

Pugni Stanek, T.; González González, I.; Tomé, J.; Esteban Cava, J.; Algeet Abaquero, N.; Recuero Pavon, L.; Merino De Miguel, S. Clasificación automatizada basada en series temporales y variables estructurales a partir de sensores remotos (S2, S1 y LiDAR).

Clasificación automatizada basada en series temporales y variables estructurales a partir de sensores remotos (S2, S1 y LiDAR).

Pugni Stanek, Tomas Luis ^{1,2} González González, Isabel ¹ Tomé, Jose Luis ¹ Esteban Cava, Jessica ¹ Algeet Abaquero, Nur ¹ Recuero Pavon, Laura ² Merino De Miguel, Silvia ²

¹ Agresta Sociedad Cooperativa, C/ Duque de Fernán Núñez, 2, 1º, 28012, Madrid, España.

² Grupo de investigación Geo-QuBiDy, Universidad Politécnica de Madrid, Avda. Ramiro de Maeztu. s/n 28040 Madrid. España

ORCID: Pugni Stanek 0009-0000-2527-1657 González González 0009-0002-1455-9174 Tomé 0000-0003-2298-9115 Esteban Cava 0000-0001-8350-5343 Algeet Abaquero 0000-0003-0819-6549 Recuero Pavon 0000-0001-5474-4900 Merino De Miguel 0000-0002-4764-5311

Correspondencia: tomas.pugni@upm.es igonza@agresta.org jltome@agresta.org jesteban@agresta.org nalgeet@agresta.org laura.recuero@upm.es silvia.merino@upm.es

RESUMEN

La clasificación precisa de las masas forestales es esencial para la gestión sostenible y la modelización de riesgos naturales. Este estudio evalúa la capacidad de discriminación de tres especies forestales (*Pinus sylvestris*, *Quercus pyrenaica* y *Fagus sylvatica*) en el Monte de Utilidad Pública de Villoslada de Cameros (La Rioja), mediante un enfoque de teledetección multitemporal y multisensor. Se integraron datos ópticos (Sentinel-2), de radar (Sentinel-1) y estructurales (LiDAR-PNOA) para caracterizar la fenología y estructura de la vegetación. La metodología partió de parcelas monoespecíficas del Inventario Forestal Nacional (IFN4) y datos locales, aplicando un sobremuestreo espacial mediante puntos aleatorios para corregir el desbalanceo de clases entre coníferas y frondosas. Tras generar un vector de 416 características, se utilizó el algoritmo de Eliminación Recursiva de Características con Validación Cruzada (RFECV) para seleccionar las 20 variables más robustas, optimizando el modelo. Los resultados mostraron una precisión global del 96,1%, identificando como predictores críticos la banda del Red-Edge (B6), el índice de Verdor del Tasseled Cap (TCG) y la polarización VH. Aunque la separación entre coníferas y frondosas fue excelente, se observó una ligera confusión entre las especies de frondosas debido a su similitud fenológica. Se concluye que las series temporales y las variables estructurales podrían ser determinante para mejorar la precisión en la automatización de la identificación de especies en inventarios forestales a gran escala.





Palabras clave: Fusión de Datos, Inventario nacional, Fenología, Periodograma, RFECV, Retrodispersión

Fecha de recepción: 18 febrero 2026 · **Fecha de aceptación:** 18 febrero 2026




Clasificación automatizada basada en series temporales y variables estructurales a partir de sensores remotos (S2, S1 y LiDAR).

Pugni Stanek, Tomas Luis ^(1,2), González González, Isabel ⁽¹⁾, Tomé, Jose Luis ⁽¹⁾, Esteban Cava, Jessica ⁽¹⁾, Algeet Abaquero, Nur ⁽¹⁾, Recuero Pavon, Laura ⁽²⁾, Merino De Miguel, Silvia ⁽²⁾

⁽¹⁾ Agresta Sociedad Cooperativa, 1º, España.

