

Munuera Irurzun, S.; Gómez, C.; Oli, A.; Domingo, D.; Rodríguez Puerta, F. Avances en la clasificación de tipos de bosque en el Sistema Ibérico español: comparación entre series temporales Landsat y AlphaEarth Foundations de Google

Avances en la clasificación de tipos de bosque en el Sistema Ibérico español: comparación entre series temporales Landsat y AlphaEarth Foundations de Google

Munuera Irurzun, Saioa ¹ Gómez, Cristina ^{1,2} Oli, Aman ¹ Domingo, Darío ^{1,3}
Rodríguez Puerta, Francisco ¹

¹ iuFOR, EifAB, Universidad de Valladolid, Soria, España

² Department of Geography and Environment, School of Geosciences, University of Aberdeen, UK.

³ GEOFOREST-IUCA, Department of Geography, University of Zaragoza, España

ORCID: Munuera Irurzun 0009-0008-0512-1465 Gómez 0000-0002-2756-0863 Oli 0000-0003-0748-2607 Domingo 0000-0002-8362-7559 Rodríguez Puerta 0000-0002-4844-1759

Correspondencia: saioa.munuera@uva.es cgomez@uva.es amanwoli1@gmail.com dario.domingo@uva.es
francisco.rodriquez.puerta@uva.es

RESUMEN

El estudio compara la clasificación de nueve tipos de bosque en el Sistema Ibérico Norte mediante dos fuentes de datos de teledetección: compuestos estacionales Landsat 8 y 9 y embeddings de AlphaEarth Foundations (AEF) de Google. Los tipos de bosque se distinguen por la morfología foliar (aciculiforme o planifolia) y el hábito foliar (perennifolio, marcescente o caducifolio), incluyendo formaciones puras y mixtas. Se generó una máscara forestal a partir del Mapa Forestal de España y se seleccionaron muestras de entrenamiento. En el caso de Landsat, se calcularon índices espectrales y compuestos estacionales como variables predictoras. Los embeddings de AEF integran información multifuente (óptica, radar, LiDAR y climática) en vectores latentes que sintetizan la dinámica anual y la estructura de la vegetación. Ambas bases de datos se clasificaron con Random Forest y los resultados se compararon píxel a píxel mediante una matriz de confusión. Los resultados muestran mayor estabilidad en las clases puras, especialmente coníferas y caducifolias, y mayor confusión en las clases mixtas. Landsat captó mejor los patrones de hábito foliar, mientras que AEF proporcionó una delimitación espacial más precisa y mejor discriminación estructural.

Palabras clave: fenología, morfología foliar, embedding, Landsat


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

Avances en la clasificación de tipos de bosque en el Sistema Ibérico español: comparación entre series temporales Landsat y AlphaEarth Foundations de Google


Munuera Irurzun, Saioa ⁽¹⁾, Gómez, Cristina ^(1,2), Oli, Aman ⁽¹⁾, Domingo, Darío ^(1,3), Rodríguez Puerta, Francisco ⁽¹⁾

⁽¹⁾ iuFOR, EifAB, Universidad de Valladolid, España


 0009-0008-0512-1465, saioa.munuera@uva.es ;  0000-0003-0748-2607, amanwoli1@gmail.com

 0000-0002-4844-1759, francisco.rodriguez.puerta@uva.es

⁽²⁾ Department of Geography and Environment, School of Geosciences, University of Aberdeen.

