Universitat Autònoma de Barcelona

Análisis de factores determinantes de las transiciones forestales y el abandono de cultivos en el ámbito mediterráneo. Un modelo de dinámica de cubiertas del suelo basado en Teledetección, SIG y Boosted Regression Trees

Juan José Vidal Macua

Análisis de factores determinantes de las transiciones forestales y el abandono de cultivos en el ámbito mediterráneo

Un modelo de dinámica de cubiertas del suelo basado en Teledetección, SIG y *Boosted Regression Trees*

Juan José Vidal Macua

Tesis Doctoral en Geografía
Universitat Autònoma de Barcelona

Universitat Autònoma de Barcelona

Tesis doctoral

Programa de Doctorado en Geografía

Departament de Geografia

Análisis de factores determinantes de las transiciones forestales y el abandono de cultivos en el ámbito mediterráneo

---

Un modelo de dinámica de cubiertas del suelo basado en Teledetección, SIG y *Boosted Regression Trees*

Juan José Vidal Macua

Diciembre de 2017

Directores:

Xavier Pons Fernández

Miquel Ninyerola Casals

Alaitz Zabala Torres

Fotos de portada:

Autor: Juan José Vidal Macua
Valle de Estós (Pirineo de Huesca, Aragón)
Valle del Río Ara y Peña Montañesa (Pirineo de Huesca, Aragón)

Diseño de iconos de portada:

Noun Project. Icons for everything
https://thenounproject.com/

Agradecimientos

AGRADECIMIENTOS

En primer lugar, dar las gracias al grupo de investigación GRUMETS por darme la oportunidad de hacer la tesis doctoral dentro de su equipo, y en especial a su director, el Dr. Xavier Pons, por haber confiado en mí desde el principio, por sus consejos y ayuda a lo largo de estos cuatro años. Ha sido una experiencia profesional y personal formidable el haber trabajado con un grupo tan competente, sin duda un referente en el ámbito de las tecnologías de la información geográfica. También agradecer la información y la colaboración prestada por el grupo para la elaboración de esta tesis. A Xavier, Miquel y Alaitz: gracias por vuestra ayuda y disponibilidad durante todo este tiempo, y por ser los responsables del proyecto que ha financiado mi tesis.

Agradecer al Ministerio de Economía y Competitividad el haber financiado mi trabajo de investigación a través de la beca FPI BES-2013-063766 en el contexto de los proyectos (concedidos a GRUMETS) DinaCliVe (*Análisis espaciotemporal de las cubiertas del suelo y del estrés de la vegetación en la Península Ibérica a la luz de medio siglo de dinámica climática y sus anomalías*) y ACAPI (*Análisis del Cambio Global en la Península Ibérica: un laboratorio integrado del clima y las cubiertas del suelo 1950-2030*), y por haberme concedido la ayuda para realizar la estancia de investigación en el centro *Institute of Landscape Ecology of the Slovak Academy of Sciences* (Eslovaquia).

Al Departament de Geografia de la Universitat Autònoma de Barcelona, por las instalaciones y espacios de los que he estado haciendo uso. A los compañeros de departamento por el intercambio de conocimientos y buenos ratos, y la satisfacción de haber participado en un departamento tan dinámico y puntero. A mis compañeros de doctorado, gracias también por el intercambio de ideas y por vuestra amistad.

Al equipo de programadores de GRUMETS, en especial a Xavier Pons, Xavier Calaf y Abel Pau, por el desarrollo y mejora de herramientas en momentos importantes. A mis compañeros, también miembros del grupo de investigación, Oscar González, Xavier Calaf, Guillem Closa, Cristian Padró, Alaitz Zabala, Maria Mira, Cristina Cea, Cristina Domingo, Lluis Pesquer y Pere Serra por los momentos en los que he pedido vuestra colaboración, pero sobre todo, por el buen ambiente y la sensación de confianza que me habéis transmitido. Alaitz, gracias por iluminar la oscura senda de la maquetación de textos en el programa que tú ya sabes.

Al equipo del SIOSE del Ministerio de Fomento e Instituto Geográfico Nacional, por su amabilidad y diligencia a la hora de proporcionar información referente a esta base de datos.

I would like to thank to the Institute of Landscape Ecology of the Slovak Academy of Sciences for giving me the opportunity of doing my international research stay within its highly competent research team. I especially thank Dr. Lubos Halada for having coordinated the stay and for his kindness.

Al Departamento de Geografía de la Universidad de Zaragoza y a sus profesores les agradezco el haber fortalecido en su momento mi vocación por la Geografía, y la seguridad de haber estudiado esta carrera en uno de los mejores centros del estado. Y especialmente al Dr. Juan de la Riva, por haber facilitado en 1999 mi primer contacto con el ámbito profesional a través del Instituto Aragonés de Fomento.

Finalmente quiero agradecer a mi familia y amigos por animarme a lo largo de estos años. Y muy especialmente a Susana, gracias por estar siempre ahí, por toda la ilusión y el respaldo que me has dado, y la paciencia que has tenido durante este tiempo.

Índice

ÍNDICE

Resumen/Summary

RESUMEN

La dinámica de cualquier fenómeno geográfico, en relación a su cambio de patrón espacial a lo largo del tiempo, responde a una serie de agentes que directa o indirectamente actúan como impulsores de cambio. La configuración espacial que los sistemas naturales y socioeconómicos tienen sobre el territorio depende por tanto de la interacción de uno o más agentes, de tal forma que un cambio prolongado en el tiempo sobre ellos puede derivar en un nuevo escenario territorial. La vegetación y la actividad agraria son componentes fundamentales del territorio, cuya dinámica temporal puede ser monitorizada en relación a una serie de agentes o factores de cambio. Si bien ambos están relacionados, dado que la expansión o abandono de la actividad agraria influye directamente en la distribución de la vegetación, las condiciones que favorecen un cambio en su dinámica temporal pueden estar caracterizadas por diferentes factores explicativos. El cese de la actividad agraria en las regiones geográficas más periféricas del ámbito mediterráneo ha dado lugar a la revegetación natural de las tierras abandonadas, al mismo tiempo que ha habido una intensificación de los usos agrarios en otras regiones más accesibles. Por un lado, el análisis de la dinámica de la vegetación permite identificar qué condiciones son las más favorables para la expansión y colonización del espacio por parte de los diferentes tipos fisionómicos de vegetación. Por otro lado, un análisis del abandono agrícola ayuda a determinar la vulnerabilidad o idoneidad de la actividad agrícola en relación a una serie de factores ambientales y socioeconómicos. La comprensión de todo ello es fundamental para la definición de políticas eficaces de gestión del territorio.

Los análisis basados en modelos de dinámica de usos y cubiertas del suelo suponen sin duda la aproximación más coherente para la definición de patrones de cambio en relación a la influencia de una serie factores potencialmente explicativos. Las transiciones a lo largo del tiempo entre diferentes cubiertas de vegetación así como el abandono de usos agrarios, pueden monitorizarse realizando una cartografía temática de la superficie terrestre en diferentes momentos temporales. En este sentido, la teledetección espacial es la técnica más eficaz para la obtención de mapas de usos y cubiertas del suelo; no sólo porque permite un procedimiento semi-automático, sino también por la resolución temporal que ofrece y la posibilidad de analizar grandes superficies de terreno, características fundamentales para la definición de patrones a escala regional o continental. Junto con la teledetección, son necesarias otras técnicas que han de ser capaces de gestionar grandes volúmenes de información proveniente de las imágenes satelitales y, a su vez, definir con solidez los patrones que relacionan las transiciones entre cubiertas y sus factores determinantes. Dentro de este grupo de técnicas se encuentran los Sistemas de Información Geográfica, la geocomputación y los modelos estadísticos basados en aproximaciones *machine-learning*.

El ámbito mediterráneo de la Península Ibérica es una región geográfica de gran interés, dado que desde mediados del s. XX ha habido un intenso proceso de abandono agrícola, especialmente en las zonas montañosas, lo que ha llevado consigo un recuperación de la dinámica natural de la vegetación en estos espacios. Este ámbito se caracteriza por una amplia variabilidad en cuanto a contextos topográficos. Además, es una región que durante las últimas décadas se ha visto afectada por un incremento de las sequías, y donde los modelos climáticos pronostican una continuidad de dicha tendencia.

El núcleo principal de esta tesis lo componen tres artículos científicos, dos de ellos publicados y un tercero aceptado, aunque todavía no publicado en el momento de depósito de la tesis. En el primer artículo se presenta la metodología para la obtención de series temporales de mapas de usos y cubiertas del suelo para grandes superficies de terreno en esta región geográfica. Para ello se ha implementado un esquema de clasificación de imágenes del satélite Landsat. La información de referencia utilizada para la obtención

de áreas de entrenamiento y test ha sido la base de datos SIOSE 2005 (Sistema de Información sobre Ocupación del Suelo de España). El artículo incluye un procedimiento para la obtención de estas áreas en fechas que no disponen de tal información de referencia. Las clasificaciones obtenidas tienen un porcentaje de acierto global superior al 90% y un acierto por categorías temáticas por encima del 85% en la mayoría de los casos.

En el segundo artículo se analizan las principales etapas de sucesión y transiciones de vegetación entre las siguientes categorías de cubiertas de vegetación: prados, matorrales, bosques de coníferas, bosques de frondosas perennifolias y bosques de frondosas caducifolias. Los ámbitos de estudio se corresponden con tres escenas Landsat y el periodo de estudio es 1987-2012. Como factores explicativos de estas transiciones se ha utilizado un conjunto de variables de las que destacan las derivadas de la topografía y las variables que cuantifican la recurrencia de periodos secos, obtenidas a partir de un SIG (Sistema de Información Geográfica). Para determinar qué factores son los más importantes y de qué manera influyen en cada etapa de sucesión y transición se ha utilizado el método *machine-learning* denominado *Boosted Regression Trees* (BRT). De los numerosos resultados obtenidos cabe destacar que las transiciones a frondosas mediterráneas y sub-mediterráneas son más probables en aquellas zonas más afectadas por la repetición de periodos secos, mientras que las transiciones a coníferas son más probables en aquellas zonas con menor ocurrencia de estos eventos, pero que existen importantes interacciones con el contexto topográfico que hay que tener en cuenta. Del análisis se desprende que el decline en la tasa de transición a coníferas y el incremento en la tasa de transición a frondosas se relaciona positivamente con el aumento de periodos secos y que, por tanto, si se cumplen los pronósticos de los modelos climáticos, en algunas zonas podría haber un cambio de especie dominante ante la expansión de las frondosas mediterráneas y sub-mediterráneas.

El tercer artículo es un análisis de los factores internos de abandono de, por un lado, cultivos de secano y, por otro, cultivos de regadío también mediante teledetección, SIG y BRT. En este caso son dos ámbitos de estudio (periodo 1987-2012): el Pirineo central español y la cuenca central del Ebro. En este análisis se han incorporado variables que cuantifican el coste de desplazamiento a núcleos urbanos y carreteras, y otras que incluyen información socioeconómica a nivel municipal. De este artículo son también varios los resultados, de los que cabe destacar los siguientes: los limitantes físicos tienen un peso específico muy importante, excepto para los cultivos de regadío en la cuenca central, donde la mayor influencia la tiene acceso a infraestructuras; la actividad agrícola en el Pirineo es más dependiente de mercados internos a nivel local; los cultivos de regadío son más vulnerables a la ocurrencia de sequías; la diversificación económica puede ayudar a mantener la actividad agraria, especialmente en el Pirineo. La probabilidad de abandono no siempre tiene una relación lineal con los factores explicativos, encontrando en algunos casos patrones de respuesta polarizados que reflejan, para un mismo factor, diferentes condiciones que favorecen el abandono.

SUMMARY

The dynamics of any geographical phenomenon, in relation to its spatial-pattern change over time, responds to a series of agents that directly or indirectly act as drivers of change. Therefore, the natural and socioeconomic systems spatial configuration over the territory depends on the interaction of one or more agents, so that a prolonged change over time in them can lead to a new territorial scenario. Vegetation and agricultural activity are fundamental components of terrestrial ecosystems and society, whose temporal dynamics can be monitored in relation to a series of agents or change factors. Although these two components are related two each other, given that the expansion or abandonment of agricultural activity directly influences on vegetation distribution, the conditions leading to a change in their temporal dynamics can be characterized by different explanatory factors. In the most peripheral geographical regions of the Mediterranean area, the cessation of agricultural activity has led to the natural re-vegetation of abandoned lands, while there has been land-use intensification in the more accessible regions. On the one hand, the analysis of vegetation dynamics allows identifying which conditions are the most favorable for the expansion of different vegetation physiognomic groups. On the other hand, the analysis of agricultural abandonment helps to determine the vulnerability or suitability of agricultural activity in relation to a series of environmental and socioeconomic factors. All these issues are very important for the development of territorial management effective policies.

Undoubtedly, the analysis based on land-use and land-cover dynamics models represent the most coherent approach for defining change patterns in relation to the influence of a series of potentially explanatory factors. The transitions over time between different vegetation covers and the agricultural abandonment can be monitored by developing thematic land-cover maps at different times. In this sense, remote-sensing based techniques are the most efficient in obtaining land-cover/land-use maps; not only because their semi-automatic procedure, but also because of their temporal resolution and because they allow analyzing large geographical areas, both main features for defining patterns at a regional or continental scale. In addition to remote sensing, other techniques being capable of managing large volumes of satellite data and, in turn, defining solid relational-patterns between land-cover changes and their determinant factors are needed, as for example, Geographical Information Systems, geocomputation and statistical models based on machine-learning approaches.

The Mediterranean fringe of the Iberian Peninsula is a region of great interest, since during the second half of XX century there has been an intense process of agricultural abandonment, mainly in the mountainous areas, which has led to a recovery of the natural vegetation dynamics. This area is characterized by a wide variability of topographic contexts. In addition, this region has been affected by an increase in droughts during the last decades, and climate trends are expected to be characterized by a similar trend.

The main core of this thesis consists of three scientific articles, two of them published and a third one that is accepted but not yet published at the time of the thesis deposit. The first article deals with obtaining temporal series of land-use and land-cover maps for large areas of this region. To do this, a classification scheme of Landsat imagery has been implemented. Training and test areas were extracted from the Land Occupation Information System of Spain 2005 (SIOSE, Sistema de Información sobre Ocupación del Suelo en España). The article includes a procedure for obtaining these areas on dates lacking such kind of reference information. An overall accuracy of over 90% has been obtained for all maps, and accuracy by categories is above 85% in most cases.

The second article analyzes the main successional stages and transitions between the following vegetation-cover categories: grasslands, shrublands, coniferous forests, broadleaf evergreen forests and broadleaf deciduous forests. Three areas of Spain were included in this study as different scenarios for model development. Ambits correspond to land-cover classifications of three entire Landsat scene and the study period is 1987-2012. Conditions were described by topography derived variables and drought-occurrence variables, which were obtained through a Geographical Information System (GIS). The machine-learning technique Boosted Regression Trees (BRT) were used to identify the most important variables and describe the relationships between the forest vegetation transitions and key factors. Main findings: Transitions to Mediterranean and sub-Mediterranean broadleaf forests are positively associated with drought occurrence while transitions to conifers are negatively affected by drought; Important interactions between topography derived variables and drought have been found. The study provides robust evidence that drought occurrence plays an important role in the decline of conifers and the expansion of Mediterranean and sub-Mediterranean broadleaves, which could become the dominant species in many areas of the Mediterranean if climate model forecasts are met.

The third article focuses on the abandonment of rainfed and irrigated herbaceous crops in order to derive specific explanations according to the crop for two different study areas: the Central Spanish Pyrenees and the Central Ebro Basin. The methodology is also based on remote sensing, GIS and BRT. Variables quantifying the cost distance to urban centers and roads and other variables related to socioeconomic data at a municipal level have been included in this analysis. Topography derived variables are the main determinants, except for irrigated crops in the Ebro Basin, where locational factors play a more important role. BRT models have allowed identifying other significant patterns such as: the vulnerability of irrigated crops to drought; the higher dependence of agricultural activity in the Pyrenees on internal networks; pattern shifts of land abandonment in the analyzed sub-periods, and; evidence of the importance of economic diversification for maintaining cropland. The abandonment probability does not always has a linear relationship with the explanatory factors and, in some cases, it shows polarized response patterns that reflect, for a single factor, different contexts prone to be abandoned.

Introducción

1. INTRODUCCIÓN

Durante las últimas décadas, el desarrollo e implementación de las ciencias y tecnologías de la información geográfica han contribuido de forma notable e inequívoca al conocimiento de los sistemas naturales y socio-territoriales. No se podría entender de otra forma el progreso realizado en la definición y cuantificación de las estructuras y los procesos naturales y antrópicos, tanto a escala local como regional o global. A ello hay que añadir el esfuerzo de las instituciones públicas por ofrecer en acceso abierto numerosa información de tipo ambiental y socioeconómico, y el desarrollo de sensores tanto *in-situ* como *ex-situ* que han permitido la elaboración de numerosas bases de datos. El incremento de bases de datos espaciales y otro de tipo de información estadística a disposición de la comunidad científica ha ido en paralelo con el aumento de la capacidad de procesamiento computacional. Todo ello ha favorecido el desarrollo de nuevos algoritmos muy eficaces en el análisis de grandes volúmenes de datos. Dichos algoritmos, diseñados sobre la base de métodos estadísticos, han permitido individualizar patrones robustos de comportamiento de fenómenos naturales y socioeconómicos. El resultado son modelos que definen de manera más fiable los fenómenos geográficos, dada la representatividad espacial de los datos, variedad de factores tenidos en cuenta y sofisticación del algoritmo empleado.

Si bien la cobertura de datos para grandes superficies del territorio es fundamental para determinar patrones espaciales a escala regional o continental, el registro de su variabilidad temporal ayuda a entender su evolución y pronosticar su dinámica futura, es decir, el análisis de cambios en el patrón espacial de un fenómeno a lo largo del tiempo. La limitada disponibilidad de observaciones empíricas registradas cronológicamente y sistemáticamente hasta hace escasas décadas, hizo que los estudios sobre el medio natural o socioeconómico fueran fundamentalmente de carácter sincrónico (Masalles & Vigo 1987). En dichos estudios, los cambios en un fenómeno geográfico determinado, como pueda ser el patrón de distribución de la vegetación, se estudiaban atendiendo a las diferentes fases de evolución encontrados a lo largo y ancho de un territorio para un momento concreto. De esta forma, la particularidad del contexto ambiental o socioeconómico en cada fase encontrada permitía establecer conclusiones acerca de sus características. Con el acceso en abierto de series temporales de información estadística y el desarrollo de planes de observación continua de la superficie terrestre (como puedan ser los planes nacionales para la generación de ortofotos, o el lanzamiento de satélites de observación de la Tierra) han aumentado las aportaciones científicas de carácter diacrónico, es decir, aquellas que analizan un determinado fenómeno a través de su evolución temporal, pudiendo conocerse así las condiciones en las que se producen determinados cambios.

Una de las aproximaciones más relevantes en la reciente literatura científica, son los modelos predictivos basados en el análisis de la dinámica de los usos y cubiertas del suelo (Verburg et al. 2015; Maglioca et al. 2015). Estos trabajos muestran la presión ejercida sobre el planeta, teniendo en cuenta que los cambios de cubiertas y usos del suelo es uno de los factores clave del cambio global ambiental que está acaeciendo, especialmente en aspectos relacionados con el albedo terrestre, ciclo del carbono, clima, etc. (Lambin et al. 2003; Verburg et al. 2015). Así mismo, la evolución del modelo de ocupación humana del territorio se traduce en perturbaciones en la estructura del paisaje, con consecuencias en la biodiversidad y el patrimonio paisajístico derivado de los usos históricos de la tierra (Farina 2008). Las variaciones en la distribución espacial de cubiertas del suelo como la vegetación, responden a las diferentes condiciones ecológicas del territorio, a la intersección con los usos antrópicos y a los cambios en el tiempo en ambos factores, ya sea a causa de agentes externos o por la propia dinámica natural. En este sentido, los análisis

retrospectivos permiten identificar las condiciones impulsoras de estos cambios, tanto en lo que se refiere a la dinámica de las cubiertas naturales (vegetación, masas de hielo, suelo desnudo, etc.) como a la de los usos antrópicos (cultivos, áreas urbanizadas, etc.). Como ejemplo cabe destacar la atención que algunos estudios han prestado a la dinámica del clima y su influencia en la evolución de las cubiertas del suelo, y en especial, en la dinámica de la vegetación (Walther et al. 2002; Schuur 2003; Kelly & Goulden 2008).

Buena parte de los cambios en la estructura de los sistemas naturales y, en general, del paisaje, responden a la dinámica de la vegetación y de los usos del suelo. El conocimiento preciso de los factores que controlan esta dinámica es esencial para entender la configuración del paisaje y el modelo socio-territorial en un determinado momento. Son las perspectivas relacionadas con la ciencia del paisaje las que de manera destacada han contribuido en este sentido. En este tipo de estudios, factores de tipo ambiental y socioeconómico son incluidos como variables explicativas de los cambios en el paisaje (Schneeberger et al. 2006; Mottet et al. 2006; Serra et al. 2008; Álvarez-Martínez et al. 2011; Pazur et al. 2014; Lieskovsky et al. 2016). En función de la respuesta a dichas variables se pueden extraer patrones generales de cambios entre usos y cubiertas del suelo (Burgi et al. 2004), es decir, qué factores tienen más peso y de qué manera influyen en dichos cambios o transiciones. Dado el carácter diacrónico del enfoque, la determinación de dichos patrones puede utilizarse para prever dinámicas futuras en función del cambio pronosticado en alguna de las variables explicativas. Por ejemplo, si la actual distribución espacial de la vegetación responde al gradiente altitudinal de la temperatura, el incremento o descenso pronosticado de las temperaturas podría explicar la transición entre dos tipos fisionómicos de vegetación en función de su tolerancia a las nuevas condiciones de temperatura. Otro ejemplo lo encontramos en el caso de un modelo que relacione el descenso (entre dos momentos temporales) de la carga ganadera extensiva con la transición de cubiertas herbáceas a cubiertas leñosas. Podría entonces valorarse la implantación de un plan estratégico que potencie la ganadería extensiva; ya que el recrecimiento de la vegetación leñosa reduce la escorrentía superficial, lo que a su vez influye en el rendimiento hídrico de la cuenca (López-Moreno et al. 2011; García-Ruiz et al. 2011), del que dependen, entre otras cosas, los sistemas de abastecimiento de agua. En definitiva, se trata de modelos que permiten cuantificar la probabilidad de ocurrencia de un cambio de uso o cubierta del suelo a partir de la influencia, positiva o negativa, de una serie de factores potencialmente explicativos. El hecho de registrar la variabilidad espacial y temporal de los factores permite, además, adaptarse a futuros escenarios, mejorando así la eficacia de cualquier proceso de toma de decisiones en el ámbito de la planificación territorial.

1.1. Detección de cambios entre usos y cubiertas del suelo

Una de las metodologías más empleadas para registrar la dinámica de las cubiertas del suelo es la generación de cartografía en diversos momentos temporales, de tal forma que puedan identificarse cambios entre un momento y otro. La interpretación de fotografías aéreas es una técnica comúnmente utilizada, especialmente cuando se dispone de una amplia cobertura histórica de imágenes aéreas. Esta metodología ha permitido obtener información detallada sobre los tipos de cubiertas del suelo y de los cambios en su patrón espacial para períodos de estudio de en torno a 40-50 años (Allen et al. 1998; Sluiter & de Jong, 2007; Alados et al. 2004; Gerard et al. 2010). No obstante, la interpretación visual de fotogramas aéreos es una tarea que consume un tiempo considerable si han de analizarse grandes superficies de territorio, lo cual es esencial para definir tendencias a escala regional o continental. En cambio, las metodologías automáticas o semi-automáticas basadas en el procesamiento de imágenes de satélite

ofrecen una visión más completa de la distribución de las cubiertas del suelo. La teledetección permite generar cartografía para grandes extensiones de terreno en tiempos de procesamiento relativamente cortos, aunque su resolución espacial (unidad mínima de análisis, habitualmente definida como el tamaño de píxel) ha sido hasta hace relativamente poco menor que la fotografía aérea (por ejemplo, desde los 10 m de lado de píxel en las imágenes proporcionadas por el satélite Sentinel o 30 m las del satélite Landsat, en comparación con la resolución de 0.5-1 m que pueden llegar a proporcionar los clásicos fotogramas aéreos); de ahí, y del hecho que la interpretación visual acostumbre a permitir la diferenciación de numerosas clases basadas en usos (no en cubiertas), que el nivel de desagregación de categorías de usos y cubiertas del suelo suela ser menor comparado con las leyendas derivadas de la fotointerpretación. Otras plataformas, como es el caso del satélite Ikonos, proporcionan imágenes con una resolución espacial de entre 1 y 4 m, pero no disponen de de una serie temporal de imágenes (por el momento en que se pusieron en órbita) tan amplia como la que, por ejemplo, proporciona el satélite Landsat (en órbita desde la década de los 70).

No obstante, las técnicas basadas en la teledetección espacial registran la radiación electro-magnética de las cubiertas terrestres en un amplio rango de longitudes de onda (Chuvieco 1998, 2002; Pons & Arcalís 2012), en comparación con las ortofotos aéreas, de las que generalmente sólo puede derivarse la correspondiente a la longitud de onda del espectro visible[1]. En este sentido, en diversos estudios sobre biodiversidad, como los realizados por la Agencia Europea del Medio Ambiente (EEA, 2014), se deduce que una mayor resolución espectral es mucho más importante que una mejor resolución espacial. Una mayor resolución espectral ayuda a individualizar la fisionomía y fisiología de los diferentes tipos de vegetación a partir de sus correspondientes firmas espectrales (respuesta espectral de una cubierta del suelo a lo largo del rango de longitudes de onda) y, al mismo tiempo, permite diferenciar estados fenológicos si se cuenta con una serie temporal intraanual de imágenes de satélite, algo casi nunca disponible en el caso de imágenes aéreas.

Existen diversos productos derivados del procesamiento de imágenes de satélite para el análisis de la dinámica de cubiertas y usos del suelo a escala regional y global. Uno de ellos es el proyecto *Climate Change Initiative Land Cover Dataset*[2], de la Agencia Espacial Europea, para los años 2000, 2005 y 2010, utilizando el sensor MERIS (Arino et al. 2007; Kirches et al. 2013); y también, los productos anuales del sensor MODIS para el periodo 2001-2013 (Strahler et al. 1999; Friedl et al. 2010). Sin embargo, el inconveniente de estos productos es que tienen una resolución espacial moderada (500 m en MODIS y 300 m en MERIS en nadir, siendo mayor en los extremos de la imagen), y tal tamaño de píxel no permite una comprensión detallada de la dinámica de las cubiertas terrestres para, por ejemplo, capturar cambios graduales en el espacio en intervalos de 1-5 años. Se necesita un tamaño de píxel menor para detectar fenómenos de dimensión espacial más modesta y para identificar fases intermedias dentro de los cambios fluctuantes en el paisaje (Mucher et al. 2000), sobre todo en áreas con una elevada fragmentación del paisaje, como es el caso del ámbito mediterráneo. Aunque las imágenes capturadas por el sensor MODIS, a bordo de los satélites TERRA y AQUA, tienen una repetición de órbita de entorno a 2 días (es decir, en torno 182 imágenes al año para una misma localización) para latitudes medias (hacia los polos la disponibilidad de imágenes es mayor, incluso de varias al día, debido al solapamiento de las órbitas), los sensores del

[1] El espectro visible tiene un rango espectral de 0.4 a 0.7 µm de longitud de onda, comprendiendo las bandas azul, verde y rojo del espectro electro-magnético, mientras que, por ejemplo, el sensor del satélite Landsat registra bandas espectrales que van desde el espectro visible hasta los 12500 µm de la banda del infrarrojo térmico.
[2] https://www.esa-landcover-cci.org/ o http://maps.elie.ucl.ac.be/CCI/viewer/index.php

satélite Landsat (con una repetición de órbita de en torno a 16 días) permiten identificar con mayor especificidad los patrones espaciales, dada la resolución espacial de sus imágenes (píxel de 30 m). Una resolución temporal alta no es tan importante como una resolución espacial alta a la hora de generar productos de cambio a escala regional o continental, teniendo en cuenta que los principales vectores de cambio se identifican en el intervalo de 1-5 años (Wulder et al. 2008). Las imágenes proporcionadas por la constelación de satélites SPOT tienen una resolución de hasta 10 m de lado de píxel en las imágenes multiespectrales (llegando a 5 m en el pancromático monoespectral del SPOT 5), aunque se trata de un satélite comercial y la adquisición de las imágenes es a demanda, por lo que el registro de la serie temporal de éstas no tiene la regularidad que dispone el satélite Landsat. Con Landsat, la disponibilidad potencial es de aproximadamente una imagen cada 16 días, lo que aumenta la posibilidad de tener mayor número de imágenes sin cobertura de nubes. Además, esta frecuencia de registro es suficiente para capturar las diferencias fenológicas a lo largo del año en la vegetación y cultivos. La base de datos del organismo *NASA-USGS Global Land Survey* (Gutman et al. 2013), compuesta principalmente por imágenes Landsat y disponible de forma gratuita para la comunidad científica y el público en general desde 2008, es una fuente de información muy importante para la cartografía de detalle desde el año 1972 hasta la actualidad. El acceso a estos datos permite realizar estudios de cambio regional o global a una resolución relativamente detallada, y facilita una mayor comprensión de la complejidad de la dinámica del paisaje (Vogelmann et al. 1998; Homer et al. 2004; Wulder et al. 2008; Pons et al. 2014).

1.1.1. Técnicas basadas en el procesamiento de imágenes de satélite

La monitorización de los ecosistemas a partir de series temporales de imágenes de satélite puede abordarse mediante diferentes esquemas metodológicos (Coppin et al. 2004), dentro de los cuales uno de los aspectos más importantes es la técnica de detección de cambios (Lu et al. 2002; Coppin et al. 2004). Para esta finalidad pueden establecerse dos enfoques distintos: las técnicas que identifican los cambios entre dos momentos temporales y las que definen trayectorias temporales. Las primeras pueden basarse en la transición entre dos cubiertas de suelo distintas, de tal forma que una de ellas sustituye a la otra (por ejemplo, la sustitución de un bosque de pinos por uno de encinas, o el cambio de la especie dominante en un bosque compuesto por estas dos especies); o también pueden basarse en la detección de modificaciones, donde la clase de cubierta puede ser la misma, pero uno de sus atributos ha podido cambiar entre las dos fechas, identificándose así un tipo de perturbación más gradual (por ejemplo, la variación en el contenido de humedad de un bosque, el grado de defoliación del mismo entre una fecha y otra, el aclarado del bosque, etc). La técnica denominada como post-clasificación es una de las más habituales en la detección de transiciones entre categorías diferentes de cubiertas del suelo. Dicha técnica se basa en la obtención de una clasificación de cubiertas del suelo para cada fecha, en la que la asignación de categorías temáticas puede ser a partir del vector estadístico de los píxeles (clasificación “píxel a píxel”), o a partir de la segmentación de la imagen en base a objetos (agrupaciones previas de píxeles que presentan una correlación espacial en cuanto a su rango espectral, textura, etc., y que después se vinculan con una categoría temática) (Chen et al. 2012). Posteriormente se realiza una intersección de los dos mapas resultantes para identificar los cambios entre unas cubiertas y otras. La interpretación de los resultados pues, es relativamente fácil, dado que se trata de categorías conocidas de una leyenda previa. En cuanto a las técnicas basadas en la detección de modificaciones, una de las más utilizadas es el análisis del vector de cambios (CVA, *Change-Vector Analysis*), que transforma la información espectral de las imágenes de

satélite en las variables físicas de brillo, verdor y humedad, también denominada como transformación *Tasseled Cap* (Kauth et al. 1976). Una vez extraídos estos componentes para las dos fechas, el CVA identifica la dirección del cambio (en términos de componente) y la magnitud o intensidad de dicho cambio (Lambin & Strahler 1994; Baker et al. 2007). En esta línea, durante los últimos años se han desarrollado algoritmos más sofisticados (Keogh et al. 2001; Hermosilla et al. 2016) y aunque, al igual que CVA, son técnicas que cuantifican de manera efectiva los cambios en relación a la respuesta espectral de las cubiertas, requieren de interpretaciones adicionales para seleccionar los umbrales de magnitud relacionados con las tipologías de cambio. Por otro lado, las técnicas basadas en trayectorias temporales analizan las cubiertas terrestres a partir de perfiles o curvas de evolución de alguno de sus atributos, tales como la reflectancia o el estado fenológico a partir del *Normalized Difference Vegetation Index* (NDVI). Dicha técnica requiere de una densa secuencia temporal de imágenes para definir con solidez las curvas de tendencia en cada píxel y, al igual que en el caso anterior, la asignación de una tipología de perturbación depende de la definición de una serie de umbrales de magnitud, en este caso ponderados por la tendencia de la serie temporal completa.

En el caso de la clasificación estadística de imágenes de satélite, método utilizado en la presente tesis para la generación de mapas de cubiertas del suelo en diversas fechas, la calidad de los resultados depende de diversos factores. En este sentido, la selección del tipo de clasificador se considera un aspecto no tan decisivo como otras partes del esquema de trabajo (Lu & Weng 2007). En cambio, la determinación de las variables más adecuadas para que el clasificador pueda discriminar estadísticamente los tipos de cubiertas del suelo es una fase fundamental. Complementariamente a las variables espectrales, constituidas por las imágenes de satélite, y las variables que pueden calcularse a partir de éstas (índices de vegetación y humedad o los componentes derivados de la transforación *Tasseled Cap*) pueden incluirse otras de tipo topo-climático tales como la altitud, la pendiente o la radiación solar. Tal volumen de datos hace necesario identificar un conjunto idóneo de variables para, por un lado, reducir los tiempos de procesamiento del clasificador y, por otro lado y lo que es más importante, eliminar variables correlacionadas estadísticamente. La presencia de colinealidad entre variables puede afectar a las tasas de acierto de la clasificación, que es la proporción de píxeles bien clasificados, ya sea del conjunto de la imagen (acierto global) o de alguna categoría específica. Por lo tanto, es recomendable aplicar algún procedimiento para determinar qué variables de un conjunto inicial son las más útiles, y cuáles son las que pueden eliminarse debido a que no introducen mejoras en cuanto al acierto global y/o por categorías. Otro aspecto importante es la determinación de los píxeles que conforman las áreas de entrenamiento y de test necesarias para hacer la clasificación y validar los resultados. Las áreas de entrenamiento y test constituyen la información de referencia denominada “verdad-terreno”, es decir, información ya existente a partir de la cual definir las clases temáticas de la leyenda. En este caso la verdad-terreno se ha obtenido de bases oficiales de cubiertas del suelo, aunque en otros casos se pude también obtener mediante trabajo de campo, fotointerpretación, o una combinación de diversos de éstos métodos. A partir de esta información de verdad-terreno ha de extraerse un conjunto de píxeles representativos de cada cubierta del suelo, asegurando cierta homogeneidad en la respuesta espectral de cada cubierta, es decir, tratando de individualizar el patrón estadístico (según los valores de las variables) que mejor la define. Una parte de estos píxeles se utiliza para la fase de entrenamiento del clasificador, o como conjunto de patrones estadísticos a partir de los cuales clasificar el resto de la imagen, y otra para la validación de resultados (test). Una cuestión que es fundamental abordar es el aprovechamiento de estas áreas de entrenamiento y test para la clasificación de fechas que no disponen de información de referencia, o fechas en que la información de referencia disponible puede no tener la misma calidad que la utilizada en otras fechas. La

intención es identificar los píxeles de esas áreas de verdad-terreno que mantienen un mismo patrón estadístico (o muy similar en términos de distancia estadística) a lo largo de las fechas que componen el periodo de estudio y, por tanto, se pueden usar como verdad terreno en fechas para las que no se dispone de otra base de datos de referencia contemporánea a las imágenes de la cual extraer las áreas de entrenamiento y test. Todos estos pasos requieren el máximo grado de automatización dado el volumen de datos que se gestiona. Finalmente, cada clasificador dispone de un cierto grado de flexibilidad a la hora de aplicar su algoritmo en base a una serie de parámetros y opciones específicas. Por tanto, ha de encontrarse una configuración de parámetros que optimice la fase entrenamiento y los resultados en la fase de validación.

La clasificación de imágenes de satélite puede proporcionar datos de forma masiva acerca de la distribución de las cubiertas y usos del suelo, aunque se ha de prestar atención a la posible incertidumbre de estos productos. Esto supone evitar, en la medida de lo posible, la inclusión de píxeles mal clasificados en la fase de análisis de cambios entre cubiertas del suelo. Una vez tenido en cuenta este aspecto, podrán obtenerse áreas más fiables de transición o cambio entre unas cubiertas y otras a partir de la superposición de las clasificaciones.

1.2. Modelo estadístico para la interpretación cambios

Si bien el método de obtención de mapas de cubiertas del suelo es importante para identificar las zonas de cambio o transición, también se ha de contar con un método que explique las condiciones que han favorecido dichas transiciones. Previamente, la información georreferenciada referente a los posibles factores explicativos (obtenida a partir de fuentes institucionales, otros trabajos de investigación, derivada de modelos digitales de elevaciones, etc.) ha de vincularse a las zonas de cambio. Normalmente, la relación entre los posibles factores explicativos y una determinada transición entre dos cubiertas es determinada mediante análisis estadísticos basados en la regresión lineal, tales como la regresión logística (Serneels and Lambin 2001; Serra et al., 2008). Dicho análisis se basa en la predicción de una variable respuesta (variable dependiente) de tipo binomial (normalmente el valor 1 se vincula a las observaciones en las zonas de cambio, y el valor 0 a las observaciones en zonas estables o de no-cambio) a partir de los valores observados en una serie de variables o factores potencialmente explicativos (variables independientes), obteniéndose así la probabilidad de ocurrencia, en este caso, de un cambio entre dos usos o cubiertas del suelo. Otros enfoques menos utilizados son los basados en las cadenas de Markov (Balzter 2000; Camacho et al. 2013), que también modelizan la probabilidad de transición entre dos cubiertas del suelo, pero que implican normalmente aproximaciones más supervisadas (Camacho et al. 2013). En la presente tesis, la importancia de los factores y cómo éstos influyen en las transiciones entre cubiertas ha sido analizada mediante *Boosted Regression Trees* (BRT), método también conocido como *stochastic gradient boosting* (Friedman 2001, 2002; Hastie et al. 2009), en el que también se modela una variable respuesta en relación a una serie de variables explicativas. Este método relativamente nuevo basado en técnicas *machine-learning* (Breiman 2001) ha demostrado tener una alta capacidad de predicción (Kawakita et al. 2005; Elith et al. 2006, 2008; Leathwick et al. 2006; Crase et al. 2008), fundamentalmente por las siguientes razones: es un enfoque que no asume ninguna distribución o modelo de datos, sino que busca determinar patrones dominantes de respuesta a los factores combinando muchos árboles de clasificación (Breiman et al. 1984); identifica variables relevantes y relaciones e interacciones complejas; está mucho menos influenciado por información correlacionada o variables irrelevantes que otros enfoques estadísticos;

produce predicciones estables (reducción de la varianza); y proporciona representaciones gráficas de la relación entre la variable respuesta y los factores mostrando, en el caso de la dinámica de las cubiertas, dónde están las condiciones más o menos favorables (en relación al rango de valores de los factores explicativos) para que se produzca un determinado tipo de transición. Este método es particularmente útil para escenarios multi-causales, dada su capacidad para definir el papel desempeñado por los factores explicativos, incluso cuando la contribución de alguno de ellos es baja.

BRT es un método basado en árboles de decisión combinado con la técnica *boosting* (Breiman et al. 1984; Hastie et al. 2009), y que representa una optimización del método *Random Forest* (Breiman 1996; James et al. 2013). El algoritmo de árbol de decisión divide el rango de valores de una variable explicativa a partir de un punto de corte óptimo, para encontrar así regiones donde la probabilidad de ocurrencia de la variable respuesta (en este caso el evento de cambio) sea la máxima. Pueden combinarse diversas variables, de tal forma que haya diferentes regiones y subregiones a partir de diferentes puntos de corte. Mediante BRT, cientos o miles de árboles de decisión (también denominados como árboles de clasificación) se ajustan de manera secuencial y progresiva. Con el algoritmo de *boosting*, cada árbol nuevo se genera usando una fracción aleatoria de las observaciones (observaciones distribuidas aleatoriamente a lo largo las zonas de cambio) y un subconjunto distinto de variables explicativas. A diferencia de *Random Forest*, los árboles no se ajustan de forma individual, sino de manera aditiva, ya que cada nuevo árbol es ajustado sobre los residuales del árbol anterior. El procedimiento para ajustar secuencialmente árboles de decisión en un sólo modelo se basa en una serie de reglas que ponderan las observaciones en función de su tasa de error después de haber ajustado el árbol anterior (Friedman 2002; Hastie et al. 2009).

La presente tesis utiliza la clasificación de imágenes de satélite y los modelos BRT para analizar la dinámica de las cubiertas del suelo en extensas áreas de la Península Ibérica, centrando la investigación en los factores determinantes en la transición entre diferentes tipos fisionómicos de vegetación y en el proceso de abandono agrícola. El núcleo central de la tesis se compone de tres artículos, dos de los cuales están publicados en revistas ISI (*International Scientific Indexing*) del primer cuartil, y un tercero que en el momento del depósito de la tesis está aceptado también en una revista ISI del primer cuartil, pero todavía no publicado. Los artículos son los siguientes:

- *Developing spatially and thematically detailed backdated maps for land cover studies*. Revista: *International Journal of Digital Earth.*

- *Factors affecting forest dynamics in the Iberian Peninsula from 1987 to 2012. The role of topography and drought*. Revista: *Forest Ecology and Management.*

- *Environmental and socioeconomic factors of abandonment of rainfed and irrigated crops in northeast Spain*. Revista: *Applied Geography.*

En los apartados 2.1 y 2.2 se presentan de forma sintetizada los aspectos más importantes de las metodologías empleadas. Los apartados 3.1 y 3.2 contextualizan la importancia de los casos de estudio referentes a la dinámica de la vegetación y el abandono de cultivos, respectivamente, así como el interés de las zonas de estudio. Finalmente, el apartado 4 presenta los objetivos de la tesis a partir de cada uno de los artículos que la componen.

Metodología

2. METODOLOGÍA

La Figura 1 sintetiza el esquema metodológico que se describe a continuación, así como otras partes del flujo de trabajo que están descritas con mayor detalle en los artículos de investigación.

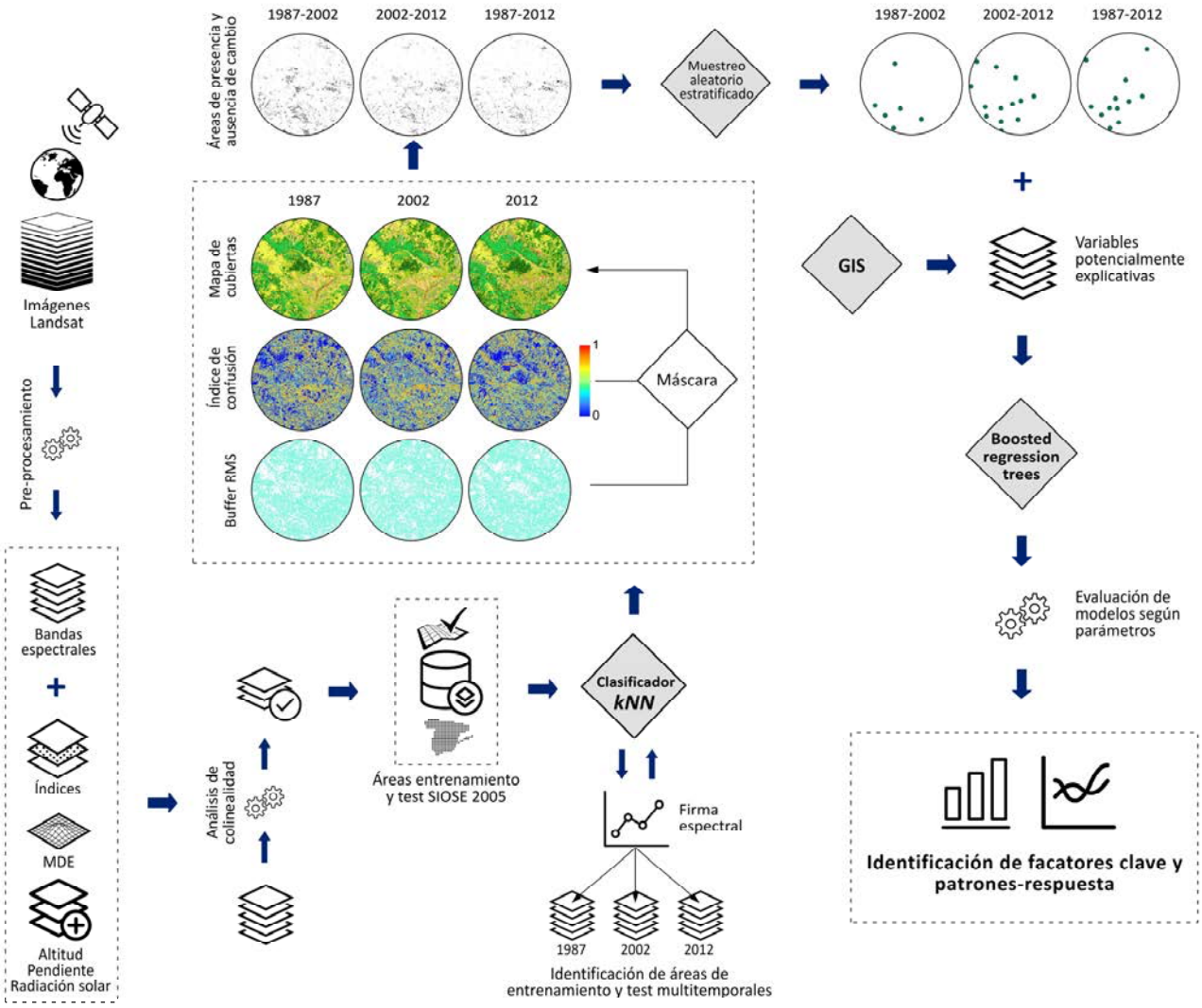


Figura 1. Síntesis del esquema metodológico (Iconos negros diseñados por Noun Project (2017)).

2.1. Obtención de mapas de usos y cubiertas del suelo

En este apartado se explica de forma sintetizada el esquema de clasificación y metodología seguida para la generación de series temporales de mapas de usos y cubiertas del suelo. El flujo de trabajo detallado queda recogido en el primer artículo científico de esta tesis: *Developing spatially and thematically detailed backdated maps for land cover studies*. Información detallada referente a las áreas de estudio, fechas que componen la serie temporal, descripción del algoritmo de clasificación y parámetros que lo controlan, puede encontrarse en dicho artículo.

Cada mapa generado hace referencia a un periodo de 5 años, por lo que cada clasificación está compuesta por un conjunto de imágenes correspondientes a diversas fechas, distribuidas desde abril a septiembre, con el objetivo de registrar la variabilidad fenológica de los tipos fisionómicos de vegetación y de los cultivos. Complementariamente a las imágenes espectrales del satélite Landsat (bandas 2, 3, 4, 5 y 7), se incluyeron en el análisis las siguientes variables auxiliares: altitud, pendientes, radiación solar, y los índices NDVI (*Normalized Difference Vegetation Index*), NDWI2 (*Normalized Difference Water Index*) y SAVI (*Soil Adjusted Vegetation Index*). En total se partió de un conjunto inicial de 30 bandas radiométricas, 18 índices derivados de éstas y las tres variables de carácter topo-climático (51 variables en total).

La corrección radiométrica de las imágenes de satélite se realizó siguiendo la metodología de Pons et al. (2014) que, entre otros parámetros, utiliza áreas seudo-invariantes (PIA, *pseudo-invariants areas*) como referencia espectral, en este caso obtenidas por el equipo de investigación Grumets (Universitat Autònoma de Barcelona) a partir de productos de reflectancia del satélite MODIS (disponibles desde el siguiente link: http://www.ogc3.uab.cat/DINACLIVE/WMS/descarregues_ESP.html).

La información de referencia para generar los píxeles de entrenamiento y test fue el Sistema de Información sobre Ocupación del Suelo de España 2005 (SIOSE). La obtención de dichos píxeles se hizo mediante un Sistema de Información Geográfica (SIG) en un procedimiento semi-automático que constó de tres fases. En una primera fase, a partir de criterios de selección basados en sentencias SQL (*Structured Query Language*) se extrajo un primer conjunto de polígonos SIOSE. De este modo se identificaron los polígonos con coberturas más puras, estableciendo como tales aquellos que tienen porcentajes de recubrimiento suficientemente representativos. En una segunda fase se realizó un filtrado de los polígonos a partir de umbrales basados en percentiles de distribución estadística de los valores de NDVI. Los umbrales utilizados se definieron de forma general para cada cubierta, en función del porcentaje de recubrimiento seleccionado en la primera fase, pero se aplicaron por separado a la distribución de los valores de los píxeles que había dentro de cada polígono (dado que para un mismo porcentaje de recubrimiento, el rango del NDVI puede variar según el contexto topo-climático). La finalidad es evitar la posible mezcla espectral de diferentes cubiertas dentro de un mismo polígono. En la última fase se utilizó el algoritmo kNN (*k-nearest neighbor*) del SIG MiraMon (Pons 2006) para identificar píxeles (de los obtenidos en la anterior fase) que mantienen un patrón estadístico similar en periodos en los que no se dispone de información de referencia (Figura 2), por lo que ha de compararse un mismo patrón de fechas y variables (V_1, V_2, V_3... V_n) entre periodos (T_{0-1}, T_0, T_{0+1}). Dicho análisis puede aplicarse de forma retrospectiva o prospectiva y permite obtener el subconjunto de píxeles de entrenamiento y de test del periodo de referencia (T_0) que es adecuado utilizar como referencia en todos los periodos (T_{0-1}, T_0, T_{0+1}).

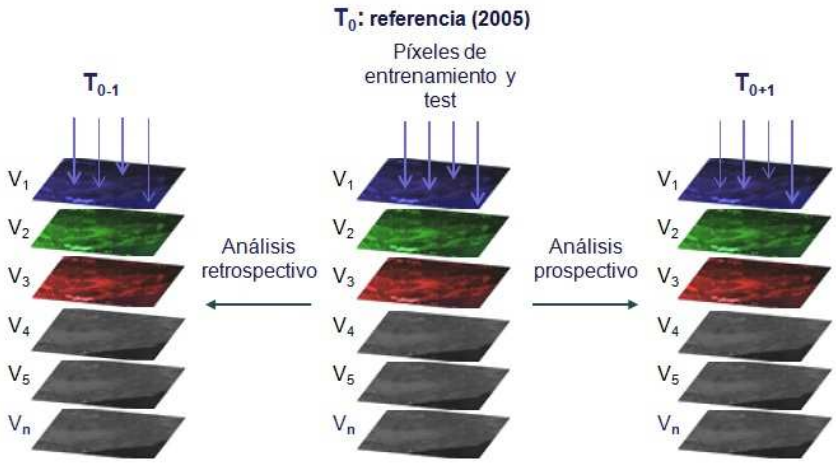


Figura 2. Análisis retrospectivo y prospectivo de los píxeles de entrenamiento y test.

El algoritmo kNN calcula la distancia euclidiana (en el espacio estadístico) entre cada píxel objeto de ser clasificado y los píxeles de referencia (de entrenamiento y test) de los que se conoce la categoría a la que pertenecen. Una vez computadas las distancias de cada píxel a cada uno de los pixeles de referencia, se selecciona un número de *k* píxeles, que en este caso son los *k* píxeles más cercanos. Por ejemplo, *k*=10 querría decir que para cada pixel se seleccionarían los 10 píxeles de referencia más cercanos, y dentro de éstos, el pixel analizado se clasificará en la categoría modal (la categoría que más se repite). Una segunda opción o variante del algoritmo, es que esos 10 píxeles más cercanos se ponderen por el inverso de su distancia estadística al cuadrado (*W*): $W=1/d(x, x')^2$, donde *d* es la distancia euclidiana entre el píxel a ser clasificado (*x*) y el píxel de referencia (*x'*); por lo que el número de veces que una categoría se repite dentro de esos *k* píxeles se pondera por *W* (por ejemplo, tres repeticiones de una categoría: $[1/d(x, x'_1)^2]+[1/d(x, x'_2)^2]+[1/d(x, x'_3)^2]$), siendo la categoría que mayor valor obtenga la categoría "ganadora". En la presente tesis se ejecutaron diversas pruebas, utilizando los píxeles de entrenamiento y test obtenidos en la segunda fase de filtrado, para determinar cuál de las dos opciones optimizaba los resultados en cuanto a porcentajes de acierto. Los resultados mostraron diferencias muy poco apreciables, también en lo que respecta al aspecto visual de los mapas, por lo que el conjunto de clasificaciones se llevó a cabo utilizando el método de la moda, dado que esta variante supone un menor tiempo de procesado en comparación con el cálculo del inverso de la distancia. Para la determinación del valor *k* óptimo se ejecutaron, así mismo, un banco de pruebas que queda descrito en el correspondiente artículo de investigación.

Previo al proceso de clasificación final, se realizó un análisis de colinealidad de las 51 variables basado en el cálculo de coeficientes de correlación. Una vez identificado el conjunto óptimo de variables se procedió a hacer la clasificación definitiva. Inicialmente, el proceso de clasificación se basó en un esquema híbrido constituido por una fase inicial no supervisada (IsoMM del SIG MiraMon), en la que se generan agrupaciones o clústeres de píxeles (en el espacio espectral) a partir de unos centroides iniciales (mediante un proceso iterativo basado en una optimización del algoritmo IsoData), y una fase supervisada (ClsMix de MiraMon) en la que dichas agrupaciones de clústeres espectrales son asignadas a categorías temáticas de cubiertas del suelo según unos parámetros de representatividad estadística. Los resultados también se compararon con otros clasificadores de las plataformas Envi (2012) y ArcGIS (ESRI 2012).

A partir de las áreas de test, la validación de los resultados se hizo mediante matrices de confusión, en las que se cuantifica el acierto global, así como el acierto por categorías en base a la exactitud del productor y del usuario (Pons & Arcalís 2012).

Con posterioridad a la publicación del artículo dedicado a esta metodología, la fase de clasificación definitiva se continuó llevando a cabo con el clasificador kNN de MiraMon, ya que tras su paralelización (opción que permite utilizar más de un procesador en ordenadores de 32 y 64 bits durante su ejecución) por parte del equipo de Grumets, los tiempos de procesado se redujeron a menos de una quinta parte (en ordenadores de 64 bits), mientras que los porcentajes de acierto se mantenían o aumentaban ligeramente, tanto a nivel global como por categorías respecto al esquema híbrido de clasificación.

La mayoría del flujo de trabajo fue automatizado mediante su integración en archivos de proceso por lotes (archivos tipo *batch*), dado que las funciones del SIG MiraMon pueden configurarse mediante comandos basados en la consola CMD de los sistemas operativos Windows de Microsoft.

En total se clasificaron 5 escenas Landsat para los siguientes quinquenios: 1985-1989, 1991-1995, 2000-2004 y 2010-2014; por lo que se obtuvieron un total de 20 mapas de usos y cubiertas del suelo (4 fechas por cada escena). No obstante, con la intención de sintetizar los resultados, en los artículos de investigación no se han incluido los datos referentes a las clasificaciones de 1991-1995, aunque en la Figura A1 del Anexo se muestran algunos ejemplos para este quinquenio. La Figura 3 muestra las escenas Landsat clasificadas referentes a los ámbitos de estudio de la tesis.

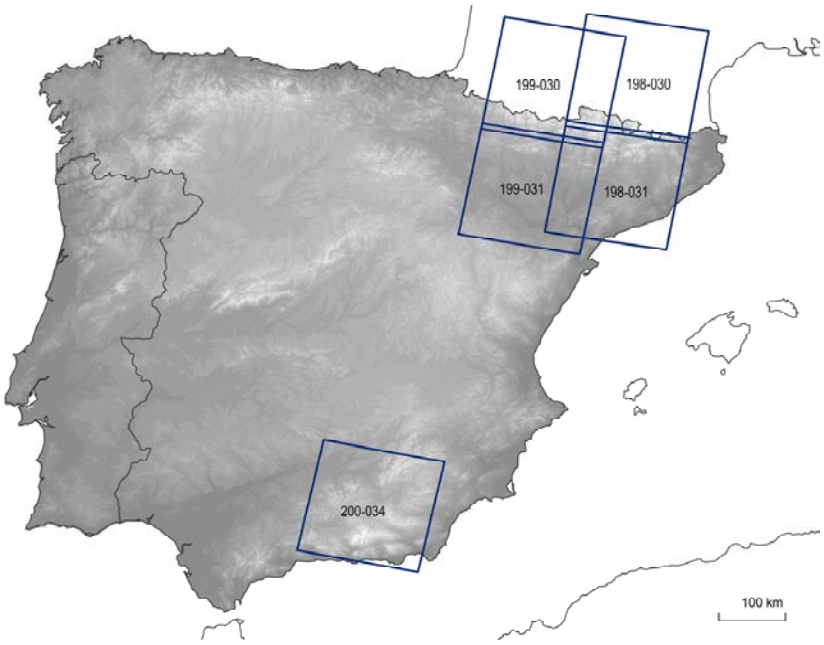


Figura 3. Escenas Landsat clasificadas. El código hace referencia al número de trayectoria y fila de la órbita.

2.2. Análisis de factores determinantes en los cambios entre cubiertas del suelo

2.2.1. Obtención de las zonas de cambio entre cubiertas

Una vez realizada la superposición de los mapas temáticos, los cambios de cubierta del suelo fueron considerados como eventos de ausencia/presencia de cambio. Las áreas donde la cubierta inicial permanecía estable entre dos fechas fueron tratadas como ausencias, y aquellas donde había un cambio de cubierta en la dirección a estudiar como presencias. Una vez establecidas estas áreas, se realizó un muestreo aleatorio por tipo de evento, intentando mantener la prevalencia entre eventos (mismo número de puntos de ausencias y presencias) para poder comparar modelos entre sí.

2.2.2. Variables explicativas

Se seleccionaron una serie de variables para incluirlas como factores explicativos (predictores) en los modelos. La intención es identificar las variables más importantes y describir, a través de inferencia estadística, la probabilidad de un cambio de cubierta según el rango de valor de las variables. El conjunto de factores potencialmente explicativos lo forman, entre otras, variables de tipo topográfico, variables de coste de desplazamiento a infraestructuras, variables relativas a información socioeconómica a nivel municipal, o variables que contabilizan la recurrencia de sequías. Las variables derivadas de la topografía (pendiente, radiación solar, coste de desplazamiento, etc.) se obtuvieron a partir de un modelo digital de elevaciones a 10 m de resolución (elaborado a partir de hojas topográficas a escala 1:5000), por un lado para el ámbito correspondiente a las regiones de Aragón, Cataluña y Navarra, y por otro, para el ámbito de las regiones de Andalucía, Castilla la Mancha y Murcia. El cálculo de las variables se realizó mediante scripts y herramientas propias de los SIG MiraMon 8.1 (Pons 2006), ArcGis 10.1 (ESRI 2012) y SAGA 2.1.2 (Conrad et al. 20015). Algunos ejemplos de estas variables se muestran en la Figuras A2, A3 y A4 del Anexo. Las variables de sequía se calcularon en base al índice SPEI (*Standardized Precipitation Evapotranspiration Index*) (Vicente-Serrano et al. 2010) a partir de los datos del Atlas Climático Digital de la Península Ibérica (Ninyerola et al. 2005) según la aproximación descrita en Domingo-Marimon (2016). El cálculo del número de eventos secos, según diferentes duraciones de sequía y periodos de tiempo, se realizó mediante herramientas del entorno de MiraMon.

Si bien la recurrencia de incendios puede considerarse como un factor explicativo en la dinámica de ciertas comunidades vegetales, especialmente en el ámbito mediterráneo, la representatividad espacial de las bases de datos que proporcionan esta información no facilita una localización precisa del foco y distribución de los incendios (de la Riva et al. 2004). De forma similar sucede con otras bases de datos que ofrecen información referente a las perturbaciones relacionadas con las plagas de insectos, así como la información relativa a la gestión forestal (ICP Forest Inventory 2006, IDF 2016).

Los puntos del muestreo en las zonas de cambio se vincularon a la información georreferenciada de todas las variables mediante un proceso automático utilizando funciones de ArcGIS. La base de datos resultante se exportó al correspondiente formato para realizar la fase de análisis dentro del *software* R (R Core Team 2017).

2.2.3. Ajuste de los modelos Boosted Regression Trees

Como se dijo en la introducción, *Boosted Regression Trees* (BRT) es un método *machine-learning* sobre la base de árboles de clasificación que son ajustados secuencialmente en un modelo aditivo mediante la técnica *boosting* (Friedman 2001; Friedman 2002). Los parámetros principales que controlan los modelos BRT son: 1) Complejidad de árbol clasificación, que es el número de subdivisiones o puntos de cortes permitido en cada árbol, y que a su vez se relaciona con el número de variables que interactúan entre sí; 2) Tasa de aprendizaje, que es un valor que actúa como las técnicas *shrinkage* en otros modelos lineales (Hastie et al 2009), es decir, escalando la contribución de cada nuevo término (de cada árbol nuevo en este caso) en el modelo aditivo mediante un factor de ponderación entre 0 y 1; y 3) Número óptimo de árboles para generar el modelo aditivo completo (generalmente entre cientos y miles), que dependerá en gran parte de los dos parámetros anteriores. Generalmente, tasas de aprendizaje bajas, entre 0.001 y 0.01, favorecen la adición de árboles más complejos (Elith et al. 2008; Hastie et al. 2009). La finalidad es encontrar una combinación de parámetros óptima que evite el sobreajuste del modelo, reduciendo así la varianza en los resultados, y obtener un buen grado de acierto tanto con las observaciones de entrenamiento como con las de validación. En esta tesis se generó un modelo para todas las combinaciones posibles entre las tasas de aprendizaje 0.001, 0.0025, 0.005, 0.01, 0.015 y 0.03, y complejidades del árbol de 1 a 6, lo que significa que se evaluaron un total de 36 modelos en cada transición de cubiertas. Para cada uno de estos modelos el número óptimo de árboles se identificó mediante validación cruzada (Elith et al. 2008) a lo largo del proceso iterativo en el que el algoritmo BRT va añadiendo árboles (en este caso de 50 en 50) al modelo aditivo.

El rendimiento predictivo de cada modelo se analizó mediante su desviación estadística o *deviance* (Friedman 2002; Ridgeway 2007; Elith et al. 2008; Hastie et al. 2009), que es una función que actúa como la suma del cuadrado de los residuos en otros modelos lineales. Se calculó la desviación estadística en el conjunto de observaciones de test y el modelo con menor desviación se seleccionó para interpretar los resultados. Adicionalmente, los resultados se analizaron utilizando la curva ROC (*Receiver Operating Characteristics*) y calculando su valor AUC (*Area Under the Curve*) en las observaciones de test. Este análisis permitió evaluar la solidez de los modelos seleccionados, verificando el equilibrio entre verdaderos-positivos (proporción de eventos de presencia de cambio que han sido correctamente identificados) y falsos-positivos (proporción de eventos de ausencia de cambio que han sido incorrectamente clasificados como eventos de presencia de cambio) (Hanley & McNeil 1982; Fawcett 2004). La curva ROC se utiliza para calcular y representar dicho balance, utilizando multitud de umbrales de probabilidad con el objetivo de encontrar el umbral óptimo que maximice la tasa de verdaderos-positivos y minimice la tasa de falsos-positivos (Pontius & Parmentier 2014), mientras que el valor de AUC es una métrica-resumen que cuantifica ese balance en un rango de 0 a 1, es decir, el rendimiento predictivo del modelo según el valor de AUC.

La importancia de cada variable en BRT es estimada contabilizando su contribución en el modelo aditivo, según el número de veces que dicha variable es seleccionada para construir un árbol en el modelo, y la mejora en términos de predicción que la variable proporciona (Friedman & Meulman 2003; Hastie et al. 2009). Para la interpretación de resultados se utilizaron los gráficos de dependencia parcial. Éstos son una salida gráfica del modelo mostrando la relación entre la variable respuesta y una variable explicativa después de haber tenido en cuenta la contribución promedio del resto de variables explicativas (Friedman 2001; Ridgeway 2004). Las funciones de ajuste de estos gráficos guardan relación con los puntos de corte

que se han seleccionado para determinar las regiones (en el espacio estadístico a lo largo del rango de valor de las variables) que maximizan la probabilidad de ocurrencia de los eventos de presencia (donde tiene lugar la transición) o de los eventos de ausencia (donde la cobertura inicial permanece estable entre dos fechas, por lo que la transición no tiene lugar), a medida que el algoritmo va construyendo los árboles de clasificación y añadiéndolos en el modelo aditivo. Por tanto, la mayor o menor idoneidad del valor de una variable depende del número de veces (dentro del modelo aditivo BRT) que este valor ha quedado en una región con una mayor proporción de eventos de presencia o en una región con una mayor proporción de eventos de ausencia. En estas funciones de ajuste, las predicciones se modelan a través del *logit* de la regresión logística (James et al. 2003; Elith et al. 2008), pero las predicciones también se pueden mostrar en un rango de probabilidad de 0 a 1. En cualquier caso, son unidades relativas. En definitiva, estos gráficos muestran cuáles son las condiciones más o menos favorables para que se produzca un determinado cambio de cubierta.

La automatización del proceso de evaluación y validación de los modelos BRT, así como la generación de las series de gráficos de dependencia, se hizo mediante el desarrollo de *scripts* basados en el lenguaje de programación del *software* R.

Dinámica de usos y cubiertas del suelo

3. ANÁLISIS DE LA DINÁMICA DE USOS Y CUBIERTAS DEL SUELO

3.1. Dinámica de la vegetación

El análisis de las transiciones entre diferentes tipos de vegetación (derivados de la leyenda de la clasificación de cubiertas del suelo) pretende dar una visión general de la sucesión ecológica y la competencia entre especies forestales en la franja mediterránea de la Península Ibérica. Dicha área geográfica tiene un especial interés dada la diversidad de condiciones topo-climáticas derivadas de su topografía y porque durante las últimas décadas una parte significativa de su superficie se ha visto afectada por un incremento de los eventos de sequía (IPCC 2007, Vicente-Serrano 2014).

La eficacia de la vegetación en la colonización del espacio depende de la influencia que los factores del medio tienen sobre las distintas especies. De esta forma, la dirección en la sucesión y las transiciones entre especies forestales responde a una serie de estrategias ecológicas (Blanco et al. 1997) mediante las cuales, los diferentes tipos fisionómicos de vegetación se adaptan a la variabilidad de los factores ambientales. En este sentido, el análisis de cada etapa o tipo de transición permite identificar qué tipo de vegetación es más dinámica o vulnerable en función de los diferentes contextos ecológicos. Analizar la sucesión o los patrones de cambio forestales puede ayudar a predecir nuevas situaciones y a definir protocolos más eficaces en la gestión del medio natural. En el presente trabajo de tesis se han analizado desde fases iniciales de sucesión, que parten de cubiertas herbáceas o arbustivas, hasta transiciones entre etapas relativas a estadios maduros de bosque (Apartado 3.1.1.).

