



UNIVERSIDAD  
**COMPLUTENSE**  
M A D R I D



Universidad  
Rey Juan Carlos



**POLITÉCNICA**

# Máster Universitario en Restauración de Ecosistemas

## Predicción de rasgos funcionales foliares de *Pinus nigra* y *Pinus sylvestris* con espectroscopía Vis-NIR

### Presentado por:

D. Javier Pajares Pérez

### Directores:

Dra. Ana Isabel García-Cervigón Morales y Dr. David Sánchez Pescador

### Tutora académica:

Dra. Ana Isabel García-Cervigón Morales

En Madrid a 27 de enero de 2022

# ÍNDICE

1. RESUMEN .....	1
ABSTRACT.....	2
2. INTRODUCCIÓN.....	3
3. MÉTODOS .....	7
<b>3.1. Descripción de la zona y especies de estudio.....</b>	<b>7</b>
<b>3.2. Diseño del experimento.....</b>	<b>8</b>
<b>3.3. Procesamiento y obtención de datos.....</b>	<b>10</b>
<b>3.4. Análisis estadístico y construcción de modelos.....</b>	<b>12</b>
4. RESULTADOS.....	14
<b>4.1. <i>Pinus nigra</i> .....</b>	<b>15</b>
4.1.1. Espesor foliar .....	16
4.1.2. LDMC .....	16
4.1.3. SLA .....	17
<b>4.2. <i>Pinus sylvestris</i> .....</b>	<b>17</b>
4.2.1. Espesor foliar .....	18
4.2.2. LDMC .....	19
4.2.3. SLA .....	19
5. DISCUSIÓN .....	20
<b>5.1. Comparación entre algoritmos .....</b>	<b>22</b>
<b>5.2. Comparación entre pretratamientos quimiométricos .....</b>	<b>22</b>
<b>5.3. Comparación entre rangos espectrales .....</b>	<b>23</b>
6. CONCLUSIONES .....	25
7. PROYECCIONES FUTURAS .....	26
8. AGRADECIMIENTOS .....	27
9. BIBLIOGRAFÍA .....	28

## 1. RESUMEN

El conocimiento de los rasgos funcionales de las plantas ha resultado ser una herramienta útil para comprender sus estrategias ecológicas y sus relaciones con el ecosistema. En este contexto, la variabilidad funcional interespecífica ha sido fundamental para explicar el ensamblaje de comunidades, sin embargo, estudios recientes recalcan la importancia de la variabilidad funcional intraespecífica (ITV) para entender la diversidad de los ecosistemas y la coexistencia de las plantas, especialmente en aquellas comunidades monoespecíficas o con pocas especies. En el marco de la ITV surge el concepto de fenotipo como la integración de múltiples rasgos funcionales, resultado del ajuste ecológico de un genotipo a las condiciones del entorno. El fenotipado masivo de una comunidad puede suponer un gran reto, sin embargo, estudios recientes han planteado la construcción de modelos de calibración con espectroscopía Vis-NIR para fenotipar comunidades enteras. Basándonos en esto, este TFM planteó la construcción de modelos capaces de predecir tres rasgos funcionales foliares (espesor foliar, contenido en materia seca -LDMC- y área foliar específica -SLA-) de dos especies de pinos coexistentes a partir de los espectros de absorción lumínica obtenidos en el laboratorio. Para ello, se probaron tres algoritmos (PLS, SVM-L y RF), 11 pretratamientos quimiométricos y 7 rangos espectrales diferentes con el objetivo de probar la capacidad de predicción de los modelos, así como el algoritmo, pretratamiento y rango más eficaces para predecir cada rasgo funcional en cada especie. Los modelos construidos tuvieron una capacidad de predicción baja o muy baja en todos los casos, probablemente debido a la presencia de *outliers*, aunque también pudo deberse al tipo de rasgos medidos. Es, por tanto, necesario desarrollar más estudios en este campo para mejorar el conocimiento sobre esta prometedora técnica en el ámbito de la ecología.

Palabras clave: Área foliar específica, contenido en materia seca, espesor foliar, espectrometría, fenotipado masivo, variabilidad funcional intraespecífica.

## ABSTRACT

Knowledge of plant functional traits has turned out to be a useful tool to understand plant ecological strategies and their relationships with the ecosystem. In this context, interspecific trait variability has been key to explain the assembly of communities, however, recent studies emphasize the importance of intraspecific trait variability (ITV) to understand the diversity of ecosystems and the coexistence of plants, especially in those monospecific or species-poor communities. Within the framework of ITV, the concept of phenotype emerges as the integration of multiple functional traits that are the result of the ecological adjustment of a genotype to the environmental conditions. The massive phenotyping of a community is a great challenge, but recent studies have proposed the development of calibration models with Vis-NIR spectroscopy to phenotype whole communities. Based on this, this TFM proposed the development of models capable of predicting three foliar functional traits (leaf thickness, leaf dry matter content -LDMC- and specific leaf area -SLA-) of two coexisting pine species from their light absorption spectra. For this purpose, three algorithms (PLS, SVM-L and RF), 11 chemometric pretreatments and 7 different spectral ranges were tested to evaluate the prediction ability of the models, as well as the most effective algorithm, pretreatment and range to predict each functional trait in each species. Adjusted models had a low or very low prediction ability in all cases, probably due to the presence of outliers, although it could be due to the type of traits that were measured. It is, therefore, necessary to develop more studies in this field to improve knowledge about this promising technique in ecology.

**Keywords:** Intraspecific functional variability, leaf dry matter content, leaf thickness, specific leaf area, spectrometry

## 2. INTRODUCCIÓN

Dentro de la ecología vegetal, se define rasgo funcional como cualquier característica morfológica, fisiológica o fenológica de las plantas, medible a nivel de individuo, con un efecto en su eficacia biológica o *fitness* a través del crecimiento, supervivencia y reproducción (Violle *et al.* 2007). Así se han propuesto multitud de rasgos funcionales que actúan como indicadores de las estrategias ecológicas de las plantas, por ejemplo, la altura máxima de la planta, la densidad de la madera, el tamaño de las semillas, o el área foliar específica (Cornelissen *et al.* 2003; Salgado-Negret, 2016).

El área foliar específica (SLA) es uno de los rasgos funcionales foliares más utilizados en los estudios de este ámbito (Shipley *et al.* 2016). En muchos casos, el SLA de una planta nos puede dar pistas sobre su tasa de crecimiento relativo o su tasa fotosintética. Mientras que unos valores de SLA bajos se suelen relacionar con inversiones altas en “defensas estructurales” de las hojas, asociados comúnmente a los ambientes con baja disponibilidad de recursos, un SLA elevado se suele asociar con entornos ricos en recursos y con bajos niveles de estrés (Cornelissen *et al.* 2003). Además, el SLA, al igual que el grosor o espesor de las hojas, está relacionado con la capacidad fotosintética de las plantas (Reich *et al.* 1997).

Otro rasgo ampliamente utilizado, correlacionado negativamente con el SLA, es el contenido en materia seca de la hoja (LDMC). Este rasgo se relaciona con el tipo de crecimiento del individuo, su vida útil y la capacidad de retención de agua de las hojas. Valores elevados de LDMC se relacionan con hojas más “duras”, por lo tanto, más resistentes a los peligros físicos y de crecimiento más lento que los individuos que presentan valores de LDMC más reducidos, que normalmente se asocian a entornos más productivos. En muchos casos el LDMC se emplea cuando el área foliar es difícil de medir, ya que ambos rasgos miden funciones similares (Cornelissen *et al.* 2003).

Todo ello hace que mediante la caracterización de estos rasgos funcionales se pueda obtener gran cantidad de información sobre cómo el entorno abiótico y biótico afecta a las plantas al estar condicionadas por la disponibilidad de recursos dentro de los ecosistemas (Shipley *et al.* 2016). Además, los diferentes rasgos capacitan a las plantas para proveer diferentes funciones ecosistémicas y servicios asociados. Por ejemplo, las propiedades del dosel y la hojarasca

tienen una gran influencia en la distribución de las precipitaciones, la infiltración y la escorrentía (Westoby & Wright, 2006). Por lo tanto, el conocimiento de estos rasgos es fundamental para conocer el estado de los organismos, no solo a nivel de individuo y comunidad, sino a nivel de ecosistema.

En este contexto, las especies han desarrollado distintos rasgos funcionales y estrategias para coexistir dentro de los ecosistemas, sirviendo, por tanto, esta diversidad funcional o variabilidad de rangos funcionales entre especies para explicar la coexistencia y el ensamblaje de comunidades (MacArthur & Levins, 1967; Mason & de Bello, 2013). Esta variabilidad interespecífica de rasgos se basa en la utilización, a nivel taxonómico, de valores medios de los rasgos funcionales para generalizar a nivel de especie (Violle *et al.* 2012).

Por otro lado, el entorno también condiciona que los individuos de una misma especie puedan tener rasgos funcionales diferentes, fruto de la adaptación del entorno donde se encuentren. Este concepto se conoce como variabilidad intraespecífica de rasgos (ITV) (Siefert *et al.* 2015). Así, por ejemplo, en un reciente estudio, Benavides y colaboradores (2021) analizaron la variabilidad de rasgos entre poblaciones de *Pinus sylvestris* en distintas zonas de Europa y descubrieron diferencias significativas en la talla de los individuos, el tamaño de sus hojas y el contenido en nutrientes a lo largo del rango térmico y en función de las precipitaciones.

Aunque tradicionalmente la ecología basada en rasgos funcionales dejaba en un segundo plano la variabilidad intraespecífica de rasgos, centrándose en la variabilidad de rasgos entre especies para explicar el ensamblaje de comunidades y la separación de nichos (MacArthur & Levins, 1967), multitud de estudios recientes están poniendo en valor la importancia de la ITV para explicar la coexistencia de los individuos de una misma especie (Albert *et al.* 2011; Violle *et al.* 2012; Pescador *et al.* 2015). En este aspecto, la escala es un punto crucial, ya que, a pequeña escala, la variabilidad funcional intraespecífica tiene un papel más relevante para explicar cómo se forman las comunidades, al predominar las interacciones planta-planta. En cambio, cuando aumenta la escala, la variabilidad interespecífica parece dirigir el ensamblaje de comunidades.

Por lo tanto, la variabilidad de rasgos, tanto intra como interespecífica, es una pieza clave para entender la diversidad de los ecosistemas. En ecosistemas muy ricos y diversos en especies, los individuos de una misma especie tienden a ser más similares entre sí, dominando, por lo tanto, la variabilidad interespecífica de rasgos. Por el contrario, en las comunidades monoespecíficas

o con pocas especies, los individuos de estas especies tendrán una mayor variabilidad intraespecífica fruto de una especialización individual de nicho, ya que los individuos intentarán minimizar la competencia con los otros individuos de su especie diferenciando su propio nicho (Escudero *et al.* 2021).

