

Morales-Martín, A.; Torres-Sánchez, J.; Vázquez-Piqué, J.; Mesas-Carrascosa, F.; Gutiérrez, P.; Montoro Girona, M.; Pardos, M.; Calama, R. Evaluación de la combinación de variables dasométricas y de teledetección en la predicción de la vecería en *Pinus pinea* L. mediante Machine Learning

# Evaluación de la combinación de variables dasométricas y de teledetección en la predicción de la vecería en *Pinus pinea* L. mediante Machine Learning

Morales-Martín, Alejandro <sup>1</sup> Torres-Sánchez, Jorge <sup>2</sup> Vázquez-Piqué, Javier <sup>3</sup> Mesas-Carrascosa, Francisco Javier <sup>2</sup> Gutiérrez, Pedro Antonio <sup>1</sup> Montoro Girona, Miguel <sup>3,4</sup> Pardos, Marta <sup>5</sup> Calama, Rafael <sup>5</sup>

<sup>1</sup> Universidad de Córdoba. Departamento de Ciencia de la Computación e Inteligencia Artificial. Edif. Albert Einstein (C2)–3ª Planta. Campus de Rabanales Ctra. N-IV km. 396 C.P. 14071, Córdoba, España.

<sup>2</sup> Universidad de Córdoba. Departamento de Ingeniería Gráfica y Geomática. Edif. Gregor Mendel (C5)–2ª Planta. Campus de Rabanales Ctra. N-IV km. 396 C.P. 14071, Córdoba, España.

<sup>3</sup> Universidad de Huelva. Departamento de Ciencias Agroforestales. Avda. Fuerzas Armadas S/N. C.P. 21007, Huelva, España.

<sup>4</sup> Groupe de Recherche en Écologie de la MRC Abitibi (GREMA), Forest Research Institute (IRF), Université du Québec en Abitibi-Témiscamingue (UQAT), 341 Rue Principale Nord, Amos, Québec J9T 2L8, Canada.

<sup>5</sup> Instituto de Ciencias Forestales (ICIFOR-INIA), CSIC, Ctra. de la Coruña km 7,5 C.P. 28040, Madrid, España.

ORCID: Morales-Martín 0000-0003-1133-237X Torres-Sánchez 0000-0003-1420-0145 Vázquez-Piqué 0000-0002-7133-4142 Mesas-Carrascosa 0000-0002-5674-1292 Gutiérrez 0000-0002-2657-776X Montoro Girona 0000-0002-6916-3639 Pardos 0000-0002-5567-5406 Calama 0000-0002-2598-9594

Correspondencia: [g52moma@uco.es](mailto:g52moma@uco.es) [o22tosaj@uco.es](mailto:o22tosaj@uco.es) [jpique@uhu.es](mailto:jpique@uhu.es) [ig2mecaf@uco.es](mailto:ig2mecaf@uco.es) [pagutierrez@uco.es](mailto:pagutierrez@uco.es) [miguel.montoro@uqat.ca](mailto:miguel.montoro@uqat.ca) [pardos@inia.csic.es](mailto:pardos@inia.csic.es) [rcalama@inia.csic.es](mailto:rcalama@inia.csic.es)

## RESUMEN

Uno de los principales desafíos en la estimación cuantitativa de la producción de fruto es la elevada variabilidad interanual que presentan numerosas especies leñosas, como el pino piñonero (*Pinus pinea* L.). Discriminar entre parcelas en producción y sin producción de piña a lo largo de los años resulta complejo, ya que los factores que intervienen no están completamente esclarecidos. Este estudio analiza la influencia de incorporar variables obtenidas mediante teledetección a las variables dasométricas tradicionalmente empleadas en la modelización de la producción de piña. Para ello, se evaluaron tres combinaciones de variables: dasométricas (SD), índices de vegetación (SIV) y su combinación (ST). Con dichos conjuntos de datos se desarrollaron modelos de clasificación binaria usando Random Forest (RF) y XGBoost (XGB). Los resultados muestran que ST con XGB ofrece el mejor rendimiento, alcanzando un F1 (media armónica de Precision y Recall) del 67,6%. En futuros estudios es conveniente incorporar los índices de vegetación como predictores relevantes en la clasificación de parcelas en producción de piña.