En las montañas mediterráneas de la Península Ibérica la sucesión ecológica de la vegetación ha progresado tras el abandono general de las actividades tradicionales (García-Ruiz 1990; Lasanta-Martínez et al. 2005). La actividad antrópica y su gestión del territorio han sido y siguen siendo determinantes importantes en la dinámica de la vegetación (Gehrig Fasel et al. 2007; Lasanta-Martínez et al. 2005). No obstante, otros factores han de tenerse en cuenta para entender mejor la colonización del espacio por parte de la vegetación y las transiciones forestales. El patrón espacial derivado de la topografía es un factor estructural que explica en gran parte la distribución de las comunidades vegetales. Las variables derivadas de la topografía, como la altitud, la pendiente, la radiación solar o la influencia de la topografía en los procesos hidrológicos, son algunas de las más utilizadas en el análisis de los patrones de la vegetación (Pons & Solé Sugrañes 1994; Florinsky & Kuryakova 1996; Serra-Díaz et al. 2011). El factor climático es bien conocido como determinante, y la dinámica del clima es responsable de la dinámica y transiciones entre especies en muchas partes del planeta (Kelly & Goulden 2008). En gran parte de la Península Ibérica ha habido un considerable aumento de las temperaturas y una disminución de las precipitaciones durante las últimas décadas (López-Moreno et al. 2010; El Kenawy et al. 2012), lo que ha supuesto una intensificación de las sequías, especialmente en la franja mediterránea (IPCC 2007; Vicente-Serrano 2014). La respuesta de la vegetación a los fenómenos de sequía es una cuestión de creciente interés y en trabajos recientes se ha mostrado su efecto en varias especies de la cuenca mediterránea y Europa (Pasho et al. 2011; Carnicer et al. 2011; Camarero et al. 2011; Vilà Cabrerà et al. 2013; Martínez-Vilalta & Lloret 2016). Algunas de estas especies se encuentran dentro de su límite más meridional en el ámbito mediterráneo, lo que explica su vulnerabilidad al incremento térmico (Andreu et al. 2007; Sánchez-Salguero et al. 2016). Su declive podría ser la causa de transiciones permanentes, y especies tolerantes a la sequía podrían pasar a ser dominantes en la composición de algunas comunidades vegetales.

3.1.1. Caso de estudio

En este trabajo de investigación se han analizado 13 tipos de transición (Tabla 1) en tres grandes áreas de la vertiente mediterránea de la Península Ibérica, cuyos ámbitos corresponden con tres escenas Landsat de aproximadamente 32500 km² cada una. El análisis se centra en la influencia de las variables topográficas y la recurrencia de sequías en cada una de las transiciones. El periodo de análisis es 1987-2012, incluyendo los sub-periodos 1987-2002 y 2002-2012. Se han ejecutado 13 modelos BRT (uno por cada tipo de transición) para cada ámbito (3) y periodo (3), lo que supone que un total de 117 modelos han sido evaluados y validados. Por criterios de extensión de la revista, en el artículo dedicado a este análisis, "*Factors affecting forest dynamics in the Iberian Peninsula from 1987 to 2012. The role of topography and drought*", no se han incluido las transiciones desde formaciones herbáceas (prados). Los resultados del análisis de este tipo de transiciones quedan recogidos en el apartado final de resultados de la tesis (Apartado 8).

Cubierta de vegetación inicial	**Cubierta de cambio**
De formaciones de prado a	Matorrales
	Bosque de coníferas
	Bosque de frondosas perennifolias
	Bosque de frondosas caducifolias
De formaciones de matorral a	Bosque de coníferas
	Bosque de frondosas perennifolias
	Bosque de frondosas caducifolias
De bosque de coníferas a	Bosque de frondosas perennifolias
	Bosque de frondosas caducifolias
De bosque de frondosas perennifolias	Bosque de coníferas
	Bosque de frondosas caducifolias
De bosque de frondosas caducifolias	Bosque de coníferas
	Bosque de frondosas perennifolias

Tabla 1. Tipos de transición entre cubiertas de vegetación estudiadas.

3.2. Abandono agrícola

El abandono agrícola es un tipo de fenómeno fundamentalmente relacionado con factores externos de tipo socioeconómico (Fielding 1989; Collantes & Pinilla 2004; Benayas et al. 2007). No obstante existen factores internos (Lasanta et al., 2017), de tipo ambiental y socioeconómico, que determinan el grado de vulnerabilidad en relación a este fenómeno dentro de un mismo contexto territorial.

En zonas de montaña, el modelo tradicional agrícola está considerado un régimen de perturbación del paisaje con efectos positivos en la biodiversidad (Farina 2006; Bezák & Halada 2010; Halada et al. 2011), y también desde el punto de vista del patrimonio cultural (Farina 2006; Marín-Yaselí & Lasanta 2003). Por otro lado, otros enfoques consideran que el abandono agrícola tiene efectos positivos en la estructura del paisaje, ya que incrementa su conectividad y el desarrollo de especies forestales (Farina 2006; Beilin et al. 2014; Navarro & Pereira 2015). Los efectos negativos del abandono agrícola habitualmente hacen referencia al incremento de la erosión del suelo, principalmente en zonas áridas y semiáridas (y litología más deleznable) o con un recubrimiento escaso de vegetación (Gallart & Llorens 2004; García-Ruiz 2010; Kosmas et al. 2015), y el riesgo de incendios (Lloret et al. 2002; Moreira et al. 2011). No obstante, del

abandono de tierras también se derivan otro tipo de consecuencias que trascienden más allá de los efectos a corto o medio plazo. El abandono general de usos tradicionales de la tierra, como pueda ser la agricultura y la ganadería extensiva en las zonas de montaña, puede tener repercusiones en otros ámbitos territoriales y temáticos, como por ejemplo el abastecimiento de agua de zonas urbanas y tierras irrigadas en zonas de valle central (García-Ruiz 2011; López-Moreno 2011). Recientemente se han empezado a diseñar estrategias que intentan recuperar dichas actividades tradicionales para fijar la población y aumentar las rentas en el medio rural de montaña[3]. Comprender cómo los factores ambientales y socioeconómicos han influido en el abandono de tierras, puede proporcionar información útil para la implementación de dichas estrategias, así como para gestionar las inversiones, ya sean del sector público o privado.

En España, el Pirineo es una de las zonas más afectadas por el abandono agrícola como consecuencia del éxodo rural y las limitaciones físicas a la modernización (Benayas 2007, Collantes y Pinilla 2004). Después del declive del sistema tradicional, basado en una economía de subsistencia, los cultivos de cereal fueron menos competitivos como recurso alimenticio para la población y se convirtieron en una fuente importante de forraje para ganado. Otros cultivos forrajeros y prados antrópicos han aumentado su área debido al abandono de la trashumancia en la mayoría de las comarcas (García-Ruiz & Lasanta 1993, IAEST 2017, IDESCAT 2017). La expansión del turismo en esta región ha llevado a reajustes económicos en las familias del ámbito rural. Muchas áreas mantienen una simbiosis equilibrada entre el turismo y la producción agrícola (Cánoves et al. 2004), pero la ganadería y los cultivos han disminuido claramente en las áreas con mayor desarrollo turístico (García-Ruiz et al. 1996, 2011). En cambio, la modernización (alta mecanización y mano de obra limitada) y la intensificación agrícola han sido la tendencia en otras regiones geográficas, como la depresión del Ebro, una región semiárida donde se han desarrollado modernos sistemas de riego desde mediados del siglo XX (Pinilla 2006). Durante la liberalización económica y el desarrollo capitalista de España, esta región estaba en mejores condiciones para beneficiarse de la expansión de la industrialización y los mercados externos, y la modernización del regadío aumentó la productividad de los cultivos. Al mismo tiempo, este proceso contribuyó a la despoblación y al abandono agrícola en el Pirineo, debido a la construcción de embalses en los valles para abastecer las tierras irrigadas (Pinilla 2006, Duarte et al. 2014). Por otro lado, el cereal de invierno sigue siendo el cultivo de secano predominante, y aunque ocupa una mayor extensión, la producción por hectárea de los cultivos de regadío (principalmente maíz) es 5-6 veces mayor (Pinilla 2006).

3.2.1. Caso de estudio

El análisis del abandono agrícola queda recogido en el artículo "*Environmental and socioeconomic factors of abandonment of rainfed and irrigated crops in northeast Spain*". En dicho trabajo se analiza la influencia de una serie de factores ambientales y socioeconómicos en el abandono agrícola en el noreste de España durante el periodo 1987-2012. Con el fin de determinar particularidades con respecto a tipos de cultivo y contexto territorial, el análisis diferencia entre el abandono de cultivos de secano y de regadío en los Pirineos centrales y en la cuenca central del Ebro. Como en el caso de estudio de la dinámica de la vegetación, se han incluido dos sub-periodos en el análisis, 1987-2002 y 2002-2012, para tener en cuenta la variabilidad de los factores humanos y climáticos. Por tanto, se evaluaron 2 modelos de abandono agrícola (secano y regadío) por periodo (3) y área de estudio (2) generado un total de 12 modelos BRT. Las dos áreas

[3] http://www.eldiario.es/aragon/economia/entidades-recuperar-ganaderia-extensiva Pirineo_0_693281700.html

de estudio quedan dentro del ámbito correspondiente a la clasificación de cuatro escenas Landsat. Los criterios para delimitar cada una de las áreas quedan definidos en el correspondiente artículo. Se consideró como zonas de abandono aquellas parcelas agrícolas que habían sido colonizadas por vegetación leñosa (es decir, cambio de cubierta de cultivo a categoría de cubierta arbustiva o forestal) entre las dos fechas del correspondiente periodo de estudio.

Objetivos

4. OBJETIVOS DE LA TESIS

Tal y como se ha establecido la introducción, el núcleo de la presente tesis lo conforman tres artículos de investigación dedicados al análisis de la dinámica de las cubiertas del suelo y a la determinación de sus factores explicativos. Así pues, este sería el objetivo general, enfocado en el análisis en los casos de estudio de la dinámica de la vegetación y del abandono de cultivos. La coherencia temática del conjunto del trabajo queda reflejada en el orden y contenido de los artículos: un primer artículo dedicado a la metodología para la obtención de series temporales de mapas de usos y cubiertas del suelo; y un segundo y tercer artículos donde se determina la influencia de factores en las transiciones de vegetación y en el abandono de cultivos, respectivamente. Por tanto, los objetivos específicos de cada uno de los tres artículos son los siguientes:

- *Developing spatially and thematically detailed backdated maps for land cover studies*. Uno de los objetivos principales de este artículo es establecer un esquema metodológico y de clasificación óptimo para la obtención de mapas de usos y cubiertas del suelo para grandes superficies de terreno, con tal de contar con patrones regionales en una segunda fase de análisis. Ello supone obtener mapas con un nivel de desagregación temática suficiente como para abordar un análisis detallado de la dinámica de las cubiertas del suelo. Así mismo, otro de los objetivos principales es concretar un procedimiento que permita generar mapas en momentos temporales en los que no se dispone de información de referencia. Finalmente, la intención es obtener el máximo porcentaje de acierto (tanto a nivel global como por categorías) en la validación de las clasificaciones, lo que a su vez influirá en la fiabilidad de la siguiente fase de detección de cambios.

- *Factors affecting forest dynamics in the Iberian Peninsula from 1987 to 2012. The role of topography and drought*. Este artículo recoge de forma detallada el flujo de trabajo seguido para la obtención de las zonas de cambio en cada una de las transiciones de vegetación, así como el procedimiento de ajuste, validación y evaluación de los modelos *Boosted Regression Trees* (BRT). Con ello, los objetivos principales de dicho artículo son, por un lado, identificar los factores claves en cada una de las transiciones de vegetación, considerando factores clave a aquellos que permiten diferenciar las estrategias o preferencias ecológicas de cada tipo fisionómico de vegetación. Y, por otro lado, determinar si la ocurrencia de sequías es un factor clave y, si es así, identificar las interacciones más importantes con otros factores clave.

- *Environmental and socioeconomic factors of abandonment of rainfed and irrigated crops in northeast Spain*. Este artículo contiene en una primera parte la metodología seguida para la identificación de zonas agrícolas abandonadas así como una descripción sintetizada de los parámetros que controlan los modelos BRT, su evaluación y validación. Una vez establecida la metodología, el objetivo principal es la determinación de factores clave para identificar particularidades en los patrones de abandono según tipo de cultivo (secano y regadío) y contexto territorial (Pirineo central y cuenca central del Ebro).

5. ARTÍCULO 1: *Developing spatially and thematically detailed backdated maps for land cover studies*

Vidal-Macua, J. J., Zabala, A., Ninyerola, M., & Pons, X. (2017). Developing spatially and thematically detailed backdated maps for land cover studies. *International Journal of Digital Earth*, 10(2), 175-206. https://doi.org/10.1080/17538947.2016.1213320

INTERNATIONAL JOURNAL OF DIGITAL EARTH, 2017
VOL. 10, NO. 2, 175–206
http://dx.doi.org/10.1080/17538947.2016.1213320

Developing spatially and thematically detailed backdated maps for land cover studies

Juan José Vidal-Macua[a], Alaitz Zabala[a], Miquel Ninyerola[b] and Xavier Pons[a]

[a]Dep Geografia, Grumets Research Group, Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain; [b]Dep Biologia Animal, Vegetal i Ecologia, Grumets Research Group, Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain

ABSTRACT

Global or regional land cover change on a decadal time scale can be studied at a high level of detail using the availability of remote sensing data such as that provided by Landsat. However, there are three main technical challenges in this goal. First, the generation of land cover maps without reference data is problematic (backdating). Second, it is important to maintain high accuracies in land cover change map products, requiring a reasonably rich legend within each map. Third, a high level of automation is necessary to aid the management of large volumes of data. This paper describes a robust methodology for processing time series of satellite data over large spatial areas. The methodology includes a retrospective analysis used for the generation of training and test data for historical periods lacking reference information. This methodology was developed in the context of research on global change in the Iberian Peninsula. In this study we selected two scenes covering geographic regions that are representative of the Iberian Peninsula. For each scene, we present the results of two classifications (1985–1989 and 2000–2004 quinquennia), each with a legend of 13 categories. An overall accuracy of over 92% was obtained for all 4 maps.

ARTICLE HISTORY
Received 19 January 2016
Accepted 9 July 2016

KEYWORDS
Land cover mapping; Landsat; statistical classifiers; backdating; Iberian Peninsula

1. Introduction

The study of land cover change has become central to the study of global change, and land cover maps are some of the most widely used products in such studies. Land cover maps are also useful to land managers (planning, conservation, etc.) as well as for scientific research (landscape analysists, macroecologists, etc.). Vegetation and land cover dynamics have a key role in landscape structure; indeed, changes in the distribution of plant communities or human alteration of biotic and abiotic components are main drivers of landscape evolution (Cowen and Jensen 1998; Serra, Pons, and Sauri 1999; Pèlachs et al. 2009; Sexton et al. 2013b). Furthermore, analysis of land cover changes allows researchers to understand the interactions between ecological and climatic systems (Bonan 2008; Herrera and Doblas-Miranda 2012; Xian, Homer, and Aldridge 2012; Beltrán et al. 2014). A high level of detail (i.e. legend complexity and spatial resolution) is needed to improve understanding of landscape dynamics in global or regional change studies.

CONTACT Juan José Vidal-Macua juanjo.vidal@uab.cat Dep Geografia, Grumets Research Group, Universitat Autònoma de Barcelona, Edifici B, Bellaterra 08193, Catalonia, Spain

Ecosystem matrices and anthropogenic occupation models assume gradual dynamics over time and space. However, the Earth's surface is not homogeneously affected by climate anomalies (Parry 2007), resulting in localized distribution patterns of indicators of change, so analyses need to be applied at a fine spatial resolution. There are some useful products available for global change studies: these include the Climate Change Initiative Land Cover dataset of the European Space Agency for the years 2000, 2005 and 2010 using the Medium Resolution Imaging Spectrometer (MERIS) instrument (Arino et al. 2007; Kirches et al. 2013), and the Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover products whose temporal coverage is from 2001 to 2012 (Strahler et al. 1999; Friedl et al. 2010). However, the drawback of these global dataset products is that they have a moderate spatial resolution (500 m in MODIS and 300 m in MERIS), and such large pixels do not permit a detailed understanding of land cover dynamics in small 'patches'. Higher spatial detail is needed to detect phenomena with a modest spatial dimension and identify intermediate categories within fluctuating changes in the landscape (Mucher et al. 2000).

Not long ago, insufficient computational capabilities and low data availability made it difficult to carry out high-resolution studies of large areas. Nowadays, increased computing power and access to a large amount of detailed spatial data (Craglia et al. 2012) permit studies of large spatial areas using finer units of measurement. Using big data approaches, such as process parallelization and 64-bit processors, allows achievement of high-quality results within a reasonable amount of time.

The NASA–USGS Global Land Survey dataset (Gutman et al. 2013), primarily comprising of Landsat images (30 m resolution) and available via open access to the science community and general public since 2008, is a very important resource for detailed mapping from 1972 to present. This dataset allows the study of global change at a relatively fine resolution for large areas, and has already facilitated a greater understanding of the complexity of landscape dynamics (Pons et al. 2014a).

Global or regional land cover maps provide high-quality products, generated with human-aided mapping techniques (field surveys and/or photointerpretation) or with automated methodologies (remote sensing classification). Human-aided maps provide detailed land use legends, but updates or new mapping programs are constrained by high costs, and it is difficult to achieve normalization of methodologies used by individual photointerpreters. One example of human-aided mapping is Corine Land Cover, coordinated by the European Environment Agency. On the other hand, automated methodologies provide consistent and comprehensive overviews of the variables of interest, but the legends are quite broad and only distinguish a few types of vegetation or categories of agricultural or urban land uses. Along these lines, in studies of biodiversity by the European Environment Agency (2014), it was deduced that higher spectral resolution is much more important than improved spatial resolution.

Satellite imagery offers higher spectral and temporal resolution than aerial imagery, allowing better differentiation of vegetation types (Vogelmann et al. 1998; Moré et al. 2005) or types of anthropogenic land use such as crops (irrigated, rainfed, herbaceous, woody crops, etc.) or urbanized areas. Once the workflow has been set, satellite image classification can provide regional land cover and land cover change cartography products (Vogelmann et al. 2001; Gutman et al. 2013; Glanz et al. 2014) having sufficient detail to deal with most global change issues.

Several studies with high accuracy and relatively complex legends (e.g. Chen et al. (2015), with 10 categories and Lu and Weng (2006), with 5 categories) or large spatial representativeness (Hansen et al. 2013; Sexton et al. 2013a; Gaveau et al. 2014; Chen et al. 2015) have been carried out using satellite imagery. As discussed above, alternative data sources from sensors such as MODIS or MERIS have limited spatial resolution for the appreciation of certain change phenomena, and are also limited to relatively short time series (beginning in the year 2000) as compared to Landsat (http://landsathandbook.gsfc.nasa.gov/data_properties/prog_sect6_5.html). Moreover, although MODIS imagery has a 2-day orbital repeat, the Landsat satellite (16-day orbital repeat) can provide a better identification of more specific spatial patterns due to its higher pixel resolution. High temporal resolution is not as important as a high spatial resolution for finer change products in global-scale studies, whereas major vectors of change are identified in the 1–5-year range (Wulder et al.

2008). With the Landsat datasets, phenological differences can be captured, and the spatial representation of landscape diversity is adequately represented.

The availability of land cover products based on Landsat imagery has increased over the last few decades. One such project is the National Land Cover Database for the United States (NLCD) which has developed intermediate-scale (30 m) land cover products for the years 1992, 2001, 2006 and 2011 (Vogelmann et al. 1998, 2001; Homer et al. 2004; Xian, Homer, and Fryc 2009; Jin et al. 2012, 2013). In this project, the approach to land cover classification was based on unsupervised classifiers for the initial NLCD 1992, and for later periods a decision tree classification method was implemented. Recently, a global land cover mapping at 30 m resolution for years 2000 and 2010 was presented in Chen et al. (2015), which achieved an overall accuracy (OA) over 80% for a 10-category legend, and included the integration of several pixel classifiers, object-based identification, and integration with web-based reference data.

The NLCD began classifying the 1992 data using reference data available for that quinquennium, and the same methodology was followed for later quinquennia classifications. The 2001 NLCD has an OA of 85.3% for Level I (16-category legend, excluding classes found in Alaska), and regional accuracies (10 geographic regions) range from 79% to 91% (Wickham et al. 2010; http://www.mrlc.gov/finddata.php). Accuracy of the 1992 NLCD was lower (http://landcover.usgs.gov/accuracy/index.php), and results for 2006 and 2011 have not been published yet. Therefore, there are cases in which OA does not allow the achievement of low error rates in change analysis, considering that in order to assess the accuracy resulting from overlapping two land cover maps (i.e. different quinquennia classifications), the overall accuracies of the two maps must be multiplied (Serra, Pons, and Saurí 2003). For instance, if we overlap two maps from different dates with respective accuracies of 80% and 82%, the resulting land cover change map will have an accuracy of only 65.6% ($0.80 \times 0.82 = 65.6$). This is why it is important to achieve accuracies over 90% in land cover classifications, permitting overall accuracies for land cover change products of over 80%.

These issues are especially relevant for an area such as the Iberian Peninsula which presents a case of high heterogeneity for the purposes of land cover classification due to its rugged relief (from 0 to 3478 m), spatial variability of climate, and its long and complex human history (García-Ruiz 1990; Tullot 2000; Gil-Olcina and Gómez-Mendoza 2001). Moreover, the Iberian Peninsula (and the Mediterranean more generally) has had an intense land cover dynamics which have led to high fragmentation of the landscape. The region is also especially influenced by climate change dynamics (Parry 2007). As said, our final purpose will be to analyze changes in recent decades in a detailed manner using high-accuracy land change maps.

Studying the land cover dynamics of the whole Iberian Peninsula requires a big data approach. Indeed, Landsat datasets provide a large number of spectral variables and derived indices for whole time series, while ancillary data, such as high-resolution digital elevation models (DEM) and climatic datasets, can also be used. Detailed training and test reference information is also needed. Processing such volumes of data and dimensionality increase the computational complexity of image classification, requiring maximum levels of automation applied to data preprocessing, to training data extraction, and algorithm parameterization.

Another issue which needs to be solved is the availability of training and test areas required for mapping historical periods (Townshend et al. 2012), or in other words, for backdating. The most recent land cover classifications have typically used whatever reference information may be available from official institutions (orthophotos, thematic maps, spatial databases, etc.), and, in cases where reference data is lacking, information-gathering is complemented by field work. Nevertheless, classification for periods without reference data is much more difficult because training areas must be extracted from visual analysis of older aerial imagery (when available), a very time-consuming task, especially when studying large areas. Furthermore, reference information may be compiled from different sources or dates and hence developed using diverse methodologies, leading to inconsistencies in training and validation data (Townshend et al. 2012) as well as in comparisons between classifications.

In the context of the Digital Earth era (Craglia et al. 2012), rigorous and extensive land cover information and the ability to deal with the temporal component will be key for designing future earth observation applications. Although land cover has usually been a key component of studies at local, regional and global scales, greater progress needs to be made with the temporal component, especially when backdating is necessary for understanding the dynamics and resilience of natural and socioeconomic systems. In this sense, obtaining reliable and coherent reference information for land cover time series is essential for the development of monitoring systems at all geographical scales. In particular, working at a fine level of spatial detail will increase the confidence of conclusions drawn from studies of land cover dynamics. This study addresses challenges which are inherent to understanding a dynamic planet, and promotes the capacities of the Digital Earth and Big Data to support sustainable development (Huadong 2016).

This paper describes methodologies which are designed to handle the problems related to data volume and backdating, developed within the context of a project to map land cover dynamics in the Iberian Peninsula over a period of decades. The workflow is implemented in a geographic information system (MiraMon) which has the required consistency and automation capabilities for such an endeavor.

2. Objectives

The main objective of this study is to lay the groundwork for the production of highly detailed and accurate (over 90% to guarantee adequate land cover change products) land cover maps of large areas (by using full frames of about 30,000 km^2). High detail and accuracy are made possible by the relatively high spatial resolution of the Landsat program, paired with commonly available ancillary data. The aim of the proposed method is to obtain a spatially continuous land cover product at five-year intervals (quinquennia) over an entire Landsat time series whenever possible, and that it results in relative richness of legend entries. One important issue of this work is the tradeoff between obtaining high accuracies and avoiding unclassified areas, since limiting the uncertainty of the classification leads to a less-classified area.

A second objective is reduce computational processing times by reducing the number of variables used, since better classification results can be obtained by employing intra-annual phenological information data, thereby including data from several dates.

A third objective is to test a simple method for semi-automatically refining the features of training data so that they can be used in the classification of other quinquennia (typically used for backdating).

Our proposal also aims to establish several standard configurations for when similar reference information is available, thereby offering a reproducible methodology. We consider that the parameterization techniques described in this paper will be of great interest to earth observation researchers in cases in which this condition of available reference information is met.

Using two of the most diverse Landsat full frames of the Iberian Peninsula, we show how our complete method produces a thorough radiography of land covers in the complex Iberian geographical area for the last decades.

3. Study area and data

3.1. Study area

Two study areas in Spain were used (Figure 1). Study area 1 is located in the northeast part of the Iberian Peninsula, and corresponds to path 199, row 31, of the Landsat Worldwide Reference System-2 (WRS-2) scenes (hereafter 199031 or northern scene). This scene comprises most of provinces of Huesca and Zaragoza, the western part of Lleida and Tarragona, and the north of Teruel, for a total of 30,848 km^2. Maximum altitude in this region is 2490 m at the Turbón massif in the Pyrenees. The

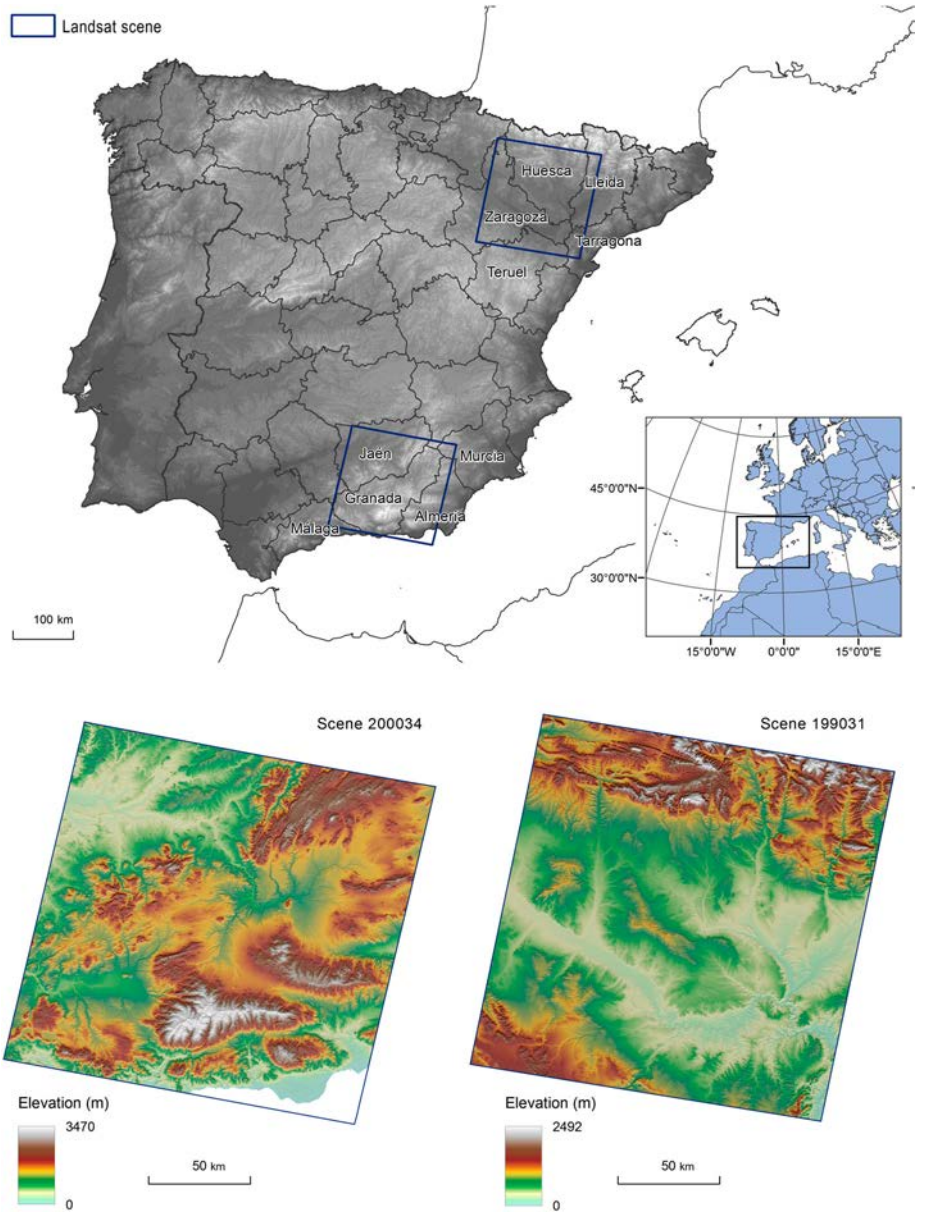


Figure 1. Study areas.

central Ebro basin occupies a large part of the scene; this basin, with an altitudinal range of mainly between 100 and 300 m, contains the lowest altitudes, reaching a minimum of 30 m in the province of Tarragona near the Ebro delta. Study area 2 is located in the southeast part of the Iberian Peninsula, and corresponds to path 200, row 34 (hereafter 200034 or southern scene). This scene has a continental area of 29,259 km^2, and includes the province of Granada, almost all of Jaén, the western half of Almería, the eastern end of Málaga and the western end of Murcia. This scene has a more irregular topography and the altitude ranges from 0 m in coastal areas to 3478 m at the Mulhacén peak in the Sierra Nevada massif.

These Landsat scenes comprise very large ambits of study. Also, they have been chosen because of their heterogeneity of altitudinal ranges (combining flat areas with very complex topography), diversity of biogeographical regions (including Alpine, Eurosiberian and Mediterranean areas) and land occupation patterns (a diversity of land uses including several types of crops and configurations of landownership such as small holdings or large plots).

3.2. Data

We applied the methodology in two quinquennia, 1985–1989 and 2000–2004. For the 2000–2004 quinquennium we used the SIOSE (Land Occupation Information System of Spain) reference information. This dataset, produced in 2005, has the advantage that it is the first continuous, detailed government dataset available for all of Spain. This dataset provided a large number of training and test areas which could also be used in the land cover classification of the 1985–1989 quinquennium using the method explained in Section 4.7.

Following visual examination of the available Landsat imagery for the 1985–1989 and 2000–2004 periods, we selected a set of dates with relatively low cloud cover. In order to identify training and test areas containing useful reference data for several quinquennia (see Section 4.7), it was necessary to maintain equal or similar phenological patterns between quinquennia. A further selection of dates closer to the central years (1987 and 2002 for the first and second quinquennium, respectively) was made for the final set of images (Table 1).

Each quinquennium is treated as a whole, and it is understood that the most suitable representation of seasonal phenological variation of natural vegetation and crops is found in the central year; when use of the central year is not possible due to cloud cover, it is assumed that its variation with neighboring years is small. Winter dates were discarded due to topographic shadows and strong phenological changes (i.e. deciduous vegetation). For each date, analysis included the spectro-temporal variables in addition to the following ancillary variables: altitude, slope, summer solar radiation (RAD), and the indices NDVI (Normalized Difference Vegetation Index), NDWI (Normalized Difference Water Index) and SAVI (Soil Adjusted Vegetation Index; see Section 3.3.4 for details). These auxiliary variables aid classification processes by providing common statistical patterns which can be also linked with thematic classes (Moré, Serra, and Pons 2006; Pons et al. 2014c).

In total, 51 variables were included in the classification process for each scene: bands 2, 3, 4, 5 and 7 of Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) for each date, 18 indices, and 3 topo-climatic variables. Band 1 was excluded due to the high influence of atmospheric aerosols which can introduce statistical 'noise', and preliminary trials showed that this band did not contribute to classification accuracy. Thermal bands 6 and 9 were also excluded. Solar summer radiation was only considered because it was the most useful for classification among the seasonal solar radiation datasets. The variables of spectral indices, altitude and solar radiation were scaled between 0 and 100 so that their ranges would be similar to those of reflectance and slope.

Table 1. Dates by scene and quinquennium, sorted by month.

	1985–1989	2000–2004
Southern scene (200034)	14 April 1987	12 April 2001
	17 June 1987	30 May 2001
	30 June 1986	10 June 2002
	16 July 1986	25 July 2001
	20 August 1987	21 August 2002
	2 September 1986	1 September 2003
Northern scene (199031)	20 April 1986	27 April 2003
	14 May 1989	29 May 2003
	20 June 1985	8 June 2001
	6 July 1985	2 July 2001
	17 July 1989	24 July 2003
	23 August 1985	25 August 2003

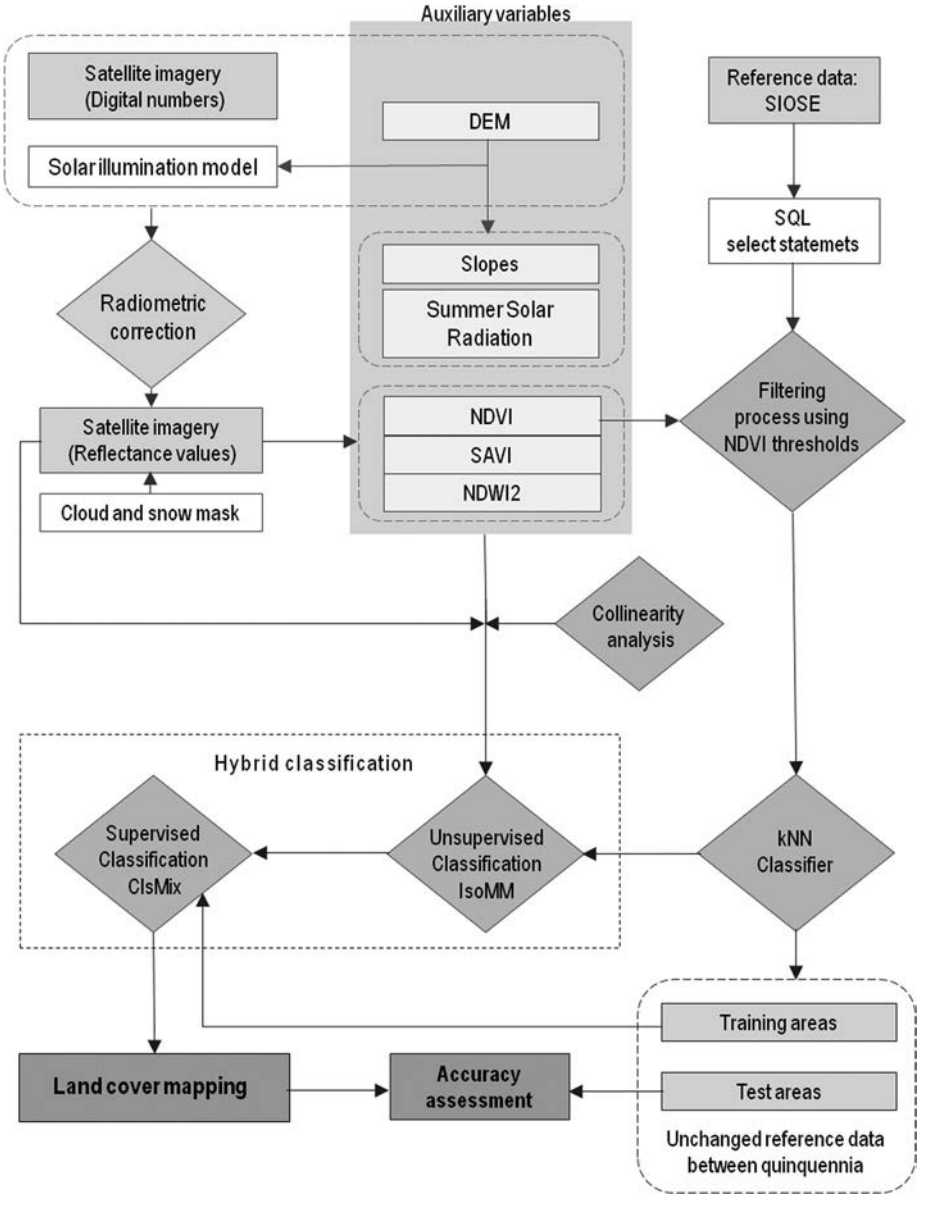


Figure 2. Methodological workflow.

Figure 2 shows the methodological workflow of the study.

4. Pre-classification processing

4.1. Digital terrain models

A 30 m DEM was generated using 1:5000 map sheets containing light detection and ranging data from the Aerial Orthophotography National Plan 2010 (PNOA). The following variables were calculated from the 30 m DEM using MiraMon software (Pons 2006):

- A solar illumination model generated using the Illum and Ombra MiraMon modules for each date. This parameter is required for radiometric correction and is related to illumination and topographic shadows for the specific day and hour when the image is captured (incident angles, self-shadows and cast-shadows).
- A slope variable generated using the Pendent module which accounts for eight neighboring pixels.
- A summer solar radiation variable generated using the InsolDia module which computes the total amount of incident solar radiation for each pixel at any given day of the year following the methodology of Pons and Ninyerola (2008).

4.2. Geometric and radiometric correction

Satellite images had the LT1 geometric correction level based on control points and DEM data; these usually provide accuracies of about half a pixel (http://landsat.usgs.gov/Landsat_Processing_Details.php). The spatial reference system and projection was Universal Transverse Mercator WGS84.

Radiometric correction was applied to convert digital numbers to reflectance values following Pons et al. (2014b) using the MiraMon CorRad module. This tool integrates, among other parameters, pseudo-invariants areas (PIA) derived from MODIS reference values, an illumination model (cast-shadows and self-shadows), sun elevation, sensor calibration parameters, DEM, etc.

4.3. Snow, clouds and cloud-shadow masking

Snow cover masks were generated using the Normalized Difference Snow Index (NDSI; Hall, Riggs, and Salomonson 1995), from which we obtained a threshold value to discriminate snow surfaces (Dozier 1989; Cea, Cristóbal, and Pons 2007). We used a threshold of NDSI > 3 for our study areas. In addition, a 30 m buffer mask was applied to exclude pixels within that range in order to account for snow-covered pixels' spectral influence on their nearest neighbors. Cloud and cloud shadow masks were manually refined after a first automatic procedure (Zhu and Woodcock 2012; Zhu, Wang, and Woodcock 2015).

Statistical inconsistencies resulting from anomalous sensor behavior were also manually masked.

4.4. Vegetation and moisture indices

Vegetation and moisture variables were calculated from the spectral bands. NDVI is the most widely used index for analysis of vegetation and crops (Kriegler et al. 1969; Baret and Guyot 1991; Serra and Pons 2013), and in this study it was computed from bands 3 and 4. The SAVI index minimizes influence of the soil on reflectance, helping to dissociate spectral response of the soil from vegetation in discontinuous covers (Huete 1988; Baret and Guyot 1991; Serra and Pons 2013); in this study it was computed from bands 3 and 4, using a value of 0.5 for the L constant. Finally, the NDWI index, based on the $NDWI_3$ proposed by Ouma and Tateishi (2006) as (B5−B4)/(B5 + B4), was calculated. This index enhances the identification of water bodies such as lakes and reservoirs and results in a better delineation of the waterfront.

4.5. Source of training and test areas

Training and test data were obtained from the Land Occupation Information System of Spain 2005 (SIOSE, Sistema de Información sobre Ocupación del Suelo en España), a database and vector layer developed at a 1:25,000 scale by photointerpretation of 2004, 2005 and 2006 orthophotos and Earth Observation Satellite (SPOT, Satellite Pour l'Observation de la Terre) imagery, in addition to information from other reference databases (http://www.siose.es/). It is an official governmental data

Table 2. Land cover classes.

1	Water bodies
2	Coniferous forests
3	Broadleaf deciduous forests
4	Broadleaf evergreen forests
5	Shrublands
6	Grasslands
7	Bare soils
8	Urbanized areas
9	Irrigated herbaceous crops
10	Rainfed herbaceous crops
11	Irrigated woody crops
12	Rainfed woody crops
13	Rice crops[a]
14	Greenhouses[b]

[a]Only present in scene 199031.
[b]Only present in scene 200034.

source with full coverage of Spain and with well-defined disaggregated levels and land cover percentages inside each polygon.

In this paper, the legend derived from the SIOSE nomenclature has 14 thematic classes (Table 2). Some of these thematic classes involve grouping two or more SIOSE simple covers (Table 3).

4.6. Refinement of training and test areas

The main data fields in the SIOSE principal database include the SIOSE_CODE field and several COVER_FRACTION fields (Table 4). The SIOSE_CODE field contains a code whose nomenclature depicts the polygon composition in terms of land cover type, cover fraction and spatial distribution typology in composite covers; examples of SIOSE_CODE entries include association (A), regular mosaic (R), etc. The other fields contain the cover fraction for each single cover; this is entered as a percentage value for categories such as grasslands (PST), broadleaf evergreen trees (FDP), coniferous forests (CNF), shrublands (MTR), Bare soil (SDN), Broadleaf deciduous forests (FDC), etc. Some cover types are also included, for example, 'dehesa' (DHS).

Using Structured Query Language (SQL) statements, the selection criteria for SIOSE polygons that became part of the initial reference data was established. Polygons with purer covers were first identified based on cover fraction, and we also preferentially selected polygons without mixing of certain covers (see Appendix for details). We randomly split the polygon set into two groups, a training set containing 70% of the polygons and a validation set with the remaining 30%. This process was performed separately for each category.

In a second stage, polygons obtained in the previous phase were filtered based on NDVI thresholds. We filtered polygons with broadleaf forests covers of 45–75% and coniferous forest covers below 90% in order to isolate the tree layer from shrubs, pastures, bare land, or any other type of cover inside the initial polygon. A similar procedure was used in the case of shrubland

Table 3. Thematic classes based on two or more SIOSE simple covers.

Thematic class	SIOSE simple cover labels
Bare soil or rock	Bare soil (SDN), rocky outcrops and crags (ARR), scree (CCH), extractive and pouring areas (ZEV) and mining areas (PMX)
Water bodies	Lakes and ponds (LAA, ALG) and reservoirs (AEM)
Woody crops[a]	Olive grove (LOL), vineyard (LVI), citrus fruit (LFC), no citrus fruit (LFN) and other woody crops (LOC)
Urbanized areas	Old town (UCS), urban widening (UEN), discontinuous urban areas (UDS), organized industrial zones (IPO), unordered industrial area (IPS), isolated industry (IAS), office and commercial areas (TCO), administrative and institutional facilities (EAI), road network (NRV) and Roads, parking or pedestrian areas without vegetation (VAP)

[a]In SIOSE, woody crops have the same label, adding a suffix for rainfed (sc) and irrigated (rr).

Table 4. An abbreviated example record in the SIOSE database.

SIOSE_CODE	DHS	PST	FDP	FDC	CNF	MTR	SDN
A(30FDP_25MTR_15DHS(35PST_35MTR_30FDP)_15FDC_10SDN_05PST)	15	10.3	34.5	15	0	30.3	10

and grassland polygons in order to remove bare soil pixels to the greatest extent possible. Also, we set an NDVI threshold for urban areas in order to avoid confounding between bare soils and areas with sparse vegetation. A threshold was set for water bodies in order to determine the purest water pixels in situations with seasonal changes in water levels at reservoirs and natural lakes. This approach attempts to minimize spectral heterogeneity within thematic categories (Toll 1984; Cushnie 1987; Chuvieco 1996).

Using map algebra, each training pixel is given the maximum value of the entire NDVI time series. For polygons with 45–75% of forest cover, pixels with values above the median were selected first, effectively retaining 50% of the pixel set corresponding to the highest values, since the highest NDVI values within each polygon are associated with forest layers (not shrubs, pastures or bare land). For polygons with coniferous forest covers below 90%, we selected pixels above the first quartile. For shrubland and grassland, values above the first quartile were selected in order to remove low values associated with bare soil. Finally, for water bodies and urban areas values below the median were selected. Next, we performed a manual check for appropriateness of selected thresholds by comparing selections with regional orthophotos (2005) at 1:5000 scale. This was done in sample group of 15% of total training areas, which were selected at random and were homogeneously distributed throughout the scene. Finally, broadleaf forests were assigned threshold value corresponding to the median plus one standard deviation, and in shrubland polygons an upper threshold (corresponding to the third quartile plus one average absolute deviation) was established due to the presence of a dense tree layer in several polygons. NDVI thresholds were: ≥0.63 for coniferous forests, ≥0.62 for broadleaf evergreen forests (≥0.56 southern scene), ≥0.70 for broadleaf deciduous forests, between 0.34 and 0.62 for shrublands, ≥0.25 for grassland, ≤0.17 for urban areas and ≤0.09 for water bodies. The shrublands range was established so that it would account for areas with scattered trees. The different thresholds for broadleaf evergreen forests of the northern and southern scenes were established because broadleaf evergreen forests in the southern scene have a less compact structure and are mixed with grasslands, shrublands and bare soils due to impacts from livestock. In Section 6 we show some examples of the influence of thresholds on classifications.

4.7. Obtaining multi-temporal training and test areas

Once training and test areas for the 2000–2004 quinquennium were filtered based on the established NDVI thresholds, we proceeded to a retrospective analysis. The objective of this analysis is to identify training and test pixels with an invariant statistical pattern for the 1985–1989 quinquennium. The k-nearest neighbor (kNN) classifier implemented in MiraMon software (Pons 2006) was used for this purpose. First, this algorithm calculates the Euclidean distance (in the data space) from each target pixel to each reference pixel (of training areas). Next it selects the kNNs (determined by the user), and finally each target pixel is classified to the modal category in those k pixels. In this process we only consider the reflectance variables. The pixels from the 1985–1989 period were the target pixels, clipped to the training and test areas obtained after the previous filtering process, whereas the pixels from 2000–2004 period were the reference pixels clipped to the same area. A value of 15 was established for the kNNs parameter based on a set of testbeds (in Section 6 we summarize the procedure that led to the determination of this value). After this process we kept pixels that remained unchanged (pertaining to the same category in 1985–1989 and 2000–2004 quinquennia) to be used as training and test data when running land cover classifications of the complete scenes in both quinquennia. The ratio between training and test pixels was maintained for all categories (imbalances did not exceed 3% compared to the original split).

5. Classification approach

Our study used a hybrid classification method for land cover mapping. The workflow followed these steps:

(1) Collinearity analysis to reduce the number of correlated variables and processing times.
(2) Unsupervised classification to identify a large number of statistical classes having multivariate statistical meaning.
(3) Supervised stage to assign statistical classes to thematic classes (land cover categories).

The same dataset of variables was used in a support vector machine (SVM) classifier (Hsu, Chang, and Lin 2003) implemented in Envi software. Results are shown in Section 6.

5.1. Two-block classification

In order to reduce processing time and statistical ranges, classifications were performed separately for crop areas (CRA) and natural and urban areas (NUA). Corresponding masks were generated from a kNN classification of whole scenes using training data (obtained as described in Section 4.7) grouped into two categories (CRA and NUA). The kNN classifier demands much more computing time than unsupervised classification. Nevertheless, using two categories or reducing the number of target pixels significantly reduces processing times.

In order to avoid misclassifications related to the urban areas category (13), NUA blocks were classified in a three-step procedure. In the first step, pixels classified as urban areas were overlapped with an urban mask obtained from the SIOSE database. This mask, extracted using SQL statements, depicts the maximum growth achieved in urban areas up to 2005. Thus, the overlay process restricts area classified as urban in the 1985–1989 quinquennium to the area where urban growth corridors have been developed (we dismiss the possibility of large urban areas that have been eliminated before 2005 and are located outside of this growth area). Also, this overlay discards urban pixels classified outside of the urban growth area for both quinquennia, avoiding commission errors related to the bare soils category (7) or areas with sparse vegetation (mostly occurring in the grasslands category). A second urban mask (UM2) is obtained from the common pixels in both layers, that is, the area of intersection between urban area classified in the first NUA block and the SIOSE urban mask. A second classification of the NUA block (NUA2) was performed on the area not covered by the second urban mask (UM2) including all categories except for urban areas. In the third step a mosaic of NUA2 and UM2 was created in order to obtain the definitive NUA block classification. Finally, a validation test including all categories was performed for the mosaic composed of the two blocks (definitive NUA block and CRA block).

As we will see in later sections, this complex procedure proved useful, taking into account that most of the workflow can be integrated in batch processes, achieving high levels of accuracy in relatively short processing times.

5.2. Collinearity analysis

In order to reduce data size and processing time, and also to determine the most useful variables for the classification process, we used correlation coefficients to quantify associations between pairs of variables. Variables which did not contribute to the success of classification were removed.

We consider that quinquennia have a similar collinearity structures because we assume that differences between quinquennia do not have sufficient spatial relevance, and also because the selected dates (or phenological pattern) between quinquennia are similar. Therefore, we took the 2000–2004 quinquennium as the reference; correlation coefficients were only calculated for this period and as a consequence the deleted variables were the same for both periods.

Several classifications were performed depending on the number of variables used. Results were analyzed in terms of OA, weighted overall accuracy (OA_w, calculated according to the area of each category in each of the maps and considering pixels classified as NoData inside the test areas), averaged producer accuracy (APA), classified area (CA) and percentage of processing time (T). In the case of processing time, please note that 100% is in reference to the set of 51 variables.

Maintaining the goal of obtaining results of similar quality (based on accuracy indices and CA), a minimum was reached with a subset of 11 variables. The criteria for elimination of one of two highly correlated variables were as follows:

(1) Remove first band 2, then band 5, and then band 7; this is because band 2 has more atmospheric influence, bands 5 and 7 provide less information than the preserved bands 3 and 4, and because band 7 is less correlated with band 3.
(2) Keep the NDVI when it is similar to other indices (SAVI or NDWI).
(3) Keep dates in the interest of maintaining phenological diversity.
(4) Keep dates with the least NoData values.

5.3. Hybrid classification: unsupervised phase

Unsupervised classification was performed using the IsoMM module of the MiraMon software (Pons 2006), based on the classic IsoData classifier (Duda and Hart 1973). This module allows the definition of thousands of clusters on mixed data type rasters (real, 2 byte integer, etc.) and also permits the classification of pixels having NoData values in some variables (for example, due to clouds on some dates). IsoMM begins by selecting initial centroids, systematically distributed in the statistical and geographic space, in order to cluster pixels that are close in the multidimensional data space (Pons 2006; Moré, Serra, and Pons 2006; Pons and Arcalís 2012). Initial clusters are generated using Euclidean distance. The process continues by adjusting centroids at each iteration, based on the pixel values clustered in the previous iteration. A convergence threshold is established in order to limit the number of iterations, based on the maximum percentage of pixels that are allowed to change between two iterations, that is, percentage of unstable pixels remaining at each iteration. As a complementary criterion, the user can set the maximum number of iterations to be executed. In this study the convergence threshold was set to 2%, while the iterations limit was set to 35. The choice for iterations limit and convergence threshold was made after setting the number of iterations to 50 and a convergence threshold of 1% for the 2000–2004 quinquennium in the southern scene. During the iterative process the convergence threshold remained in the 2–3% range between iterations 31 and 50, so we decided to apply the same setting in all cases.

At this phase the two other key parameters are the minimum number of pixels in a cluster and the minimum similarity between statistical classes (relevant for merging two clusters into the same statistical class). The number of pixels in a cluster was set to 50 to ensure representativeness (a minimum of 4.5 ha per statistical class). With this choice, the objective was to attempt to isolate deviances in reflectance behavior due to particular contexts, since the spectral variability for any land cover in a heterogeneous landscape is influenced by different topo-climatic variables as well as human factors. This choice also eliminates noise-prone statistical classes. The minimum similarity between statistical classes was set to 1, ensuring a large number of spectral classes, since two classes must be very similar to each other (accepting average reflectance differences of less than 0.14%) for them to be merged into one. Finally, the parameter for tolerance of NoData values (the maximum number of bands with a NoData value allowed in a pixel in order to classify it using the remaining variables) was set to 32 according to the following criterion: a pixel can be classified using a minimum of 2 dates (5 bands per date plus indices) and the related auxiliary variables (DEM, solar radiation and slopes). This implies that classifying a pixel requires a minimum of 19 variables.

IsoMM and kNN can perform as parallelized processes, allowing the use of more than one thread in 32 and 64 bit computers. Processing time was reduced to less than one-fifth using the 64 bit parallelized version.

5.4. Hybrid classification: supervised phase

Using the MiraMon ClsMix tool, statistical classes are grouped by thematic classes (land use and land cover categories) using a set of training areas. Each thematic category is defined by a set of statistical classes which are associated because the user has indicated that these statistical classes constitute different 'facets' of the thematic class. A statistical class can be left unclassified if none of its pixels intersect geographically with training areas.

Fidelity and representativeness are the main setting parameters of ClsMix (Serra, Pons, and Saurí 2003; Moré, Serra, and Pons 2006; Pons 2006; Pons et al. 2014c). We established a value of 40 for fidelity, which means that a statistical class belongs to a thematic class if at least 40% of its pixels are inside the training areas of that thematic class. We empirically determined this threshold since ClsMix is a very fast process and a massive testbed can be easily applied. Representativeness of a statistical class for a thematic class is the minimum proportion of the statistical class within the thematic class according to training areas; we set it to 0.001, thereby accepting even those statistical classes that represented a modest contribution to the thematic class. Some examples illustrating these parameterization decisions are shown in the following section.

6. Results and discussion

NDVI thresholds applied to the training areas allowed a better distinction between vegetation layers and other covers such as bare soil. Two examples for deciduous evergreen forest are shown in Figure 3. Upper images correspond to a SIOSE polygon with 45% broadleaf evergreen forest cover, 40% shrubland cover, and 15% grassland cover. Using the NDVI thresholds, scattered trees are classified as shrubland, but we prefer to isolate dense forest layers due to the fact that when thresholds are not applied most of the shrublands are classified as broadleaf evergreen forest. Bottom images correspond to a SIOSE polygon with 55% broadleaf, 30% shrubland, 10% grassland and 5% bare soil. Although this example shows an overestimation of the shrubland layer, we still prefer a category defined as shrublands/scattered trees for later analysis of land cover change as opposed to accepting nearly 100% of the area as broadleaf evergreen forest.

In Figure 4 we can see the validity of this approach. In these examples, the SIOSE polygons have not been used as training areas, but land cover classification results are consistent with database covers: 45% broadleaf evergreen forest, 20% broadleaf deciduous forest, 15% grassland, 15% shrubland and 5% bare soil in upper images; 45% broadleaf evergreen forest, 30% coniferous forests, 15% shrubland, 5% grassland and 5% bare soil in bottom images.

After the filtering process using NDVI thresholds, training and test areas of coniferous forests, broadleaf evergreen forests, broadleaf deciduous forests, shrublands, grasslands, urban areas and water bodies were only composed of pixels within the set range (Figure 5). Once the geometry and extent of these areas were adjusted, thematic assignment for statistical classes was also defined (Figure 6) using the ClsMix-assigned thematic categories describe above.

In order to assess the fidelity parameter, we built several testbeds evaluating OA, APA and CA for the 2000–2004 quinquennium in both scenes (Figure 7). After testing we decided to apply a threshold of 40% because this value resulted in maximum accuracy in most cases. The other parameter, representativeness, was set to 0.001% because, although this value may seem to be a modest representation of the whole scene, some clusters assigned to thematic categories using this threshold represented a considerable geographical area (e.g. 1155 ha for cluster 1696 in NUA bloc of the northern scene, or 1533 ha for cluster 2036).

After filtering with NDVI thresholds, the number of training and test pixels in the southern scene was 7,186,736 (20.5% of the scene area) and there were 3,899,873 in northern scene (11.4% of the

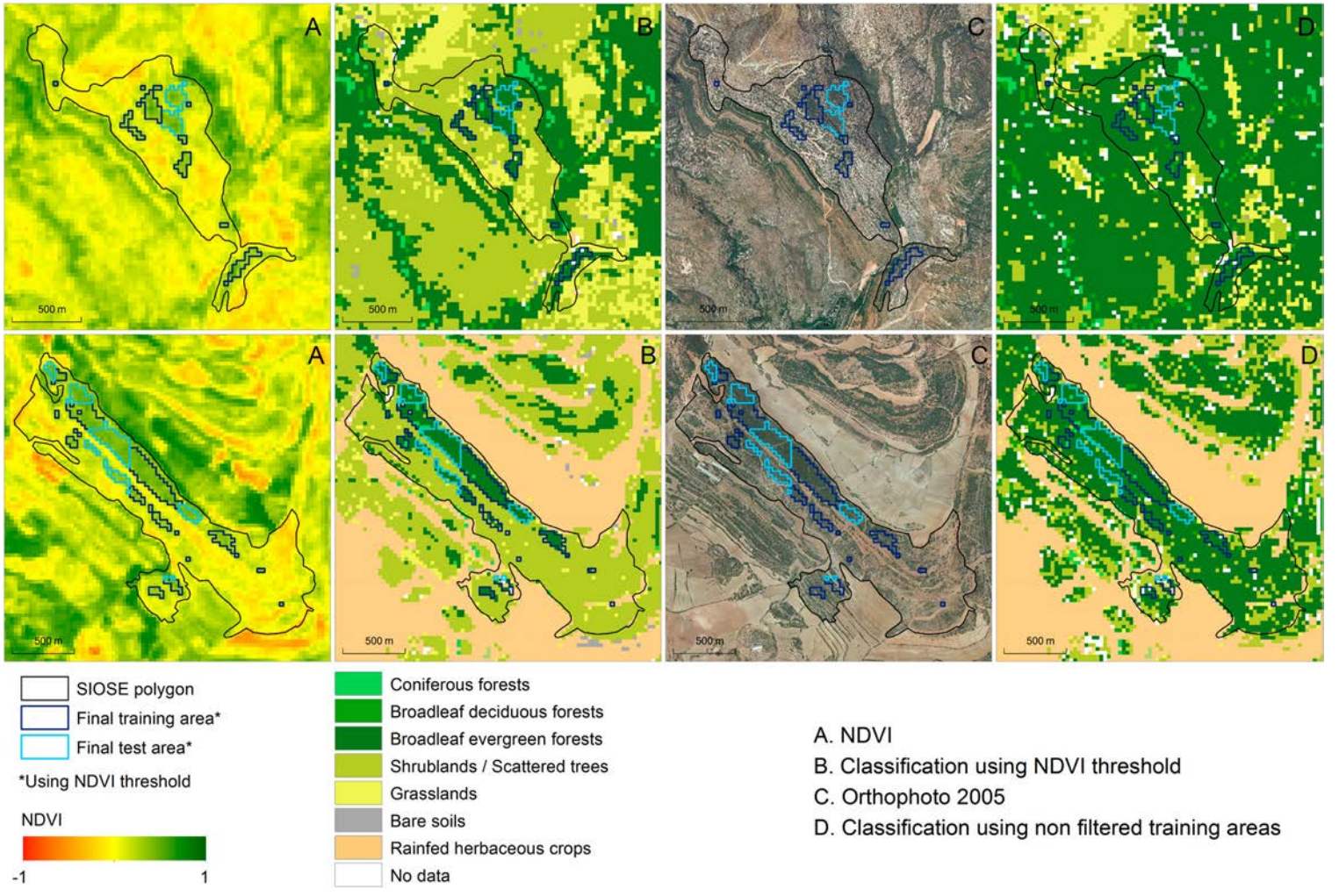


Figure 3. Differences in classification using thresholds vs. non-filtered areas for two polygons (top and bottom rows).

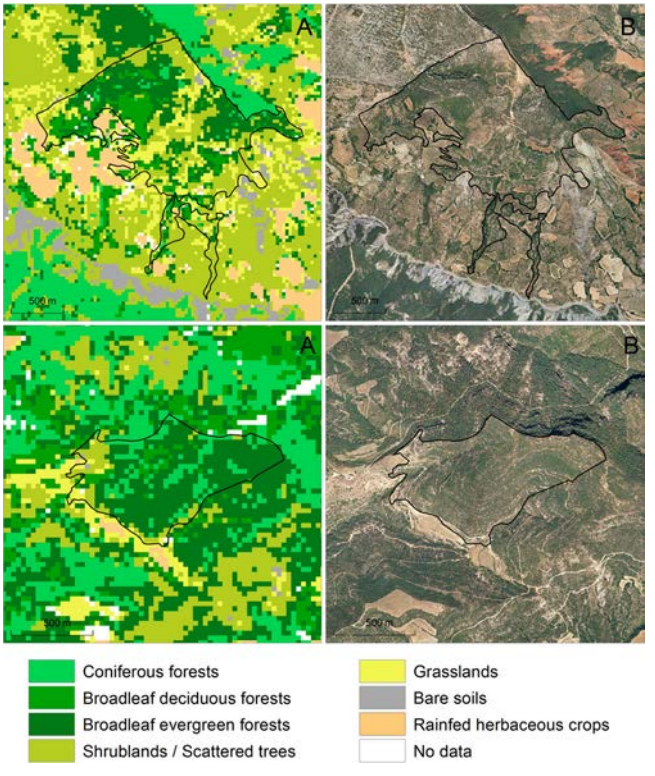


Figure 4. Land cover classification in areas where the SIOSE polygon has not been used for the training process in two polygons (top and bottom rows).

scene area). Once the kNN classifier was run to identify common pixels (belonging to the same category) in both quinquennia, the set of remaining pixels (having an invariant statistical pattern) was reduced to 4,744,942 and 2,837,801 for the southern and northern scenes, respectively, or 13.5% and 8.3% of the scene areas, respectively.

For determination of the k value, we used a testbed composed of several kNN classifications, using values from 2 to 120, every k = 2 from 2 to 20 and every k = 5 from 20 to 120 (i.e. 29 kNN classifications). In this procedure we only used pixels from the 2000–2004 quinquennium clipped to the ambit of the training areas because this significantly reduced processing times of kNN. Differences in accuracy leading from the use of different k values were small; OA was always above 94%. For both

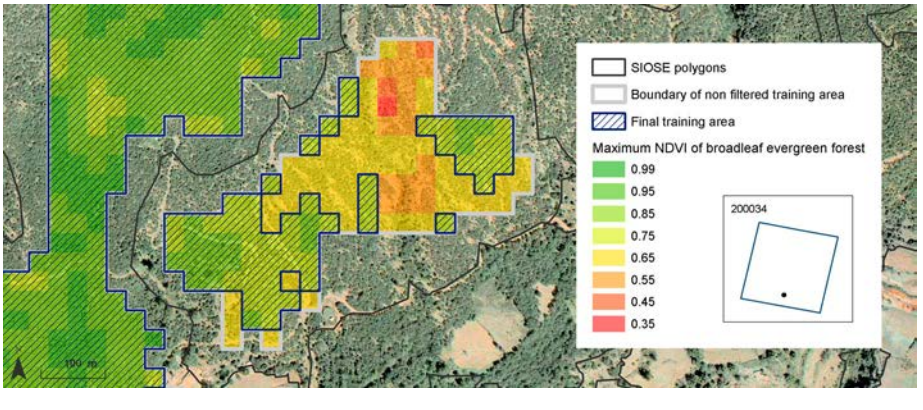


Figure 5. Selection of pixels in training areas according to NDVI thresholds (background: PNOA 2005 orthophoto).

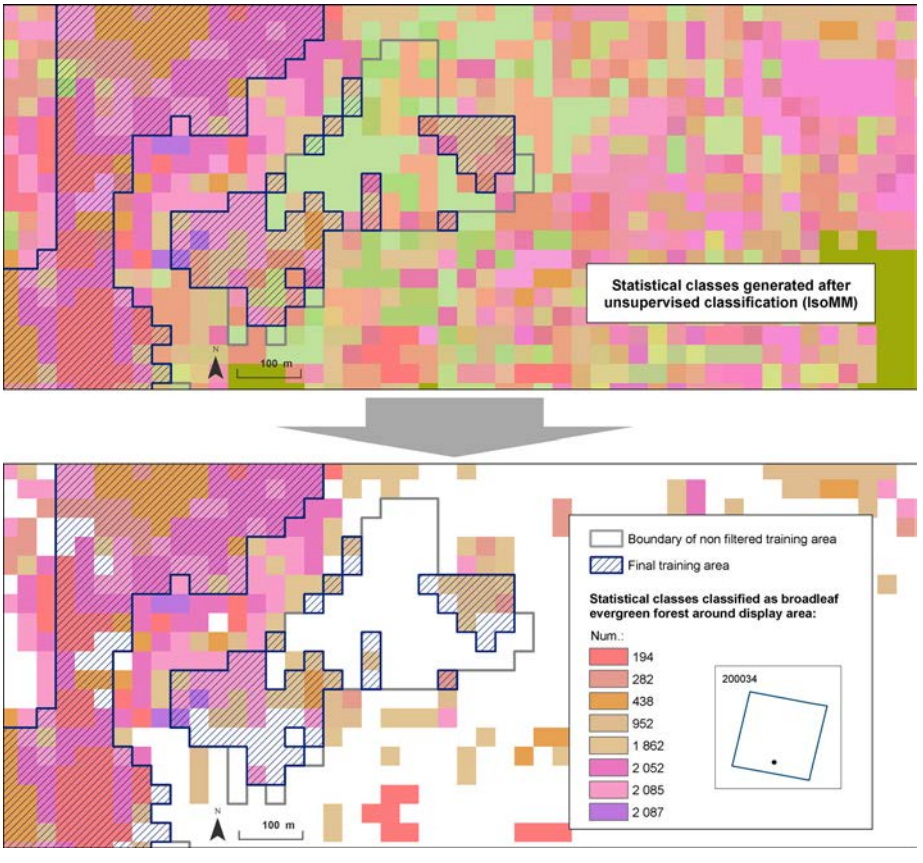


Figure 6. Statistical classes assigned to the broadleaf evergreen forest category. A legend of statistical class color is only shown in the bottom image due the large number of classes in the upper image.

scenes, average producer accuracy was 85–88% from k = 2 to k = 15 and progressively decreased with higher k values, reaching 75–80% for k values above 90. Although differences between k = 2 and k = 15 were very small, we decided to use k = 15, assuming that this would provide superior robustness for a large number of training pixels.

After exploring collinearity, classifications with different subsets of variables were performed for the 2000–2004 period. In order to select the optimum dataset for each classification, we identified the inflexion point at which accuracy indexes (OA, APA and OA_w) and CA clearly decreased (Figure 8). For NUA blocs that point was identified around subdatasets 31 and 33 (removing 20 and 18 variables, respectively). For CRA blocs, only the southern scene (200034) showed a more apparent downtrend from subdataset 41. The tendencies in the CRA bloc of northern scene (199031) were more stable than the other cases, so we decided to select subdataset 33 due to the fact it had the highest APA value. Note that OA_w, which accounts for NoData pixels in test areas, is highly related with percentage of CA. When making these decisions, we also took into consideration that when more variables are removed, a pixel with NoData values in any of its variables is less likely to be classified using the remaining variables according to the NoData threshold of the IsoMM module.