En este contexto, es necesario introducir el concepto de fenotipo. El fenotipo se puede definir como la expresión integrada de múltiples rasgos funcionales, resultado del ajuste ecológico de un genotipo a las condiciones bióticas y abióticas del entorno (Benavides *et al.* 2020; Escudero *et al.* 2021). El fenotipado, entonces, sería el conjunto de metodologías y protocolos empleados para medir el crecimiento, arquitectura y composición de las plantas, últimamente asociado a tecnologías no invasivas (Fiorani & Schurr, 2013).

Por tanto, puede parecer un enorme desafío plantear el fenotipado masivo de una comunidad vegetal, entendiendo este como el fenotipado de todos los individuos que conforman el conjunto de la comunidad. Es evidente que el muestreo, medición y análisis en laboratorio de los rasgos de todos los individuos de una comunidad implicaría un tiempo y esfuerzo prácticamente inasumible para la mayoría de los casos, por lo que en los últimos años se han desarrollado técnicas que pueden agilizar el proceso de fenotipado masivo (Carmona *et al.* 2015; Escudero *et al.* 2021). Una de estas técnicas es la utilización de sensores remotos e imágenes tomadas con helicópteros, grúas o tractores, y se ha aplicado especialmente al fenotipado de cultivos, aunque estas técnicas suelen generar problemas con la resolución de las imágenes, lo que dificulta las correlaciones con rasgos funcionales, además de los problemas de mantenimiento y costes que implican los sensores fijos con cables y la invasividad en el campo que implica el uso de vehículos y grúas (White *et al.* 2012).

Otra técnica innovadora en este campo es la construcción de modelos de calibración con espectrometría visible y de infrarrojo cercano (Vis-NIR) a partir de mediciones obtenidas directamente en muestras de plantas. La espectrometría es el registro de las propiedades de la luz después de la irradiación de un material, lo que permite obtener información sobre los rasgos de las plantas, ya que la reflectancia, transmitancia y absorbancia reflejadas indican el tamaño, densidad y forma de los tejidos vegetales, además de su composición química (Costa *et al.* 2018). Mediante análisis quimiométricos sobre los espectros de absorbancia obtenidos en el laboratorio y el uso de modelos de calibración, por ejemplo, mediante una regresión de mínimos cuadrados parciales (PLS), es posible llegar a predecir rasgos funcionales de miles de

individuos y posteriormente fenotiparlos, aunque todavía no hay suficientes estudios al respecto (Escudero *et al.* 2021).

Esta técnica, además de no ser destructiva con los individuos, ahorra tiempo y costes. Es por esto que ha sido utilizada mayoritariamente en el ámbito de la agricultura para fenotipar especies cultivadas (Ecanot *et al.* 2013; Sánchez *et al.* 2014; Gutiérrez *et al.* 2016), aunque últimamente también ha sido aplicada en ecología, con notables resultados, para el estudio de comunidades naturales y plantaciones forestales, utilizándose principalmente para rasgos morfológicos y químicos de la madera (Gaspar *et al.* 2011; Fernandes *et al.* 2017; Dias *et al.* 2020) y las hojas (Serbins *et al.* 2019; Kothari *et al.* 2021).

Así pues, estos modelos predictivos basados en la espectroscopía Vis-NIR pueden ser una técnica innovadora y eficaz para generar conocimiento científico en el ámbito de la diversidad funcional, lo que, además de ser un valor en sí mismo, es una herramienta indispensable para la conservación y restauración de los ecosistemas. Este conocimiento sobre la diversidad funcional de las comunidades y los ecosistemas puede ser aprovechado de forma práctica, a través de la restauración ecológica, para recuperar, no solo la cobertura vegetal de ecosistemas degradados a través de la selección de especies, sino para la recuperación de los procesos y servicios ecosistémicos, además, también puede ayudar a entender los mecanismos de las invasiones biológicas y contribuir, de esta forma, al manejo de los ecosistemas (Salgado-Negret, 2016; Cogollo *et al.* 2020). Asimismo, esta técnica podría tener aplicaciones prácticas en el ámbito de la silvicultura y la restauración forestal, mejorando la comprensión de los bosques y sus funciones ecosistémicas (Wang *et al.* 2021).

El objetivo principal de este Trabajo de Fin de Máster es probar la eficacia de la espectroscopía Vis-NIR para predecir tres rasgos funcionales foliares (espesor foliar, SLA y LDMC) de dos especies arbóreas, *P. nigra* y *P. sylvestris*, que coexisten en un entorno natural (Parque Natural del Alto Tajo), mediante diferentes análisis quimiométricos y modelos predictivos. Los objetivos específicos, son: probar cuál es el algoritmo que predice mejor cada rasgo en cada especie, establecer cuál es el pretratamiento más eficaz para predecir cada rasgo en cada especie y, por último, determinar qué rango espectral está más relacionado con cada rasgo en cada especie.



### 3. MÉTODOS

#### 3.1. Descripción de la zona y especies de estudio

La zona de estudio se sitúa en el Parque Natural del Alto Tajo, en la provincia de Guadalajara, entre los municipios de Zaorejas, Villanueva de Alcorón y Peñalén, (40°43'28.6"N 2°07'00.4"W, 1200 m s.n.m.) (Figura 1). Este Parque Natural posee extensas masas forestales y alberga numerosos hábitats de interés comunitario y una enorme biodiversidad botánica y florística (Ferrero *et al.* 2006).

El clima de la zona de estudio es mediterráneo continental, caracterizado por inviernos fríos con frecuentes heladas, y veranos secos y calurosos (Bastias *et al.* 2013). La precipitación media anual es de 468 mm y la temperatura media anual, de 10,2 °C, siendo mayo el mes más lluvioso (66 mm) y julio el más seco (22 mm), coincidiendo este último como el mes más caluroso (20,2 °C), mientras que el mes más frío es enero (2,5 °C) [Estación Meteorológica de Molina de Aragón (40°50'30"N 1°52'44"O), 1981 – 2010, Agencia Estatal de Meteorología ([www.aemet.es](http://www.aemet.es))]. Los suelos son pobres y poco desarrollados, principalmente de naturaleza caliza (Bastias *et al.* 2013).



Figura 1. Localización de la zona de estudio y aspecto de la misma. Fuente: Google Earth.

La zona de estudio es un pinar mixto de *Pinus nigra* J.F Arnold subsp *salzmannii* (Dunal) Franco y *Pinus sylvestris* L., con clara predominancia de *Buxus sempervirens* L. y presencia de

*Juniperus thurifera* L., *Quercus faginea* Lam. subsp *faginea* y *Juniperus communis* L. en el estrato inferior.

Las especies de estudio son *P. nigra* y *P. sylvestris* ya que representan el estrato arbóreo con mayor cobertura. En primer lugar, *P. nigra* es un árbol de hasta 40 metros de altura, con la corteza cenicienta/plateada en los individuos jóvenes y castaño oscuro en los de mayor edad, acículas flexibles de 6 a 16 cm de largo, agrupadas en fascículos de 2, y de color verde claro. Generalmente prefieren suelos pedregosos o margosos-calizos y su intervalo altitudinal se encuentra entre los 500 y 1800 m (Castroviejo, 2020). La especie tiene una distribución circunmediterránea, concretamente, la subespecie *salzmannii* se localiza en la Península Ibérica, sureste de Francia y el Norte de África (Alejano & Montes, 1996). A nivel local, *P. nigra* es la especie forestal más abundante en el Parque Natural del Alto Tajo, cuyos pinares están considerados como los más extensos y mejor conservados de la Península Ibérica (Plan de Gestión del Alto Tajo, 2017).

Por otro lado, *P. sylvestris* es un árbol que también puede alcanzar los 40 metros de altura, con tronco pardo-rojizo y característico ritidoma anaranjado-rojizo en la parte superior, delgado y que se separa en placas papiráceas. Las acículas son rígidas, de 3 a 7 cm de longitud, en fascículos de 2, y de color verde-glaucos. Generalmente son indiferentes al sustrato y su intervalo altitudinal se encuentra entre los 500 y 2000 m (Castroviejo, 2020). Se trata de una especie con una gran plasticidad ecológica, siendo la especie arbórea con mayor área de distribución de Europa, extendiéndose desde la Península Ibérica hasta Siberia (García & Gonzalez, 1999). En el Parque Natural del Alto Tajo, *P. sylvestris* es la segunda especie forestal más abundante (Plan de Gestión del Alto Tajo, 2017).

### **3.2. Diseño del experimento**

La zona de estudio se seleccionó teniendo en cuenta la diversidad y equitatividad de individuos de *P. nigra* y *P. sylvestris*. En concreto se delimitó una parcela experimental de 125 x 125 m, dividida en 25 subparcelas de 25 x 25 m para facilitar el marcaje y muestreo de los ejemplares (Figura 2).

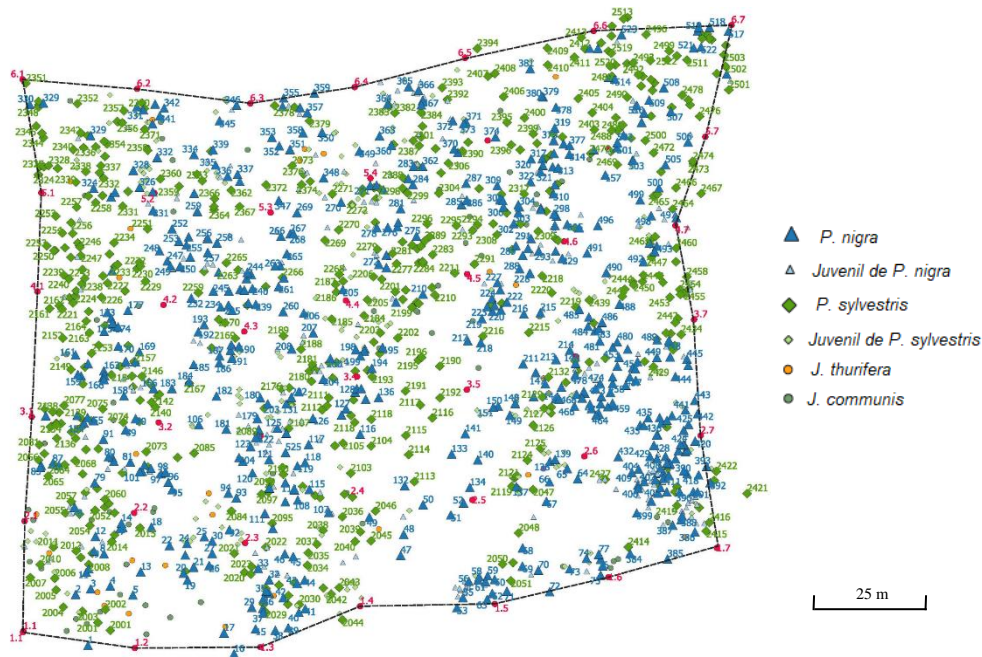


Figura 2. Mapa de parcelas y subparcelas experimentales con los ejemplares de *P. nigra*, *P. sylvestris*, *J. comunis* y *J. thurifera*. Los puntos rojos marcan el límite entre subparcelas de 25 x 25 m.

