



8º CONGRESO FORESTAL ESPAÑOL

La **Ciencia forestal** y su contribución a
los **Objetivos de Desarrollo Sostenible**



8CFE

Edita: Sociedad Española de Ciencias Forestales

Cataluña | Catalunya - 27 junio | juny - 1 julio | juliol 2022

ISBN 978-84-941695-6-4

© Sociedad Española de Ciencias Forestales

Organiza



Uso de la distancia de Mahalanobis para el cálculo de la similitud ecológica. Aplicación en poblaciones forestales

TRANQUE PASCUAL, F. J.¹, VALLEJO SANZ, F. J.¹, ALÍA MIRANDA, R.², ALONSO PONCE, R.³

¹ Dirección General de Patrimonio Natural y Política Forestal. Junta de Castilla y León.

² INIA (CSIC).

³ Förä Forest Technologies.

Resumen

Disponer de información sobre similitud ecológica entre poblaciones forestales permite hacer recomendaciones de uso del material a emplear para asegurar la adaptación y conservación, seleccionando las procedencias idóneas según el objetivo de la plantación. Existen diferentes métodos para el cálculo de esta proximidad ecológica, destacando sobre todos ellos los indicadores basados en distancias o índices de similitud.

Se propone una metodología de cálculo de similitud entre puntos mediante la distancia de Mahalanobis, determinando la proximidad climática entre los puntos de distribución de una determinada especie forestal y la zona de plantación. Análogamente, se podrían identificar las zonas de mayor similitud al origen natural de un determinado material seleccionado o mejorado, facilitando así su adecuada utilización según el objetivo previsto. Para el cálculo se emplean una serie de variables climáticas y fitoclimáticas, previamente seleccionadas, y que, con alta influencia en la distribución de la especie considerada, pueden utilizarse como vectores de un espacio multivariante. Así, mediante una sencilla aplicación en una hoja de cálculo se puede obtener la disimilitud entre vectores de acuerdo a la distancia de Mahalanobis, orientando las recomendaciones de uso de material forestal.

Palabras clave

Disimilitud, procedencia, homologación ecológica.

1. Introducción

El conocimiento de la proximidad ecológica entre poblaciones forestales permite orientar las recomendaciones de uso del material forestal de reproducción (MFR), seleccionando la procedencia idónea de una manera más eficiente y acorde a los fines deseados. Esta información es de suma importancia porque, junto con la adecuada selección de las especies y su adaptación a la estación, puede ser determinante para el objetivo perseguido, bien sea de conservación o productivo. La existencia de variación genética entre las poblaciones forestales en caracteres de crecimiento y la producción refuerza la importancia práctica de la elección de los materiales utilizados en las repoblaciones (VAN ANDEL, 1998), influyendo la procedencia de la semilla en muchas características relacionadas con la adaptación y rendimiento futuro de las plantas. Las recomendaciones de uso, por tanto, intentan establecer una relación entre zonas de suministro de los materiales forestales (regiones de procedencia según la normativa europea de comercialización de MFR) y las zonas de utilización (ver BUIJTENEN VAN, 1992, para una discusión más general).

Algunas comunidades autónomas han establecido recomendaciones más o menos estrictas en programas de forestación y normas de comercialización para el uso de los materiales forestales, indicando las especies y procedencias a emplear en las diferentes zonas de utilización, y con mayor o menor flexibilidad para emplear alternativas homologables. Aunque en general se suele aceptar que la semilla de fuentes locales producen planta adaptada a esas condiciones concretas (KAWECKI & EBERT, 2004), eso no tiene por qué traducirse en masas

más productivas, si ese es nuestro objetivo, o que sean las de mejor adaptación a las condiciones locales (NAMKOONG, 1969; SAVOLAINEN et al, 2007). Por este motivo, se debe recurrir a ensayos de campo (medio y largo plazo) y recomendaciones basadas en la similitud ecológica entre la fuente de la semilla y el lugar de utilización.

La medida de proximidad entre puntos o poblaciones puede determinarse por medio de diferentes métodos que analizan, conjuntamente, las diversas variables que las caracterizan, como los indicadores basados en la distancia o los índices de similitud. Los criterios de distancia cuantifican el grado de semejanza (o su opuesto, disimilitud) entre puntos en base a una serie de variables que caracterizarían cada zona concreta, considerándose como vectores en un espacio multivariante (DAGNELIE, 1975). Se determina lo próximo o alejados que se encuentran los vectores de acuerdo a la distancia entre ambos; cuanto mayor sea el valor de esta distancia, menor será la proximidad entre los vectores en el espacio multivariante que se compara. De esta manera se puede calcular la distancia, por ejemplo, entre una zona de plantación y el resto de puntos de distribución actual o potencial de una especie, disponiendo de una información que oriente las recomendaciones de uso del material a emplear y escogiendo las procedencias más adecuadas.

La marcada correlación entre variables abióticas, y muy especialmente entre variables climáticas, obliga a una selección previa para utilizar las que mayor importancia tienen en la distribución natural y/o potencial de la especie forestal, y que configura el espacio multivariante de cálculo. La disimilitud entre vectores se puede determinar mediante la distancia de Mahalanobis, que presenta las ventajas de ser invariante ante los cambios de escala de las variables y de tener en cuenta su correlación mediante ponderación con la matriz de covarianzas (MAHALANOBIS, 1936). Así pues, a partir de variables fitoclimáticas se puede calcular la disimilitud de las distintas zonas donde vegeta una determinada especie forestal con respecto al punto donde se plantea su utilización. Análogamente se podría determinar las zonas de mayor similitud al origen natural, o su lugar de ensayo, de un determinado material de base seleccionado (rodal selecto, por ejemplo), o mejorado.

2. Objetivo

El objetivo de este trabajo es presentar una metodología sencilla y de fácil aplicación para determinar la disimilitud ecológica entre puntos, usados como vectores de multivariables climáticas, que facilite las recomendaciones de uso del material forestal. Así, implementada en una hoja de cálculo, permite calcular la distancia entre el punto, o el conjunto de puntos, que determinan la estación climática del lugar de implantación forestal, con respecto a todas las localizaciones de la especie, pudiendo restringir puntos por autoctonía o incrementarlos por potencialidad de la especie.