Collinearity analysis enables us to formulate some general criteria for reducing variable number, to be applied in later classifications for other scenes of the Iberian Peninsula. The results show that band 2, band 7 and SAVI index can be removed for NUA blocks; also, it is possible to remove dates between 1 July and 15 August in time series containing more than one date within that range. However, these criteria could not be broadly applied for CRA blocks. Therefore, in the northern scene, the

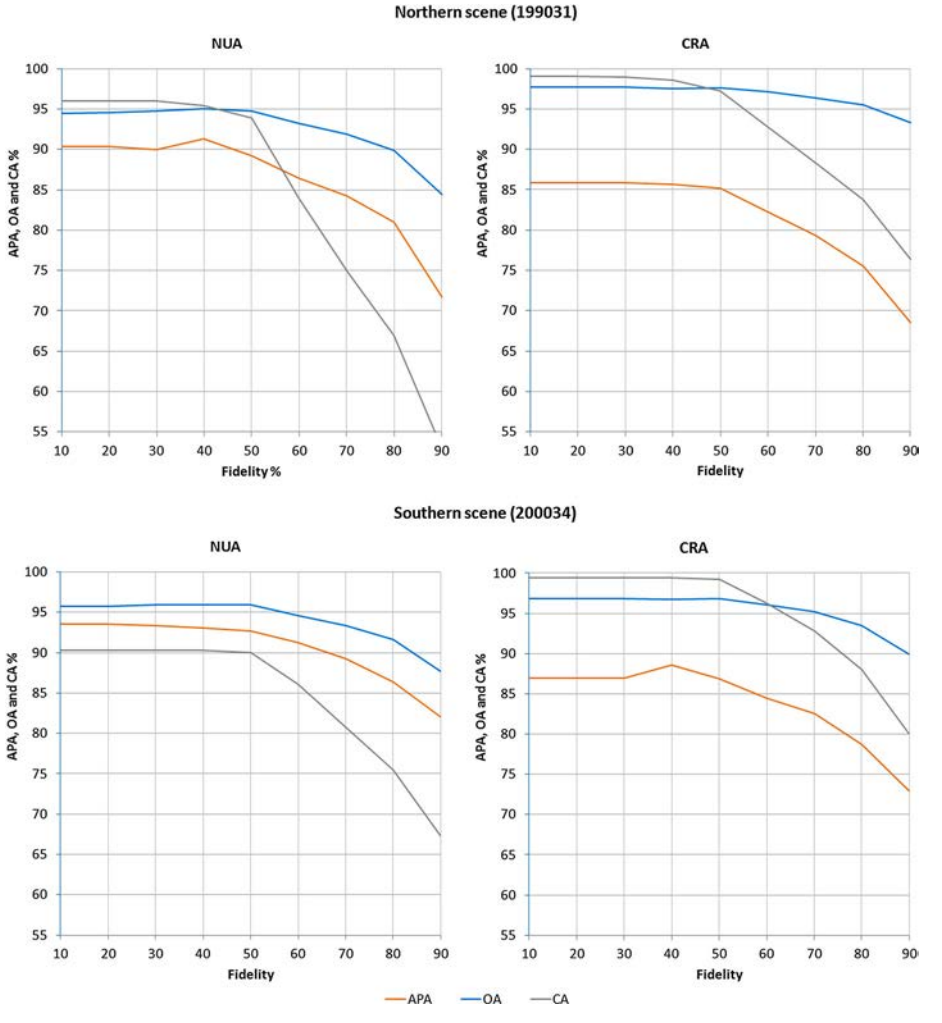


Figure 7. APA, OA and CA based on the fidelity parameter.

same variables removed in NUA blocks have been removed in CRA blocks, with the exception that all summer dates have been kept. Only 10 variables were removed in the southern scene because some categories were especially sensitive to variable removal (e.g. irrigated woody crops). In CRA blocks, SAVI index was the only variable removed at all dates. Accuracy in crop classification for the southern scene is more affected when a large number of variables are removed, possibly due to more complex phenological changes. The most meaningful variables were bands 3, 4 and 5, NDVI, NDWI2 indexes, and topo-climatic indexes. Band 7 remained in almost all dates for CRA blocs.

Classifications of two blocks (NUA and CRA) were produced by generating statistical classes (unsupervised classification stage) beginning in each scene with 32,767 initial centroids systematically distributed every 1100 meters in the geographic space. The number of statistical classes obtained in the unsupervised classification is shown in Table 5.

Once the supervised classification was finished, more than 90% of the total area was classified in southern scene and more than 94% in northern scene. Classifications by blocks had an OA of around 95% or higher (see Appendix for details). Before applying the urban mask procedure, producer accuracy in the bare soil category in the northern scene was 72.2% and 70.5% for the 1985–1989

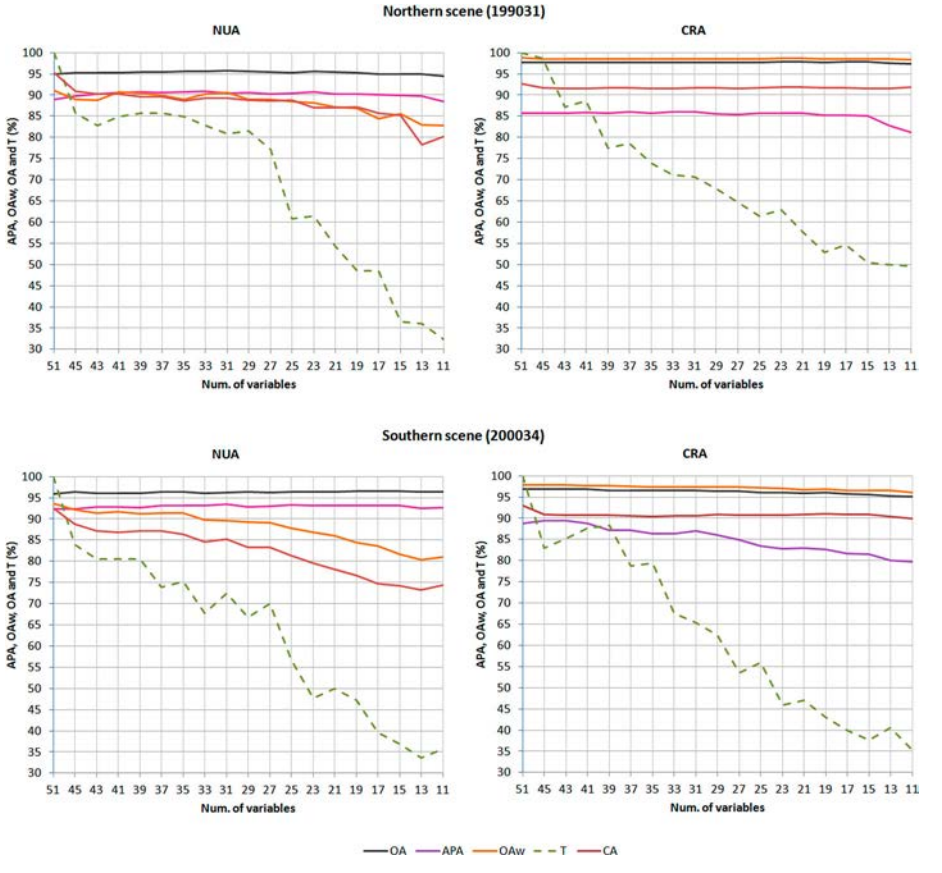


Figure 8. APA, OA, OA_w, T and CA based on the number of variables used in the classification process.

and 2000–2004 periods, respectively, and in the southern scene accuracies were 79.9% and 84.45% in each period, respectively. After this step, producer accuracies in this category in the northern scene were 83% and 84.3% for 1985–1989 and 2000–2004 periods, respectively, and in the southern scene 86.1% and 90.9% for each period, respectively.

A final map for each scene and quinquennium was created through a mosaic of the two blocks incorporating all categories. Then, a validation test was performed using confusion matrices (Tables 6–9); overall accuracies were greater than 92% in all four final maps. A second weighted overall accuracy (OA_{w2}) was calculated considering only those pixels classified in thematic categories, and this accuracy was over 95% in all cases.

The dataset of 51 variables was used to run a SVM classifier for the 2000–2004 quinquennium in both scenes. In this case, all variables were included in order to let the machine learning algorithm to determine the importance of each variable using the same training data. The SVM achieved the best results using a third degree polynomial function, although the differences between other polynomial and linear functions were very small. For the southern scene (200034) OA was 96.7% and OA_{w2} was 97.1%, while for northern scene OA was 96.7% and OA_{w2} was 98.5%. It can be seen that accuracies of the Hybrid classifier and the SVM classifier are very similar, and they are also very similar visually

Table 5. Statistical classes generated after the unsupervised phase.

	Northern scene (199031)		Southern scene (200034)	
	NUA – 31 variables	CRA – 33 variables	NUA – 33 variables	CRA – 41 variables
Number of statistical classes	4645	10,892	3866	6447

Table 6. Southern scene 200034. 1985–1989 quinquennium.

	Reference data															
Classified data	1	2	3	4	5	6	7	8	9	10	11	12	14	Total	CE (%)	UA (%)
NoData	6	26	2	23	55	92	50	16	33	209	315	16,449	69	17,345		
(1) Water bodies	11,518	0	0	0	0	0	5	0	0	8	0	117	0	11,648	1.1	98.9
(2) Coniferous forests	0	24,131	0	946	112	0	0	0	0	0	20	249	0	25,458	5.2	94.8
(3) Broadleaf deciduous forests	0	3	1644	39	0	0	0	0	84	0	112	3	0	1885	12.8	87.2
(4) Broadleaf evergreen forests	0	437	16	5311	170	3	0	0	0	0	50	536	0	6523	18.6	81.4
(5) Shrublands	0	87	0	137	3748	257	7	0	0	0	11	2275	0	6522	42.5	57.5
(6) Grasslands	0	0	2	2	221	11,795	151	12	0	208	105	19,729	0	32,225	63.4	36.6
(7) Bare soils	1	0	0	0	0	49	1315	96	0	320	8	7714	42	9545	86.2	13.8
(8) Urbanized areas	0	0	0	0	0	0	0	7884	0	0	0	0	0	7884	0.0	100.0
(9) Irrigated herbaceous crops	0	0	6	0	0	0	0	0	5262	26	183	481	0	5958	11.7	88.3
(10) Rainfed herbaceous crops	4	0	0	0	0	10	4	18	8	130,789	315	30,822	0	161,970	19.3	80.7
(11) Irrigated woody crops	0	0	0	0	0	0	0	0	97	41	20,175	8056	2	28,371	28.9	71.1
(12) Rainfed woody crops	4	0	2	3	1	21	7	48	251	15,970	12,299	1,092,098	9	1,120,713	2.6	97.4
(14) Greenhouses	0	0	0	0	0	0	0	18	0	0	0	13	22,617	22,648	0.1	99.9
Total	11,533	24,684	1672	6461	4307	12,227	1539	8092	5735	147,571	33,593	1,178,542	22,739	1,458,695		OA (%) = 92.8%
OE (%)	0.1	2.2	1.7	17.8	13.0	3.5	14.6	2.6	8.2	11.4	39.9	7.3	0.5			OA_W (%) = 95.9%
PA (%)	99.9	97.8	98.3	82.2	87.0	96.5	85.4	97.4	91.8	88.6	60.1	92.7	99.5			K = 0.8

Notes: OE = omission error, CE = commission error, PA = producer's accuracy, UA = user's accuracy, K = Kappa statistics. The shaded values are number of correctly classified pixels in the test areas.

Table 7. Southern scene 200034. 2000–2004 quinquennium.

	Reference data															
Classified data	1	2	3	4	5	6	7	8	9	10	11	12	14	Total	CE (%)	UA (%)
NoData	0	0	7	9	43	113	38	4	68	515	227	9604	95	10,723		
(1) Water bodies	11,532	0	0	0	0	0	0	0	0	2	0	8	0	11,542	0.1	99.9
(2) Coniferous forests	0	24,091	1	556	34	0	0	0	0	0	3	54	0	24,739	2.6	97.4
(3) Broadleaf deciduous forests	0	2	1554	34	0	0	0	0	183	0	61	10	0	1844	15.7	84.3
(4) Broadleaf evergreen forests	0	572	87	5671	181	8	0	0	0	1	28	781	0	7329	22.6	77.4
(5) Shrublands	0	19	0	181	3681	414	1	0	0	1	66	3768	0	8131	54.7	45.3
(6) Grasslands	0	0	1	8	323	11,545	99	0	0	721	30	11,282	0	24,009	51.9	48.1
(7) Bare soils	0	0	0	0	2	40	1372	1	1	719	32	13,197	73	15,437	91.1	8.9
(8) Urbanized areas	0	0	0	0	0	0	0	8003	0	0	0	0	0	8003	0.0	100.0
(9) Irrigated herbaceous crops	0	0	20	0	0	0	0	0	5199	131	23	280	0	5653	8.0	92.0
(10) Rainfed herbaceous crops	1	0	2	0	0	51	7	7	93	137,816	106	12,356	4	150,443	8.4	91.6
(11) Irrigated woody crops	0	0	0	0	0	0	0	1	22	21	18,566	8283	9	26,902	31	69.0
(12) Rainfed woody crops	0	0	0	2	43	56	20	45	167	7644	14,448	1,118,913	15	1,141,353	2.0	98.0
(14) Greenhouses	0	0	0	0	0	0	2	31	2	0	3	6	22,543	22,587	0.2	99.8
Total	11,533	24,684	1672	6461	4307	12,227	1539	8092	5735	147,571	33,593	1,178,542	22,739	1,458,695		OA (%) = 94.6%
OE (%)	0.0	2.4	7.1	12.2	14.5	5.6	10.9	1.1	9.3	6.6	44.7	5.1	0.9			OA_W (%) = 96.9%
PA (%)	100.0	97.6	92.9	87.8	85.5	94.4	89.1	98.9	90.7	93.4	55.3	94.9	99.1			$K = 0.8$

Notes: OE = omission error, CE = commission error, PA = producer's accuracy, UA = user's accuracy, K = Kappa statistics. The shaded values are number of correctly classified pixels in the test areas.

Table 8. Northern scene 199031. 1985–1989 quinquennium.

	Reference data															
Classified data	1	2	3	4	5	6	7	8	9	10	11	12	13	Total	CE (%)	UA (%)
NoData	14	11	22	12	81	156	17	80	85	355	71	106	78	1088		
(1) Water bodies	20,809	0	0	0	0	0	82	5	0	0	0	0	3	20,899	0.4	99.6
(2) Coniferous forests	0	24,774	89	549	258	7	0	0	0	2	0	0	1	25,680	3.5	96.5
(3) Broadleaf deciduous forests	0	64	9,004	177	1	67	0	0	3	1	5	0	0	9322	3.4	96.6
(4) Broadleaf evergreen forests	0	488	130	3390	22	6	0	0	0	0	0	0	0	4036	16.0	84.0
(5) Shrublands	0	200	2	28	5020	786	15	7	1	93	0	50	1	6203	19.1	80.9
(6) Grasslands	0	7	27	7	414	11,328	76	108	6	2676	9	423	10	15,091	24.9	75.1
(7) Bare soils	21	0	0	0	20	74	921	162	3	267	5	26	11	1510	39.0	61.0
(8) Urbanized areas	0	0	0	0	0	0	0	9045	0	0	0	0	0	9045	0.0	100.0
(9) Irrigated herbaceous crops	0	0	10	0	0	1	2	4	57,697	2082	1064	59	627	61,546	6.3	93.7
(10) Rainfed herbaceous crops	0	0	0	0	1	50	173	306	3041	670,177	143	1198	94	675,183	0.7	99.3
(11) Irrigated woody crops	0	0	512	0	1	9	0	20	1440	40	18,565	182	159	20,928	11.3	88.7
(12) Rainfed woody crops	0	0	0	0	2	10	2	105	95	793	168	7609	120	8904	14.5	85.5
(13) Rice Crops	8	0	4	0	0	1	0	19	352	43	121	65	6333	6946	8.8	91.2
Total	20,852	25,544	9800	4163	5820	12,495	1288	9861	62,723	676,529	20,151	9718	7437	866,381		OA (%) = 97.6%
OE (%)	0.2	3.0	8.1	18.6	13.7	9.3	28.5	8.3	8.0	0.9	7.9	21.7	14.8			OA_W (%) = 98.9%
PA (%)	99.8	97.0	91.9	81.4	86.3	90.7	71.5	91.7	92.0	99.1	92.1	78.3	85.2			K = 0.9

Notes: OE = omission error, CE = commission error, PA = producer's accuracy, UA = user's accuracy, K = Kappa statistics. The shaded values are number of correctly classified pixels in the test areas.

Table 9. Northern scene 199031. 2000–2004 quinquennium.

	Reference data															
Classified data	1	2	3	4	5	6	7	8	9	10	11	12	13	Total	CE (%)	UA (%)
NoData	3	21	11	10	27	49	10	47	251	485	174	130	75	1293		
(1) Water bodies	20,841	0	0	0	0	0	81	11	0	0	0	0	0	20,933	0.4	99.6
(2) Coniferous forests	0	24,835	133	526	154	22	0	0	1	6	3	0	0	25,680	3.3	96.7
(3) Broadleaf deciduous forests	0	119	9140	211	2	63	0	0	41	3	85	0	4	9668	5.5	94.5
(4) Broadleaf evergreen forests	0	387	123	3366	109	38	0	0	0	0	0	0	0	4023	16.3	83.7
(5) Shrublands	0	168	1	38	4887	1009	25	0	9	291	9	138	1	6576	25.7	74.3
(6) Grasslands	0	10	42	12	615	11,053	61	23	65	3814	29	438	17	16,179	31.7	68.3
(7) Bare soils	8	0	0	0	12	77	931	165	30	628	13	21	44	1929	51.7	48.3
(8) Urbanized areas	0	0	0	0	0	0	0	9294	0	0	0	0	0	9294	0.0	100.0
(9) Irrigated herbaceous crops	0	0	52	0	2	6	2	2	55,366	3255	1619	182	772	61,258	9.6	90.4
(10) Rainfed herbaceous crops	0	0	8	0	8	151	173	306	4184	666,666	128	1884	196	673,704	1.0	99.0
(11) Irrigated woody crops	0	2	287	0	3	10	1	1	2170	202	17,745	308	249	20,978	15.4	84.6
(12) Rainfed woody crops	0	0	0	0	1	17	4	12	173	1097	264	6591	39	8198	19.6	80.4
(13) Rice Crops	0	2	3	0	0	0	0	0	433	82	82	26	6040	6668	9.4	90.6
Total	20,852	25,544	9800	4163	5820	12,495	1288	9861	62,723	676,529	20,151	9718	7437	866,381		OA (%) = 96.7%
OE (%)	0.1	2.8	6.7	19.1	16.0	11.5	27.7	5.7	11.7	1.5	11.9	32.2	18.8			OA_W (%) = 98.8%
PA (%)	99.9	97.2	93.3	80.9	84.0	88.5	72.3	94.3	88.3	98.5	88.1	67.8	81.2			$K = 0.9$

Notes: OE = omission error, CE = commission error, PA = producer's accuracy, UA = user's accuracy, K = Kappa statistics. The shaded values are number of correctly classified pixels in the test areas.

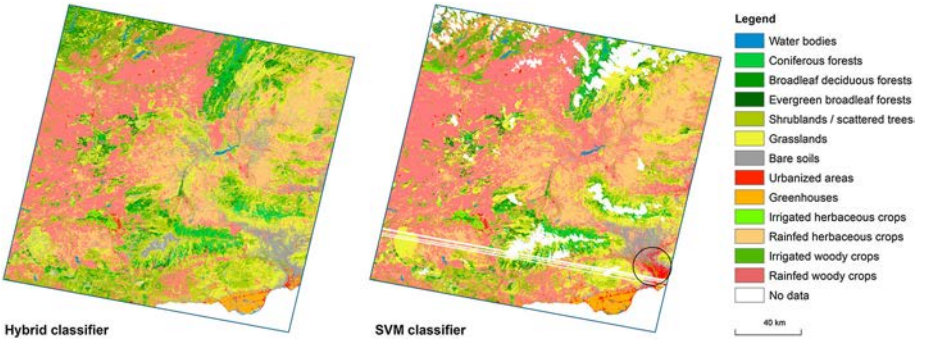


Figure 9. Comparison of results of the hybrid (using the urban mask procedure) and SVM classifiers in the southern scene. In the SVM classification, the black circle indicates the area where the commission errors related to bare soils and urban categories occurs. White areas correspond to pixels left unclassified.

(Figures 9 and 10); nevertheless, pixels with NoData values in any of the variables could not be classified using this particular SVM classifier, making the Hybrid classifier product the preferred choice (see the large unclassified areas in Figures 9 and 10 for the SVM classifier, obtained using typical SVM parameters). Misclassifications, such as the commission errors related to bare soils and urban categories (i.e. without overlapping with urban mask) were also found in SVM land cover classifications (Figure 9): producer accuracies in bare soil categories were 73.2% in the southern scene and 63.5% in the northern scene.

In addition to OA, it is prudent to analyze the errors for each thematic category since there are certain categories that deserve special attention. These cases are identified by high omission error (i.e. for a given category, the percentage of pixels in test areas which have been classified in a different category on the map) or high commission error (i.e. for a given category the percentage of pixels on the map which incorrectly appears in one or more test areas). These error types assess the producer's accuracy (omission errors) and the user's accuracy (commission errors) (Pons and Arcalís 2012; Olofsson et al. 2013; Sexton et al. 2013b).

Broadleaf evergreen forest (4) had an omission error of 12.2–19.1% and commission error of 16.0–22.6%. For the northern scene, the broadleaf deciduous forest (3) and coniferous forest (2) are the main categories contributing to the errors. This is to some degree expected since mixed forests have a large extension throughout the northern part of the scene. Pyrenean foothills represent a transition area between Mediterranean and Atlantic influences, and this is reflected in confounding of forest categories. There could also be some photointerpretation error in SIOSE polygons since it is difficult to separate textures of the three cover types in contexts of high mixture, and determination of precise percentages and cover limits would require exhaustive fieldwork. In the southern scene,

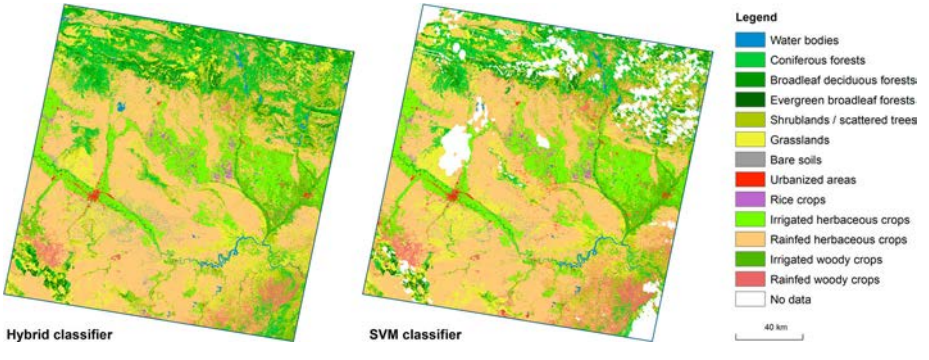


Figure 10. Comparison of results of the hybrid (using the urban mask procedure) and SVM classifier in the northern scene. White areas correspond to pixels left unclassified.

broadleaf deciduous forests (3) have a much lower inference on the classification of broadleaf evergreen forest (4) due to a low environmental suitability at these latitudes. Instead, in the southern scene shrublands (5) and rainfed woody crops (12) are responsible for a notable proportion of total error.

We also found occurrence of (somewhat expected) omission errors confounding *grasslands* (6) with *shrublands* (5) and vice versa. This can be explained by ecological succession processes (abandoned crop areas, tree line region, areas of expansion of forest plantations, etc.) and coexistence of both categories due to climacic conditions. For these categories we found most notable commission errors (42.5–63.4% in the southern scene and 19.1–31.7% in the northern scene). In the southern scene, the commission error was associated with the rainfed woody crops category (12). These errors have been analyzed by photointerpretation (1984–1985 and 2004 orthophotos), and we found that most of plots were either smallholdings or plots surrounded by hedgerows, both considered to be part of the traditional cultural landscape (fine-grained landscape); such features have a very compartmentalized structure and lower accessibility than bigger plots, and are also characterized by crops located at heterogeneous relief contexts. In light of the above, we assume that the natural vegetation signal has greater intensity in situations of ecological succession after land abandonment, or in areas with large hedgerows surrounding small plots. On the other hand, in the northern scene rainfed herbaceous crops (10) had more commission error in the grasslands (6) and shrublands (5) categories, especially in the 2000–2004 quinquennium (Table 9). This may be related to fallow periods of rainfed herbaceous crops or due to some differences between the moment of image acquisition and the updating of the SIOSE 2005 database.

Omission errors in the *bare soil* category (7) have higher percentages in the northern scene due to confounding with rainfed herbaceous crops (12). This may be due to the spectral influence of soil, which is much more intense in these crops during the fallow period or after harvest, and can cause commission errors. However, the greatest error percentages (up to 91.1%) are commissions with the rainfed woody crops category (12) in the southern scene (Tables 6 and 7). This is a foreseeable effect due to a greater distance between treetops in large plots or in young-growth crops (as ascertained through visual examination of the orthophoto), which increases the soil spectral response. Nevertheless, these commission errors had little influence on OA since rainfed woody crops represent a very large area compared to areas where the commission error occurs.

In the northern scene, *rice crops* (13) had an omission error of 14.8–18.8% because there are irrigated herbaceous crops (9) with a traditional flood irrigation system, and the spectral response of both categories is similar. *Rainfed woody crops* (12) had a 21.7% omission error in the 1985–1989 quinquennium and 32.2% for 2000–2004, owing to confounding with rainfed herbaceous crops (10). By intersecting test areas and the classified map (thus obtaining a common area between the two map layers) we were able to spatially isolate this error. After making a second layer based on intersection with SIOSE 2005, it was found that 15% of the common area was grapevine, 46% was non-citrus fruit rainfed crops, and the remaining was rainfed olive crops. After photointerpretation analysis using 2005 orthophotos of the two latter crop types (grouped into the rainfed woody crops category) in areas where confusion occurred, it was observed that the width between treetops may have increased the spectral response of herbaceous vegetation, so this may have been the source of error.

Finally, a 39.9–44.7% omission error was found in *irrigated woody crops* (11) in the southern scene. In this case, confounding occurs with rainfed woody crops (12) and almost exclusively affects the olive crops. As previously mentioned, these errors may be due to a time lag between Landsat imagery and SIOSE database development. Nevertheless, we must take into account that, according to the Andalusian Government (Junta de Andalucía 2002), a plot is defined as irrigated in cadastral information 'when the whole olive yard is irrigated or if it is any subplot'. Therefore, due to the fact that SIOSE uses cadastral information to define agricultural plots (http://www.siose.es/web/guest/metodologia-de-produccion), some polygons defined as irrigated may contain rainfed areas, even in a large proportion and, consequently, these differences are not really 'errors'.

Figures 11 and 12 show the final land cover maps.

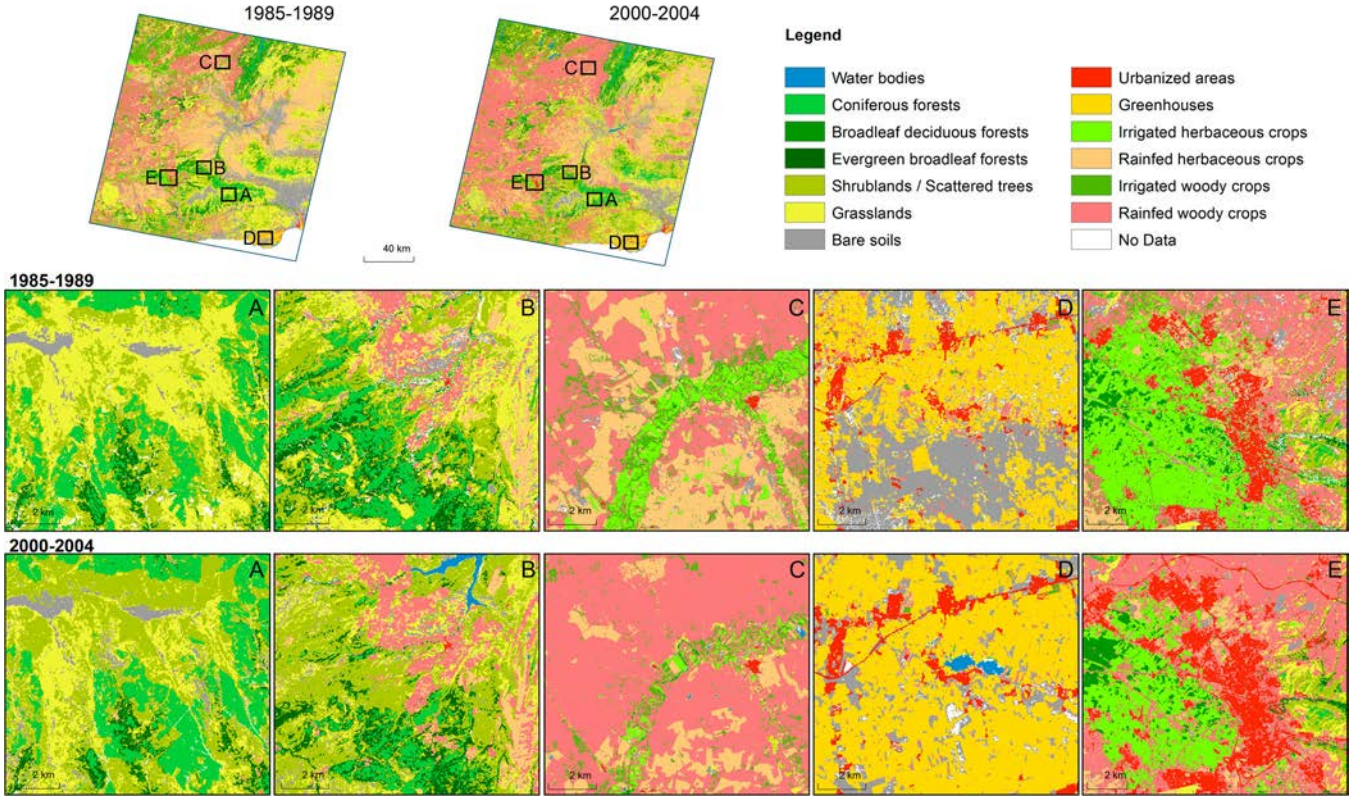


Figure 11. Land cover maps of the southern scene (200034).

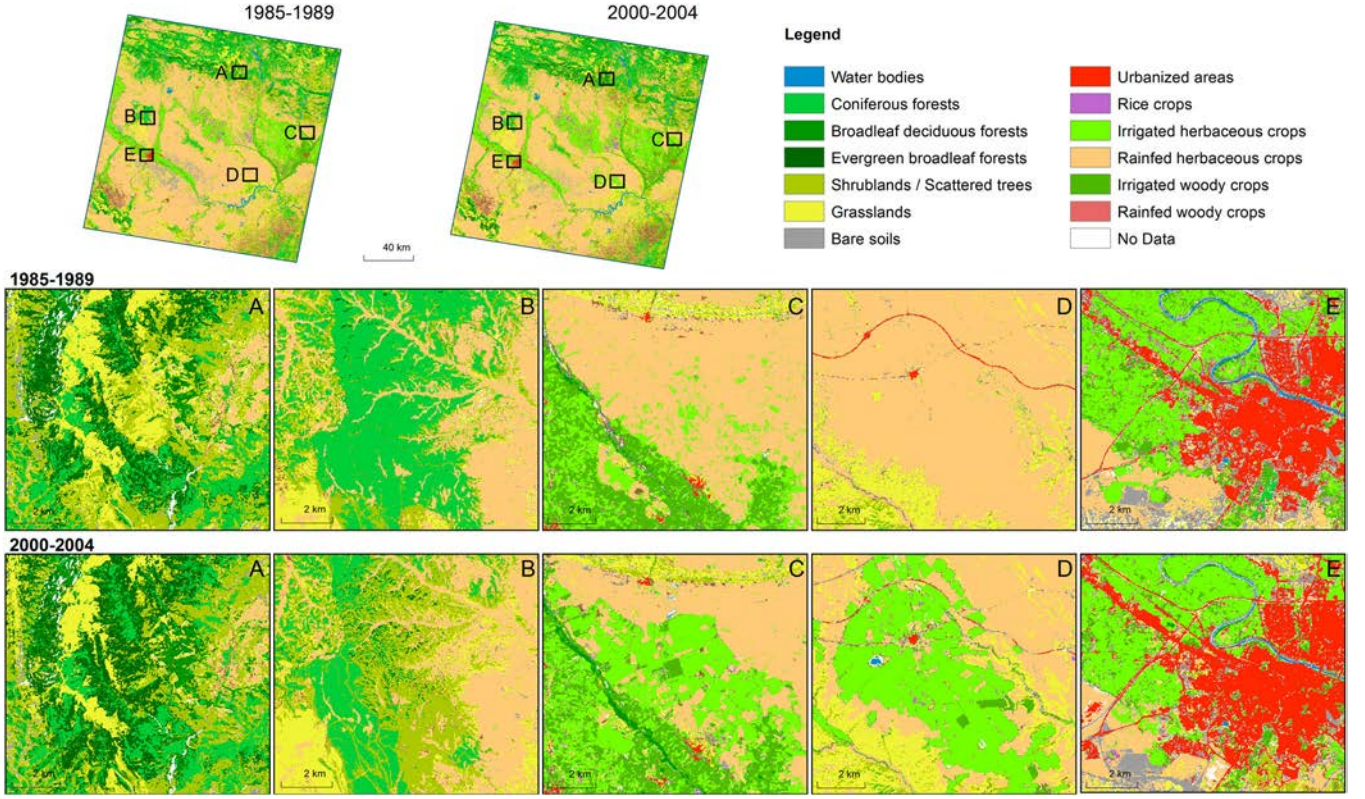


Figure 12. Land cover maps of the northern scene (199031).

7. Conclusions

Classification of multi-temporal Landsat TM and ETM+ satellite imagery with high spatial and spectral resolution allowed mapping of land cover in two large and complex areas with very high levels of accuracy. In this study, this was also accomplished with the production of a very rich legend, demonstrating that the technique is complementary to traditional procedures of manual photointerpretation at the regional scale, and leads to the generation of a sizable number of objective land cover categories. Especially when the chosen regions are very diverse, our method can be very useful for future projects analyzing land cover change because it is capable of comparing land cover changes in diverse regions, such as the Iberian Peninsula, with global or broad regional approaches.

A detailed database such as SIOSE 2005 allows training of the classifier and also informs the user about the quality of the output map after a filtering process. In this study, we considered it necessary to sieve this reference data because most of the records in the original database had a high disaggregation, making it difficult to configure a more simple legend.

The provision of reasonably pure training data for large areas of the Iberian Peninsula allowed the analysis of its temporal variability. In order to classify periods with no reference information, the kNN classifier was used to identify invariant pixels over the different quinquennia, solving the particular problem of lacking this reference information.

Backdating process identifies invariant pixels in both periods, which means that pure or nearly pure pixels are used for validation. Accuracy values may be distorted in areas where landscape structure is highly heterogeneous, but if the main goal of a land cover change analysis is to identify the main vectors of change, we believe that our land cover maps are suitable products.

Filtering processes and backdating for training areas can be reproducible methodologies for other areas of the world where detailed, up to date databases are available from national or international institutions. Although this caliber of digital information may not be available for many areas of the world, our procedure could be used after a manual process for obtaining training areas (through visual interpretation of aerial imagery of only one of the dates) to give these training areas greater coherence.

Computer processing of such a large volume of data requires appropriate techniques, and the nonparametric approach involving a Hybrid classifier (IsoMM + ClsMix) provides results in relatively short processing times. High levels of accuracy can be achieved with decision-making flexibility after conveniently discarding some of the original variables, permitting high accuracy and conservation of a large proportion of the CA. Results of the hybrid and SVM classifiers showed very slight differences, but we still prefer to use the Hybrid classifier because it is less of a 'black box' model in terms of its parameterization. Also, in the hybrid approach, the IsoMM + ClsMix classifier allows fast and direct classification of pixels with NoData values. Collinearity analysis allowed a reduction of the number of variables used; this improved performance and will also help improve workflows in future classifications because they can be performed with fewer variables.

All of the described steps using the IsoMM + ClsMix model or a similar Hybrid classifier are reproducible for the entire area of Spain, and can also be reproduced in other areas with a homogeneous distribution of reliable training areas. Although several gigabytes must be processed for each quinquennium and scene, our approach will be useful once integrated into the larger project associated with our study, whose aim is to obtain land cover and land cover change products for the entire Iberian Peninsula.

Acknowledgements

The authors acknowledge the support from Lluís Pesquer (Grumets Research Group, CREAF, Universitat Autònoma de Barcelona) and Oscar González-Guerrero (Grumets Research Group, Geography Department, Universitat Autònoma de Barcelona) for processing pseudoinvariant areas in the radiometric correction phase. Xavier Pons is recipient of an ICREA Academia Excellence in Research Grant (2016–2020).

Funding

This work was supported by the Spanish Ministry of Economy and Competitiveness [grant number BES-2013-063766]; European Union's Horizon 2020 Programme [ECOPOTENTIAL (641762-2)]; Spanish Ministry of Economy and Competitiveness [ACAPI (CGL2015-69888-P MINECO/FEDER)], [DinaClive (CGL2012-33927)]; Catalan Government [SGR2014-1491].

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCiD

Juan José Vidal-Macua http://orcid.org/0000-0002-9897-7383
Alaitz Zabala http://orcid.org/0000-0002-3931-4221
Miquel Ninyerola http://orcid.org/0000-0002-1101-0453
Xavier Pons http://orcid.org/0000-0002-6924-1641

References

Arino, O., D. Gross, F. Ranera, L. Bourg, M. Leroy, P. Bicheron, J. Latham, et al. 2007. "GlobCover: ESA Service for Global Land Cover from MERIS." Paper presented at the Geoscience and Remote Sensing Symposium, (IGARSS) 2412–2415. Barcelona, July 23–27. doi:10.1109/IGARSS.2007.4423328.

Baret, F., and G. Guyot. 1991. "Potentials and Limits of Vegetation Indices for LAI and APAR Assessment." *Remote Sensing of Environment* 35 (2): 161–173. doi:10.1016/0034-4257(91)90009-U.

Beltrán, B. J., J. Franklin, A. D. Syphard, H. M. Regan, L. E. Flint, and A. L. Flint. 2014. "Effects of Climate Change and Urban Development on the Distribution and Conservation of Vegetation in a Mediterranean Type Ecosystem." *International Journal of Geographical Information Science* 28 (8): 1561–1589. doi:10.1080/13658816.2013.846472.

Bonan, G. B. 2008. "Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests." *Science* 320 (5882): 1444–1449. doi:10.1126/science.1155121.

Cea, C., J., Cristóbal, and X. Pons. 2007. "Determinación de la Superficie Nival del Pirineo Catalán mediante Imágenes Landsat y MODIS." Paper presented at XII Congreso Nacional de Tecnologías de la Información Geográfica. Granada, September 19–22.

Chen, Jun, Jin Chen, A. Liao, X. Cao, L. Chen, X. Chen, C. He, et al. 2015. "Global Land Cover Mapping at 30m Resolution: A POK-based Operational Approach." *ISPRS Journal of Photogrammetry and Remote Sensing* 103: 7–27. doi:10.1016/j.isprsjprs.2014.09.002.

Chuvieco, Emilio. 1996. *Fundamentos de Teledetección espacial*. Madrid: Rialp.

Cowen, D. J., and J. R. Jensen. 1998. "Extraction and Modeling of Urban Attributes Using Remote Sensing Technology." In *People and Pixels: Linking Remote Sensing and Social Science*, edited by D. Liverman, E. F. Moran, R. R. Rindfuss, and P. C. Stern, 164–188. Washington, DC: National Academy Press. doi:10.17226/5963.

Craglia, M., K. de Bie, D. Jackson, M. Pesaresi, G. Remetey-Fülöpp, C. Wang, A. Annoni, et al. 2012. "Digital Earth 2020: Towards the Vision for the Next Decade." *International Journal of Digital Earth* 5 (1): 4–21. doi:10.1080/17538947.2011.638500.

Cushnie, J. L. 1987. "The Interactive Effect of Spatial Resolution and Degree of Internal Variability within Land-cover Types on Classification Accuracies." *International Journal of Remote Sensing* 8 (1): 15–29. doi:10.1080/01431168708948612.

Dozier, J. 1989. "Spectral Signature of Alpine Snow Cover from the Landsat Thematic Mapper." *Remote Sensing of Environment* 28: 9–22. doi:10.1016/0034-4257(89)90101-6.

Duda, R. O., and P. E. Hart. 1973. *Pattern Classification and Scene Analysis*. New York: Wiley. doi:10.1016/s0019-9958(76)90486-1.

EEA (European Environment Agency). 2014. *Terrestrial Habitat Mapping in Europe: An Overview*. EEA Technical report No 1/2014. Joint MNHN-EEA report. doi:10.2800/11055.

Friedl, M. A., D. Sulla-Menashe, B. Tan, A. Schneider, N. Ramankutty, A. Sibley, and X. Huang. 2010. "MODIS Collection 5 Global Land Cover: Algorithm Refinements and Characterization of New Datasets." *Remote Sensing of Environment* 114 (1): 168–182. doi:10.1016/j.rse.2009.08.016.

García-Ruiz, J. M. 1990. *Geoecología de las Áreas de Montaña*. Logroño: Geoforma Ediciones.

Gaveau, D. L., S. Sloan, E. Molidena, H. Yaen, D. Sheil, N. K. Abram, M. Ancrenaz, et al. 2014. "Four Decades of Forest Persistence, Clearance and Logging on Borneo." *Plos One* 9 (7): e101654. doi:10.1371/journal.pone.0101654.

Gil-Olcina, A., and J. Gómez-Mendoza. 2001. *Geografía de España*. Vol. 675. Barcelona: Ariel.

Glanz, H., L. Carvalho, D. Sulla-Menashe, and M. A. Friedl. 2014. "A Parametric Model for Classifying Land Cover and Evaluating Training Data Based on Multi-temporal Remote Sensing Data." *ISPRS Journal of Photogrammetry and Remote Sensing* 97: 219–228. doi:10.1016/j.isprsjprs.2014.09.004.

Gutman, G., C. Huang, G. Chander, and P. Noojipady. 2013. "Assessment of the NASA–USGS Global Land Survey (GLS) Datasets." *Remote Sensing of Environment* 134: 249–265. doi:10.1016/j.rse.2013.02.026.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, et al. 2013. "High-resolution Global Maps of 21st-century Forest Cover Change." *Science* 342 (6160): 850–853. doi:10.1126/science.1244693.

Hall, D. K., G. A. Riggs, and V. V. Salomonson. 1995. "Development of Methods for Mapping Global Snow Cover Using Moderate Resolution Imaging Spectroradiometer Data". *Remote Sensing of Environment* 54 (2): 127–140.

Herrera, J. M., and E. Doblas-Miranda. 2012. "Land-cover Change Effects on Trophic Interactions: Current Knowledge and Future Challenges in Research and Conservation." *Basic and Applied Ecology* 14 (1): 1–11. doi:10.1016/j.baae.2012.11.008.

Homer, C., C. Huang, L. Yang, B. Wylie, and M. Coan. 2004. "Development of a 2001 National Land-cover Database for the United States." *Photogrammetric Engineering & Remote Sensing* 70 (7): 829–840. doi:10.14358/PERS.70.7.829.

Hsu, C. W., C. C. Chang, and C. J. Lin. 2003. "A Practical Guide to Support Vector Classification". National Taiwan University. http://ntu.csie.org/~cjlin/papers/guide/guide.pdf.

Huadong, G. 2016. "Digital Earth and Future Earth. International Journal of Digital Earth". *International Journal of Digital Earth* 9 (1): 1–2. doi:10.1080/17538947.2015.1135667.

Huete, A. R. 1988. "A Soil-adjusted Vegetation Index (SAVI)." *Remote Sensing of Environment* 25 (3): 295–309. doi:10.1016/0034-4257(88)90106-X.

Jin, H., P. Li, T. Cheng, and B. Song. 2012. "Land Cover Classification Using CHRIS/PROBA Images and Multi-temporal Texture." *International Journal of Remote Sensing* 33 (1): 101–119. doi:10.1080/01431161.2011.584077.

Jin, S., L. Yang, P. Danielson, C. Homer, J. Fry, and G. Xian. 2013. "A Comprehensive Change Detection Method for Updating the National Land Cover Database to Circa 2011." *Remote Sensing of Environment* 132: 159–175. doi:10.1016/j.rse.2013.01.012.

Junta de Andalucía. 2002. *El Olivar Andaluz*. Unidad de Prospectiva de la Consejería de Agricultura y Pesca de la Junta de Andalucía y Empresa Pública de Desarrollo Agrario y Pesquero de Andalucía. http://www.juntadeandalucia.es/agriculturaypesca/prospectiva/Olivar4_doc_sinAnexo.pdf.

Kirches, G., O. Arino, M. Boettcher, S. Bontemps, C. Brockmann, O. Danne, P. Defourny, et al. 2013. "CCI Land Cover Pre-processing. Challenges of Pre-processing for Land Cover Classification." Poster presented at ESA Living Planet Symposium. Edinburgh, September 9–13. http://www.researchgate.net/publication/261295278_LandCover_CCI_Pre-processing.

Kriegler, F. J., W. A. Malila, R. F. Nalepka, and W. Richardson. 1969. "Preprocessing Transformations and their Effects on Multispectral Recognition." *Proceedings of the Sixth International Symposium on Remote Sensing of Environment* VI (1): 97–131. http://www.researchgate.net/publication/253819057_Preprocessing_Transformations_and_Their_Effects_on_Multispectral_Recognition.

Lu, D., and Q. Weng. 2006. "Use of Impervious Surface in Urban Land-use Classification." *Remote Sensing of Environment* 102 (1): 146–160. doi:10.1016/j.rse.2006.02.010.

Moré, G., X. Pons, J. A. Burriel, R. Castells, J. J. Ibàñez, and X. Roijals. 2005. "Diferenciación de Cubiertas Forestales para el MCSC a partir de la Clasificación de Imágenes Landsat." *Cuadernos de la Sociedad Española de Ciencias Forestales* 19: 153–162. http://secforestales.org/publicaciones/index.php/cuadernos_secf/article/viewFile/9500/9418.

Moré, G., P. Serra, and X. Pons. 2006. "Improvements on Classification by Tolerating NoData Values - Application to a Hybrid Classifier to Discriminate Mediterranean Vegetation with a Detailed Legend Using Multitemporal Series of Images." Paper presented at the Geoscience and Remote Sensing Symposium (IGARSS), 192–195. doi:10.1109/IGARSS.2006.54.

Mucher, C. A., K. T. Steinnocher, F. P. Kressler, and C. Heunks. 2000. "Land Cover Characterization and Change Detection for Environmental Monitoring of Pan-Europe." *International Journal of Remote Sensing* 21 (6–7): 1159–1181. doi:10.1080/014311600210128.

Olofsson, P., G. M. Foody, S. V. Stehman, and C. E. Woodcock. 2013. "Making Better Use of Accuracy Data in Land Change Studies: Estimating Accuracy and Area and Quantifying Uncertainty Using Stratified Estimation." *Remote Sensing of Environment* 129: 122–131. doi:10.1016/j.rse.2012.10.031.

Ouma, Y. O., and R. Tateishi. 2006. "A Water Index for Rapid Mapping of Shoreline Changes of Five East African Rift Valley Lakes: an Empirical Analysis Using Landsat TM and ETM+ Data." *International Journal of Remote Sensing* 27 (15): 3153–3181. doi:10.1080/01431160500309934.

Parry, M. L. 2007. *Climate Change 2007: Impacts, Adaptation and Vulnerability: Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Vol. 4. Cambridge University Press. https://www.ipcc.ch/pdf/assessment-report/ar4/wg2/ar4_wg2_full_report.pdf.

Pèlachs, A., R. Pérez-Obiol, M. Ninyerola, and J. Nadal. 2009. "Landscape Dynamics of Abies and Fagus in the Southern Pyrenees During the Last 2200 Years as a Result of Anthropogenic Impacts." *Review of Palaeobotany and Palynology* 156 (3): 337–349. doi:10.1016/j.revpalbo.2009.04.005.

Pons, X. 2006. *Geographical Information System and Remote Sensing Software*. Bellaterra: CREAF Research Center. http://www.creaf.uab.es/miramon/.

Pons, X., and A. Arcalís. 2012. *Diccionari Terminològic de Teledetecció*. Barcelona: Enciclopèdia Catalana i Institut Cartogràfic de Catalunya.

Pons, X., and M. Ninyerola. 2008. "Mapping a Topographic Global Solar Radiation Model Implemented in a GIS and Refined with Ground Data." *International Journal of Climatology* 28 (13): 1821–1834. doi:10.1002/joc.1676.

Pons, X., M. Ninyerola, C. Cea, O. González-Guerrero, P. Serra, A. Zabala, L. Pesquer, et al. 2014a. "It's Time for a Crisper Image of the Face of the Earth: Landsat and Climate Time Series for Massive Land Cover & Climate Change Mapping at Detailed Resolution." Poster presented at EGU General Assembly, Vienna, April 27–May 2. In EGU General Assembly Conference Abstracts 16: 13640. http://adsabs.harvard.edu/abs/2014EGUGA..1613640P.

Pons, X., L. Pesquer, J. Cristóbal, O. González-Guerrero. 2014b. "Automatic and Improved Radiometric Correction of Landsat Imagery Using Reference Values from MODIS Surface Reflectance Images." *International Journal of Applied Earth Observation and Geoinformation* 33: 243–254. doi:10.1016/j.jag.2014.06.002.

Pons, X., E. Sevillano, G. Moré, P. Serra, D. Cornford, and M. Ninyerola. 2014c. "Distribución Espacial de la Incertidumbre en Mapas de Cubiertas Obtenidos Mediante Teledetección." *Revista de Teledetección* 42: 1–10. doi:10.4995/raet.2014.3059.

Serra, P., and X. Pons. 2013. "Two Mediterranean Irrigation Communities in Front of Water Scarcity: A Comparison Using Satellite Image Time Series." *Journal of Arid Environments* 98: 41–51. doi:10.1016/j.jaridenv.2013.07.011.

Serra, P., X. Pons, and D. Sauri. 1999. "Anàlisi dels Usos del Sòl de la Plana de l'Alt Empordà i la seva Localització a través de la Teledetecció (1977-1993). Documents d'Analisis Geogràfica" 36: 63–89. http://www.raco.cat/index.php/DocumentsAnalisi/article/viewFile/31714/31548.

Serra, P., X. Pons, and D. Saurí. 2003. "Post-classification Change Detection with Data from Different Sensors: Some Accuracy Considerations." *International Journal of Remote Sensing* 24 (16): 3311–3340. doi:10.1080/0143116021000021189.

Sexton, J. O., X. P. Song, M. Feng, P. Noojipady, A. Anand, C. Huang, D. H. Kim, et al. 2013a. "Global, 30-m Resolution Continuous Fields of Tree Cover: Landsat-based Rescaling of MODIS Vegetation Continuous Fields with LIDAR-based Estimates of Error." *International Journal of Digital Earth* 6 (5): 427–448. doi:10.1080/17538947.2013.786146.

Sexton, J. O., D. L. Urban, M. J. Donohue, and C. Song. 2013b. "Long-term Land Cover Dynamics by Multi-Temporal Classification across the Landsat-5 Record." *Remote Sensing of Environment* 128: 246–258. doi:10.1016/j.rse.2012.10.010.

Strahler, A., D. Muchoney, J. Borak, F. Gao, M. Friedl, S. Gopal, J. Hodges, et al. 1999. *MODIS Land Cover Product, Algorithm Theoretical Basis Document (ATBD) Version 5.0*. Center for Remote Sensing, Department of Geography, Boston University. http://modis.gsfc.nasa.gov/data/atbd/atbd_mod12.pdf.

Toll, D. L. 1984. "An Evaluation of Simulated Thematic Mapper Data and Landsat MSS Data for Discriminating Suburban and Regional Land Use and Land Cover." *Photogrammetric Engineering and Remote sensing* 50: 1713–1724. http://ntrs.nasa.gov/search.jsp?R=19850051406

Townshend, J. R., J. G. Masek, C. Huang, E. F. Vermote, F. Gao, S. Channan, J. O. Sexton, et al. 2012. "Global Characterization and Monitoring of Forest Cover Using Landsat Data: Opportunities and Challenges." *International Journal of Digital Earth* 5 (5): 373–397. doi:10.1080/17538947.2012.713190.

Tullot, I. F. 2000. *Climatologia de España y Portugal*. Vol. 76. Salamanca: Universidad de Salamanca.

Vogelmann, J. E., S. M. Howard, L. Yang, C. R. Larson, B. K. Wylie, and N. Van Driel. 2001. "Completion of the 1990s National Land Cover Data Set for the Conterminous United States from Landsat Thematic Mapper Data and Ancillary Data Sources." *Photogrammetric Engineering and Remote Sensing* 67 (6): 650–662. http://www.mrlc.gov/nlcd1992.php.

Vogelmann, J. E., T. L. Sohl, P. V. Campbell, and D. M. Shaw. 1998. "Regional Land Cover Characterization Using Landsat Thematic Mapper Data and Ancillary Data Sources." *Environmental Monitoring and Assessment* 51 (12): 415–428. doi:10.1023/A:1005996900217.

Wickham, J. D., S. V. Stehman, J. A. Fry, J. H. Smith, and C. G. Homer. 2010. "Thematic Accuracy of the NLCD 2001 Land Cover for the Conterminous United States". *Remote Sensing of Environment* 114 (6): 1286–1296. doi:10.1016/j.rse.2010.01.018.

Wulder, M. A., J. C. White, S. N. Goward, J. G. Masek, J. R. Irons, M. Herold, W. B. Cohen, T. R. Loveland, and C. E. Woodcock. 2008. "Landsat Continuity: Issues and Opportunities for Land Cover Monitoring." *Remote Sensing of Environment* 112 (3): 955–969. doi:10.1016/j.rse.2007.07.004.

Xian, G., C. G. Homer, and C. L. Aldridge. 2012. "Effects of Land Cover and Regional Climate Variations on Long-term Spatiotemporal Changes in Sagebrush Ecosystems." *GIScience & Remote Sensing* 49 (3): 378–396. doi:10.2747/1548–1603.49.3.378.

Xian, G., C. Homer, and J. Fryc. 2009. "Updating the 2001 National Land Cover Database Land Cover Classification to 2006 by Using Landsat Imagery Change Detection Methods." *Remote Sensing of Environment* 113 (6): 1133–1147. doi:10.1016/j.rse.2009.02.004.

Zhu, Z., and C. E. Woodcock. 2012. "Object-based Cloud and Cloud Shadow Detection in Landsat Imagery." *Remote Sensing of Environment* 118: 83–94. doi:10.1016/j.rse.2011.10.028.

Zhu, Z., S. Wang, and C. E. Woodcock. 2015. "Improvement and Expansion of the Fmask Algorithm: Cloud, Cloud Shadow, and Snow Detection for Landsats 4–7, 8, and Sentinel 2 Images." *Remote Sensing of Environment* 159: 269–277. doi:10.1016/j.rse.2014.12.014.

Appendix

(1) *Filtering training and test areas*

Using SQL queries, polygons with the densest covers were identified first: dense covers are defined as a cover fraction higher or equal to 80% for vegetation categories and 90% for crop categories. Broadleaf deciduous forest and broadleaf evergreen forests had relatively patchy spatial distributions, mainly in the southern scene (200034), so the cover fraction criterion was established at 45%. Also, since coniferous forest, broadleaf evergreen forest and broadleaf deciduous forests were not mixed, a more uniform spectral response was obtained for these areas. Percentages for bare soil and urban areas of these initial polygons were set at 80% and 75% respectively. Soil outcroppings with discontinuous and sparse vegetation were also included in the bare soil category.

Because most SIOSE polygons are composed of several land cover types, each with a different cover percentage, the SQL code excludes polygons with covers having a spectral response that could mask the spectral behavior of interest. For instance, the coniferous SQL query filters polygons where conifers coexist with crops, broadleaf forests or other related categories:

("CNFpl" > = 80 OR "CNF" > = 80) AND ("SIOSE_CODE" NOT LIKE '%FDP%' AND "SIOSE_CODE" NOT LIKE '%FDC%' AND "SIOSE_CODE" NOT LIKE '%LOL%' AND "SIOSE_CODE" NOT LIKE '%LFN%' AND "SIOSE_CODE" NOT LIKE '%CHL%' AND "SIOSE_CODE" NOT LIKE '%LFC%' AND "SIOSE_CODE" NOT LIKE '%LVI%' AND "SIOSE_CODE" NOT LIKE '%LOC%')*

Note: * pl suffix = plantation forests, FDP = broadleaf evergreen forest, FDC = broadleaf deciduous forest. Note that the NOT LIKE syntax avoids use of the percentage fields (for instance, by writing "SIOSE_CODE" NOT LIKE '%LOL%' we ensure that "LOLsc" = 0 and "LOLrr" = 0 are included in the selection).

Selected polygons were eroded with a 45 m buffer in order to prevent mixing or spectral 'contamination' at boundaries between land covers as well as location errors derived from vector cartography (Serra, Pons, and Saurí 2003). Polygons smaller than 1 ha were deleted due to low spatial and spectral representativeness (Moré et al. 2005; Pons et al. 2014a). The resulting polygons were fragmented into 4 ha (or less) squares to ensure that large polygons had areas reserved for both the classification and validation (test) processes. We randomly split the polygon set into two groups, a training set containing 70% of the polygons, and a validation set containing the remaining 30%. This process was performed separately for each category. The filtering process used NDVI thresholds as explained in Section 4.6.

(2) *Classifications by blocks*: *crop areas* (*CRA*) *and natural and urban areas* (*NUA*)

A two-category legend was defined for classifications by blocks. Categories 1–8 (see Table 2 in main text) were integrated into the natural and urban areas category (NUA) while categories 9–14 were integrated into the crop areas category (CRA). We performed a kNN classification for each scene and quinquennium using training areas obtained in Section 4.7 grouped into the two thematic categories. The resulting classifications (with an OA above 96%) were used to extract NUA and CRA masks. Map areas classified as CRA and NUA were used to mask spectral variables, indexes (NDVI, SAVI and $NDWI_3$) and topo-climatic variables. Once variables were masked into these two blocks, classification processes were performed separately for CRA categories and NUA categories in each quinquennium. The OA of this process was around 95% or higher as seen in Tables A1–A4.

Table A1. Southern scene (200034). NUA block.

	1985–1989				2000–2004			
	OE	PA	CE	UA	OE	PA	CE	UA
Water bodies (1)	0.0	100.0	0.0	100.0	0.0	100.0	0.0	100.0
Coniferous forests (2)	2.2	97.8	4.2	95.8	2.4	97.6	2.4	97.6
Broadleaf deciduous forests (3)	1.2	98.8	2.5	97.5	5.8	94.2	2.3	97.7
Broadleaf evergreen forests (4)	17.8	82.2	10.5	89.5	12.2	87.8	13.0	87.0
Shrublands (5)	13.0	87.0	11.5	88.5	13.7	86.3	14.3	85.7
Grasslands (6)	3.3	96.7	3.2	96.8	4.7	95.3	3.6	96.4
Bare soils (7)	13.9	86.1	10.0	90.0	9.1	90.9	3.0	97.0
Urbanized areas (8)	1.5	98.5	0.0	100.0	0.0	100.0	0.0	100.0
OA		96.1%				96.3%		
Kappa statistics		0.9				0.9		

Note: OE = omission error, CE = commission error, PA = producer's accuracy, UA = user's accuracy.

Table A2. Southern scene (200034). CRA block.

	1985–1989				2000–2004			
	OE	PA	CE	UA	OE	PA	CE	UA
Irrigated herbaceous crops (9)	6.8	93.2	11.6	88.4	5.6	94.4	7.7	92.3
Rainfed herbaceous crops (10)	10.9	89.1	19.2	80.8	5.4	94.6	8.4	91.6
Irrigated woody crops (11)	38.8	61.2	28.9	71.1	44.0	56.0	31.0	69.0
Rainfed woody crops (12)	3.5	96.5	2.5	97.5	1.8	98.2	2.0	98.0
Greenhouses (14)	0.1	99.9	0.1	99.9	0.1	99.9	0.0	100.0
OA		94.9%				96.8%		
Kappa statistics		0.8				0.9		

Note: OE = omission error, CE = commission error, PA = producer's accuracy, UA = user's accuracy.

Table A3. Northern scene (199031). NUA block.

	1985–1989				2000–2004			
	OE	PA	CE	UA	OE	PA	CE	UA
Water bodies (1)	0.2	99.8	0.4	99.6	0.1	99.9	0.4	99.6
Coniferous forests (2)	3.0	97.0	3.5	96.5	2.8	97.2	3.3	96.7
Broadleaf deciduous forests (3)	2.9	97.1	3.3	96.7	3.2	96.8	4.1	95.9
Broadleaf evergreen forests (4)	18.6	81.4	16.0	84.0	19.1	80.9	16.3	83.7
Shrublands (5)	13.7	86.3	17.1	82.9	15.8	84.2	20.3	79.7
Grasslands (6)	8.8	91.2	5.3	94.7	10.2	89.8	6.5	93.5
Bare soils (7)	17.0	83.0	23.1	76.9	15.7	84.3	22.0	78.0
Urbanized areas (8)	3.8	96.2	0.0	100	2.3	97.7	0.0	100.0
OA		95.6%				95.2%		
Kappa statistics		0.9				0.9		

Note: OE = omission error, CE = commission error, PA = producer's accuracy, UA = user's accuracy.

Table A4. Northern scene (199031). CRA block

	1985–1989				2000–2004			
	OE	PA	CE	UA	OE	PA	CE	UA
Irrigated herbaceous crops (9)	8.0	92.0	6.2	93.8	11.5	88.5	9.5	90.5
Rainfed herbaceous crops (10)	0.5	99.5	0.7	99.3	0.7	99.3	0.9	99.1
Irrigated woody crops (11)	7.8	92.2	8.9	91.1	11.2	88.8	14.2	85.8
Rainfed woody crops (12)	16.9	83.1	13.4	86.6	27.5	72.5	19.3	80.7
Rice crops (13)	14.2	85.8	8.4	91.6	18.0	82.0	9.4	90.6
OA		98.5%				97.7%		
Kappa statistics		0.9				0.9		

Note: OE = omission error, CE = commission error, PA = producer's accuracy, UA = user's accuracy.

6. ARTÍCULO 2: *Factors affecting forest dynamics in the Iberian Peninsula from 1987 to 2012. The role of topography and drough*

Vidal-Macua, J. J., Ninyerola, M., Zabala, A., Domingo-Marimon, C., & Pons, X. (2017). Factors affecting forest dynamics in the Iberian Peninsula from 1987 to 2012. The role of topography and drought. *Forest Ecology and Management*, 406, 290-306. https://doi.org/10.1016/j.foreco.2017.10.011

Forest Ecology and Management 406 (2017) 290–306

Contents lists available at ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco

Factors affecting forest dynamics in the Iberian Peninsula from 1987 to 2012. The role of topography and drought

Juan José Vidal-Macua[a,*], Miquel Ninyerola[b], Alaitz Zabala[a], Cristina Domingo-Marimon[c], Xavier Pons[a]

[a] Grumets research group, Dep Geografia, Edifici B, Universitat Autònoma de Barcelona, 08193 Bellaterra, Catalonia, Spain
[b] Grumets research group, Dep Biologia Animal, Vegetal i Ecologia, Edifici C, Universitat Autònoma de Barcelona, 08193 Bellaterra, Catalonia, Spain
[c] Grumets research group, CREAF, Campus UAB, Edifici C, Cerdanyola del Vallès 08193, Catalonia, Spain

ARTICLE INFO

Keywords:
Forest dynamics
Ecological succession
Forest shifts
Iberian Peninsula
Driving factors
Drought
Boosted regression trees
Conifer species
Broadleaf evergreen species
Broadleaf deciduous species

ABSTRACT

In southern Europe, climate trends are expected to be characterized by an increase in temperatures and less water availability. Analyzing the role of structural factors and the influence of a changing climate provides insights into the evolution of forest ecosystems in regions with similar environmental conditions. The Mediterranean fringe of the Iberian Peninsula is of particular interest due to its diverse topo climatic conditions and the increase in drought episodes during the last decades. This work studies forest dynamics in large areas of this geographical region by analyzing nine forest transitions. Vegetation covers were classified from three Landsat scenes for the period 1987–2012, and sub-periods 1987–2002 and 2002–2012. Conditions were described by topography derived variables, human factors and drought-occurrence variables. Boosted regression trees were used to identify the most important variables and describe the relationships between the forest dynamics and key factors. Variables such as solar radiation, topographic wetness index and tolerance to drought have been shown to be key factors in forest succession and when comparisons are made between vegetation groups. Main findings: The transition rate to Mediterranean and sub-Mediterranean broadleaf forests has increased during the analyzed period, while the transition rate to coniferous forests has decreased; Transitions to Mediterranean and sub-Mediterranean broadleaf forests are positively associated with drought occurrence while transitions to conifers are negatively affected by drought; Transitions from shrublands to forest stages are more vulnerable to factors controlling water availability; Important interactions between topography derived variables and drought have been found. The study provides robust evidence that drought occurrence plays an important role in the decline of conifers and the expansion of broadleaves, which could become the dominant species in many areas of the Mediterranean if climate model forecasts are met.

1. Introduction

Vegetation disturbances affecting all biome types have been observed in recent decades at global and continental scales (Hansen et al., 2013). Therefore, precise knowledge of the factors controlling vegetation dynamics is essential for improving ecosystem monitoring. For this purpose, analyzing vegetation transitions can provide significant insights, both from a methodological perspective and due to the findings.

The use of aerial photographs is an approach that is used widely in these studies, especially when a large coverage of historical aerial imagery is available. This approach has made it possible to obtain detailed information about vegetation types and changes in their spatial pattern over periods of 40–50 years (Allen et al., 1998; Sluiter and de Jong, 2007; Alados et al., 2004). However, human-aided mapping is a time consuming task when large areas are under study, which is essential for defining regional or global trends. This is especially important for analyzing climate-induced vegetation shifts because it is necessary to cover a spatially representative area (Martínez-Vilalta and Lloret, 2016). In this sense, the broad scale of national forest inventories has made it possible to identify transitions between species at a regional scale (Coll et al., 2013; Vayreda et al., 2016; Monleon and Lintz, 2015), although the density of plots and the period covered between surveys (~10 years) mean that definitive conclusions about climate-driven vegetation shifts cannot be made (Martínez-Vilalta and Lloret, 2016). Finally, the use of remote-sensing techniques is considered the most appropriate approach for identifying robust trends in

* Corresponding author.
E-mail addresses: Juanjo.Vidal@uab.cat (J. José Vidal-Macua), Miquel.Ninyerola@uab.cat (M. Ninyerola), Alaitz.Zabala@uab.cat (A. Zabala), cristina.domingo@uab.cat (C. Domingo-Marimon), Xavier.Pons@uab.cat (X. Pons).

http://dx.doi.org/10.1016/j.foreco.2017.10.011
Received 7 June 2017; Received in revised form 26 September 2017; Accepted 5 October 2017
Available online 18 October 2017

vegetation dynamics both spatially and temporarily (McDowell et al., 2015). Land–cover mapping based on satellite imagery can provide massive datasets (Wulder et al., 2008; Hermosilla et al., 2016; Vidal-Macua et al., 2017), although special attention should be paid to the possible uncertainty of these products (Pons et al., 2003, 2014; Pontius et al., 2004; Álvarez-Martínez et al., 2010) in order to obtain reliable land-cover change products and to avoid, as far as possible, the inclusion of noisy observations in the statistical analysis. The NASA–USGS Global Land Survey dataset (Gutman et al., 2013), primarily comprised of Landsat images (30 m resolution), is a very important resource for ecosystem monitoring from 1972 to present. Nevertheless, complementary methods are required to determine the proximate causes of large vegetation changes (McDowell et al., 2015).