Se identificaron y georreferenciaron la totalidad de los individuos de *P. nigra* y *P. sylvestris* presentes en la parcela (un total de 525 ejemplares adultos de la primera especie y 520 de la segunda), marcando con una chapa metálica cada ejemplar adulto (diámetro del tronco a la altura del pecho (DBH) superior a 7,5 cm) (Figura 3A). La georreferenciación de los individuos adultos y juveniles (DBH entre 2,5 y 7,5 cm) se realizó con un sistema GPS “Leica Viva GS15” de alta precisión ( $\pm 5$  mm en la horizontal). Además, se tomaron las coordenadas de los ejemplares de otras especies (*J. thurifera* y *J. communis*). El marcaje y mapeado de los pinos se realizó durante los meses de mayo y junio de 2021.

Para construir los modelos predictivos se seleccionaron 100 individuos de *P. nigra* y 102 de *P. sylvestris* de los 525 y 520 totales respectivamente. El proceso de selección fue aleatorio para obtener rasgos de individuos con la mayor variabilidad fenotípica posible. Se planteó un muestreo durante el mes de junio de 2021 donde se recogieron muestras de ramas de los 202 individuos seleccionados utilizando una pértiga extensible de 11 metros de longitud con tijeras de poda acopladas en el extremo (Figura 3B). Estas ramas fueron recogidas y rotuladas con el código numérico del individuo correspondiente para posteriormente seleccionar muestras de hojas de cada rama (Figura 3C). De cada individuo se tomaron cuatro submuestras de 10 acículas cada una (A, B, C y D), haciendo un total de 40 acículas de cada pino (Figura 3D).

Todas las submuestras fueron transportadas al laboratorio Nutrilab, perteneciente al Campus de Móstoles de la Universidad Rey Juan Carlos, para ser analizadas.

### 3.3. Procesamiento y obtención de datos

Una vez en el laboratorio, la submuestra “D” se reservó para la determinación de pigmentos para estudios paralelos, almacenándose a -80 °C. Las otras tres submuestras se destinaron a tomar las medidas para calcular los rasgos funcionales y obtener, mediante espectrofotometría, los espectros de absorbancia de las acículas con el objetivo de realizar los modelos de calibración y, de esta forma, relacionar los rasgos de los individuos con su correspondiente espectro. Por lo tanto, se almacenaron en nevera para evitar la pérdida de humedad.

En primer lugar, se pesaron las submuestras “A”, “B” y “C” con una báscula de precisión para calcular el peso fresco o saturado. De forma paralela se midió el grosor de las acículas con un medidor de esfera (Mitutoyo Co., Aurora, IL, USA) y se escanearon con un escáner Epson Perfection 4870 Photo para calcular el área foliar. Una vez escaneadas, se calculó el área foliar en el entorno de R (R Core Team, 2021), versión 4.0.5, utilizando el paquete *LeafArea* (función “un.ij”) (Katabuchi, 2015) que utiliza las funcionalidades del *software* libre de procesamiento de imagen digital “ImageJ” (Rasband, 1997).

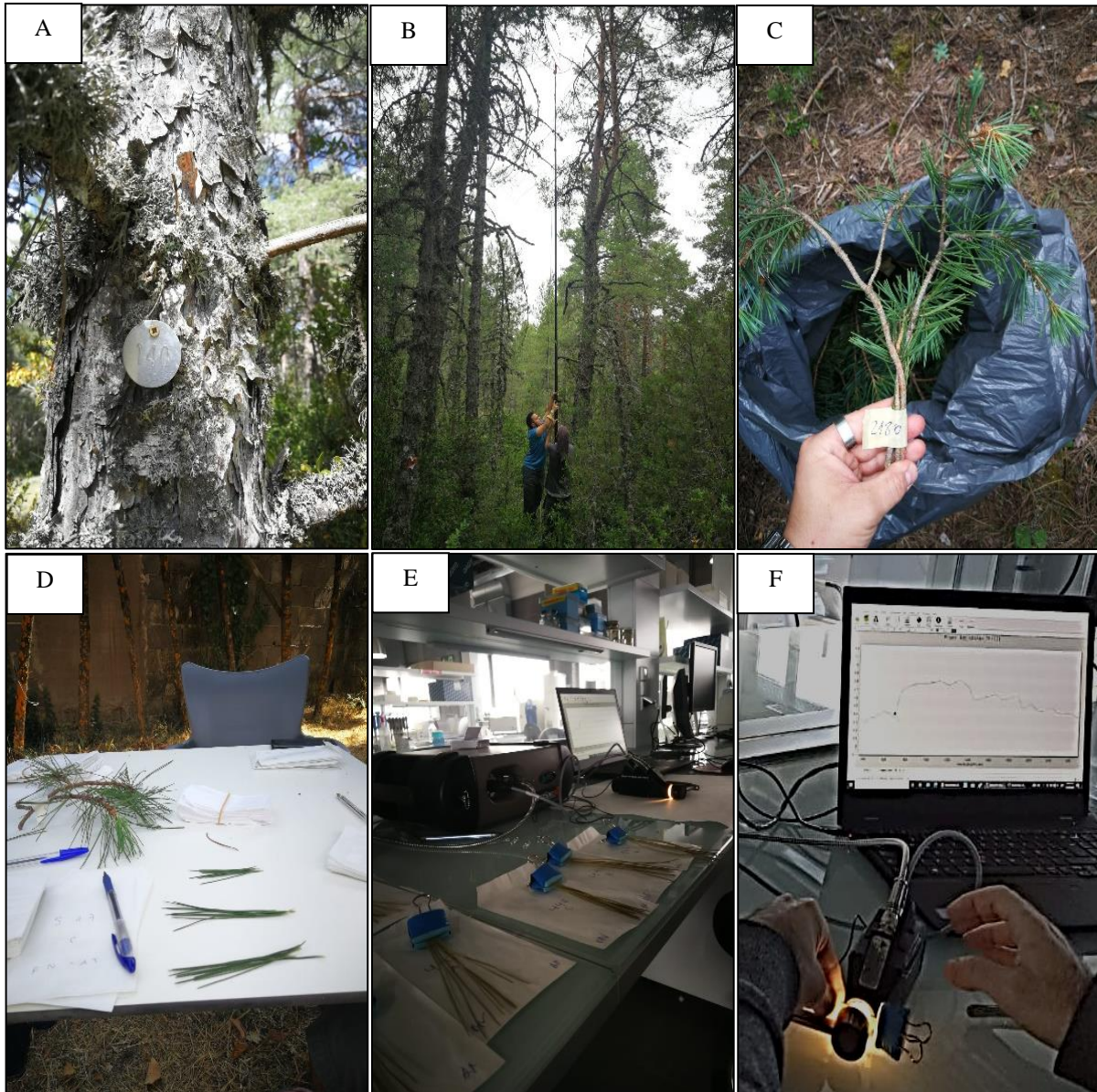
Después, estas muestras se dejaron secar en estufa a 60 °C durante 3 días y se volvieron a pesar de nuevo para obtener el peso seco de cada submuestra de acículas. A partir de estas medidas, se obtuvieron los rasgos funcionales de interés para llevar a cabo el fenotipado de los pinos. El estudio se centró en tres rasgos funcionales de las hojas, concretamente, el espesor de la hoja ( $\mu\text{m}$ ), el área foliar específica (SLA) ( $\text{mm}^2/\text{mg}$ ), calculada de acuerdo con la fórmula [1] y el contenido en materia seca (LDMC) ( $\text{mg/g}$ ), calculado de acuerdo con la fórmula [2] y siguiendo los protocolos estandarizados (ver Cornelissen *et al.* 2003).

$$\text{SLA} = \text{área foliar (mm}^2\text{)} / \text{peso seco (g)} \quad [1]$$

$$\text{LDMC} = \text{peso fresco (mg)} / \text{peso seco (g)} \quad [2]$$

Para finalizar, una vez secas, se midieron los espectros de absorción lumínica de las muestras con un espectrofotómetro “ASD LabSpec 4 (Malvern Panalytical, Malvern, Reino Unido)”,

agrupando las 10 acículas de cada submuestra y sometiéndolas a excitación por una fuente de luz producida por una sonda de reflectancia (Figuras 3E y 3F). Se obtuvieron dos espectros de cada submuestra, uno para el haz de las acículas y otro para el envés, resultando un total de 1212 espectros brutos (202 pinos, tres submuestras por cada pino, dos espectros por submuestra) (Figura 4A).



*Figura 3. Reportaje fotográfico de la obtención de datos. A: Ejemplar de P. nigra con la chapa identificativa; B: Muestreo de ramas con pértiga extensible; C: Muestra de rama de P. sylvestris; D: Proceso de selección de acículas; E: Preparación de acículas de P. nigra para obtención de espectro de absorbancia con espectrofotómetro "ASD LabSpec 4"; F: Obtención de espectro de absorbancia con espectrofotómetro "ASD LabSpec 4" mediante excitación de las acículas por fuente de luz provocada por sonda de contacto.*

### 3.4. Análisis estadístico y construcción de modelos

A partir de los espectros de absorción obtenidos de las acículas de los pinos, se planteó la construcción de los modelos predictivos. En primer lugar, se realizó el promediado de los espectros de haz y envés de cada submuestra para obtener un único espectro de cada submuestra. En casos puntuales, cuando la desviación entre ambos espectros era muy elevada, es decir, cuando había mucha diferencia entre ellos, se seleccionó únicamente el espectro con los picos más pronunciados.

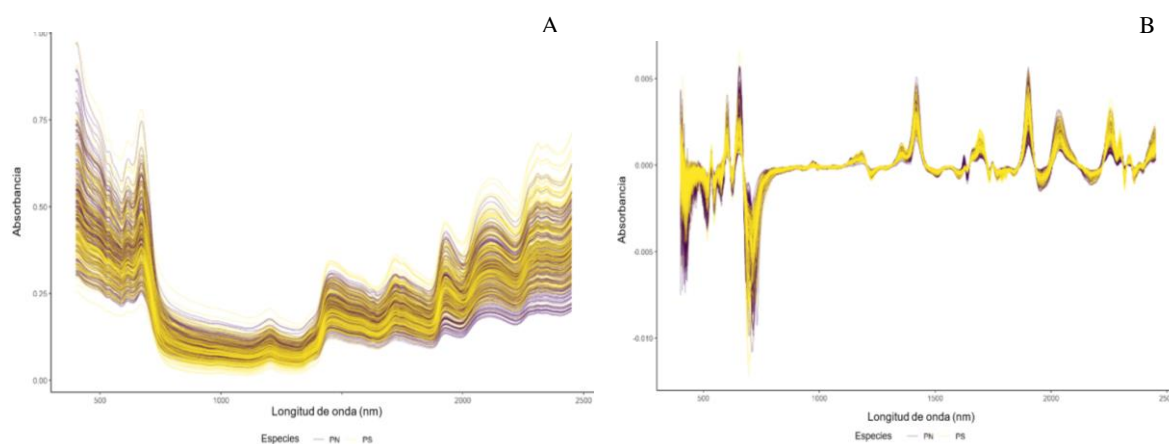


Figura 4. Espectros de absorbancia en el Vis-NIR de muestras foliares de *P. nigra* (morado) y *P. sylvestris* (amarillo). A: Espectros brutos; B: Espectros con pretratamiento de la primera derivada.