 0009-0002-1455-9174, igonzalez@agresta.org ;  0000-0003-2298-9115, jltome@agresta.org
 0000-0001-8350-5343, jesteban@agresta.org ;  0000-0003-0819-6549, nalgeet@agresta.org

⁽²⁾ Grupo de investigación Geo-QuBiDy, Universidad Politécnica de Madrid.

 0009-0000-2527-1657, tomas.pugni@upm.es ;  0000-0001-5474-4900, laura.recuero@upm.es
 0000-0002-4764-5311, silvia.merino@upm.es

Resumen: La clasificación precisa de las masas forestales es esencial para la gestión sostenible y la modelización de riesgos naturales. Este estudio evalúa la capacidad de discriminación de tres especies forestales (*Pinus sylvestris*, *Quercus pyrenaica* y *Fagus sylvatica*) en el Monte de Utilidad Pública de Villoslada de Cameros (La Rioja), mediante un enfoque de teledetección multitemporal y multisensor. Se integraron datos ópticos (Sentinel-2), de radar (Sentinel-1) y estructurales (LiDAR-PNOA) para caracterizar la fenología y estructura de la vegetación. La metodología partió de parcelas mono-específicas del Inventario Forestal Nacional (IFN4) y datos locales, aplicando un sobremuestreo espacial mediante puntos aleatorios para corregir el desbalanceo de clases entre coníferas y frondosas. Tras generar un vector de 416 características, se utilizó el algoritmo de Eliminación Recursiva de Características con Validación Cruzada (RFECV) para seleccionar las 20 variables más robustas, optimizando el modelo. Los resultados mostraron una precisión global del 96,1%, identificando como predictores críticos la banda del *Red-Edge* (B6), el índice de Verdor del Tasseled Cap (TCG) y la polarización VH. Aunque la separación entre coníferas y frondosas fue excelente, se observó una ligera confusión entre las especies de frondosas debido a su similitud fenológica. Se concluye que las series temporales y las variables estructurales podrían ser determinante para mejorar la precisión en la automatización de la identificación de especies en inventarios forestales a gran escala.

Palabras clave: Fusión de Datos, Inventario nacional, Fenología, Periodograma, RFECV, Retrodispersión.

Advances in National Forest Inventory: Automated Classification Based on Time Series and Structural Variables Using Remote Sensing (S2, S1, and LiDAR).

Abstract: Accurate classification of forest stands is essential for sustainable management and natural hazard modeling. This study evaluates the discrimination capability of three forest species (*Pinus sylvestris*, *Quercus pyrenaica*, and *Fagus sylvatica*) in the "Monte de Utilidad Pública" of Villoslada de Cameros (La Rioja), using a multi-temporal and multi-sensor remote sensing approach. Optical (Sentinel-2), radar (Sentinel-1), and structural (LiDAR-PNOA) data were integrated to characterize vegetation phenology and structure. The methodology was based on monospecific plots from the National Forest Inventory (IFN4) and local data, applying spatial oversampling with random points to correct class imbalance between conifers and broadleaves. After generating a vector of 416 features, the Recursive Feature Elimination with Cross-Validation (RFECV) algorithm was used to select the 20 most robust variables, optimizing model parsimony. The results showed an overall accuracy of 96.1%, identifying the Red-Edge band (B6), the Tasseled Cap Greenness index (TCG), and VH polarization as critical predictors. Although the separation between conifers and broadleaves was excellent, slight confusion was observed among broadleaf species due to their phenological similarity. It is concluded that time series and structural variables are decisive for automating large-scale forest inventory.

Keywords: Data Fusion, Phenology, Forest Inventory, RFECV, Backscatter.

1. INTRODUCCIÓN

La clasificación de masas forestales mediante teledetección ha evolucionado más allá del inventario tradicional, convirtiéndose en un recurso clave para la gestión de riesgos naturales (Shaban, 2022). Disponer de información detallada sobre la composición específica y la estructura de la vegetación es crítico para la modelización de incendios forestales, permitiendo estimar la propagación y la intensidad de llama potencial. Asimismo, estos datos son fundamentales para la planificación de tratamientos de silvicultura preventiva y para obtener la caracterización precisa del territorio.