**Palabras clave:** imágenes de satélite, índices de vegetación, modelo predictivo, clasificación binaria.

Fecha de recepción: 7 febrero 2026 · Fecha de aceptación: 7 febrero 2026

# Evaluación de la combinación de variables dasométricas y de teledetección en la predicción de la vecería en *Pinus pinea* L. mediante Machine Learning

Morales-Martín, Alejandro <sup>(1)</sup>, Torres-Sánchez, Jorge <sup>(2)</sup>, Vázquez-Piqué, Javier <sup>(3)</sup>, Mesas-Carrascosa, Francisco Javier <sup>(2)</sup>, Gutiérrez, Pedro Antonio <sup>(1)</sup>, Montoro Girona, Miguel <sup>(3,4)</sup>, Pardos, Marta <sup>(5)</sup>, Calama, Rafael <sup>(5)</sup>


<sup>(1)</sup> Universidad de Córdoba. Departamento de Ciencia de la Computación e Inteligencia Artificial, España.

 0000-0003-1133-237X, g52momaa@uco.es;  0000-0002-2657-776X, pagutierrez@uco.es.


<sup>(2)</sup> Universidad de Córdoba. Departamento de Ingeniería Gráfica y Geomática., España.

 0000-0003-1420-0145, o22tosaj@uco.es;  0000-0002-5674-1292, ig2mecaf@uco.es

<sup>(3)</sup> Universidad de Huelva. Departamento de Ciencias Agroforestales., España.

 0000-0002-7133-4142, jpique@uhu.es

<sup>(4)</sup> Groupe de Recherche en Écologie de la MRC Abitibi (GREMA), Forest Research Institute (IRF), Université du Québec en Abitibi-Témiscamingue (UQAT) Nord, Amos, Québec J9T 2L8, Canada.

 0000-0002-6916-3639, miguel.montoro@uqat.ca

<sup>(5)</sup> Instituto de Ciencias Forestales (ICIFOR-INIA), CSIC, 5 C.P. 28040, España.

 0000-0002-5567-5406, pardos@inia.csic.es;  0000-0002-2598-9594, rcalama@inia.csic.es

**Resumen:** Uno de los principales desafíos en la estimación cuantitativa de la producción de fruto es la elevada variabilidad interanual que presentan numerosas especies leñosas, como el pino piñonero (*Pinus pinea* L.). Discriminar entre parcelas en producción y sin producción de piña a lo largo de los años resulta complejo, ya que los factores que intervienen no están completamente esclarecidos. Este estudio analiza la influencia de incorporar variables obtenidas mediante teledetección a las variables dasométricas tradicionalmente empleadas en la modelización de la producción de piña. Para ello, se evaluaron tres combinaciones de variables: dasométricas ( $S_D$ ), índices de vegetación ( $S_{IV}$ ) y su combinación ( $S_T$ ). Con dichos conjuntos de datos se desarrollaron modelos de clasificación binaria usando *Random Forest* (RF) y *XGBoost* (XGB). Los resultados muestran que  $S_T$  con XGB ofrece el mejor rendimiento, alcanzando un  $F1$  (media armónica de *Precision* y *Recall*) del 67,6%. En futuros estudios es conveniente incorporar los índices de vegetación como predictores relevantes en la clasificación de parcelas en producción de piña.

**Palabras clave:** imágenes de satélite, índices de vegetación, modelo predictivo, clasificación binaria.

## ***Evaluation of the combination of dasometric and remote sensing variables in the prediction of *Pinus pinea* L. plots with cone production using Machine Learning***

**Abstract:** One of the main challenges in the quantitative estimation of fruit production is the high interannual variability exhibited by numerous woody species, such as stone pine (*Pinus pinea* L.). Discriminating between plots in production and without cone production over the years is complex, as the factors involved are not completely understood. This study analyzes the influence of incorporating variables obtained through remote sensing to the dasometric variables traditionally used in modeling cone production. To this end, three combinations of variables were evaluated: dasometric ( $S_D$ ), vegetation indices ( $S_{IV}$ ), and their combination ( $S_T$ ). With these datasets, binary classification models were developed using *Random Forest* (RF) and *XGBoost* (XGB). The results show that  $S_T$  with XGB offers the best performance, achieving an  $F1$  (harmonic mean of *Precision* and *Recall*) of 67.6%. In future studies, it is advisable to incorporate vegetation indices as relevant predictors in the classification of plots in cone production.