3. Metodología

Datos climáticos

Para el cálculo de la distancia de similitud se han utilizado variables climáticas y fitoclimáticas (ALLUÉ, 1990) obtenidos del modelo de diagnosis fitoclimático para la España peninsular (GONZALO, 2010), y que se presentan en la Tabla 1.

Distribución natural y potencial de las especies forestales

La información de la distribución actual de las especies forestales consideradas en el presente trabajo para el cálculo de la disimilitud climática se tomó fundamentalmente del mapa forestal de España (MFE50) y el tercer Inventario Forestal Nacional (IFN3). Ciertas localizaciones puntuales de algunas especies, como *Prunus avium* o *Sorbus domestica*, son propias de los autores y basadas en observaciones de campo. En las especies del género *Pinus* spp. se disponía de información del carácter autóctono, o no, del material.

La distribución potencial utilizada de las especies fue elaborada por el INIA-CIFOR a partir de un modelo de aproximación de máxima entropía (PHILLIPS et al, 2006) mediante el programa MAXENT, y calculado según probabilidad logística de existencia de cada una de las especies de trabajo en función de una serie de parámetros climáticos de entrada y aplicado sobre la distribución natural. Se estableció como valor de adecuación aquel que fuera mayor o igual que el percentil 10 establecido por el modelo MAXENT (AUÑÓN y SÁNCHEZ DE RON, 2010; AUÑÓN et al, 2011).

Tabla 1. Variables climáticas y fitoclimáticas empleadas para el cálculo de disimilitud

Variable	Descripción
TM	Temperatura media (°C)
TMC	Temperatura media mensual más alta (°C)
TMAXMCAL	Temperatura media de las máximas del mes más cálido (°C)
TMF	Temperatura media mensual más baja (°C)
TMINMFRI	Temperatura media de las mínimas del mes más frío (°C)
OSC	Oscilación térmica (TMC-TMF) (°C)
PREC_PRIM	Precipitación de primavera (mm)
PREC_VER	Precipitación de verano (mm)
PREC_OTO	Precipitación de otoño (mm)
PREC_INV	Precipitación de invierno (mm)
PREC_ANUAL	Precipitación total anual (mm)
PE	Precipitación mensual estival mínima (mm)
HSEG	Número de meses de helada segura
DIAS_5G	Número de días al año con más de 5°C
MESES_5G	Número de meses al año con más de 5°C
A	Duración de la aridez (nº meses 2Ti>Pi)
K	Intensidad de la aridez (As/Ah)

Mediante una aplicación GIS se generó una malla de cuadrículas de 1x1 km de lado para toda la península ibérica, asignando toda la información indicada al centro de la cuadrícula.

Selección de variables

Numerosas variables ecológicas pueden tener marcada influencia sobre la distribución de las especies y poblaciones forestales, siendo necesaria una preselección de las que se incluyen en los modelos para evitar la sobreparametrización y la redundancia de la información, especialmente si las variables están muy correlacionadas o presentan multicolinealidad.

En la metodología propuesta en este trabajo se utilizaron variables de tipo fitoclimático, que presentan una fuerte correlación en la mayoría de los casos, y que así se ha puesto de manifiesto en estudios de diagnosis fitoclimática, como es el caso de GONZALO (2010) donde el cálculo de las correlaciones de treinta factores fitoclimáticos obtenidos a partir de estaciones termopluviométricas de toda España peninsular mostró correlaciones significativas, así como un alto valor de la medida de adecuación muestral (KMO). En la Figura 1 se muestra gráficamente, y a modo de ejemplo, la matriz de correlación de las variables fitoclimáticas en el ámbito de distribución natural de *Pinus pinaster*, donde se puede apreciar la fuerte correlación significativa existente entre la mayoría de las variables. Estos resultados son similares en el resto de especies analizadas.

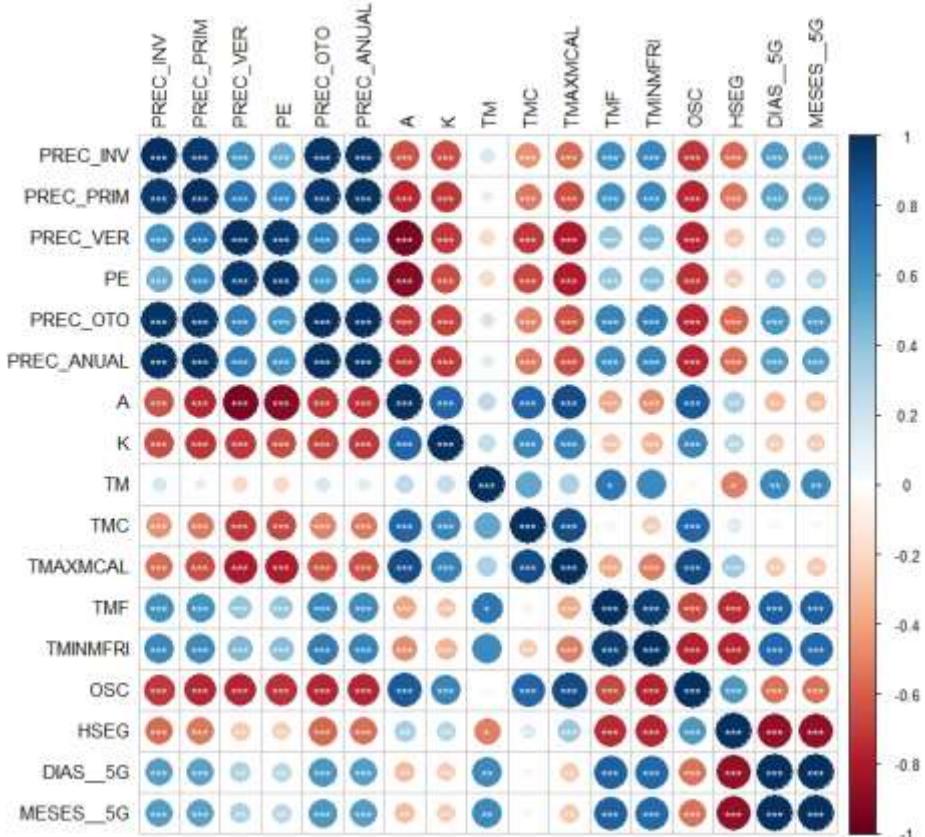


Figura 1. Matriz de correlación de las variables fitoclimáticas de la distribución natural de *Pinus pinaster* en la península ibérica. Correlaciones positivas y negativas se muestran en azul y rojo, respectivamente. Tamaño e intensidad de color de cada círculo son proporcionales a los coeficientes de correlación. Niveles de significancia ($p<0.05^*$, $p<0.01^{**}$ y $p<0.001^{***}$).