In the Mediterranean mountains of Spain, ecological succession of vegetation has advanced after a general abandonment of traditional activities (García-Ruiz, 1990; Lasanta-Martínez et al., 2005; Jimenez-Olivencia et al., 2006; Cohen et al., 2011). Land-use management and, in general, human activities have been, and continue to be, the main driving force in the vegetation dynamics (Pino et al., 2000; Lasanta-Martínez et al., 2005; Gehrig-Fasel et al., 2007; Améztegui et al., 2010). However, other factors have to be taken into account to understand the vegetation colonization processes more fully as well as the shifts between species and between entire plant communities. The spatial pattern derived from the topography is a structural factor that explains much of the vegetation distribution. Interactions between topography–derived variables, such as altitude, slope, solar radiation or topographic effect on hydrological processes, are some of the factors most widely used in differentiating vegetation patterns (Pons and Solé-Sugrañes, 1994; Florinsky and Kuryakova, 1996; Burrough et al., 2001; Allen et al., 2004; Bennie et al., 2006; Serra-Díaz et al., 2011; Moeslund et al., 2013). The role of fire occurrence is another important issue for analyzing vegetation spatial-patterns and stability of certain plant communities (Diaz-Delgado and Pons, 2001; Salvador et al., 2005; Pausas and Keely, 2009). Climatic factors are also known to be main determinants in vegetation distribution, and climate dynamics are considered to drive vegetation shifts in many areas of the world (Walther et al., 2002; Schuur, 2003; Kelly and Goulden, 2008). In the Iberian Peninsula there has been a marked increase in temperatures and a decrease in precipitations over the last decades (López-Moreno et al., 2010; del Río et al., 2011, 2012; El Kenawy et al., 2012), which has led to an increase in the severity of droughts, especially in the Mediterranean area (IPCC, 2007, 2013; Gonzalez-Hidalgo et al., 2009; Vicente-Serrano, 2014; GECCC, 2016). The response of vegetation to droughts has become a matter of growing interest (Breshears et al., 2005; Allen et al., 2010; Vicente-Serrano et al., 2013; Martínez-Vilalta and Lloret, 2016; Norman et al., 2016) and recent works have shown the effects of drought on several Mediterranean and European species (Bigler et al., 2006; Weber et al., 2007; Pasho et al., 2011; Camarero et al., 2011; Carnicer et al., 2011; Vilà-Cabrerà et al., 2013; Galiano et al., 2013; Lévesque et al., 2013; Vicente-Serrano et al., 2015). Some of these species represent southernmost populations in the Mediterranean ambit, which explains their vulnerability to the warmer conditions (Andreu et al., 2007; Sánchez-Salguero et al., 2016). Declines of these species in forests could lead to long–term shifts (Peñuelas and Boada, 2003; Rigling et al., 2013; Galiano et al., 2010) and drought–tolerant species could become dominant in community compositions. The carbon stock of forests can be affected by these vegetation shifts (Vayreda et al., 2012), and the changes in flammability of vegetation can influence the fire regime, which is a very important issue in southern Europe (Pausas and Fernández-Muñoz, 2012; Moreira et al., 2012; Ganteaume and Jappiot, 2013).

Here, we analyze the influence of a series of physical, climatic and human factors on the forest dynamics in the Iberian Peninsula from 1987 to 2012. Two sub-periods have been included in the analysis, 1987–2002 and 2002–2012, in order to take into account the variability in human and climatic factors. The study focuses on progressive succession (Glenn-Lewin et al., 1992) and forest transitions, analyzing the response to factors in nine different forest cover changes, from shrublands to shifts in mature developing stages between conifer, broadleaf evergreen and broadleaf deciduous forests. For this purpose we have selected three large representative ambits, taking into account their particular climatic context and topographic characteristics. These ambits correspond to refined land-cover classifications of three entire Landsat scenes ($\sim$32,400 km^2). A set of variables derived from the SPEI (Standardized Precipitation-Evapotranspiration Index) was determined for spatially quantifying the occurrence of droughts. As far as we know, this is the first time that drought-occurrence variables have been used as explanatory factors of vegetation transitions.

Usually, the relationship between the explanatory variables and a given land cover transition is determined by linear regression analysis, such as logistic or multiple logistic regression (Serneels and Lambin, 2001; Serra et al., 2008), or by less commonly used approaches like Markov chains (Balzter, 1999). In our case, the importance of variables and how they influence each of the nine forest cover changes was analyzed using boosted regression trees (BRT), also known as stochastic gradient boosting (Friedman, 2001, 2002; Hastie et al., 2009). This relatively new machine-learning technique (Breiman, 2001), in which hundreds or thousands of decision trees (Breiman et al., 1984) are sequentially and progressively fitted, has been demonstrated to be particularly suited to predicting species distributions (Kawakita et al., 2005; Elith et al., 2006, 2008, 2009; De'ath, 2007; Leathwick et al., 2006; Crase et al., 2012) for the following main reasons: this approach does not assume any data distributions or data models, rather it tries to determine dominant patterns by combining many classification trees; it identifies relevant variables and complex interactions; it is much less influenced by correlated information or irrelevant variables than other statistical approaches; it produces stable predictions (variance reduction); and it provides graphical depictions of the relationship between the response variable and predictors. A separate BRT model was developed by forest cover change, period and study ambit, meaning that 81 models were evaluated.

The use of BRT is relatively new in ecology (Leathwick et al., 2006, 2008; Moisen et al., 2006; Sankaran et al., 2008; Levers et al., 2014; Verkerk et al., 2015), but we believe that they have not yet been applied to analyze vegetation transitions. Finally, we would like to highlight the particular relevance of the three study areas in the Iberian Peninsula: the three ambits are located on the Mediterranean fringe of the Peninsula, where a drying trend has been observed over the last decades, especially in the northeast (De Luis et al., 2010). Our intention was to provide insights into the conditions that lead to forest succession and transitions between vegetation groups, focusing on topographic variables, human–derived factors and drought occurrence. Thus, the main objectives of this study were to determine the following: (1) the most important factors in each of the forest cover transitions; (2) the role of the key factors for comparisons of the main forest species; (3) whether drought occurrence influences succession stages and forest transitions; and 4) whether drought occurrence has a clear influence, to determine the main interactions with other key factors.

2. Material and methods

2.1. Study areas

Three areas of Spain were included in this study as different scenarios for model development. Ambits correspond to land-cover classifications of three entire Landsat scenes, identified by their path-row: 198-031, 199-031 and 200-034 (Fig. 1). These large study areas were chosen because of their heterogeneity of biogeographical regions, including Alpine, Eurosiberian and Mediterranean areas, and, as previously explained, due to their climatic context. Representative mountainous regions of northeast and southeast Spain are included in these areas.

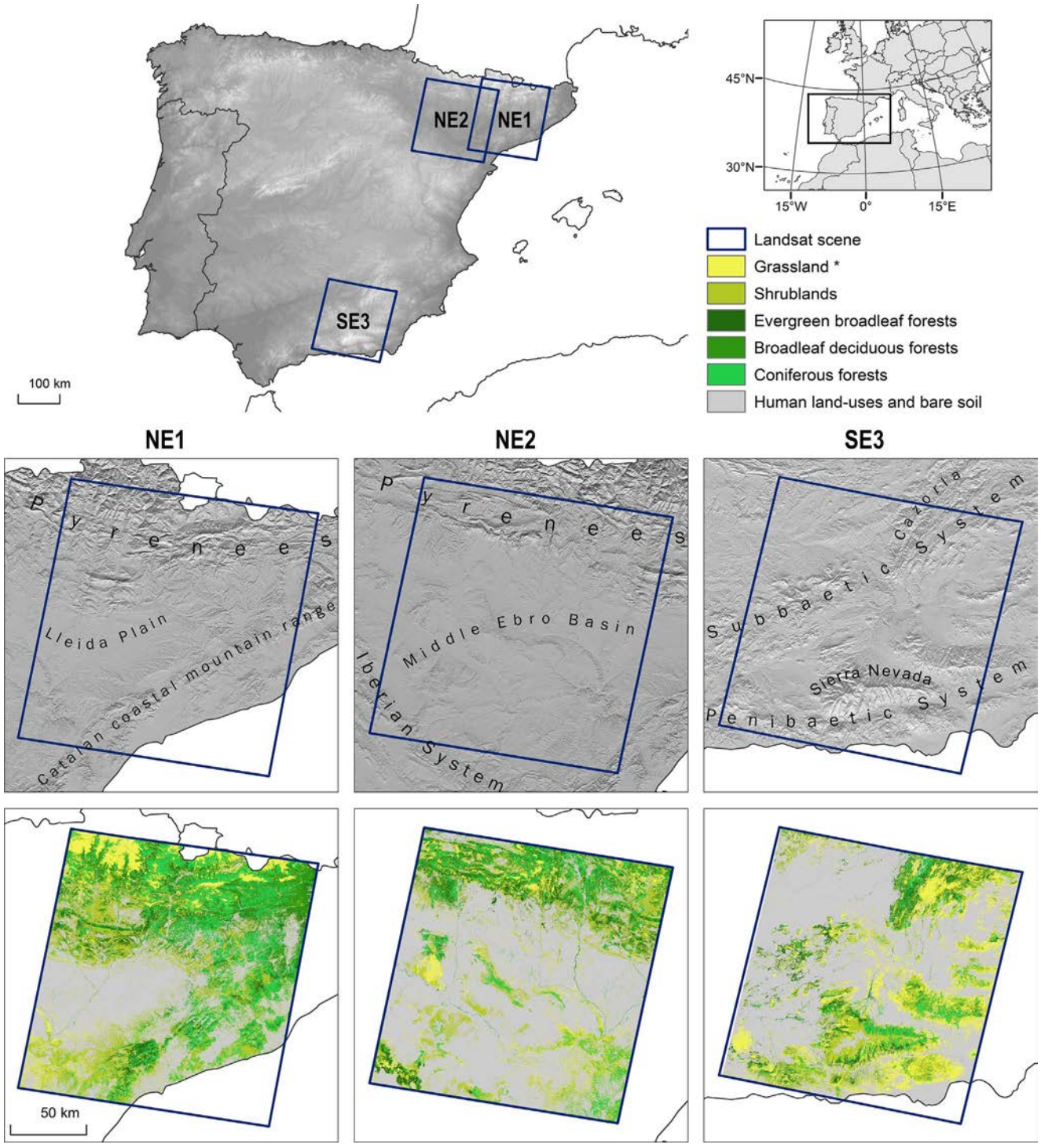


Fig. 1. Study areas. * Grassland category includes semi-arid and mountain ecotypes, and may include scattered shrubs or trees.

Scenes 198-031 and 199-031 are located in the northeast of the Iberian Peninsula and referred to hereafter as ambits NE1 and NE2 respectively. These two areas are spatially contiguous but have different environmental features, mainly because NE2 is more continental and arid. In addition, there is abundant reference information for these areas and some studies that can be related to ours.

The eastern ambit NE1 is influenced by a Mediterranean climate on the coast as well as the Catalan coastal mountain range, and by a continental climate in the Lleida Plain (Ebro basin) and Pyrenees (although the eastern Pyrenees also have a Mediterranean influence). Vegetation land-covers occupy a broad altitudinal range (from 0 to 3300 m), from sites near the coast to sites in the alpine and subalpine zones of the Pyrenees. Ambit NE2 has a greater continental influence in the Pyrenees and northern slopes of the Iberian System, with a semi-arid influence in the Middle Ebro Basin. In this ambit, which has less altitudinal range (from 50 to 2500 m), vegetation extends up to the upper montane zone and a very narrow area of the subalpine zone.

Scene 200-034 is located in southeastern Spain, and is hereafter referred to as ambit SE3. The coastal mountains in this ambit have a Mediterranean climate with pronounced aridity: less precipitations and higher temperatures than the northeastern ambits. Mountainous massifs of the eastern Baetic System, like the Sierra Nevada and Cazorla

mountain range, have a continental climate also with higher temperatures than the northern ambits. The Sierra Nevada has a complete altitudinal zonation, while there are environments from foothills to upper montane areas in most of these mountains.

2.2. Vegetation types

The study analyzes transitions between four vegetation classes derived from the land cover classification: shrublands, coniferous forests, broadleaf evergreen forests and broadleaf deciduous forests. The land cover class shrublands includes several formations ranging from temperate-xerophytic and temperate to alpine climate regions. Evergreen broadleaf sclerophyllous and evergreen needleleaf species are generally dominant in these areas (either as early transitional or permanent stages such as maquis). On the other hand, broadleaf deciduous shrubs are only dominant as permanent stages in cold-limited alpine environments. Coniferous forests are mainly composed by *Pinus halepensis*, *Pinus nigra* and *Pinus sylvestris*. In northern areas, *P. halepensis* is the dominant Mediterranean coniferous species below the upper montane zone, and in the southern area it is mixed with *P. nigra*, which, together with *P. sylvestris*, reaches the upper forest limit. Above the lower montane zone, Eurosiberian coniferous forests are composed mostly of *P. sylvestris*, with *P. nigra* at the lower limit and *Abies alba* and *P. uncinata* in the middle and upper areas (subalpine zone) respectively in the northern ambits. Broadleaf evergreen forests are dominated by *Quercus ilex* in the three ambits, although *Quercus coccifera* is widespread in NE2. *Q. ilex* extends from sites near the coast to the upper montane limit, and *Q. coccifera* spreads as continentality increases. Broadleaf deciduous forests are mainly composed by *Quercus pubescens*, *Quercus faginea* and *Fagus sylvatica*. *Q. pubescens* and *Q. faginea* are the dominant species below the upper montane zone in northern ambits, although in ambit NE2 they have a more patchy distribution. In the southern ambit, these species together with *Quercus pyrenaica* are restricted to sheltered locations. In the northern ambits, *F. sylvatica* is the most important Eurosiberian broadleaf deciduous species, although it is not present in subalpine zones.

2.3. Obtaining and refining land-cover maps

Vegetation cover maps were obtained by classifying Landsat imagery at 30-m resolution following the methodology described in Vidal-Macua et al. (2017). In our case, the k-nearest neighbor (kNN) classifier was used to obtain both multi-temporal training and test areas and to use them in final classifications. Each land-cover map refers to a five-year period, so classifications are composed of a set of dates in which the central years are 1987, 2002 and 2012. Training and test areas were extracted from the Land Occupation Information System of Spain 2005 (SIOSE, Sistema de Información sobre Ocupación del Suelo en España), a spatial database developed at a 1:25,000 scale (http://www.siose.es/). The disaggregation level of vegetation categories in the SIOSE database was used to determine the four vegetation classes defined in Section 2.2, except the grassland cover which is not included in the present work. Once a first set of SIOSE polygons were selected, we implemented a filtering process based on NDVI (Normalized Difference Vegetation Index) thresholds to reduce confusion between categories within polygons (Vidal-Macua et al., 2017). In a second stage, the kNN algorithm was applied to identify training and test pixels with an invariant statistical pattern for the 1987–2012 period (Vidal-Macua et al., 2017). Once these pixels had been identified for three dates, we used them as the final set to train the classifier and validate the results. The overall accuracy was greater than 90% for all three dates and ambits. The producer accuracy and user accuracy of vegetation land cover classes are shown in Table S1 of the Supplementary Material.

In order to avoid, as far as possible, including misclassified pixels in the statistical analysis, classifications were filtered using a confusion index (Burrough et al., 1997, 2000; Lewis et al., 2000; Gorsevski, 2005; Tapia et al., 2005; Álvarez-Martínez et al., 2010). This index was calculated with the *kNN* algorithm in which the nearest training pixels (*k*-nearest neighbors) to each target pixel (to be classified) are identified by means of the Euclidean distance. Class membership values (M_c) are assigned to each target pixel according to the number of times a category C appears within the nearest training pixels (N_c) weighted by the squared inverse distance, following the expression:

$$M_c = \frac{\sum_{i=1}^{N_c} w_{c_i}}{\sum_{i=1}^{N_k} w_{c_i}}$$

where C is a land-cover category within the nearest training pixels and W is the inverse of the squared distance between the target pixel (x) and a training pixel (x') labeled with the class C: $W = 1/d(x, x')^2$; and N_k is the number of k nearest training pixels. Thus, a target pixel has a membership value for each land-cover category within the k nearest training pixels. Vidal-Macua et al., (2017) describes the procedure for determining the optimum value of k.

Finally, the uncertainty associated with a target pixel was calculated with the confusion index (CI) equation:

$$CI = 1-(M_c\text{max}-M_c max_2)$$

where M_C max is the maximum membership value of land-cover categories within the k nearest training pixels, and M_C max_2 is the second maximum membership value.

Values of this index range from 0 to 1, so that values close to 1 indicate high confusion between at least two classes, and values close to 0 represent high certainty for a classified pixel. We used a threshold of 0.5 so that pixels classified with a CI over this value where masked to exclude them from the land-cover maps.

A second mask was created to remove misregistration in land-cover polygon boundaries. These situations are related to the location inaccuracy of border pixels after the geometric correction, and therefore, to false positive or negative changes between land-cover classifications (Pons et al., 2003). To solve this, classification polygon boundaries were eroded with a 20-m buffer mask according to the average RMS error of the imagery geometric correction. More details about this methodology can be found in Pons et al. (2003).

As our work is focused on natural ecological succession, we also used a mask to remove forest plantation areas, extracted from the SIOSE (Land Occupation Information System of Spain) database. Burned areas (REDIAM, 2016; Gobierno de Aragón, 2016; Generalitat de Catalunya, 2016) in each period were also masked. Like wildfires, insect outbreaks and windstorms are other important disturbances; however, in our analysis we focus on isolating, as much as possible, the influence of topography, drought and other variables described in later chapters on forest dynamics.

Finally, classifications were overlaid to obtain 9 forest-cover changes (Table 1) for the periods 1987–2012, 1987–2002 and

Table 1
Forest cover transitions.

Initial land-cover	Land-cover change	Code
Shrublands to	Coniferous forests	1. SRB-CNF
	Broadleaf evergreen forests	2. SRB-BEF
	Broadleaf deciduous forests	3. SRB-BDF
Coniferous forests to	Broadleaf evergreen forests	4. CNF-BEF
	Broadleaf deciduous forests	5. CNF-BDF
Broadleaf evergreen forests to	Coniferous forests	6. BEF-CNF
	Broadleaf deciduous forests	7. BEF-BDF
Broadleaf deciduous forests to	Coniferous forests	8. BDF-CNF
	Broadleaf evergreen forests	9. BDF-BEF

2002–2012. These transitions define the 9 models for each ambit and period; however, as explained below, some models were removed due to their low number of observations. In the present work, transitions from shrublands to forest stages are related to early development stages, so in later sections this term will be used to refer to these transitions (SRB-CNF, SRB-BEF and SRB-BDF in Table 1).

2.4. Sampling

We considered land-cover changes as absence/presence events. Areas where the initial land–cover remained stable between two dates were treated as absences, and those where there were changes in land-cover were treated as presences. For instance, absence areas (no-change events) in shrubland to coniferous forest cover are those where shrubland polygons in 1987 and 2002 match; and presence areas (change events) are those where shrubland polygons in 1987 and coniferous forest polygons in 2002 match, and so on for each land-cover change and period.

Once the change and no-change areas were established, stratified random sampling was carried out, attempting to maintain the prevalence (the same number of points) between the two types of events. As a general rule, 3000 points were randomly selected for each event class with a minimum distance of 250 m between each point. Prevalence was not achieved in all cases, especially in the southern ambit, where certain vegetation formations, like broadleaf deciduous forest, lack dynamism (compared with other formations) due to climatic conditions and the rear-edge location. In these cases we kept the resulting proportions between presences and absences since it has been shown that BRT models perform well with unbalanced samples (Edith et al., 2008; Sankaran et al., 2008; Edith and Graham, 2009; Crase et al., 2012). Nevertheless, extremely unbalanced models were rejected (see Section 3). Each dataset (composed by 6000 points when there is prevalence) was randomly split into two subsets, one made up of 75% of the samples for fitting the model, and a validation subset with the remaining 25% of samples.

2.5. Explanatory variables

In order to interpret the forest transitions, a set of variables was chosen (Table 2) to be included as explanatory factors in the models.

Table 2
Explanatory variables.

Variable code	Description	Units
Altitude	Altitude	Meters
Slope	Slope	Degrees
Curv	Overall terrain curvature	Dimensionless
Pf_Curv	Profile curvature	Dimensionless
Pl_Curv	Plan curvature	Dimensionless
VRM	Terrain roughness	Dimensionless
TWI	Topographic Wetness Index	Dimensionless
Win_SRad	Winter solar radiation	10 kJ m^{-2} day^{-1}
Sum_SRad	Summer solar radiation	10 kJ m^{-2} day^{-1}
Wind	Wind	Meters/second
Lithology	Acidity or basicity of the geologic substrate	Categorical
Dist_UrbA	Euclidean distance to urban areas	Meters
Dist_SecRo	Euclidean distance to secondary roads	Meters
Dist_MajRo	Euclidean distance to major roads	Meters
PopDen	Population density	Inhabitants/km^2
Pop_Dyn	Population gain or loss	Categorical
Liv_Units	Livestock units	Livestock units
S6_80_02_5 (example)	Every drought variable is coded as follows: "S6" – First two digits indicate the SPEI time-scale (6 or 24) "80_02" – Span of years used to calculate drought occurrence "5" – Last digit indicates drought episode duration	Number of drought episodes

The intention was to identify the most important variables and use statistical inference to describe the likelihood of a forest-cover change according to the value range of the variables. The MiraMon 8.2, ArcGis 10.1 and SAGA 2.1.2 software were used to calculate these variables.

2.5.1. Topography-derived variables

These variables are related to physical and environmental factors that could influence vegetation succession and competition within forest stages. The intention was to determine how the different vegetation types adapt to the environmental variability derived from the topography and to identify what vegetation types are more dynamic or vulnerable in the different topo-climatic contexts.

A 10 m DEM was created from 1:5000 map sheets of the Aerial Orthophotography National Plan 2010 (PNOA) and the following variables were derived from it:

- Altitude in meters above sea level.
- Slope in degrees.
- Terrain curvature (dimensionless), which includes overall curvature, profile curvature (in the slope direction) and plane curvature (perpendicular to the slope direction). The 0 value means no curvature in a typical range from −5 to 5, where negative values represent concave curves and positive values indicate convex curves.
- Terrain roughness (dimensionless), using the Vector Ruggedness Measure (Sappington et al., 2007), for which values close to 0 represent flat areas in an approximate range from 0 to 0.20 in our ambits.
- Topographic Wetness Index (dimensionless), which is a DEM-based soil moisture index (Beven and Kirkby, 1979; Sørensen et al., 2006; Kopecký and Čížková, 2010), and generally ranges from 3 to 30, where higher values indicate higher moisture availability.
- Winter and summer solar radiation (units in 10 kJ m^{-2} day^{-1}), which computes the total amount of incident solar radiation for each pixel at winter and summer solstice dates, following the methodology of Pons and Ninyerola (2008).

2.5.2. Wind

Wind data came from the Webservice-Energy platform (http://www.webservice-energy.org/) and CENER (National Renewable Energy Centre), and refer to mean wind speed in meters per second in a 4-km resolution raster. Wind can be a restrictive factor to plant growth and can influence seed dispersal.

2.5.3. Lithology

Lithology data may refer to the acidity or basicity of the geologic substrate. After obtaining geological information layers from several institutions (Spain, Catalonia, Aragon and Andalusia Governments) we reclassified the lithological groups into the following classes: acidic, basic and mixed. This is related to the tolerance of vegetation to low pH (silicates, more abundant in acidic rocks) or high pH (carbonates, more common in basic rocks).

2.5.4. Distance variables

Three variables were obtained as a measure of landscape accessibility: Euclidean distance to urban areas (cities, towns and villages), Euclidean distance to major roads, and Euclidean distance to secondary roads. We used the same road network and urban areas layers (obtained from the same governmental institutions) to create these variables for all periods, firstly because we did not find information about building dates or any other older layers, and secondly because we assumed that there has been few changes in the entirety of these infrastructures. We relate this variable both to the influence of isolation from infrastructures on forest management abandonment, and to the possible disturbance in natural dynamics due to higher accessibility.

2.5.5. Municipality-level information

Three variables were elaborated from INE (Spanish Statistical Office) data at a municipality level as factors that can disturb the natural succession:

- Population density (inhabitants per square kilometer). The central year of each analysis period was chosen as the reference year. Changes in this variable and in population dynamics could influence vegetation re-growth (Parcerisas et al., 2012).
- Population dynamics. The population gain or loss was calculated between each pair of dates of an analysis period. Four classes were defined as follows: ≥50% (and less than 100%) population increase, ≥100% population increase, ≥33% population decrease and no relevant changes.
- Livestock units, which account for cattle, sheep, goats and horses. Extensive and stabled livestock are placed in a single category in the INE data. Extensive livestock is more widespread in mountain areas where vegetation cover changes are taking place. This activity can affect vegetation succession, especially in shrubland covers, but also in evergreen forests because goats and sheep eat the acorns of *Quercus ilex*. In general terms, during 1987–2012, livestock units have increased by approximately 40% in the southern ambit and have not varied significantly in the northern ambits, although different trends can be found depending on the municipality.

2.5.6. Variables representing recurrent drought episodes based on the standardized precipitation evapotranspiration index

A set of variables indicating recurrent drought episodes was generated based on the Digital Topo–climatic Drought Atlas of the Spanish Iberian Peninsula (Domingo-Marimon, 2016). The Atlas includes a set of SPEI (Standardized Precipitation-Evapotranspiration Index) maps at 100-m spatial resolution for the entire Iberian Peninsula from 1950 to 2012. The SPEI (Vicente-Serrano et al., 2010a), based on precipitation and mean temperature (to estimate potential evapotranspiration), is an index that quantifies water deficits for multiple timescales. The values are standard deviations for which negative values indicate less than average precipitation, i.e. drought events, while positive values indicate greater than average precipitation, i.e. wet events. A threshold of SPEI $= < -1$ is selected to identify drought conditions, which end as soon as SPEI > -1 again.

The index was computed at several timescales corresponding to drought specific conditions. A first set of variables was generated using SPEI at a 6-month timescale from 1980 to 2012 as indicative of the medium-term moisture condition, which first identifies anomalies in the water streamflow. A second set of variables was generated using SPEI at a 24-month timescale from 1980 to 2012 as indicative of the long-term moisture condition, which identifies reservoir level and ground water level anomalies. For both sets, the number of drought episodes (SPEI $= < -1$) with durations of a minimum of 4, 5, 7 or 8 consecutive months, as representative lengths that may cause harmful effects, were counted for the period 1980–2012. The number of drought episodes was also counted by year spans of 15, 10 and 5 years for the 1987–2012 period. Therefore, the year spans used were: 1980–2012, 1997–2012, 2002–2012 and 2007–2012. In addition, the analysis was performed using two shorter sub-periods, 1987–2002 and 2002–2012, and their corresponding year spans: 1980–2002, 1987–2002, 1992–2002, 1997–2002, and 1980–2012, 1997–2012, 2002–2012 and 2007–2012 respectively. Including several year spans allows us to analyze how a greater or lesser drought frequency influences forest dynamics and whether recent droughts (during the last 5–10 years) have had an effect on transitions. The final dataset consisted of 32 drought variables in each model.

2.6. Variable subset selection

A collinearity analysis was carried out before the models were run to avoid the presence of correlated quantitative variables and to reduce processing times. The initial set of 49 variables was resized for each model using an alternative way to correlate coefficient estimates. We used Variance Inflation Factors (VIF) and a VIF threshold of 5 as other authors recommend (Zuur et al., 2009; O'brien, 2007; James et al., 2013). The analysis begins by making a regression of each variable on the other variables, and then calculating the VIF value: $VIF = 1/(1-R^2_i)$; where R^2_i is the R^2 of the regression of a variable onto all other variables. High values of R^2 (close to 1) mean that a variable is correlated with one or more variables, which in turn will lead to a high VIF value. After that, the variable with the highest VIF is removed. The analysis continues iteratively, recalculating VIF with the regressions of the remaining variables, until all variables have a VIF ≤ 5.

2.7. Data analysis

To identify the most important variables and to quantify their influence on forest transitions we used boosted regression trees (BRT), also known as stochastic gradient boosting (Friedman, 2001; Friedman, 2002; Hastie et al., 2009). BRT is a tree-based method combined with the strength of boosting (Breiman et al., 1984; Hastie et al., 2009; James et al., 2013). Classification trees segment the value range of an explanatory variable in order to determine regions that maximize the occurrence probability of a class inside them. Several variables can be combined to construct a tree depending on the interactions between them; for instance, from a tree of a single variable with a cutpoint and two regions, to a more complex tree where one of these regions can be split into two subregions based on a cutpoint in a second variable. The algorithm finds the best candidate variables in order to minimize the error rate while looking for a tree complexity that reduces the variance and the risk of overfit (James et al., 2013). The random forest algorithm (Breiman, 2001) is a substantial improvement over decision trees because it introduces bagging (Breiman, 1996) as a procedure for reducing variance and the error rate. By bagging, many bootstrapped samples (in the order of hundreds or thousands without replacement) are selected randomly from the training set to fit the same number of trees using a different random subset of variables in each of them (Hastie et al., 2009; James et al., 2013). The prediction for each observation is the average over all models.

Boosting (Freund and Schapire, 1996; Ridgeway, 1999; Friedman, 2002) is the optimization method in BRT. Like bagging, in the boosting algorithm hundreds or thousands of trees are built using a random fraction of the observations (without replacement) in each new tree. The main difference is that trees are not fitted individually but additively because each new tree is fitted to the residuals of the previous tree. The procedure to sequentially build trees is based on a series of rules that weigh observations depending on their error rate after fitting the previous tree (Friedman, 2002; Hastie et al., 2009). Each new tree is focused then on observations that have been poorly predicted.

2.7.1. Model fitting and evaluation

The procedure followed for fitting the BRT models is described in Section S1 of the Supplementary Material.

The initial results showed that altitude is the variable that contributed most notably in most models, so we fitted an additional model that did not include this predictor. The altitude variable can be interpreted as a climatic variable because it introduces a temperature and precipitation gradient. We assumed that variables interacting with altitude are more likely to be selected as important contributors, so removing it would allow other interacting schemes to fit the models.

2.7.2. Model inference

To interpret the results in each forest transition model, we determined the importance of the predictors in the BRT models by considering their contributions in the additive model. The relative importance was measured based on the number of times a predictor was

selected for splitting, weighted by the squared improvement to the model as a result of each split, and averaged by the number of trees (Friedman and Meulman, 2003; Hastie et al., 2009). Then, results were scaled between 1 and 100.

The response of a forest transition in the predictor values was interpreted using partial dependence plots. This graphic output shows the relationship between the response and an explanatory variable after the average effects of all other variables are accounted for, using the weighted tree traversal method described in Friedman (2001) (Ridgeway, 2004). Thus, these plots depict where the suitable conditions are for a transition to take place. Some interactions between variables are shown using three-dimensional partial dependence plots (Elith et al., 2008), which are based on predictions for each variable pair.

3. Results

3.1. Model validation

Information about the model validation is provided in Section S2 of the Supplementary Material.

3.2. Estimate of the transition rate between vegetation types (summary)

Although most of the area corresponded to zones that remained stable, there are some trends which are worth highlighting. In the northern ambits (Tables S5 and S6), while the percentage of change towards coniferous forest has decreased from one period to the next in all transitions, transition rates towards broadleaf evergreen forest have increased (except in the BDF-BEF transition), especially in ambit NE2. In these ambits, the transition rates towards broadleaf deciduous forest have also increased, although to a lesser degree. On the other hand, in the southern ambit (Table S7) all transition rates have decreased from one period to the next, especially in CNF-BEF shifts.

3.3. Explanatory factors in forest cover changes

The collinearity analysis (VIF) allowed reducing models size. The variables removed were mainly drought indices (from an initial set of 32 to 12–14 drought variables). In most cases, summer solar radiation was removed in the VIF analysis. The relative importance of explanatory variables for each forest cover change, period and ambit is shown in Tables S8–S10 of the supplementary material. In order to synthesize the results, our analysis mainly focuses on the six most important variables in each forest cover change and ambit (Table 3). To do this, we calculated the average importance of each variable, considering the six most important variables in each period, including and excluding altitude. However, other variables were taken into account to improve species comparisons and the reliability of explanations. Partial dependence plots can be consulted in Figs. S19–S30. In general, topography-derived variables and drought indexes showed the clearest patterns; therefore, we emphasize the role of these variables when types of transitions are compared.

3.3.1. Transitions from shrubland to forest stages (SRB-CNF, SRB-BEF and SRB-BDF)

We have selected four topography-derived variables to compare changes from shrubland covers in the 1987–2012 period (Fig. 2): altitude, slope, solar radiation and soil moisture (TWI). A similar pattern for altitude can be detected in NE1 and NE2; however, given that the altitudinal range in NE2 has a lower upper limit, comparing trends from 1600–1700 m upwards could be misinterpreted. This pattern represents the altitudinal zonation of vegetation. Thermo and meso–Mediterranean conifers are more competitive on foothills (below 500–600 m) and Eurosiberian conifers above the lower montane zone (1400–1500 m). Broadleaf forests are more dynamic from the super-Mediterranean zone (600–700 m) to the upper montane zone (1700–1800 m). Ambit SE3 shows similar patterns but with an upward shift due to a more xeric climatic context. For this reason, succession to upper-montane conifers (*Pinus sylvestris*) are less likely events than in northern ambits, and transitions to broadleaf deciduous are very rare ones.

In all cases, locations with lower solar radiation are more suitable, although SRB-CNF dynamics seem to be more tolerant to higher solar radiation. All transitions are dependent on soil moisture (TWI), especially in transitions to broadleaf forests in northern ambits.

There are significant findings with respect to drought indices. In ambit NE1, long-term drought occurrence negatively affects SRB-CNF changes, which are more likely at sites where the SPEI at a 24-month timescale and at least 8-month duration occurs at most once (Figs. S19 and S21). On the other hand, medium-term droughts (SPEI at a 6–month timescale) of 4 and 5-month duration have a positive influence on transitions to broadleaf evergreen forest (SRB–BEF) in northern ambits (Figs. S21 and S25). We identified important interactions regarding these indices, and some examples are shown in Fig. 3. In ambit NE1, SRB-BEF transitions are occurring at sites more affected by drought events, low wind velocity and low solar radiation values (Fig. 3A). The same interaction regarding solar radiation can be seen at ambit NE2 together with high TWI values (Fig. 3B).

Drought occurrence also has a positive influence on SRB–BDF changes in ambit NE1. In this case medium-term drought events occurred during the last five years (2007–2012) (Fig. S21). Looking at the most important drought variables, SRB–BDF transitions are more likely at sites affected by, at least, 1–2 occurrences of medium-term droughts of 8–month duration and at sites affected by 3–4 occurrences of medium-term droughts of 4–month duration. We used reference information (Catalonia Land–Cover Map 1:25,000, 2009) to calculate the average distance between SRB-BDF presence events and vegetation cover polygons according to the dominant species (Table S11 of supplementary material). We found that the closest species were *Pinus sylvestris* and *Quercus pubescens*, thus, the latter is a more drought–tolerant species. The most important interactions with drought occurrence are again variables that can influence the water evaporation (Fig. 4), although medium–term droughts of 4–month duration show less pronounced interactions.

3.3.2. Transitions between conifer and broadleaf evergreen species (BEF-CNF and CNF-BEF)

We have selected the following variables to analyze forest transitions: altitude, slope, solar radiation, soil moisture (TWI) and wind. The response to variables in shifts between conifers and broadleaf evergreen species are shown in Fig. 5. In general, broadleaves are better adapted to steep slopes, which is also reflected in the other forest transitions. As in the previous section, patterns concerning altitude can be easily interpreted by relating them to the altitudinal zonation of vegetation groups. In ambit NE2, higher probabilities of BEF–CNF shifts are concentrated above the lower montane zone, where conifer species are mainly composed of *Pinus sylvestris*. In northern ambits, BEF-CNF shifts are more likely to occur as solar radiation decreases, especially in ambit NE2. In NE1, the “u” shape in the dependence on soil moisture plot (TWI) could reflect the presence of the two types of conifers in BEF–CNF shifts: *P. halepensis*, which is more adapted to lower values, and *Pinus sylvestris*, which is more dependent on water availability. In NE2, *P. sylvestris* is more dynamic than evergreen *Quercus* spp. as solar radiation decreases and TWI increases. Unlike in ambit NE1, shifts to evergreen *Quercus* spp. (CNF–BEF) in ambit NE2 are more likely as solar radiation increases and seem to be more tolerant to soil water scarcity. On the other hand, in ambit SE3, the density of presence events and probability regarding the altitude indicate that Mediterranean conifers are predominant in BEF–CNF transitions. This ambit does not show remarkable differences regarding solar radiation. Instead, response to soil moisture shows that CNF-BEF transitions are more likely than

Table 3
The six most important variables by forest cover change and ambit. M.I.V = Most important variables; Freq. = frequency (number of times the variable is within the 6 most important variables); A.R.I = Average relative importance. The following models were not used in the analysis due the scarcity of presence events: BDC-CNF in the 2002–2012 period in NE2, and CNF-BDF and BDF-CNF in three periods in SE3.

	Ambit NE1				Ambit NE2				Ambit SE3			
	M.I.V.	Freq.	M.R.I.		M.I.V.	Freq.	M.R.I.		M.I.V.	Freq.	M.R.I.	
SRB-CNF	S24_07_12_8	4	19.56	± 1.27	Wind	6	12.97	± 1.78	Altitude	3	20.53	± 1.74
	Altitude	3	17.39	± 1.20	Altitude	3	9.08	± 1.41	Slope	6	9.04	± 2.65
	Liv_Units	2	7.61	± 0.34	Liv_Units	6	8.98	± 0.90	Liv_Units	6	7.88	± 1.88
	PopDen	6	7.51	± 0.90	PopDen	6	7.44	± 1.07	Wind	6	7.50	± 1.43
	Slope	4	6.56	± 1.64	Dist_MajRo	6	7.26	± 0.51	Dist_MajRo	6	7.46	± 0.99
	Wind	6	6.28	± 0.93	Win_SRad	4	7.09	± 0.48	PopDen	3	7.37	± 1.26
SRB-BEF	Altitude	3	29.01	± 11.37	Altitude	3	41.18	± 12.51	Altitude	3	26.71	± 3.02
	Win_SRad	6	11.54	± 1.87	S6_80_02_5	1	17.11	± 0	Win_SRad	4	8.97	± 1.01
	S6_80_02_4	1	8.97	± 0	Win_SRad	4	9.65	± 3.30	Lithology	1	8.74	± 0
	Wind	5	6.95	± 2.53	S6_80_12_5	1	8.59	± 0	Dist_MajRo	6	7.96	± 1.31
	Liv_Units	2	6.67	± 0.82	PopDen	2	8.50	± 1.22	Dist_UrbA	2	7.72	± 2.21
	VRM	6	6.20	± 1.49	VRM	3	7.33	± 0.11	PopDen	6	7.17	± 0.80
SRB-BDF	Altitude	3	33.86	± 6.29	Altitude	3	20.49	± 1.52	Altitude	3	22.53	± 2.15
	S6_07_12_8	2	14.91	± 3.83	Wind	6	11.34	± 3.52	TWI	6	13.09	± 4.56
	S6_07_12_4	2	10.46	± 6.57	Win_SRad	6	10.78	± 3.09	S6_07_12_5	1	12.43	± 0
	Liv_Units	5	8.64	± 4.20	TWI	5	10.64	± 1.77	S6_80_02_5	1	11.77	± 0
	Win_SRad	6	8.47	± 1.77	PopDen	6	9.67	± 2.51	Dist_MajRo	4	7.62	± 1.42
	S6_97_02_8	1	8.23	± 0	Dist_MajRo	3	6.66	± 1.78	S6_07_12_4	2	7.54	± 1.32
CNF-BEF	Altitude	3	32.02	± 1.91	Altitude	3	24.42	± 11.95	Altitude	3	15.47	± 1.14
	Win_SRad	6	9.41	± 2.13	Wind	6	10.85	± 1.48	Dist_MajRo	6	9.56	± 0.47
	Wind	5	7.30	± 1.30	Dist_MajRo	6	7.11	± 1.36	Wind	6	8.81	± 2.31
	Slope	6	7.29	± 0.78	PopDen	4	6.95	± 1.44	Dist_UrbA	6	8.33	± 1.40
	S6_80_12_7	1	6.86	± 0	Win_SRad	5	6.51	± 0.96	Liv_Units	3	7.28	± 1.85
	Liv_Units	5	6.43	± 1.79	Dist_UrbA	5	6.32	± 0.53	PopDen	3	7.04	± 0.39
CNF-BDF	Altitude	3	17.79	± 3.70	Altitude	3	18.04	± 10.48				
	Liv_Units	6	9.66	± 3.11	Slope	6	14.07	± 3.54				
	S6_07_12_8	4	8.27	± 0.86	Win_SRad	6	11.58	± 3.70				
	Slope	6	8.04	± 1.24	Lithology	5	11.26	± 4.84				
	PopDen	2	7.39	± 0.50	PopDen	6	8.04	± 1.90				
	Wind	2	7.31	± 1.12	Wind	3	6.84	± 3.06				
BEF-CNF	Slope	6	14.55	± 0.37	S24_97_02_8	2	14.97	± 0.60	Slope	6	11.14	± 1.81
	Win_SRad	6	9.30	± 0.70	Wind	6	11.23	± 0.87	Altitude	3	10.85	± 3.20
	Altitude	3	8.82	± 0.80	S6_92_02_4	2	11.19	± 1.32	Liv_Units	6	9.26	± 0.98
	Wind	6	7.49	± 1.03	Altitude	3	10.14	± 4.71	Dist_MajRo	6	8.83	± 0.90
	Liv_Units	3	6.87	± 0.96	Slope	6	9.36	± 2.56	Wind	1	7.31	± 0
	S24_07_12_7	1	6.62	± 0	Dist_MajRo	4	8.30	± 0.70	Dist_SecRo	2	6.93	± 1.04
BEF-BDF	Win_SRad	6	14.34	± 4.91	Win_SRad	6	18.45	± 6.74	Lithology	6	23.97	± 7.25
	PopDen	3	9.28	± 3.97	Altitude	3	11.38	± 2.59	Dist_UrbA	6	11.72	± 6.35
	S6_07_12_8	1	8.78	± 0	Wind	6	9.37	± 4.09	TWI	6	9.42	± 2.40
	Liv_Units	6	8.61	± 0.77	Slope	6	9.00	± 4.25	Altitude	3	9.24	± 3.57
	Slope	4	7.43	± 0.83	Liv_Units	2	5.66	± 0.05	Slope	4	7.03	± 0.63
	Wind	5	6.99	± 1.05	Dist_MajRo	5	5.50	± 0.99	Wind	3	6.71	± 1.11
BDF-CNF	Altitude	3	20.70	± 14.88	Dist_MajRo	4	14.09	± 9.18				
	Slope	6	10.93	± 1.12	Altitude	2	11.95	± 2.96				
	Dist_SecRo	1	10.09	± 0	Slope	4	11.02	± 1.82				
	Liv_Units	4	9.73	± 2.43	Liv_Units	2	9.29	± 0.70				
	VRM	4	9.61	± 2.23	Dist_UrbA	4	7.98	± 0.85				
	Wind	3	8.51	± 2.05	Wind	4	7.64	± 2.05				
BDF-BEF	Altitude	3	24.05	± 2.16	Win_SRad	4	16.25	± 0.93	Altitude	3	32.41	± 4.52
	Liv_Units	6	10.70	± 3.53	Altitude	3	10.69	± 1.42	Sum_SRad	3	9.34	± 3.63
	Win_SRad	5	9.44	± 1.36	Slope	6	9.22	± 1.08	Liv_Units	1	9.06	± 0
	Slope	5	9.17	± 2.85	Wind	6	8.68	± 0.96	Dist_UrbA	4	8.85	± 2.79
	Wind	6	7.00	± 1.21	Dist_UrbA	1	8.16	± 0	Dist_MajRo	6	8.60	± 3.37
	PopDen	3	6.69	± 0.48	PopDen	1	7.40	± 0	Wind	4	8.23	± 0.89

BEF–CNF as TWI values increase. Wind velocity does not show clear patterns; however, the lower wind velocities seem to be more suitable for *Pinus sylvestris* and evergreen *Quercus* spp. in NE2 and SE3 respectively.

Shifts between conifers and broadleaf evergreen species show significant associations with drought variables. BEF-CNF transitions in ambit NE1 are more likely at sites where long-term droughts of at least 7-month duration occurs at most once (Fig. S20). In ambit NE2, these shifts are more likely when long-term droughts of at least 8-month duration do not occur (Figs. S24 and S26). In addition, occurrence of medium-term droughts of 4–month duration in NE2 (Figs. S24 and S26) seems to be less suitable for BEF-CNF transitions. Changes from conifers to evergreen *Quercus* spp. (CNF-BEF) in ambit NE1 are associated with a decrease in the number of occurrences of medium-term droughts of 7–month duration (Fig. S21). It is worth noting that, unlike in NE1, this shift in NE2 is more likely as occurrences of medium-term droughts of

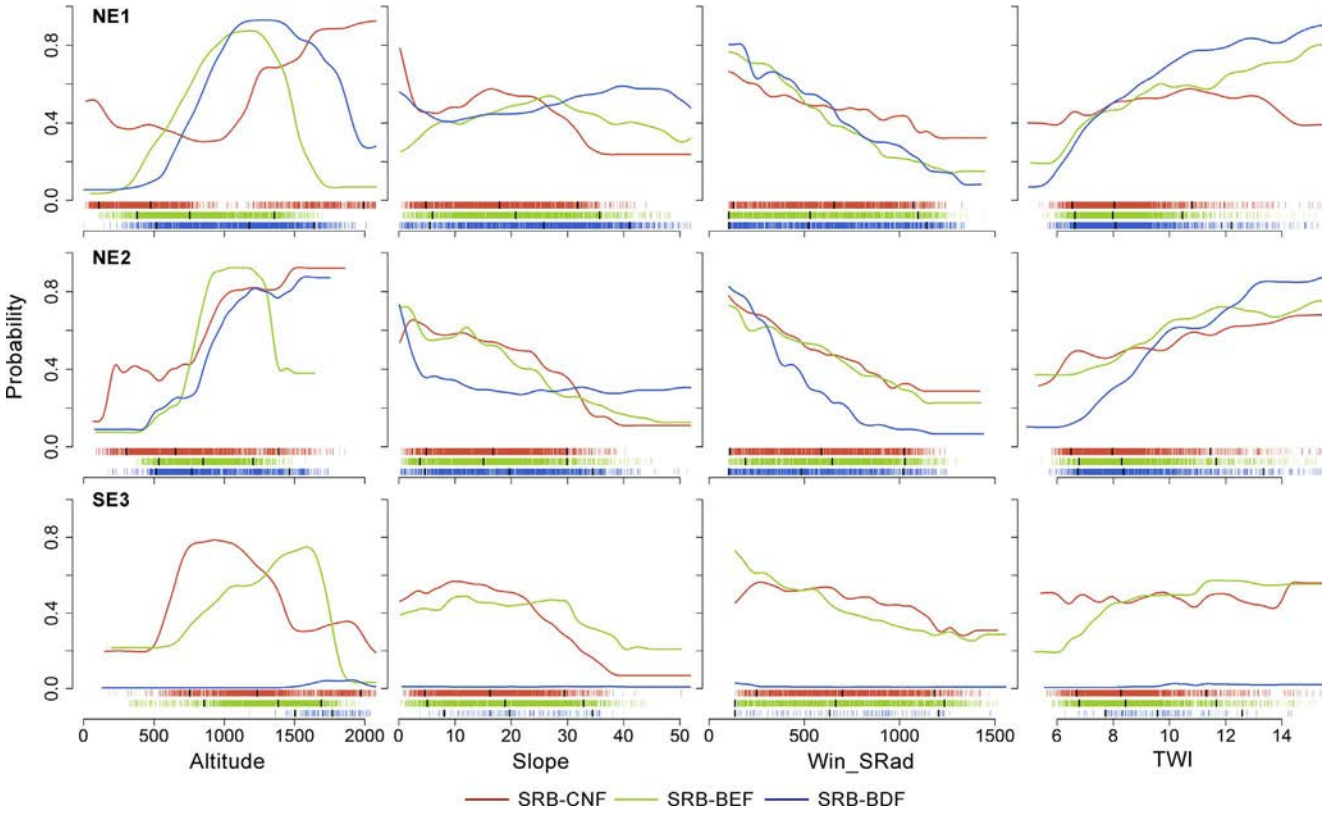


Fig. 2. Response of SRB-CNF, SRB-BEF and SRB-BDF transitions to altitude, slope, solar radiation and soil moisture (TWI) for the three ambits in the 1987–2012 period. Density of presence events is represented by vertical lines above the x-axes, and overlaid solid black ticks depict the 0.05, 0.5 and 0.95 percentiles.

5–month duration increase (Fig. S25). We have plotted the response of CNF-BEF to different drought durations, which can be compared in the two northern ambits (Fig. 6). A similar pattern between ambits can be observed in terms of density of observations (vertical bars). In terms of suitability, patterns match well if we consider that a very low number of observations in B3 with high probabilities at 5–6 repetitions can be related to small statistical artifacts (derived from the method used to interpolate climatic data or classify land–cover) or exceptional topo–-climatic conditions. This suggests that, the greater importance of slope in ambit NE1 (Table S8), compared with NE2 (Table S9), could explain the greater importance of the drought of 7–month duration in NE1 and its negative effects.

3.3.3. Transitions between conifer and broadleaf deciduous species (BDF-CNF and CNF-BDF)

For transitions between conifers and broadleaf deciduous species we can only report the results for the northern ambits (Fig. 8) because these models were rejected in ambit SE3 (Table S4). We kept the dependence curve of BDF–CNF shifts in ambit NE2 although it is a less reliable model for establishing comparisons because it is a more unbalanced sample (Table S3). CNF-BDF shift patterns differ markedly in response to altitude. We found out that higher probabilities below 500 m in ambit NE2 are related to the presence of an important anastomosed river (Cinca River) with extensive fluvial deposits where vegetation dynamics depend on stability and the age of deposits (Ojeda, 2007). Looking at the response to altitude in this ambit, *Fagus sylvatica* seems to be the species with most dynamism in CNF-BDF shifts (above 1500 m). Water evaporation is observed as an important factor in the dynamics between these species. Broadleaf deciduous species tend to progress as solar radiation decreases and soil moisture increases. In addition, lower wind velocities are more suitable for these species in NE1. In mixed landscapes of conifers and broadleaf deciduous species, the conifers are less vulnerable to solar radiation and more dynamic with lower soil moisture values.

In contrast to vulnerability to soil moisture availability, medium-term droughts of at least 7–8–month duration have a positive association with CNF-BDF shifts in NE1 (Figs. S19 and S21). *Quercus pubescens* is the predominant broadleaf deciduous species in these transitions and *Pinus sylvestris* the coniferous one (Table S11). This is consistent with the partial dependence on altitude (Fig. 8), where there are higher probabilities in a range from 1000 to 1200 m. Thus, the results indicate that in mixed forests of these two species, *Quercus pubescens* is less sensitive to drought and this is reflected as a shift.

3.3.4. Transitions between broadleaf evergreen and broadleaf deciduous species (BDF-BEF and BEF-BDF)

Dynamics between broadleaf evergreen species and broadleaf deciduous species are shown in Fig. 9. Both transitions in SE3 and BDF–BEF transitions in NE2 correspond to more unbalanced samples, which it is reflected in lower probabilities. In terms of the response pattern to variables, the common trend regarding the altitude is that BEF–BDF shifts are more likely above the lower montane zone and BDF-BEF shifts are more likely in foothills and Mediterranean zones. Tolerance to water evaporation is again a differential factor in these transitions. Broadleaf evergreen species are more dynamic than broadleaf deciduous species as solar radiation increases and TWI values decrease. Instead, broadleaf deciduous species are more dynamic on north faces and with soil moisture availability. Patterns in dependence to wind velocity are less clear, although BEF-BDF transitions seem to be more likely with low wind velocities.

Quercus pubescens is the most abundant broadleaf deciduous species regarding these transitions in NE1 (Table S11). This species again seems to show a positive association with drought occurrence in mixed landscapes with *Quercus ilex* under medium-term drought conditions (drought variable that contributed most: SPEI at a 6–month timescale and 8-month duration) (Table 3 and Figs. S20 and S22). It is worth

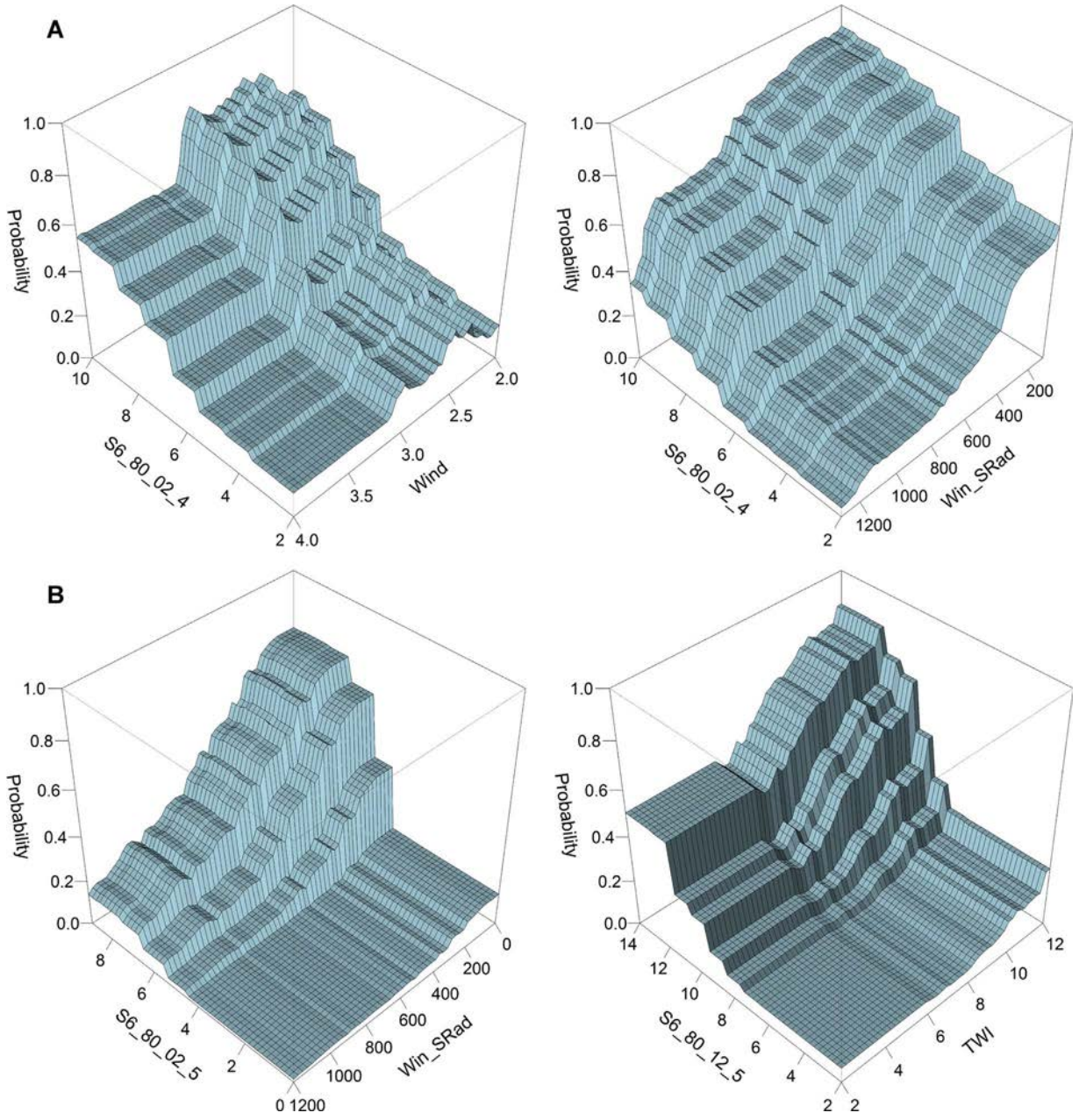


Fig. 3. Variables that interact with drought occurrence in SRB-BEF transitions: wind and solar radiation in NE1, 1987–2002 period (A); solar radiation and soil moisture (TWI) in NE2, 1987–2002 and 1987–2012 periods (B).

keeping in mind that solar radiation is the most important variable; therefore, looking at the response to this variable it is reasonable to presume that under medium-term droughts conditions *Quercus pubescens* is more dynamic in mixed forests on north faces.

4. Discussion

In general terms, transition rates derived from classifying the three Landsat scenes indicate that broadleaf forests have increased their area in the northern ambits, while coniferous forests have followed a declining trend from one period to the next, which is consistent with former studies (Vayreda et al., 2016, Carnicer et al., 2014). In contrast, the magnitude of the transitions in the southern region suggests that broadleaf evergreen species, like *Quercus ilex*, are becoming less dynamic, in agreement with Vayreda et al. (2016), who found a contraction of the latitudinal range of this species at its southernmost limit. In addition, and in accordance with these authors, the expansion of evergreen broadleaves appears to be greater in the more xeric environments of the northern regions (NE2).

All transitions from shrubland are vulnerable to high solar radiation, indicating that water evaporation is a controlling factor in early stages of forest dynamics. Hence, soil water availability also increases recruitment, although coniferous species are more tolerant to low soil moisture values, and to higher solar radiation, which can be explained by the heliophilous character of pines (Blanco et al., 1997; Broncano et al., 1998, Ameztegui and Coll, 2011). It is reasonable to think that early developing stages are more dependent on topo–climatic conditions like soil moisture, which is driven by solar radiation and the TWI, among other factors. These results are in line with other studies reporting the susceptibility of seedling growth of the evergreen *Quercus* spp. (Espelta et al., 1995; Benayas, 1998) and *Pinus sylvestris* (Castro et al., 2004) to large evapotranspiration rates in open spaces and canopy gaps.

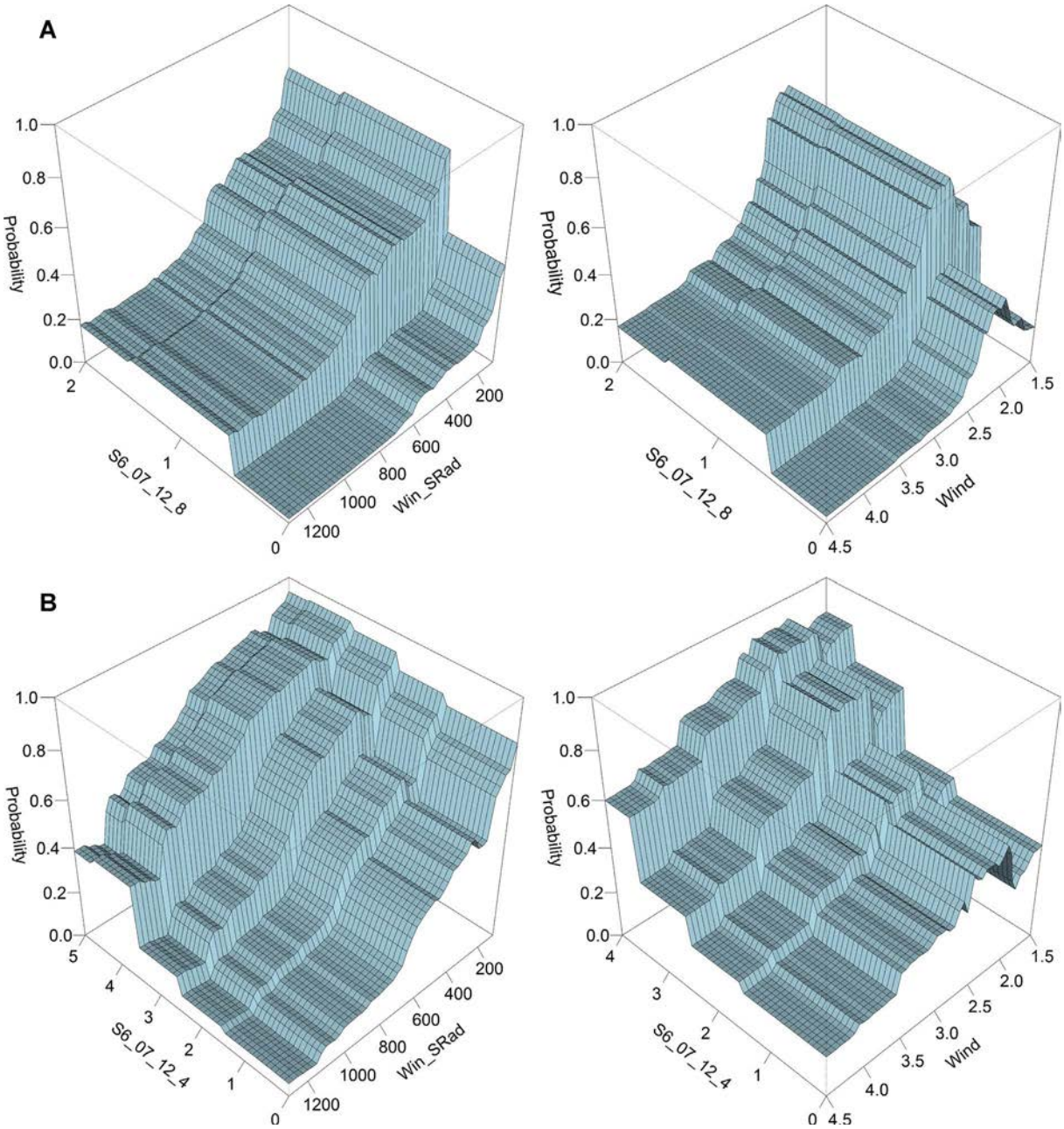


Fig. 4. Variables interacting with drought occurrence in SRB-BDF shifts, ambit NE1: solar radiation and wind in the 2002 2012 period (A); solar radiation and wind in the 1987–2012 period (B).

Concerning competition within forest stages, broadleaf deciduous species are the vegetation group that is most dependent on topographic factors controlling moisture retention (i.e. transitions to these species are more likely as solar radiation decreases and TWI values increase). Non–Mediterranean conifers, such as *Pinus sylvestris*, benefit from soil moisture retention and low solar radiation in competition with broadleaf evergreen species, which is clearly manifested in a drier environment such as the ambit NE2. On the other hand, Mediterranean conifers and broadleaf evergreen species do not show a clear preference regarding solar radiation. However, our results suggest that broadleaf evergreen species have a greater plasticity, as they are more competitive in a more xeric ambit with lower water availability (Fig. 5, ambit NE2) and under drought conditions (Fig. S29), which is in parallel with other works (Vicente-Serrano et al., 2010b, Pasho et al., 2011). Nonetheless, there is another factor that could influence these dynamics: *Quercus coccifera*, which is more extended in NE2 than in NE1, has less water requirements than *Quercus ilex* (Blanco et al., 1997). Besides, our results corroborate that under warmer climate conditions (Fig. 5, ambit SE3), broadleaf evergreen species tend to be much more competitive with soil moisture availability (Blanco et al., 1997), also because its roots are able to penetrate into the deep water table (Lloret et al., 2004). In general, the more developed root system of broadleaves also explains their better adaptation to steep slopes.

Response to drought events has been shown to be an important driving factor in vegetation succession and forest shifts. The results suggest that under drought conditions, in mixed landscapes composed of Mediterranean and sub-Mediterranan broadleaves and conifers, shifts to broadleaves are more likely to occur. These findings are consistent with previous works showing that evergreen *Quercus* spp. are more adapted to drought periods than pines (Martínez-Ferri et al., 2000;

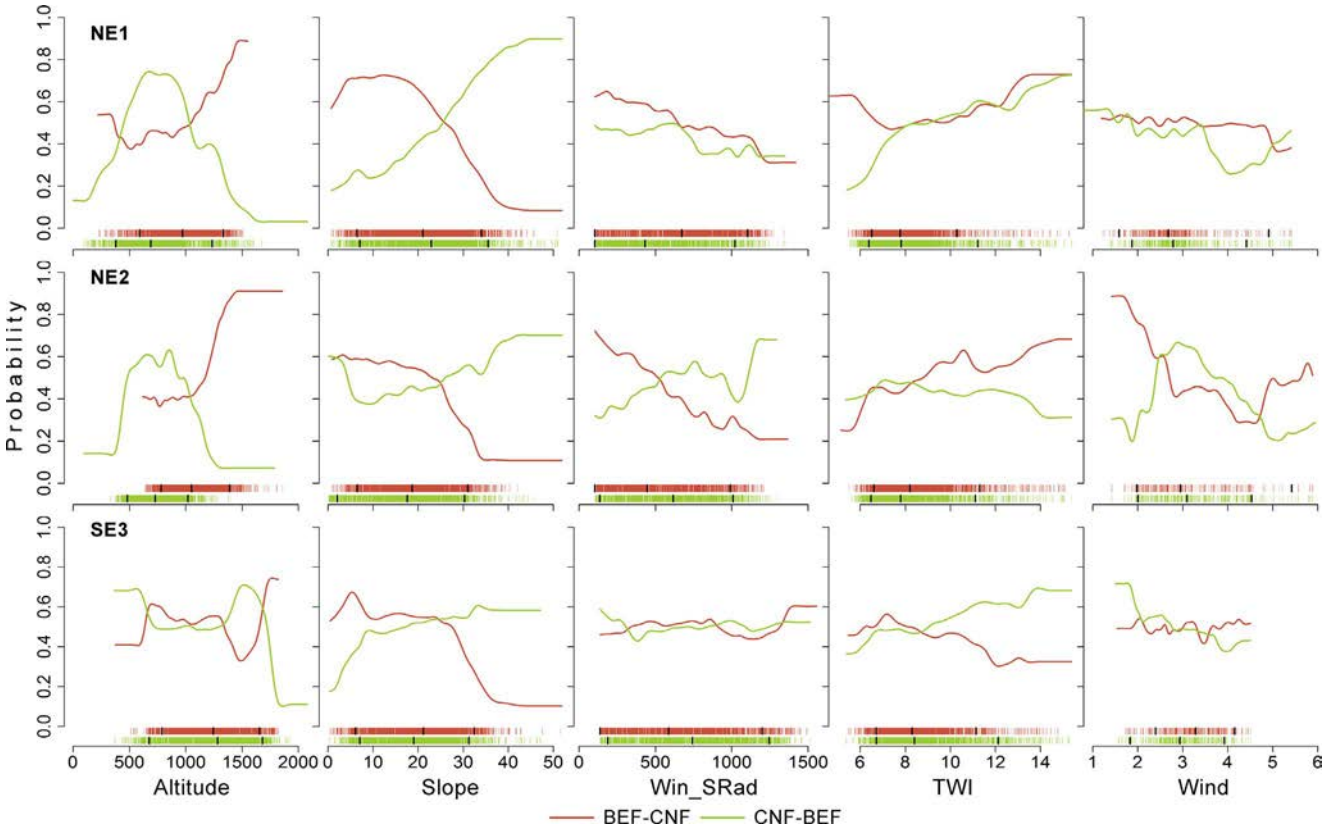


Fig. 5. Response of BEF-CNF and CNF-BEF transitions to altitude, slope, solar radiation, soil moisture (TWI) and wind, for the three ambits in the 1987–2012 period. The density of presence events is represented by vertical lines above x-axes, and overlaid solid black ticks depict 0.05, 0.5 and 0.95 percentiles.

