantes, poda anticipada	Alto rendimiento, sensible a hiperparámetros
LightGBM	Boosting (Leaf-wise)	Crecimiento hoja a hoja, eficiente en grandes volúmenes	Rápido y preciso; riesgo de sobreajuste
AdaBoost	Boosting (Stumps)	Aumenta peso de errores, pondera modelos por precisión	Sencillo y efectivo con datos limpios
CatBoost	Boosting especializado	Codificación avanzada de variables categóricas, robustez a ruido	Ideal para datos heterogéneos
Gradient Boosting Decision Trees (GBDT)	Boosting	Árboles secuenciales ajustados a residuos	Buen rendimiento; mayor coste de entrenamiento
K-Nearest Neighbors (KNN)	Basado en instancia	No requiere entrenamiento, predice por proximidad	Sensible a escala y outliers
Multi-Layer Perceptron (MLP)	Red neuronal	Modela relaciones no lineales complejas	Requiere normalización y regularización
Support Vector Regression (SVR)	Kernel y márgenes	Predicción dentro de tolerancia ϵ , uso de kernels no lineales	Robusto; elevado coste computacional
SVM con kernel	SVM no lineal	Maximiza margen, admite distintos kernels (RBF, polinomial, etc.)	Alta precisión; sensible a hiperparámetros
Bayesian Neural Network / Ridge Regression	Probabilístico / Bayesiano	Predicción con incertidumbre, estimación automática de hiperparámetros	Útil para inferencia y regularización
Naive Bayes (Gaussian, Multinomial, Bernoulli)	Probabilístico	Asume independencia condicional, rápido y simple	Eficaz en texto y alta dimensionalidad

Tabla 4.4: Resumen de modelos de aprendizaje supervisado aplicados

4.7. Supuestos de elegibilidad y verificación externa

Para que un proyecto forestal sea elegible en programas de *créditos de carbono*, debe cumplir requisitos técnicos establecidos por marcos regulatorios internacionales [? ?].

A continuación se resume cada criterio y la forma en que se aborda en este estudio:

- **Intervención humana directa.** El incremento de carbono debe proceder de actuaciones planificadas (reforestación, restauración o manejo sostenible). En nuestro caso, el modelo se entrena sobre datos observacionales (IFN3–IFN4); por tanto, la *verificación de intervención* no se deduce del modelo, sino que se contempla como *condición externa* de elegibilidad del proyecto a evaluar.
- **Permanencia mínima de 30 años.** Este es el criterio más restrictivo en relación

con los datos disponibles, dado que el intervalo entre inventarios IFN3 e IFN4 no supera los 18 años. En consecuencia, el modelo se entrena y valida en horizontes de 15–18 años, evaluando su estabilidad temporal mediante validación cruzada estratificada por periodo.

Aunque los datos no cubren los 30 años requeridos por los estándares de proyectos de carbono, los resultados obtenidos permiten plantear una *extrapolación cautelosa* siempre que se cumplan ciertas condiciones. En particular, un modelo que mantenga un desempeño estable y coherente dentro del rango observado puede extenderse a horizontes mayores si:

1. se **reporta la incertidumbre** asociada a las predicciones, mediante intervalos de confianza o bandas de predicción;
2. se realiza un **análisis de sensibilidad temporal**, explorando escenarios conservadores y pesimistas;
3. y se verifica externamente la **validez de las condiciones del proyecto** —superficie, altura mínima e intervención humana— conforme a los marcos regulatorios aplicables.

Este enfoque combina una base empírica sólida (validación cruzada en horizontes observados) con una verificación externa de elegibilidad, permitiendo aplicar el modelo a contextos de mayor duración sin comprometer la trazabilidad ni la consistencia metodológica.

- **Superficie mínima de 1 ha.** Este criterio se considera *externo* al alcance del modelo predictivo, ya que el aprendizaje se realiza a nivel de parcela e inventario y no sobre polígonos de superficie total. En la práctica, la verificación de la superficie se realiza *ex ante*, sobre la geometría declarada del proyecto forestal. En los terrenos forestales generados a partir de intervención humana directa —como plantaciones o repoblaciones—, la extensión suele presentar una estructura homogénea, con una especie dominante, edades coetáneas y densidades estandarizadas. Bajo estas condiciones, el carbono total es proporcional a la superficie: duplicar el área de una masa forestal homogénea implica aproximadamente duplicar su carbono almacenado. Por tanto, la variable de superficie no afecta al ajuste interno del modelo y su cumplimiento puede evaluarse fácilmente a nivel de proyecto, sin comprometer la validez de las predicciones.
- **Fracción mínima de cabida cubierta del 20 %.** La base de datos dispone de f_{ccarb} (arbórea) y f_{cctot} (total). Este umbral se aplica como *filtro de elegibilidad* previo o posterior al modelado, sin modificar la arquitectura del modelo.
- **Altura mínima de 3 m en la madurez.** Este requisito se refiere a la altura que alcanzan los árboles en su fase de pleno desarrollo, y no a la altura inicial de los plantones. Por tanto, las mediciones realizadas durante las etapas tempranas de crecimiento no determinan la elegibilidad del proyecto, siempre que las especies seleccionadas sean capaces de superar los 3 metros en la madurez. En nuestro conjunto de datos, la altura no se registra explícitamente, por lo que este criterio se evalúa de forma *externa* al modelo, mediante la selección de especies forestales adecuadas y la verificación con fuentes auxiliares (catálogos silvícolas o tipologías de masa). En la práctica, el cumplimiento del requisito depende de una decisión de diseño del proyecto —*no plantar especies cuyo tamaño adulto sea inferior a 3 metros*— más que del ajuste predictivo del modelo. Por ello, la altura no interviene directamente en el entrenamiento, aunque sí condiciona la

elegibilidad final del proyecto forestal.

5. IMPLEMENTACIÓN DE LOS MODELOS

El desarrollo y evaluación de los modelos predictivos se realizó íntegramente en **Python** y el entorno de ejecución fue local, en un equipo con procesador Intel Core i7 y 32 GB de RAM, lo que permitió realizar experimentos de forma eficiente con un conjunto de datos de tamaño considerable (~80.000 muestras).

El preprocesamiento de datos se llevó a cabo mediante la librería `scikit-learn`, utilizando `Pipeline` y `ColumnTransformer` para combinar transformaciones numéricas y categóricas. En particular, las variables numéricas se imputaron con la mediana y se escalaron con `StandardScaler`, mientras que las variables categóricas se trataron mediante imputación por moda y codificación *one-hot*. La función objetivo a predecir fue el carbono total (CZ) acumulado en cada parcela.