Posteriormente, se sometió a los espectros resultantes a 11 pretratamientos quimiométricos, es decir, transformaciones de los datos, para observar cuál de ellos era más útil para procesar los datos, después, se seleccionaron 7 rangos espectrales para determinar qué región del espectro estaba más asociada con cada rasgo funcional y, por último, se probaron tres algoritmos matemáticos (técnicas de regresión y sistemas *machine learning*), para construir los modelos y predecir los rasgos funcionales.

Los 11 pretratamientos quimiométricos fueron los siguientes: Primera derivada (D1) (Figura 4B), segunda derivada (D2), *Detrend* (Dt), *Moving averages* + *Savitzky-Golay* (MA+SG), *Mean centering* (MC), *Mean centering-scaling* (MCS), *Multiple scatter correction* (MSC), *Moving averages* (MVA), *Moving averages* + primera derivada (MVA+D1), *Savitzky-Golay* (SG), *Standard Normal Variate* (SNV). Además, se utilizaron los datos brutos (Raw). Después, para determinar la región del espectro más relacionada con cada rasgo funcional, se establecieron 7 rangos espectrales de cada uno de los espectros resultantes: 400 - 2450 nm (espectro completo con eliminación de las colas para evitar ruido), 400 - 700 nm (rango

visible), 400 - 1000 nm (rango del detector VNIR), 701 - 2450 nm (infrarrojo cercano), 1001 - 2450 nm (rango de detectores SWIR), 1001 - 1800 nm (rango de detector SWIR 1) y 1801 - 2450 nm (rango de detector SWIR 2).

Para la construcción de los modelos se utilizó como variable dependiente los rasgos funcionales y como variable independiente, los espectros de absorción de las acículas. Además, se validaron un total de tres algoritmos diferentes usados para construir los modelos predictivos: Regresión de mínimos cuadrados parciales (PLS), *Supporting Vector Machine* con kernel lineal (SVM-L) y *Random Forest* (RF) con el objetivo de evaluar cuál de estos era más eficaz prediciendo los rasgos funcionales estudiados.

El primero, la regresión PLS es un método de clasificación que generaliza y combina características del análisis de componentes principales y la regresión múltiple para predecir un conjunto de variables dependientes a partir de un conjunto de variables independientes utilizando unos predictores o componentes, 15 en este caso (Abdi, 2003). Por otro lado, SVM-L es un sistema *machine learning* que funciona como método de clasificación discriminativo separando grupos de observaciones mediante un hiperplano óptimo para diferenciar los datos por clases. Para el entrenamiento y predicción de los modelos se utilizó el método kernel lineal (Vapnik Vladimir, 1999). Por último, *Random Forest* es un sistema *machine learning* que funciona como un árbol de decisiones, de modo que cada árbol depende de los valores de tres vectores aleatorios muestreados de forma independiente y con la misma distribución para todos los árboles del bosque. Este algoritmo posee mayor rendimiento ya que permite eliminar la correlación de los datos y que ningún árbol trabaje con todos los datos de entrenamiento (Breiman, 2001).

Para validar cada combinación de pretratamiento, rango espectral y algoritmo se llevaron a cabo un total de 20 simulaciones en cada una de las cuales se seleccionó aleatoriamente el 70% de las muestras para entrenar cada modelo y el 30% restante para validar las predicciones. Esto resultó en un total de 5040 modelos en total para cada rasgo funcional estudiado, 1680 modelos PLS, 1680 modelos SVM-L y 1680 modelos RF (12 pretratamientos  $\times$  7 rangos espectrales  $\times$  20 simulaciones para cada uno) que fueron calibrados y validados. La validación de las predicciones se realizó comparando los valores predichos con los observados, es decir, con los datos de los rasgos funcionales obtenidos en el laboratorio, calculando el coeficiente de validación  $R^2$  para determinar la eficacia de las predicciones. El coeficiente de validación  $R^2$

tiene un valor de 0 a 1 e indica el grado en que la variable independiente (en este caso, los espectros de absorbancia predichos en los modelos) es capaz de explicar la variable dependiente (en este caso, los datos de los rasgos funcionales medidos).

Todos los análisis estadísticos y la construcción de los modelos predictivos se realizaron en el entorno de R (R Core Team, 2021), versión 4.0.5, utilizando principalmente los siguientes paquetes y funciones: *asdreader* (Roudier, 2017) para leer los espectros obtenidos con el espectrofotómetro; *prospectr* (funciones “movav” y “detrend”) (Stevens & Ramirez-Lopez, 2021), *base* (funciones “diff” y “scale”) (R Core Team, 2021), *pls* (función “msc”) (Liland *et al.* 2021) y *mdatools* (función “prep.snv”) (Kucheryavskiy, 2020) para los pretratamientos quimiométricos; *caret* (función “train”) (Kuhn, 2020), *kermlab* (función “ksvm”) (Karatzoglou *et al.* 2004) y *ranger* (función “ranger”) (Wright & Ziegler, 2017) para ajustar los algoritmos PLS, SVM-L y RF respectivamente; y *ggplot2* (Wickham, 2016) para crear los gráficos.

#### 4. RESULTADOS

En primer lugar, para observar la distribución de los datos y detectar posibles datos atípicos u *outliers*, se realizó un diagrama de cajas y bigotes con los datos brutos de los rasgos funcionales para cada especie de pino estudiada (Figura 5). Excepto el espesor foliar de *P. nigra*, el resto de los rasgos funcionales presentaron datos atípicos, especialmente el LDMC de *P. sylvestris*. Por lo tanto, se decidió comparar los valores promedio de  $R^2$  obtenidos en los modelos en lugar de los valores máximos para evitar una sobreestimación de la capacidad de predicción asociada a datos atípicos.

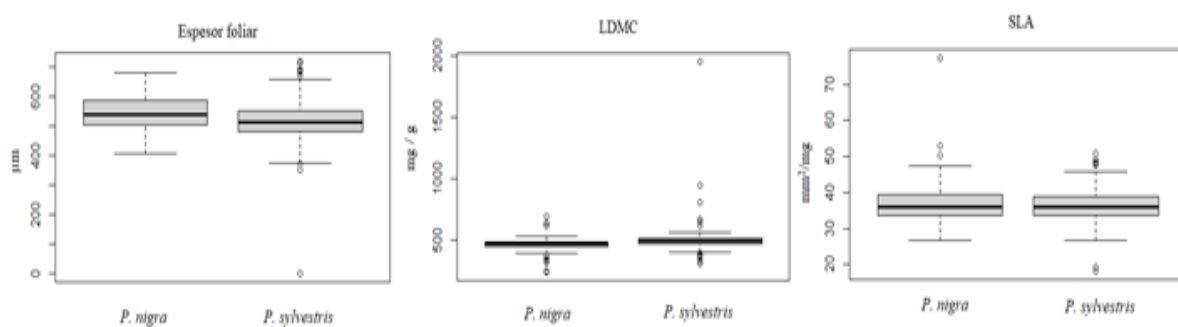


Figura 5. Diagramas de cajas y bigotes para cada rasgo funcional y especie. La caja y los “bigotes” muestran la mediana y los cuartiles de los datos. Los puntos indican la presencia de datos atípicos.



#### 4.1. *Pinus nigra*

Como tendencias generales, en los modelos de PLS y SVM-L, los pretratamientos MSC y SNV fueron los más eficaces prediciendo el conjunto de los rasgos de *P. nigra*, mientras que el pretratamiento de la segunda derivada fue el menos eficaz (Figura 6). En RF, con los datos brutos y los pretratamientos MVA y MC, se obtuvieron valores de  $R^2$  bastante inferiores al resto de pretratamientos para todos los rasgos, mientras que los pretratamientos MVA+D1 y MA+SG fueron los más eficaces prediciendo los rasgos funcionales. En cuanto a los rangos espectrales, el rango de 400 - 700 nm, a nivel general, obtuvo valores medios de  $R^2$  muy reducidos en todos los modelos para todos los rasgos funcionales estudiados, mientras que los rangos de 701 a 2450 nm y de 1001 a 2450 nm fueron los que mayor eficacia de predicción tuvieron para la mayoría de los rasgos estudiados.