Vicente-Serrano et al., 2010b; Pasho et al., 2011; Büntgen et al., 2013), especially in the case of *Pinus Sylvestris* (Bigler et al., 2006; Sánchez-Salguero et al., 2012; Vilà-Cabrera et al., 2013) which is in the southernmost populations of its distribution. The adaptive strategy of sclerophyllia in evergreen *Quercus* spp. and its larger and deeper roots suggest a greater resistance to water shortage. In the same line, other researches show that seedling and recruitment of *Quercus pubescens* (sub–Mediterranean broadleaf deciduous tree) benefits from drought conditions in mixed forests with *Pinus sylvestris* (Galiano et al., 2010, 2013; Rigling et al., 2013). This does not mean that broadleaves sprout as a consequence of drought. It is logical to suppose that they have been in the understory or that they were not the dominant species, and after canopy defoliation or tree dieback of *Pinus sylvestris*, the dynamics of broadleaves remain stable. Our study also points out that *Quercus pubescens* could be more dynamic than evergreen *Quercus* spp. under drought conditions (especially on north faces), supporting the results obtained in two works by Galiano et al., (2010, 2013). These authors reported the association between *Quercus* spp. seedling abundance and drought occurrence, and in both works they found that seedling recruitment of *Quercus pubescens* was significantly greater than *Quercus ilex*.

However, the interaction with topo-climatic factors must be taken into account when drought tolerance is analyzed. For instance, successions from shrubland towards Mediterranean broadleaf forests are driven by drought tolerance, but under favorable topo–climatic conditions of water availability (low solar radiation levels, high soil moisture values and low wind velocities), indicating that juvenile plants have narrower niches to withstand a disturbance event (Martínez-Vilalta and Lloret, 2016, Jackson et al., 2009). This is consistent with other works that reported the drought tolerance of *Quercus ilex* recruitment and seedlings, especially when soil water availability is not a limiting factor (Blanco et al., 1997; Benayas, 1998; Corcuera et al., 2004). On the other hand, evergreen broadleaves on very steep slopes are vulnerable to longer droughts (Fig. 7), suggesting that the combination of abiotic stress effects reduces the competitive ability of these species.

Drought-induced vegetation decline has been reported as a global trend (Allen et al., 2010). The Mediterranean basin and particularly the Iberian Peninsula are geographical regions where defoliation and tree mortality of the main forest species are associated with the recent increase in drier conditions (Carnicer et al., 2011; Neuman et al., 2017). However, our results clearly suggest that broadleaves are better able to withstand drought disturbances; and that the observed decline in shifts to conifers could be driven by the increase in drought frequency, as is also reported for other regions (Mueller et al., 2005; Dolanc et al., 2013; McDowell et al., 2016). These forest dynamics may have effects on the carbon storage of Mediterranean forests because it has been proven that the increase in broadleaves in mixed forests with conifers enhances carbon sequestration (Vayreda et al., 2012). At the same time, and according to climate models (Giorgi and Lionello, 2008), long-term drought-induced vegetation shifts could have effects on fires, whose frequency and severity have increased in southern Europe during the last decades (Pausas et al., 2009; Moreira et al., 2011; Ganteaume and Jappiot, 2013). In this sense, some studies have shown that pines are more highly flammable than evergreen broadleaves (Ganteaume et al., 2011; Kauf et al., 2014), and there is evidence of lower fire hazard in mixed forests of broadleaves and pines and mature evergreen broadleaf forests than in pure pine forests (Fernandes, 2009; Moreira et al., 2009; Fernandes et al., 2010). On the other hand, Ganteaume et al. (2011) found a higher ignition frequency and shorter time-to-ignition in leaf litter of mixed stands composed of Mediterranean pines and broadleaves than in litter of pure pine stands.

We are aware that historical land management has had a strong influence on the vegetation dynamics in the Iberian Peninsula, and that land abandonment and the replacement of firewood by fossil fuels has contributed to the expansion of broadleaves (Vayreda et al., 2016). Nevertheless, the consistency that our approach provides, because it

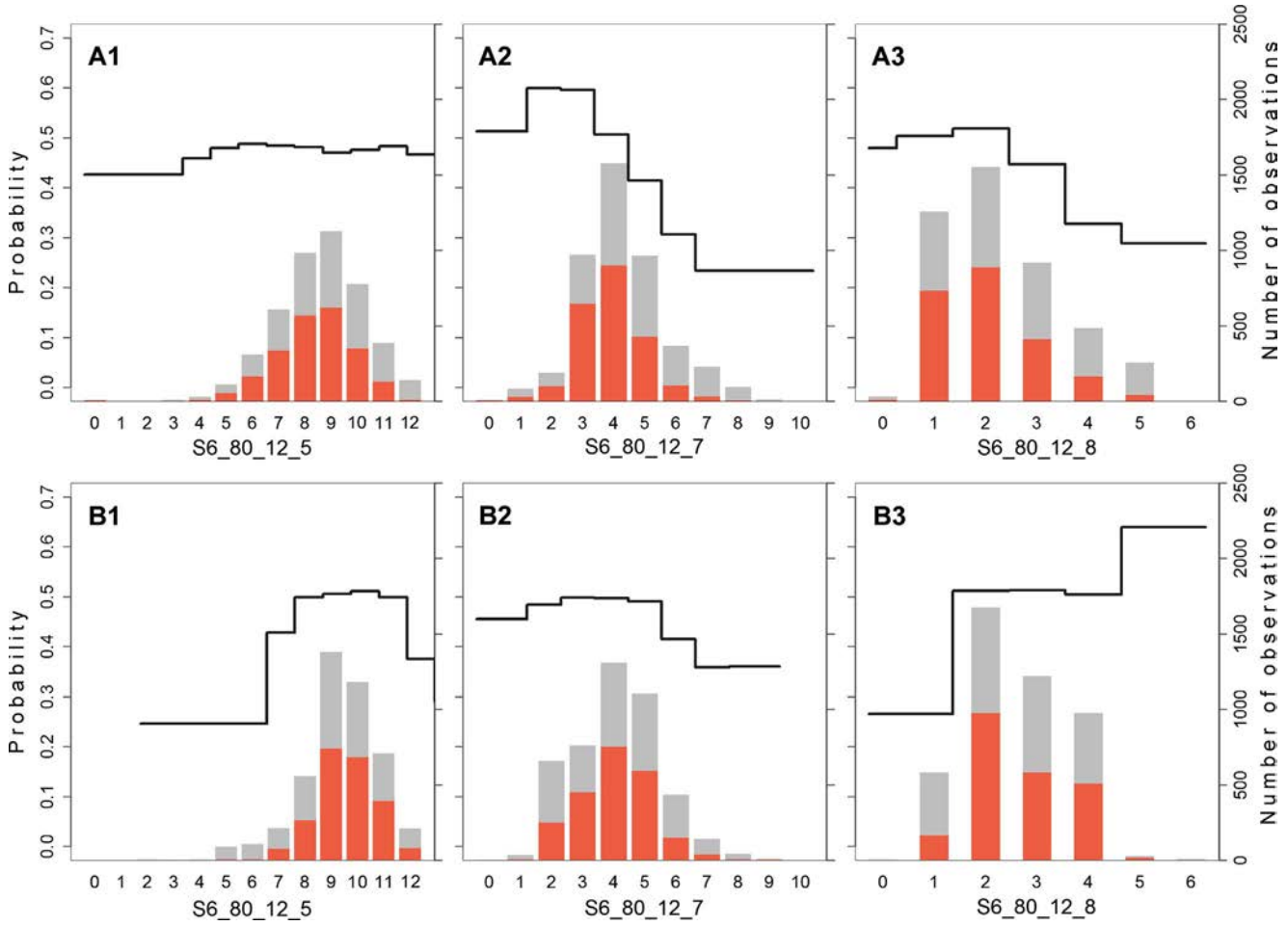


Fig. 6. Response of CNF-BEF transitions to different drought durations. A1-A3: ambit NE1, 1987 2012 period. B1-B3: ambit NE2, 1987–2012 period. Black solid line depicts occurrence probability and vertical bars depict density of observations in the training set by number of occurrences (presence events in red, absence events in grey).

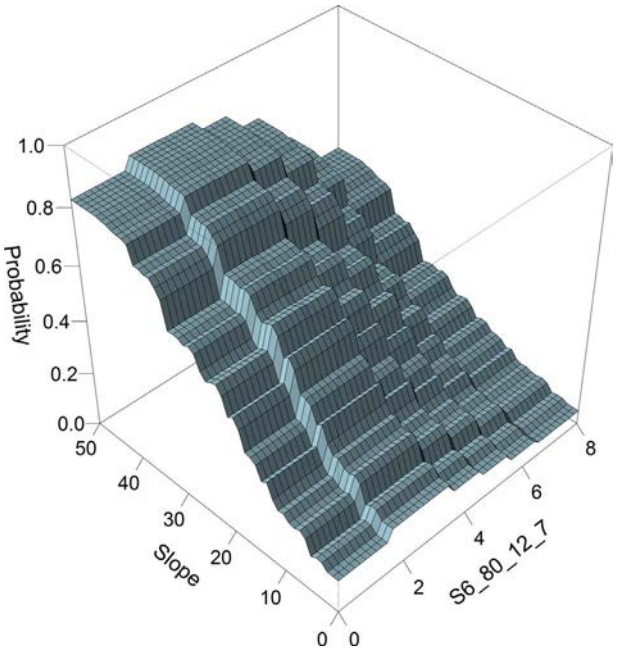


Fig. 7. Interaction between slope and the SPEI at a 6 month timescale and 7-month duration in CNF-BEF transitions, ambit NE1, 1987–2012 period.

covers a broad and diverse region and the robustness of BRT models, reinforces the idea that drought occurrence plays a very important role in vegetation shifts of this geographical region.

5. Conclusions

Forest transitions in three large areas of the Iberian Peninsula have been analyzed considering the role of several variables in vegetation groups dynamics, from early developing stages to successions within forest vegetation stages. Topography-related variables, such as solar radiation, soil moisture and slope, as well as drought tolerance have been shown as key factors in determining certain transitions. Drought occurrence has a negative influence on transitions to coniferous forest and a positive association with transitions towards Mediterranean and sub-Mediterranean broadleaf species. However, the study shows that interaction with topo-climatic factors, as well as drought duration, plays an important role in withstanding drought events.

Our work is in line with previous studies that indicate the vulnerability of species such as *Pinus sylvestris* (Martínez-Vilalta and Piñol, 2002; Galiano et al., 2013; Vilà-Cabrera et al., 2013; Sánchez-Salguero, 2016) to the increase in drier conditions predicted for the future (IPCC, 2013). Populations of this species in the Mediterranean area are situated at their southernmost limit, which may indicate that this limit is moving northwards due to recurrent droughts. This could lead to an expansion of Mediterranean and sub-Mediterranean *Quercus* spp. and

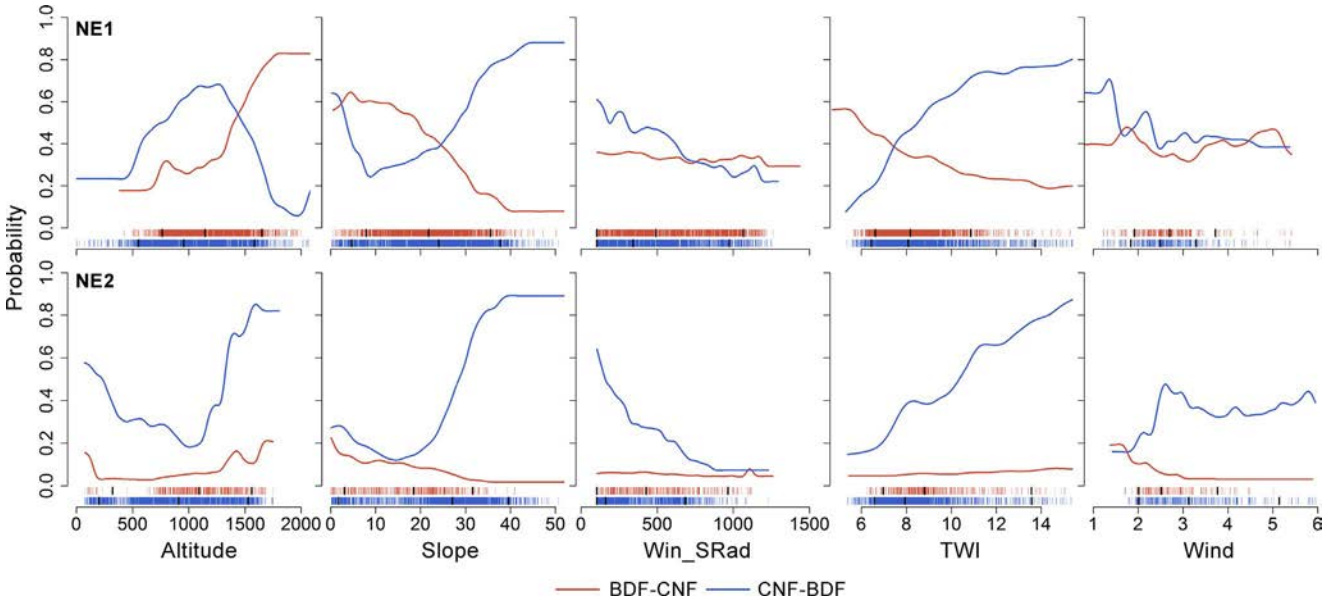


Fig. 8. Response of BDF-CNF and CNF-BDF transitions to altitude, slope, solar radiation, soil moisture (TWI) and wind, in ambits NE1 and NE2 in the 1987–2012 period. Density of presence events is represented by vertical lines above the x-axes, and overlaid solid black ticks depict 0.05, 0.5 and 0.95 percentiles.

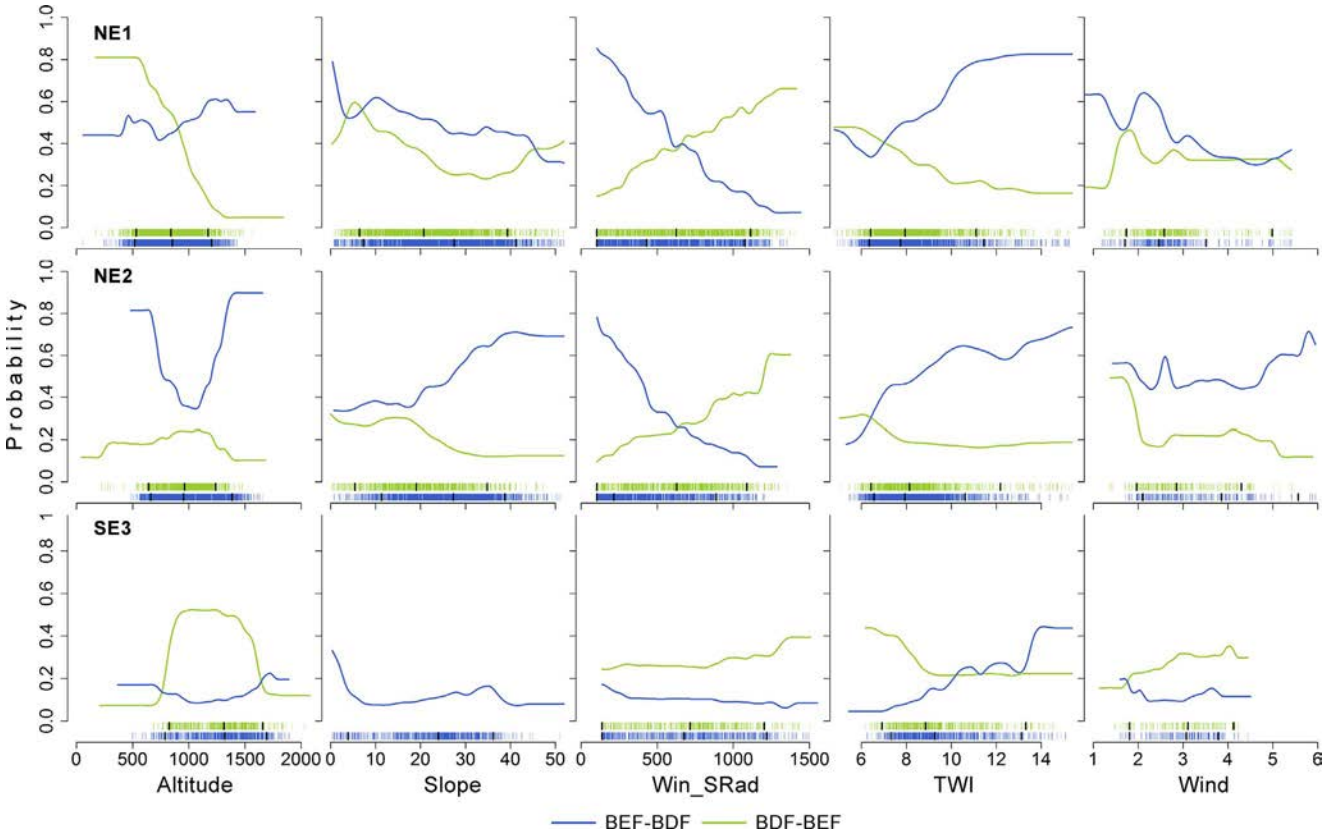


Fig. 9. Response of BDF-BEF and BEF-BDF transitions to altitude, slope, solar radiation, soil moisture (TWI) and wind, 1987–2012 period. Slope variable is not shown in BDF-BEF of ambit SE3 because after the VIF analysis it was removed. Density of presence events is represented by vertical lines above the x-axes, and overlaid solid black ticks depict 0.05, 0.5 and 0.95 percentiles.

an upward shift of their altitudinal range (Peñuelas and Boada, 2003; Gimmi et al., 2010; Vilà-Cabrera et al., 2013; Rigling et al., 2013; Vayreda et al., 2016). With these considerations, shifts detected in the present work should be monitored in future works to assess whether they are eventual shifts, stages of ecological succession, or shifts consolidated over time (as mature vegetation).

Acknowledgements

The authors acknowledge the support from Lluís Pesquer (Grumets Research Group, CREAF, Universitat Autònoma de Barcelona) and Oscar González-Guerrero (Grumets Research Group, Geography Department, Universitat Autònoma de Barcelona) for processing pseudoinvariant areas in the radiometric correction phase. Xavier Pons is

recipient of an ICREA Academia Excellence in Research Grant (2016–2020). This work was supported by the Spanish Ministry of Economy and Competitiveness [grant number BES-2013-063766 to Juan José Vidal-Macua]; European Union's Horizon 2020 Programme [ECOPOTENTIAL (641762-2)]; Spanish Ministry of Economy and Competitiveness [ACAPI (CGL2015-69888-P MINECO/FEDER)]; Catalan Government [SGR2014-1491].

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.foreco.2017.10.011.

References

Allen, C.D., Breshears, D.D., 1998. Drought-induced shift of a forest–woodland ecotone: rapid landscape response to climate variation. Proc. Natl. Acad. Sci. 95 (25), 14839–14842.
Allen, C.D., Macalady, A.K., Chenchouni, H., Bachelet, D., McDowell, N., Vennetier, M., et al., 2010. A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. For. Ecol. Manage. 259 (4), 660–684.
Álvarez-Martínez, J.M., Stoorvogel, J.J., Suárez-Seoane, S., de Luis Calabuig, E., 2010. Uncertainty analysis as a tool for refining land dynamics modelling on changing landscapes: a case study in a Spanish Natural Park. Landscape Ecol. 25 (9), 1385–1404.
Améztegui, A., Brotons, L., Coll, L., 2010. Land-use changes as major drivers of mountain pine (Pinus uncinata Ram. expansion in the Pyrenees. Glob. Ecol. Biogeogr. 19 (5), 632–641.
Ameztegui, A., Coll, L., 2011. Tree dynamics and co-existence in the montane–sub-alpine ecotone: the role of different light-induced strategies. J. Veg. Sci. 22 (6), 1049–1061.
Andreu, L., Gutiérrez, E., Macias, M., Ribas, M., Bosch, O., Camarero, J.J., 2007. Climate increases regional tree-growth variability in Iberian pine forests. Glob. Change Biol. 13 (4), 804–815.
Alodos, C.L., Pueyo, Y., Barrantes, O., Escós, J., Giner, L., Robles, A.B., 2004. Variations in landscape patterns and vegetation cover between 1957 and 1994 in a semiarid Mediterranean ecosystem. Landscape Ecol. 19 (5), 543–559.
Allen, T.R., Walsh, S.J., Cairns, D.M., Messina, J.P., Butler, D.R., Malanson, G.P., 2004. Geostatistics and spatial analysis: characterizing form and pattern at the alpine treeline. In: Geographic Information Science and Mountain Geomorphology. Springer Science & Business Media, pp. 190–218.
Benayas, J.M.R., 1998. Growth and survival in Quercus ilex L. seedlings after irrigation and artificial shading on Mediterranean set-aside agricultural land. Ann. des Sci. For. 55 (7), 801–807.
Bennie, J., Hill, M.O., Baxter, R., Huntley, B., 2006. Influence of slope and aspect on long-term vegetation change in British chalk grasslands. J. Ecol. 94 (2), 355–368.
Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. Hydrol. Sci. J. 24 (1), 43–69.
Bigler, C., Bräker, O.U., Bugmann, H., Dobbertin, M., Rigling, A., 2006. Drought as an inciting mortality factor in Scots pine stands of the Valais, Switzerland. Ecosystems 9 (3), 330–343.
Blanco, E., Casado, M.A., Costa, M., Escribano, R., García, M., Génova, M., Gómez, F., Moreno, J.C., Morla, C., Regat, P., Sáinz, H., 1997. Los Bosques Ibéricos, una interpretación geobotánica. Planeta, Barcelona.
Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees. Wadsworth & Brooks, Monterey.
Breiman, L., 1996. Bagging predictors. Mach. Learn. 24 (2), 123–140.
Breiman, L., 2001. Statistical modeling: The two cultures (with comments and a rejoinder by the author. Stat. Sci. 16 (3), 199–231.
Breshears, D.D., Cobb, N.S., Rich, P.M., Price, K.P., Allen, C.D., Balice, R.G., Romme, W.H., Kastens, J.H., Floyd, M.L., Belnap, J., Anderson, J.J., Myers, O.B., Anderson, J.J., Meyer, C.W., 2005. Regional vegetation die-off in response to global-change-type drought. Proc. Natl. Acad. Sci. U. S. A. 102 (42), 15144–15148.
Broncano, M., Riba, M., Retana, J., 1998. Seed germination and seedling performance of two Mediterranean tree species, holm oak (shape Quercus ilex L. and Aleppo pine (shape Pinus halepensis Mill.): a multifactor experimental approach. Plant Ecol. 138 (1), 17–26.
Büntgen, U., Martínez-Peña, F., Aldea, J., Rigling, A., Fischer, E.M., Camarero, J.J., Hayes, M.J., Fatton, V., Egli, S., 2013. Declining pine growth in Central Spain coincides with increasing diurnal temperature range since the 1970s. Glob. Planet. Change 107, 177–185.
Burrough, P.A., van Gaans, P.F., Hootsmans, R., 1997. Continuous classification in soil survey: spatial correlation, confusion and boundaries. Geoderma 77 (2), 115–135.
Burrough, P.A., van Gaans, P.F., MacMillan, R.A., 2000. High-resolution landform classification using fuzzy k-means. Fuzzy Sets Syst. 113 (1), 37–52.
Burrough, P.A., Wilson, J.P., Van Gaans, P.F., Hansen, A.J., 2001. Fuzzy k-means classification of topo-climatic data as an aid to forest mapping in the Greater Yellowstone Area, USA. Landscape Ecology 16 (6), 523–546.
Camarero, J.J., Bigler, C., Linares, J.C., Gil-Pelegrín, E., 2011. Synergistic effects of past historical logging and drought on the decline of Pyrenean silver fir forests. For. Ecol. Manage. 262 (5), 759–769.
Carnicer, J., Coll, M., Ninyerola, M., Pons, X., Sánchez, G., Peñuelas, J., 2011. Widespread crown condition decline, food web disruption, and amplified tree mortality with increased climate change-type drought. Proc. Natl. Acad. Sci. 108 (4), 1474–1478.
Carnicer, J., Coll, M., Pons, X., Ninyerola, M., Vayreda, J., Peñuelas, J., 2014. Large-scale recruitment limitation in Mediterranean pines: the role of Quercus ilex and forest successional advance as key regional drivers. Glob. Ecol. Biogeogr. 23 (3), 371–384.
Castro, J., Zamora, R., Hódar, J.A., Gómez, J.M., 2004. Seedling establishment of a boreal tree species (Pinus sylvestris. at its southernmost distribution limit: consequences of being in a marginal Mediterranean habitat. J. Ecol. 92 (2), 266–277.
Coll, M., Carnicer, J., Peñuelas, J., Pons, X., Ninyerola, M., 2013. Multivariate effect gradients driving forest demographic responses in the Iberian peninsula. For. Ecol. Manage. 303 (1), 195–209.
Corcuera, L., Camarero, J.J., Gil-Pelegrín, E., 2004. Effects of a severe drought on Quercus ilex radial growth and xylem anatomy. Trees 18 (1), 83–92.
Crase, B., Liedloff, A.C., Wintle, B.A., 2012. A new method for dealing with residual spatial autocorrelation in species distribution models. Ecography 35 (10), 879–888.
Cohen, M., Varga, D., Vila, J., Barrassaud, E., 2011. A multi-scale and multi-disciplinary approach to monitor landscape dynamics: a case study in the Catalan pre-Pyrenees (Spain). Geogr. J. 177 (1), 79–91.
De'Ath, G., 2007. Boosted trees for ecological modeling and prediction. Ecology 88 (1), 243–251.
De Luis, M., Brunetti, M., Gonzalez-Hidalgo, J.C., Longares, L.A., Martín-Vide, J., 2010. Changes in seasonal precipitation in the Iberian Peninsula during 1946–2005. Global Planet. Change 74 (1), 27–33.
Del Río, S.D., Herrero, L., Fraile, R., Penas, A., 2011. Spatial distribution of recent rainfall trends in Spain (1961–2006. Int. J. Climatol. 31 (5), 656–667.
Del Río, S., Cano-Ortiz, A., Herrero, L., Penas, A., 2012. Recent trends in mean maximum and minimum air temperatures over Spain (1961–2006. Theoret. Appl. Climatol. 109 (3–4), 605–626.
Díaz-Delgado, R., Pons, X., 2001. Spatial patterns of forest fires in Catalonia (NE of Spain. along the period 1975–1995: analysis of vegetation recovery after fire. For. Ecol. Manage. 147 (1), 67–74.
Domingo-Marimon, C., 2016. Contributions to the Knowledge of the Multitemporal Spatial Patterns of the Iberian Peninsula Droughts from a Geographic Information Science Perspective. Universitat Autònoma de Barcelona (Ph.D. thesis).
Dolanc, C.R., Thorne, J.H., Safford, H.D., 2013. Widespread shifts in the demographic structure of subalpine forests in the Sierra Nevada, California, 1934 to 2007. Glob. Ecol. Biogeogr. 22 (3), 264–276.
Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129–151.
Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77 (4), 802–813.
Elith, J., Graham, C.H., 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography 32 (1), 66–77.
El Kenawy, A., López-Moreno, J.I., Vicente-Serrano, S.M., 2012. Trend and variability of surface air temperature in northeastern Spain (1920–2006): linkage to atmospheric circulation. Atmos. Res. 106, 159–180.
Espelta, J.M., Riba, M., Javier, R., 1995. Patterns of seedling recruitment in West-Mediterranean Quercus ilex forest influenced by canopy development. J. Veg. Sci. 6 (4), 465–472.
Fernandes, P.M., 2009. Combining forest structure data and fuel modelling to classify fire hazard in Portugal. Ann. For. Sci. 66 (4), 1–9.
Florinsky, I.V., Kuryakova, G.A., 1996. Influence of topography on some vegetation cover properties. Catena 27 (2), 123–141.
Freund, Y., Schapire, R.E., 1996. Experiments with a new boosting algorithm. In: International Conference of Machine Learning. New Jersey, 1996.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 1189–1232.
Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal. 38 (4), 367–378.
Friedman, J.H., Meulman, J.J., 2003. Multiple additive regression trees with application in epidemiology. Stat. Med. 22 (9), 1365–1381.
Galiano, L., Martínez-Vilalta, J., Lloret, F., 2010. Drought-induced multifactor decline of Scots pine in the Pyrenees and potential vegetation change by the expansion of co-occurring oak species. Ecosystems 13 (7), 978–991.
Galiano, L., Martínez-Vilalta, J., Eugenio, M., Granzow-de la Cerda, Í., Lloret, F., 2013. Seedling emergence and growth of Quercus spp. following severe drought effects on a Pinus sylvestris canopy. J. Veg. Sci. 24 (3), 580–588.
García-Ruiz, J.M., 1990. Geoecología de las Áreas de Montaña. Geoforma Ediciones, Logroño.
Ganteaume, A., Marielle, J., Corinne, L.M., Thomas, C., Laurent, B., 2011. Effects of vegetation type and fire regime on flammability of undisturbed litter in Southeastern France. For. Ecol. Manage. 261 (12), 2223–2231.
Ganteaume, A., Jappiot, M., 2013. What causes large fires in Southern France. For. Ecol. Manage. 294, 76–85.
Gehrig-Fasel, J., Guisan, A., Zimmermann, N.E., 2007. Tree line shifts in the Swiss Alps: climate change or land abandonment? J. Veget. Sci. 18 (4), 571–582.
Giorgi, F., Lionello, P., 2008. Climate change projections for the Mediterranean region. Glob. Planet. Change 63 (2), 90–104.
Glenn-Lewin, D.C., van der Maarel, E., 1992. Patterns and Processes of Vegetation Dynamics. Plant Succession: Theory and Prediction, pp. 11–59.

Gimmi, U., Wohlgemuth, T., Rigling, A., Hoffmann, C.W., Bürgi, M., 2010. Land-use and climate change effects in forest compositional trajectories in a dry Central-Alpine valley. Ann. For. Sci. 67 (7), 701.
Gutman, G., Huang, C., Chander, G., Noojipady, P., 2013. Assessment of the NASA-USGS Global Land Survey (GLS) Datasets. Remote Sens. Environ. 134, 249–265.
Gobierno de Aragón, 2016. Departamento de Desarrollo Rural y Sostenibilidad. Dirección General de Medio Forestal.
GECCC, Grup d'Experts en Canvi Climàtic de Catalunya. Generalitat de Catalunya, 2016. Departament d'Agricultura, Ramaderia, Pesca i Alimentació.
Gonzalez-Hidalgo, J.C., López-Bustins, J.A., Štepánek, P., Martín-Vide, J., de Luis, M., 2009. Monthly precipitation trends on the Mediterranean fringe of the Iberian Peninsula during the second-half of the twentieth century (1951–2000). Int. J. Climatol. 29 (10), 1415–1429.
Gorsevski, P.V., Jankowski, P., Gessler, P.E., 2005. Spatial prediction of landslide hazard using Fuzzy k-means and Dempster-Shafer theory. Trans. GIS 9 (4), 455–474.
Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-resolution global maps of 21st-century forest cover change. Science 342 (6160), 850–853.
Hastie, T., Tibshirani, R.I., 2009. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer, Berlin.
Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., Campbell, L.B., 2016. Mass data processing of time series Landsat imagery: pixels to data products for forest monitoring. Int. J. Dig. Earth 9 (11), 1035–1054.
IPCC, 2007. Climate change 2007: the physical science basis. Contribution of Working Group I to the Forth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge.
IPCC, 2013. Climate change 2013: the physical science basis. In: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge and New York.
Jackson, S.T., Betancourt, J.L., Booth, R.K., Gray, S.T., 2009. Ecology and the ratchet of events: climate variability, niche dimensions, and species distributions. Proc. Natl. Acad. Sci. 106 (Supplement 2), 19685–19692.
James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning. Springer, New York.
Jiménez Olivencia, Y., Porcel Rodríguez, L., Píñar Álvarez, A., 2006. Evolución histórica de los paisajes del Parque Nacional de Sierra Nevada y su entorno. Proyectos de investigación en Parques Nacionales 2009, 109–128.
Kauf, Z., Fangmeier, A., Rosavec, R., Španjol, Ž., 2014. Testing vegetation flammability: the problem of extremely low ignition frequency and overall flammability score. J. Combust 2014.
Kawakita, M., Minami, M., Eguchi, S., Lennert-Cody, C.E., 2005. An introduction to the predictive technique AdaBoost with a comparison to generalized additive models. Fish. Res. 76 (3), 328–343.
Kelly, A.E., Goulden, M.L., 2008. Rapid shifts in plant distribution with recent climate change. Proc. Natl. Acad. Sci. 105 (33), 11823–11826.
Kopecký, M., Čížková, Š., 2010. Using topographic wetness index in vegetation ecology: does the algorithm matter? Appl. Veg. Sci. 13 (4), 450–459.
Lasanta-Martínez, T., Vicente-Serrano, S.M., Cuadrat-Prats, J.M., 2005. Mountain Mediterranean landscape evolution caused by the abandonment of traditional primary activities: a study of the Spanish Central Pyrenees. Appl. Geogr. 25 (1), 47–65.
Leathwick, J.R., Elith, J., Francis, M.P., Hastie, T., Taylor, P., 2006. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. Mar. Ecol. Prog. Ser. 321, 267–281.
Leathwick, J.R., Elith, J., Chadderton, W.L., Rowe, D., Hastie, T., 2008. Dispersal, disturbance and the contrasting biogeographies of New Zealand's diadromous and non-diadromous fish species. J. Biogeogr. 35 (8), 1481–1497.
Levers, C., Verkerk, P.J., Müller, D., Verburg, P.H., Butsic, V., Leitão, P.J., Kuemmerle, T., 2014. Drivers of forest harvesting intensity patterns in Europe. For. Ecol. Manage. 315, 160–172.
Lévesque, M., Saurer, M., Siegwolf, R., Eilmann, B., Brang, P., Bugmann, H., Rigling, A., 2013. Drought response of five conifer species under contrasting water availability suggests high vulnerability of Norway spruce and European larch. Glob. Change Biol. 19 (10), 3184–3199.
Lewis, H.G., Brown, M., Tatnall, A.R.L., 2000. Incorporating uncertainty in land cover classification from remote sensing imagery. Adv. Space Res. 26 (7), 1123–1126.
Lloret, F., Siscart, D., Dalmases, C., 2004. Canopy recovery after drought dieback in holm-oak Mediterranean forests of Catalonia (NE Spain). Global Change Biol. 10 (12), 2092–2099.
López-Moreno, J.I., Vicente-Serrano, S.M., Angulo-Martínez, M., Beguería, S., Kenawy, A., 2010. Trends in daily precipitation on the northeastern Iberian Peninsula, 1955–2006. Int. J. Climatol. 30 (7), 1026–1041.
McDowell, N.G., Coops, N.C., Beck, P.S., Chambers, J.Q., Gangodagamage, C., Hicke, J.A., et al., 2015. Global satellite monitoring of climate-induced vegetation disturbances. Trends Plant Sci. 20 (2), 114–123.
McDowell, N.G., Williams, A.P., Xu, C., Pockman, W.T., Dickman, L.T., Sevanto, S., et al., 2016. Multi-scale predictions of massive conifer mortality due to chronic temperature rise. Nat. Clim. Change 6 (3), 295–300.
Martínez-Vilalta, J., Piñol, J., 2002. Drought-induced mortality and hydraulic architecture in pine populations of the NE Iberian Peninsula. For. Ecol. Manage. 161 (1), 247–256.
Martínez-Vilalta, J., Lloret, F., 2016. Drought-induced vegetation shifts in terrestrial ecosystems: the key role of regeneration dynamics. Glob. Planet. Change 144, 94–108.
Martínez-Ferri, E., Balaguer, L., Valladares, F., Chico, J.M., Manrique, E., 2000. Energy dissipation in drought-avoiding and drought-tolerant tree species at midday during the Mediterranean summer. Tree Physiol. 20 (2), 131–138.
Moisen, G.G., Freeman, E.A., Blackard, J.A., Frescino, T.S., Zimmermann, N.E., Edwards, T.C., 2006. Predicting tree species presence and basal area in Utah: a comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. Ecol. Model. 199 (2), 176–187.
Moeslund, J.E., Arge, L., Bøcher, P.K., Dalgaard, T., Ejrnæs, R., Odgaard, M.V., Svenning, J.C., 2013. Topographically controlled soil moisture drives plant diversity patterns within grasslands. Biodivers. Conserv. 22 (10), 2151–2166.
Monleon, V.J., Lintz, H.E., 2015. Evidence of tree species' range shifts in a complex landscape. PLoS One 10 (1).
Moreira, F., Vaz, P., Catry, F., Silva, J.S., 2009. Regional variations in wildfire susceptibility of land-cover types in Portugal: implications for landscape management to minimize fire hazard. Int. J. Wildland Fire 18 (5), 563–574.
Fernandes, P., 2010. Scientific Knowledge and Operational Tools to Support Prescribed Burning: Recent Developments. Towards Integrated Fire Management Outcomes of the European Project Fire Paradox, 151.
Moreira, F., Viedma, O., Arianoutsou, M., Curt, T., Koutsias, N., Rigolot, E., et al., 2011. Landscape–wildfire interactions in southern Europe: implications for landscape management. J. Environ. Manage. 92 (10), 2389–2402.
Mueller, R.C., Scudder, C.M., Porter, M.E., Talbot Trotter, R., Gehring, C.A., Whitham, T.G., 2005. Differential tree mortality in response to severe drought: evidence for long-term vegetation shifts. J. Ecol. 93 (6), 1085–1093.
Neumann, M., Mues, V., Moreno, A., Hasenauer, H., Seidl, R., 2017. Climate variability drives recent tree mortality in Europe. Glob. Change Biol.
Norman, S.P., Koch, F.H., Hargrove, W.W., 2016. Review of broad-scale drought monitoring of forests: Toward an integrated data mining approach. For. Ecol. Manage. 380, 346–358.
O'brien, R.M., 2007. A caution regarding rules of thumb for variance inflation factors. Qual. Quant. 41 (5), 673–690.
Ojeda, A.O., 2007. Territorio fluvial: diagnóstico y propuesta para la gestión ambiental y de riesgos en el Ebro y los cursos bajos de sus afluentes. Bakeaz/Fundación Nueva Cultura del Agua, Bilbao.
Parcerisas, L., Marull, J., Pino, J., Tello, E., Coll, F., Basnou, C., 2012. Land use changes, landscape ecology and their socioeconomic driving forces in the Spanish Mediterranean coast (El Maresme County, 1850–2005. Environ. Sci. Policy 23, 120–132.
Pasho, E., Camarero, J.J., de Luis, M., Vicente-Serrano, S.M., 2011. Impacts of drought at different time scales on forest growth across a wide climatic gradient in north-eastern Spain. Agric. For. Meteorol. 151 (12), 1800–1811.
Pausas, J.G., Llovet, J., Rodrigo, A., Vallejo, R., 2009. Are wildfires a disaster in the Mediterranean basin?–a review. Int. J. Wildland Fire 17 (6), 713–723.
Pausas, J.G., Keeley, J.E., 2009. A burning story: the role of fire in the history of life. Bioscience 59 (7), 593–601.
Pausas, J.G., Fernández-Muñoz, S., 2012. Fire regime changes in the Western Mediterranean Basin: from fuel-limited to drought-driven fire regime. Clim. Change 110 (1), 215–226.
Peñuelas, J., Boada, M., 2003. A global change-induced biome shift in the Montseny mountains (NE Spain. Glob. Change Biol. 9 (2), 131–140.
Pons, X., Solé-Sugrañes, L., 1994. A simple radiometric correction model to improve automatic mapping of vegetation from multispectral satellite data. Remote Sens. Environ. 48 (2), 191–204.
Pons, X., Serra, P., Saurí, D., 2003. A Rigorous Protocol for Post-classification Land Cover and Land Use Change Detection. Geoinformation for European-wide integration. Millpress Science, Rotterdam.
Pons, X., Ninyerola, M., 2008. Mapping a topographic global solar radiation model implemented in a GIS and refined with ground data. Int. J. Climatol. 28 (13), 1821–1834.
Pons, X., Sevillano, E., Moré, G., Serra, P., Cornford, D., Ninyerola, M., 2014. Distribución espacial de la incertidumbre en mapas de cubiertas obtenidos mediante teledetección. Revista de Teledeteccion 42, 1–10.
Pontius, R.G., Shusas, E., McEachern, M., 2004. Detecting important categorical land changes while accounting for persistence. Agr. Ecosyst. Environ. 101 (2), 251–268.
Pino, J., Rodà, F., Ribas, J., Pons, X., 2000. Landscape structure and bird species richness: implications for conservation in rural areas between natural parks. Landscape Urban Plann. 49 (1), 35–48.
Red de información Ambiental de Andalucía (REDIAM), 2016. Consejería de Medio Ambiente y Ordenación del Territorio. Junta de Andalucía.
Ridgeway, G., 1999. The state of boosting. Comput. Sci. Stat. 172–181.
Ridgeway, G., Ridgeway, M. G., 2004. The gbm package. R Foundation for Statistical Computing, Vienna, Austria vol. 5(3).
Rigling, A., Bigler, C., Eilmann, B., Feldmeyer-Christe, E., Gimmi, U., Ginzler, C., Graf, U., Mayer, P., Vacchiano, G., Weber, P., Wohlgemuth, T., Zweifel, Z., Dobbertin, M., 2013. Driving factors of a vegetation shift from Scots pine to pubescent oak in dry Alpine forests. Glob. Change Biol. 19 (1), 229–240.
Salvador, R., Lloret, F., Pons, X., Pinol, J., 2005. Does fire occurrence modify the probability of being burned again? A null hypothesis test from Mediterranean ecosystems in NE Spain. Ecol. Model. 188 (2), 461–469.
Sánchez-Salguero, R., Navarro-Cerrillo, R.M., Swetnam, T.W., Zavala, M.A., 2012. Is drought the main decline factor at the rear edge of Europe? The case of southern Iberian pine plantations. For. Ecol. Manage. 271, 158–169.
Sánchez-Salguero, R., Camarero, J.J., Gutiérrez, E., González-Rouco, F., Gazol, A., Sangüesa-Barreda, G., Andreu-Hayles, L., Linares, J.C., Seftigen, K., 2016. Assessing forest vulnerability to climate warming using a process-based model of tree growth: bad prospects for rear-edges. Glob. Change Biol.
Sankaran, M., Ratnam, J., Hanan, N., 2008. Woody cover in African savannas: the role of resources, fire and herbivory. Glob. Ecol. Biogeogr. 17 (2), 236–245.

Sappington, J.M., Longshore, K.M., Thompson, D.B., 2007. Quantifying landscape ruggedness for animal habitat analysis: a case study using bighorn sheep in the Mojave Desert. J. Wildlife Manage. 71 (5), 1419–1426.
Schuur, E.A., 2003. Productivity and global climate revisited: the sensitivity of tropical forest growth to precipitation. Ecology 84 (5), 1165–1170.
Serneels, S., Lambin, E.F., 2001. Proximate causes of land-use change in Narok District, Kenya: a spatial statistical model. Agric. Ecosyst. Environ. 85, 65–81.
Serra, P., Pons, X., Saurí, D., 2008. Land-cover and land-use change in a Mediterranean landscape: a spatial analysis of driving forces integrating biophysical and human factors. Appl. Geogr. 28 (3), 189–209.
Serra-Díaz, J.M., Cristobal, J., Ninyerola, M., 2011. A classification procedure for mapping topo-climatic conditions for strategic vegetation planning. Environ. Model. Assess. 16 (1), 77–89.
Sluiter, R., de Jong, S.M., 2007. Spatial patterns of Mediterranean land abandonment and related land cover transitions. Landscape Ecol. 22 (4), 559–576.
Sörensen, R., Zinko, U., Seibert, J., 2006. On the calculation of the topographic wetness index: evaluation of different methods based on field observations. Hydrol. Earth System Sci. Discuss. 10 (1), 101–112.
Tapia, R., Stein, A., Bijker, W., 2005. Optimization of sampling schemes for vegetation mapping using fuzzy classification. Remote Sens. Environ. 99 (4), 425–433.
Vayreda, J., Martínez-Vilalta, J., Gracia, M., Retana, J., 2012. Recent climate changes interact with stand structure and management to determine changes in tree carbon stocks in Spanish forests. Glob. Change Biol. 18 (3), 1028–1041.
Vayreda, J., Martínez-Vilalta, J., Gracia, M., Canadell, J.G., Retana, J., 2016. Anthropogenic-driven rapid shifts in tree distribution lead to increased dominance of broadleaf species. Global Change Biol. 22 (12), 3984–3995.
Verkerk, P.J., Levers, C., Kuemmerle, T., Lindner, M., Valbuena, R., Verburg, P.H., Zudin, S., 2015. Mapping wood production in European forests. For. Ecol. Manage. 357, 228–238.
Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., 2010a. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. J. Clim. 23 (7), 1696–1718.
Vicente-Serrano, S.M., Lasanta, T., Gracia, C., 2010b. Aridification determines changes in forest growth in Pinus halepensis forests under semiarid Mediterranean climate conditions. Agric. For. Meteorol. 150 (4), 614–628.
Vicente-Serrano, S.M., Gouveia, C., Camarero, J.J., Beguería, S., Trigo, R., López-Moreno, J.I., Azorín-Molina, C., Pasho, E., Lorenzo-Lacruz, J., Revuelto, J., Morán-Tejeda, E., Morán-Tejeda, E., 2013. Response of vegetation to drought time-scales across global land biomes. Proc. Natl. Acad. Sci. 110 (1), 52–57.
Vicente-Serrano, S.M., López-Moreno, J.I., Beguería, S., Lorenzo-Lacruz, J., Sanchez-Lorenzo, A., García-Ruiz, J.M., Azorín-Molina, C., Morán-Tejeda, E., Revuelto, J., Trigo, R., Coelho, F., Espejo, F., 2014. Evidence of increasing drought severity caused by temperature rise in southern Europe. Environ. Res. Lett. 9 (4), 044001.
Vicente-Serrano, S.M., Camarero, J.J., Zabalza, J., Sangüesa-Barreda, G., López-Moreno, J.I., Tague, C.L., 2015. Evapotranspiration deficit controls net primary production and growth of silver fir: Implications for Circum-Mediterranean forests under forecasted warmer and drier conditions. Agric. For. Meteorol. 206, 45–54.
Vidal-Macua, J.J., Zabala, A., Ninyerola, M., Pons, X., 2017. Developing spatially and thematically detailed backdated maps for land cover studies. Int. J. Dig. Earth 10 (2), 175–206.
Vilà-Cabrera, A., Martínez-Vilalta, J., Galiano, L., Retana, J., 2013. Patterns of forest decline and regeneration across Scots pine populations. Ecosystems 16 (2), 323–335.
Walther, G.R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J., Fromentin, J.M., Hoegh-Guldberg, O., Bairlein, F., 2002. Ecological responses to recent climate change. Nature 416 (6879), 389–395.
Weber, P., Bugmann, H., Rigling, A., 2007. Radial growth responses to drought of Pinus sylvestris and Quercus pubescens in an inner-Alpine dry valley. J. Veg. Sci. 18 (6), 777–792.
Wulder, M.A., White, J.C., Goward, S.N., Masek, J.G., Irons, J.R., Herold, M., Woodcock, C.E., 2008. Landsat continuity: issues and opportunities for land cover monitoring. Remote Sens. Environ. 112 (3), 955–969.
Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., Smith, G.M., 2009. Mixed Effects Models and Extensions in Ecology with R. Spring Science and Business Media, New York.

6S. ARTÍCULO 2*: Supplementary material*

S1. Model fitting and evaluation

The bag fraction is the BRT parameter that controls the randomly selected observations at each new tree. This stochasticity avoids overfitting and reduces computation time, but introduces variance between models (Friedman, 2002; Elith et al., 2008). We used a bag fraction of 0.5 (50% of the observations) as is suggested in Friedman (2002) and Elith et al. (2008) after applying several testbeds. The other main parameters of the BRT models are: tree complexity, learning rate and number of trees. Tree complexity is the number of splits allowed for fitting at each iteration (tree) and is related to the interaction level between variables. A tree complexity of 1 means that all trees are composed of single candidate variable (which might differ from one tree to another) and no interaction with other variables is allowed. Instead, an overly complex tree (involving many variables) could overfit the training data and devalue predictions in a test dataset. The learning rate parameter works like shrinkage techniques in other linear models (James et al., 2013; Hastie et al., 2009), scaling the contribution of each new term (i.e. the constituent trees) in the additive model by a factor between 0 and 1. A low learning rate allows more complex trees to be added gradually without overfitting the model, thus diverse interaction structures are taken into account in the additive model. Finally, the optimal number of trees depends on the choice of the two previous parameters. Normally, low learning rates, in a range of 0.001 to 0.01, imply increasing the number of trees to achieve a good performance (James et al., 2013; Elith et al., 2008; Hastie et al., 2009), although a validation procedure is recommended (Ridgeway 2007). Prediction for a single observation is the sum of predictions over all trees multiplied by the learning rate.

BRT were applied in R software using the *dismo* package (Elith et al., 2008), which is based on the *gbm* package (Ridgeway, 2007). In order to determine the best parameterization for each forest-cover change model, we ran a model for all possible combinations of learning rates 0.001, 0.0025, 0.005, 0.01, 0.015, 0.03 and tree complexities from 1 to 6. This led to a total of 36 models for each forest cover change. For each of these models the optimal number of trees was identified by cross-validation (Elith et al., 2008): by an iterative process, models were developed increasing the number of trees by 50, applying a 10-fold cross-validation at each step, and setting the maximum to a 10000-tree model. Predictive performance is recorded (by the cross-validation) at each step using the deviance (Friedman, 2002; Ridgeway, 2007; Elith et al., 2008; Hastie et al., 2009), which is a loss function that can be seen as the RSS (residual sum of squares) used in other linear approaches. When the average deviance of the five recent iterations is higher than the five previous ones, the deviance has been minimized and the optimal number of trees is identified to fit the final model.

We also calculated the deviance in the test dataset for the 36 models and the one with the lowest deviance value was selected to obtain and interpret the results. In addition, we analyzed the results using the ROC (receiver operating characteristics) curve and by calculating the AUC (area under the curve) score. This analysis allowed us to evaluate the robustness of the selected models by checking the balance between true and false positives (Hanley and McNeil, 1982; Fawcett, 2004; Pontius and Parmentier, 2014).

S2. Model validation

We decided to keep models that had more than 100 presence events (Tables S2, S3 and S4). In general, all models performed well (Figures S1 to S18) with mean AUC scores of 0.897 (±0.029), 0.894 (±0.043) and 0.859 (±0.053) for NE1, NE2 and SE3 respectively. Almost all models that did not include altitude performed worse, especially in ambit SE3. There were a few cases where the probability thresholds used for classifying change and non-change events had a greater overlap between them, like BDF-BEF in NE2 and SE3 (Figure S7, Figure S10 and Figure S16) and CNF-BEF transitions in SE3 (Figure S14, Figure S17 and Figure S18). And a few cases showed a more dispersed distribution, like BDF-CNF in NE1 (Figure S2) and SRB BDF transitions in SE3 (Figure S15 and Figure S17).

References:

Elith, J., Leathwick, J. R., & Hastie, T. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77(4): 802-813.

Hanley, J. A., & McNeil, B. J. 1982. The meaning and use of the area under a receiver operating characteristic (ROC curve). *Radiology* 143(1): 29-36.

Friedman, J. H. 2002. Stochastic gradient boosting. *Computational Statistics & Data Analysis* 38(4): 367-378.

Hastie, T., Tibshirani, R.I. 2009. *The elements of statistical learning: data mining, inference and prediction*. Springer, Berlin

James, G., Witten, D., Hastie, T., & Tibshirani, R. 2013. *An introduction to statistical learning*. Springer, New York.

Pontius, R. G., & Parmentier, B. 2014. Recommendations for using the relative operating characteristic (ROC. *Landscape Ecology* 29(3): 367-382.

Ridgeway, G. 2007. Generalized Boosted Models: A guide to the gbm package. *Update* 1(1): 2007.

	NE1						NE2						SE3					
	1987		2002		2012		1987		2002		2012		1987		2002		2012	
	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA
SRB	92.1	96.2	85.8	93.4	91.6	95.3	90.2	86.4	83.9	81.3	89.8	96	95.6	93.8	91.7	91.6	92.4	89.8
CNF	97.1	93.8	96.2	97.9	99	98.5	95.3	99.5	97.1	99.1	98	98.8	97.8	98.5	94.6	92.9	94.5	96.9
BEF	91.5	74.9	92.4	68.8	93	89.9	95.6	69.1	96.2	70.9	96.5	89.6	98.9	86.5	95.3	91.3	91.1	93.5
BDF	98.8	95.3	98.2	95.8	99	97.8	95.1	93	95.2	84.2	96.7	91.3	98	86.6	96.8	95.3	97	86.9

Table S1. Producer accuracy (PA) and user accuracy (UA) by vegetation cover category. SRB=Shrubland; CNF=Coniferous forest; BEF=Broadleaf evergreen forest; BDF=Broadleaf deciduous forest

	1987-2002				2002-2012				1987-2012			
Vegetation-cover changes	**P**	**A**	**AUC Alt**	**AUC NoAlt**	**P**	**A**	**AUC Alt**	**AUC NoAlt**	**P**	**A**	**AUC Alt**	**AUC NoAlt**
1. SRB-CNF	2998	2998	0.906	0.889	3000	2998	0.909	0.912	3000	2996	0.919	0.906
2. SRB-BEF	3000	2998	0.903	0.862	3000	2996	0.895	0.880	3000	2996	0.916	0.876
3. SRB-BDF	1592	2997	0.941	0.918	2997	2998	0.968	0.959	2991	2994	0.955	0.949
4. CNF-BEF	3000	3000	0.892	0.872	2999	2999	0.877	0.878	3000	3000	0.883	0.865
5. CNF-BDF	1831	3000	0.915	0.914	2967	2999	0.922	0.910	2982	3000	0.928	0.909
6. BEF-CNF	3000	2999	0.869	0.866	2999	3000	0.867	0.874	3000	3000	0.873	0.858
7. BEF-BDF	1051	2999	0.911	0.899	2770	2999	0.914	0.913	2963	3000	0.914	0.917
8. BDF-CNF	2868	2997	0.848	0.843	242	2998	0.919	0.829	2142	3000	0.872	0.869
9. BDF-BEF	1947	3000	0.875	0.870	638	2998	0.923	0.902	2133	2999	0.906	0.896

Table S2. Number of presence and absence events, and AUC score. Ambit NE1. AUC Alt= AUC in models including altitude. AUC NoAlt= AUC in models not including altitude. Rejected models in grey.

	1987-2002				2002-2012				1987-2012			
Vegetation-cover changes	**P**	**A**	**AUC Alt**	**AUC NoAlt**	**P**	**A**	**AUC Alt**	**AUC NoAlt**	**P**	**A**	**AUC Alt**	**AUC NoAlt**
1. SRB-CNF	2975	2997	0.871	0.859	2999	2999	0.907	0.908	2994	3000	0.919	0.906
2. SRB-BEF	3000	2998	0.952	0.919	2998	2997	0.903	0.890	2999	2994	0.910	0.886
3. SRB-BDF	652	2999	0.913	0.907	1118	2996	0.938	0.906	2195	2994	0.937	0.927
4. CNF-BEF	3000	2999	0.868	0.882	2996	2998	0.871	0.887	2998	2996	0.874	0.858
5. CNF-BDF	508	2997	0.955	0.955	2626	2997	0.939	0.941	2280	2996	0.953	0.948
6. BEF-CNF	3000	3000	0.936	0.909	2770	3000	0.924	0.924	3000	3000	0.926	0.926
7. BEF-BDF	2059	3000	0.886	0.863	2999	3000	0.920	0.897	3000	2999	0.914	0.895
8. BDF-CNF	538	2988	0.851	0.833	98	2975			409	2985	0.863	0.879
9. BDF-BEF	1255	2988	0.780	0.771	870	2980	0.870	0.831	1043	2986	0.828	0.805

Table S3. Number of presence and absence events, and AUC score. Ambit NE2. AUC Alt= AUC in models including altitude. AUC NoAlt= AUC in models not including altitude. Rejected models in grey.

	1987-2002				2002-2012				1987-2012			
Vegetation-cover changes	**P**	**A**	**AUC Alt**	**AUC NoAlt**	**P**	**A**	**AUC Alt**	**AUC NoAlt**	**P**	**A**	**AUC Alt**	**AUC NoAlt**
1. SRB-CNF	2134	3000	0.872	0.861	3000	3000	0.884	0.873	3000	3000	0.883	0.854
2. SRB-BEF	3000	3000	0.889	0.865	3000	3000	0.830	0.803	2999	3000	0.891	0.843
3. SRB-BDF	131	3000	0.927	0.900	154	3000	0.923	0.909	198	3000	0.947	0.954
4. CNF-BEF	2050	3000	0.845	0.809	1174	2998	0.776	0.729	2924	3000	0.835	0.780
5. CNF-BDF	18	3000			13	2999			55	2999		
6. BEF-CNF	2003	3000	0.840	0.818	1961	2999	0.825	0.827	2990	2999	0.841	0.814
7. BEF-BDF	563	2999	0.949	0.940	649	3000	0.888	0.902	883	2999	0.927	0.913
8. BDF-CNF	5	1125			10	1459			27	1137		
9. BDF-BEF	530	1145	0.868	0.802	394	1455	0.834	0.759	584	1148	0.842	0.848

Table S4. Number of presence and absence events, and AUC score. Ambit SE3. AUC Alt= AUC in models including altitude. AUC NoAlt= AUC in models not including altitude. Rejected models in grey.

	1987-2002 %	2002-2012 %	1987-2012 %
SRB *	94.7	92.1	82.5
SRB-CNF	2.6	1.5	5.9
SRB-BEF	0.6	1.8	3.5
SRB-BDF	0.1	0.3	0.4
CNF *	98.4	98.1	95.6
CNF-BEF	0.5	0.9	1.8
CNF-BDF	0.1	0.3	0.8
BEF *	96.2	98.0	94.2
BEF-CNF	2.7	1.2	3.9
BEF-BDF	0.3	0.6	1.1
BDF *	98.2	99.8	98.5
BDF-CNF	0.6	0.0	0.6
BDF-BEF	0.3	0.1	0.3

Table S5. Percentage of change by vegetation transition and period in NE1. The percentage of change has been calculated from the masked land-cover maps. *Area that remained stable.

	1987-2002 %	2002-2012 %	1987-2012 %
SRB *	94.3	89.6	83.5
SRB-CNF	0.8	0.5	2.0
SRB-BEF	1.3	5.3	7.4
SRB-BDF	0.1	0.1	0.3
CNF *	96.9	90.9	83.8
CNF-BEF	0.9	6.9	12.5
CNF-BDF	0.1	0.8	0.8
BEF *	85.0	92.9	82.2
BEF-CNF	11.9	2.7	5.8
BEF-BDF	1.9	3.6	10.4
BDF *	96.6	97.1	96.5
BDF-CNF	0.6	0.1	0.5
BDF-BEF	0.9	0.4	0.9

Table S6. Percentage of change by vegetation transition and period in NE2. The percentage of change has been calculated from the masked land-cover maps. *Area that remained stable.

	1987-2002 %	2002-2012 %	1987-2012 %
SRB *	96.0	96.3	90.5
SRB-CNF	0.5	0.7	1.9
SRB-BEF	1.0	1.0	2.8
SRB-BDF	0.0	0.0	0.0
CNF *	89.4	98.0	89.4
CNF-BEF	3.3	1.2	5.5
CNF-BDF	0.0	0.0	0.1
BEF *	95.1	98.3	93.6
BEF-CNF	1.1	0.8	2.6
BEF-BDF	0.3	0.2	0.5
BDF *	77.8	82.5	71.2
BDF-CNF	0.0	0.0	0.1
BDF-BEF	2.5	2.2	4.3

Table S7. Percentage of change by vegetation transition and period in SE3. The percentage of change has been calculated from the masked land-cover maps. *Area that remained stable.