Se implementaron y optimizaron diversos modelos de regresión supervisada, incluyendo:

- **Modelos basados en árboles:** `RandomForestRegressor`, `XGBoost`, `LightGBM`, `CatBoost`, `GradientBoosting`, `AdaBoost` y `Bagging`.
- **Modelos basados en instancias:** `KNeighborsRegressor`.
- **Modelos de redes neuronales:** `MLPRegressor`.
- **Modelos de soporte vectorial:** `SVR`.
- **Modelos probabilísticos:** `BayesianRidge`.

La optimización de hiperparámetros se realizó mediante `RandomizedSearchCV`, con validación cruzada de 5 particiones y búsqueda en espacios definidos manualmente para cada modelo. Los modelos fueron evaluados en términos de R^2 y **RMSE**, tanto en el conjunto de entrenamiento como en el de prueba.

Con el objetivo de mejorar el rendimiento predictivo, se evaluaron además varias configuraciones de `StackingRegressor`, combinando distintos subconjuntos de modelos base (previamente entrenados) con diversos meta-modelos (`LinearRegression`, `Ridge`, `GradientBoosting`, `SVR`, `MLP`, entre otros). Estas combinaciones permitieron comparar sinergias entre modelos complementarios.

El tiempo de entrenamiento varió según el modelo y la configuración de hiperparámetros. Gracias al uso de `n_jobs=-1` se aprovechó el paralelismo multinúcleo para acelerar la optimización.

6. RESULTADOS

Presentar los resultados obtenidos al aplicar el modelo a los datos de entrada. Incluir gráficos y tablas que ayuden a ilustrar el rendimiento del modelo.

7. DISCUSIÓN

Interpretar los resultados obtenidos, comparándolos con investigaciones previas. Discutir las limitaciones del modelo y las posibles áreas de mejora.

8. CONCLUSIONES

Recapitular las conclusiones más importantes del estudio. Resaltar la relevancia del modelo desarrollado y su aplicación en proyectos de forestación y reforestación.

9. RECOMENDACIONES PARA FUTURAS INVESTIGACIONES

Sugerir áreas que podrían beneficiarse de estudios adicionales o mejoras en la metodología.

AGRADECIMIENTOS

Agradecer a las personas o instituciones que hayan apoyado tu trabajo, como supervisores, financiadores, etc.

Investigación financiada por la subvención **TSI-100933-2023-1** de la **Convocatoria de Cátedras Universidad-Empresa (Cátedras ENIA 2022)**, **Ministerio de Transformación Digital y Función Pública de España**, y el **Plan de Recuperación y Resiliencia de la UE** (*NextGenerationEU/PRTR*).

10. ANEXOS

Anexo: Códigos de provincias de España

Código	Provincia	Código	Provincia
01	Álava	27	Lugo
02	Albacete	28	Madrid
03	Alicante	29	Málaga
04	Almería	30	Murcia
05	Ávila	31	Navarra
06	Badajoz	32	Ourense
07	Baleares (Illes)	33	Asturias
08	Barcelona	34	Palencia
09	Burgos	35	Palmas (Las)
10	Cáceres	36	Pontevedra
11	Cádiz	37	Salamanca
12	Castellón	38	Santa Cruz de Tenerife
13	Ciudad Real	39	Cantabria
14	Córdoba	40	Segovia
15	Coruña (A)	41	Sevilla
16	Cuenca	42	Soria
17	Girona	43	Tarragona
18	Granada	44	Teruel
19	Guadalajara	45	Toledo
20	Guipúzcoa	46	Valencia
21	Huelva	47	Valladolid
22	Huesca	48	Vizcaya
23	Jaén	49	Zamora
24	León	50	Zaragoza
25	Lleida	51	Ceuta
26	La Rioja	52	Melilla

Tabla 10.1: Relación de códigos numéricos utilizados para las provincias españolas.

Anexo: Rocosidad

Se considerará el conjunto de la parcela clasificando la rocosidad según la siguiente codificación:

1. **Sin pedregosidad:** la superficie de la parcela está completamente cubierta de vegeta-

ción.

2. **Poco pedregoso:** cuando la superficie de la parcela cubierta por rocas coherentes es menor del 25 %.
3. **Pedregoso:** cuando la superficie rocosa está comprendida entre el 25 % y el 50 %.
4. **Muy pedregoso:** cuando la superficie rocosa se sitúa entre el 50 % y el 75 %.
5. **Roquedo:** cuando la superficie de rocas es mayor del 75 %. En este caso, no se tomará ningún dato adicional correspondiente a suelos.

Anexo: Tipo de Suelo

Se utilizará la siguiente codificación para el tipo de suelo, diferenciando tres variables:

Tipo de suelo (I): Presencia de sales, yesos o hidromorfía

1. **No se observan sales, yesos ni procesos de hidromorfía.**
2. **Suelo salino.** Si presenta al menos dos de las siguientes características:
 - Presencia de eflorescencias en la superficie o a distintas profundidades.
 - Existencia de plantas halófitas.
 - Zonas llanas o endorreicas con climas secos que provocan gran evaporación.
3. **Suelo yesífero.** Si presenta alguna de las siguientes características:
 - Presencia de materia yesífera en superficie o a distintas profundidades.
 - Existencia de plantas gipsófilas.
4. **Suelo hidromorfo.** Si el suelo presenta síntomas de hidromorfía acusada, cumpliendo al menos dos de las siguientes:
 - Zona encharcada permanente o casi permanentemente de forma natural.
 - Zona llana o endorreica con climas húmedos.
 - Grietas en verano si no hay encharcamiento.
 - Presencia de vegetación indicadora de hidromorfismo.

Identificándose las siguientes:

- Formaciones vegetales indicadoras de hidromorfía:
 - Ribereñas: *saucedas*, *mimbreras*, *alisedas*.
 - Brezales con *Erica ciliaris*, *Erica tetralix*.
 - Turberas arboladas (excepto Cornisa Cantábrica y Pirineos).
 - Turberas de montaña con *Sphagnum*, *Erica tetralix*.
 - Cervunales con *Nardus stricta*.
 - Carrizales y espadañares (*Phragmites*, *Typha*, *Cladium*).
 - Juncales (*Scirpus*, *Juncus*).
 - Pastizales con cárices (*Carex spp.*).
 - Marismas.
- Formaciones vegetales gipsófilas:
 - Aznallar: matorral de *Ononis tridentata*.
 - Tomillares gipsófilos con:
 - *Lepidium subulatum*
 - *Gypsophila spp.*
 - *Matthiola fruticulosa*
- Formaciones vegetales indicadoras de suelos salinos:
 - Salicorniales: matas leñosas crasas (*Salicornia*, *Arthrocnemum*, *Halozyllon*).
 - Bosques halófitos del género *Tamarix*.
 - Saladar o sosar: predominio de *Suaeda vera*.