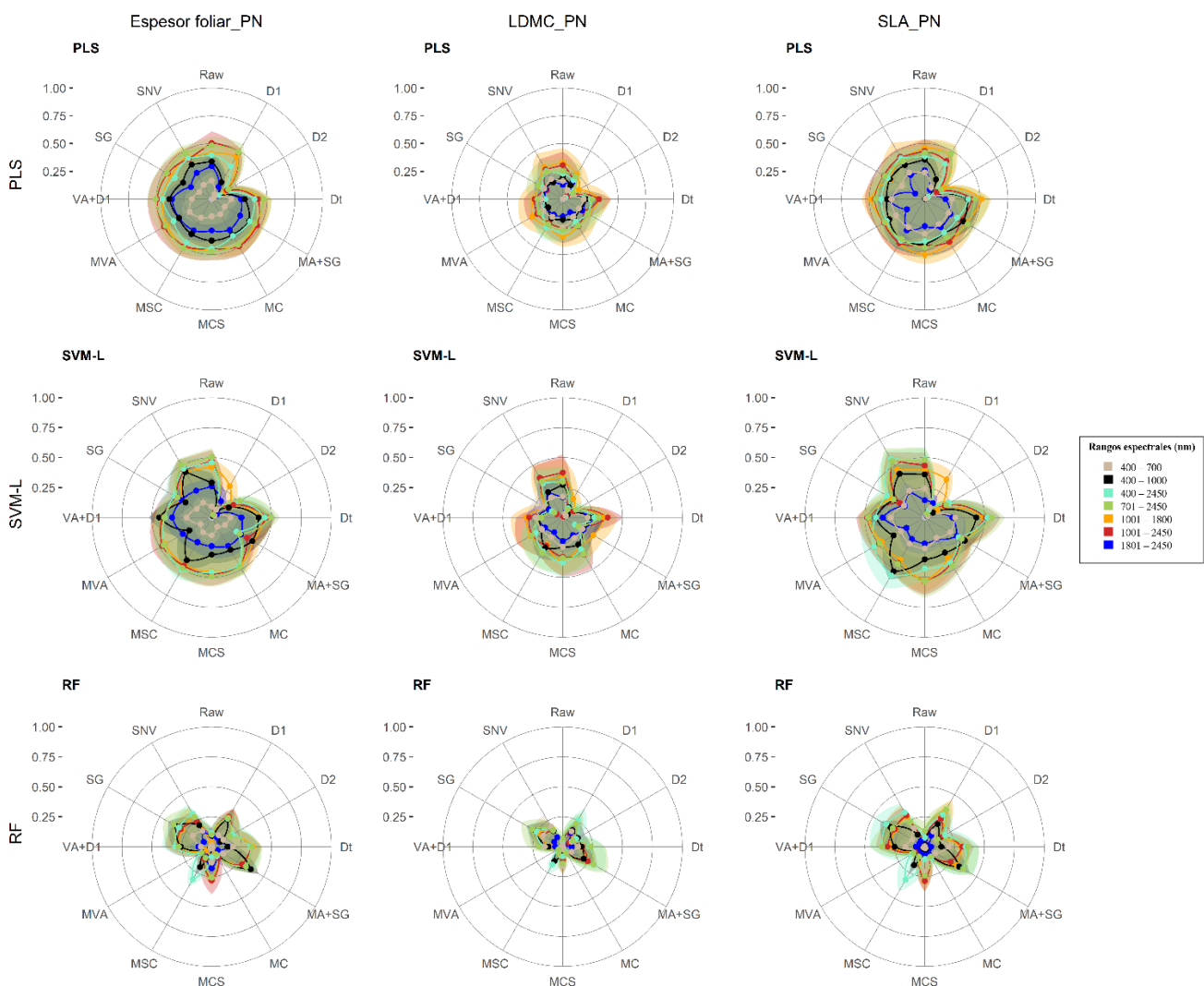


Figura 6. Gráficos polares de los resultados de los modelos predictivos generados con los algoritmos PLS, SVM-L y RF para los rasgos funcionales de *P. nigra*. Se representa el valor de  $R^2$  obtenido por rasgo funcional para cada algoritmo, pretratamiento y rango espectral.

#### 4.1.1. Espesor foliar

Entre los tres algoritmos probados, PLS y SVM-L tienen valores promedio de  $R^2$  similares, siendo mayor en PLS con un valor de 0,34 (en comparación con el 0,32 de SVM-L). RF es el algoritmo que menor media de  $R^2$  tuvo con un valor 0,18 (Figura 6). En cuanto a los pretratamientos con el algoritmo PLS, fue con el MSC cuando se obtuvieron mayores valores promedio de  $R^2$  (i.e. 0,40), aunque, exceptuando el pretratamiento de la segunda derivada, los datos brutos y el resto de pretratamientos obtuvieron valores generales bastante similares de  $R^2$  (entre 0,30 y 0,40). En cuanto al rango espectral con el algoritmo PLS, para el espesor foliar de *P. nigra*, es de 701 a 2450 nm donde se obtuvieron valores medios de  $R^2$  más elevados (0,45). En los modelos de SVM-L, tanto MSC como SNV obtuvieron el valor medio de  $R^2$  más elevado con 0,39 para ambos pretratamientos. El rango espectral más identificado con el espesor de *P. nigra* con el algoritmo SVM-L fue, al igual que con PLS, el de 701 a 2450 nm. Para los modelos generados con RF fue SG el pretratamiento más eficaz prediciendo el espesor foliar, con un valor medio de  $R^2$  de 0,267. En cuanto al rango espectral, para RF fue el de 400 a 2450 nm el que obtuvo resultados más elevados de  $R^2$ , con 0,23 de media (Figura 6).

#### 4.1.2. LDMC

Para el LDMC de *P. nigra* se obtuvieron medias de  $R^2$  inferiores a las del resto de rasgos para los tres algoritmos. En este caso, SVM-L es, con 0,22, el que mayor media de  $R^2$  presentó, seguido de cerca por PLS (0,21). RF fue el algoritmo que menor media de  $R^2$  obtuvo con 0,11, siendo la más baja para todos los rasgos funcionales estudiados (Figura 6). En los modelos generados con PLS el pretratamiento con mayor media de  $R^2$  fueron Dt y SG con 0,25. El rango espectral más asociado con el LDMC de *P. nigra* en los modelos de PLS fue el de 1001 a 1800 nm con un valor medio de  $R^2$  de 0,28. En los modelos de SVM-L, el pretratamiento con mayor media de  $R^2$  fue MCS con un valor de 0,30. Para el rango espectral con el algoritmo SVM-L, al igual que en los modelos de PLS, fue el de 1001 a 1800 nm el que mayor media de  $R^2$  obtuvo. En cambio, con el algoritmo RF, fue MA+D1 el pretratamiento con mayor media de  $R^2$  con 0,21. En cuanto al rango espectral, tanto de 701 a 2450 nm como de 400 a 2450 nm se obtuvieron los mayores valores promedio de  $R^2$  (0,16 y 0,15 respectivamente) usando el algoritmo RF (Figura 6).

#### 4.1.3. SLA

En cuanto al SLA, al igual que con el espesor, tanto PLS como SVM-L tuvieron un valor promedio de  $R^2$  muy similar, con 0,34 en PLS y 0,33 en SVM-L. También RF fue el algoritmo que menor media de  $R^2$  tuvo (0,17), con bastante diferencia con los otros algoritmos probados (Figura 6). En los modelos de PLS, el pretratamiento con mayor media de  $R^2$  fue MSC, con un valor de 0,40. En relación con el rango espectral más asociado con el SLA en *P. nigra*, para los modelos de PLS fue de 1001 a 1800 nm con un valor promedio de  $R^2$  de 0,43. Para los modelos generados con SVM-L, el pretratamiento que obtuvo mayor valor promedio de  $R^2$  fue SNV con 0,43. Los rangos espectrales con mayor media de  $R^2$  usando el algoritmo SVM-L fueron de 400 a 2450 nm y de 701 a 2450 nm con 0,40 en ambos casos. Por último, para los modelos de RF, los pretratamientos más eficaces para predecir el SLA en *P. nigra* fueron MVA+D1 y MA+SG con valores medios de  $R^2$  de 0,25, mientras en cuanto al rango espectral, fue el de 400 a 2450 nm el que mayor valor medio de  $R^2$  obtuvo con 0,25 (Figura 6).

#### 4.2. *Pinus sylvestris*

En el caso del *P. sylvestris*, únicamente se observa un patrón en los pretratamientos en los modelos de RF, que coincide con el de *P. nigra*, donde los datos brutos y los pretratamientos SG y MA+SG fueron los que mayores valores promedio de  $R^2$  reflejaron, mientras que fueron MVA y MC los que menor capacidad de predicción tuvieron para todos los rasgos funcionales (Figura 7). En el caso de los rangos espectrales, fue el rango de 400 a 700 nm el que tuvo peor capacidad de predicción de todos los rasgos funcionales, mientras que el rango de 1001 a 1800 nm fue el que mayores valores promedio de  $R^2$  obtuvo.

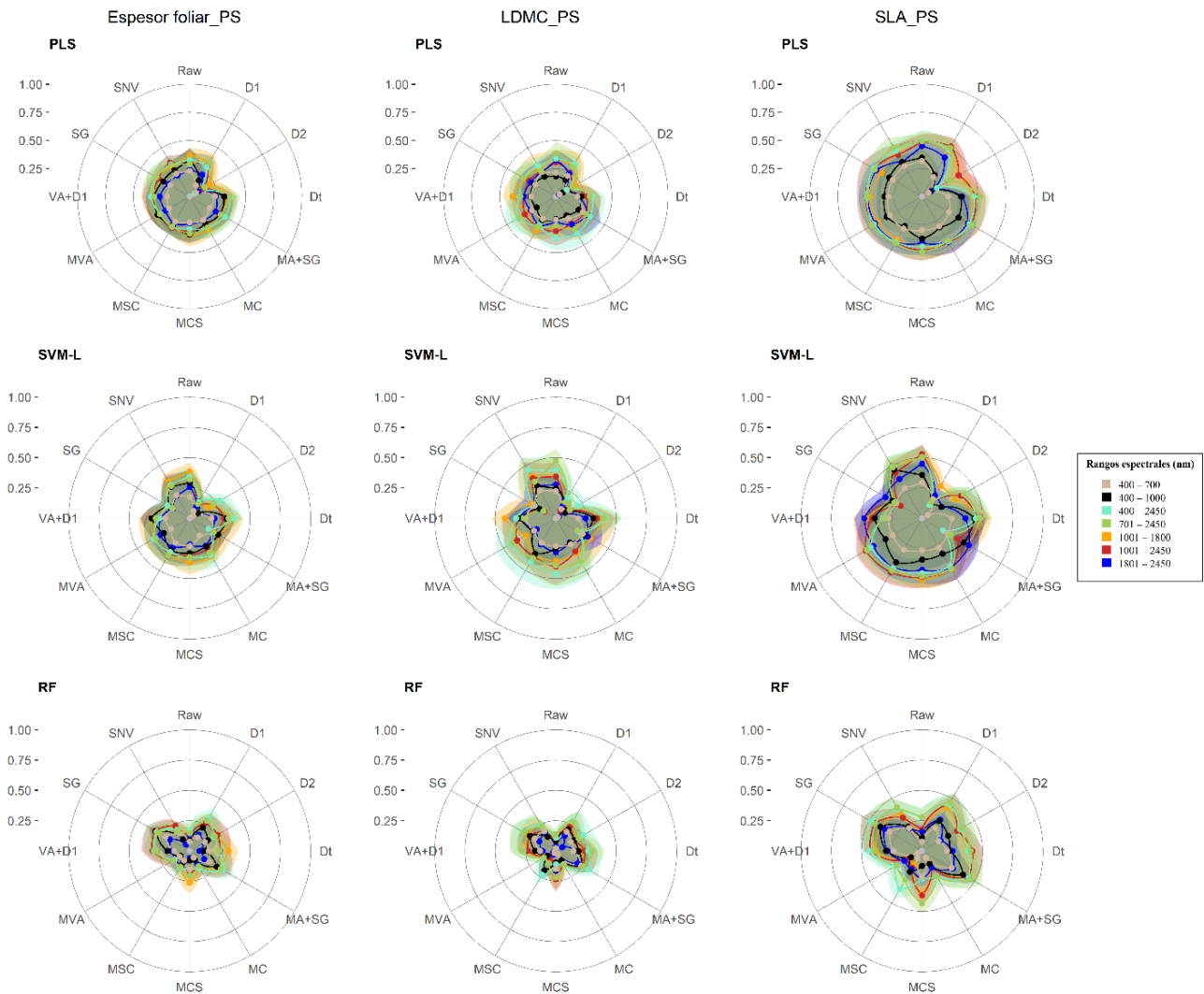


Figura 7. Gráficos polares de los resultados de los modelos predictivos generados con los algoritmos PLS, SVM-L y RF para los rasgos funcionales de *P. sylvestris*. Se representa el valor de  $R^2$  obtenido por rasgo funcional para cada algoritmo, pretratamiento y rango espectral.