	Including altitude			Not including altitude		
	1987-2012	**2002-2012**	**1987-2002**	**1987-2012**	**2002-2012**	**1987-2002**
SRB-CNF	S24_07_12_8 18.88	S24_07_12_8 18.41	Altitude 18.49	S24_07_12_8 21.32	S24_07_12_8 19.62	PopDen 8.48
	Altitude 16.12	Altitude 17.57	Liv_Units 7.85	PopDen 8.13	PopDen 6.97	Slope 8.17
	PopDen 6.59	PopDen 8.34	Slope 7.74	Wind 5.97	Wind 6.67	Wind 7.96
	Wind 5.36	Wind 5.70	Win_SRad 6.78	Dist_UrbA 5.24	Dist_UrbA 5.18	Liv_Units 7.36
	Win_SRad 4.75	Slope 5.46	PopDen 6.55	Win_SRad 4.45	Slope 4.86	Win_SRad 7.05
	Dist_UrbA 4.68	Dist_UrbA 5.36	Wind 6.04	Pop_Dyn 4.31	VRM 4.59	Dist_UrbA 6.29
SRB-BEF	Altitude 25.27	Altitude 19.99	Altitude 41.79	Win_SRad 10.81	Win_SRad 12.44	Wind 10.61
	Win_SRad 13.93	Win_SRad 12.71	Win_SRad 10.70	VRM 7.95	Wind 7.79	S6_80_02_4 8.97
	VRM 6.02	VRM 6.12	Slope 4.14	Liv_Units 7.25	VRM 7.42	Win_SRad 8.64
	PopDen 5.67	PopDen 6.08	Wind 3.89	Wind 6.94	PopDen 7.01	PopDen 6.96
	TWI 5.14	Wind 5.51	PopDen 3.70	PopDen 6.68	TWI 6.69	VRM 6.02
	Slope 5.00	TWI 5.37	VRM 3.66	TWI 6.60	Liv_Units 6.09	Slope 5.97
SRB-BDF	Altitude 40.93	Altitude 28.88	Altitude 31.78	S6_07_12_4 15.10	S6_07_12_8 17.62	Dist_SecRo 10.01
	Win_SRad 10.57	S6_07_12_8 12.20	Dist_SecRo 7.74	Liv_Units 13.63	Liv_Units 11.28	Liv_Units 9.64
	Wind 6.75	Wind 7.58	Win_SRad 6.88	Win_SRad 10.44	Win_SRad 8.91	S6_97_02_8 8.23
	TWI 5.02	Win_SRad 7.51	TWI 5.83	Wind 8.31	Wind 8.25	Wind 7.66
	Dist_SecRo 4.57	S6_02_12_8 5.06	Dist_UrbA 4.80	TWI 4.94	S6_80_12_4 7.51	TWI 6.78
	Liv_Units 3.86	Liv_Units 4.81	Wind 4.74	Dist_SecRo 4.50	S6_07_12_4 5.81	Win_SRad 6.53
CNF-BEF	Altitude 29.81	Altitude 33.08	Altitude 33.17	Win_SRad 8.20	Win_SRad 10.26	Liv_Units 9.17
	Win_SRad 9.49	Win_SRad 12.93	Win_SRad 9.04	Slope 7.58	Wind 7.86	Wind 9.12
	Slope 8.12	Slope 7.87	Wind 6.38	Wind 7.35	Slope 7.16	Slope 7.12
	Wind 5.79	PopDen 4.66	Slope 5.91	Liv_Units 7.15	PopDen 5.54	Dist_UrbA 6.65
	TWI 4.95	Dist_UrbA 4.55	Liv_Units 5.90	S6_80_12_7 6.86	Dist_MajRo 5.49	Win_SRad 6.57
	Liv_Units 4.63	TWI 4.52	Dist_UrbA 4.43	PopDen 6.49	Liv_Units 5.30	PopDen 6.27
CNF-BDF	Altitude 14.50	Altitude 17.07	Altitude 21.79	S6_07_12_8 9.53	Slope 9.17	Liv_Units 15.24
	Slope 7.86	Slope 9.38	Liv_Units 11.30	Liv_Units 8.47	Liv_Units 8.40	Wind 8.10
	S6_07_12_8 7.57	S6_07_12_8 7.95	PopDen 7.03	Slope 8.35	S6_07_12_8 8.04	PopDen 7.74
	Liv_Units 7.49	Liv_Units 7.05	Wind 6.52	Win_SRad 7.50	Win_SRad 7.26	TWI 7.70
	TWI 6.63	S6_07_12_7 6.49	TWI 6.02	TWI 5.75	TWI 6.37	Slope 7.46
	Win_SRad 6.23	TWI 6.27	Slope 6.00	S24_07_12_4 5.43	S6_07_12_7 5.29	Dist_MajRo 6.59
BEF-CNF	Slope 14.98	Slope 14.40	Slope 14.36	Slope 15.01	Slope 14.09	Slope 14.48
	Win_SRad 9.25	Win_SRad 8.67	Win_SRad 10.05	Win_SRad 9.13	Win_SRad 8.48	Win_SRad 10.20
	Altitude 8.61	Altitude 8.14	Altitude 9.71	Wind 8.24	Wind 8.00	Wind 8.54
	Dist_SecRo 6.65	Wind 7.30	Wind 7.18	Dist_SecRo 7.48	Dist_SecRo 6.74	Liv_Units 7.75
	PopDen 5.76	S24_07_12_7 6.62	Liv_Units 7.01	Dist_MajRo 5.85	Dist_MajRo 5.90	Dist_SecRo 6.60
	Wind 5.69	Dist_SecRo 5.72	Dist_SecRo 6.32	PopDen 5.69	Liv_Units 5.85	Dist_MajRo 6.52
BEF-BDF	Win_SRad 15.94	Win_SRad 18.17	PopDen 11.15	Win_SRad 16.85	Win_SRad 18.81	PopDen 11.97
	Wind 7.48	Liv_Units 8.88	Liv_Units 9.17	S6_07_12_8 8.78	Liv_Units 9.48	Slope 8.55
	Liv_Units 7.42	Slope 6.83	Win_SRad 8.39	Liv_Units 8.68	Slope 7.57	Wind 8.40
	S6_07_12_8 6.96	Wind 5.85	S6_80_02_5 7.43	Wind 7.16	Wind 6.07	Liv_Units 8.01
	S24_07_12_7 4.79	S6_07_12_8 4.87	Dist_UrbA 6.91	S24_07_12_7 5.78	S6_07_12_8 5.33	Win_SRad 7.90
	PopDen 4.72	S24_07_12_4 4.83	Slope 6.76	Dist_UrbA 5.00	S24_07_12_4 4.91	Dist_UrbA 7.54
BDF-CNF	Slope 12.04	Altitude 37.88	Altitude 12.45	Slope 11.65	VRM 12.89	Liv_Units 12.15
	Altitude 11.78	Slope 11.94	Liv_Units 11.44	VRM 9.04	Slope 10.51	Slope 9.24
	VRM 8.55	VRM 7.96	Slope 10.21	Liv_Units 7.21	Wind 10.39	Wind 8.82
	Liv_Units 8.12	Win_SRad 7.13	Dist_UrbA 6.35	Win_SRad 7.01	Dist_SecRo 10.09	PopDen 8.44
	Win_SRad 7.31	PopDen 3.82	Wind 6.32	Dist_MajRo 6.50	Win_SRad 8.06	Dist_UrbA 6.86
	Dist_UrbA 6.30	S6_80_12_8 2.95	PopDen 5.93	Dist_UrbA 6.19	S6_80_12_7 5.96	Win_SRad 6.02
BDF-BEF	Altitude 22.85	Altitude 26.54	Altitude 22.75	Slope 10.51	Liv_Units 16.31	Liv_Units 12.21
	Win_SRad 10.17	Liv_Units 12.03	Liv_Units 8.05	Win_SRad 9.93	Slope 13.03	Wind 9.02
	Liv_Units 6.55	Win_SRad 10.63	Win_SRad 7.17	Liv_Units 9.07	Win_SRad 9.27	Dist_MajRo 7.38
	Slope 5.96	Slope 9.53	Dist_UrbA 6.92	Wind 7.29	PopDen 7.06	Dist_UrbA 6.99
	Wind 5.26	Wind 6.76	Wind 6.73	PopDen 6.16	Wind 6.91	PopDen 6.86
	Dist_MajRo 4.64	Dist_UrbA 4.36	Dist_MajRo 6.44	Dist_MajRo 5.67	Dist_UrbA 5.00	Slope 6.83

Table S8. Six most important variables by vegetation-cover change and period. Ambit NE1

	Including altitude			Not including altitude		
	1987-2012	**2002-2012**	**1987-2002**	**1987-2012**	**2002-2012**	**1987-2002**
SRB-CNF	Wind 10.90	Wind 12.22	Wind 13.15	Wind 12.19	Wind 13.19	Wind 16.17
	Altitude 10.72	Liv_Units 9.38	Liv_Units 8.81	Liv_Units 8.27	Liv_Units 9.93	Liv_Units 9.82
	Liv_Units 7.67	Altitude 8.25	Altitude 8.28	Dist_MajRo 7.37	PopDen 9.49	Dist_MajRo 7.87
	Dist_MajRo 7.03	PopDen 7.52	Win_SRad 7.31	PopDen 7.04	Dist_MajRo 7.81	Win_SRad 7.63
	Win_SRad 6.89	Dist_MajRo 6.83	Dist_MajRo 6.63	Win_SRad 6.54	Slope 6.98	PopDen 7.34
	PopDen 6.73	Slope 6.28	PopDen 6.53	Dist_UrbA 6.00	Dist_UrbA 6.11	Slope 6.36
SRB-BEF	Altitude 32.66	Altitude 35.34	Altitude 55.54	TWI 8.82	Wind 9.70	S6_80_02_5 17.11
	Slope 8.30	Slope 7.53	Win_SRad 13.13	S6_80_12_5 8.59	TWI 8.36	Win_SRad 11.58
	Win_SRad 8.03	TWI 6.22	Dist_MajRo 2.96	VRM 7.30	PopDen 7.64	PopDen 9.36
	Dist_MajRo 7.28	Wind 6.05	Wind 2.92	Slope 7.21	VRM 7.45	VRM 7.25
	TWI 4.91	Win_SRad 5.88	Slope 2.85	Wind 6.95	Dist_MajRo 7.21	Wind 6.20
	Wind 4.70	Dist_MajRo 5.75	Dist_SecRo 2.85	Dist_MajRo 6.91	Slope 6.87	Liv_Units 4.19
SRB-BDF	Altitude 18.76	Altitude 21.61	Altitude 21.09	Win_SRad 14.61	PopDen 13.32	Wind 16.94
	Win_SRad 14.68	TWI 10.67	Wind 11.53	Wind 12.87	TWI 11.06	PopDen 11.18
	TWI 10.92	Win_SRad 9.28	PopDen 8.83	TWI 12.72	Wind 10.83	Dist_MajRo 8.27
	Wind 9.51	PopDen 7.78	Win_SRad 8.29	PopDen 10.55	Win_SRad 10.05	TWI 7.81
	PopDen 6.38	Wind 6.36	Dist_MajRo 6.97	Dist_SecRo 4.86	Liv_Units 5.63	Win_SRad 7.78
	Dist_SecRo 5.26	Dist_MajRo 4.75	Dist_SecRo 6.37	S6_07_12_7 4.27	Dist_UrbA 4.78	Liv_Units 5.44
CNF-BEF	Altitude 18.67	Altitude 16.43	Altitude 38.16	Wind 11.66	Wind 9.47	Wind 12.32
	Wind 12.42	Wind 10.23	Wind 9.01	PopDen 8.38	Dist_MajRo 9.16	Dist_MajRo 7.64
	Win_SRad 7.33	Dist_MajRo 7.40	Dist_MajRo 5.47	Dist_MajRo 7.30	PopDen 7.84	Liv_Units 6.21
	Dist_UrbA 6.11	Win_SRad 6.67	Win_SRad 4.86	Dist_UrbA 7.16	Win_SRad 6.94	Dist_UrbA 5.81
	Dist_MajRo 5.71	PopDen 6.40	Slope 4.34	Win_SRad 6.73	Liv_Units 6.64	S6_87_02_5 5.77
	PopDen 5.19	Dist_UrbA 6.02	TWI 3.96	Liv_Units 5.73	Dist_UrbA 6.52	S6_80_02_4 5.72
CNF-BDF	Slope 17.05	Win_SRad 13.41	Altitude 30.05	Slope 19.00	Slope 13.67	Lithology 14.98
	Win_SRad 14.69	Slope 13.36	Lithology 17.65	Win_SRad 14.19	Win_SRad 13.48	Slope 12.37
	Altitude 13.34	Altitude 10.74	Slope 8.97	PopDen 10.07	PopDen 8.12	Wind 10.28
	Lithology 8.36	PopDen 7.29	Win_SRad 6.84	Lithology 9.14	Dist_MajRo 6.19	PopDen 10.21
	PopDen 7.33	Dist_MajRo 5.90	PopDen 5.20	TWI 6.38	Lithology 6.17	Win_SRad 6.85
	TWI 6.26	TWI 5.31	Wind 4.42	Wind 5.82	TWI 6.12	Dist_UrbA 6.15
BEF-CNF	Wind 11.83	Altitude 15.38	S24_97_02_8 14.54	Wind 12.32	Wind 11.22	S24_97_02_8 15.39
	Altitude 8.79	Wind 10.60	Slope 12.33	Dist_MajRo 9.25	Liv_Units 9.71	Slope 12.84
	Dist_MajRo 8.37	Slope 7.83	S6_92_02_4 10.26	Slope 8.67	Dist_MajRo 8.00	S6_92_02_4 12.12
	Slope 7.16	Dist_MajRo 7.61	Wind 9.89	Lithology 7.47	Slope 7.34	Wind 11.53
	Liv_Units 6.94	Liv_Units 6.99	Altitude 6.26	Liv_Units 7.43	Win_SRad 5.90	PopDen 4.31
	Win_SRad 6.78	Win_SRad 5.64	Liv_Units 4.22	Win_SRad 6.49	PopDen 5.71	Liv_Units 3.99
BEF-BDF	Win_SRad 24.42	Win_SRad 18.44	Wind 13.80	Win_SRad 26.23	Win_SRad 20.57	Wind 15.39
	Altitude 10.39	Altitude 14.32	Win_SRad 9.81	Slope 7.46	Slope 15.21	Win_SRad 11.24
	Wind 6.79	Slope 13.54	Altitude 9.43	Wind 6.83	Wind 6.92	Dist_MajRo 6.78
	Slope 6.40	Wind 6.46	Dist_MajRo 6.27	TWI 5.01	TWI 5.24	Liv_Units 5.69
	Lithology 5.26	Dist_MajRo 4.57	Slope 5.87	Dist_MajRo 4.64	Dist_MajRo 5.24	Slope 5.51
	TWI 4.26	Dist_UrbA 4.31	Liv_Units 5.62	Lithology 4.43	VRM 5.00	Lithology 5.16
BDF-CNF	Altitude 14.04		Dist_MajRo 20.70	Slope 11.05		Dist_MajRo 23.27
	Slope 13.53		Altitude 9.86	Wind 10.10		Slope 10.19
	Liv_Units 8.79		Slope 9.31	Liv_Units 9.78		Dist_UrbA 8.62
	Wind 8.51		Dist_UrbA 7.48	Dist_UrbA 8.77		Wind 6.28
	Dist_UrbA 7.04		Win_SRad 5.85	Dist_MajRo 6.28		Dist_SecRo 6.23
	Dist_MajRo 6.11		Wind 5.67	Win_SRad 6.04		Win_SRad 5.88
BDF-BEF	Win_SRad 16.71	Win_SRad 14.99	Altitude 12.16	Win_SRad 17.13	Win_SRad 16.18	Wind 10.03
	Wind 9.39	Altitude 10.59	Wind 8.86	Slope 9.20	Slope 10.72	Slope 8.91
	Altitude 9.32	Slope 9.96	Slope 7.54	Dist_MajRo 8.49	Wind 8.29	TWI 8.09
	Slope 8.96	Wind 7.34	TWI 7.47	Dist_UrbA 8.16	Liv_Units 7.12	Dist_MajRo 7.94
	Liv_Units 7.32	Dist_SecRo 6.37	Dist_SecRo 7.00	Wind 8.14	TWI 6.82	Dist_SecRo 7.41
	Dist_MajRo 6.65	TWI 6.17	Dist_MajRo 6.81	Liv_Units 6.89	Dist_MajRo 6.75	PopDen 7.40

Table S9. Six most important variables by vegetation-cover change and period. Ambit NE2

	Including altitude			Not including altitude		
	1987-2012	**2002-2012**	**1987-2002**	**1987-2012**	**2002-2012**	**1987-2002**
SRB-CNF	Altitude 19.70	Altitude 22.52	Altitude 19.36	Wind 8.95	Slope 12.28	Liv_Units 11.08
	Slope 7.54	Slope 12.53	Liv_Units 8.86	Slope 8.11	Dist_MajRo 8.59	Wind 8.93
	Wind 6.91	Dist_MajRo 6.92	PopDen 7.12	Dist_MajRo 7.75	Wind 7.93	PopDen 8.74
	Liv_Units 6.79	Liv_Units 5.70	Wind 7.09	Liv_Units 7.27	Liv_Units 7.58	Dist_MajRo 8.49
	Dist_MajRo 6.12	Wind 5.18	Slope 7.01	Win_SRad 7.20	Dist_SecRo 7.02	Dist_SecRo 7.54
	Win_SRad 5.89	Dist_SecRo 4.35	Dist_MajRo 6.87	Dist_SecRo 6.80	PopDen 6.26	Slope 6.77
SRB-BEF	Altitude 29.79	Altitude 26.59	Altitude 23.76	Dist_MajRo 8.47	Dist_MajRo 10.24	Win_SRad 10.36
	Win_SRad 8.42	Dist_MajRo 7.68	Win_SRad 9.05	Win_SRad 8.07	Dist_UrbA 9.28	Lithology 8.74
	Dist_MajRo 7.26	PopDen 6.59	PopDen 7.62	Dist_SecRo 7.27	PopDen 7.53	PopDen 8.27
	PopDen 6.03	TWI 6.18	Dist_MajRo 6.40	PopDen 7.00	TWI 7.43	Dist_MajRo 7.72
	TWI 5.25	Dist_UrbA 6.16	Wind 5.91	Wind 6.08	Wind 7.33	Dist_SecRo 7.44
	Slope 4.97	Slope 5.47	Liv_Units 5.43	Slope 6.06	Dist_SecRo 6.42	Wind 6.85
SRB-BDF	Altitude 24.48	Altitude 20.22	Altitude 22.89	TWI 12.83	TWI 19.13	TWI 10.49
	TWI 10.67	TWI 17.93	S6_80_02_5 11.77	Wind 8.59	Dist_MajRo 8.37	Dist_MajRo 8.49
	S6_07_12_4 8.47	S6_07_12_5 12.43	TWI 7.50	Dist_MajRo 8.10	PopDen 6.95	Wind 7.90
	Win_SRad 7.43	Pl_Curv 6.60	Liv_Units 6.35	Win_SRad 7.84	Wind 6.76	Dist_SecRo 7.21
	Wind 5.52	Dist_UrbA 6.37	Win_SRad 6.14	S6_07_12_4 6.60	Pl_Curv 6.45	Win_SRad 7.05
	Dist_MajRo 5.50	PopDen 5.06	Dist_UrbA 5.53	Dist_UrbA 5.86	Lithology 5.19	Lithology 6.57
CNF-BEF	Altitude 15.87	Altitude 14.17	Altitude 16.35	Dist_UrbA 10.38	Wind 12.47	Dist_UrbA 9.82
	Dist_MajRo 9.25	Dist_MajRo 10.16	Dist_MajRo 9.17	Dist_MajRo 10.10	Dist_MajRo 9.58	Dist_MajRo 9.10
	Dist_UrbA 7.78	Wind 10.14	Dist_UrbA 7.38	Wind 8.34	Dist_UrbA 7.47	Liv_Units 9.04
	Wind 6.35	PopDen 7.30	Wind 6.56	Liv_Units 7.46	PopDen 7.23	Wind 8.98
	Win_SRad 6.21	Dist_UrbA 7.13	Win_SRad 6.01	Dist_SecRo 6.43	Win_SRad 7.21	PopDen 6.59
	Slope 5.40	Win_SRad 7.00	Liv_Units 5.36	VRM 6.19	Dist_SecRo 6.98	Slope 6.44
BEF-CNF	Slope 9.85	Slope 13.73	Altitude 14.50	Slope 10.96	Slope 12.96	Slope 10.08
	Altitude 9.55	Liv_Units 9.72	Slope 9.24	Liv_Units 10.93	Dist_MajRo 9.09	Dist_MajRo 9.91
	Liv_Units 8.89	Altitude 8.50	Liv_Units 8.05	Dist_MajRo 9.54	Liv_Units 8.78	Liv_Units 9.20
	Dist_MajRo 8.87	Dist_MajRo 7.97	Dist_MajRo 7.58	Dist_UrbA 6.59	Dist_SecRo 7.67	Dist_UrbA 7.35
	Win_SRad 6.35	Dist_UrbA 6.96	Dist_UrbA 6.31	Dist_SecRo 6.20	Dist_UrbA 7.51	Wind 7.31
	Dist_UrbA 6.32	PopDen 6.39	PopDen 6.30	Win_SRad 6.15	PopDen 6.54	PopDen 6.24
BEF-BDF	Lithology 23.38	Dist_UrbA 18.09	Lithology 33.84	Lithology 25.10	Dist_UrbA 20.51	Lithology 29.69
	TWI 10.40	Lithology 16.41	Altitude 13.29	Dist_UrbA 11.05	Lithology 15.39	Wind 7.96
	Dist_UrbA 9.93	TWI 11.75	Wind 6.33	TWI 9.51	TWI 11.75	Dist_MajRo 7.53
	Slope 6.70	Altitude 7.90	TWI 6.04	Dist_SecRo 6.99	Slope 7.34	Liv_Units 7.37
	Altitude 6.52	Slope 7.75	Dist_MajRo 5.67	Slope 6.34	Wind 5.83	TWI 7.08
	Dist_SecRo 5.68	Liv_Units 5.60	Dist_UrbA 5.40	Dist_MajRo 5.89	Dist_MajRo 5.49	Dist_UrbA 5.37
BDF-BEF	Altitude 34.51	Altitude 27.22	Altitude 35.49	Dist_SecRo 8.80	Sum_SRad 12.35	Dist_MajRo 15.23
	TWI 7.52	Sum_SRad 10.35	Wind 8.82	TWI 8.45	Dist_UrbA 12.33	VRM10.03
	Wind 6.93	Dist_SecRo 8.54	Dist_MajRo 7.67	Dist_MajRo 8.41	Liv_Units 9.06	Dist_UrbA 9.62
	Dist_MajRo 6.23	TWI 7.69	PopDen 5.96	Wind 8.36	Dist_MajRo 7.86	Wind 8.79
	Sum_SRad 5.31	Dist_MajRo 6.19	VRM 4.99	PopDen 7.76	TWI 7.72	PopDen 8.38
	Pf_Curv 4.73	Dist_UrbA 5.84	Win_SRad 4.97	Dist_UrbA 7.59	Dist_SecRo 7.12	TWI 6.02

Table S10. Six most important variables by vegetation-cover change and period. Ambit SE3

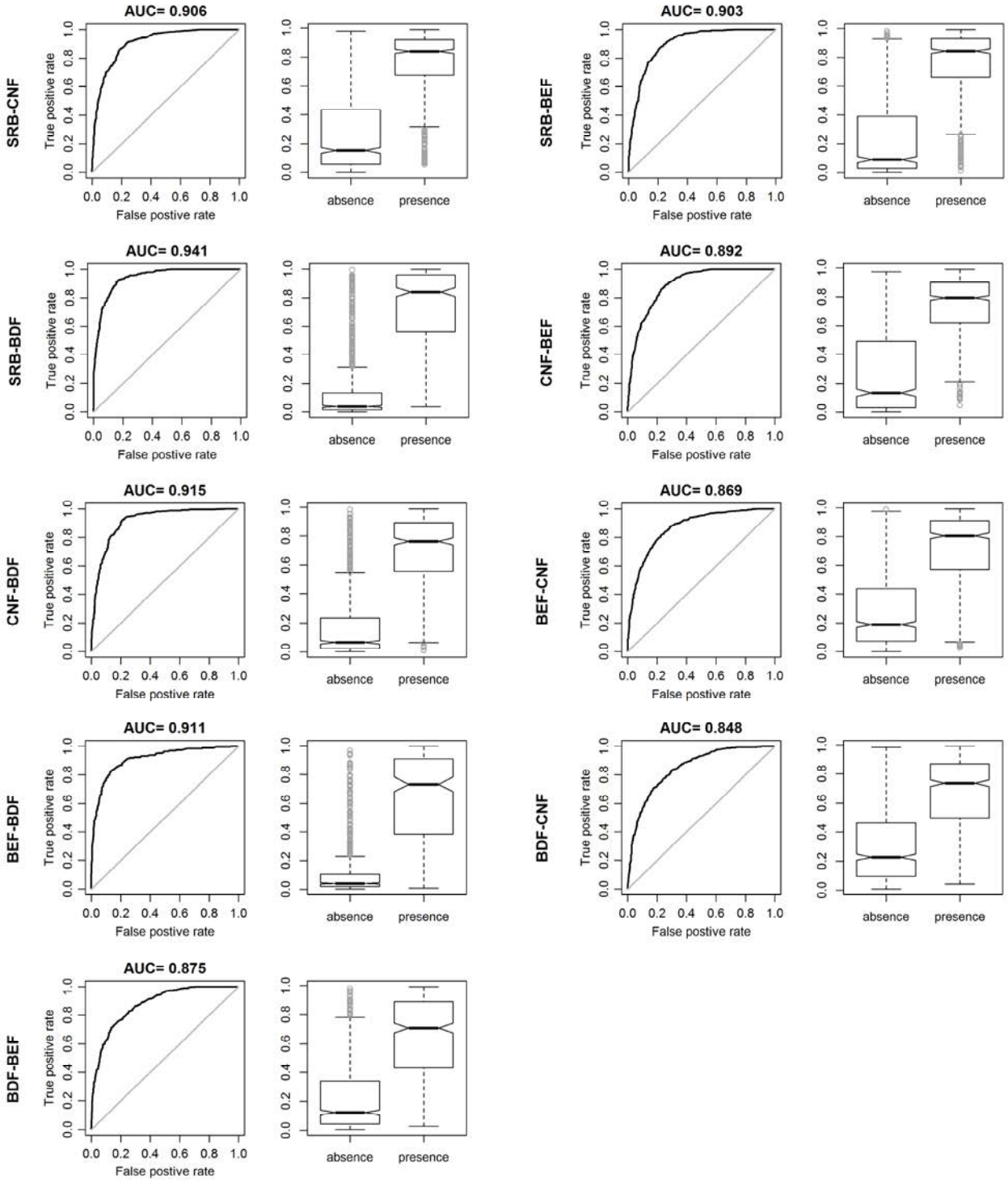


Figure S1. ROC curves and distribution of probability thresholds for presence and absence events. Ambit NE1, period 1987-2002, including altitude.

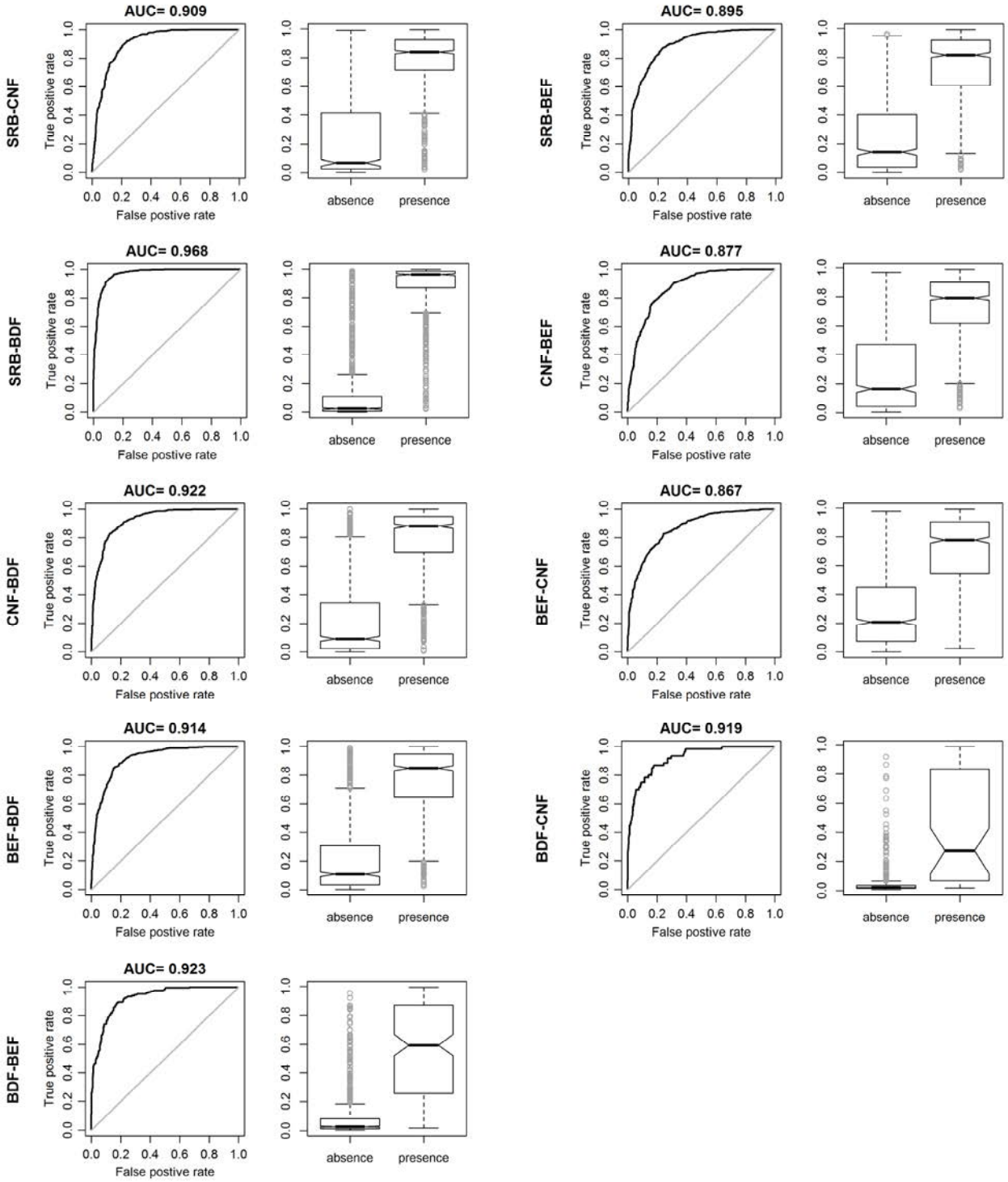


Figure S2. ROC curves and distribution of probability thresholds for presence and absence events. Ambit NE1, period 2002-2012, including altitude.

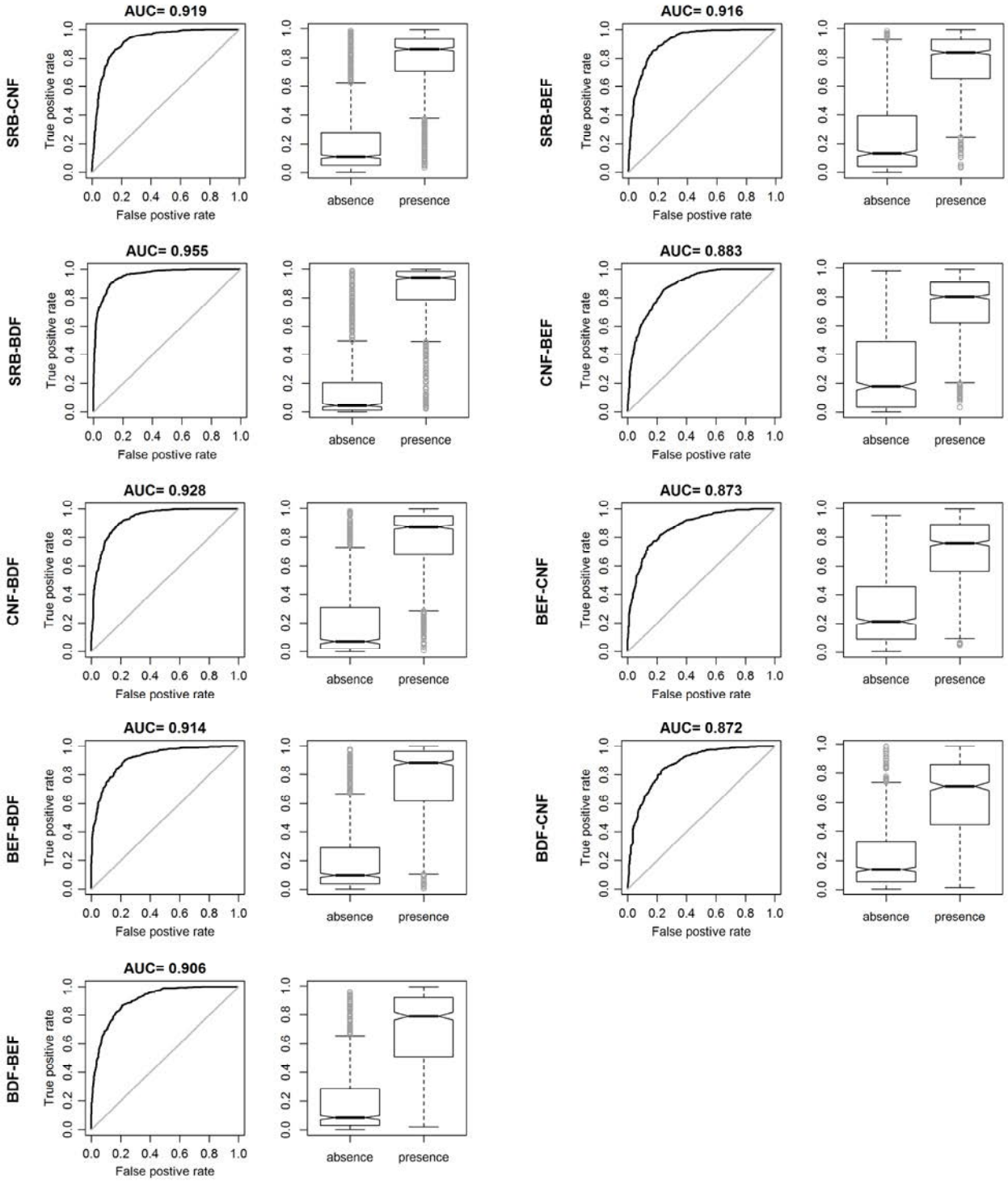


Figure S3. ROC curves and distribution of probability thresholds for presence and absence events. Ambit NE1, period 1987-2012, including altitude.

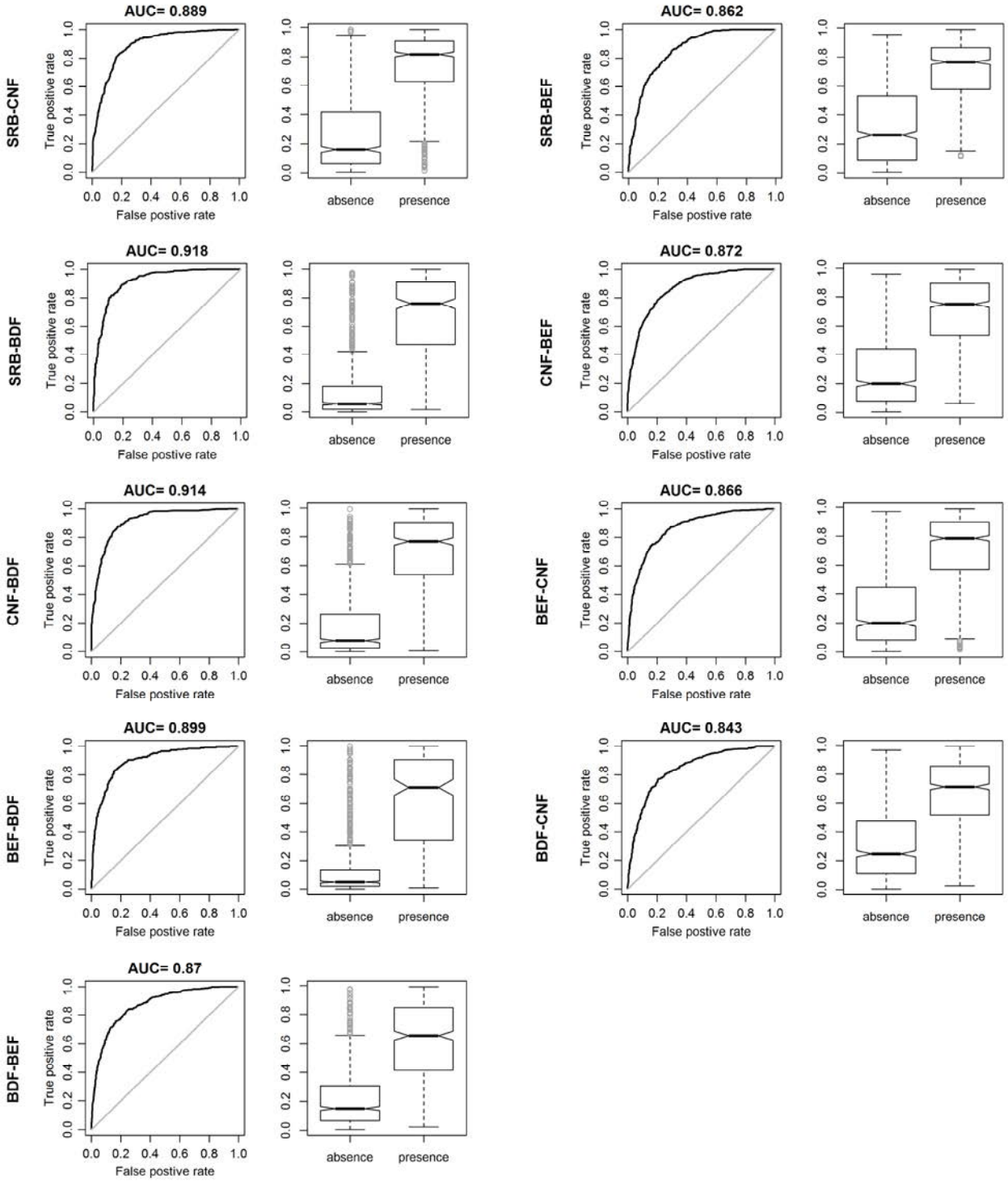


Figure S4. ROC curves and distribution of probability thresholds for presence and absence events. Ambit NE1, period 1987-2002, not including altitude.

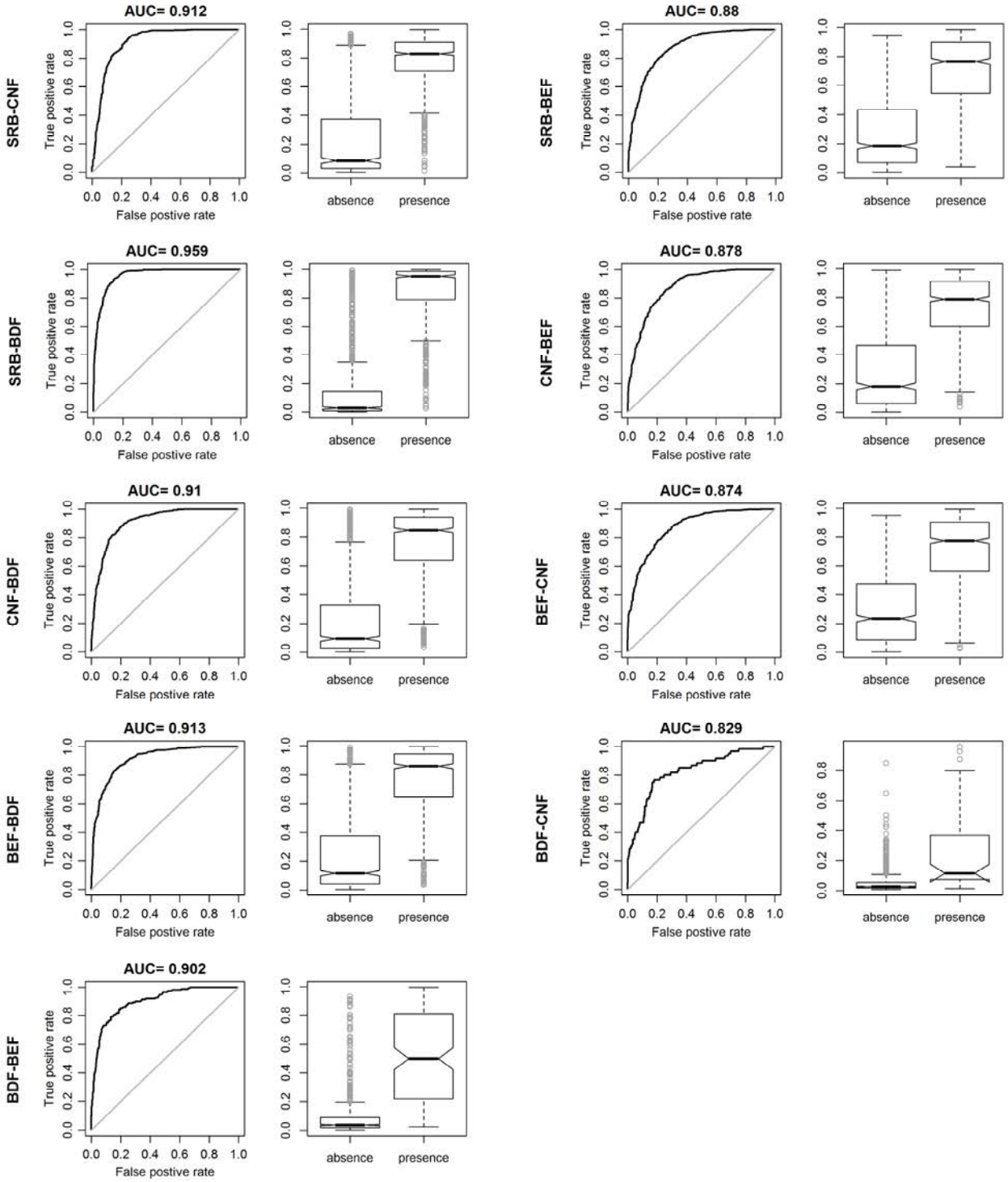


Figure S5. ROC curves and distribution of probability thresholds for presence and absence events. Ambit NE1, period 2002-2012, not including altitude.

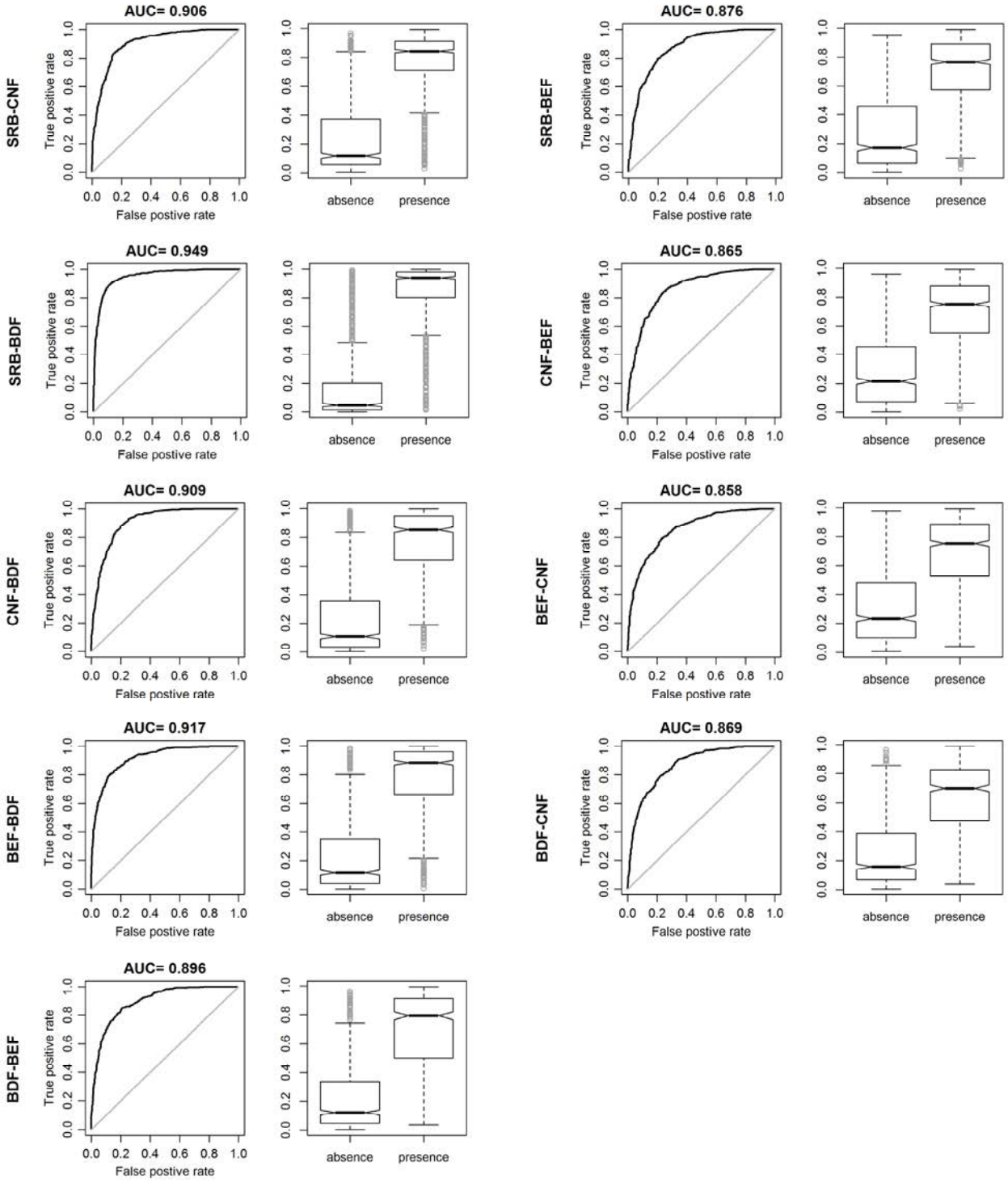


Figure S6. ROC curves and distribution of probability thresholds for presence and absence events. Ambit NE1, period 1987-2012, not including altitude.

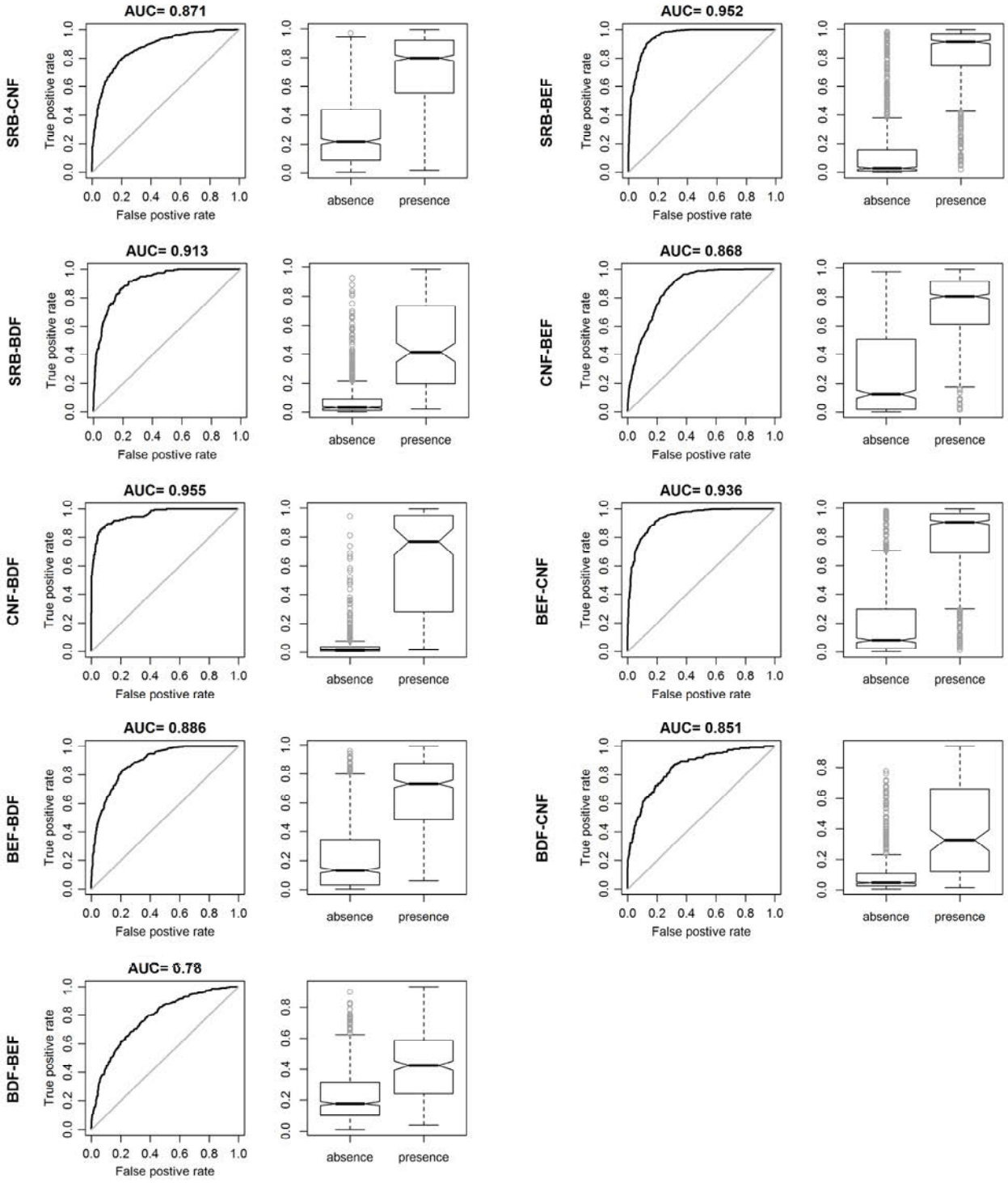


Figure S7. ROC curves and distribution of probability thresholds for presence and absence events. Ambit NE2, period 1987-2002, including altitude.

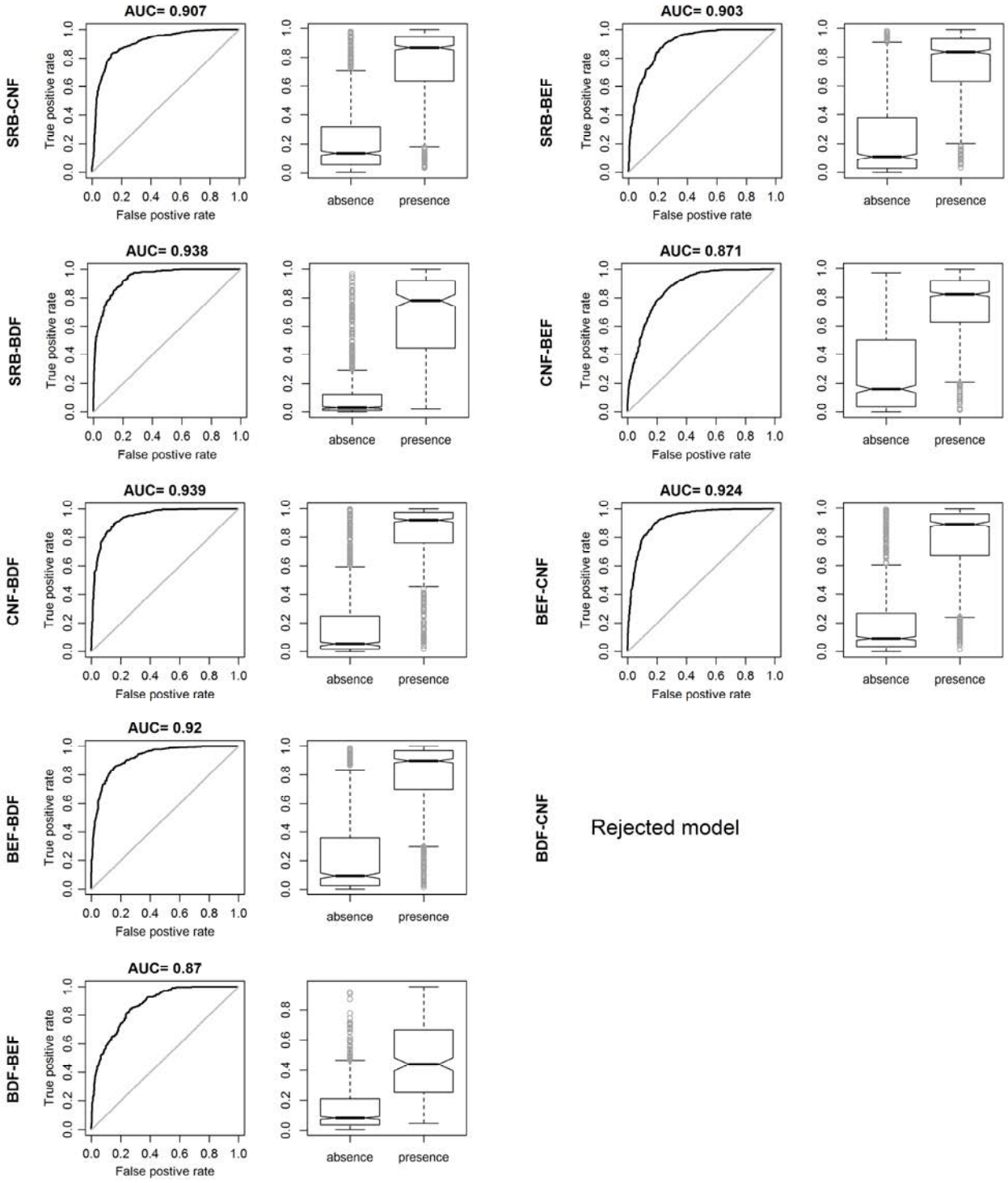


Figure S8. ROC curves and distribution of probability thresholds for presence and absence events. Ambit NE2, period 2002-2012, including altitude, including altitude.

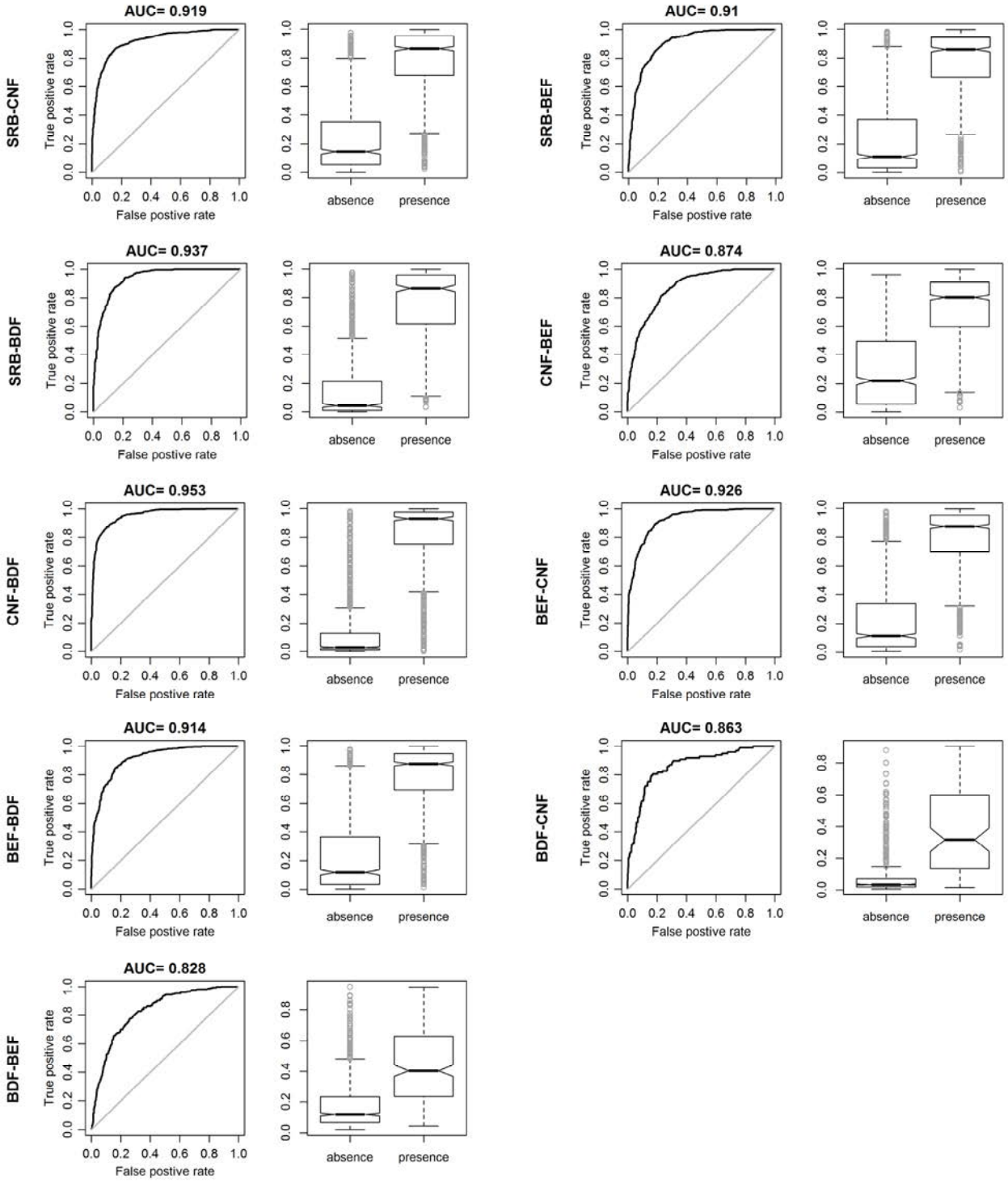


Figure S9. ROC curves and distribution of probability thresholds for presence and absence events. Ambit NE2, period 1987-2012, including altitude.

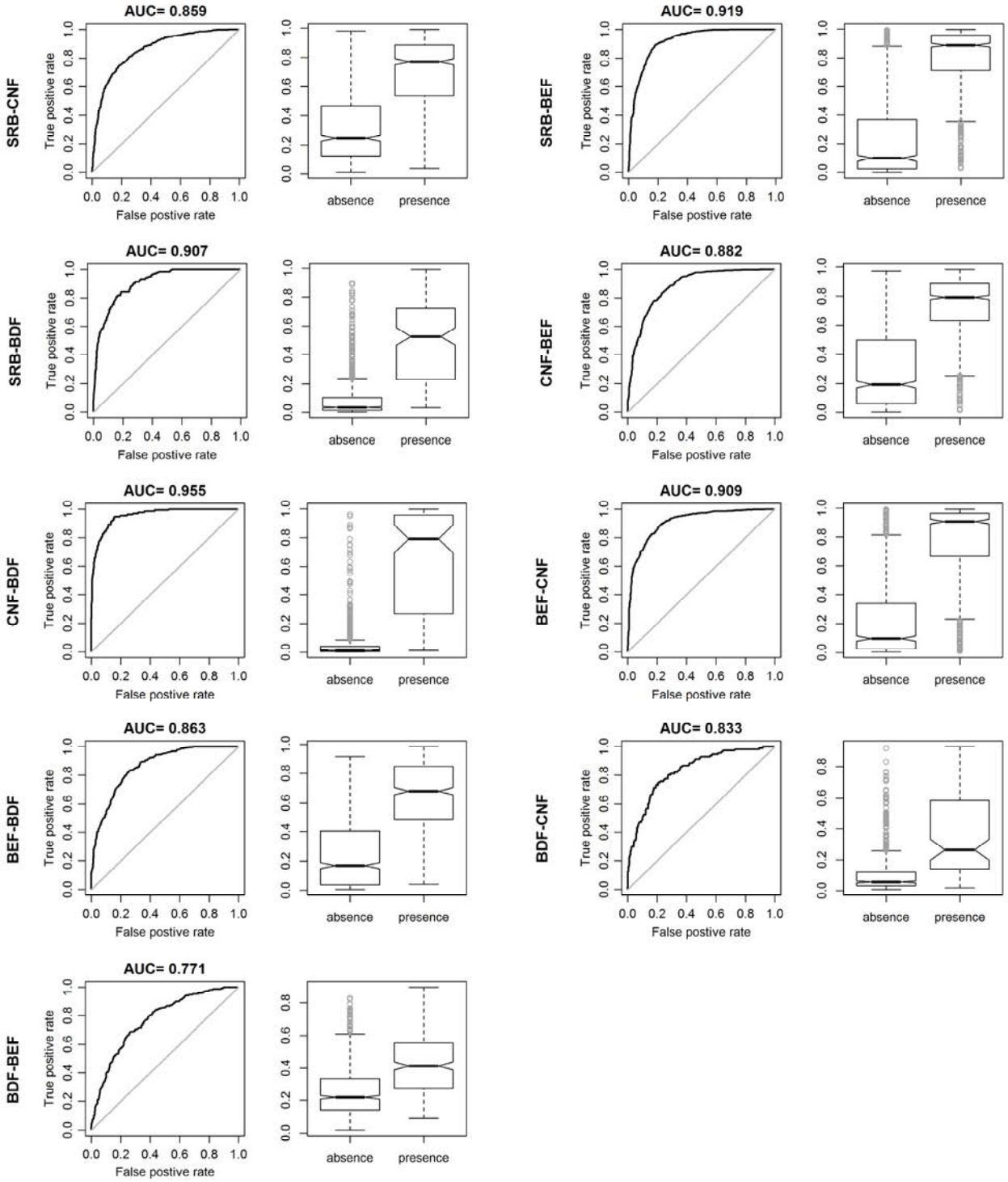


Figure S10. ROC curves and distribution of probability thresholds for presence and absence events. Ambit NE2, period 1987-2002, not including altitude.

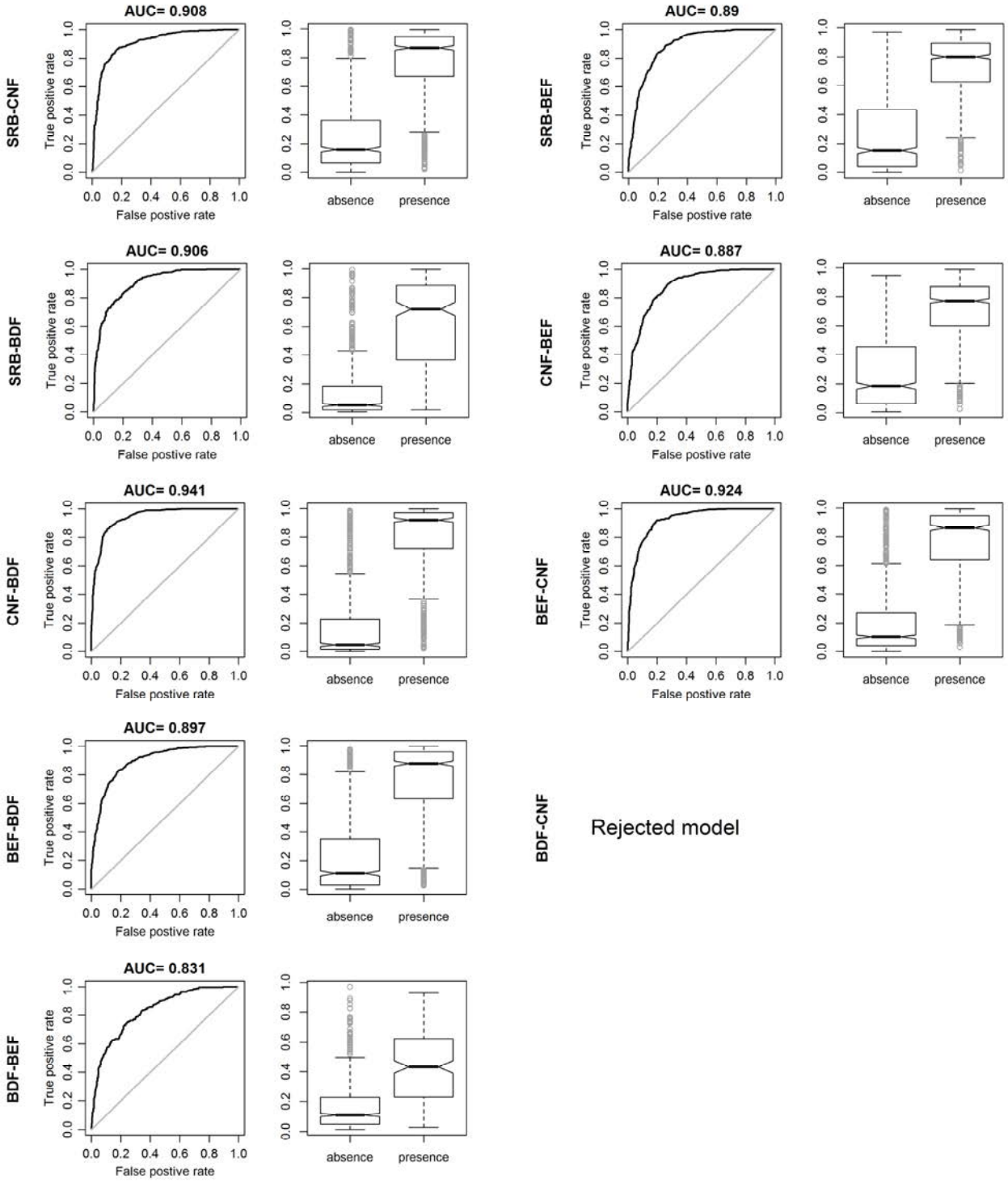


Figure S11. ROC curves and distribution of probability thresholds for presence and absence events. Ambit NE2, period 2002-2012, not including altitude.

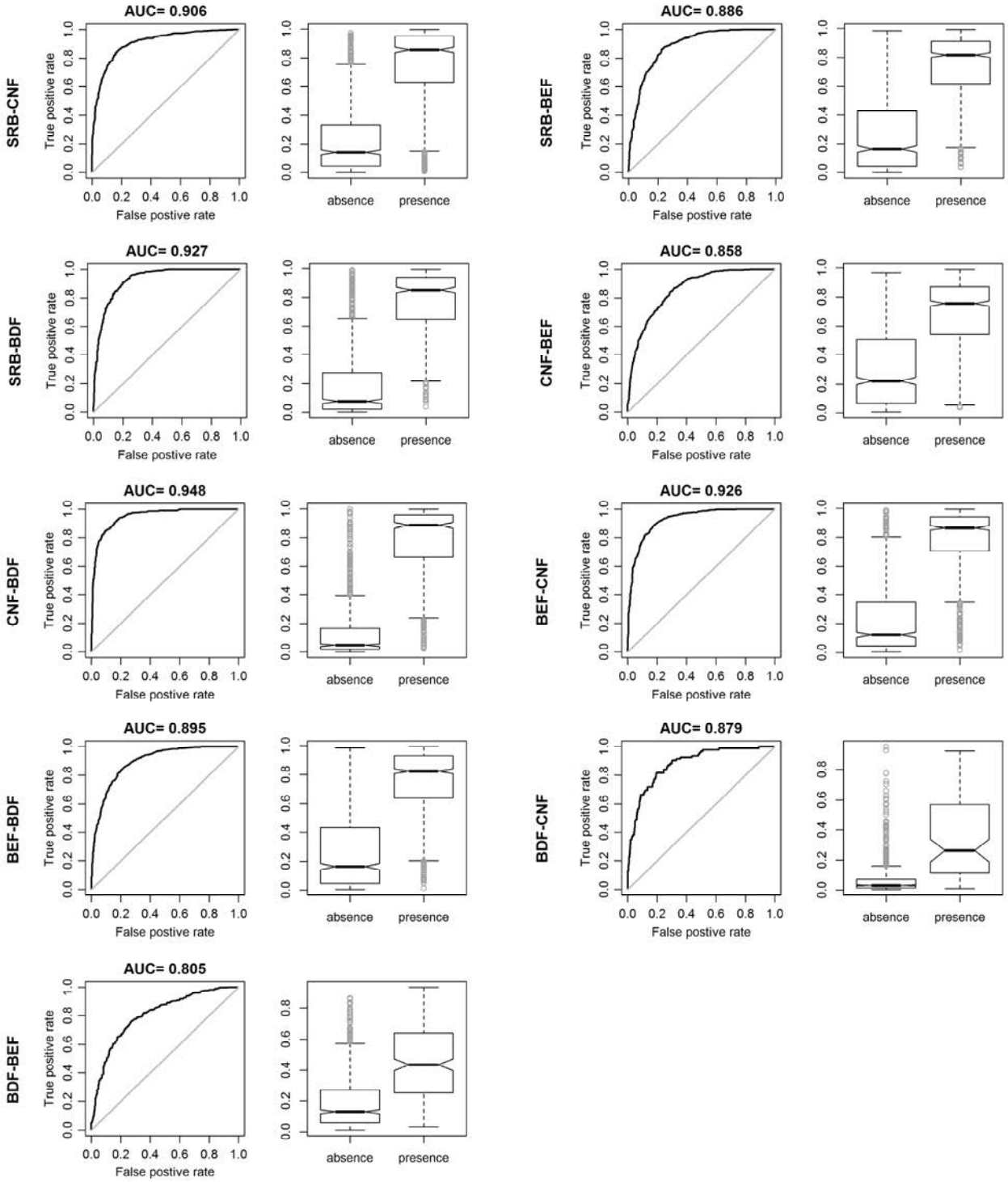


Figure S12. ROC curves and distribution of probability thresholds for presence and absence events. Ambit NE2, period 1987-2012, not including altitude.

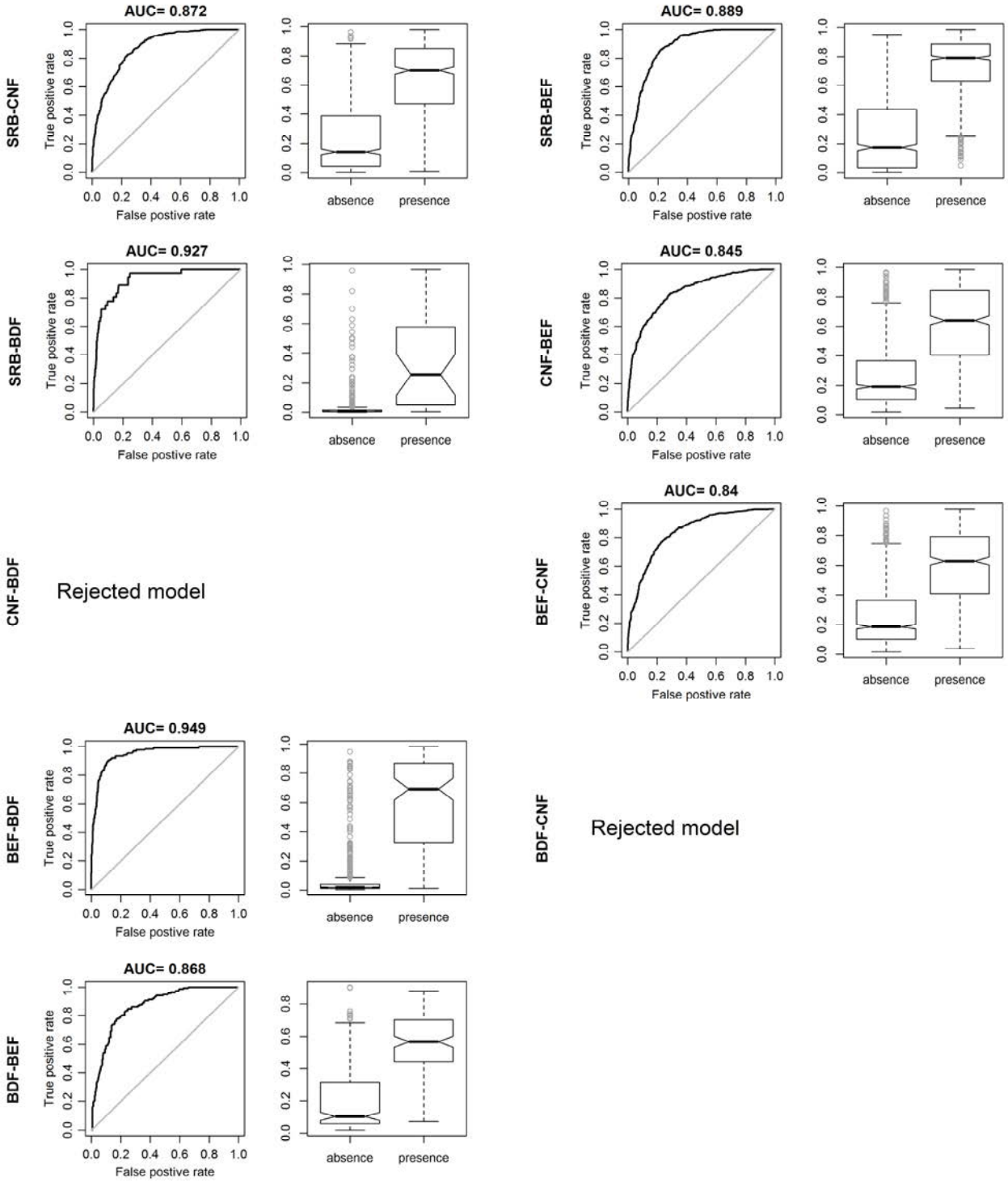


Figure S13. ROC curves and distribution of probability thresholds for presence and absence events. Ambit SE3, period 1987-2002, including altitude.