- Saladar blanco: predominio de *Atriplex halimus*.

Tipo de suelo (II y III): Composición del suelo (calizo o silíceo)

1. **Suelo calizo.** Más del 50 % de la vertical del perfil da efervescencia con ácido clorhídrico.
 - **Moderadamente básico:** pH en superficie 8.5.
 - **Fuertemente básico:** pH en superficie >8.5.
2. **Suelo silíceo.** Menos del 50 % de la vertical del perfil da efervescencia.
 - **Moderadamente ácido:** pH 5.5.
 - **Fuertemente ácido:** pH <5.5.

Anexo: Manifestaciones Erosivas

Se observará la parcela y sus alrededores hasta una distancia de 60 metros desde el centro, y se codificará la existencia de manifestaciones erosivas según la siguiente clave:

1. **No hay ninguna manifestación.**
2. **Cuellos de raíces al descubierto:** los cuellos de las raíces están visibles, con acumulación de residuos aguas arriba de los tallos y obstáculos, así como abundancia superficial de piedras.
3. **Presencia de regueros:** canales paralelos de erosión con una profundidad máxima de un palmo (aproximadamente 20 cm).
4. **Cárcavas y barrancos en V:** erosión lineal más profunda que los regueros, con forma de “V”.
5. **Cárcavas y barrancos en U:** erosión avanzada con formas suavizadas y amplias en “U”.
6. **Deslizamientos del terreno:** desplazamientos de masas de tierra, ladera o materiales del suelo.

Anexo: Distribución Espacial

La disposición de la vegetación en el espacio se clasificará según la siguiente codificación:

1. **Uniforme.** Cuando el estrato arbóreo presenta continuidad en el espacio.
2. **Diseminada en bosquetes aislados.** Cuando la masa arbórea se encuentra dividida en porciones que tienen una superficie inferior a 0,5 ha.
3. **Diseminada en individuos aislados.** Cuando se trata de dehesas.
9. **Otras o no se sabe.** En caso diferente a los anteriores o si se desconoce el dato exacto.

Anexo: Composición Específica

En función de las especies presentes:

1. **Masas homogéneas o puras.** Masas monoespecíficas con una única especie arbórea. La normativa española precisa que una masa es monoespecífica o pura cuando al menos el 90 % de los pies pertenecen a la misma especie.
2. **Masas heterogéneas o mezcladas pie a pie.** Masas de diferentes especies que se juntan o bien se entremezclan por golpes o grupos, siempre que tengan una altura similar.
3. **Masas heterogéneas o mezcladas con subpiso.** Las dos o más especies mezcladas, cuando alcancen el estado adulto y la estabilidad, presentarán alturas diferentes.
9. **Otras o no se sabe.** En caso diferente a los anteriores o desconocer el dato exacto.

Anexo: Textura del Suelo

Se clasificará en función de la siguiente codificación:

1. **Suelo arenoso.** Si los cilindros se deshacen sin apenas formarse.
2. **Suelo franco.** Es posible hacer cilindros gruesos pero no delgados.
3. **Suelo arcilloso.** Se consiguen cilindros de unos 5 mm de diámetro.

Anexo: Nivel de usos del suelo

1. **Monte.** Toda superficie en la que vegetan especies arbóreas, arbustivas, de matorral o herbáceas, ya sea espontáneamente o procedan de siembra o plantación, siempre que no sean características de cultivo agrícola o fueran objeto del mismo.
2. **Agrícola.** Territorio o ecosistema poblado con siembras o plantaciones de herbáceas y/o leñosas, anuales o plurianuales que se laborean con una fuerte intervención humana, puede estar poblado por especies forestales de fruto (flor, hojas o en el futuro biomasa) siempre que la intervención humana sea importante. Incluye las dehesas, montes huecos o montes adehesados de base cultivo, siempre que la fracción de cabida cubierta de los árboles sea inferior al 5 %.
3. **Artificial.** Territorio o ecosistemas dominado por edificios, parques urbanos (aunque estén poblados de árboles), viveros fuera de los montes (aunque sean de especies forestales), carreteras (salvo las vías de servicio de los montes) u otras construcciones humanas que tengan superficies continuas.
4. **Humedal.** Lo constituyen las lagunas, charcas, zonas húmedas, marismas y corrientes discontinuas de agua en las que, al menos durante 6 meses del año, esté presente dicho líquido.
5. **Agua.** Es la parte de la tierra constituida por ríos, lagos, embalses, canales o estanques con superficies continuas de más de 0.26 ha y con agua prácticamente todo el año.

Anexo: Nivel morfoestructural

Para el nivel de usos del suelo Monte se definirán los siguientes niveles morfoestructurales.

1. **Monte arbolado.** Territorio o ecosistema con especies forestales arbóreas como manifestación vegetal de estructura vertical dominante y con una fracción de cabida cubierta igual o superior al 20%; incluye dehesas con base cultivo o pastizal con labores siempre que la fracción arbolada supere el 20%, y excluye terrenos con fuerte intervención humana para obtener frutos, hojas, flores o varas.
2. **Monte arbolado ralo.** Terreno de uso forestal con especies arbóreas forestales dominantes y fracción de cabida cubierta entre el 10 % y 20 % (incluido el 10 %, excluido el 20 %); también aplica a terrenos con matorral o pastizal natural como dominantes, pero con presencia importante de árboles forestales, incluyendo dehesas de base de cultivo.
3. **Monte temporalmente desarbolado.** Terreno que fue monte arbolado recientemente y que casi con seguridad volverá a estar cubierto de árboles en un futuro próximo.
4. **Monte desarbolado.** Terreno con matorral y/o pastizal natural o débil intervención humana como cobertura dominante, con fracción de cabida cubierta por árboles forestales inferior al 5 %.
5. **Monte sin vegetación superior.** Terreno de uso forestal que no está poblado por vegetales superiores debido a condiciones actuales de suelo, clima o topografía, aunque podría estarlo en otras circunstancias.