#### 4.2.1. Espesor foliar

Para el espesor foliar en *P. sylvestris*, el algoritmo con mayor valor promedio de  $R^2$  resultó ser PLS con 0,28. El segundo algoritmo con mayor media de  $R^2$  fue SVM con 0,25 y el de menor valor promedio de  $R^2$  obtuvo fue RF con 0,19 (Figura 7). En los modelos de PLS, el pretratamiento más eficaz prediciendo el espesor foliar de *P. sylvestris* fue Dt, con un valor promedio de  $R^2$  de 0,31 (aunque el resto de pretratamientos, exceptuando las derivadas, tuvieron valores similares, en torno al 0,29) mientras que, en cuanto al rango espectral, fue el de 1001 a 1800 nm el que mayor valor medio de  $R^2$  obtuvo con 0,32. Con el algoritmo SVM-

L, los datos brutos tuvieron, con 0,30, mayor media de  $R^2$  que los pretratamientos, aunque los pretratamientos MC, MCS, MVA y SNV obtuvieron valores similares (en torno a 0,29). El rango espectral con mayor valor medio de  $R^2$  fue el de 1001 a 1800 nm con 0,30 para los modelos de SVM-L. En los modelos de RF, fue SG el pretratamiento con mayor valor promedio de  $R^2$  con 0,28, mientras que, usando este mismo algoritmo, fue el rango espectral de 1001 a 1800 nm el que obtuvo mayor media de  $R^2$ , con 0,23 al igual que en modelos de PLS y SVM-L.

#### 4.2.2. LDMC

En cuanto al LDMC, SVM-L y PLS obtuvieron valores promedio de  $R^2$  similares, en torno a 0,27. De nuevo, RF resultó el algoritmo que menor media de  $R^2$  obtuvo con 0,17 (Figura 7). En los modelos generados con PLS, el pretratamiento más eficaz prediciendo el LDMC de *P. sylvestris* fue MA+SG con un valor promedio de  $R^2$  de 0,31 mientras que el rango espectral con mayor media de  $R^2$  utilizando este algoritmo fue el de 400 a 2450 nm, con un valor de 0,32. En los modelos de SVM-L, fue SNV el pretratamiento con mayor media de  $R^2$ , con un valor de 0,35. El rango espectral más eficaz prediciendo el LDMC en esta especie fue, al igual que con PLS, el de 400 a 2450 nm con una media de  $R^2$  de 0,35 con este algoritmo. Con RF, el pretratamiento con mayor media de  $R^2$ , con un valor de 0,25, fue MA+SG. En cuanto al rango espectral usando los modelos de RF, al igual que los modelos generados con los algoritmos anteriores, fue el de 400 a 2450 nm el que mayor media de  $R^2$  obtuvo, con 0,22.

#### 4.2.3. SLA

El SLA en *P. sylvestris* fue el rasgo funcional que, a nivel general, mayores valores de  $R^2$  obtuvo para los tres algoritmos. PLS fue el algoritmo con mayor media de  $R^2$  con un valor de 0,41, seguido de SVM-L con 0,37. RF obtuvo, con 0,28 de media, valores de  $R^2$  menores que los otros dos algoritmos (Figura 7). En los modelos de PLS el pretratamiento más eficaz en predecir el SLA de *P. sylvestris* fue MVA, con un valor promedio de  $R^2$  de 0,45. Aunque los datos brutos y el pretratamiento de MA+SG obtuvieron resultados similares, ambos con 0,44. En cuanto al rango espectral con este mismo algoritmo, fue el de 701 a 2450 nm el que mayor media de  $R^2$  obtuvo con un valor de 0,48 (seguido de cerca por el de 1001 a 2450 nm, con 0,46). Para los modelos de SVM-L, fue MSC el pretratamiento con mayor media de  $R^2$ , con

0,45. El rango espectral más asociado al SLA con este algoritmo fue el de 1001 a 1800 nm con un valor promedio de  $R^2$  de 0,43. Para finalizar, en los modelos de RF, el pretratamiento con mayor media de  $R^2$  fue SG con un valor de 0,41 (aunque MA+SG obtuvo un valor similar, de 0,40). En cuanto al rango espectral, fue el de 701 a 2450 nm el que mayor media de  $R^2$  obtuvo con 0,37 con los modelos de RF.

## 5. DISCUSIÓN

Los modelos construidos para tratar de predecir los rasgos funcionales de las hojas de las dos especies de pino a partir de sus espectros de absorción lumínica en el rango Vis-NIR no fueron eficaces en ninguna de las dos especies estudiadas. Aunque se obtuvieron unos valores máximos de  $R^2$  aceptables en algunos casos (0,6 – 0,75), pudiendo indicar una capacidad de predicción moderada, hay que ser muy cauteloso a la hora de tener en cuenta estos valores máximos porque no siempre son indicativos ya que pueden estar relacionados con datos atípicos, por lo cual, se decidió tomar como referencia los valores promedios de  $R^2$  aunque estos indicaran una capacidad de predicción baja o muy baja de los modelos construidos. Estos resultados contrastan con la mayoría de estudios de predicción de rasgos funcionales con tecnología Vis-NIR revisados, donde los modelos tienen una capacidad de predicción moderada o alta (Gutiérrez *et al.* 2016; Streher *et al.* 2020). Es importante destacar que, en la revisión bibliográfica, prácticamente no se encontró ningún estudio en este ámbito que obtuviera como resultado valores de  $R^2$  inferiores a 0,53 en el peor de los casos (Costa *et al.* 2018). Esto no implica necesariamente que no se hayan realizado estos estudios, sino que podría deberse, más bien, a que los estudios con resultados negativos tienen menos posibilidades de ser publicados, lo que se conoce como sesgo de publicación (Easterbrook *et al.* 1991). De hecho, Matosin y colaboradores (2014) afirmaron que el 90% de las publicaciones en revistas científicas se correspondían con resultados positivos. En el caso de este estudio, la baja o muy baja capacidad de predicción de los modelos construidos pudo deberse a la presencia de datos atípicos en los rangos funcionales (Figura 5), probablemente debidos a errores en la toma de datos en el laboratorio.

Por otro lado, al estudiar rasgos foliares morfológicos, hay que tener en cuenta que estos rasgos tienden a predecirse peor que los rasgos químicos según los resultados de varios estudios

(Ecartot *et al.* 2013; Ge *et al.* 2019), probablemente debido a que los rasgos químicos, como los pigmentos, influyen en la absorbancia de regiones espectrales específicas y son más sencillos de medir que las propiedades estructurales como el SLA o el LDMC. Además, en estudios donde se predicen rasgos foliares y rasgos de la madera, los modelos de rasgos foliares obtienen peores resultados que los construidos para los rasgos de la madera y la corteza (Costa *et al.* 2018). En su estudio, Costa y colaboradores exponen que estos peores resultados en rasgos foliares pueden deberse, en primer lugar, a sobreestimaciones de los valores más altos en estos rasgos y, en segundo lugar, a la no homogeneidad de la estructura de las hojas, lo que podría aumentar la varianza entre lecturas de la misma planta en el espectrofotómetro. En el presente estudio, este último motivo parece improbable, ya que para evitar errores de este tipo se tomaron 3 submuestras diferentes de cada individuo y se calculó el espectro promedio de haz y envés.

En cuanto a las especies estudiadas, aunque en los últimos años está habiendo un avance en este sentido, la mayoría de las publicaciones que estudian el fenotipado mediante modelos predictivos con espectroscopía Vis-NIR se centran en especies herbáceas cultivables, siendo menos comunes los estudios en plantas leñosas de porte arbóreo, por lo que las comparaciones con los resultados en especies leñosas deben ser cuidadosas. En este sentido, de las dos especies de estudio, *P. sylvestris* ha sido más estudiada para la construcción de modelos predictivos de sus rasgos funcionales, tanto foliares como de la madera (Petisco *et al.* 2006; Fernades *et al.* 2017; Escudero *et al.* 2021) mientras que los estudios centrados en *P. nigra* son más escasos y están dirigidos principalmente a los rasgos de la madera (Dias *et al.* 2020).

Por otro lado, los resultados obtenidos parecen indicar tendencias similares en los valores de  $R^2$  para los distintos rasgos funcionales en los diferentes modelos generados (Figuras 6 y 7). Esto podría deberse a que los datos brutos de los rasgos medidos en las hojas están correlacionados entre sí, como se ha descrito en Cornelissen y colaboradores (2003). Para comprobarlo, se realizó un test de Pearson para obtener el coeficiente de correlación entre rasgos y, efectivamente, el resultado en *P. nigra* fue una correlación negativa significativa entre SLA y espesor foliar ( $r = -0,63$ ,  $P < 0,001$ ) y entre SLA y LDMC ( $r = -0,54$ ,  $p < 0,001$ ), al igual que en *P. sylvestris*, donde se obtuvo una correlación negativa significativa entre SLA y espesor foliar ( $r = -0,45$ ,  $P < 0,001$ ) y entre SLA y LDMC ( $r = -0,42$ ,  $P < 0,001$ ).

## 5.1. Comparación entre algoritmos

Las comparaciones en este estudio entre los tres algoritmos probados para la construcción y validación de los modelos tuvieron como resultado que PLS y SVM-L fueron los que mayor capacidad de predicción tuvieron (siendo con PLS ligeramente mayor que con SVM-L) mientras que RF fue el que arrojó valores promedio de  $R^2$  menores para todos los rasgos de las dos especies de estudio. En este sentido, este Trabajo de Fin de Máster es novedoso, ya que compara entre tres algoritmos diferentes, mientras que en la mayoría de los estudios únicamente se utiliza un algoritmo para construir los modelos, siendo el más común la regresión PLS (Costa et al. 2018; Streher et al. 2020; Kothari et al. 2021). De los pocos estudios que comparan la capacidad de predicción con dos algoritmos diferentes, Ge y colaboradores (2019) obtuvieron valores bastante similares de  $R^2$  en los modelos de PLS y SVM al predecir rasgos foliares morfológicos (aunque en este caso, con resultados ligeramente superiores en SVM). El algoritmo *Random Forest*, en cambio, prácticamente no ha sido utilizado en este ámbito, por lo que sería necesario el desarrollo de más estudios para poder realizar análisis comparativos.