Actualmente, la disponibilidad de datos provenientes de sensores multiespectrales como Sentinel-2 y Landsat (Shaban, 2022), hiperespectrales como PRISMA y de radar de apertura sintética (SAR) como Sentinel-1, facilita la monitorización de la estructura y fenología forestal. Sin embargo, la integración de estas fuentes masivas requiere metodologías precisas para evitar la redundancia de información y optimizar los modelos predictivos mediante algoritmos de Machine Learning (Maxwell *et al.*, 2018).

Estudios previos han demostrado la eficacia de combinar series temporales con variables estructurales (Immitzer *et al.* 2016). Complementariamente, los sensores SAR permiten caracterizar la estructura forestal y el contenido de humedad mediante el análisis de sus distintas polarizaciones, siendo capaces de detectar cambios independientemente de la cobertura nubosa (Woodhouse *et al.*, 2012). Finalmente, la tecnología LiDAR se consolida como la herramienta más precisa para la modelización tridimensional de la estructura del bosque (Wulder *et al.*, 2012).

En este contexto, Immitzer *et al.* (2016) evaluaron la clasificación de especies basada en firmas espectrales, obteniendo una precisión global del 63,5% mediante un enfoque basado en píxeles. No obstante, la inclusión de series temporales resulta determinante, ya que permite capturar las dinámicas fenológicas particulares de cada especie, aportando información que va más allá de la respuesta espectral en una única fecha (Persson *et al.*, 2018).

El objetivo de este estudio es clasificar las especies forestales arboladas presentes en el Monte de Utilidad Pública de Villoslada de Cameros (La Rioja), con el fin evaluar el rendimiento de la integración de datos de diferentes sensores, concretamente Sentinel 1, Sentinel 2 y LiDAR, en la discriminación entre distintas especies forestales.

2. MATERIAL Y MÉTODOS

2.1. Zona de Estudio

La zona de estudio (Figura 1) corresponde al Monte de Utilidad Pública n.º 141, ubicado en el término municipal de Villoslada de Cameros (La Rioja) con una superficie total de 14.646,88 ha. Las principales formaciones forestales presentes en el área son el pinar (*Pinus*

sylvestris), el rebollar (*Quercus pyrenaica*) y el hayedo (*Fagus sylvatica*). **Figura 1**

2.2. Materiales

2.2.1. Datos de campo (Verdad terreno)

Para garantizar una muestra representativa, el estudio combina datos del Cuarto Inventario Forestal Nacional (IFN4) con parcelas propias de AGRESTA, que complementan la información oficial.

2.2.2. Sensores remotos

Se han integrado tres fuentes de datos complementarias. **Sentinel-2 (Óptico):** Datos multiespectrales para obtener reflectancia de la cubierta (2018-2025), **Sentinel-1 (Radar):** Datos de retrodispersión (SAR) en banda C (2018-2025), y **LiDAR (PNOA):** Datos de la tercera cobertura del Plan Nacional de Ortofotografía Aérea (2024).

2.3. Métodos

El estudio se fundamenta en el análisis multitemporal y multisensor de parcelas dominadas por una única especie. Para entrenar el modelo *Random Forest*, se aplicó un algoritmo de Eliminación Recursiva de Características con Validación Cruzada (RFECV), seleccionando las variables explicativas óptimas a partir del vector de características inicial. Debido al desbalanceo de las muestras detectado (dominancia de *Pinus sylvestris* sobre *Quercus pyrenaica* y *Fagus sylvatica*), se realizó un sobremuestreo (oversampling) mediante la generación de puntos aleatorios en rodales puros de las citadas especies obtenidos a partir de la cartografía de la ordenación del monte en 2024 obteniendo un total de 486 puntos. Este procedimiento garantiza la equitatividad de la muestra y la robustez estadística del análisis posterior (**Figura 1**).