 0000-0002-2756-0863, cgomez@uva.es

⁽³⁾ GEOFOREST-IUCA, Department of Geography, University of Zaragoza, España

 0000-0002-8362-7559, dario.domingo@uva.es

Resumen: El estudio compara la clasificación de nueve tipos de bosque en el Sistema Ibérico Norte mediante dos fuentes de datos de teledetección: compuestos estacionales Landsat 8 y 9 y *embeddings* de *AlphaEarth Foundations* (AEF) de Google. Los tipos de bosque se distinguen por la morfología foliar (aciculiforme o planifolia) y el hábito foliar (perennifolio, marcescente o caducifolio), incluyendo formaciones puras y mixtas. Se generó una máscara forestal a partir del Mapa Forestal de España y se seleccionaron muestras de entrenamiento. En el caso de Landsat, se calcularon índices espectrales y compuestos estacionales como variables predictoras. Los *embeddings* de AEF integran información multifuente (óptica, radar, LiDAR y climática) en vectores latentes que sintetizan la dinámica anual y la estructura de la vegetación. Ambas bases de datos se clasificaron con Random Forest y los resultados se compararon pixel a pixel mediante una matriz de confusión. Los resultados muestran mayor estabilidad en las clases puras, especialmente coníferas y caducifolias, y mayor confusión en las clases mixtas. Landsat captó mejor los patrones de hábito foliar, mientras que AEF proporcionó una delimitación espacial más precisa y mejor discriminación estructural.

Palabras clave: fenología, morfología foliar, teledetección, embedding

Advancing forest type classification in the Spanish Iberian System: a comparison between Landsat time series and Google's Alpha Earth Foundations

Abstract: The study compares the classification of nine forest types in the Northern Iberian System using two remote sensing sources: seasonal composites from Landsat 8 and 9 and *embeddings* from *Google's AlphaEarth Foundations* (AEF). Forest types were defined according to leaf morphology (acicular or planifolia) and leaf habit (evergreen, marcescent, or deciduous), including pure and mixed formations. A forest mask was generated from the Forest Map of Spain, and training samples were selected. In the case of Landsat, spectral indices and seasonal composites were calculated as predictor variables. AEF *embeddings* integrate multi-source information (optical, radar, LiDAR, and climatic) into latent vectors that synthesize the annual dynamics and structure of vegetation. Both databases were classified with Random Forest and compared pixel by pixel using a cross-confusion matrix. The results show greater stability in pure classes, especially conifers and deciduous trees, and greater confusion in mixed classes. Landsat better captured phenological patterns, while AEF provided more accurate spatial delimitation and better structural discrimination.

Keywords: spectrophenology, leaf morphology, remote sensing, embedding

1. INTRODUCCIÓN

Conocer y cuantificar la distribución de los tipos de bosque clasificados por su morfología foliar (aciculiforme

o plana), y el tipo de hábito foliar (perennifolia, marcescente o caducifolia), es relevante para comprender el funcionamiento y la dinámica de los ecosistemas forestales (Wang *et al.* 2025) y su

adaptabilidad a cambios ambientales (Wang *et al.*, 2024). La morfología foliar refleja compromisos ecofisiológicos en la adquisición y conservación de recursos y condiciona la respuesta frente a sequías y heladas (Kim *et al.* 2021). El hábito foliar determina la duración de la temporada fotosintética y por tanto la productividad primaria y los ciclos de carbono y nutrientes, influyendo directamente en el balance energético e hídrico (Reich *et al.*, 1995). La distribución espacial de los tipos de morfología y hábitos foliares está regulada por los regímenes térmicos y las propiedades del suelo (Ma *et al.*, 2023). Identificar la fenología dominante contribuye a mejorar la modelización de la productividad, los balances de carbono y refuerza la base científica para una gestión forestal resiliente (Andreatta *et al.*, 2025).

Las series temporales de teledetección reflejan la dinámica de la fenología foliar, lo que permite realizar clasificaciones coherentes a escala regional y global (Zhang *et al.*, 2003; Bolton *et al.*, 2020). Para cartografiar la fenología y los tipos de bosque según su hábito foliar se han empleado datos procedentes de múltiples sensores, entre ellos MODIS, Landsat, Sentinel-2 y Sentinel-1 (e.g. Caparrós *et al.*, 2019; Bolton *et al.*, 2020). Sin embargo, debido a su distribución restringida principalmente al ámbito mediterráneo, los bosques marcescentes han recibido una atención limitada en la literatura. De forma similar, la complejidad estructural y espectral de los bosques mixtos ha dificultado su análisis y clasificación detallada. La creciente disponibilidad de datos, junto con nuevas fuentes y productos derivados, ofrece una oportunidad para avanzar en el nivel de detalle y la especificidad de las clasificaciones espectrales, con el objetivo de generar cartografías más precisas y diferenciadas.