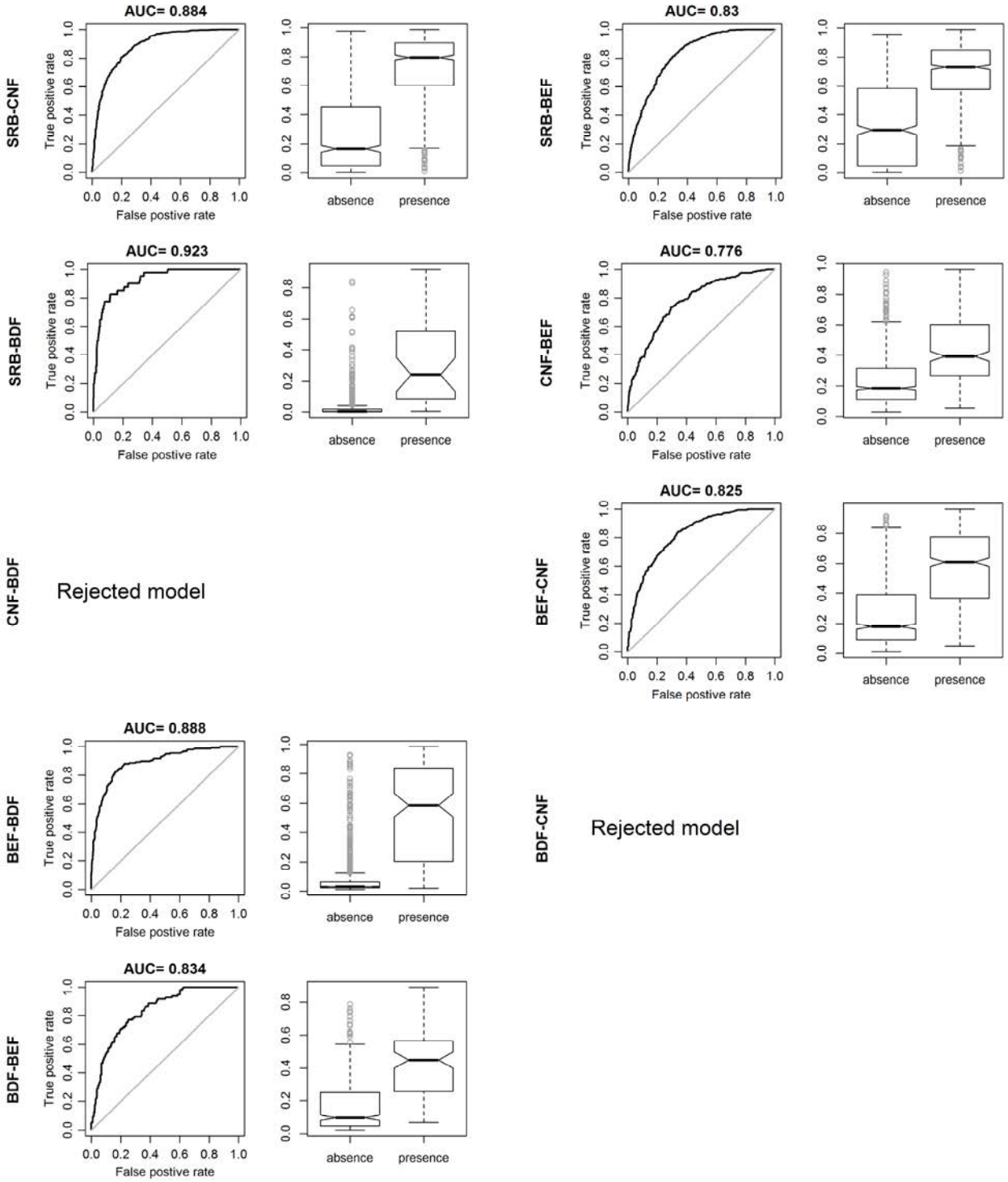


Figure S14. ROC curves and distribution of probability thresholds for presence and absence events. Ambit SE3, period 2002-2012, including altitude.

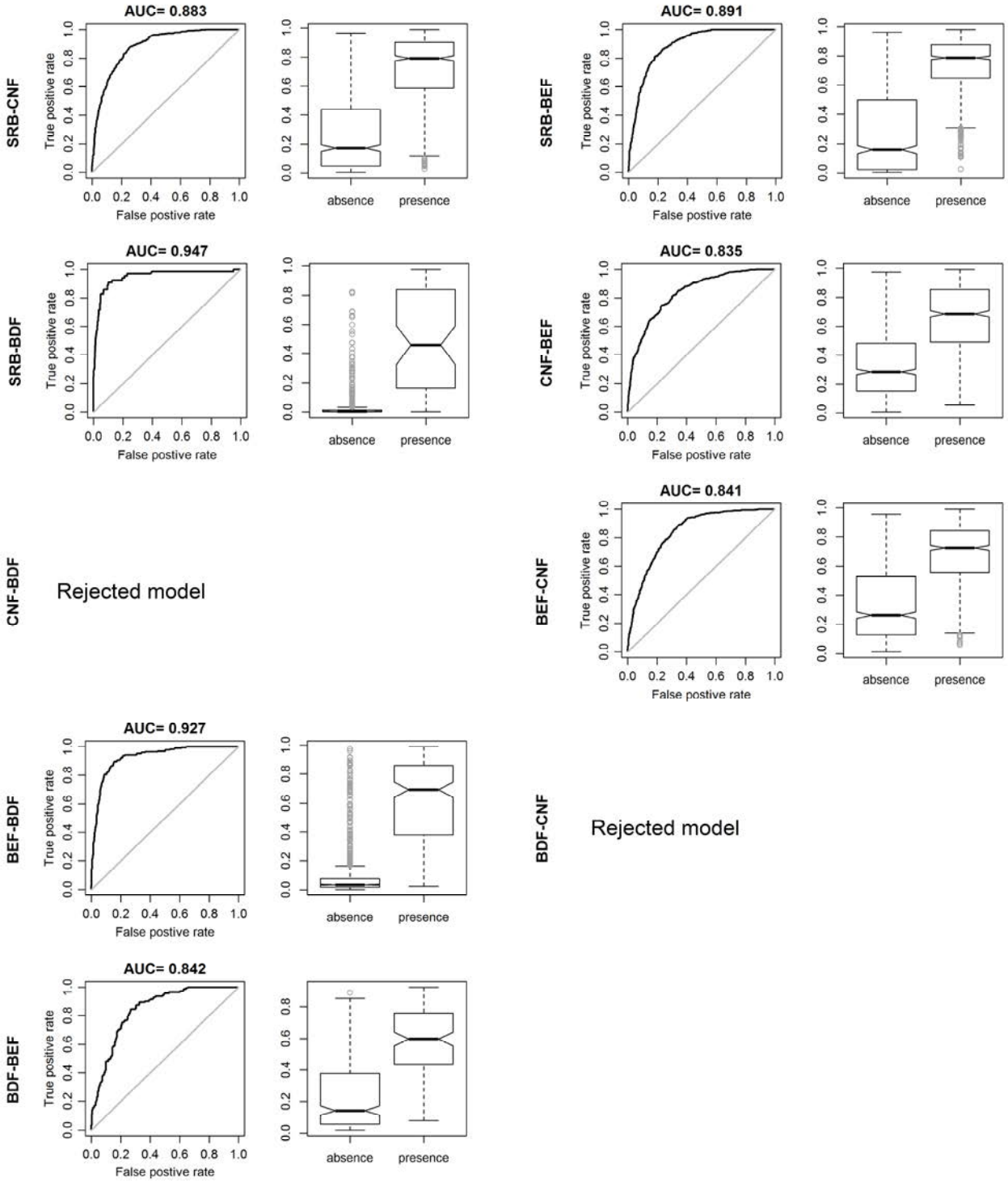


Figure S15. ROC curves and distribution of probability thresholds for presence and absence events. Ambit SE3, period 1987-2012, including altitude.

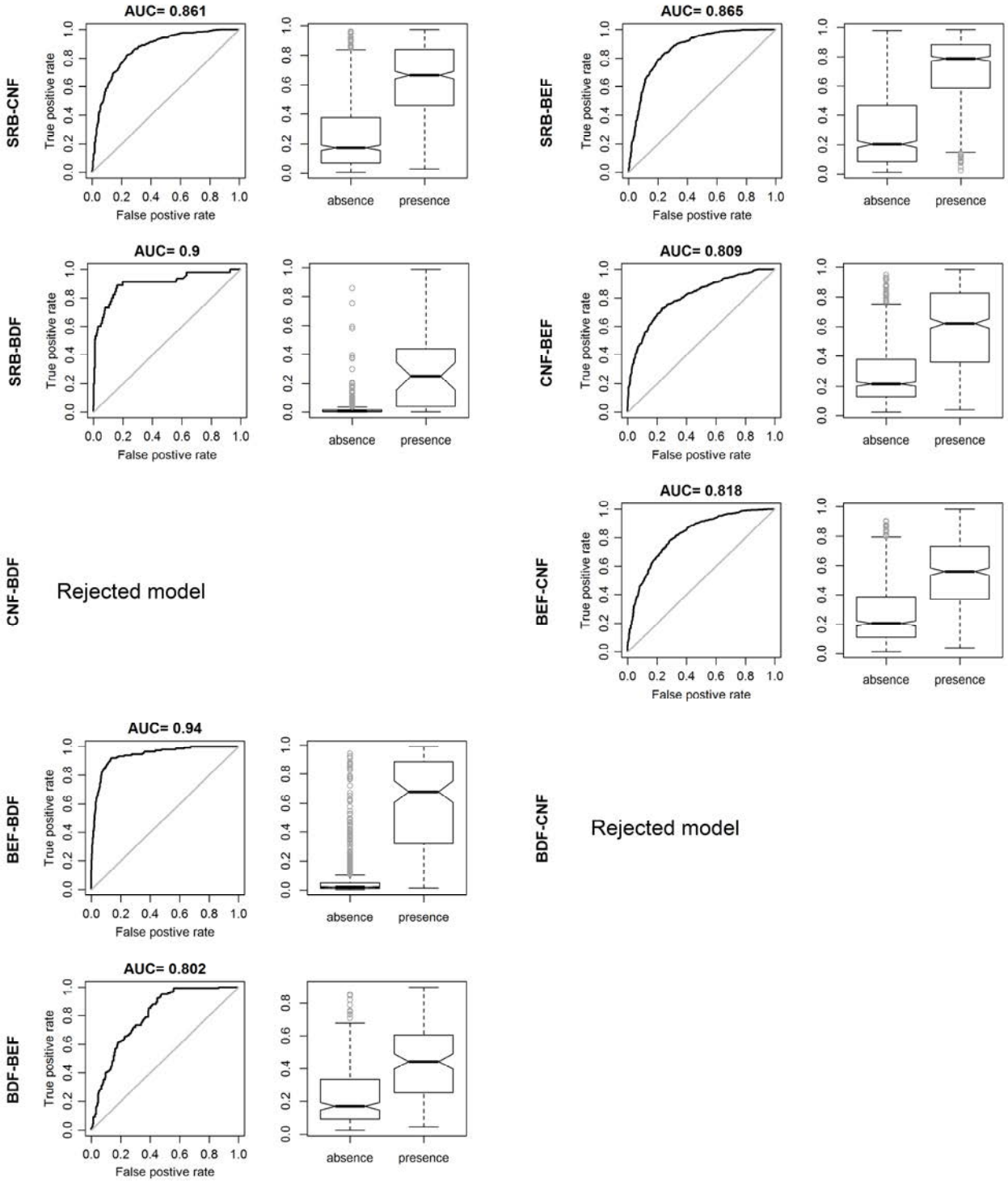


Figure S16. ROC curves and distribution of probability thresholds for presence and absence events. Ambit SE3, period 1987-2002, not including altitude.

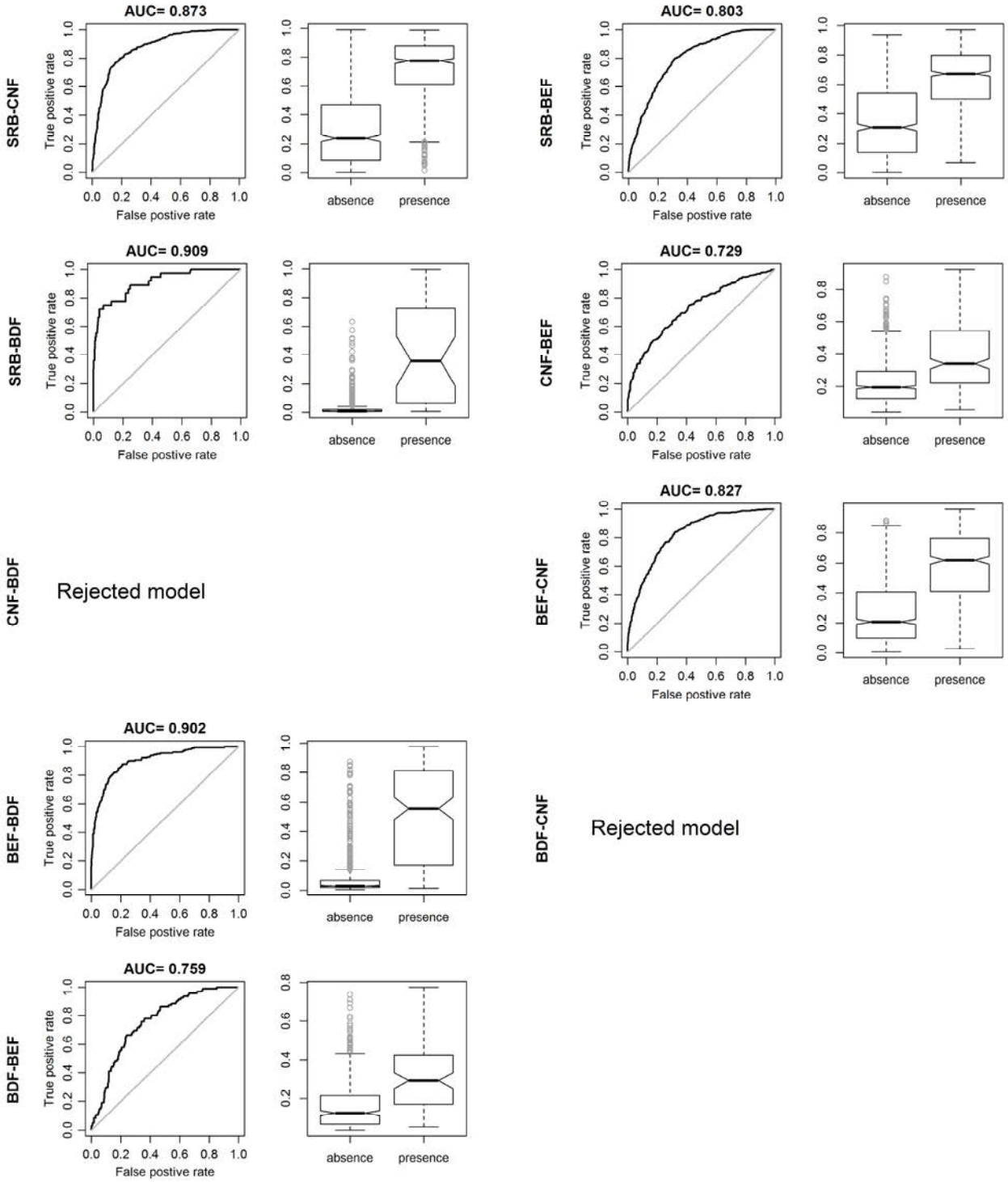


Figure S17. ROC curves and distribution of probability thresholds for presence and absence events. Ambit SE3, period 2002-2012, not including altitude.

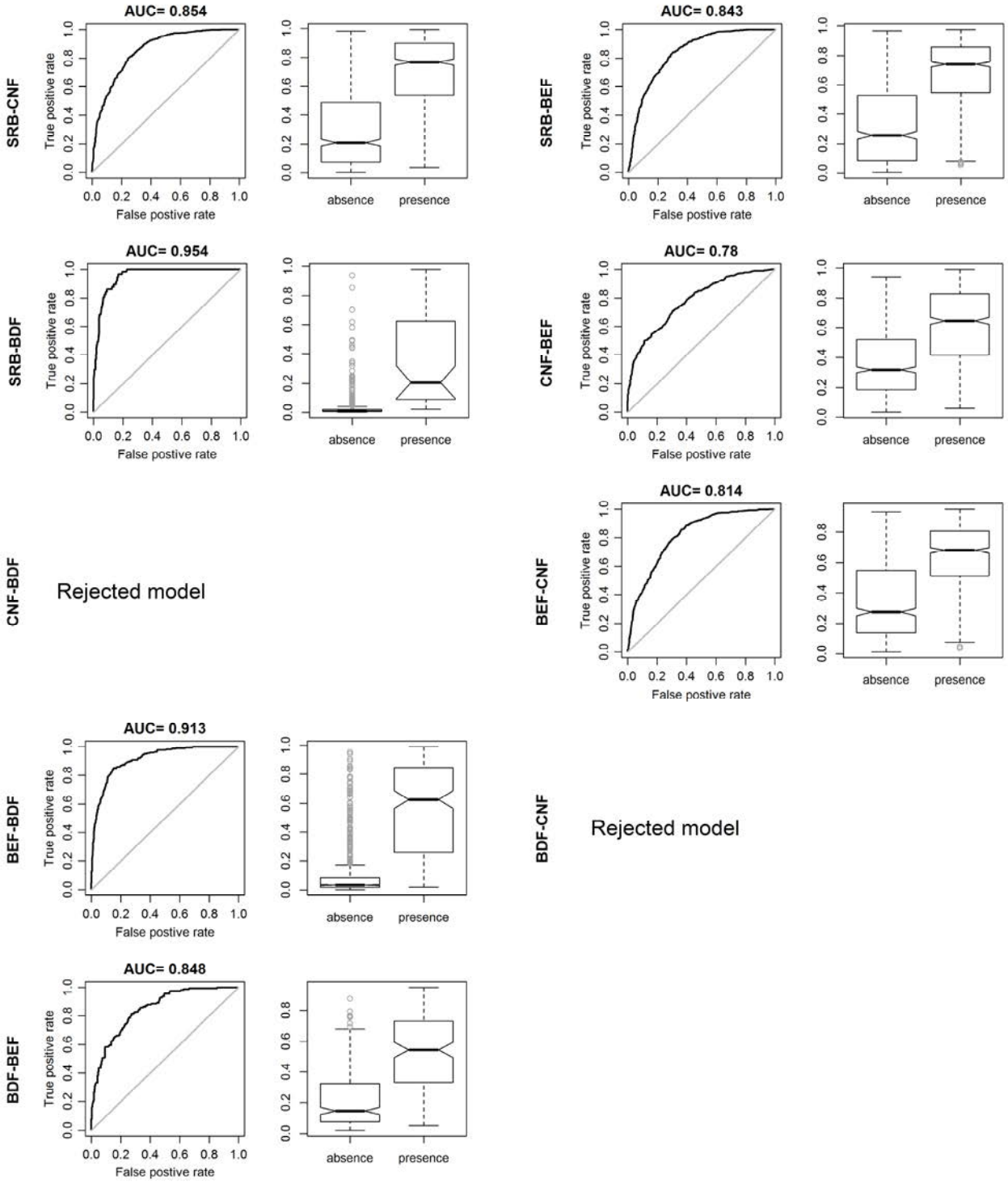


Figure S18. ROC curves and distribution of probability thresholds for presence and absence events. Ambit SE3, period 1987-2012, not including altitude.

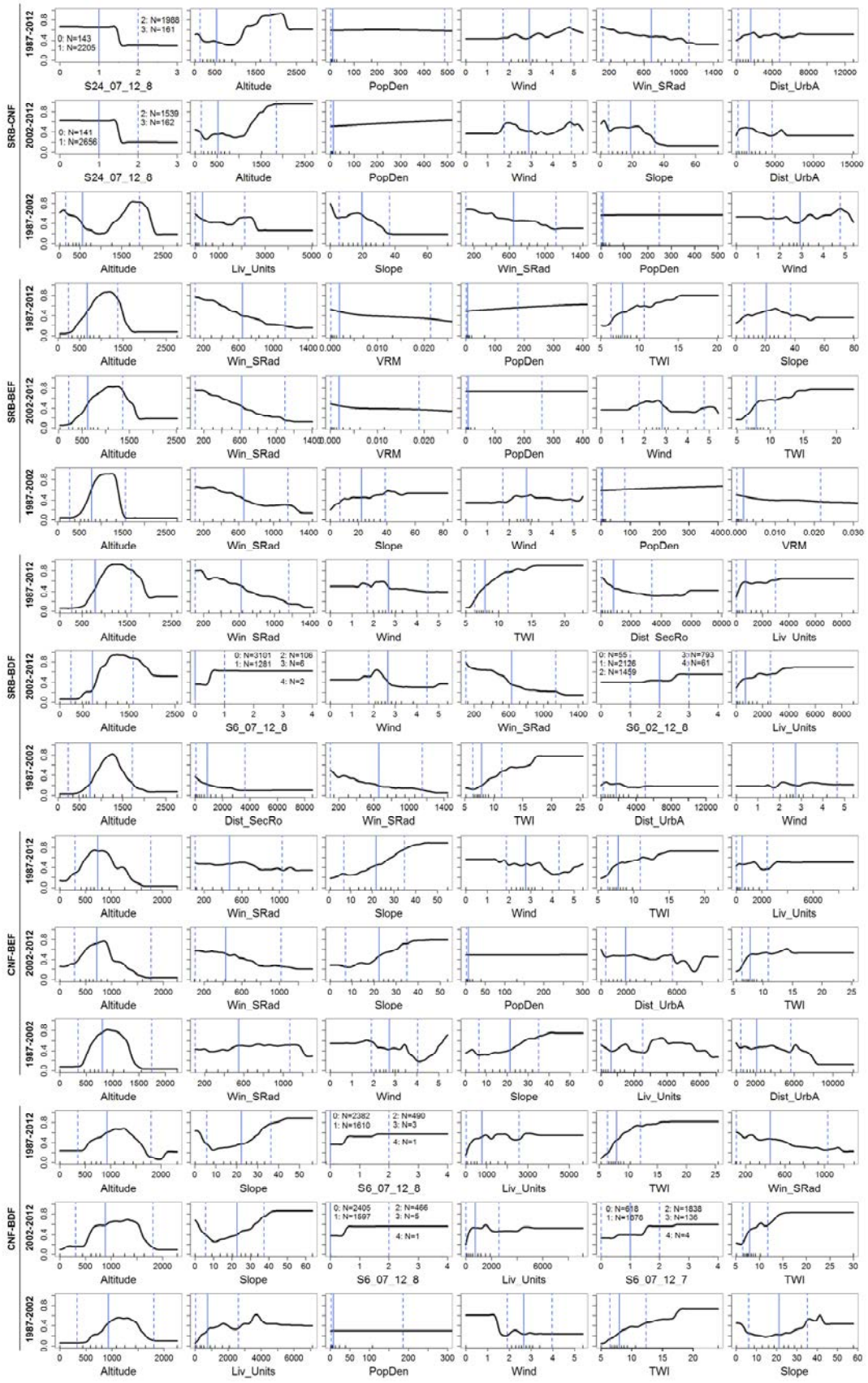


Figure S19. Partial dependence plots for the six most important variables. Cover changes 1 to 5. NE1, including altitude. Y-axis: probability. Black vertical lines above x-axes: deciles. Blue vertical lines: percentiles 0.05, 0.50, 0.95.

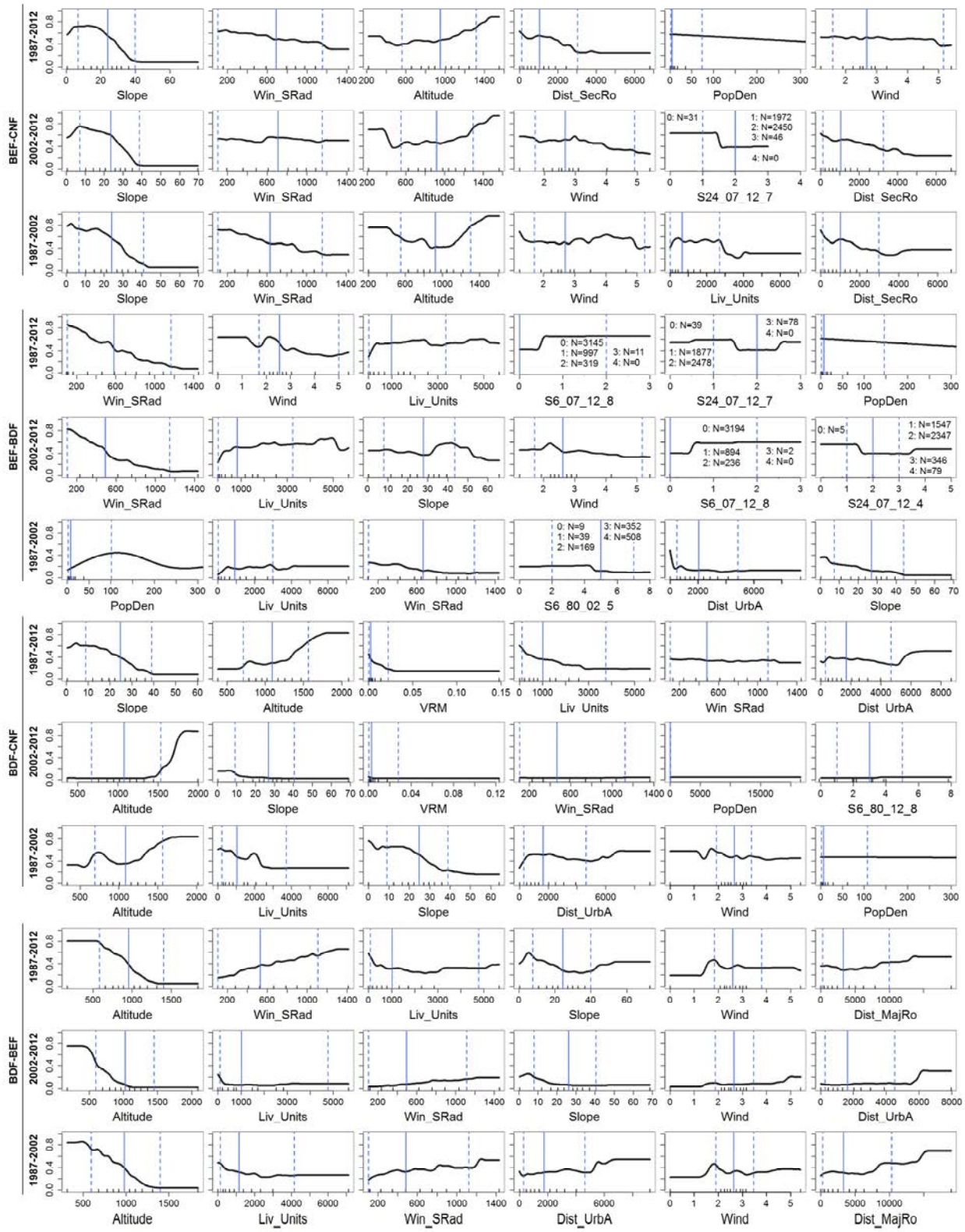


Figure S20. Partial dependence plots for the six most important variables. Cover changes 6 to 9. NE1, including altitude. Y-axis: probability. Black vertical lines above x-axes: deciles. Blue vertical lines: percentiles 0.05, 0.50 and 0.95.

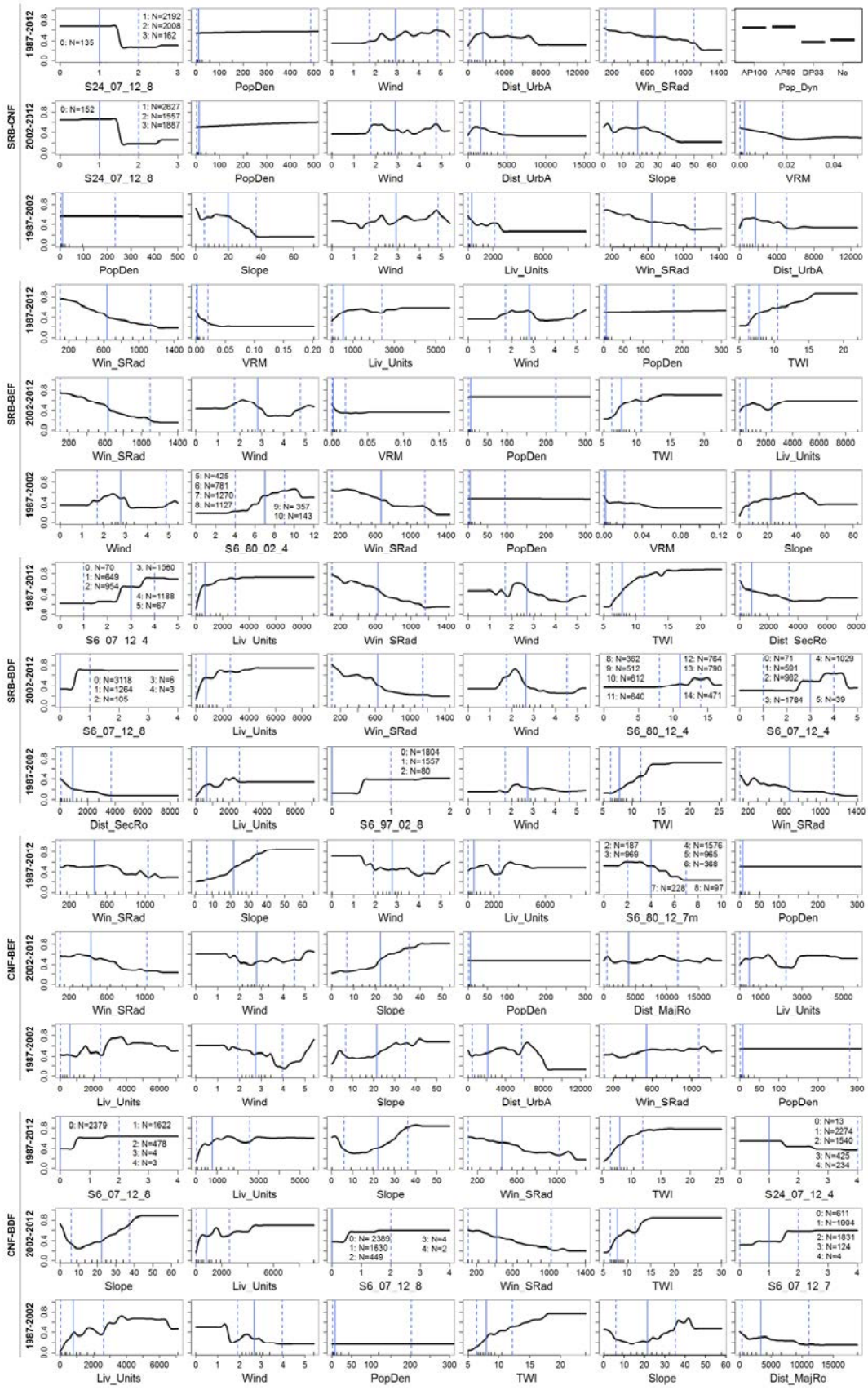


Figure S21. Partial dependence plots for the six most important variables. Cover changes 1 to 5. NE1, not including altitude. Y-axis: probability. Short black lines above x-axes: deciles. Blue vertical lines: percentiles 0.05, 0.50 and 0.95.

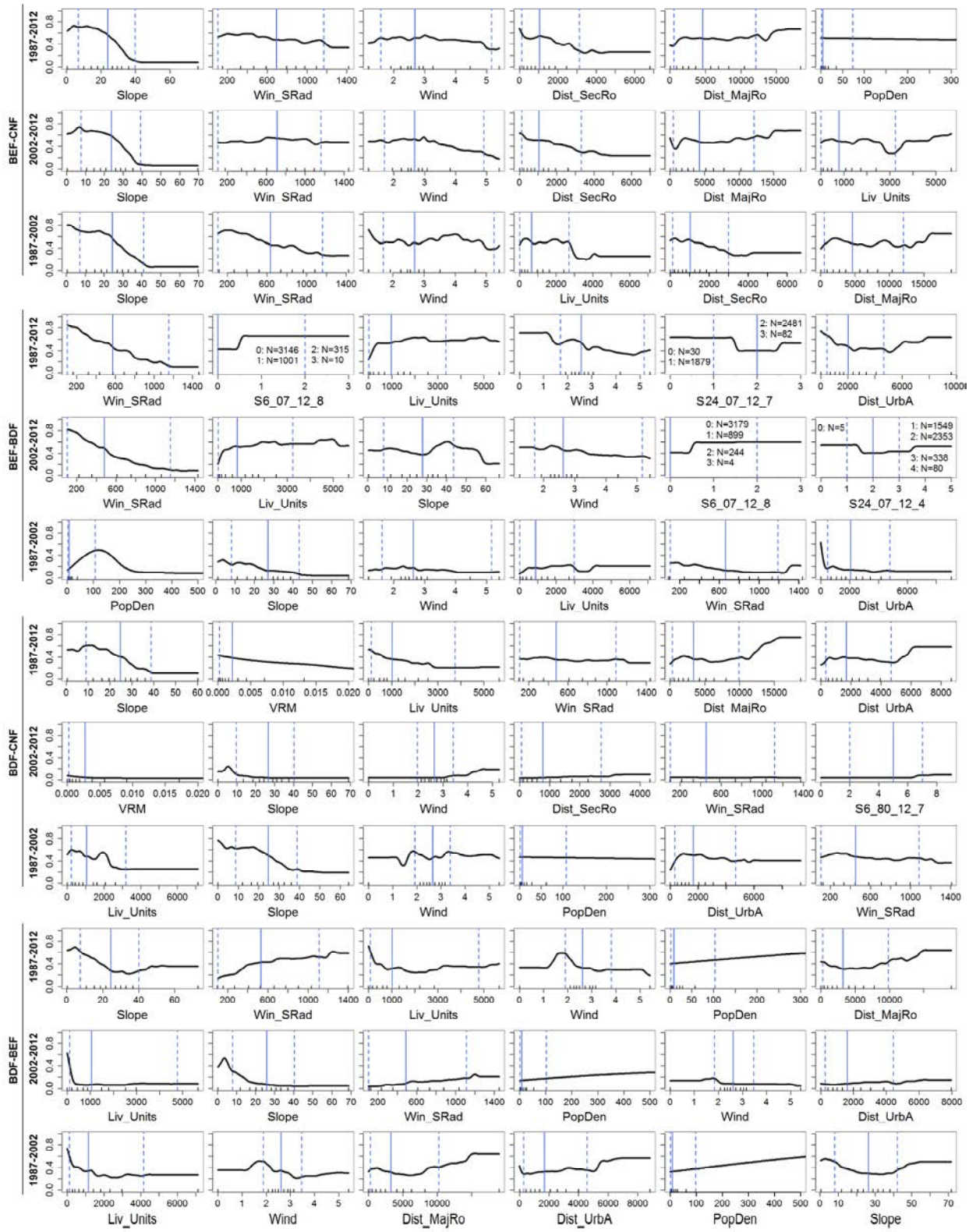


Figure S22. Partial dependence plots for the six most important variables. Cover changes 6 to 9. NE1, not including altitude. Y-axis: probability. Short black lines above x-axes: deciles. Blue vertical lines: percentiles 0.05, 0.50 and 0.95.

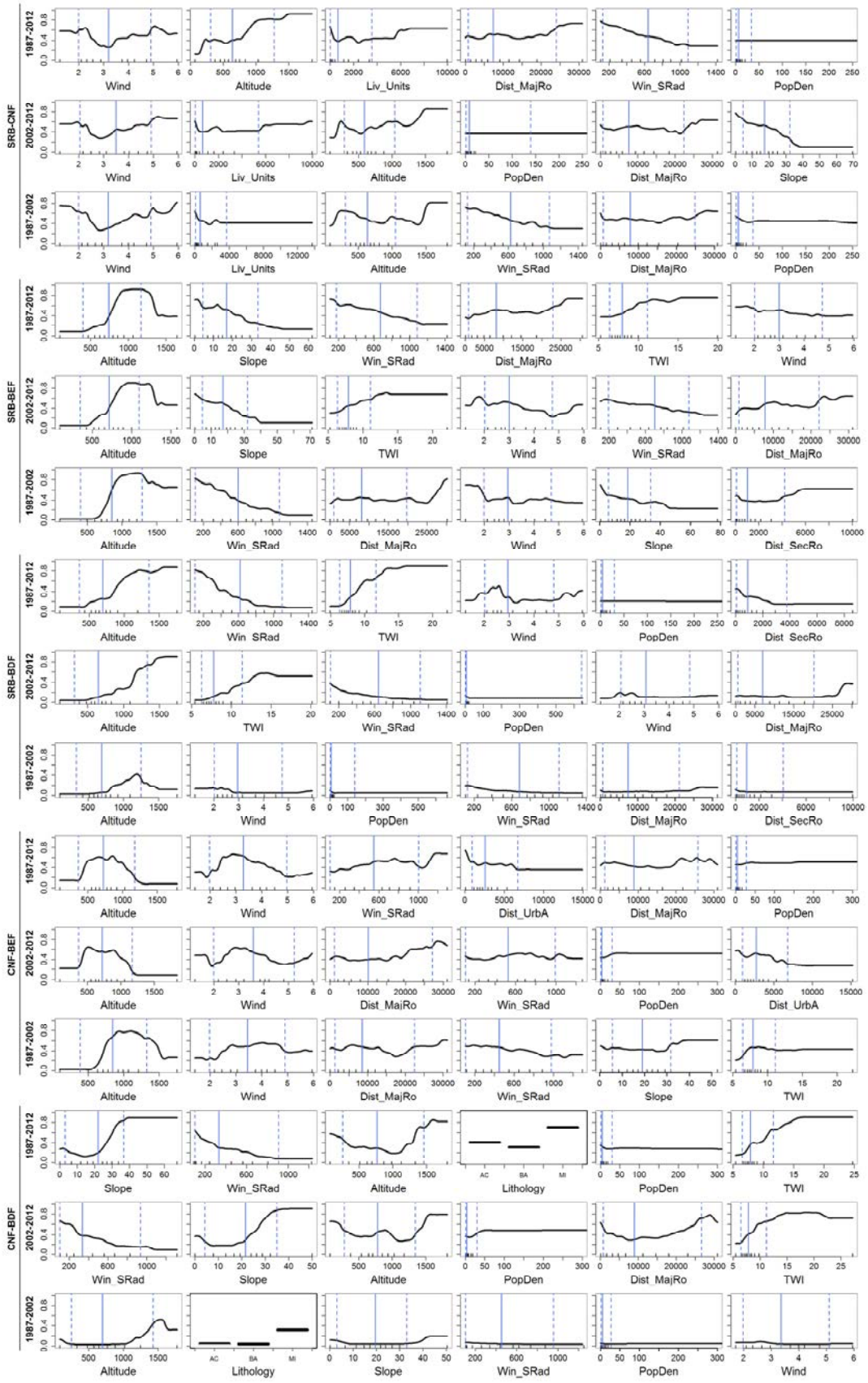


Figure S23. Partial dependence plots for the six most important variables. Cover changes 1 to 5. NE2, including altitude. Y-axis: probability. Short black lines above x-axes: deciles. Blue vertical lines: percentiles 0.05, 0.50 and 0.95.

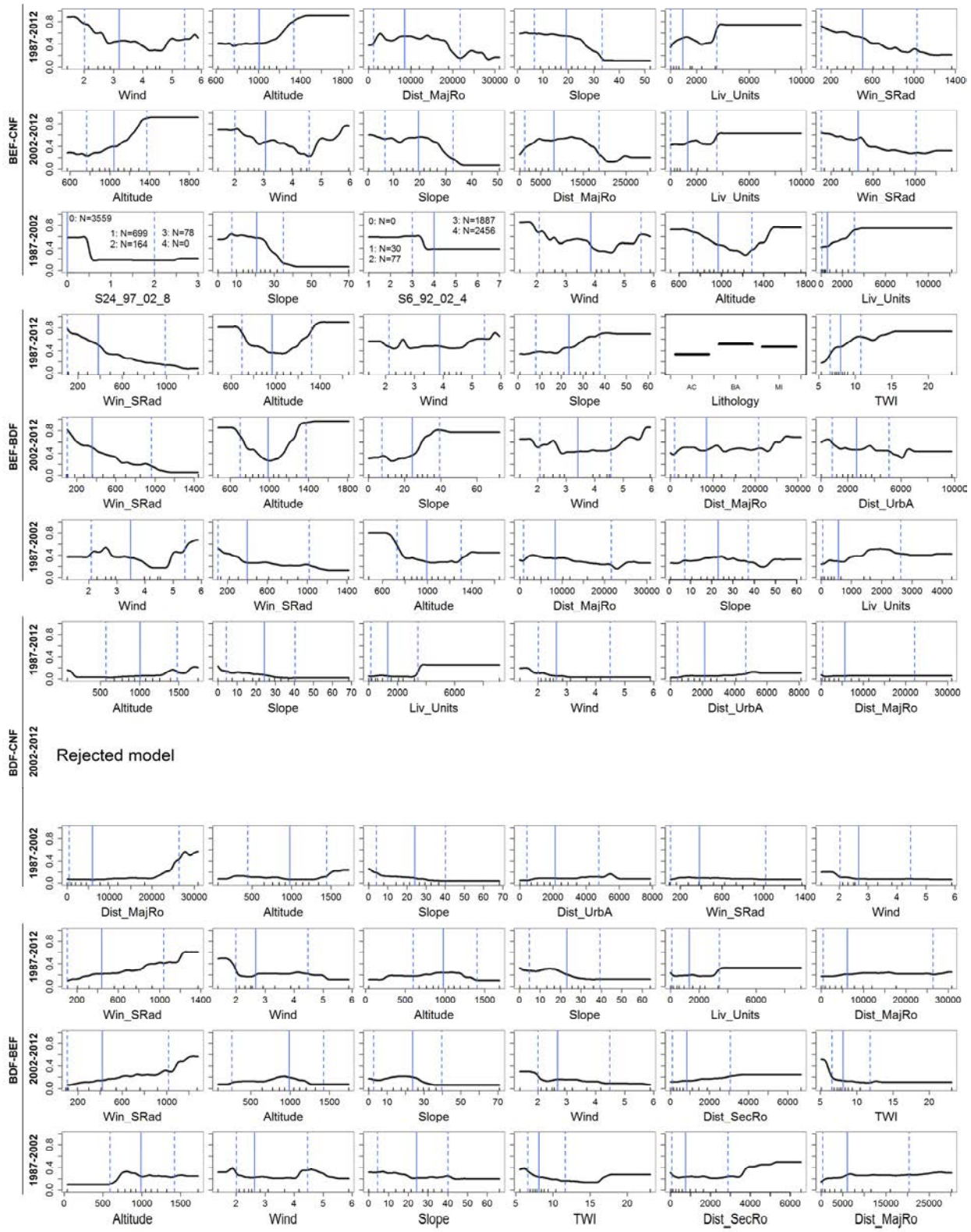


Figure S24. Partial dependence plots for the six most important variables. Cover changes 6 to 9. NE2, including altitude. Y-axis: probability. Short black lines above x-axes: deciles. Blue vertical lines: percentiles 0.05, 0.50 and 0.95.

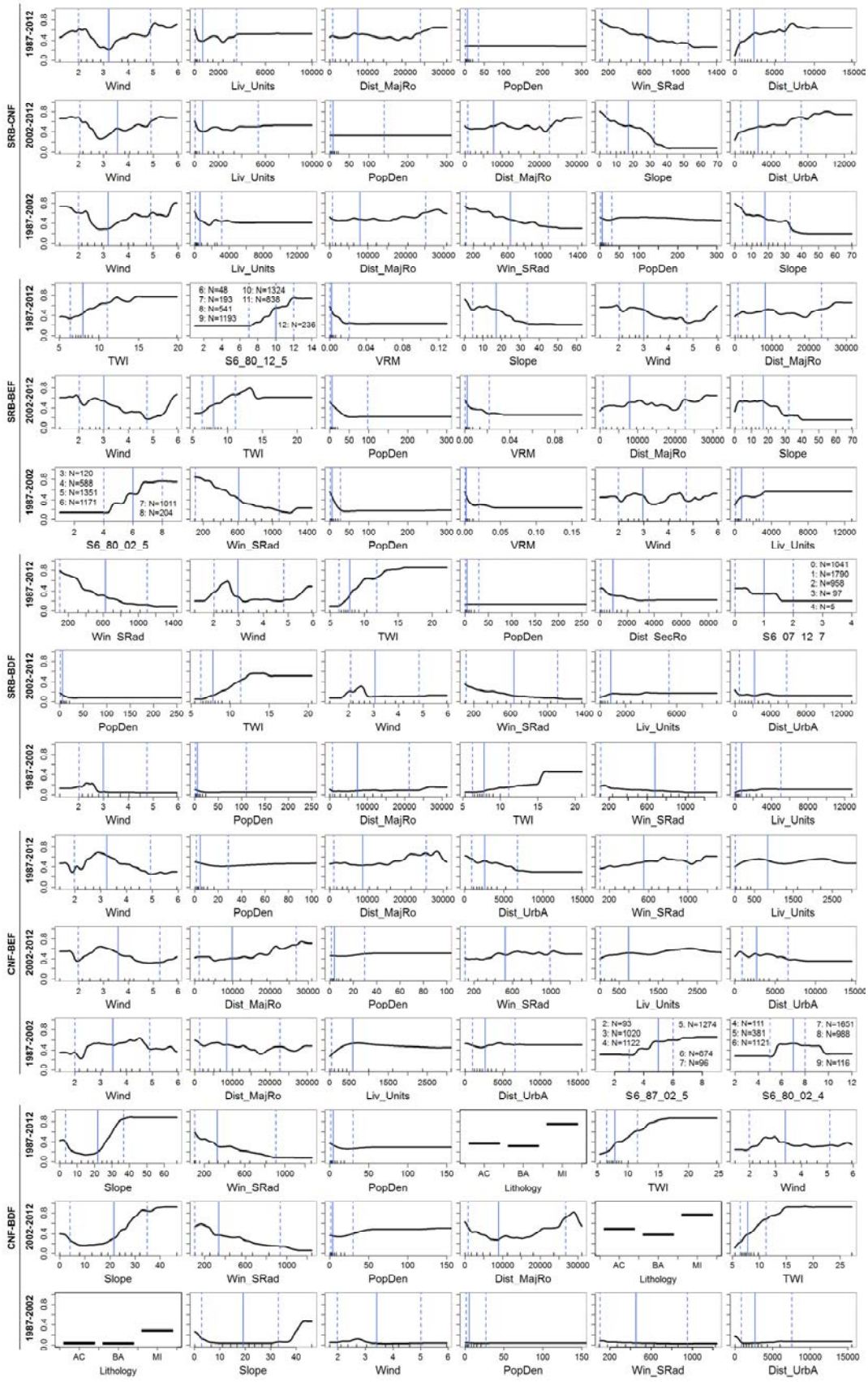


Figure S25. Partial dependence plots for the six most important variables. Cover changes 1 to 5. NE2, not including altitude. Y-axis: probability. Short black lines above x-axes: deciles. Blue vertical lines: percentiles 0.05, 0.50 and 0.95.

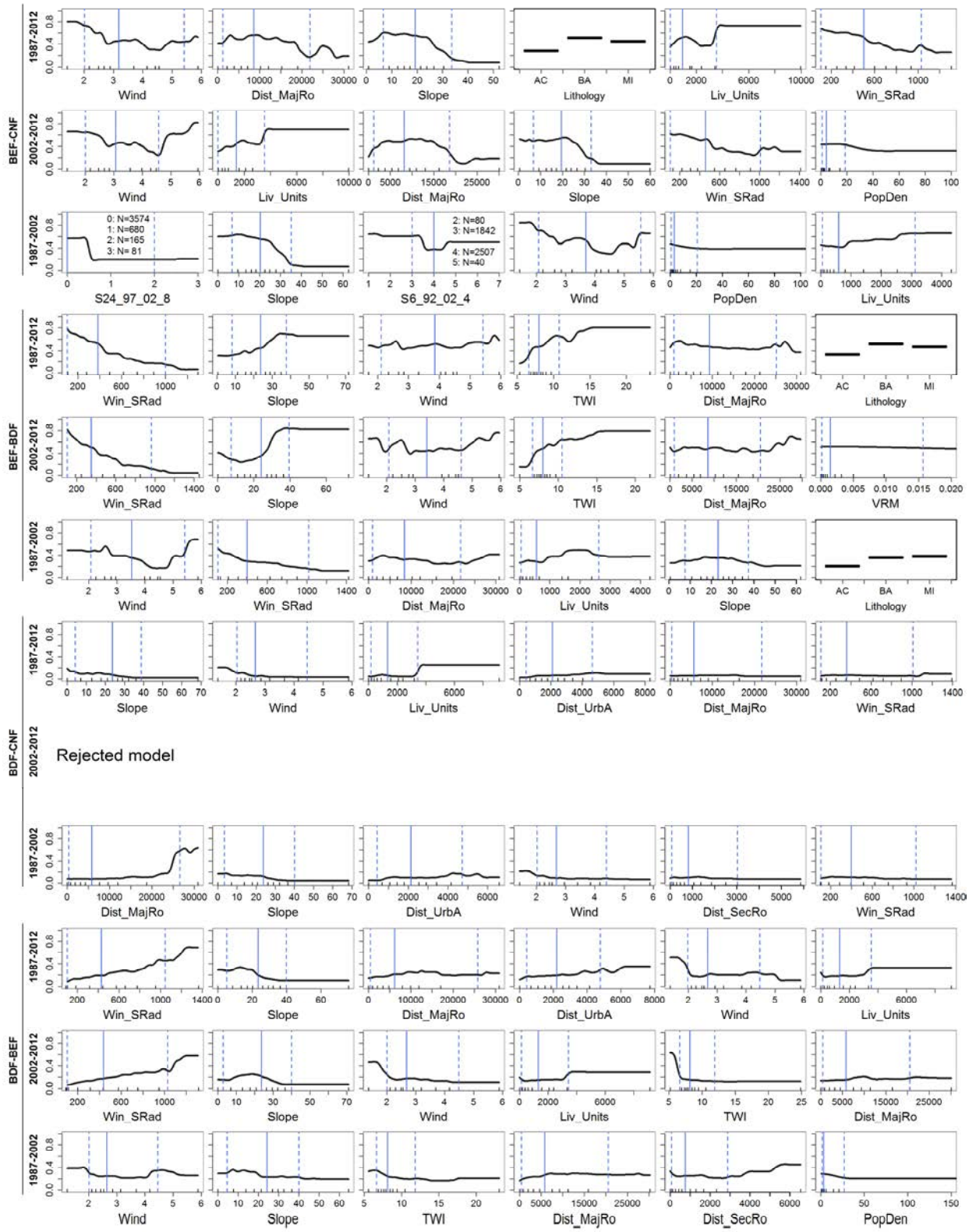


Figure S26. Partial dependence plots for the six most important variables. Cover changes 6 to 9. NE2, not including altitude. Y-axis: probability. Short black lines above x-axes: deciles. Blue vertical lines: percentiles 0.05, 0.50 and 0.95.

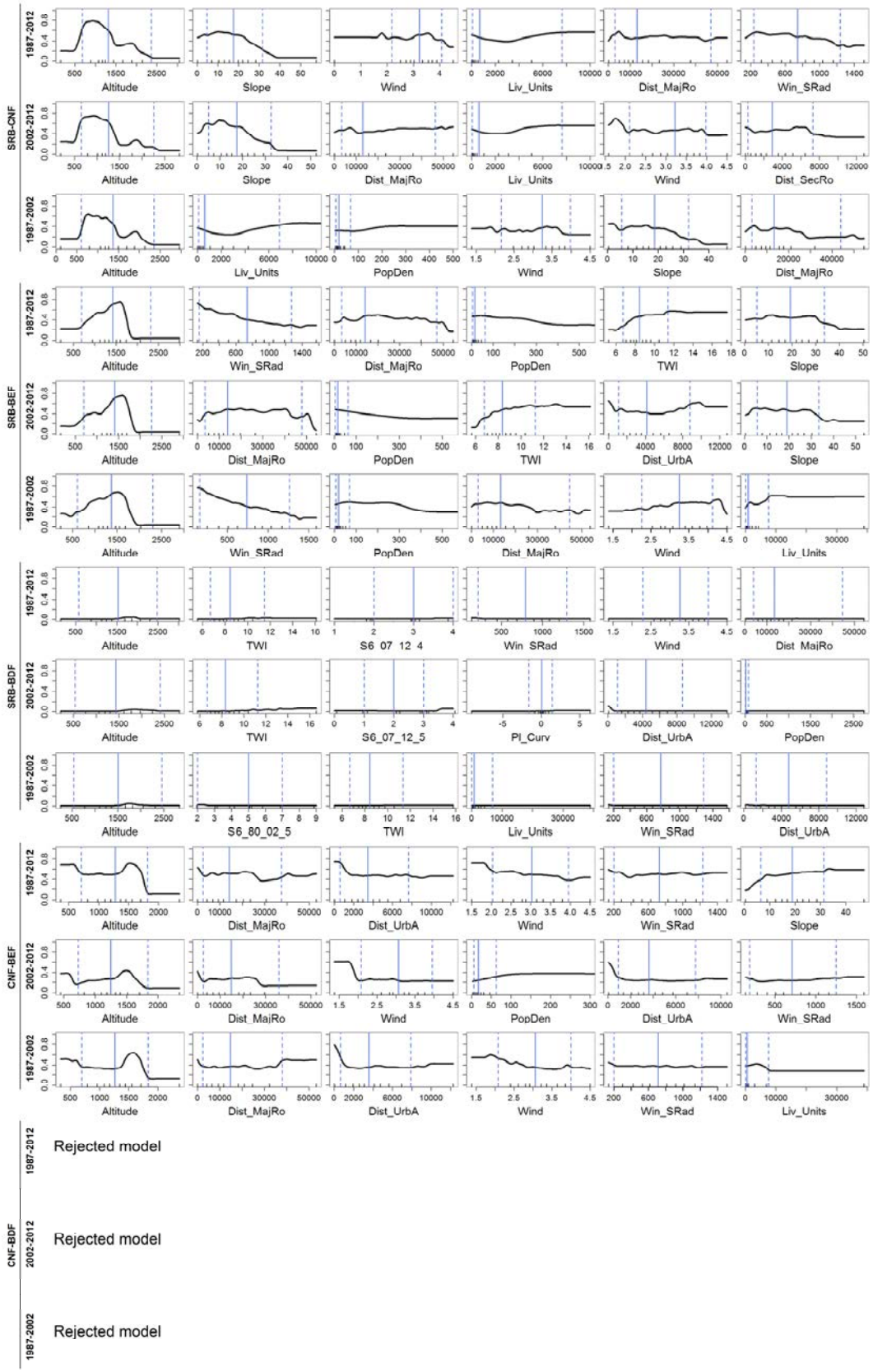


Figure S27. Partial dependence plots for the six most important variables. Cover changes 1 to 5. SE3, including altitude. Y-axis: probability. Short black lines above x-axes: deciles. Blue vertical lines: percentiles 0.05, 0.50 and 0.95.

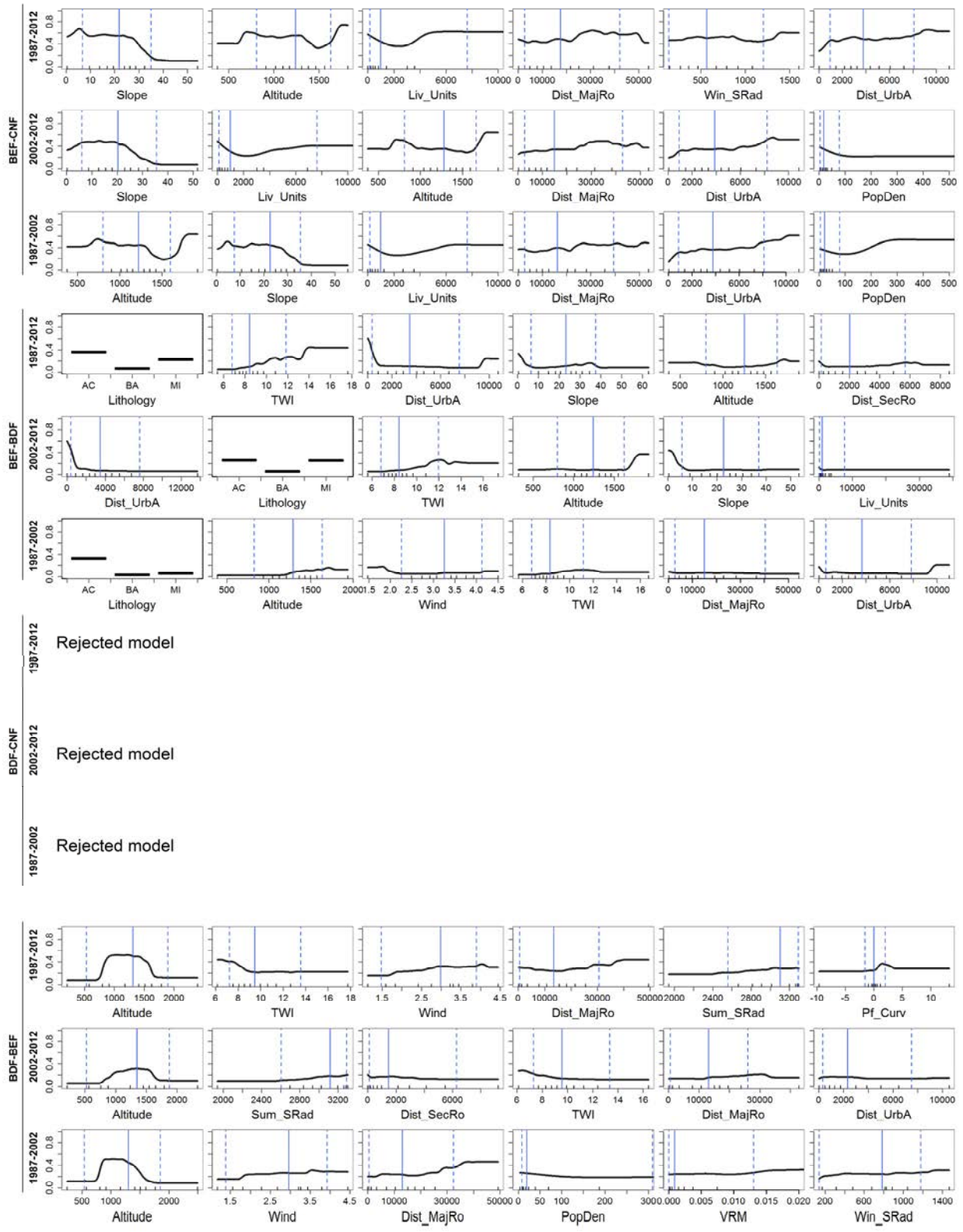


Figure S28. Partial dependence plots for the six most important variables. Cover changes 6 to 9. SE3, including altitude. Y-axis: probability. Short black lines above x-axes: deciles. Blue vertical lines: percentiles 0.05, 0.50 and 0.95.

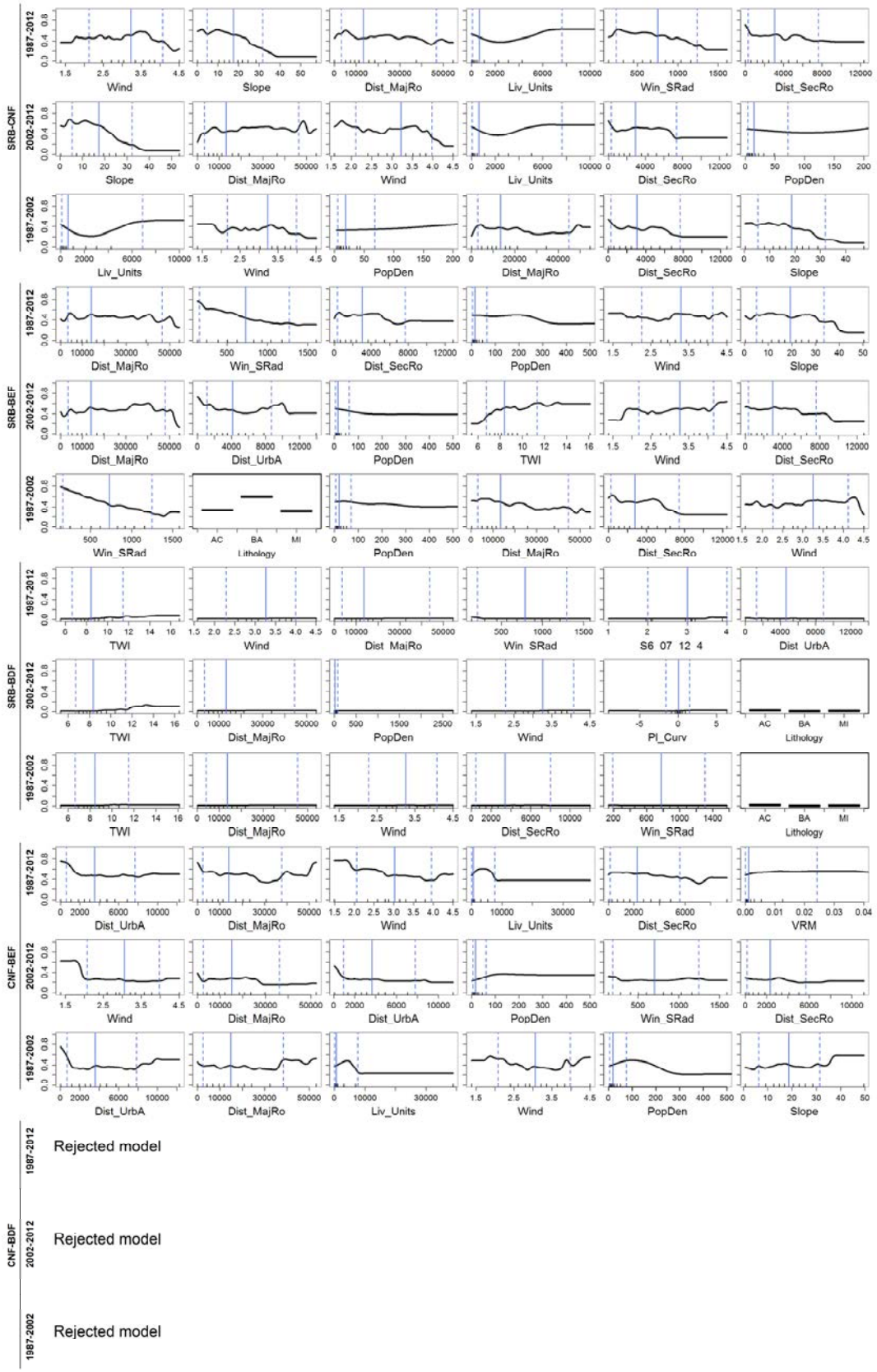


Figure S29. Partial dependence plots for the six most important variables. Cover changes 1 to 5. SE3, not including altitude. Y-axis: probability. Short black lines above x-axes: deciles. Blue vertical lines: percentiles 0.05, 0.50 and 0.95.

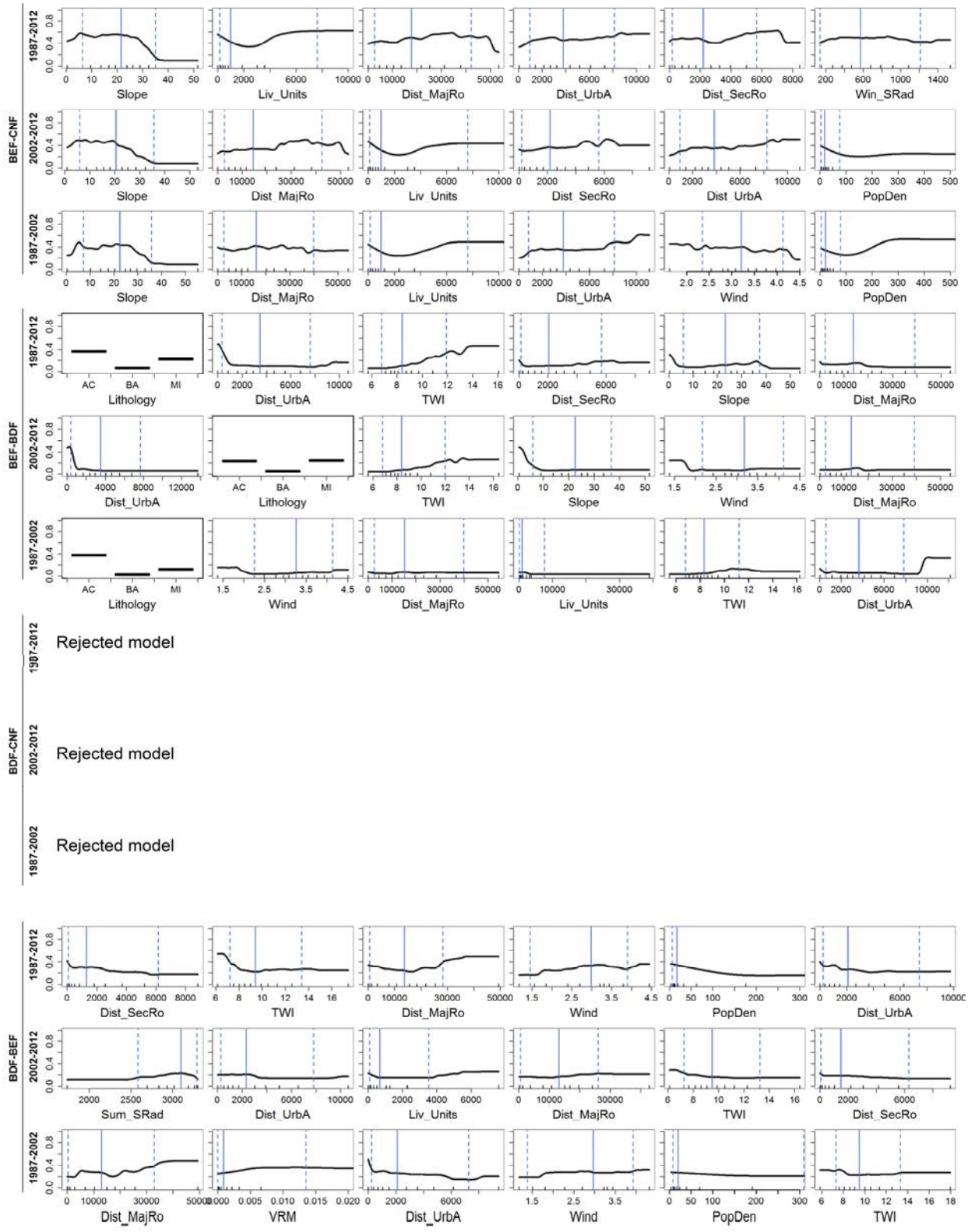


Figure S30. Partial dependence plots for the six most important variables. Cover changes 6 to 9. SE3, not including altitude. Y-axis: probability. Short black lines above x-axes: deciles. Blue vertical lines: percentiles 0.05, 0.50 and 0.95.

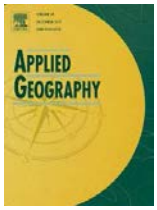
Vegetation-cover change	Pinus halepensis	Pinus nigra	Pinus sylvestris	Pinus uncinata	Quercus ilex	Quercus faginea	Quercus humilis	Quercus petraea	Fagus sylvatica
1. GRS-SRB	6223.00	3890.94	3639.77	21506.45	1662.14	4773.71	3274.23	32143.73	15299.55
2. GRS-CNF	13885.49	9489.33	1069.44	2587.57	6431.68	15451.00	2634.06	10242.02	8971.03
3. GRS-BEF	6099.09	2290.97	2307.73	16202.36	532.92	5142.49	1542.01	27963.23	12035.60
4. GRS-BDF	11423.70	8466.64	625.76	2997.05	5128.29	12801.94	1439.04	9033.15	6904.49
5. SRB-CNF	3669.12	7433.21	11441.59	59977.46	2203.09	7837.14	4595.41	40550.51	42098.02
6. SRB-BEF	3598.66	2987.77	5989.22	39200.66	369.98	3629.77	3252.59	37549.67	25473.60
7. SRB-BDF	8151.83	5663.93	1368.68	10767.14	2811.31	8621.95	1468.38	15677.14	9027.41
8. CNF-BEF	2853.16	2998.89	3813.21	37321.48	480.47	4609.44	2194.71	22879.55	20716.11
9. CNF-BDF	6606.93	6566.37	459.71	13621.50	3350.25	9693.04	540.03	14767.88	4728.68
10. BEF-CNF	5951.52	2651.01	1196.45	27032.60	830.15	6762.85	1582.65	27001.13	22217.71
11. BEF-BDF	5502.86	5071.05	748.77	20209.10	1217.81	7914.88	461.08	17997.50	8798.21
12. BDF-CNF	9336.07	7223.88	173.79	7725.85	4388.72	11933.81	734.78	12600.61	5893.04
13. BDF-BEF	5883.07	4129.01	1487.26	24931.98	1156.85	6557.28	1043.64	22761.24	15812.83

Table S11. Average distance from presence events in ambit NE1 to main dominant species in Catalonia. Reference information: Catalonia Land-Cover Map 1:25000, year 2009.