Existen multitud de métodos para identificar las variables más explicativas en un modelo, evitando así redundancias, pero sin duda dicha selección debería tener también un cierto carácter subjetivo según la relevancia ecológica de la variable para la especie considerada, siendo preferibles aquellas con un mayor sentido biológico e integrador. Las técnicas de análisis multivariantes, por ejemplo, reducen la dimensionalidad permitiendo seleccionar aquellas variables que retienen una alta proporción de la información original (TABACHNICK & FIDELL, 2007). Mediante un análisis de componentes principales se pueden detectar los componentes que acumulan la mayor parte de la varianza y utilizarlos como nuevas variables construidas como combinación lineal, o bien, seleccionar aquellas variables originales que, representando en gran medida la variación total del conjunto, no estén correlacionadas entre sí. También se puede recurrir a métodos basados en modelos lineales generalizados (GLM) para una selección automatizada. Por ejemplo, la función *multGLM* de la biblioteca *fuzzySim* (BARBOSA, 2015) del paquete estadístico R, permite seleccionar las variables para un conjunto de registros de presencia/ausencia de especies a partir de un GLM binomial.

En este trabajo se seleccionaron las variables mediante un análisis de componentes principales de cada especie (Figura 2), escogiendo aquellas que más saturaban (mayor autovector) los componentes principales con autovalor superior a 0.7 (Jolliffe, 1972; 1973), y que en su conjunto representaran, al menos, el 80% de la varianza acumulada. De esta primera selección se excluyeron las variables así sugeridas por la función *multGLM* con los siguientes argumentos: *corSelect*, de exclusión por correlación (umbral de 0.75); *step*, para realizar una selección de variable paso a paso (criterio de información de Akaike) comenzando por modelo completo y en ambos sentidos; *trim*, para una eliminación de variables no significativas (*p-*

valor ≥ 0.05). Se analizó la multicolinealidad de las variables con la función *multicoll* de la misma biblioteca de R, desestimando en todos los casos las variables *precipitación anual* y *oscilación térmica* al ser una combinación lineal de otras variables analizadas. Con la unión de ambos métodos, se escogieron entre tres y cinco variables para cada especie que son las que se utilizan en los cálculos posteriores de distancia. Un resumen de las variables seleccionadas para algunas especies forestales, y calculadas según su distribución natural en Castilla y León o la España peninsular, se muestra en la Tabla 2.

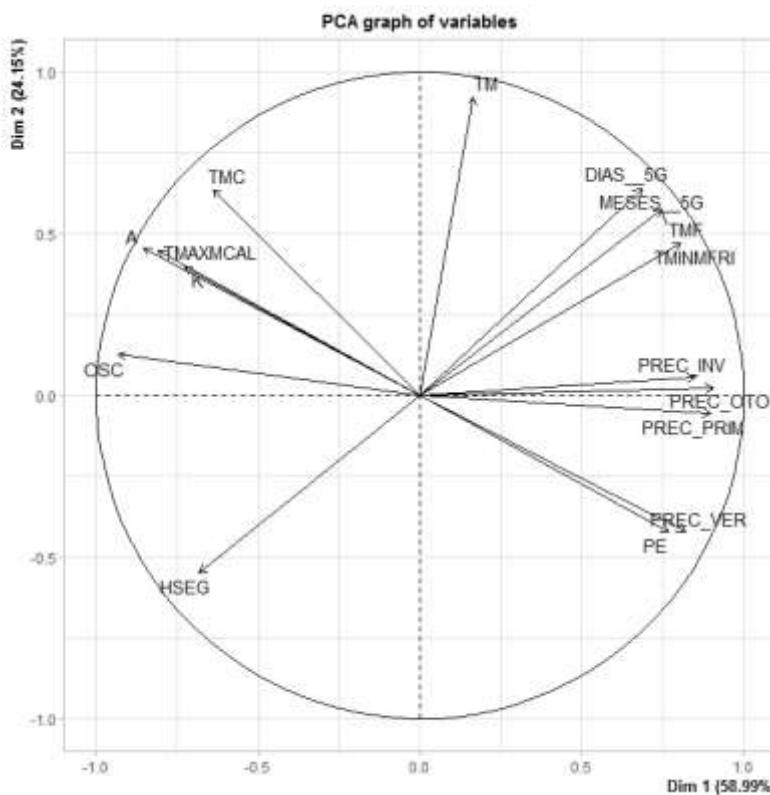


Figura 2. Gráfica de las dos primeras componentes principales de las variables fitoclimáticas de la distribución peninsular de *Pinus pinaster*.

Cálculo de la distancia

Para el cálculo de la disimilitud entre puntos se ha empleado la distancia estadística de Mahalanobis, muy utilizada en ecología y recomendable con variables aleatorias multidimensionales altamente correlacionadas (LEGENDRE & LEGENDRE, 1998). Esta distancia (MAHALANOBIS, 1936) tiene en cuenta la correlación existente entre las variables aleatorias reduciendo la redundancia y además es invariante a los cambios de escala, no dependiendo, por tanto, de las unidades de medida.