## 5.2. Comparación entre pretratamientos quimiométricos

A nivel general y salvo excepciones, los pretratamientos quimiométricos obtuvieron valores de  $R^2$  similares entre sí y con los datos brutos. Aun así, destacan los siguientes: *Multiple scatter correction* (MSC), *Standard Normal Variate* (SNV), *Moving averages* (MVA) y *Detrend* (Dt). En este sentido, las comparaciones con otros estudios no son posibles, ya que, de nuevo, se trata de una novedad en este tipo de fenotipados con espectrometría Vis-NIR. Si bien es cierto que varios estudios suelen utilizar algún pretratamiento para modificar los espectros obtenidos, siendo los más comunes la primera y segunda derivada, Savitzky-Golay, o *Multiple scatter correction* (Ecanot et al. 2013; Serbin et al. 2014; Kothani et al. 2019), prácticamente ninguno compara la eficacia de los pretratamientos para construir modelos predictivos, mucho menos utilizando 11 pretratamientos, a excepción de Escudero y colaboradores (2021), cuyo estudio fue realizado por el mismo grupo de investigación y englobado en el mismo proyecto que este TFM.



### 5.3. Comparación entre rangos espectrales

En cuanto a la comparación entre los rangos del espectro Vis-NIR más asociados con los rasgos funcionales de cada pino, hay que destacar que, aunque se encuentran en la bibliografía de manera más común que las comparaciones entre algoritmos y entre pretratamientos, no se han encontrado estudios que prueben tantos rangos de longitud de onda diferentes. En este sentido, los modelos construidos indicaron, mayoritariamente, una mejor capacidad de predicción del espesor foliar de *P. nigra* en el rango espectral de 701 a 2450 nm. A pesar de ser un rango bastante amplio que abarca casi todo el espectro NIR, no coincide con los resultados obtenidos por Costa y colaboradores (2018), que indicaron en su estudio que el espesor y la densidad de las hojas están asociadas con longitudes de onda superiores a 2400 nm. El resultado del presente estudio, por el contrario, coincide más con los obtenidos en otra publicación, donde el grosor de la cutícula y el mesófilo, características asociadas al espesor de la hoja, se relacionan positivamente con la reflectancia NIR en 800 nm (Slaton *et al.* 2001). En el caso del espesor foliar de *P. sylvestris*, los resultados obtenidos no coinciden con ninguno de estos estudios, ya que los valores promedio más altos de  $R^2$  se alcanzan en el rango espectral de 1001 a 1800 nm en todos los modelos probados.

Los modelos de PLS y SVM probados para el LDMC de *P. nigra* tuvieron mayor capacidad predictiva en el rango de 1001 a 2450 nm, coincidiendo, en este caso, con Curran (1989) quien determinó que las regiones espectrales SWIR (1000 – 2500 nm) contienen características de absorción relacionadas con rasgos estructurales de las hojas, como el LDMC o el contenido interno de agua de la hoja. En cambio, con *P. sylvestris* se obtuvieron las mejores predicciones de LDMC en el rango espectral completo (400 – 2450 nm) en todos los modelos estudiados, lo que se consideró un rango demasiado amplio como para extraer información concluyente.

En cuanto al SLA, Costa y colaboradores (2018) asocian este rasgo funcional con un pico del espectro alrededor de 2180 nm, relacionándolo con la absorción de proteínas. Comparando estos resultados con los del presente estudio, se podrían asemejar a los obtenidos en los modelos de PLS para el SLA de *P. sylvestris*, aunque el rango espectral con mayor capacidad de predicción en este caso fue bastante amplio (700 – 2450 nm). En cambio, para *P. nigra*, esta afirmación no se podría aplicar al asociarse con el rango de 1001 a 1800 nm.

Por otro lado, en relación con el tipo de planta analizado, varios estudios coinciden en asociar las acículas de los árboles perennes, con estructuras más densas y adaptadas a entornos con pocos recursos, con una reflectancia en el espectro NIR más baja, en contraposición con las especies planifolias, que están asociadas en mayor medida al rango SWIR (1000 – 2500 nm) (Ollinger, 2011; Serbin *et al.* 2014). Esta afirmación no es sencilla de aplicar a los resultados de este estudio, ya que los modelos predictivos reflejaron resultados variables en cuanto al rango espectral, muy dependientes del rasgo funcional y la especie estudiados.

Para finalizar, siguiendo el hilo de la asociación entre rangos espectrales del Vis-NIR y los rasgos funcionales foliares de las plantas, Serbin y colaboradores (2019) descubrieron que las relaciones entre rangos espectrales y rasgos funcionales, en este caso, la masa foliar por área (LMA), son asimilables entre biomas a pesar de las variaciones de LMA entre los propios biomas (variabilidad funcional). Esta afirmación implica que las técnicas basadas en la espectroscopía Vis-NIR pueden ser de gran ayuda para comprender la variabilidad funcional de las plantas provocada por la variabilidad climática y, por tanto, también para prever la posible variabilidad de las plantas frente al cambio global (Shiklomanov *et al.* 2019).

## 6. CONCLUSIONES

Respecto al objetivo general de este Trabajo de Fin de Máster, con los resultados obtenidos no es posible concluir que los modelos basados en la espectroscopía Vis-NIR sean efectivos para predecir el espesor foliar, el SLA ni el LDMC de *P. nigra* ni *P. sylvestris*, ya que los modelos construidos mostraron una baja capacidad de predicción.

En cuanto a los objetivos específicos, en relación con el algoritmo que mejor predice cada rasgo para cada especie se puede concluir que, para *P. nigra*, la regresión de mínimos cuadrados parciales (PLS) fue el algoritmo que mejor predijo el espesor foliar y el SLA, mientras que *Support Vector Machine* (SVM-L) fue el más eficaz prediciendo el LDMC. Para *P. sylvestris*, el PLS fue el algoritmo con mejor capacidad de predicción para todos los rasgos funcionales foliares estudiados, aunque en el caso del LDMC, los modelos construidos con el algoritmo SVM-L obtuvieron una capacidad de predicción igual a la de los modelos de PLS.

Respecto al pretratamiento quimiométrico de los espectros más eficaz prediciendo los rasgos funcionales de cada pino, se pueden concluir que, para *P. nigra*, el pretratamiento que resultó más eficaz en la predicción del espesor foliar y el LDMC fue el *Multiple scatter correction* (MSC), mientras que en los modelos predictivos del SLA, el pretratamiento más eficaz fue el *Standard Normal Variate* (SNV). Para *P. sylvestris*, el pretratamiento que mejor predijo el espesor foliar fue *Detrend* (Dt), SNV para el LDMC y *Moving averages* (MVA) para el SLA.

En relación con los rangos espectrales Vis-NIR más asociados con cada rasgo funcional en cada especie, para *P. nigra*, los modelos generados para el espesor foliar obtuvieron su mayor capacidad de predicción en el rango espectral de 701 a 2450 nm, mientras que tanto el LDMC como el SLA se asociaron, en mayor medida con el rango de 101 a 1800 nm. Para *P. sylvestris*, en cambio, el espesor foliar fue predicho en mayor medida en el rango de 1001 a 1800 nm, el LDMC en el rango completo (400 a 2450 nm) y el SLA, en el rango de 701 a 2450 nm.

## 7. PROYECCIONES FUTURAS

Al tratarse de un estudio, por un lado, novedoso y, por otro lado, con resultados negativos, las posibles proyecciones futuras son muy numerosas. En primer lugar, como proyección futura más inmediata, sería interesante intentar aumentar la capacidad de predicción de los modelos, por ejemplo, mediante la eliminación o el tratamiento de los datos atípicos u *outliers*. También podría ser interesante construir modelos probando la eficacia de otros algoritmos como la regresión múltiple lineal (MLR) que han obtenido buenos resultados en otros estudios. Además, podría ser interesante comparar rangos espectrales más reducidos.

Por otro lado, está la posibilidad de la aplicación de los modelos generados para realizar un fenotipado masivo de todos los individuos de *P. nigra* y *P. sylvestris* de la parcela muestreada. De hecho, dentro de los trabajos de campo llevados a cabo para obtener las muestras necesarias para este estudio, también se obtuvieron muestras de ramas del resto de individuos (1045 en total). De estas muestras foliares se obtuvieron los espectros de absorción con el objetivo de aplicar los modelos predictivos construidos en este estudio. Es decir, mediante los modelos predictivos construidos con los datos de los rasgos funcionales y espectros obtenidos en el laboratorio a partir de los 202 individuos seleccionados, el paso siguiente sería predecir los rasgos funcionales de los 1045 individuos de la parcela a partir de sus espectros de absorbancia.

Además, como parte del mismo proyecto, se ha planteado la construcción de modelos para predecir otros rasgos funcionales diferentes a los de este estudio, tanto foliares (p. ej. contenido en nutrientes y pigmentos) como de la madera (p. ej. densidad de la madera y grosor de la corteza) y relacionados con la talla de la planta (p. ej. altura del árbol y DBH). Sería interesante comprobar si para estos rasgos la capacidad predictiva de los modelos es mejor.

Para finalizar, otra posible proyección futura podría dirigirse a la posibilidad de estudiar las interacciones entre las dos especies de pinos coexistentes, aprovechando la georreferenciación realizada durante los trabajos de muestreo para analizar el componente espacial.

## 8. AGRADECIMIENTOS

Este trabajo ha sido financiado por la Agencia Estatal de Investigación a través del proyecto Phenotypes (PGC2018-099115-B-I00).

En primer lugar, quiero agradecer a Adrián Escudero la oportunidad de participar en este proyecto. Después, agradecer a Ana Isabel García-Cervigón y David S. Pescador su disponibilidad en todo momento, así como su infinita paciencia para aclarar mis dudas, además de su participación en el diseño del experimento, muestreo y, especialmente, en el análisis estadístico. Dar las gracias también a Sergio Núñez, Manuel Rojo y Carlos Díaz por la confianza y hacer más amenas las jornadas de campo y laboratorio, al personal técnico e investigador del Área de Biodiversidad y Conservación de la URJC por acogerme como uno más durante los trabajos de laboratorio, a la buena gente de Zaorejas (Guadalajara) por acogernos y dejarnos ensuciarles el hotel con nuestras ramas de pinos.

Como no, agradecer el apoyo a mis compadres y comadres, los que son y siempre están ahí.

Por último, sobre todo, quiero dar las gracias a mi padre, mi madre y mi hermano, por su apoyo incondicional y por dejarse el alma y los huesos para que haya podido llegar hasta aquí, sin vosotros no habría sido posible.