El objetivo de este trabajo es comparar el desempeño de los compuestos estacionales de Landsat y los *embeddings* de *AlphaEarth Foundations* (AEF) de Google para la clasificación de áreas forestales dominadas por nueve tipos de bosque. Estos tipos se definen en función de la morfología foliar (aciculiforme o plana) y del hábito foliar (perennifolia, caducifolia o marcescente), incluyendo además formaciones mixtas resultantes de la combinación de las categorías descritas.

2. MATERIAL Y MÉTODOS

2.1. Área de estudio

Este estudio se desarrolla en el Sistema Ibérico Norte, ámbito que abarca sectores de las provincias de Burgos, Soria y La Rioja, sobre una superficie de aproximadamente 1.5×10^6 hectáreas. La región presenta diversidad de formaciones arbóreas, con existencia de especies aciculifolias y planifolias. Entre las perennifolias predominan coníferas como el pino albar (*Pinus sylvestris*), el pino marítimo (*Pinus pinaster*) y la sabina (*Juniperus thurifera*), así como frondosas esclerófilas como la encina (*Quercus ilex*). Son también abundantes las planifolias marcescentes, como el rebollo (*Quercus pyrenaica*) y el quejigo (*Q. faginea*), junto con planifolias caducifolias como el haya (*Fagus sylvatica*) y el roble albar (*Q. petraea*). La frecuente coexistencia de estas especies da lugar a bosques mixtos con una amplia

variedad de combinaciones florísticas y estructurales, lo que incrementa la complejidad ecológica y espectral del territorio.

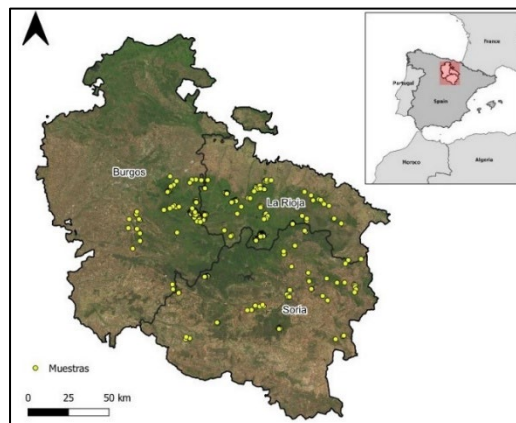


Figura 1. Localización de la zona de estudio.

2.2. Materiales

2.2.1. Máscara forestal

Se generó una máscara forestal a partir del Mapa Forestal de España (MFE) (MITECO, 2023) seleccionando los polígonos de las clases “arbolado” y “arbolado ralo” para representar las áreas forestales arboladas y excluir el resto de las coberturas. Esta máscara se aplicó para seleccionar los píxeles a considerar en la clasificación de los tipos de bosque.

2.2.2. Muestreo

En el área de estudio se consideraron nueve tipos de bosque. Cuatro de estos tipos están dominados por un solo tipo de especie en cuanto a su morfología y hábito foliar: coníferas (C), frondosas perennifolias (E), frondosas marcescentes (M), y frondosas caducifolias (D). Los otros cinco tipos de bosque presentan una combinación de especies con dos tipos fundamentales de hábito foliar y son los considerados en este trabajo como bosques mixtos, en los cuales las especies dominantes son: conífera y frondosa perennifolia (CE), conífera y frondosa marcescente (CM), conífera y frondosa caducifolia (CD), frondosa caducifolia y frondosa perennifolia (DE), frondosa caducifolia y frondosa marcescente (DM) (Tabla 1).

Tabla 1. Tipos de bosque considerados en este estudio.