La distancia de Mahalanobis (DM) entre los puntos i y j vendría definida por la siguiente expresión: $DM(i, j) = [(x_i - x_j)' S^{-1} (x_i - x_j)]^{1/2}$ donde x son los vectores de variables fitoclimáticas de los distintos puntos i y j , siendo uno el de la zona de plantación y el otro cada uno de los puntos de distribución, y S^{-1} la inversa de la matriz de varianza-covarianza de las variables. Esta distancia puede interpretarse como una medida aproximada de la disimilitud de cada punto considerado de la distribución la especie respecto a la zona de utilización indicada, de acuerdo a las variables empleadas en el cálculo. Cuanto menor sea el valor de esta distancia mayor será la similitud fitoclimática.

Tabla 2. Parámetros fitoclimáticos recomendados, según lo indicado en apartado de selección de variables, a utilizar en el ámbito de la península ibérica (SPA) o de Castilla y León (CYL) para las especies consideradas.

Especie	TM	TMC	TMAX MCAL	TMF	TMIN MFRI	OSC	PREC_ PRIM	PREC_ VER	PREC_ OTO
<i>Pinus pinaster</i>	SPA	CYL			CYL			SPA	CYL
<i>Quercus ilex</i>	CYL/SPA						CYL	CYL/SPA	
<i>Pinus sylvestris</i>	CYL/SPA						CYL/SPA	CYL/SPA	
<i>Quercus pyrenaica</i>	CYL/SPA							CYL/SPA	
<i>Juniperus thurifera</i>	SPA							CYL	CYL
<i>Juniperus communis</i>	SPA	CYL					SPA	SPA	CYL
<i>Juniperus oxycedrus</i>			SPA	CYL/SPA					
<i>Pinus nigra</i>	SPA						CYL	CYL/SPA	
<i>Juniperus phoenicea</i>			SPA		SPA		CYL		
<i>Prunus avium</i>	SPA		SPA		CYL				CYL
<i>Sorbus domestica</i>			SPA	SPA	CYL				
<i>Quercus faginea</i>	SPA				CYL			SPA	
<i>Fraxinus excelsior</i>	SPA				CYL		SPA	SPA	CYL
<i>Juglans regia</i>	CYL/SPA		SPA				SPA		CYL

Especie	PREC_ INV	PREC_ ANUAL	PE	HSEG	DIAS_ 5G	MESES _5G	A	K
<i>Pinus pinaster</i>	SPA			SPA				CYL/SPA
<i>Quercus ilex</i>	SPA			SPA				CYL/SPA
<i>Pinus sylvestris</i>				SPA				CYL/SPA
<i>Quercus pyrenaica</i>	CYL/SPA			SPA				CYL/SPA
<i>Juniperus thurifera</i>	SPA			CYL/SPA			SPA	CYL/SPA
<i>Juniperus communis</i>				CYL/SPA				CYL/SPA
<i>Juniperus oxycedrus</i>	SPA			SPA			SPA	CYL/SPA
<i>Pinus nigra</i>	SPA			CYL/SPA				CYL/SPA
<i>Juniperus phoenicea</i>	SPA		CYL	CYL			SPA	CYL
<i>Prunus avium</i>	SPA						SPA	CYL
<i>Sorbus domestica</i>	CYL/SPA		CYL					CYL/SPA
<i>Quercus faginea</i>	CYL/SPA		CYL					CYL/SPA
<i>Fraxinus excelsior</i>				SPA				CYL/SPA
<i>Juglans regia</i>				SPA	CYL		SPA	CYL

4. Resultados y discusión

El resultado final obtenido, tras aplicar los cálculos de la metodología propuesta, es un listado de distancias de Mahalanobis entre la zona de plantación indicada y todos y cada uno de los puntos de distribución seleccionada de la especie (autóctona, no autóctona y/o potencial), en el ámbito escogido. Estas distancias calculadas, que pueden representarse gráficamente en un GIS a partir de las coordenadas UTM de cada punto favoreciendo así su interpretación, suponen una medida de proximidad ecológica que pueden utilizarse para recomendaciones de uso y homologaciones ecológicas de las diferentes especies forestales al identificar las zonas de mayor y menor similitud a un punto de referencia considerado como zona de plantación.

Para una mayor claridad y mejor interpretación, los valores de distancia pueden reclasificarse en grupos empleando para ello unos criterios que pueden ser más o menos subjetivos, o tener cierta base ecológica, estadística o incluso normativa. Por ejemplo, en datos multivariantes distribuidos normalmente la distancia al cuadrado de Mahalanobis sigue aproximadamente una distribución chi-cuadrado (χ^2) con tantos grados de libertad como variables empleadas (Figura 3). Es por ello que esta distancia se emplea como método para identificar valores atípicos en distribuciones multidimensionales (CABANA et al, 2019; FILZMOSER, 2016; GARCÍA y URIBE, 2013); por tanto, valores de distancia que superen un

determinado umbral crítico ($D^2 > \chi^2_{n;1-\alpha}$), se pueden interpretar como atípicos. Así pues, se puede determinar las distancias que se encuentran por encima de un determinado valor crítico de la distribución χ^2 , considerado como umbral superior, y que pueden interpretarse como puntos de menor similitud ($D^2 > \chi^2_{n;1-\alpha}$). Análogamente, puntos con valores de distancia inferiores a un valor de probabilidad considerado como umbral inferior ($D^2 < \chi^2_{n;1-\alpha}$) pueden considerarse como los más próximos al punto objetivo de análisis.

Es importante reseñar que cuanto más alejado esté el punto de plantación del centroide de la distribución (actual o potencial) de la especie, valores más elevados de distancia arrojará la aplicación, pudiendo llegar a obtener escasos (o incluso ninguno) puntos de similitud aceptable según el valor crítico de la distribución chi-cuadrado.