**Keywords:** satellite images, vegetation indices, predictive model, binary classification.

## 1. INTRODUCCIÓN

La producción de fruto en numerosas especies leñosas se caracteriza por episodios de fructificación altamente variables. En particular, el pino piñonero (*Pinus pinea* L.), especie característica de la cuenca mediterránea, muestra una marcada irregularidad en la producción anual de piña, alternando años de alta producción y años con escasa o nula fructificación (Montero *et al.*, 2008). Esta variabilidad se ve además acentuada por su capacidad para crecer bajo condiciones ambientales limitantes, como suelos arenosos y climas continentales extremos (Calama *et al.*, 2011). Como consecuencia, se genera una elevada incertidumbre en la planificación y gestión de este recurso forestal no maderero de alto valor económico, lo que hace necesario disponer de herramientas que permitan anticipar la ocurrencia de la producción a una escala espacial y temporal relevantes para su gestión (Scarascia-Mugnozza *et al.*, 2000). Así, discernir entre parcelas en producción y sin producción de piñas es un paso inicial fundamental para la estimación cuantitativa de la producción anual.

Aunque en las últimas décadas se han desarrollado numerosos modelos para explicar la variabilidad espacial de la producción de piña a distintas escalas (Calama *et al.*, 2008; Calama *et al.*, 2011), integrándose incluso en modelos de crecimiento y producción como PINEA2 (Calama *et al.*, 2007), la descripción y modelización explícita de los patrones que afectan a la producción sigue siendo un desafío.

Paralelamente, la integración de imágenes Landsat con mediciones en campo ha demostrado ser una herramienta muy eficaz para estudiar los ecosistemas forestales, al proporcionar información única y objetiva a múltiples escalas, difícil de alcanzar por otros medios. Índices de vegetación (IV) como el índice de vegetación de diferencia normalizada (NDVI), el índice de humedad de diferencia normalizada (NDMI), y los índices normalizados de áreas quemadas (NBR y NBR<sub>2</sub>) (Henrich *et al.*, 2012) permiten capturar las variaciones en la actividad fotosintética, el contenido hídrico y el estado estructural de la vegetación, aspectos estrechamente relacionados con los procesos productivos y reproductivos de las especies forestales (Long *et al.*, 2023).

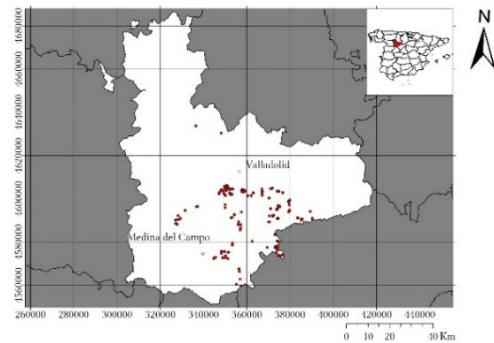
Este estudio pretende analizar el comportamiento de estos cuatro IVs en la clasificación de parcelas con y sin producción de piña y evaluar el rendimiento de dos algoritmos de *Machine Learning* (ML) utilizando como datos de entrada los datos de campo (variables dasométricas), los IVs y su combinación. La hipótesis de partida es que el uso de IVs derivados de imágenes Landsat permiten clasificar entre parcelas en producción y sin producción de piña con buenos niveles de precisión.

## 2. MATERIAL Y MÉTODOS

### 2.1. Área de estudio

En otoño de 1995 se instaló en la provincia de Valladolid, España (Fig.1) una red de 140 parcelas de muestreo en rodales coetáneos de pino piñonero para desarrollar

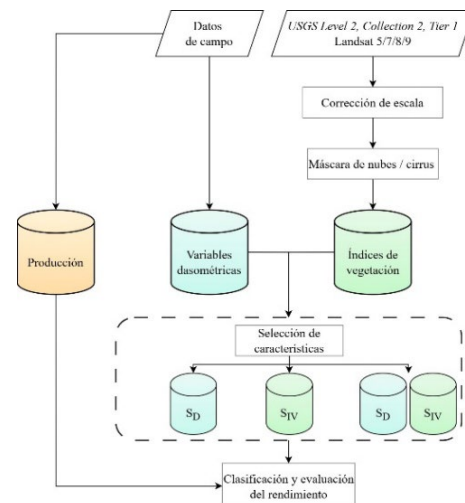
modelos de crecimiento y rendimiento de madera y piña (Calama *et al.*, 2008). Estas parcelas son circulares, de tamaño variable, con 20 árboles cada una.



**Figura 1.** Localización de las parcelas de muestreo de *Pinus pinea* L. en la provincia de Valladolid (ETRS89/UTM zona 30N).