Identificador	Tipo foliar	Especies dominantes ejemplo
C	Perenne aciculifolia	<i>P. sylvestris</i> , <i>P. pinaster</i> , <i>J. thurifera</i>
E	Perenne planifolia	<i>Q. ilex</i>
M	Marcescente planifolia	<i>Q. pyrenaica</i> , <i>Q. faginea</i>
D	Decidua planifolia	<i>F. sylvatica</i> , <i>Q. petraea</i>
CE	C + E	<i>J. thurifera</i> + <i>Q. ilex</i>
CM	C + M	<i>P. pinaster</i> + <i>Q. faginea</i>
CD	C + D	<i>P. sylvestris</i> + <i>F. sylvatica</i>
EM	E + M	<i>Q. ilex</i> + <i>Q. faginea</i>
DM	D + M	<i>F. sylvatica</i> + <i>Q. pyrenaica</i>

Para la clasificación de los nueve tipos de bosque, se diseñó un muestreo aleatorio simple de 250 muestras, distribuidas proporcionalmente a la superficie de cada cobertura dentro del MFE, garantizando su representatividad. Las muestras de entrenamiento y validación se seleccionaron en zonas representativas mediante inspección visual de imágenes de Google Earth, procurando una distribución espacial homogénea para captar la variabilidad interna de cada clase y minimizar sesgos espaciales (Figura 1).

2.2.3. Fuentes de datos

Se emplearon imágenes Landsat 8 y 9 Collection 2, Level 2 de reflectancia superficial disponibles en *Google Earth Engine* (GEE) correspondientes al periodo comprendido entre diciembre de 2024 y noviembre de 2025. Las imágenes, con tamaño de píxel 30x30m, fueron procesadas para eliminar nubes, sombras y nieve mediante el enmascaramiento de los píxeles identificados en la banda QA_PIXEL (valores 1-5). Posteriormente se calcularon cuatro índices espectrales (NDVI, EVI, NDMI y NBR2) y se generaron compuestos estacionales de invierno (diciembre-febrero), primavera (marzo-mayo), verano (junio-agosto) y otoño (septiembre-noviembre), utilizando la mediana de los datos de cada estación. De este modo, se obtuvo un conjunto de 40 variables predictoras, correspondientes a 10 variables por cada una de las cuatro estaciones.

Como segunda fuente de datos se emplearon los *embeddings* generados por el modelo AEF de Google disponibles en GEE. Estos *embeddings* sintetizan trayectorias temporales coherentes a escala anual, integrando de forma conjunta múltiples fuentes de información geoespacial. Entre ellas se incluyen datos ópticos y térmicos procedentes de satélites como Sentinel-2 y Landsat 8 y 9, datos de radar (Sentinel-1C, ALOS-PALSAR-2), mediciones tridimensionales derivadas de LiDAR (GEDI), modelos digitales de elevación (GLO-30), variables climáticas (ERA5-Land), campos gravitacionales (GRACE) y descripciones textuales asociadas al territorio (Brown *et al.*, 2025). Cada píxel de 10x10 m se caracteriza mediante un vector de 64 dimensiones sin unidades físicas directas; cada dimensión forma parte de una representación latente aprendida y no es interpretable de manera independiente.

2.2.4. Procesamiento y clasificación de los datos de teledetección

La clasificación espectral se realizó mediante el algoritmo Random Forest parametrizado con 300 árboles, seleccionando 7 variables aleatorias por nodo, una fracción de muestreo (bagging) de 0.7 y una semilla fija (42) para garantizar la reproducibilidad. Las muestras se dividieron aleatoriamente en 70 % para entrenamiento y 30 % para validación. El modelo se aplicó para clasificar los tipos forestales en el área de estudio, asignando valores numéricos del 1 al 9 a las clases forestales.

El mismo procedimiento de clasificación se aplicó al conjunto de datos AEF, entrenando igualmente un modelo Random Forest con la misma configuración. Este enfoque permitió comparar el rendimiento de ambos

conjuntos de datos y determinar cuál ofrecía mejores resultados en la clasificación de tipos de bosque.

2.2.5. Comparación de resultados

Con el fin de analizar las diferencias espaciales entre la clasificación generada mediante el modelo del conjunto de *embeddings* y la clasificación derivada de datos Landsat, se realizó una comparación píxel a píxel entre ambos mapas temáticos. Los datos raster clasificados se armonizaron en términos de sistema de referencia, resolución espacial, extensión geográfica y asegurando la correspondencia espacial exacta entre píxeles. A partir de esta superposición se construyó una matriz de confusión, en la que las filas representan las clases del modelo *embeddings* y las columnas las clases del modelo Landsat. Cada celda de la matriz recoge el número de píxeles asignados simultáneamente a la clase *i* en la primera clasificación y a la clase *j* en la segunda.