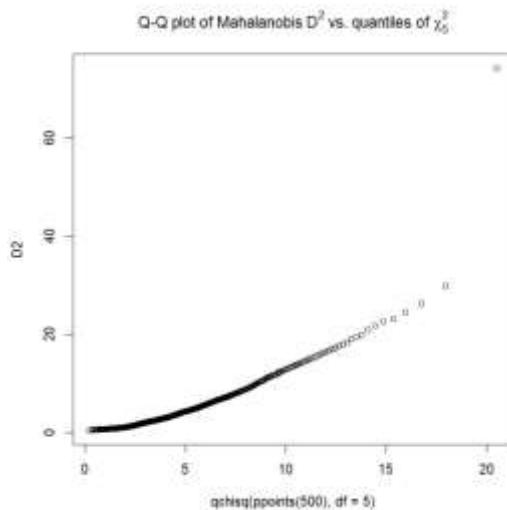


Figura 3. Gráfica que representa la distancia de Mahalanobis al cuadrado frente a los cuantiles de la distribución chi-cuadrado con cinco grados de libertad.

Otro método de reclasificación, que complementa o sustituye al anterior, puede utilizar un criterio más *normativo* basado por ejemplo en los esquemas de comercialización de los materiales forestales de reproducción. Así pues, considerando las regiones de procedencia (RIU en el método divisivo) como unidades homogéneas ecológicamente (Real Decreto 289/2003, de 7 de marzo, sobre comercialización de los materiales forestales de reproducción), podemos emplear la distribución de distancias que la especie tiene en la región de uso para obtener grupos de similitud. Así, distancias de los puntos, en la RIU donde se realiza el análisis, que se sitúen por encima de un valor de compromiso, como puede ser el tercer cuartil (Q3) más 1.5 veces el rango intercuartílico (Q3-Q1), se pueden considerar como más alejados, y aquellos valores inferiores al tercer cuartil, como puntos más próximos. Entre ambos se situarían los valores intermedios de similitud. Evidentemente, si la especie no vegeta en la RIU donde se hace la simulación esta reclasificación no se puede realizar. No obstante, el usuario puede utilizar su propia reclasificación *ad hoc* si así lo estima oportuno.

Complementariamente, esta misma metodología descrita se podría emplear para calcular los valores de distancia con respecto a la estación natural de procedencia, o lugar de ensayo, de un determinado material de base seleccionado o mejorado (rodal selecto o clon, por ejemplo), y así determinar la aptitud de dicho material en su ámbito de distribución. En este sentido, en la Figura 4 se ejemplifica la idoneidad de dos clones de cerezo, aprobados en el registro nacional de materiales de base (CYL-01 y CYL-04), en el territorio de Castilla y León a partir de las distancias de Mahalanobis con respecto a los puntos que determinan el origen natural de dichos clones.

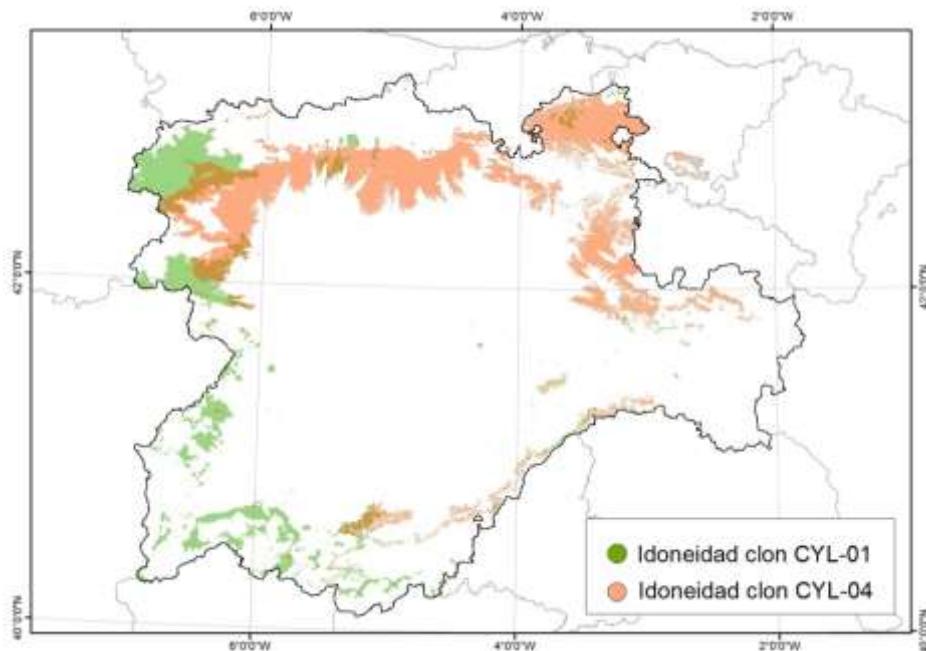


Figura 4. Mapa de la idoneidad de los clones de cerezo CYL-01 y CYL-04 en Castilla y León. Se muestran solamente los valores de mayor aptitud (1^{er} tercio de distancias) y exclusivamente en el ámbito modelizado como viable por ALONSO PONCE et al (2009) para la especie.

A continuación se presenta, mediante un ejemplo, el proceso de introducción de datos en la aplicación desarrollada para el cálculo de la distancia de Mahalanobis objeto de este trabajo, así como las tablas de resultados que se obtienen con las reclasificaciones de similitud comentadas anteriormente.

Aplicación práctica mediante hoja de cálculo

Programando en *visual basic* todas operaciones se puede utilizar una hoja de cálculo de tipo Excel para obtener una aplicación que permita determinar, a partir de diferentes tablas de entrada preestablecidas, la distancia de Mahalanobis entre la zona de implantación forestal y todas las localizaciones de la distribución de una especie, pudiendo restringir puntos por autoctonía o incrementarlos por potencialidad de la especie.

Los datos que se deben incluir en la aplicación son:

(1) las coordenadas UTM de la zona de plantación. Incluyendo un radio de afección se puede definir la estación climática de referencia contra la que calcular las distancias,

(2) la especie forestal a evaluar, según listado desplegable,

(3) los datos de distribución con los que se desea realizar el análisis, que puede ser la distribución autóctona, no autóctona y/o potencial,

(4) el ámbito espacial de análisis, que puede ser Castilla y León o toda la España peninsular,

(5) las variables fitoclimáticas con las que se desean realizar los cálculos. La aplicación dispone de unas variables recomendadas a utilizar para cada especie (Tabla 2), calculadas según lo explicado en el apartado de “selección de variables”, si bien el usuario puede utilizar otras que tengan un mayor sentido ecológico o biológico para el caso concreto.