## 9. BIBLIOGRAFÍA

- Abdi, H. (2003). Partial least square regression (PLS regression). *Encyclopedia for research methods for the social sciences*, 6(4), 792-795.
- Agencia Estatal de Meteorología. (2022). Valores climatológicos normales. Molina de Aragón. Agencia Estatal de Meteorología - AEMET. Gobierno de España. Revisado 23/10 de <https://www.aemet.es/es/serviciosclimaticos/datosclimatologicos/valoresclimatologicos?l=3013&k=clm>
- Albert, C. H., Grassein, F., Schurr, F. M., Vieilledent, G., & Violle, C. (2011). When and how should intraspecific variability be considered in trait-based plant ecology?. *Perspectives in Plant Ecology, Evolution and Systematics*, 13(3), 217-225.
- Alejano, R., & Montes, E. M. (1996). Distribución de *Pinus nigra* Arn. Subsp. *salzmanii* en las Sierras Béticas. *Ecología*, (10), 231-242
- Benavides, R., Carvalho, B., Matesanz, S., Bastias, C. C., Cavers, S., Escudero, A., ... & Valladares, F. (2021). Phenotypes of *Pinus sylvestris* are more coordinated under local harsher conditions across Europe. *Journal of Ecology*.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Carmona, C. P., Rota, C., Azcárate, F. M., & Peco, B. (2015). More for less: sampling strategies of plant functional traits across local environmental gradients. *Functional Ecology*, 29(4), 579-588
- Castroviejo, S. (2020). Flora ibérica: plantas vasculares de la Península Ibérica e Islas Baleares. *Flora ibérica*, 1-784.
- Cogollo Calderón, A. M., Velasco Linares, P., & Manosalva, L. (2020). Caracterización funcional de plantas y su utilidad en la selección de especies para la restauración ecológica de ecosistemas altoandinos. *Biota colombiana*, 21(1), 1-15

- Cornelissen, J. H. C., Lavorel, S., Garnier, E., Díaz, S., Buchmann, N., Gurvich, D. E., ... & Poorter, H. (2003). A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. *Australian journal of Botany*, 51(4), 335-380.
- Costa, F. R., Lang, C., Almeida, D. R., Castilho, C. V., & Poorter, L. (2018). Near-infrared spectrometry allows fast and extensive predictions of functional traits from dry leaves and branches. *Ecological Applications*, 28(5), 1157-1167.
- Curran, P. J. (1989). Remote sensing of foliar chemistry. *Remote sensing of environment*, 30(3), 271-278.
- Dias, A., Carvalho, A., Silva, M. E., Lima-Brito, J., Gaspar, M. J., Alves, A., ... & Lousada, J. L. (2020). Physical, chemical and mechanical wood properties of *Pinus nigra* growing in Portugal. *Annals of Forest Science*, 77(3), 1-11.
- Dirección General de Política Forestal y Espacios Naturales, Viceconsejería de Medio Ambiente, Consejería de Agricultura, Medio Ambiente y Desarrollo Rural, & Junta de Comunidades de Castilla-La Mancha. (2017, septiembre). Plan de Gestión de ALTO TAJO (Guadalajara y Cuenca), ES4240016/ES0000092.
- Easterbrook, P. J., Gopalan, R., Berlin, J. A., & Matthews, D. R. (1991). Publication bias in clinical research. *The Lancet*, 337(8746), 867-872.
- Ecartot, M., Bączyk, P., Tessarotto, L., & Chervin, C. (2013). Rapid phenotyping of the tomato fruit model, Micro-Tom, with a portable VIS–NIR spectrometer. *Plant physiology and biochemistry*, 70, 159-163.
- Escudero, A., Matesanz, S., Pescador, D. S., de la Cruz, M., Valladares, F., & Cavieres, L. A. (2021). Every bit helps: The functional role of individuals in assembling any plant community, from the richest to monospecific ones. *Journal of Vegetation Science*, 32(5), e13059.
- Fernandes, C., Gaspar, M. J., Pires, J., Alves, A., Simões, R., Rodrigues, J. C., ... & Lousada, J. L. (2017). Physical, chemical and mechanical properties of *Pinus sylvestris* wood at five sites in Portugal. *IForest*.

- Ferrero, L. M., Montouto, O., & Herranz, J. M. (2006). Flora amenazada y de interés del Parque Natural del Alto Tajo. Junta de comunidades de Castilla-la-Mancha, Toledo, Spain.)
- Fiorani, F., & Schurr, U. (2013). Future scenarios for plant phenotyping. *Annual review of plant biology*, 64, 267-291.
- García, F. M., & González, M. Á. C. (1999). Los bosques de " *Pinus sylvestris*" L. del Sistema Central español: distribución, historia, composición florística y tipología. Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria.
- Gaspar, M. J., Alves, A., Louzada, J. L., Morais, J., Santos, A., Fernandes, C., ... & Rodrigues, J. C. (2011). Genetic variation of chemical and mechanical traits of maritime pine (*Pinus pinaster* Aiton). Correlations with wood density components. *Annals of Forest Science*, 68(2), 255-265.
- Ge, Y., Atefi, A., Zhang, H., Miao, C., Ramamurthy, R. K., Sigmon, B., ... & Schnable, J. C. (2019). High-throughput analysis of leaf physiological and chemical traits with VIS–NIR–SWIR spectroscopy: a case study with a maize diversity panel. *Plant methods*, 15(1), 1-12.
- Gutiérrez, S., Tardaguila, J., Fernández-Navales, J., & Diago, M. P. (2016). Data mining and NIR spectroscopy in viticulture: applications for plant phenotyping under field conditions. *Sensors*, 16(2), 236.
- Karatzoglou, A., Smola, A., Hornik, K., Zeileis, A. (2004). kernlab - An S4 Package for Kernel Methods in R. *Journal of Statistical Software* 11(9), 1-20. URL <http://www.jstatsoft.org/v11/i09/>
- Katabuchi, M. (2015). LeafArea: an R package for rapid digital image analysis of leaf area. *Ecological Reserach* 30:1073-1077.
- Kothari, S., Beauchamp-Rioux, R., Laliberté, E., & Cavender-Bares, J. (2021). Reflectance spectroscopy allows rapid, accurate, and non-destructive estimates of functional traits from pressed leaves. *bioRxiv*.



- Kucheryavskiy S. (2020). "mdatools - R package for chemometrics." *\_Chemometrics and Intelligent Laboratory Systems\_*, \*198\*. <URL: <https://doi.org/10.1016/j.chemolab.2020.103937>>.
- Kuhn, M. (2021). caret: Classification and Regression Training. R package version 6.0-90. <https://CRAN.R-project.org/package=caret>
- Liland, K. H., Bjørn-Helge Mevik, B., and Wehrens, R. (2021). pls: Partial Least Squares and Principal Component Regression. R package version 2.8-0. <https://CRAN.R-project.org/package=pls>
- MacArthur, R., & Levins, R. (1967). The limiting similarity, convergence, and divergence of coexisting species. *The american naturalist*, 101(921), 377-385.
- Matosin, N., Frank, E., Engel, M., Lum, J. S., & Newell, K. A. (2014). Negativity towards negative results: a discussion of the disconnect between scientific worth and scientific culture. *Disease models & mechanisms*, 7(2), 171-173.
- Ollinger, S. V. (2011). Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytologist*, 189(2), 375-394.
- Pescador, D. S., de Bello, F., Valladares, F., & Escudero, A. (2015). Plant trait variation along an altitudinal gradient in mediterranean high mountain grasslands: controlling the species turnover effect. *PLoS One*, 10(3), e0118876.
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rasband, W. S. (1997). ImageJ. Bethesda, Maryland, USA: US National Institutes of Health.
- Reich, P. B., Walters, M. B., & Ellsworth, D. S. (1997). From tropics to tundra: global convergence in plant functioning. *Proceedings of the National Academy of Sciences*, 94(25), 13730-13734.
- Roudier, P (2017). asdreader: Reading ASD Binary Files in R. R package version 0.1-3. <https://CRAN.R-project.org/package=asdreader>

- Salgado - Negret, B. E. (Ed.). (2016). *La ecología funcional como aproximación al estudio, manejo y conservación de la biodiversidad: protocolos y aplicaciones*. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt.
- Sánchez, T., Ceballos, H., Dufour, D., Ortiz, D., Morante, N., Calle, F., ... & Davrieux, F. (2014). Prediction of carotenoids, cyanide and dry matter contents in fresh cassava root using NIRS and Hunter color techniques. *Food chemistry*, 151, 444-451.
- Serbin, S. P., Singh, A., McNeil, B. E., Kingdon, C. C., & Townsend, P. A. (2014). Spectroscopic determination of leaf morphological and biochemical traits for northern temperate and boreal tree species. *Ecological Applications*, 24(7), 1651-1669.
- Serbin, S. P., Wu, J., Ely, K. S., Kruger, E. L., Townsend, P. A., Meng, R., ... & Rogers, A. (2019). From the Arctic to the tropics: multibiome prediction of leaf mass per area using leaf reflectance. *New Phytologist*, 224(4), 1557-1568.
- Siefert, A., Violle, C., Chalmandrier, L., Albert, C. H., Taudiere, A., Fajardo, A., ... & Wardle, D. A. (2015). A global meta-analysis of the relative extent of intraspecific trait variation in plant communities. *Ecology letters*, 18(12), 1406-1419.
- Shiklomanov, A. N., Bradley, B. A., Dahlin, K. M., M Fox, A., Gough, C. M., Hoffman, F. M., ... & Smith, W. K. (2019). Enhancing global change experiments through integration of remote-sensing techniques. *Frontiers in Ecology and the Environment*, 17(4), 215-224.
- Shipley, B., De Bello, F., Cornelissen, J. H. C., Laliberté, E., Laughlin, D. C., & Reich, P. B. (2016). Reinforcing loose foundation stones in trait-based plant ecology. *Oecologia*, 180(4), 923-931.
- Slaton, M. R., Raymond Hunt Jr, E., & Smith, W. K. (2001). Estimating near-infrared leaf reflectance from leaf structural characteristics. *American journal of botany*, 88(2), 278-284.
- Stevens, A., & Ramirez-Lopez, L. (2021). An introduction to the prospectr package. R package Vignette R package version 0.2.2.

- Streher, A. S., da Silva Torres, R., Morellato, L. P. C., & Silva, T. S. F. (2020). Accuracy and limitations for spectroscopic prediction of leaf traits in seasonally dry tropical environments. *Remote Sensing of Environment*, 244, 111828.
- Vapnik Vladimir, N. (1999). *The Nature of Statistical Learning Theory (Information Science and Statistics)*.
- Violle, C., Navas, M. L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., & Garnier, E. (2007). Let the concept of trait be functional!. *Oikos*, 116(5), 882-892.
- Violle, C., Enquist, B. J., McGill, B. J., Jiang, L. I. N., Albert, C. H., Hulshof, C., ... & Messier, J. (2012). The return of the variance: intraspecific variability in community ecology. *Trends in ecology & evolution*, 27(4), 244-252.
- Westoby, M., & Wright, I. J. (2006). Land-plant ecology on the basis of functional traits. *Trends in ecology & evolution*, 21(5), 261-268.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- White, J. W., Andrade-Sanchez, P., Gore, M. A., Bronson, K. F., Coffelt, T. A., Conley, M. M., ... & Wang, G. (2012). Field-based phenomics for plant genetics research. *Field Crops Research*, 133, 101-112.
- Wright, M. N., & Ziegler, A. (2017). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*, 77(1), 1–17. <https://doi.org/10.18637/jss.v077.i01>