### 2.2. Compilación de datos

La Figura 2 ilustra a modo de esquema el flujo de trabajo seguido en este estudio.



**Figura 2.** Proceso para la clasificación de parcelas en producción y sin producción de piña.

Se utilizaron dos tipos de datos: las variables dasométricas obtenidas de mediciones *in situ* y los IVs derivados de imágenes Landsat. Para evaluar la influencia de los IVs en la clasificación de parcelas con y sin producción de piña, se utilizaron tres combinaciones de variables: (i) exclusivamente variables dasométricas (S<sub>D</sub>), (ii) exclusivamente IVs (S<sub>IV</sub>) y (iii) fusión de las variables dasométricas y los IVs (S<sub>T</sub>). Para hacer una clasificación supervisada es necesario definir las clases presentes, en este caso «no producción» y «producción»; para ello, se etiquetaron las parcelas de acuerdo con los datos de producción como verdad terreno: «producción», cuando al menos un árbol dentro de la parcela produce piña, y «no producción», si ningún árbol presenta producción. Una vez etiquetadas las parcelas, se clasificaron mediante dos algoritmos de ML: *Random Forest* (RF) y *XGBoost* (XGB). Finalmente, se calculó la matriz de confusión y sus métricas derivadas

de evaluación (*Precision*, *Recall* y *F1*) con el fin de comparar y analizar cuantitativamente el rendimiento de cada modelo de clasificación.

### 2.2.1. Datos de campo

Entre 1996 y 2023, se recolectaron manualmente las piñas de los cinco árboles más próximos al centro de cada parcela. Se registraron las alturas y los diámetros medios de los árboles muestreados. Asimismo, la edad de la parcela se estimó a partir del valor medio del conteo de anillos de crecimiento en muestras extraídas de entre tres y cinco árboles, tomadas a la altura del tocón usando una barrena de Pressler. La Tabla 1 recoge todas las variables dasométricas empleadas en el estudio.

**Tabla 1.** Definición de las variables dasométricas utilizadas.

Atributos	VARIABLES	UNIDADES
dbh	Diámetro normal del árbol.	cm
h	Altura del árbol.	m
cw	Diámetro de la copa.	m
T	Edad del árbol.	años
ABHA	Área basimétrica por hectárea.	m <sup>2</sup> /ha
dg	Diámetro medio cuadrático.	cm
SDI	Índice de Densidad de Masa (Reineke).	-
SI	Índice de sitio.	m
Hdom	Altura dominante.	m

### 2.2.2. Datos de imágenes de satélite

Se utilizaron las colecciones USGS Landsat Level 2, Collection 2, Tier 1 de Landsat 5, 7, 8 y 9 disponibles en Google Earth Engine, con una resolución espacial de 30 m y una resolución temporal de 16 días. Adicionalmente, se aplicó el factor de escala proporcionado por la USGS para convertir los valores digitales (DN) de las bandas multiespectrales a valores físicos reales de reflectancia superficial (SR):

$$SR = (DN \times 0,0000275) - 0,2 \quad (1)$$

Por último, se aplicó una máscara de nubes para excluir aquellos píxeles que impiden la observación de la superficie terrestre.

Dado que el pino piñonero tiene un periodo de tres años desde el momento de la floración hasta la apertura de las piñas, es necesario tener en cuenta el vigor de los árboles a lo largo de este periodo. Para filtrar las imágenes de Landsat, se definió el año de referencia ( $Y_{ref}$ ) como el año en el que se evalúa la producción de piña. Una vez definido, se establecieron cuatro intervalos temporales ( $Y_0$ ,  $Y_1$ ,  $Y_2$  y  $Y_3$ ) con inicio en el mes de abril para capturar el ciclo fenológico al completo: (i)  $Y_0$ , desde abril de  $Y_{ref}$  hasta abril de  $Y_{ref+1}$ ; (ii)  $Y_1$ , desde abril de  $Y_{ref-1}$  hasta abril de  $Y_{ref}$ ; (iii)  $Y_2$ , desde abril de  $Y_{ref-2}$  hasta abril de  $Y_{ref-1}$ ; y (iv)  $Y_3$ , desde abril de  $Y_{ref-3}$  hasta abril de  $Y_{ref-2}$ .