Se presenta a continuación un ejemplo (Figura 5) simulando una plantación de *Pinus pinaster* en el término municipal de Íscar ($X=372.890$; $Y=4.582.793$), provincia de Valladolid y región de uso (RIU) nº 17.-Tierras del pan y del vino. La estación media de comparación abarca 5 kilómetros alrededor de las coordenadas indicadas. Se utilizan los puntos de distribución autóctona en toda la península ibérica. Las variables empleadas para realizar los cálculos serán

las recomendadas por la aplicación. Al ejecutar la aplicación (botón EVALUAR) se realiza una serie de cálculos que dan como resultado, principalmente, un listado de distancias de todos los puntos analizados, según distribución seleccionada, al punto o zona de implantación establecido cuyo centro de coordenadas se indica el punto (1). Esta distancia de Mahalanobis (referido en la tabla como DM) se puede interpretar como disimilitud al punto de plantación, que figura como el primero de los registros de la tabla. Teniendo en cuenta la región de utilización (RIU) como una unidad homogénea, podemos considerar como similitud alta aquellos puntos con distancia inferior al tercer cuartil (Q3), y baja las superiores al valor del tercer cuartil más 1.5 veces el rango intercuartílico (Q3+1.5*RIQ). Entre estos dos umbrales los valores se consideran medios. Otra manera alternativa de clasificar estas distancias, y que también se presenta en la tabla de resultados, es mediante las probabilidades de la distribución χ^2 con tantos grados de libertad como variables empleadas en el cálculo de la distancia. Por ejemplo, probabilidades superiores a 0.95 pueden interpretarse como puntos próximos, y aquellas inferiores a 0.05 como alejados al punto de análisis; estos umbrales pueden modificarse por parte del usuario para una mejor definición de los resultados.

Junto al valor de la distancia de Mahalanobis (DM), y la similitud asociada, la aplicación aporta otra información relativa a cada punto de distribución considerado que puede resultar de interés para una interpretación posterior de los resultados, como por ejemplo la procedencia, autoctonía del material o características generales del tipo de suelo, y que deben considerarse meramente informativas y aproximadas (Figura 6).

A partir de las coordenadas UTM de cada registro se puede cargar el archivo resultante en un programa GIS y visualizar de forma gráfica los valores de distancia obtenidos (Figura 7). La aplicación, en cualquier caso, permite exportar en un archivo *.kml los puntos de mayor similitud para una rápida visualización cartográfica de los resultados sin necesidad de un software GIS.

Complementariamente, la aplicación presenta en otra hoja un breve resumen de las características ecológicas de la zona de utilización definida por las coordenadas introducidas, pero referidas al centroide de la cuadrícula más próximo al punto indicado, y que pueden suministrar al usuario información de apoyo para la selección definitiva (Figura 8).

a) Indique un nombre para su archivo de resultados:	Pt_Iscar	
b) Indique las coordenadas del punto de implantación: (ETRS89; huso 30 extendido)	X 372.890	Y 4.582.793
c) Elija los puntos en torno al punto de implantación que quiere que se tengan en cuenta para el vector de comparación	<ul style="list-style-type: none"> - Un número concreto de puntos más próximos: - Los puntos que estén en un círculo cuyo radio (km) es: <input type="text" value="5"/> Si se indican en ambas casillas la aplicación tomará el mayor número de puntos.	
d) Seleccionar especie del desplegable:	<input type="text" value="Pinus pinaster"/> <div style="border: 1px solid #ccc; padding: 5px; margin-top: 5px;"> Pinus nigra Pinus pinaster Pinus radiata Pinus sylvestris Pinus avium Quercus faginea Quercus ilex Quercus pyrenaica </div>	
e) Indique con qué distribución de la sp se desea evaluar <small>(varias opciones son posibles)</small>	<input checked="" type="checkbox"/> Autóctono <input type="checkbox"/> No autóctono <input type="checkbox"/> Potencial	
f) Indique con qué puntos quiere realizar el análisis:	Sólo Castilla y León Toda España peninsular	<input checked="" type="checkbox"/>

g) Elija los parámetros con los que se pretende realizar el cálculo:

Grupo	ID	Descripción	Elegidos	Recomendados
Temperaturas (ºC)	TM	Temperatura media (ºC)	x	SPA
	TMC	T. media mensual más alta		CYL
	TMAXMCL	T. media de las máx. del mes más cálido		
	TMF	T. media mensual más baja		
	TMINMFRI	T. media de las mín. del mes más frío		CYL
	OSC	Oscilación térmica (TMC-TMF)		
Precipitaciones (mm)	PREC_PRIM	Precipitación de primavera (mm)		
	PREC_VER	Precipitación de verano	x	SPA
	PREC_OTO	Precipitación de otoño		CYL
	PREC_INV	Precipitación de invierno	x	SPA
	PREC_ANUAL	Precipitación total anual		
	PE	Precipitación mensual estival mínima		
Índices	HSEG	Número de meses de helada segura	x	SPA
	DIAS_5G	Número de días al año con más de 5ºC		
	MESES_5G	Número de meses al año con más de 5ºC		
	A	Duración de la aridez (en meses)		
	K	Intensidad de la aridez (Allué-Andrade)	x	CYL/SPA

EVALUAR
Borrar TODO
exportar resultados a KML (Google Earth)
[1000 pts]

Figura 5. Aspecto de la hoja de cálculo para la distancia entre puntos. Se muestran los apartados relativos al nombre del archivo de resultados, selección de punto de implantación y de su extensión para la comparación, distribución actual y/o potencial a emplear, selección de la especie, ámbito espacial de distribución y selección de variables, con indicación de las recomendadas para la especie escogida.