6. **Árboles fuera del monte.** Incluye riberas arboladas no estructuradas con los montes, bosquetes de menos de 2.500 m², alineaciones de especies arbóreas o arbustivas de menos de 25 m de anchura, y árboles sueltos en terreno forestal.
7. **Monte arbolado disperso.** Terreno forestal con especies arbóreas dominantes y fracción de cabida cubierta entre el 5 % y el 10 % (incluido el 5 %, excluido el 10 %); también terrenos con matorral o pastizal como cobertura dominante pero con presencia significativa de árboles forestales, incluyendo dehesas de base cultivo.

Anexo: Contenido en Materia Orgánica

Según la siguiente clasificación:

1. **Suelo muy húmifero.** Cuando a 15 cm la pureza es menor de 4, o cuando la capa de broza sea de espesor mayor de 5 cm y a 15 cm de profundidad la pureza sea menor de 6.
2. **Suelo moderadamente húmifero.** Cuando a 15 cm la pureza sea menor de 6 con capa de broza nula o de escaso espesor, o cuando dicha capa tenga espesor mayor de 5 cm y a 15 cm de profundidad la pureza sea igual o mayor de 6.
3. **Suelo poco húmifero.** En los restantes casos.

Anexo: Reacción del Suelo (pH)

En función del pH, el suelo se clasifica según la siguiente codificación:

Valores del pH de la solución del suelo	Clasificación del suelo	Codificación
1	Suelo extremadamente ácido	1
2	Suelo muy fuertemente ácido	2
3–4	Suelo fuertemente ácido	3
5–6	Suelo moderadamente ácido	4
7	Suelo neutro	5
8	Suelo moderadamente básico	6
9	Suelo fuertemente básico	7
10	Suelo extremadamente básico	8

Tabla 10.2: Clasificación del suelo según el pH de la solución del suelo.

Anexo: Espesor de la Capa Muerta, Césped, Musgo y Líquenes

Se anotará con la siguiente codificación:

- Espesor menor de 0,5 cm **00**
- Espesor de 0,5 a 1,4 cm **01**
- Espesor de 1,5 a 2,4 cm **02**
- Espesor de 2,5 a 3,4 cm **03**
- *Y así sucesivamente.*

Si en la parcela hay zonas con diferentes espesores de capa muerta, se anotará el valor medio estimado.

10.1. Anexo: Modelo de Combustible

Se determinará la clase de combustible que es más probable que propague el fuego si hubiese un incendio en la zona, hasta un máximo de 60m: pasto, matorral, hojarasca de bosque o desechos o restos de corta. Se determinará el modelo de combustible a partir de la siguiente clave:

GRUPO	MOD. COMBUS-TIBLE	DESCRIPCIÓN DEL MODELO
PASTOS	1	Pasto fino, seco y bajo, que recubre completamente el suelo. Puede aparecer algunas plantas leñosas dispersas ocupando menos de 1/3 de la superficie.
	2	Pasto fino, seco y bajo, que recubre completamente el suelo. Las plantas leñosas dispersas cubren de 1/3 a 2/3 de la superficie; pero la propagación del fuego se realiza por el pasto.
	3	Pasto grueso, denso, seco y alto (>1 m). Puede haber algunas plantas leñosas dispersas. Los campos de cereales son representativos de este modelo.
MATORRAL	4	Matorral o plantación joven muy densa; de más de 2 m de altura; con ramas muertas en su interior. Propagación del fuego por las copas de las plantas.
	5	Matorral disperso, denso y verde, de menos de 1 m de altura. Propagación del fuego por la hojarasca, el pasto, las ramillas y el matorral.
	6	Parecido al modelo 5, pero con especies más inflamables, de mayor talla, pudiéndose encontrar ramas gruesas en el suelo. Propagación del fuego con vientos moderados a fuertes.
	7	Matorral de especies muy inflamables; de 0.5 a 2 m de altura, situado como sotobosque en masas de coníferas.
HOJARASCA BAJO ARBOLADO	8	Bosque denso, sin matorral. Propagación del fuego por la hojarasca muy compacta, formada por acículas cortas (5 cm o menos) o por hojas planas no muy grandes.
	9	Parecido al modelo 8, pero con hojarasca menos compacta, formada por acículas largas y rígidas (P. pinaster) o follaje de frondosas de hoja grande, caducas (castaño o robles).
	10	Bosque con gran cantidad de leña y árboles caídos, como consecuencia de vendavales, plagas intensas, etc.
RESTOS DE CORTA Y OPERACIONES SELVÍCOLAS	11	Bosque claro y fuertemente aclarado. Restos de poda o aclareo ligeros (diámetro <7.5 cm).
	12	Predominio de los restos sobre el arbolado. La hojarasca y el matorral presente ayudarán a la propagación del fuego.
	13	Grandes acumulaciones de restos gruesos y pesados, cubriendo todo el suelo.

Tabla 10.3: Descripción de los modelos de combustible del Inventario Forestal Nacional, clasificados por grupo funcional.

Anexo: Categoría de Desarrollo y Densidad

La categoría de desarrollo se identifica en función de la altura y el diámetro de los pies de las diferentes especies. Cuando el 85 % de los ejemplares pertenecen a una determinada categoría, se considerarán todos dentro de la misma.

1. Pies con altura inferior a 30 cm.
2. Pies con altura comprendida entre 30 y 130 cm.
3. Pies con altura superior a 130 cm y diámetro normal menor de 2,5 cm.

4. Pies con altura superior a 130 cm y diámetro normal comprendido entre 2,5 y 7,5 cm. Corresponde a los pies menores del IFN-2.

La densidad se cuantifica según la categoría de desarrollo:

Para las categorías 1, 2 y 3 (radio de parcela = 5 m):

1. **Escasa:** De 1 a 4 pies en la parcela.
2. **Normal:** De 5 a 15 pies en la parcela.
3. **Abundante:** Más de 15 pies en la parcela.

Para la categoría 4:

- Se cuenta el número exacto de pies por especie en la subparcela de 5 m de radio. Se registra en la casilla "N" y se calcula aproximadamente la altura media total de cada grupo.

Nota: Si aparecen más de 40 pies en las categorías 1, 2 o 3, el conteo puede ser estimado. Los pies menores muertos no se contabilizan. Para brotes de cepa, cada uno se considera como una planta.