La matriz fue posteriormente normalizada por filas para obtener frecuencias relativas (%) que describen la redistribución de cada clase original entre categorías del segundo modelo. El análisis se centró en evaluar las diferencias estructurales entre clasificaciones mediante el estudio de las probabilidades de transición y la representación gráfica de las frecuencias de cambio entre clases. Este enfoque permite identificar patrones sistemáticos de discrepancia entre clasificaciones.

3. RESULTADOS Y DISCUSIÓN

La comparación píxel a píxel entre las clasificaciones derivadas de los *embeddings* y de los datos Landsat evidenció diferencias significativas entre clases. El análisis de la matriz de confusión normalizada por filas mostró mayor coincidencia en los tipos de bosque puro, destacando C (68.3%), D (58.0%). En contraste, las clases mixtas (CD, CM y DM) mostraron niveles menores de coincidencia (15% - 32%) (ver Tabla 2).

Tabla 2. Matriz de confusión normalizada por filas (%) entre la clasificación de AEF (filas) y Landsat (columnas).

	Landsat								
	C	D	E	M	CD	CM	DM	EC	EM
C	68	0	17	0.2	3.2	3.3	0	7.2	0.4
D	0.8	58	1.4	12	5	3.3	5	3.8	10
E	6.8	0.1	41	0.1	0.9	1.1	0.1	45	5.3
M	0.3	25	0.6	36	1.6	1.1	23	0.7	12
CD	29	1.5	7.8	2.2	19	15	1.2	16	9.2
CM	18	1.8	2.9	3.3	7.8	15	3.4	29	18
DM	0.2	9.4	1.4	12	2.7	5.3	32	2.8	34
EC	2.3	0	20	0.1	0.6	0.5	0.1	69	7.2
EM	0.5	0.4	4.5	1.9	0.8	0.9	3.5	42	45

Se identificaron discrepancias entre clases próximas, como E y EC: el 44.8% de los píxeles clasificados como E en los *embedding* fueron asignados como EC con Landsat. De forma similar, la clase EM presentó una elevada confusión con EC (42.2%). Estos resultados sugieren que las mayores discrepancias entre modelos

se concentran en clases mixtas o en transiciones entre formaciones estructuralmente afines.

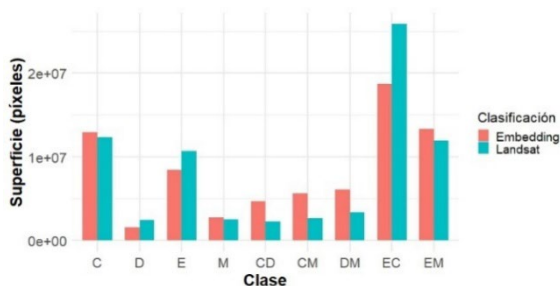


Figura 2. Comparación de la superficie total asignada a cada tipo de bosque por las clasificaciones de embeddings y Landsat.

La Figura 3 compara la clasificación obtenida con imágenes Landsat (A, C, E) y con *embeddings* (B, D, F). En A-B se observa que Landsat capta la estacionalidad y la marcescencia, mientras que estas diferencias no se reflejan en los *embeddings*. En C-D, la menor resolución de Landsat hace que los caminos aparezcan como una clase diferenciada, mientras que en los *embeddings* desaparecen al no corresponder a ninguna categoría temática. Finalmente, en E-F, Landsat confunde *Q. ilex* con *P. halepensis*, mientras que el modelo basado en *embeddings*, al incorporar información estructural, evita esta confusión y clasifica homogéneamente como pinar.

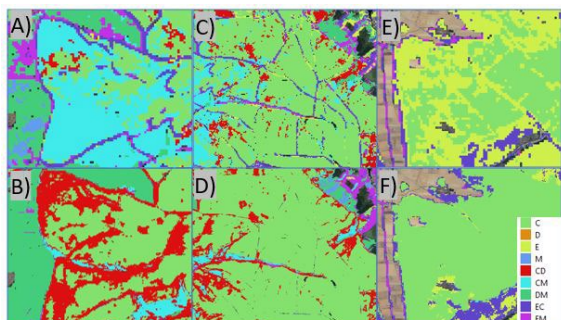


Figura 3. Comparativa entre clasificaciones; Landsat (A, C, E) y embeddings (B, D, F).