X	Y	autoctonia	RIU	altitud	Suelo (FAO)	pH	CZ_FTA	DM	similitud (RIU)	DM2	P(x<a)	similitud χ^2 (0,05/0,95)
372177	4578106		17	750	Fluvisoles eutrícos	Ácido	23		plantación	0	1,0000	próximo
367890	4576793	autóctono	17	758	Cambisoles calcáricos	Básico	23	0,1472	alta	0,0217	1,0000	próximo
365890	4576793	autóctono	17	756	Cambisoles calcáricos	Básico	23	0,1708	alta	0,0292	1,0000	próximo
370890	4568793	autóctono	17	772	Arenosoles cambícos	N	23	0,1734	alta	0,0301	1,0000	próximo
357891	4562793	autóctono	17	775	Luvisoles verticos	Ácido	23	0,1823	alta	0,0332	1,0000	próximo
367890	4571793	autóctono	17	759	Arenosoles cambícos	N	23	0,184	alta	0,0339	1,0000	próximo
369890	4569793	autóctono	17	763	Arenosoles cambícos	N	23	0,1874	alta	0,0351	1,0000	próximo
371890	4569793	autóctono	17	774	Arenosoles cambícos	N	23	0,1923	alta	0,0370	1,0000	próximo
368890	4575793	autóctono	17	760	Cambisoles calcáricos	Básico	23	0,1927	alta	0,0371	1,0000	próximo
370890	4544793	autóctono	17	848	Arenosoles cambícos	N	23	0,2029	alta	0,0412	1,0000	próximo
366890	4576793	autóctono	17	758	Cambisoles calcáricos	Básico	23	0,2059	alta	0,0424	1,0000	próximo
181460	4819254	autóctono	3	532	Podsoles húmicos	Ácido		4,1528	baja	17,2457	0,0084	alejado
28983	4723707	autóctono	1	108	Cambisoles húmicos	Ácido		4,1528	baja	17,2457	0,0084	alejado
225067	4468143	autóctono	18	599	Leptosoles distícos	Ácido		4,1529	baja	17,2466	0,0084	alejado
158270	4803864	autóctono	3	145	Rankers	Ácido		4,1529	baja	17,2466	0,0084	alejado
554891	4535791	autóctono	21	1183	Cambisoles calcáricos	Básico		4,1531	baja	17,2482	0,0084	alejado
564891	4530791	autóctono	21	1201	Cambisoles eutrícos	N		4,1531	baja	17,2482	0,0084	alejado
537891	4601791	autóctono	16	1073	Cambisoles gleícos	Ácido	19	4,1532	baja	17,2491	0,0084	alejado
642890	4459791	autóctono	22	1368	Cambisoles eutrícos	N		4,1532	baja	17,2491	0,0084	alejado

Figura 6. Salida gráfica de la tabla de resultados. Se presenta únicamente una muestra de puntos con menor distancia (alta similitud) así como de puntos de elevada distancia (baja similitud). Se indica el valor de similitud, sombreado en rojo, según los dos criterios descritos en la metodología (similitud según criterio RIU y rango de similitud según criterio χ^2 , siendo próximo si $p>0.95$ y alejado si $p<0.05$). Se refleja otra información asociada a cada punto, como las coordenadas X e Y, autoctonía y potencialidad, RIU, altitud, tipo de suelo (FAO), pH, número de cuaderno de zona de Castilla y León (CZ_FTA), distancia de Mahalanobis (DM), distancia al cuadrado (DM2) y probabilidad según distribución χ^2 ($P(x<a)$).

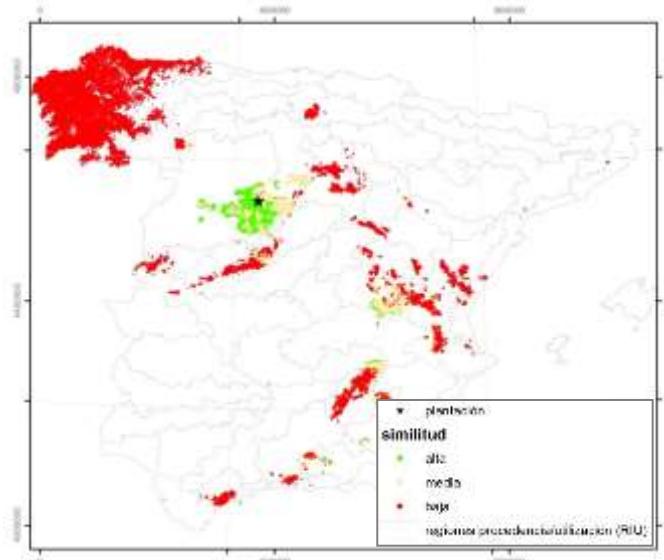


Figura 7. Similitud de los puntos de distribución actual de *Pinus pinaster* con respecto al punto de plantación, según el criterio de uniformidad de la región de uso (RIU). Similitudes altas son aquellos puntos con distancia inferior al tercer cuartil (Q3) de la RIU de plantación, y bajas, las superiores al valor del tercer cuartil más 1.5 veces el rango intercuartílico (Q3+1.5*RIQ). Entre ambos umbrales la similitud se considera media.

información de la zona de plantación (malla 1x1 km)						
Provincia	Término municipal	Región de uso (RIU)	nº zona (FTA_CyL)	Fitoclima	X (m)	Y (m)
Valladolid	Iscar	17	23	VII(V)1	372.890	4.582.793
«Distribución actual de la sp?»	Tipo Suelo (FAO)	pH	Altitud aprox (m)	«Potencialidad de la sp?»	latitud (º)	longitude (º)
Si	Cambisoles calcáricos	Básico	736	Si	41,39	-4,52
TM	TMC	TMAXCAL	TMF	TMINMFRI	OSC	
12,3	22,4	31,6	3,7	-0,7	18,7	
PREC_PRIM	PREC_VER	PREC_OTO	PREC_INV	PREC_ANUAL	PE	
126	64	133	130	453	13	
HSEG	DIAS_5G	MESES_5G	A	K		
1,80	307	10,1	3,15	0,28		

Figura 8. Salida gráfica de la tabla informativa relativa a la zona de plantación. Los datos se refieren al centroide de la cuadrícula de malla 1x1 km más próximo al punto de plantación indicado a partir de sus coordenadas UTM.