Anexo: Tipo de Regeneración

Se identifica el origen de los pies con la siguiente clave:

1. **Siembra o semilla.**
2. **Plantación.**
3. **Brote de cepa o raíz.**
4. **Desconocido.**
5. **Dudoso.**
6. **Mixto.**

Anexo: Estado de las Poblaciones (IFN3 e IFN4)

Se determinará las fases de desarrollo de las *poblaciones* codificándose de la siguiente forma:

1. **Repoblado.** Conjunto de pies que desde el estrato herbáceo llega hasta el subarborescente y los pies inician la tangencia de copas.
2. **Monte bravo.** Comprende desde el estrato y clase de edad anterior hasta el momento en que por efecto del crecimiento, los pies empiezan a perder las ramas inferiores; es decir que en esta clase de edad, las ramas se encuentran a lo largo de todo el fuste.
3. **Latizal.** Comprende desde la clase anterior hasta que los pies tienen 20 cm de diámetro normal; es decir, el diámetro de su fuste, medido a la altura de 1,30 m del suelo.
4. **Fustal.** Se caracteriza esta clase de edad, porque sus pies tienen diámetros normales superiores a 20 cm.

Anexo: Forma Principal de Masa (IFN3 e IFN4)

1. **Coetánea.** Cuando al menos el 90 % de sus pies tienen la misma edad individual. Ejemplo típico: las repoblaciones.

2. **Regular.** Cuando al menos el 90 % de sus pies pertenecen a la misma clase artificial de edad o misma clase diamétrica en su defecto.
3. **Semirregular.** Cuando al menos el 90 % de sus pies pertenecen a dos clases artificiales de edad cíclicamente contiguas o dos clases diamétricas contiguas en su defecto.
4. **Irregular.** Cuando no se cumplen las condiciones anteriores, es decir, cuando en cualquier parte de la masa existen pies más o menos mezclados, de todas las clases de edad que tiene la masa o de varias clases diamétricas en su defecto.

Anexo: Tratamiento de la Masa (IFN3 e IFN4)

1. **Monte alto.** Cuando todos los pies proceden de semilla.
2. **Monte medio.** Cuando coexisten pies de la misma especie, unos procedentes de semilla (brinzales) y otros de brote (chirpiales).
3. **Monte bajo.** Cuando todos los pies proceden de brote de cepa o de raíz.

Anexo: Origen de la Masa (IFN3 e IFN4)

1. **Natural.** Bosque desarrollado espontáneamente, sin intervención humana directa.
2. **Artificial.** Plantado intencionadamente por el ser humano.
3. **Naturalizado.** Bosque originalmente plantado pero que ha evolucionado hacia una estructura más similar a un bosque natural.

Anexo: MASA (IFN2)

- 1. Artificial.
- 2. Natural regular.
- 3. Natural irregular.
- 9. Dudoso.

Anexo: Estado (IFN2)

- 0. Repoblado.
- 1. Monte bravo-repoblado.
- 2. Monte bravo.
- 3. Latizal-monte bravo.
- 4. Latizal.
- 5. Fustal-latizal.
- 6. Fustal.

Anexo: Origen (IFN2)

1. Siembra o semilla.
2. Plantación.
3. Brote de cepa o raíz.
4. Desconocido.
5. Dudoso.
6. Mixto.

10.2. Anexo: Código de las especies

Tabla 10.4: Relación de especies con claves IFN3 e IFN4. Esta tabla se construye cruzando información de las tablas CambioEspecie de la base de datos de Sig de IFN3 e IFN4 y la tabla ESPECIES ARBÓREAS Y ARBUSTIVAS incluida tanto en el documentador de Sig del IFN4 como en el documentador de Sig del IFN3.

Clave	Nombre especie	Sinonimias	Claves IFN3	Claves IFN4
1	Heberdenia bahamensis	Heberdenia excelsa		
2	Amelanchier ovalis	Guillomo		
3	Frangula alnus	Rhamnus frangula		
4	Rhamnus alaternus	Aladierno		
5	Euonymus europaeus			
6	Myrtus communis			
7	Acacia spp.		[7, 92, 207, 307]	[7, 90, 99, 207, 307]
8	Phillyrea latifolia			[8, 90, 99, 999]
9	Cornus sanguinea			
10	Sin asignar	Sin asignar		
11	Ailanthus altissima	Ailanthus glandulosa		
12	Malus sylvestris			[12, 70, 99, 999]
13	Celtis australis			[13, 99]
14	Taxus baccata			[14, 19, 21, 22]
15	Crataegus spp.		[4, 15, 95, 215, 295, 315]	[15, 99, 215]
16	Pyrus spp.			[16, 70, 99, 395, 999]
17	Cedrus atlantica			
18	Chamaecyparis lawsoniana		[18, 319]	[18, 19, 28, 34, 35]

Clave	Nombre especie	Sinonimias	Claves IFN3	Claves IFN4
19	Otras coníferas		[14, 17, 18, 19, 21, 22, 23, 24, 25, 26, 28, 31, 32, 33, 34, 35, 36, 37, 38, 39, 217, 219, 235, 236, 237, 238, 239, 319, 336, 337, 436, 926]	
20	Pinos			[20, 26]
21	<i>Pinus sylvestris</i>		[14, 17, 19, 21, 22, 25, 31, 32, 36, 37, 237]	[19, 21, 26]
22	<i>Pinus uncinata</i>	<i>Pinus montana</i> , <i>Pinus mugo</i>	[14, 22]	[19, 21, 22]
23	<i>Pinus pinea</i>		[14, 17, 20, 23, 24, 25, 26, 36, 37, 38, 39, 219, 236, 237, 317, 319, 436]	[19, 23, 24, 26]
24	<i>Pinus halepensis</i>		[17, 20, 21, 23, 24, 25, 26, 36, 37, 38, 39, 236, 237, 336]	[19, 24]
25	<i>Pinus nigra</i>	<i>Pinus laricio</i> , <i>Pinus clusiana</i>	[14, 17, 19, 21, 25, 26, 28, 34]	[19, 21, 25, 26]
26	<i>Pinus pinaster</i>	<i>Pinus maritima</i>	[14, 17, 18, 21, 23, 24, 25, 26, 27, 28, 29, 36, 37, 38, 39, 217, 219, 236, 237, 239, 317, 319, 336, 436, 826, 926]	[19, 24, 26, 28, 526, 626, 726, 826, 926]
27	<i>Pinus canariensis</i>		[27]	
28	<i>Pinus radiata</i>	<i>Pinus insignis</i>	[17, 19, 21, 23, 24, 25, 26, 27, 28, 33, 34, 35, 36, 236, 337, 436]	[19, 28]
29	Otros pinos			