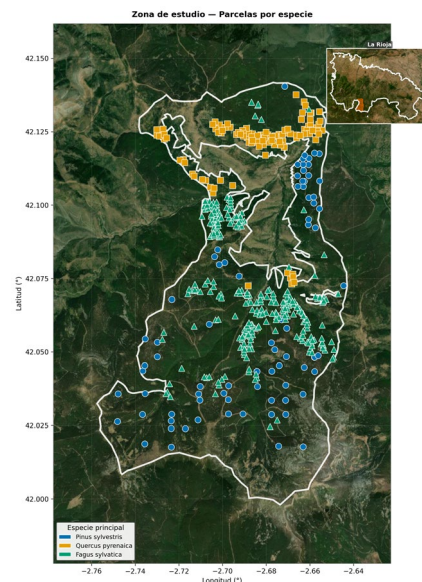


Figura 1 Mapa de la distribución de parcelas por especies.

A partir de la información de Sentinel 2, se analizaron las dinámicas de las bandas espectrales y se calcularon los índices Tasseled Cap; y a partir de los datos Sentinel 1, se utilizaron las polarizaciones ascendentes y descendentes, el Radar Vegetation Index y el Normalized

Difference Polarization Index. Se obtuvieron todas las imágenes derivadas de Sentinel-1 y 2 aplicando a este último la máscara de nubes (SCL), se regularizaron las series a una frecuencia semanal (52 imágenes por año), se imputaron los valores faltantes mediante una interpolación lineal y se descompusieron en tendencia, estacionalidad y residuo, (Figura 3).

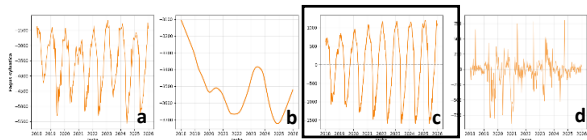


Figura 1 Serie de tiempo (a) descompuesta en tendencia (b), estacionalidad (c) y residuo (d), remarcando la estacionalidad.

La dinámica de las series temporales se analizó únicamente a través de la componente de estacionalidad utilizando dos métodos complementarios que exploran diferentes dimensiones de la serie (Box *et al.*, 2016)(Figura 4). Desde la perspectiva temporal, la función de autocorrelación se utilizó para evaluar la memoria de la serie a los retardos de 6, 12, 18 y 24 meses, caracterizando así la persistencia de la serie (Sáenz *et al.*, 2024). Además, se utilizó el test Q de Ljung-Box para evaluar el contenido de información de la componente estacional. Desde la perspectiva frecuencial, el periodograma, permitió la identificación del periodo dominante caracterizando así la estacionalidad en función del número, la amplitud y estabilidad de los ciclos estacionales. La significancia de la estacionalidad se evaluó mediante la prueba Kappa de Fisher (Recuero *et al.*, 2019).

Además, a partir de la componente de estacionalidad, se calcularon las siguientes métricas: media y desviación estándar, el área bajo la curva, los valores máximos y mínimos, y estadísticas anuales.

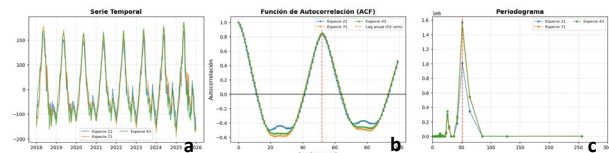


Figura 2. Ejemplo de la Componente de estacionalidad (a), función de autocorrelación (b) y periodograma (c) para la banda 5.

Finalmente, se normalizaron los datos LiDAR respecto a la trayectoria de vuelo para derivar métricas estructurales y radiométricas, incluyendo estadísticos de altura (extremos, media, percentiles), cobertura (> 2m, FHD) y momentos de la distribución de intensidad (media, desviación, asimetría y curtosis) y del primer retorno (Wulder *et al.*, 2012).