Una vez filtradas, corregidas y enmascaradas las imágenes, se calcularon la media y la mediana anual de

cada uno de los IVs (Eq. 2-5) estudiados en cada uno de los intervalos temporales y la desviación típica de los cuatro intervalos analizados.

$$NDVI = \frac{NIR - R}{NIR + R} \quad (2)$$

$$NDMI = \frac{NIR - SWIR_1}{NIR + SWIR_1} \quad (3)$$

$$NBR = \frac{NIR - SWIR_2}{NIR + SWIR_2} \quad (4)$$

$$NBR_2 = \frac{SWIR_1 - SWIR_2}{SWIR_1 + SWIR_2} \quad (5)$$

Donde R es la banda del rojo, NIR es la banda del infrarrojo cercano y SWIR es la banda del infrarrojo medio.

### 2.3. Clasificación y evaluación

El enfoque adoptado es de carácter predictivo y estático a nivel de parcela-año, donde cada observación se considera independiente. No obstante, se incorpora información multitemporal de los IVs que resumen el estado de la vegetación en los tres años previos, permitiendo capturar indirectamente la dinámica temporal del proceso reproductivo. Este problema se planteó como una clasificación binaria, donde el modelo predice entre dos clases: «no producción» y «producción».

Los datos de entrada de los algoritmos de RF y XGB fueron los vectores de características asociados a cada parcela ( $S_D$ ,  $S_{IV}$  y  $S_T$ ). Cada uno de estos sets se dividió de manera estratificada en entrenamiento (80%) y test (20%). Adicionalmente, se ajustaron los pesos en el conjunto de entrenamiento, asegurando el balanceo entre ambas clases. El ajuste de los hiperparámetros de dichos algoritmos de ML fue mediante una búsqueda en cuadrícula a través de una validación cruzada estratificada. La metodología propuesta utilizó el método de validación cruzada de 5 iteraciones, con 10 ejecuciones independientes y 10 semillas distintas.

Se aplicaron distintas métricas de evaluación, tales como *Precision*, *Recall* y *F1* (media armónica de *Precision* y *Recall*), con el fin de comparar y analizar cuantitativamente el rendimiento de cada modelo:

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

donde TP representa los verdaderos positivos, TN los verdaderos negativos, FP los falsos positivos y FN los falsos negativos.

Finalmente, se emplearon los análisis estadísticos de Kolmogorov-Smirnov y la prueba *t* de Student para evaluar el efecto de los IVs en la clasificación de parcelas con y sin producción de piñas basada en los algoritmos de ML.

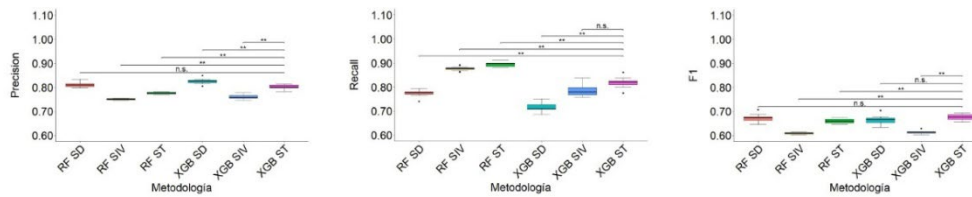


Figura 3. Efecto de los índices de vegetación en dos algoritmos de ML.

### 3. RESULTADOS Y DISCUSIÓN

En este estudio se observaron diferencias significativas en el rendimiento promedio de todas las métricas descritas, influenciadas por el tipo de datos de entrada y la metodología escogida (Fig. 3).

Basándonos en la métrica *Precision*, los modelos entrenados con el conjunto de datos S<sub>D</sub> mostraron el mejor rendimiento en ambos algoritmos de ML, alcanzando valores del 81,1% para RF y del 82,5% para XGB. En cuanto a la métrica *Recall*, los mejores resultados se obtuvieron al utilizar los datos S<sub>T</sub>, con valores del 89,6% para RF y del 81,8% para XGB. Respecto a la métrica *F1*, la combinación de los datos S<sub>T</sub> y con el algoritmo XGB presentó el mejor desempeño, con un valor del 67,6%; mientras que para el algoritmo RF, el mejor desempeño se observó utilizando los datos S<sub>D</sub>, con una puntuación del 67,2%. Adicionalmente, la combinación S<sub>D</sub> y XGB obtuvo una puntuación *F1* del 66,5%, mientras que la combinación S<sub>IV</sub> y XGB fue del 60,7%.