Clave	Nombre especie	Sinonimias	Claves IFN3	Claves IFN4
30	Mezcla de coníferas	Coníferas, excepto pinos		
31	Abies alba	Abies pectinata	[31, 33]	[19, 31]
32	Abies pinsapo			
33	Picea abies	Picea excelsa		[19, 28, 33, 34, 35]
34	Pseudotsuga menziesii	Pseudotsuga douglasii	[19, 28, 31, 33, 34, 35]	[19, 28, 34, 35]
35	Larix spp.		[33, 34, 35, 235, 335]	[19, 35]
36	Cupressus sempervirens			[19, 21, 24, 26, 36]
37	Juniperus communis			[21, 37, 237]
38	Juniperus thurifera		[37, 38, 39, 237, 239]	[26, 38]
39	Juniperus phoenicea		[36, 37, 38, 39, 237]	[26, 39, 237]
40	Quercus		[41, 42, 43, 44, 45, 46, 48, 243]	[40, 41, 44, 999]
41	Quercus robur	Quercus pedunculata	[41, 42, 43, 44, 46, 48, 243]	[41, 999]
42	Quercus petraea	Quercus sessiliflora	[41, 42, 48]	[41, 42, 43, 999]
43	Quercus pyrenaica	Quercus toza	[41, 42, 43, 44, 45, 46, 47, 48, 243, 746, 846, 946]	[43, 99, 999]
44	Quercus faginea	Quercus lusitanica var. faginea	[40, 42, 43, 44, 46, 47, 49, 243, 946]	[43, 44, 45, 999]
45	Quercus ilex ssp. ballota	Quercus rotundifolia	[2, 4, 40, 44, 45, 46, 47, 49, 66, 67, 68, 75, 91, 93, 95, 215, 269, 276, 295, 378, 476]	[41, 45, 99, 245]
46	Quercus suber		[43, 44, 45, 46, 47, 144, 846, 946]	[43, 45, 46, 99, 646, 746, 846, 946]
47	Quercus canariensis	Quercus lusitanica var. baetica	[44, 47, 49]	[47]
48	Quercus rubra	Quercus borealis		[41, 42, 43, 48, 99, 999]
49	Otros quercus			[44, 999]

Clave	Nombre especie	Sinonimias	Claves IFN3	Claves IFN4
50	Mezcla de árboles de ribera	Árboles ripícolas	[2, 3, 4, 6, 7, 9, 51, 52, 53, 54, 55, 56, 57, 58, 59, 61, 62, 63, 69, 74, 79, 92, 94, 97, 253, 255, 256, 257, 258, 297, 299, 307, 355, 357, 392, 457, 657, 757, 857, 957]	
51	Populus alba		[51, 53, 55, 57, 58, 62, 255, 257, 258, 357, 657, 757, 857, 957]	[50, 51, 58, 99, 258, 999]
52	Populus tremula			[50, 51, 52, 58, 99, 258, 999]
53	Tamarix spp.			[50, 51, 53, 999]
54	Alnus glutinosa		[7, 52, 53, 54]	[54, 99, 999]
55	Fraxinus angustifolia		[55, 255]	[50, 55, 99, 255, 999]
56	Ulmus minor	Ulmus campestris		[50, 56, 70, 99, 256, 999]
57	Salix spp.		[3, 4, 7, 9, 51, 52, 53, 54, 55, 57, 58, 79, 92, 97, 207, 257, 258, 297, 357, 457, 557, 657, 757, 857, 957]	[57, 99, 357, 657, 999]
58	Populus nigra			[50, 58, 99, 258, 999]
59	Otros árboles ripícolas			[54, 59]
60	Mezcla de eucaliptos	Eucaliptos	[61, 62, 63, 64]	
61	Eucalyptus globulus		[60, 61, 62, 63, 64, 264, 364]	[61, 99]
62	Eucalyptus camaldulensis	Eucalyptus rostrata	[60, 61, 62, 63, 64, 364]	[61, 62, 64, 99, 264]
63	Otros eucaliptos			[61, 63, 64, 99, 264]
64	Eucalyptus nitens		[62, 64]	[61, 64, 99, 264]

Clave	Nombre especie	Sinonimias	Claves IFN3	Claves IFN4
65	Ilex aquifolium			[65, 99, 999]
66	Olea europaea	Olea oleaster	[66, 67]	[45, 66, 99, 999]
67	Ceratonia siliqua			[45, 67]
68	Arbutus unedo		[2, 3, 4, 5, 8, 9, 12, 13, 15, 16, 56, 65, 66, 68, 72, 76, 93, 95, 97, 99, 215, 256, 275, 276, 295, 297, 299, 315, 369, 376, 395, 399, 578]	[45, 68, 99, 999]
69	Phoenix spp.			
70	Mezcla de frondosas de gran porte	Frondosas de gran porte (H.t. >10 m)	[11, 12, 13, 16, 42, 56, 66, 71, 72, 73, 74, 75, 76, 77, 78, 256, 273, 275, 276, 277, 278, 356, 373, 376, 377, 378, 476, 478, 576, 578, 676, 678]	
71	Fagus sylvatica			[71, 99]
72	Castanea sativa	Castanea vesca	[11, 16, 42, 46, 66, 72, 75, 92, 99, 299, 399, 858]	[72, 99, 999]
73	Betula spp.		[73, 273, 373]	[73, 99, 273, 999]
74	Corylus avellana			[74, 99, 999]
75	Juglans regia			[70, 75, 99, 256, 999]
76	Acer campestre		[76, 276, 376, 476, 576, 676]	[70, 76, 99, 576, 999]
77	Tilia spp.		[77, 277, 377]	[70, 77, 377, 999]
78	Sorbus spp.		[12, 16, 78, 278, 378, 478, 578, 678]	[50, 70, 78, 99, 278, 378, 999]
79	Platanus hispanica	Platanus hybrida		[50, 58, 79, 99, 258, 999]
80	Laurisilva			

Clave	Nombre especie	Sinonimias	Claves IFN3	Claves IFN4
81	Myrica faya			
82	Ilex canariensis		[82, 282]	
83	Erica arborea		[83, 283]	
84	Persea indica			
85	Sideroxylon marmulano			
86	Picconia excelsa	Notelaea excelsa		
87	Ocotea phoetens			
88	Apollonias barbujana	Apollonias canariensis		
89	Otras laurisilvas		[1, 86, 87, 88, 89, 95, 389, 489, 495]	
90	Mezcla de pequeñas frondosas	Frondosas de pequeño porte (H.t. ≤ 10 m)	[1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 15, 16, 65, 66, 67, 68, 69, 74, 78, 91, 92, 93, 94, 95, 96, 97, 99, 215, 269, 278, 295, 297, 299, 307, 369, 378, 395, 399, 478, 495, 499, 578, 595, 599]	
91	Buxus sempervirens			
92	Robinia pseudoacacia	Acacia robinia	[7, 56, 61, 64, 72, 73, 75, 77, 79, 92, 256, 273, 373, 377]	[50, 58, 92, 99, 207, 258, 999]
93	Pistacia terebinthus	Cornicabra		
94	Laurus nobilis	Laurel		[90, 94, 99]
95	Prunus spp.	Prunus	[4, 5, 9, 95, 295]	[45, 90, 95, 99, 395, 999]
96	Rhus coriaria	Zumaque		[50, 96]
97	Sambucus nigra	Saúco negro		[50, 90, 97, 99, 657, 999]