La integración de los sensores generó un vector inicial de 416 características por parcela. Dada la alta dimensionalidad de este conjunto de datos, que dificulta la reproducibilidad y la interpretación ecológica de las dinámicas estudiadas, el primer paso del análisis consistió en una selección rigurosa de variables. El

objetivo fue identificar el mínimo de predictores necesario para maximizar la precisión del modelo.

Se empleó el algoritmo RFECV, para seleccionar un subconjunto óptimo de 20 variables. Este método, basado en validación cruzada, maximiza la precisión predictiva evitando el sobreajuste y garantiza la escalabilidad computacional del modelo. La **Tabla 1** muestra las variables seleccionadas.

Tabla 1. Variables seleccionadas por categoría.

Categoría	Variables
Bandas Ópticas	$B6_{AMP}, B6_{VER}, B6_{IN}, B6_{PR}, B6_{MIN}, B6_{MAX}, B5_{L104}, B5_{Q1}, B5_{Q104}$
Índices Tasseled Cap	$TCG_{PR}, TCG_{OT}, TCG_{AMP}, TCG_{MAX}, TCG_{MIN}, TCW_{IN}, TCW_{VER}, TCB_{DESV-}$
Radar Sentinel-1	$VH_{PR}^{ASC}, VH_{VER}^{ASC}, VH_{IN}^{ASC}$

3. RESULTADOS Y DISCUSIÓN

La selección de variables subraya la importancia de los estadísticos estacionales, especialmente los calculados para la banda B6 (740 nm). Su ubicación en la zona de transición del *Red Edge* permite capturar variaciones en el contenido de clorofila, resultando estratégico para discriminar dinámicas fenológicas. Por su parte, la información estructural y de humedad del SWIR se integra de forma optimizada mediante los componentes del *Tasseled Cap*, complementando la caracterización fisiológica de las especies.

El índice de Verdor (TCG) del *Tasseled Cap* resultó crucial para la diferenciación de especies, especialmente durante las fases de primavera y otoño. Además, métricas derivadas como la amplitud de la estación permitieron cuantificar la *fuerza del ciclo* vegetativo (Recuero *et al.*, 2019), aportando información clave sobre la intensidad de la respuesta fenológica de cada especie.

En cuanto a la caracterización estructural, la polarización cruzada VH de Sentinel-1 (órbita ascendente) resultó determinante. Esta variable captura la dispersión volumétrica del dosel, permitiendo diferenciarla en momentos clave (primavera, verano e invierno). Por último, el análisis de autocorrelación aportó información sobre la estabilidad fenológica: la media de los valores de autocorrelación en el retardo 104 (Lag_{104}) y el estadístico Q a este mismo retardo revelaron la consistencia del ciclo a dos años vista. Los resultados del modelo, desglosados por especie y a nivel global, se detallan en la Tabla 2.

Tabla 2 Resultados del modelo tras la prueba.

Clase / Métrica	Precisión	Recall	F1-Score
Pinus sylvestris	0.94	1.00	0.97
Quercus pyrenaica	0.94	1.00	0.97
Fagus sylvatica	1.00	0.95	0.98
Macro promedio	0.96	0.98	0.97
Promedio ponderado	0.97	0.97	0.97
Accuracy global	—	—	0.97

El modelo demostró un alto rendimiento general, reflejado en métricas elevadas de Precisión, *Recall* y *F1-Score*. La discriminación a nivel de grupo funcional (coníferas vs. frondosas) fue excelente, aportando el mayor peso a la precisión global y evidenciando una baja tasa de error. Sin embargo, al descender al nivel de especie dentro de las frondosas (*Fagus sylvatica* vs. *Quercus pyrenaica*), se observó una leve disminución en la precisión. A pesar de ello, los resultados confirman que las series temporales y las bandas del Red Edge son determinantes para la caracterización forestal en el entorno de Villoslada de Cameros.