Los resultados obtenidos ponen de manifiesto la complejidad inherente a la clasificación de parcelas en producción. Si bien la combinación de las variables dasométricas y los IVs contribuye a la mejora de la capacidad predictiva de los modelos, se evidencia la necesidad de incorporar nuevas variables que permitan capturar de forma más completa los factores que condicionan la elevada variabilidad interanual en la producción de piña. Según Calama *et al.* (2011), la inclusión de los factores climáticos resulta fundamental, ya que estos explican entre un 60 y un 80% de la variación temporal en la producción de piñas a escala regional, al influir directamente en los procesos de floración y fructificación. Por tanto, futuros trabajos abordarán la inclusión de variables climáticas, el uso de la mediana del periodo de mayor crecimiento vegetativo y la estimación cuantitativa de la producción de piñas.

### 4. CONCLUSIONES

El uso combinado de los IVs y las variables dasométricas permitió mejorar el rendimiento de los algoritmos de ML en la clasificación de parcelas con y sin producción de piña. No obstante, para incrementar la robustez y precisión de los modelos, resulta necesario incorporar nuevas variables (*i.e.*, datos climáticos) que capturen de forma más completa los factores que condicionan la producción. En este marco, la contribución científica de esta investigación radica en la incorporación de los IVs como predictores relevantes para la modelización de la producción en especies leñosas, estableciéndose como base metodológica de futuros estudios.

### 5. AGRADECIMIENTOS

Esta investigación es parte del proyecto PID2022-136906OB-C21/C22, financiada por MICIU/AEI/10.13039/501100011033/ y FEDER/UE; la Comisión Europea, proyecto *Test and Experiment Facilities for the Agri-Food Domain*, AgriFoodTEF (DIGITAL-2022-CLOUD-AI-02, 101100622) y la Cátedra Internacional ENIA de Inteligencia Artificial y Agricultura (TSI-100921-2023-3). Alejandro Morales es investigador pre-doctoral FPU (FPU21/03433) y beneficiario de la ayuda de movilidad FPU 2025 (EST25/00144). Los autores agradecen a Javier Gordo y al Servicio Territorial de Medio Ambiente de Valladolid todo el apoyo prestado en el mantenimiento y seguimiento anual de la red de parcelas de producción de piña.

### 6. REFERENCIAS

- Calama, R., Sánchez-González, M., & Montero, G. (2007). Integrated management models for Mediterranean multifunctional forests: the case of stone pine (*Pinus pinea* L.). *EFI Proceedings*, 56, 57-69.
- Calama, R., Gordo, F. J., Mutke, S., & Montero, G. (2008). An empirical ecological-type model for predicting stone pine (*Pinus pinea* L.) cone production in the Northern Plateau (Spain). *Forest Ecology and Management*, 255(3-4), 660-673. <https://doi.org/10.1016/j.foreco.2007.09.079>.
- Calama, R., Mutke, S., Tomé, J., Gordo, J., Montero, G., & Tomé, M. (2011). Modelling spatial and temporal variability in a zero-inflated variable: the case of stone pine (*Pinus pinea* L.) cone production. *Ecological Modelling*, 222(3), 606-618. <https://doi.org/10.1016/j.ecolmodel.2010.09.020>.
- Henrich, V., Krauss, G., Götze, C., Sandow, C. (2012). IDB - [www.indexdatabase.de](http://www.indexdatabase.de), Entwicklung einer Datenbank für Fernerkundungsindizes. AK Fernerkundung, Bochum, 4.-5. 10. 2012.
- Long, L., Chen, Y., Song, S., Zhang, X., Jia, X., Lu, Y., & Liu, G. (2023). Remote Sensing Monitoring of Pine Wilt Disease Based on Time-Series Remote Sensing Index. *Remote Sensing*, 15(2), 360. <https://doi.org/10.3390/rs15020360>.
- Montero, G., Calama, R., & Ruiz-Peinado, R. (2008). Selvicultura de *Pinus pinea* L. In: Montero, G., Serrada, R., & Reque, J. (Eds.), *Compendio de Selvicultura de Especies*. Madrid, INIA – Fundación Conde del Valle de Salazar, pp. 431-470.
- Scarascia-Mugnozza, G., Oswald, H., Piussi, P., & Radoglou, K. (2000). Forests of the Mediterranean region: gaps in knowledge and research needs. *Forest Ecology and management*, 132(1), 97-109. [https://doi.org/10.1016/S0378-1127\(00\)00383-2](https://doi.org/10.1016/S0378-1127(00)00383-2).