4. CONCLUSIÓN

La comparación de los resultados obtenidos en la clasificación de tipos de bosque en el Sistema Ibérico Norte español, mediante compuestos estacionales de Landsat y *embeddings* de *AlphaEarth Foundations* (AEF), revela una elevada capacidad de ambas fuentes para discriminar con detalle distintas morfologías y hábitos foliares. No obstante, se identificaron diferencias atribuibles a características intrínsecas de los datos: la mayor resolución espacial de AEF permitió una delimitación más precisa de formaciones heterogéneas, mientras que la mayor riqueza temporal de los datos Landsat favoreció una mejor captación de los patrones fenológicos de los distintos tipos de bosque.

5. AGRADECIMIENTOS

Este trabajo ha sido en parte financiado con fondos de la Agencia Estatal de Investigación (AEI) a través del

proyecto PID2023-147393OB-I00. Saioa Munuera disfruta de una beca predoctoral Banco Santander (CONTPR-2024-315).

6. REFERENCIAS

- Andreatta, D., Buchmann, N., Jucker, T., *et al.* 2025. Diverging trends in plant phenology and productivity across European mountains in a warming world, *Agricultural and Forest Meteorology*, 375, 110874, <https://doi.org/10.1016/j.agrformet.2025.110874>
- Bolton, D.K., Gray, J.M., Melaas, E.K., *et al.* 2020. Continental-scale land surface phenology from harmonized Landsat 8 and Sentinel-2 imagery. *Remote Sensing of Environment*, 240, 111685, <https://doi.org/10.1016/j.rse.2020.111685>.
- Brown, C.F., Kazmierski, M.R., Pasquarella, V.J., *et al.* 2025. AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data (Version 2). arXiv. <https://doi.org/10.48550/ARXIV.2507.22291>.
- Caparros-Santiago, J.A., Rodriguez-Galiano, V. 2024. Analysing long-term spatiotemporal land surface phenology patterns over the Iberian Peninsula using 250 m MODIS EVI2 data. *Science of The Total Environment*, 954, 176453, <https://doi.org/10.1016/j.scitotenv.2024.176453>
- Kim, J., Kim, Y., Zona, D. *et al.* 2021. Carbon response of tundra ecosystems to advancing greenup and snowmelt in Alaska. *Nat Commun* 12, 6879. <https://doi.org/10.1038/s41467-021-26876-7>
- Ma, H., Crowther, T.W., Mo, L. *et al.* 2023. The global biogeography of tree leaf form and habit. *Nat. Plants* 9, 1795–1809. <https://doi.org/10.1038/s41477-023-01543-5>
- MITECO. 2023. “Most Up-to-Date Version of the Spanish Forest Map.” <https://www.miteco.gob.es/es/cartografia-y-sig/ide/descargas/biodiversidad/mfe.html>.
- Reich, P.B., Walters, M.B., Ellsworth, D.S. 1992. Leaf life-span in relation to leaf, plant, and stand characteristics among diverse ecosystems. *Ecological Monographs* 62 (3), 365-392. <https://doi.org/10.2307/2937116>
- Wang, Y., Yang, D., Vargas, G., *et al.* 2024. Leaf habit differentiation explains trait tradeoffs across savanna woody plants. *Forest Ecosystems*, 11, 100190. <https://doi.org/10.1016/j.fecs.2024.100190>.
- Wang, L., Li, P., Peng, Y., Ren, P. *et al.* 2025. The widely increasing sensitivity of vegetation productivity to phenology in northern middle and high latitudes. *Geophysical Research Letters*, 52, e2024GL113892. <https://doi.org/10.1029/2024GL113892>
- Zhang, X., Friedl, M. A., Schaaf, C. B., *et al.* 2003. Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, 84(3), 471–475. [https://doi.org/10.1016/S0034-4257\(02\)00135-9](https://doi.org/10.1016/S0034-4257(02)00135-9)