5. Conclusiones

La distancia de Mahalanobis puede emplearse, al menos en una primera aproximación, como medida de disimilitud entre puntos o poblaciones forestales. En el presente trabajo se presenta una sencilla metodología para el cálculo de esta distancia, a la vez que proporciona otra información de índole ecológica, que permite orientar las recomendaciones de uso del material forestal a emplear en repoblaciones y plantaciones forestales.

La posibilidad de implementar el procedimiento en un software de amplio uso, como una hoja de cálculo, y de personalizar los parámetros de cálculo, como son las variables empleadas o los puntos de corte de las similitudes, mejora la versatilidad de la aplicación y afina los resultados al poder utilizar para cada especie la información más idónea de la que disponga el usuario en cada caso concreto. Además, la facilidad de incorporar nuevas posibles distribuciones de las especies basadas en los escenarios de cambio climático, aumenta la potencialidad de la aplicación.

Los resultados de disimilitud entre zonas pueden servir para determinar las procedencias y materiales de base más recomendables a emplear, o para realizar homologaciones ecológicas entre zonas de uso de las distintas especies forestales. Complementariamente, permite analizar las posibilidades de uso de materiales mejorados, frecuentemente clones, valorando la aptitud ecológica del material a utilizar.

No obstante, este análisis de tipo fitoclimático debería tener en cuenta, además, otras consideraciones de igual o mayor trascendencia que pueden afinar, modificar o apoyar a posteriori la interpretación de los valores resultantes, como son la similitud en parámetros edáficos y fisiográficos, condicionantes de conservación de recursos genéticos forestales o incluso criterios de proximidad geográfica,

6. Agradecimientos

Los autores desean agradecer la colaboración de todos los compañeros que han participado testando la aplicación descrita, identificando fallos en el proceso de cálculo, así como sugiriendo mejoras. Agradecer al INIA toda la información GIS facilitada. Julio Díez y Valentín Pando (ETSIIAA-Palencia, Universidad de Valladolid) mejoraron el texto con sus revisiones.

7. Bibliografía

- ALLÚE, J. L. (1990). Atlas Fitoclimático de España. Taxonomías. Instituto Nacional de Investigaciones Agrarias. Ministerio de Agricultura, Pesca y Alimentación. Madrid.
- ALONSO PONCE, R.; CISNEROS, O.; MODREGO, P.; SÁNCHEZ-PALOMARES, O. (2009). Recomendaciones de uso para clones de cerezo (*Prunus avium* L.) en Castilla y León. 5º Congreso Forestal Español. Ávila.
- AUÑÓN, F. J.; ALÍA; R.; SÁNCHEZ DE RON; D., MIGUEL Y DEL ANGEL, J.; GARCÍA DEL BARRIO, J. M. (2011). Sig-Forest. Visor Sig de especies forestales. *Forestal*, 51, 38–43.
- AUÑÓN, F. J.; SÁNCHEZ DE RON, D. (2010). MANUAL PARA EL USO DEL VISOR ARC-GIS-SERVER (Regiones de Procedencia de las especies del RD 289/2003).
- BARBOSA, A. M. (2015). FuzzySim: applying fuzzy logic to binary similarity indices in ecology. *Methods. Ecol Evol*, 6(7), 853–858.
- BUIJTENEN VAN, J. P. (1992). Fundamental genetic principles. En: FINS, S. T.; FRIEDMAN, J.; BROTSCHOL (eds.): *Handbook of Quantitative Forest Genetics*. Pag 29-68 Kluwer Academic Publishers. Dordrecht.
- CABANA, E.; LILLO, R. E.; LANIADO, H. (2019). Multivariate outlier detection based on a robust Mahalanobis distance with shrinkage estimators. *Stat. Pap.* 62, 1583–1609.
- DAGNELIE, P. (1975). Analyse statistique à plusieurs variables. Presses Agron., Gembloux, 362 pp.
- FILZMOSER, P. (2016). Identification of Multivariate Outliers: A Performance Study. *Austrian J Stat*, 34(2), 127–138.
- GARCÍA, J. A. M.; URIBE, I. A. (2013). Técnicas para detección de outliers multivariantes.

Rev en Telecomunicaciones e Informática, 3(5), 11–25.

GONZALO, J. (2010). Diagnosis fitoclimática de la España peninsular. Organismo Autónomo Parques Nacionales, Ministerio de Medio Ambiente Rural y Marino. Madrid.

JOLLIFFE, I. T. (1972). Discarding variables in a principal component analysis. I: Artificial data. *Appl Stat*, 21(2), 160–173.

JOLLIFFE, I. T. (1973). Discarding variables in a principal component analysis. II: Real data. *Appl Stat*, 22(1), 21–31.

KAWECKI, T. J.; EBERT, D. (2004). Conceptual issues in local adaptation. *Ecol Lett*, 7(12), 1225–1241.

LEGENDRE, L.; LEGENDRE, P. (1998). Numerical ecology. Developments. *Environ Model*. 2nd English Edn. Elsevier, Amsterdam.

MAHALANOBIS, P. C. (1936). On the Generalized Distance in Statistics. *Proc. Nat. Inst. Sci., India*, 2, 49–55.

NAMKOONG, G. (1969). Nonoptimality of Local Races. Proceedings of the tenth Southern Conference on Forest Tree Improvement. Houston, Texas.

PHILLIPS, S.; ANDERSON, R.; Y SCHAPIRE, R. (2006). Maximum entropy modeling of species geographic distributions. *Ecol Model*, 190(3–4), 231–259.

SAVOLAINEN, O.; PYHÄJÄRVI, T.; Y KNÜRR, T. (2007). Gene Flow and Local Adaptation in Trees. *Annu. Rev. Ecol. Evol Syst*, 38(1), 595–619.

TABACHNICK, B. G.; FIDELL, L. S. (2007). Using multivariate statistics (Vol. 5). Pearson. Boston, MA.

VAN ANDEL, J. (1998). Intraspecific variability in the context of ecological restoration projects. *Perspect Plant Ecol Evol Syst*, 1(2), 221–237.