Clave	Nombre especie	Sinonimias	Claves IFN3	Claves IFN4
98	Carpinus betulus	Carpe	[43, 44, 46, 47, 51, 53, 54, 55, 57, 58, 60, 62, 63, 64, 72, 75, 144, 257, 357, 657]	[71, 273]
99	Otras frondosas	Otras frondosas	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 64, 65, 66, 67, 68, 69, 71, 72, 73, 74, 75, 76, 77, 78, 79, 82, 83, 84, 86, 87, 88, 89, 91, 92, 93, 94, 95, 96, 97, 99, 207, 215, 217, 243, 253, 255, 256, 257, 258, 264, 269, 273, 275, 276, 277, 278, 279, 291, 292, 293, 295, 297, 299, 307, 315, 355, 356, 357, 364, 369, 373, 375, 376, 377, 378, 389, 392, 395, 399, 415, 457, 469, 476, 478, 489, 492, 495, 499, 557, 569, 576, 578, 592, 595, 657, 676, 678, 757, 778, 857, 858, 957]	[90, 99, 999]
207	Acacia melanoxylon	Acacia melanoxylon		[7, 99, 207, 307]
215	Crataegus monogyna	Majuelo	[15, 215, 315, 415]	[15, 45, 99, 215]
217	Cedrus deodara	Cedrus deodara		[17, 26, 217]
219	Tetraclinis articulata	Tetraclinis articulata		[26, 219]
235	Larix decidua	Alerce común		[19, 34, 35, 235]

Clave	Nombre especie	Sinonimias	Claves IFN3	Claves IFN4
236	Cupressus arizonica	Ciprés arizónica		[19, 24, 26, 36, 236]
237	Juniperus oxycedrus	Enebro oxicedro	[36, 37, 38, 39, 236, 237, 239]	[21, 26, 39, 237]
238	Juniperus turbinata	Sabina canaria	[238, 337]	[38]
239	Juniperus sabina	Sabina rastrera		
243	Quercus pubescens	Quercus pubescens, Quercus humilis	[42, 43, 44, 47, 243]	[43, 243, 999]
244	Quercus lusitanica	Quercus fruticosa, Quejigueta		
245				[45, 245]
253	Tamarix canariensis	Tarajal		[53]
255	Fraxinus excelsior	Fresno excelsior	[55, 255, 355]	[50, 55, 99, 255, 999]
256	Ulmus glabra	Ulmus montana		[70, 99, 256, 999]
257	Salix alba	Sauce blanco		[57, 99, 257, 357, 657, 999]
258	Populus x canadensis	Populus x euroamericana	[58, 258]	[50, 58, 99, 258, 999]
264	Eucalyptus viminalis	Eucalipto viminalis		[64, 99, 264]
268	Arbutus canariensis	Madroño canario		[68]
273	Betula alba	Betula verrucosa, Abedul pubescens		[73, 99, 273, 999]
275	Juglans nigra	Nogal		[256, 275]
276	Acer monspessulanum	Arce de Montpelier		[45, 70, 276, 999]
277	Tilia cordata	Tilo cordata		[277, 377, 999]
278	Sorbus aria	Mostajo		[70, 99, 278, 378, 999]
279	Platanus orientalis	Plátano oriental		[58, 999]
281				
282	Ilex platyphylla	Naranjero		[82]

Clave	Nombre especie	Sinonimias	Claves IFN3	Claves IFN4
283	<i>Erica scoparia</i>	Tejo, brezo arbóreo escopario		[83]
289	<i>Pleiomeris canariensis</i>	Delfino		[89]
291	<i>Buxus balearica</i>	Boj de Baleares		
292	<i>Sophora japonica</i>	Acacia sofora		
293	<i>Pistacia atlantica</i>	Cornicabra canaria		[93]
294	<i>Laurus azorica</i>	Laurel canario	[68, 268, 294]	[94]
295	<i>Prunus spinosa</i>	Espino negro		
297	<i>Sambucus racemosa</i>	Saúco racemosa		
299	<i>Ficus carica</i>	Higuera		[90, 99, 299, 999]
307	<i>Acacia dealbata</i>	Acacia dealbata		[7, 90, 99, 207, 307]
315	<i>Crataegus laevigata</i>	Espino majuelo		[215, 315]
317	<i>Cedrus libani</i>	<i>Cedrus libani</i>		
319	<i>Thuja</i> spp.	Thuja		[19, 28, 319]
335	<i>Larix leptolepis</i>	<i>Larix kaempferi</i> , Alerce leptolepis		
336	<i>Cupressus lusitanica</i>	Ciprés lambertiana		[26, 36]
337	<i>Juniperus cedrus</i>	Enebro canario		[37]
344				
355	<i>Fraxinus ornus</i>	Fresno orno		
356	<i>Ulmus pumila</i>	Olmo pumilo		[70, 356]
357	<i>Salix atrocinerea</i>	Bardaguera		[57, 99, 357, 657, 999]
364	<i>Eucalyptus gomphocephalus</i>	Eucalipto gonfo	[62, 63, 64, 264, 364]	[64, 264, 364]
369	<i>Chamaerops humilis</i>	Palmito		
373	<i>Betula pendula</i>	<i>Betula hispanica</i> , Abedul péndula		[73, 273, 373, 999]
376	<i>Acer negundo</i>	Negundo fraxinifolia, Arce negundo		[76, 99, 276, 376, 999]