De acuerdo con los resultados obtenidos tras la clasificación multitemporal, se observa una gran capacidad de distinguir las especies dominantes en la zona de estudio. Esto se alinea con resultados obtenidos en otros estudios (Esteban Cava *et al.*, 2025; Persson *et al.*, 2018). Estos estudios muestran la capacidad de Sentinel 2 de discriminar las especies forestales mediante Random Forest. Si bien el enfoque realizado en este estudio nos muestra que el enfoque temporal y las dinámicas de las especies nos permiten alcanzar unos resultados muy buenos para la zona de estudio, es necesario ampliar la zona para obtener mayores puntos de muestreo y conocer la precisión a una escala mayor de esta metodología. También será interesante conocer como es el rendimiento en zonas de masas mixtas.

4. CONCLUSIONES

Este estudio demuestra que la integración de series temporales es determinante para la discriminación de especies forestales en el entorno de Villoslada de Cameros. Los resultados validan que las bandas del Red-Edge y los índices derivados de Tasseled Cap poseen la sensibilidad suficiente para separar especies con dinámicas similares. Futuras investigaciones se orientarán a ampliar la zona de estudio para evaluar la capacidad de generalización del modelo ante una mayor diversidad de especies.

5. AGRADECIMIENTOS

Los autores desean agradecer a Copernicus por aportar datos multiespectrales y radar, al IGN por la cesión de los datos LiDAR y a la Comunidad de Madrid por la cofinanciación del doctorado industrial de Tomás Pugni Stanek (IND2024/AMB-34193).

6. REFERENCIAS

- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time series analysis: forecasting and control*. John Wiley & Sons, Inc.
- Esteban Cava, J., González, I., & Tomé, J. L. (2025). *Discriminación de especies de pino mediante imágenes satelitales de Sentinel-2 y PRISMA en el distrito forestal Terra de Lemos*. <http://prisma-i.it/index.php/en>.
- Immitzer, M., Vuolo, F., & Atzberger, C. (2016). First experience with Sentinel-2 data for crop and tree species classifications in central Europe. *Remote Sensing*, 8(3). <https://doi.org/10.3390/rs8030166>
- Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: an applied review. *International Journal of Remote Sensing*, 39(9), 2784–2817. <https://doi.org/10.1080/01431161.2018.1433343>
- Persson, M., Lindberg, E., & Reese, H. (2018). Tree Species Classification with Multi-Temporal Sentinel-2 Data. *Remote Sensing*, 10(11), 1794. <https://doi.org/10.3390/rs10111794>
- Recuero, L., Litago, J., Pinzón, J. E., Huesca, M., Moyano, M. C., & Palacios-Orueta, A. (2019). Mapping Periodic Patterns of Global Vegetation Based on Spectral Analysis of NDVI Time Series. *Remote Sensing*, 11(21), 2497. <https://doi.org/10.3390/rs11212497>
- Sáenz, C., Cicuéndez, V., García, G., Madruga, D., Recuero, L., Bermejo-Saiz, A., Litago, J., de la Calle, I., & Palacios-Orueta, A. (2024). New Insights on the Information Content of the Normalized Difference Vegetation Index Sentinel-2 Time Series for Assessing Vegetation Dynamics. *Remote Sensing*, 16(16). <https://doi.org/10.3390/rs16162980>
- Shaban, A. (2022). *Fundamentals of Satellite Remote Sensing* (pp. 1–14). https://doi.org/10.1007/978-3-031-15549-9_1
- Woodhouse, I. H., Mitchard, E. T. A., Brolly, M., Maniatis, D., & Ryan, C. M. (2012). Radar backscatter is not a “direct measure” of forest biomass. *Nature Climate Change*, 2(8), 556–557. <https://doi.org/10.1038/nclimate1601>
- Wulder, M. A., White, J. C., Nelson, R. F., Næsset, E., Ørka, H. O., Coops, N. C., Hilker, T., Bater, C. W., & Gobakken, T. (2012). Lidar sampling for large-area forest characterization: A review. *Remote Sensing of Environment*, 121(47), 196–209. <https://doi.org/10.1016/j.rse.2012.02.001>