Clave	Nombre especie	Sinonimias	Claves IFN3	Claves IFN4
377	Tilia platyphyllos	Tilo común		[70, 377, 999]
378	Sorbus aucuparia	Serbal de cazadores		[45, 50, 70, 78, 99, 278, 378, 999]
389	Rhamnus glandulosa	Sanguino		[89]
392	Gleditsia triacanthos	Acacia gleditsia		
395	Prunus avium	Cerezo silvestre		[90, 95, 99, 395, 999]
399	Morus spp.	Morera		
415	Crataegus laciniata	Majoleto		
435	Larix x eurolepis	Alerce híbrido		
436	Cupressus macrocarpa	Ciprés americano		[26, 36, 436]
455				
456				[70, 456]
457	Salix babylonica	Sauce llorón		[57, 357, 457, 657]
464				[61, 64, 99, 464]
469	Phoenix canariensis	Palmera	[469, 569]	[69]
476	Acer opalus	Arce ópalus		[45, 70, 276, 476, 576, 999]
478	Sorbus domestica	Serbal común		[278, 478, 999]
489	Visnea mocanera	Mocan		[89]
495	Prunus lusitanica	Loro, hija		[95, 495]
499	Morus alba	Morera		
515	Crataegus azarolus	Espino		
557	Salix cantabrica	Sauce cantábrico		[357, 557]
569	Dracaena draco	Drago		
576	Acer pseudoplatanus	Arce seudoplátano		[70, 76, 99, 276, 576, 999]
578	Sorbus torminalis	Serbal torminal		[70, 278, 378, 578, 999]
595	Prunus padus	Prunus		
599	Morus nigra	Morera		
657	Salix caprea	Sauce cabruno		[57, 99, 357, 657, 999]

Clave	Nombre especie	Sinonimias	Claves IFN3	Claves IFN4
676	Acer platanoides	Arce platanoides		[70, 99, 276, 576, 676, 999]
678	Sorbus latifolia	Serbal de hoja ancha		[278, 678, 999]
757	Salix elaeagnos	Sarga		[99, 357, 657, 757, 999]
776				
778	Sorbus chamaemespilus	Serbal chame		
857	Salix fragilis	Mimbre		[357, 657, 857, 999]
858	Salix canariensis	Sauce canario		[58]
917	Cedrus spp.	Cedrus spp.		
936	Cupressus spp.	Cipres		
937	Juniperus spp.	Enebro y sabinas	[36, 37, 38, 39, 236, 237, 239, 336]	
955	Fraxinus spp.	Fresnos	[55, 255]	
956	Ulmus spp.	Olmo		
957	Salix purpurea	Mimbrera		[99, 357, 657, 957, 999]
975	Juglans spp.			
976	Acer spp.	Arces	[71, 72, 76, 276, 278, 307, 378, 476, 478, 492, 576, 578, 676]	
997	Sambucus spp.			

REFERENCIAS

- [1] MITECO. Inventario forestal nacional (ifn2, ifn3, ifn4): metodología y bases de datos. Ministerio para la Transición Ecológica y el Reto Demográfico (España), 2023. <https://www.miteco.gob.es/es/biodiversidad/servicios/banco-datos-naturaleza/informacion-disponible/ifn.aspx> (acceso 2025).
- [2] Joaquín Muñoz-Sabater, Emanuel Dutra, Anna Agustí-Panareda, Clément Albergel, Giorgio Arduini, Gianpaolo Balsamo, et al. Era5-land: A state-of-the-art global reanalysis at land surfaces. *EGU General Assembly / ECMWF (Copernicus Climate Data Store)*, 2021. <https://doi.org/10.24381/cds.e2161bac>.
- [3] USGS. Landsat collection 2 level-2 science products: Surface reflectance. U.S. Geological Survey, 2021. <https://www.usgs.gov/landsat-missions/landsat-collection-2-level-2-science-products> (acceso 2025).
- [4] Maider Araceli Urbón Jiménez, Jaime Gabriel Vegas, Ana de Luis Reboredo, Belén Pérez Lancho, and Ana-Belén Gil-González. Greenwest-db: Base de datos integrada de atributos forestales, climáticos y espectrales para España. Manuscrito en preparación. Universidad de Salamanca, Grupo BISITE, 2025.
- [5] Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. Catboost: unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2018. <https://proceedings.neurips.cc/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html>.
- [6] United Nations. United nations framework convention on climate change (unfccc). United Nations, 1992. <https://unfccc.int/resource/docs/convkp/conveng.pdf>.
- [7] United Nations Framework Convention on Climate Change. The kyoto protocol. <https://unfccc.int/resource/docs/convkp/kpeng.pdf>, 1997. Accessed: 2023-07-09.
- [8] Intergovernmental Panel on Climate Change. *Climate Change 2007: Mitigation of Climate Change*. Cambridge University Press, Cambridge, UK, 2007.
- [9] United Nations Framework Convention on Climate Change. Paris agreement. <https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement>, 2015. Accessed: 2023-07-09.
- [10] Scott J. Goetz and Ralph O. Dubayah. Remote sensing of forest structure, carbon, and biodiversity. *Remote Sensing of Environment*, 113:S1–S3, 2009.
- [11] Huan Zhang, Wei Li, Peng Wang, and Yi Chen. Carbon stock estimation in forest ecosystems using deep learning and sentinel-2 imagery. *Ecological Indicators*, 143:109372, 2022.
- [12] Tao Yang, Xin Li, Lei Zhao, and Yan Wang. Forestcarbonai: A deep learning model for high-resolution forest carbon mapping from remote sensing data. *Remote Sensing*, 15(3):765, 2023.
- [13] G. Reiersen, D. Dao, B. Lütjens, K. Klemmer, K. Amara, A. Steinegger, and X. Zhu. Reforestree: A dataset for estimating tropical forest carbon stock with deep learning and aerial imagery. *arXiv preprint*, 2022.

- [14] W. Dong, E. T. A. Mitchard, H. Yu, S. Hancock, and C. M. Ryan. Forest aboveground biomass estimation using gedi and earth observation data through attention-based deep learning. *arXiv preprint*, 2023.
- [15] Ministerio para la Transición Ecológica y el Reto Demográfico. Instrucciones de uso de la calculadora de absorciones de co2 ex ante de las especies forestales arbóreas españolas, 2023. Accedido: 2025-07-16.
- [16] USGS. Usgs Landsat 5 level 2, collection 2, tier 1, 2025. Accedido: 2025-07-08.
- [17] J. Muñoz Sabater. Era5-land hourly data from 1950 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2019. DOI: 10.24381/cds.e2161bac (Accedido: 2025-07-07).
- [18] Ministerio para la Transición Ecológica (MITECO). Guía para la estimación de absorciones de dióxido de carbono. Technical report, Ministerio para la Transición Ecológica y el Reto Demográfico, 2021. Accedido: 2025-07-10.
- [19] Intergovernmental Panel on Climate Change. *2006 IPCC Guidelines for National Greenhouse Gas Inventories*. IPCC, Geneva, Switzerland, 2006.
- [20] T. Buchholz. Artificial intelligence and machine learning in environmental science. *Environmental Modelling Software*, 56:60–69, 2014.