7. ARTÍCULO 3: *Environmental and socioeconomic factors of abandonment of rainfed and irrigated crops in northeast Spain*

Vidal-Macua, J. J., Ninyerola, M., Zabala, A., Domingo-Marimon, C., González-Guerrero, O. & Pons, X. (2017). Environmental and socioeconomic factors of abandonment of rainfed and irrigated crops in northeast Spain. *Applied Geography*. (*Artículo recientemente aceptado, aún en prensa*).

Environmental and socioeconomic factors of abandonment of rainfed and irrigated crops in northeast Spain



Juan José Vidal-Macua[1], Miquel Ninyerola[2], Alaitz Zabala[1], Cristina Domingo-Marimon[3], Oscar Gonzalez-Guerrero[1], Xavier Pons[1]

[1] Grumets research group. Dep. Geografia. Edifici B. Universitat Autònoma de Barcelona. 08193 Bellaterra, Catalonia, Spain.
[2] Grumets research group. Dep. Biologia Animal, Vegetal i Ecologia. Edifici C. Universitat Autònoma de Barcelona. 08193 Bellaterra, Catalonia, Spain
[3] Grumets research group. CREAF, Campus UAB. Edifici C, Cerdanyola del Vallès 08193, Catalonia, Spain.

ABSTRACT

Changes over the last century in the economic model of European countries and the development of the market economy have led to intense shifts in land occupation patterns. Agricultural abandonment is an important consequence of these processes and has modified natural and cultural landscapes, involving side-effects for society. Understanding how environmental and socio-economic factors influence the abandonment process can provide useful insights for managing investments, whether from the public or the private sector. In Spain, the Pyrenees and the Ebro Depression are two differentiated areas in terms of land-use dynamics, particularly in terms of the agricultural model carried out. In this paper we have analyzed the agricultural abandonment in these areas during the 1987-2012 period in relation to several potential explanatory factors. The analysis focuses on the abandonment of rainfed and irrigated herbaceous crops in order to derive specific explanations according to the crop type and geographical region. Crop covers were classified from four Landsat scenes, and conditions were described by topographic variables, human factors and drought occurrence. Boosted regression trees (BRT) were used to identify the most important variables and to describe the relationships between agricultural abandonment and key factors. Topography derived variables were found to be the main determinants, except for irrigated crops in the Ebro Basin, where locational factors play a more important role. BRT models allowed us to identify other significant patterns such as: the vulnerability of irrigated crops to drought; the higher dependence of agricultural activity in the Pyrenees on internal networks; pattern shifts of land abandonment in the analyzed sub-periods, and; evidence of the importance of economic diversification for maintaining cropland.

Keywords: agricultural abandonment; Spanish Pyrenees; Central Ebro Basin; rainfed crops; irrigated crops; Landsat classification; boosted regression trees;

1. INTRODUCTION

In the 20th century, especially during the second half, economic growth and development of mass markets in Europe resulted in a depopulation of marginal areas and a densification of those areas better placed to supply the expanding cities and industrialized zones (Fielding 1989, Collantes and Pinilla 2004). Agricultural land abandonment and intensification have continued to be widespread land-use changes since then (Ramankutty and Foley 1999), especially in the Mediterranean area (Gerard et al., 2010).

Land abandonment has been treated as a phenomenon involving negative or positive effects according to the analysis approach and/or the territorial context (MacDonald et al., 2000; Benayas et al., 2007; Keenleyside et al.; 2010). Soil erosion and reduced water yield of catchments are some of the most conspicuous issues that have been frequently addressed in the Mediterranean area, which is more prone to these problems (Gallart and Llorens 2004; López-Moreno 2006; García-Ruiz 2010; Kosmas et al., 2015). There are also associated benefits, like soil and nutrient recovery and carbon sequestration due to vegetation re-growth in abandoned lands (Tiessen et al., 1992; Molinillo et al.; 1997, Schröter et al.; 2005). The effect of cropland abandonment on landscape structure has been reported as a disturbance that can positively affect landscape connectivity and increase forest species (Farina 2006; Beilin et al., 2014; Navarro and Pereira 2015). In addition, other analyses have shown a positive association between abandonment and rural development (Beilin et al., 2014). On the other hand, several studies have reported the negative effects on biodiversity (Farina 2006; Bezák and Halada 2010; Halada et al., 2011), highlighting the importance of the seasonal-disturbance regime of croplands in maintaining the diversity and environmental range of species. Fire occurrence is another negative effect, especially regarding woodland recovery (Lloret et al., 2002; Moreira et al., 2012). Finally, during the last decade special attention has been paid to the drivers of agricultural abandonment, including environmental and socioeconomic variables as explanatory variables in statistical-based methods (Gellrich and Zimmerman 2007; Serra et al., 2008; Lieskovsky et al.; 2014, Sang et al., 2014; Pazur et al., 2014). These approaches are very useful for determining patterns of land abandonment and for identifying which places are more prone to being abandoned within a given territory.

The above-mentioned issues are especially relevant for land abandonment analyses in a country like Spain, which has high land-use heterogeneity due to its topographic and climatic spatial variability, and a long and complex human history. Spain, and the Mediterranean in general, has followed intense land cover dynamics that have led to high heterogeneity of land occupation patterns. Moreover, in recent decades the region has also been particularly influenced by climate change dynamics (IPCC 2007, 2013; Vicente-Serrano 2014). However, apart from some works describing the phenomenon through landscape metrics for relatively small areas (e.g., Bielsa et al., 2005), there are few detailed analyses of the conditions leading to agricultural abandonment in Spain, and more specifically, focused on the likelihood of abandonment based on empirical data.

In mountainous regions like the Spanish Pyrenees, the ecological succession of vegetation has advanced after a general abandonment of traditional activities (García-Ruiz 1990; Lasanta et al., 2005; Roura-Pascual et al., 2005; Cohen et al., 2011; Serra et al., 2014) driven by the combination of the rural exodus and the physical constraints to modernization (Benayas 2007, Collantes and Pinilla 2004). The cultivated area is mainly divided into cereals and fodder production. After the decline of the traditional system, based on a subsistence economy, cereals were less competitive (as a food supply) and have become an important source of fodder, while other forage crops and managed meadows have increased their area due the abandonment of transhumance in most counties (García-Ruiz and Lasanta 1993; IAEST; IDESCAT). The expansion of tourism in this region has led to economic readjustments in rural households. Many areas maintain a balanced symbiosis between tourism and agricultural production (Cánoves et al., 2004), but livestock farming and land cultivation have clearly decreased in the areas with more touristic development (García-Ruiz et al., 1996, 2011). In contrast, modernization (strong mechanization and limited labor) and agricultural intensification have taken place in other geographical regions of Spain, such as the Ebro Depression, a semi-arid region where modern irrigation systems have been developed since the middle of

the 20th century (Pinilla 2006). During the economic liberalization and capitalist development of Spain, this region was better placed to benefit from the expansion of industrialization and external markets, and the modern irrigation techniques increased the productivity of croplands. At the same time, this process contributed to the depopulation and agricultural abandonment in the Spanish Pyrenees due to the construction of reservoirs in valleys to supply the irrigated lands (Pinilla 2006; Duarte et al., 2014). Non-irrigated lands are mainly used to cultivate winter cereals, and although they occupy a greater extent of land, output per hectare for irrigated crops (mostly maize) is 5-6 times higher (Pinilla 2006). Land intensification is a common trend in the Mediterranean fringe of Spain, and although it leads to higher crop yields, its environmental and social viability depends on the pressures on the landscape and external market fluctuations (Nainggolan et al., 2012).

Although the external factors that have led to agricultural abandonment in Spain have already been described in several works, as we mentioned above, there are still few studies on the internal factors (Lasanta et al., 2017) controlling this phenomenon. Here, we analyze the influence of a series of internal environmental and socioeconomic factors on the land abandonment in northeast Spain from 1987 to 2012. In order to determine particularities regarding types of land cultivation and territorial contexts, the analysis differentiates between the abandonment of rainfed and irrigated crops in the Spanish Central Pyrenees and the Central Ebro Basin. Two sub-periods have been included in the analysis, 1987-2002 and 2002-2012, in order to take into account the variability in human and climatic factors. The ambits correspond to refined land-cover classifications of Landsat scenes. The importance of several factors and how they influence cropland abandonment was analyzed using boosted regression trees (BRT), also known as stochastic gradient boosting (Friedman 2001, 2002; Hastie et al., 2009). This relatively new machine-learning technique (Breiman 2001), in which hundreds or thousands of decision trees (Breiman et al., 1984) are sequentially and progressively fitted, has been demonstrated to have a high prediction performance (Kawakita et al., 2005; Elith et al., 2006, 2008; Leathwick et al., 2006; Crase et al., 2008) for the following main reasons: this approach does not assume any data distributions or data models, but rather it aims to determine dominant patterns by combining many classification trees; it identifies relevant variables and complex interactions; it is much less influenced by correlated information or irrelevant variables than other statistical approaches; it produces stable predictions (variance reduction); and it provides graphical depictions of the relationship between the response variable and predictors. We found this method particularly useful for multicausal scenarios, given its capacity to define the role played by the explanatory factors, even when the contribution (relative weight/importance) is low.

The general aim of this work is to better understand the internal conditions that lead to agricultural abandonment. We consider this study as groundwork for the development of sectoral studies and to improve the decision-making process for policy making and land management. Thus, the main objectives of this study are to determine: 1) the most important internal factors in the abandonment of rainfed and irrigated crops in two geographical regions, the Spanish Central Pyrenees and the Central Ebro Basin; and 2) the role of key factors in identifying particularities within each type of land cultivation and territorial context.

2. MATERIAL AND METHODS

2.1. Study areas

Two areas of northeast Spain were included in this work as different scenarios for model development. These areas are within the limits of land-cover classifications of four Landsat scenes, identified by their path-row: 198-030, 198-031, 199-030 and 199-031 (Figure 1).

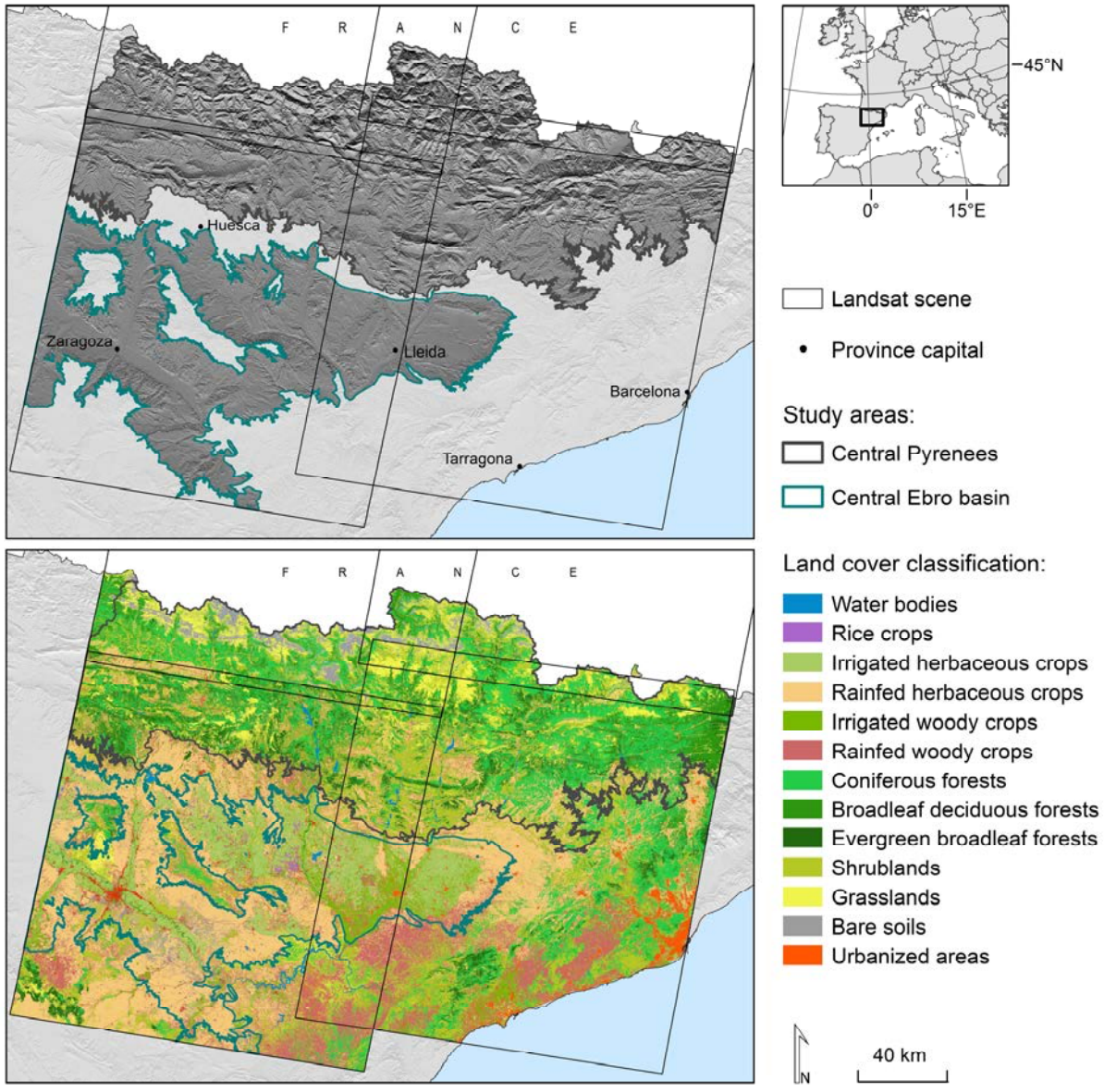


Figure 1. Study areas

The first study area corresponds to the mountainous region of the Spanish Central Pyrenees and, as previously explained, it is representative of an extensive and less-developed (less mechanized) agriculture. It includes areas with altitudes from 600 m up to the Spanish border with France (reaching altitudes of 3000 m and more), although in some locations the perimeter at its lower altitudinal limit has been adapted to natural boundaries with the Central Ebro Valley. Interior valleys below 600 m are also included within this geographical region. The area is influenced by a continental climate (although the eastern Pyrenees also have a Mediterranean influence) and the altitude introduces a temperature and precipitation gradient. The

predominant rainfed crops are barley and rye, generally located in the lower areas. Irrigated crops, such as forage plants, are mainly linked to the livestock farming and the local meat and dairy processing industry. The second area exemplifies the development and intensification of agriculture in the Central Ebro Valley. This region is defined by the area below 450 m elevation, and low-range interior mountains or those connected with mountain massifs were excluded in order to form a more homogeneous geographical region. The area has a continental climate with a marked semi-arid influence. Wheat and other feed cereals linked to intensive animal husbandry are the main rainfed crops in this area, while maize production is associated with modern irrigation systems.

2.2. Obtaining land abandonment areas and sampling

Agricultural land abandonment areas were obtained by classifying Landsat imagery at 30 m resolution following the methodology described in Vidal-Macua et al., (2017), in which a hybrid classification scheme is implemented. In our case, the k-nearest neighbor (kNN) classifier was used to obtain both multi-temporal training and test areas and to use them in final classifications (Figure 2). Each land-cover map refers to a five-year period, so classifications are composed of a set of dates in which the central years are 1987, 2002 and 2012. This allowed us to obtain a suitable representation of the seasonal phenological variation of crops because dates from March to September were included in each five-year period. The level of disaggregation for crop covers included woody and herbaceous crops, differentiating between irrigated and rainfed regimes. From the three land-cover maps (1987, 2002 and 2012), we computed land cover map changes for the periods: 1987-2002, 2002-2012 and 1987-2012.

Land abandonment was considered as woody vegetation encroachment or afforestation in agricultural plots. We selected irrigated and rainfed herbaceous crops as the land covers to be monitored due to the scarcity of abandonment events detected in woody crops. After a photointerpretation analysis of 20% randomly selected cases (orthophotos 1987, 1988, 2004, 2005, 2014 and 2015 at 1:5000 scale), many transitions to grassland cover (which could have scattered shrubs) were interpreted as a cessation of activity in the 1987-2002 period, but some of these turned out to be fallow stages and plots were cultivated again in 2012. Changes from crop to woody covers (shrublands or forests) were found to be stable transitions since woody plant encroachment in 2002 continued in 2012 (except cases related to changes to impervious surfaces in 2012). Therefore, changes from herbaceous crops to shrublands or forests covers between two dates were selected as abandoned areas. Mean producer accuracy (i.e. for a given category, the percentage of pixels in test areas which have been correctly classified) and mean user accuracy (i.e. for a given category, the percentage of pixels on the classified map which correctly appears in test areas) of the three land cover classifications (1987, 2002, 2012) are shown in Table 1, although natural vegetation covers were treated as one entire category to identify abandoned crop covers.

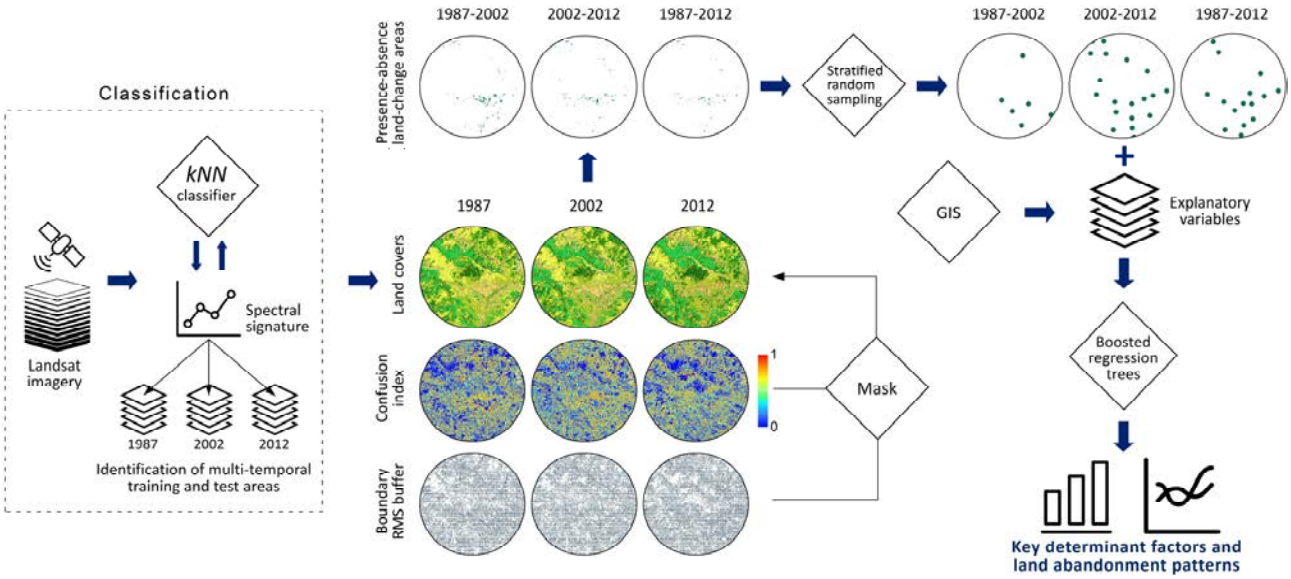


Figure 2. Methodological workflow (black icons designed by Noun Project, 2017).

	PA	UA
Irrigated herbaceous crops	88.23	92.33
Rainfed herbaceous crops	96.58	95.93
Coniferous forests	96.38	97.33
Broadleaf deciduous forests	94.76	91.78
Broadleaf evergreen forests	89.96	77.28
Shrublands	88.08	88.38

Table 1. Mean producer accuracy (PA) and mean user accuracy (UA) by land cover category

In order to avoid, as far as possible, including misclassified pixels in the statistical analysis, classifications were filtered using a confusion index (Burrough et al., 2000; Lewis et al., 2000; Tapia et al., 2005; Álvarez-Martínez et al., 2010). This index was calculated with the kNN algorithm in which the nearest training pixels (k-nearest neighbors) to each target pixel (to be classified) are identified by means of the Euclidean distance in the data space (Vidal-Macua et al., 2017). Values of this index range from 0 to 1, so that values close to 1 indicate high confusion between at least two classes, and values close to 0 represent high certainty for a classified pixel. We used a threshold of 0.5 so that pixels classified with a confusion index over this value where masked to exclude them from the land-cover maps.

A second mask was created to remove misregistration in land-cover polygon boundaries. This situation is related to some unavoidable location inaccuracy after the geometric correction (that cannot provide a zero planimetric error), and therefore, to false positive or negative changes between land-cover classifications (Pons et al., 2003) in pixels being at the borders between polygons. To solve this, classification polygon boundaries were eroded with a 20 m buffer mask according to the average RMS error of the geometric correction of the imagery. More details about this methodology can be found in Serra et al. (2003).

Finally, classifications were overlaid to identify abandonment of irrigated and rainfed crops for the two ambits (Central Pyrenees and Central Ebro Valley) and three periods (1987-2012, 1987-2002 and 2002-2012), which resulted in 12 initial models to be evaluated. We considered crop-cover changes as absence/presence events. Areas where the crop cover remained stable between two dates were treated as absences, and those where there were changes to vegetation covers were treated as presences. Once the abandoned and non-abandoned areas were established, stratified random sampling was carried out maintaining the prevalence (the same number of points) between the two types of events. Points were randomly selected for each event class with a minimum distance of 1 km between points. As expected, there were far fewer abandonment events in the Central Ebro Basin (Table 2). Therefore, for irrigated herbaceous crops in the Central Ebro Basin, we only considered the 1987-2012 period to develop the land-abandonment model. The same number of absence events as presence events were selected for all cases.

Pyrenees	Rainfed herbaceous crops Presence events	Irrigated herbaceous crops Presence events
1987-2002	1361	648
2002-2012	2197	735
1987-2012	2039	857
Central Ebro Basin		
1987-2002	392	42
2002-2012	308	46
1987-2012	505	169

Table 2. Number of presence events by ambit and period

2.3. Explanatory variables

A set of variables was chosen (Table 3) to be included as explanatory factors for agricultural abandonment in the studied areas. The intention was to identify the most important variables and use statistical inference to describe the abandonment according to the value range of the variables. The MiraMon 8.1 (Pons 2006), ArcGis 10.1 (ESRI 2012) and SAGA 2.1.2 (Conrad et al., 2015) software were used to calculate these variables.

Topography-derived variables

These variables are related to physical and environmental factors that could influence agricultural abandonment. The intention was to determine how the agricultural activity is affected by the environmental variability derived from the topography and its vulnerability in different topo-climatic contexts. Slope, curvature and terrain roughness can be used as proxies of the physical constraints to agricultural production, as well as the topography effect on the amount of incident solar radiation can help to identify thresholds in crops performance since solar radiation is related to factors such as the water evapotranspiration.

A 10 m Digital Elevation Model (DEM) was created from 1:5000 map sheets of the Aerial Orthophotography National Plan 2010 (PNOA) and the following variables were derived from it:

- Altitude in meters above sea level.
- Terrain slope in degrees.
- Overall terrain curvature (dimensionless). The 0 value means no curvature in a typical range from -5 to 5, where negative values represent concave curves and positive values indicate convex curves.
- Terrain roughness (dimensionless), using the Vector Ruggedness Measure (Sappington et al., 2007), for which values close to 0 represent flat areas in an approximate range from 0 to 0.20 in our ambits.
- Topographic Wetness Index (dimensionless), which is a DEM-based soil moisture index (Beven and Kirkby 1979; Sørensen et al., 2006; Kopecký and Čížková 2010) that generally ranges from 3 to 30, where higher values indicate higher moisture availability.
- Winter and summer solar radiation (units in 10 $kJ \cdot m^{-2} \cdot day^{-1}$), which calculates the total amount of incident solar radiation for each pixel at winter and summer solstice dates, following the methodology of Pons and Ninyerola (2008).
-

Distance variables

Four variables were obtained as a measure of land accessibility: cost distance to province capitals, urban areas, major roads and secondary roads. Cost distance is calculated for each cell as the lowest accumulative cost to the nearest target feature (urban areas, major roads, etc.) weighting the Euclidean distance by the slope (ESRI 2012) and an impedance factor to movement depending on cell values (such as major roads, secondary roads or background) given a cost surface.

Municipality-level information

Nine variables were elaborated from INE (Spanish Statistical Office) data at a municipality level as factors that can influence agricultural abandonment:

- Percentage of population between 20 and 55 years old. We relate this age group to the economically active population.
- Percentage of population aged 55 years or over. Economically inactive population taking into account possible early retirements.
- Rate of natural increase. The birth rate minus the death rate by population size. For these first three variables, the fifth year prior to the last year of each period was chosen as the reference year (1997 and 2007).
- Annual work units. This variable relates hours worked on an agricultural holding to population size (http://ec.europa.eu/eurostat/statistics-explained/index.php). These data are only available for the Agricultural Census of 1999 and 2009.
- Accommodation capacity. The number of beds in tourist accommodations divided by population size. Given the restricted availability of these data, the reference years are 2000 and 2010.
- Four variables accounting for the number of workers in the primary sector (agriculture, livestock and fishing), industry and energy sector, building sector and services sector, weighted by population size. These variables were used only in the 2002-2012 and 1987-2012 periods, evaluating additional models, because data before 2002 were unavailable.

Variables representing recurrent drought episodes based on the Standardized Precipitation Evapotranspiration Index

A set of variables indicating recurrent drought episodes was generated based on the Digital Topo Climatic Drought Atlas of the Spanish Iberian Peninsula (Domingo-Marimon, 2016). The Atlas includes a set of Standardized Precipitation Evapotranspiration Index (SPEI) maps at 100 m spatial resolution for the entire Iberian Peninsula from 1950 to 2012. The SPEI (Vicente-Serrano et al., 2010) is based on precipitation and mean temperature (to estimate potential evapotranspiration) and quantifies water deficits for multiple timescales. The values are standard deviations and negative values indicate less than average precipitation, i.e. drought events, while positive values indicate greater than average precipitation, i.e. wet events. A threshold of SPEI =<-1 was selected to identify drought conditions, which end as soon as the SPEI>-1 again. The index can be computed at several timescales corresponding to drought specific conditions.

A first set of variables was generated using SPEI at a 6-month timescale from 1980 to 2012 as indicative of the medium-term moisture condition, to first identifies anomalies in the water streamflow. A second set of variables was generated using SPEI at a 24-month timescale from 1980 to 2012 as indicative of the long-term moisture condition, to identify reservoir level and ground water level anomalies. For both sets, the number of drought episodes (SPEI=<-1) with durations of a minimum of 4, 5, 7 or 8 consecutive months, as representative lengths that may cause harmful effects, were counted for the period 1980-2012. The number of drought episodes was also counted for year spans of 15 and 10 years from 1987 to 2012. Therefore, the year spans used were: 1980-2012, 1997-2012 and 2007-2012. In addition, the analysis was performed using two shorter sub-periods, 1987-2002 and 2002-2012, and their corresponding year spans: 1980-2002, 1987-2002, 1992-2002 and 1980-2012, 1997 2012 and 2002-2012 respectively. Including several year spans allows us to analyze how a greater or lesser drought frequency influences forest dynamics and whether recent droughts (during the last 10 years) have had any effect on transition.

Variable code	Description	Units
Altitude	Altitude	Meters
Slope	Terrain slope	Degrees
Curv	Overall terrain curvature	Dimensionless
VRM	Terrain roughness	Dimensionless
TWI	Topographic Wetness Index	Dimensionless
Win_SRad	Winter solar radiation	10 $kJ \cdot m^{-2} \cdot day^{-1}$
Sum_SRad	Summer solar radiation	10 $kJ \cdot m^{-2} \cdot day^{-1}$
Dist_ProCa	Cost distance to province capitals	Meters
Dist_UrbA	Cost distance to urban areas	Meters
Dist_SecRo	Cost distance to secondary roads	Meters
Dist_MajRo	Cost distance to major roads	Meters
P_20_55	Population between 20 and 55 years old	Percentage
P_m55	Population aged 55 years or over	Percentage
RNI	Rate of natural increase	(births-deaths)·100/population size
AWU	Worked hours on an agricultural holding (1,800 hours = 1 AWU)	AWU/population size
AcCap	Accommodation capacity	nº of beds/population size
Wr_ALF*	Workers in primary sector	Wr_ALF/population size
Wr_InE*	Workers in industry and energy sector	Wr_InE/population size
Wr_Bld*	Workers in building sector	Wr_Bld/population size
Wr_Ser*	Workers in services sector	Wr_Ser/population size
S6_80_02_5 (example)	Every drought variable is coded as follows: "S6" - First two digits indicate the SPEI time-scale (6 or 24) "80_02" - Span of years used to calculate drought occurrence "5" - Last digit indicates drought episode duration	Number of drought episodes

Table 3. Explanatory variables. *These variables were only used in the periods 2002-2012 and 1987-2012 due to the unavailability of data before 2002.

2.4. Variable subset selection

A collinearity analysis was carried out before the models were run to avoid the presence of correlated drought variables and to reduce processing times. The initial set of 24 drought variables was resized for each model using an alternative way to correlate coefficient estimates. We used Variance Inflation Factors (VIF) and a VIF threshold of 5 as other authors recommend (Zuur et al., 2007; O'brien 2007; James et al., 2013).

2.5. Data analysis

To identify the most important variables and to quantify their influence on land abandonment we used boosted regression trees (BRT). BRT is a tree-based method combined with the strength of boosting (Breiman et al., 1984; Freund and Schapire 1996; Hastie et al., 2009; James et al., 2013). By boosting, hundreds or thousands of trees are built using a random fraction of the observations in each new tree, using a different random subset of variables in each of them. Trees are not fitted individually but additively because each new tree is fitted to the residuals of the previous tree. The procedure to sequentially build trees is based on a series of rules that weigh observations depending on their error rate after the previous tree is fitted (Friedman 2002; Hastie et al., 2009). Each new tree is focused, then, on observations that have been poorly predicted.

The bag fraction is the BRT parameter that controls the randomly selected observations for each new tree. We used a bag fraction of 0.5 (50% of the observations) as suggested in Friedman (2002) and Elith et al., (2008) after several testbeds were applied. The other main parameters of the BRT models are: tree complexity, learning rate and number of trees. Tree complexity is the number of splits allowed for fitting at each iteration (tree) and is related to the interaction level between variables. A tree complexity of 1 means that all trees are composed of single candidate variable (which might differ from one tree to another) and no interaction with other variables is allowed. Instead, an overly complex tree (involving many variables) could overfit the training data and devalue predictions in a test dataset. The learning rate parameter works like shrinkage techniques in other linear models (James et al., 2003; Hastie et al., 2009), scaling the contribution of each new term (i.e. the constituent trees) in the additive model by a factor between 0 and 1. A low learning rate allows more complex trees to be added gradually without overfitting the model, thus diverse interaction structures are taken into account in the additive model. Finally, the optimal number of trees depends on the choice of the two previous parameters. Normally, low learning rates, in a range of 0.001 to 0.01, imply that the number of trees needs to be increased to achieve a good performance (James et al., 2003; Elith et al., 2008; Hastie et al., 2009), although a validation procedure is recommended (Ridgeway 2007). Prediction for a single observation is the sum of predictions over all trees weighted by the learning rate.

BRTs were applied in R software using the *dismo* package (Elith et al., 2008), which is based on the *gbm* package (Ridgeway 2007). In order to determine the best parameterization for each crop-cover abandonment model, we ran a model for all possible combinations of learning rates 0.001, 0.0025, 0.005, 0.01, 0.015, 0.03 and tree complexities from 1 to 5. For each of these models the optimal number of trees was identified by a 10-fold cross-validation (Elith et al.,. 2008), setting the maximum to a 15 000-tree model. Once the optimal number of trees had been identified, the predictive performance was recorded using the deviance (Friedman 2002; Ridgeway 2007; Elith et al., 2008; Hastie et al., 2009). We calculated

the deviance by applying a 10-fold cross-validation, and the model with the lowest deviance value was selected to obtain and interpret the results. In addition, in the results section we show the prediction performance of the selected models (those with the lowest deviance) again using a 10-fold cross-validation to represent the ROC (Receiver Operating Characteristic) curve and to calculate the AUC (Area Under the Curve) score (Hanley and McNeil 1982; Fawcett 2004; Pontius and Parmentier 2014). This procedure allows to evaluate the robustness of the selected models by checking the balance between true positives (fraction of presence events that are correctly identified) and false positives (fraction of absence events that are incorrectly classified as presence events). The ROC curve represents this balance for all possible probability thresholds in order to find the optimal one that maximizes the true-positive rate and minimizes the false-positive rate. The AUC is a summary metric that quantifies this balance in a range of 0 to 1 (Fawcett 2004), where values close to 1 indicate a high predictive performance.

The initial results showed that slope and terrain roughness (*VRM*) were the variables that contributed most notably in almost all models (Figure A1 and Figure A2), so we decided to run an additional model that did not include these predictors. In other words, once the importance of slope and terrain roughness had been recognized, removing them from the analysis makes it more likely that other variables can appear as important contributors and, consequently, other interacting schemes may fit the models. Finally, we analyzed the results including the variables *Wr_ALF*, *Wr_InE*, *Wr_Bld* and *Wr_Ser* in models of the periods 2002-2012 and 1987-2012. Then, the results focused on models that did not include slope or *VRM* and models that considered the number of workers by economic sector.

To interpret the results, the importance of the predictors in the BRT models was determined by considering their contributions in the additive model. The relative importance was measured based on the number of times a predictor was selected for splitting, weighted by the squared improvement to the model as a result of each split, and averaged by the number of trees (Friedman and Meulman 2003; Hastie et al., 2009). Then, results were scaled between 1 and 100.

The probability of a land abandonment event was modeled via the *logit* of a logistic regression (James et al., 2003; Elith et al., 2008). The relationship between the land abandonment and an explanatory variable was interpreted using dependence plots, which is a visualization of the fitted functions in a BRT model. This graphical output shows the effect of an explanatory variable after the average effects of all other variables are accounted for, using the weighted tree traversal method described in Friedman 2001 (Ridgeway 2004). Thus, these plots show which areas are most likely to be abandoned regarding the value range of the variables.

3. RESULTS AND DISCUSSION

All models have AUC scores above 0.70 (Table A1 of the Appendix). Figure 3 shows the ROC curves and AUC scores calculated with the 10-fold cross-validation for the models that do not include the slope and VRM variables.

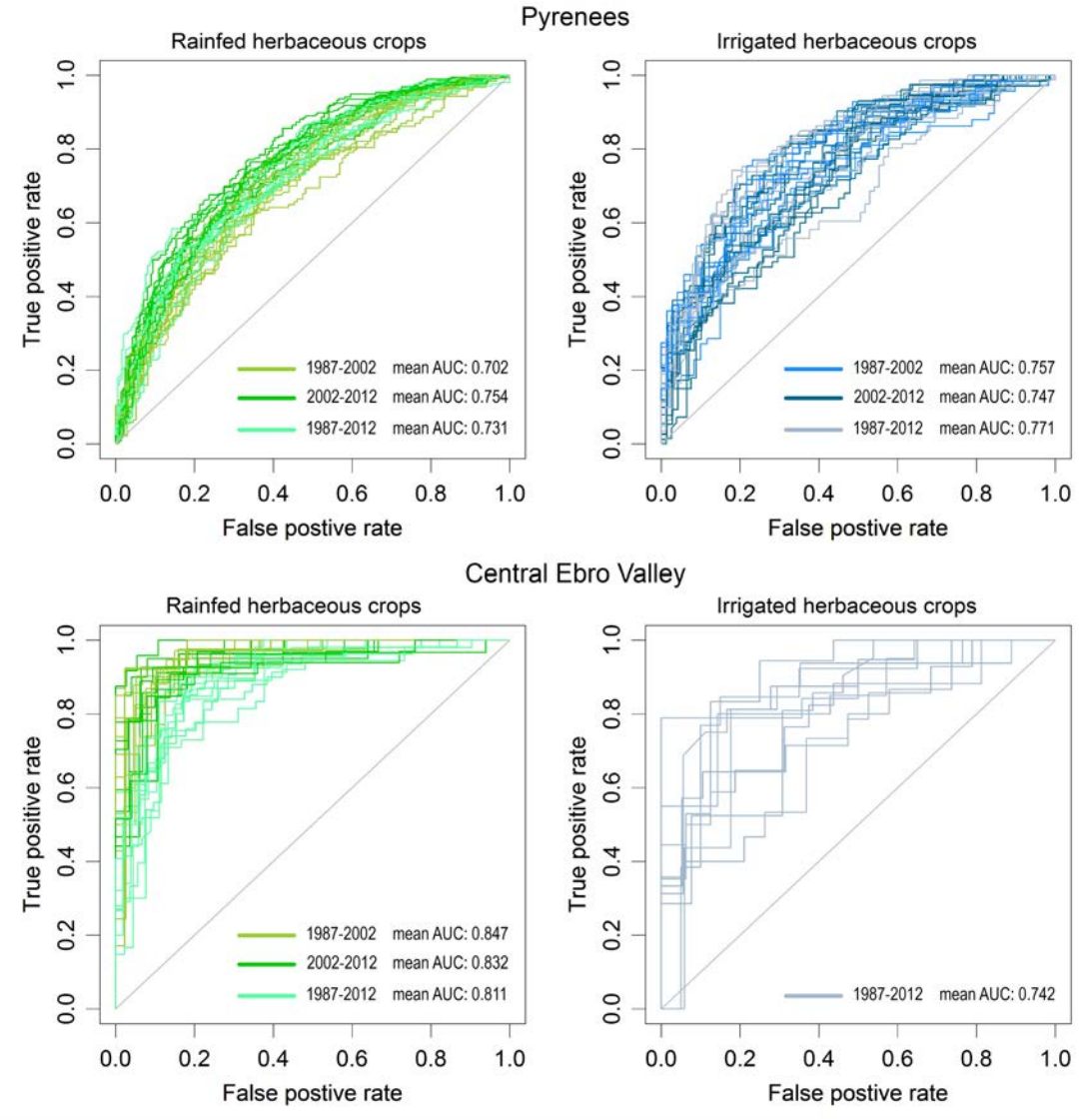


Figure 3. ROC curve and AUC score calculated with the 10-fold cross-validation. Models with the lowest deviance value that do not include Slope and VRM variables.

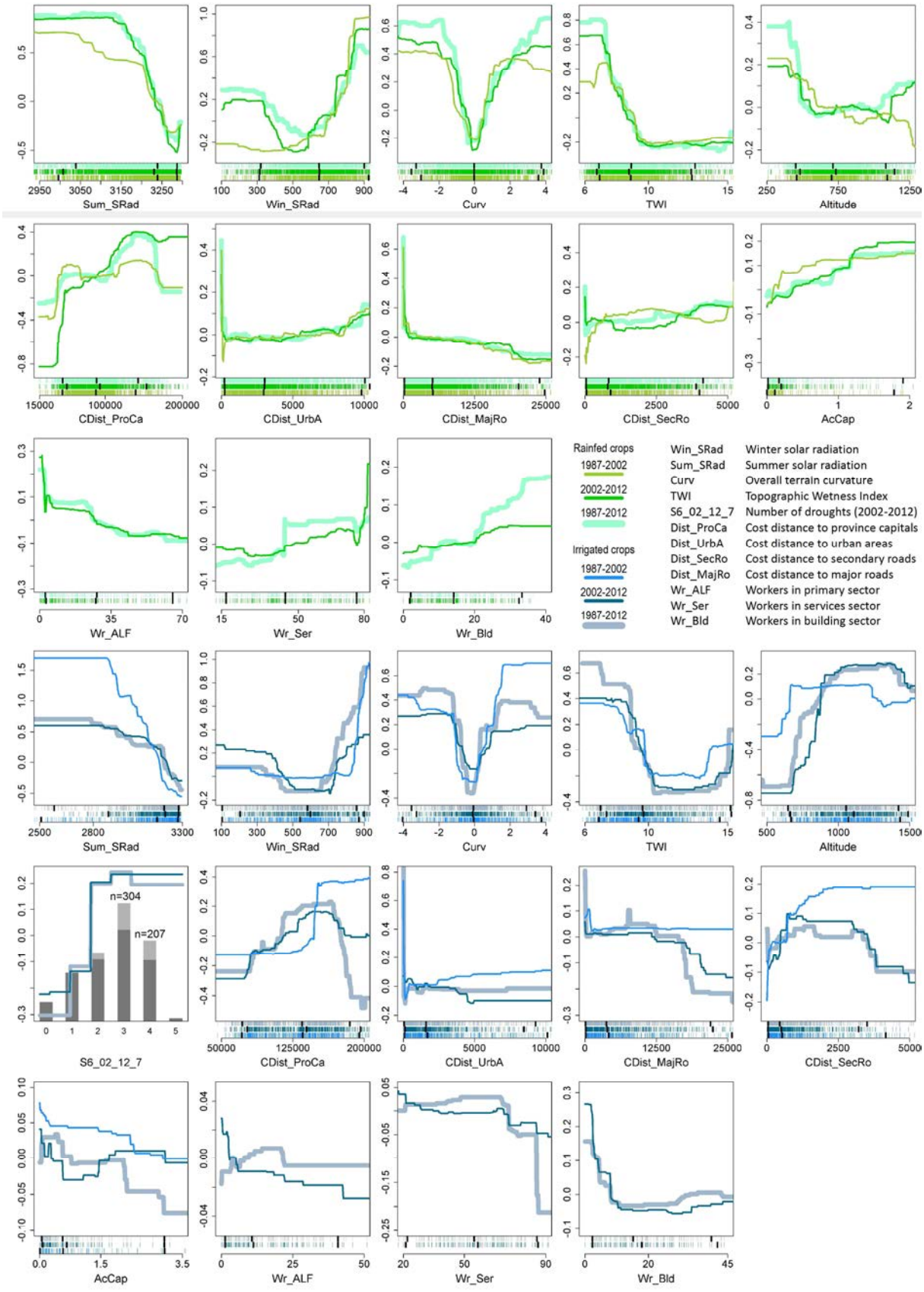


Figure 4. Dependence plots for agricultural abandonment in the Central Spanish Pyrenees. The density of presence events is represented by vertical lines above the x-axes, and overlaid solid black ticks depict the 0.05, 0.5 and 0.95 percentiles.

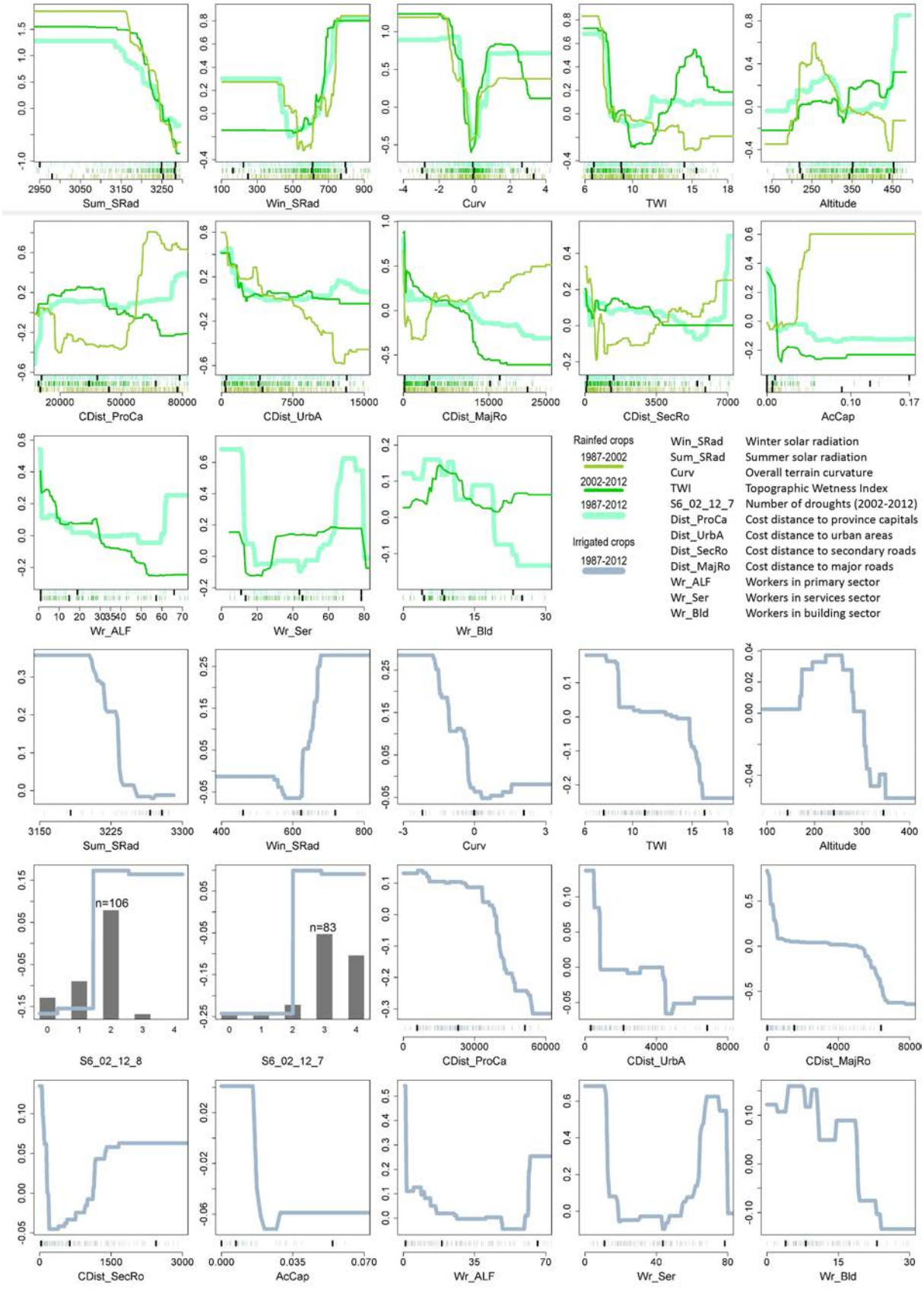


Figure 5. Dependence plots for agricultural abandonment in the Central Ebro Basin. The density of presence events is represented by vertical lines above the x-axes, and overlaid solid black ticks depict the 0.05, 0.5 and 0.95 percentiles.

To synthesize the results, our analysis focuses on the five most important environmental and socioeconomic variables (Figure A3 and Figure A4 of the Appendix), and places emphasis on those with interpretable response patterns. In addition, we also include relevant information from models that include the variables *Wr_ALF*, *Wr_InE*, *Wr_Bld* and *Wr_Ser* (Figure A5 and Figure A6). Explanations are based on the fitted function that lies between the percentiles 0.5 and 0.95 of predictor values in presence events.

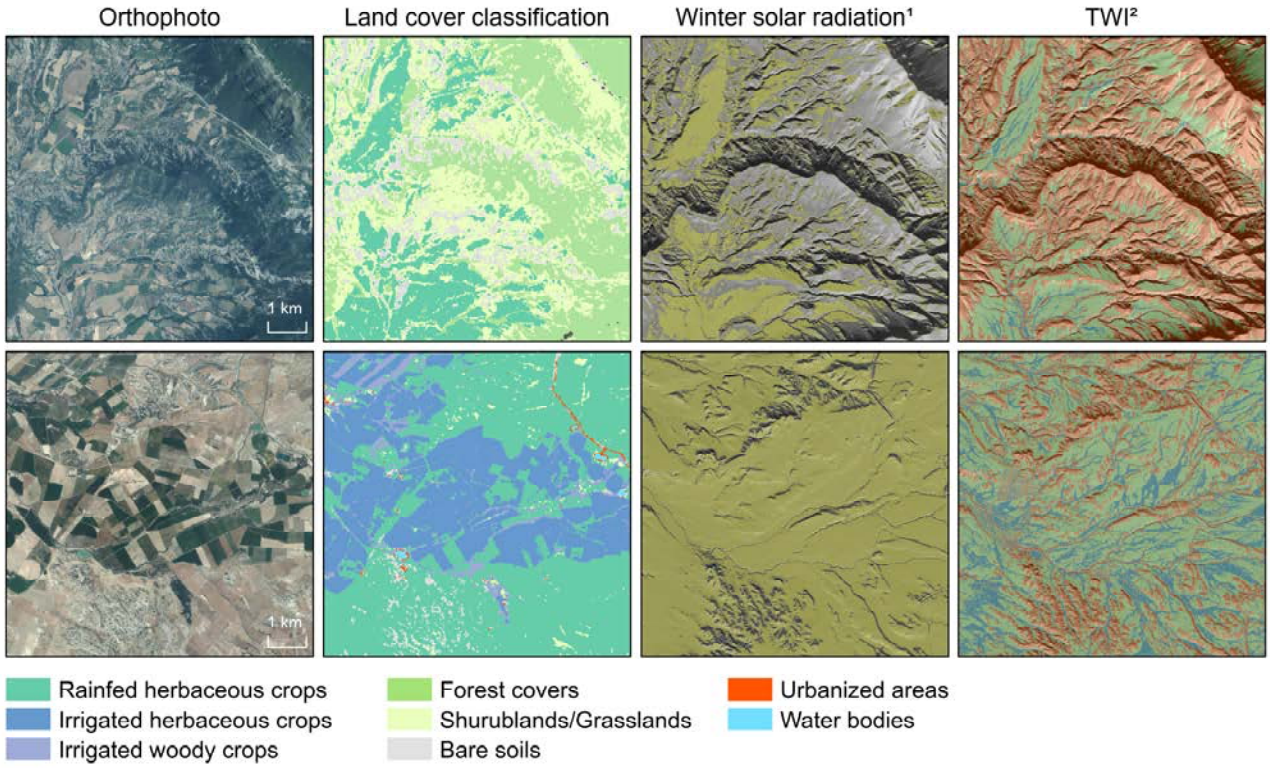


Figure 6. Land cover classification, solar radiation and TWI spatial distribution in a representative area of the Central Pyrenees (top row) and Central Ebro Basin (bottom row). ¹ *Win_SRad*: yellow color depicts low fitted values of land abandonment (between 500 and 700 10kJ·m^{-2}·day^{-1}). ² *TWI:* red color depicts high fitted values of land abandonment in all models (below 9), green color depicts low fitted values (from 9 to 12.5), and blue color depicts high fitted values of abandonment of rainfed crops in the Central Ebro Basin (above 12.5).

3.1. Environmental variables

Overall, the slope and terrain roughness (*VRM)* are the main explanatory variables in land abandonment. In models that do not include these two variables (Figure 4 and Figure 5), other environmental variables such as solar radiation (*Sum_SRad* and *Win_SRad*), terrain curvature (*Curv*) and Topographic Wetness Index (*TWI*) are the most important predictors, except for in the models of irrigated crops in the Central Ebro Basin, where *CDist_MajRo* retains 26% of the information. The response to these environmental variables is represented by patterns that depict the suitability of the terrain for agricultural activity (Gellrich and Zimmerman 2007, Stoebner and Lant, 2014, Pazur et al., 2014, Lieskovsky et al., 2014). The influence of slope, terrain roughness and curvature on agricultural abandonment is related to accessibility to the land and the constraints to mechanization, intensification and, in general, agricultural production. The *Curv*

dependence patterns for crops of the Central Basin (Figure 5) suggest that concave curvatures tend to be more abandoned, which can be related to the development of poorly drained soils and the salinity problems linked to the lithology of the Ebro Depression (Herrero and Synder 1997, Nogués and Herrero 2000, Pinilla 2006). In this area, the topographic effect on water runoff represented by the *TWI* has an influence on rainfed crops (2002-2012 period) that could be related to the former causes. Like the other cases, these crops tend to be abandoned when there are low *TWI* values (indicating that soil moisture favors agricultural production), although there is a threshold for the concentration of soil moisture above which the effects could be negative (Figure 6). Then, concave curvatures, which are prone to high soil moisture concentrations, could be an important constraint for rainfed crop production in the 2002-2012 period. The vulnerability to concave curvatures is also reflected in the irrigated crops of the Central Ebro Basin. On the other hand, flat valley bottom areas and gentle southern slopes are less likely to be abandoned in relation to the amount of incident solar radiation (Figure 6).

Finally, it is worth pointing out that variables quantifying drought occurrence have a higher relative influence on abandonment of irrigated crops than on rainfed crops in both study areas for the periods 2002-2012 and 1987-2012. The most important drought variables (within the 5 first environmental variables) are *S6_02_12_7*, with a relative importance of 6-6.5% in the Pyrenees, and *S6_02_12_7* and *S6_02_12_8*, with a relative importance of 6.2% and 6.9% respectively in the Central Ebro Basin. While for rainfed crops, the maximum relative importance of drought variables is below 2.5%. Looking at the dependence plots, this means that the occurrence of medium-term droughts (SPEI at a 6-month timescale) of at least 7-8 months' duration is positively associated with abandonment of irrigated crops during the last decade (2002-2012). We consider these results as evidence of the decrease in efficiency of agriculture in these areas (García-Ruiz et al., 2011). In the Mediterranean basin, in which the two study areas are located, an increase in temperatures has been reported while precipitation has remained stable or has decreased (López-Moreno et al., 2006; del Río et al., 2011, 2012). This has led to an increase in droughts (IPCC 2007, 2013; Vicente-Serrano 2014) and a consequent reduction in crop yields, as has also been reported for other regions (Craft et al., 2013). Similarly, it has resulted in a decrease in the water yield in the Ebro Basin (López-Moreno et al., 2011; García-Ruiz et al., 2011), which affects water supplies to irrigated lands. Moreover, high evaporation rates could increase salinity problems linked to irrigation (Metternicht & Zinck 2008).

3.2. Distance and socioeconomic variables

The influence of cost distance to roads and urban areas can be interpreted differently according to the study areas. In the Pyrenean region (Figure 4), remoteness from province capitals (*CDist_ProCa*), where external markets are located, is known to be one of the main determinants of land abandonment, which is reflected in the fitted functions for irrigated and rainfed crops. However, it should be noted that during the 1987-2002 period, the highest fitted values of abandonment for rainfed crops are related both to the nearest and the furthest distances from province capitals. Urban growth in other cities and villages (promoted by land-use conversion) seems to be the cause of the highest probabilities of abandonment for crops that are adjacent to built-up areas (*CDist_UrbA*). Outside this ring, land suitability tends to decrease as *CDist_UrbA* increases, especially for rainfed crops and during the 1987-2002 period for irrigated crops. This pattern is also reflected in the accessibility to secondary roads. It is reasonable to state therefore that these patterns reflect features of a local-based economy. The secondary road network connecting villages

with small cities in this mountainous region configures a small-scale socioeconomic network (Benayas et al., 2007, Osawa et al., 2016). Locational factors could be decisive in maintaining farm household incomes when the farming workforce is elderly and/or without successors (McDonald et al., 2000). Reducing costs leads to more isolated plots being abandoned or to land-use intensification around secondary roads and urban areas (Pointereau et al., 2008). On the other hand, the effect of isolation from province capitals and accessibility to major roads can be linked to large-scale forces of change (Benayas et al., 2007). Farm households in highlands are less influenced by modernization and agricultural competitiveness is lower due to the physical constraints. Alternatively, locations near major roads are more suitable for services and industrial activities connected with the main markets and, at the same time, they have more access to the demand from bigger cities. In this context, land owners are better able to adjust their economies to new opportunities.

The response to location factors in the Central Ebro Basin differs from that of the Pyrenees mainly due to the variability in patterns between periods. The results suggest that the land abandonment process in the Central Basin (Figure 5) has had at least two chronological stages, which is clearly manifested in the rainfed crop models. During 1982-2002, abandonment occurred mostly as cost distance to the province capital and roads increased, although the closest locations were also abandoned. In the 2002-2012 period, the pattern shifted and abandonment clearly increased as cost distance decreased. This could be explained by supposing that cropland abandonment was a gradual process where remote plots were first abandoned due to their unfavorable location with respect to the economic network. However, other socioeconomic factors must be taken into account. After Spain's accession to the European Union (EU) in 1986, the Central Ebro Basin was one of the areas most affected by land-retirement programs of the Common Agricultural Policy (CAP) during the 1990-2000 period (Meza and Albisu 1995; Comins and Navalpotro 1995, Baylis et al., 2008), which was a measure for production control and environmental protection. With these considerations, it is logical to suppose that less accessible plots would be the first ones to be abandoned. The attraction of the urban labor-market and new opportunities linked to favorable accessibility increased in importance during the 2002-2012 period as causes for croplands being abandoned, which is also reflected in irrigated crop patterns. This pattern can also be associated with the land conversion process, as is reported for other regions that have undergone urban expansion (Jiang et al. 2012, Liu et al. 2017). Thus, abandonment did not only occur in marginal areas but also in accessible locations (Gellrich et al., 2007, Hatna and Bakker 2011), which is also reflected in the *CDist_UrbA* plot. Access to secondary roads still plays an important role for irrigated crops, which may be due to the maintenance of irrigation infrastructures and production being more present in external markets (Pinilla 2006).

In the Central Pyrenees, there are significant differences between rainfed and irrigated crops regarding variables related to economic sectors. Abandonment of rainfed crops is clearly more likely in those municipalities with a higher proportion of population working in the service industry (*Wr_Ser*) and building sector (*Wr_Bld*). Moreover, the probability of abandonment increases as accommodation capacity increases (*AcCap*). Development of alternative sectors to farm activities could then replace the source of incomes in rural households (Melendez-Pastor et al., 2014). Patterns in irrigated crops are less clear and even an opposite trend can be detected. These results suggest that socioeconomic diversification could enhance the viability and the capacity to adapt of irrigated crop production, which is consistent with other studies (Dax et al., 1998, 2001; McDonald et al., 2000; Strijker 2005, Petrou et al., 2007). It could be considered that the value-added product of irrigated crops is higher than rainfed crops. As we mentioned before, irrigated crops in the Pyrenees are linked to livestock and the derived dairy products, which supply

the internal market (at a small-scale), but are also highly valued by the seasonal tourism. At the same time, the services and building sectors also benefit from tourism, and seasonal complementary activities to agriculture raise the incomes of farm households. Nonetheless, special attention should be paid to activities such as tourism in order to minimize the possible negative impacts on these traditional land-uses and the related cultural landscape (Marín-Yaselí and Lasanta 2003, Petrou et al., 2007); and like tourism, any land-use change or intensification should be developed on the basis of sustainable land-allocation decisions (Tenerelli and Carver 2012).

In the Central Basin, the development of the services sector seems to have led to an economic adaptation during the years after EU accession (1987-2002), which is reflected in a higher probability of abandonment of rainfed crops as accommodation capacity (*AcCap*) increases (Figure 5). On the other hand, the building sector (*Wr_Bld*) seems to have a significant weight in maintaining agricultural activity for both irrigated and rainfed crops, while new opportunities linked to the services sector (*Wr_Ser*) could promote agricultural abandonment. In other words, the economy of agricultural households could depend on part-time jobs related to the urban growth of the last decades and be completely re-adjusted through full-time work in the services sector. However, the U-shaped response pattern suggests that very low proportions (below the 0.5 percentile) of employees in the services sector are also associated with high land abandonment rates (Gellrich and Zimmermann 2007) so that the least diversified municipalities are less able to maintain primary sector activities.

4. CONCLUSIONS

Agricultural abandonment in northeast Spain has been analyzed using BRT multivariate models. The study differentiates between abandonment of rainfed and irrigated crops in two different territorial contexts. Physical constraints derived from topography variables were shown to be main determinants of abandonment in all models except for irrigated crops in the Central Ebro Basin, where accessibility to major roads and province capitals are the factors that contribute most. BRT fitted functions have also allowed us to gain insights into the role played by other environmental and socioeconomic variables and to detect pattern shifts between the analyzed crops and periods.

Our results suggest that the increase in drier conditions predicted for the future could lead to lower yields in irrigated lands, especially in the Central Ebro Basin, where there has been a large investment into agricultural intensification through modern irrigation systems. Moreover, vegetation re-growth after abandonment of traditional activities in the Pyrenees decreases water run-off, which reduces reservoir storage and the water yield of the Ebro Basin (García-Ruiz et al., 2011; López-Moreno 2007, 2011, 2013). Thus, planning and land-management strategies should be efficiently designed in order to reduce environmental and social costs (Fernández-Comuñas and Arrojo 2002; García-Ruiz et al., 2011).

In a general way, land suitability in terms of other environmental factors is reflected by similar patterns for the two types of crops, although whether conditions are more or less favorable may vary depending on the interactions between variables (such as the response to terrain curvature and the Topographic Wetness Index) and other site features not accounted for in the study (e.g. lithology).

Socioeconomic conditions defining where abandonment occurs can be interpreted as due to diverse underlying causes, which in many cases are reflected in polarized patterns. From our results, it appears that

agricultural activity in the Pyrenees is more dependent on accessibility to internal markets and networks, especially for rainfed crops. However, off-farm opportunities linked to favorable locations also seem to promote land abandonment. This is clearly manifested in the Central Ebro Basin where, in addition, an inflexion point between the two periods suggests that during the 2002-2012 period there was a readjustment of rural household economies (those where abandonment occurred) towards less traditional roles that are more dependent on the infrastructures and urban networks. The greater presence of CAP aids during the period 1987-2002 could explain this pattern shift, especially in the case of winter cereals, which are generally more dependent on these policies. In this sense, the integration of other socioeconomic determinants as, for example, the crop prices variability (Stoebner and Lant, 2014), could improve land-abandonment models, helping to interpret temporal and spatial pattern shifts. On the other hand, this study points out that economic diversification could contribute to the maintenance of cropland and, in general, to embed professional sectors into the rural context. This can be decisive in mountainous regions like the Pyrenees, where agriculture is more vulnerable to remoteness from external markets and the main urban areas.

Land abandonment is not always characterized by a linear relationship with the determinant factors, and it has been shown that this is even more evident when different types of land cultivation and territorial contexts are compared. Whatever the ultimate objectives in managing the rural environment, multidisciplinary approaches must be followed in order to gain an overall view, taking into account that a diverse range of territorial scales and thematic areas could be affected by the decisions made. The evaluations and perceptions of the inhabitants of these areas should be included in the decision-making process.

ACKNOWLEDGEMENTS

This work was supported by the Spanish Ministry of Economy and Competitiveness [grant number BES-2013-063766 to Juan José Vidal-Macua]; European Union's Horizon 2020 Programme [ECOPOTENTIAL (641762-2)]; Spanish Ministry of Economy and Competitiveness [ACAPI (CGL2015-69888-P MINECO/FEDER)]; Catalan Government [SGR2014-1491]. Xavier Pons is recipient of an ICREA Academia Excellence in Research Grant (2016–2020).

APPENDIX

	Rainfed herbaceous crops			Irrigated herbaceous crops		
Pyrenees	1987-2002	2002-2012	1987-2012	1987-2002	2002-2012	1987-2012
Including Slope and VRM	0.728	0.781	0.760	0.779	0.775	0.798
Not including Slope and VRM	0.702	0.754	0.731	0.757	0.747	0.771
Including Workers by sector*		0.755	0.733		0.747	0.770
Central Ebro Valley						
Including Slope and VRM	0.867	0.867	0.833			0.767
Not including Slope and VRM	0.847	0.832	0.811			0.742
Including workers by sector*		0.824	0.815			0.727

Table A1. AUC score by model. Rejected models in grey. * *Wr_ALF*, *Wr_InE*, *Wr_Bld* and *Wr_Ser* variables

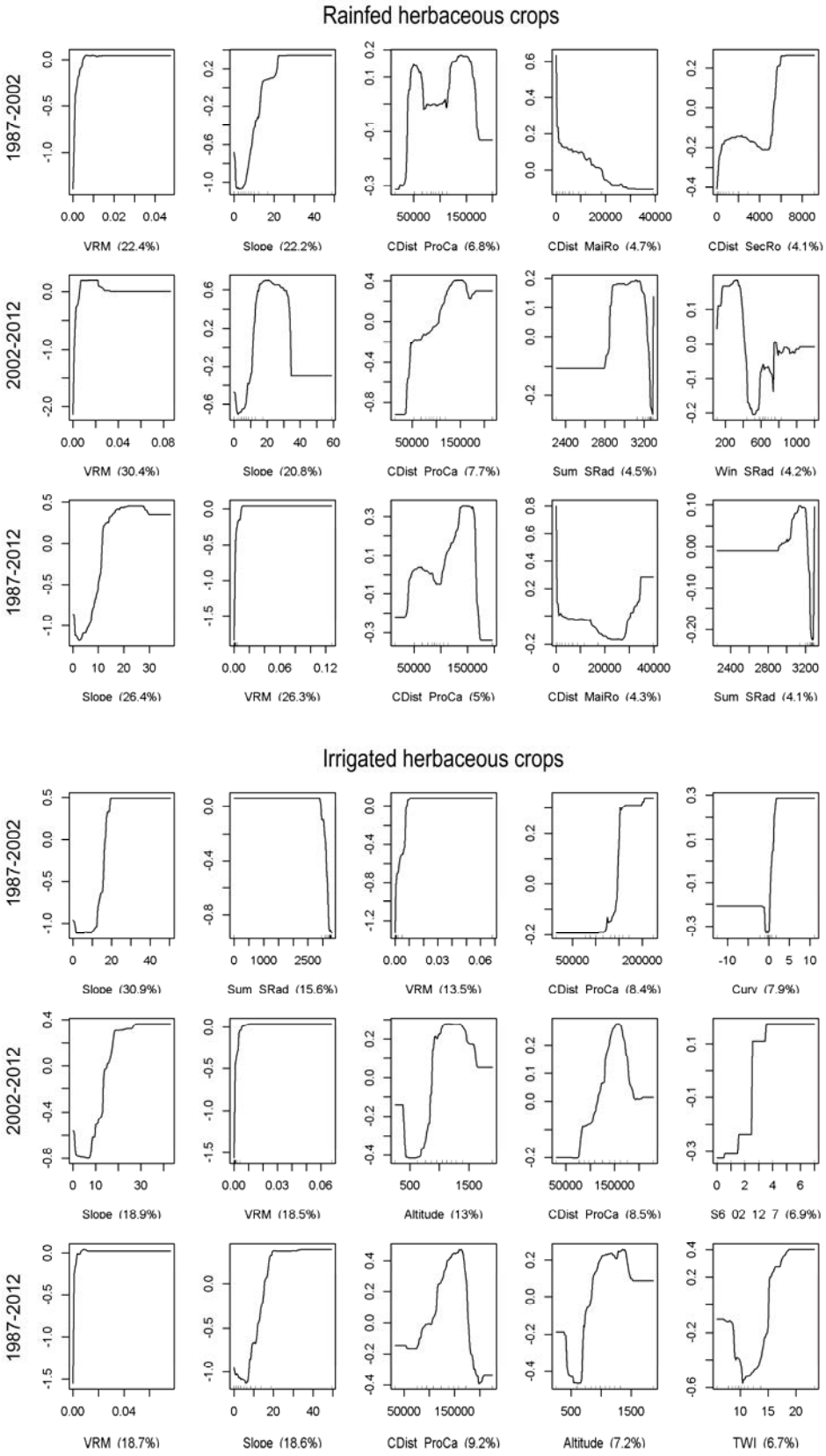


Figure A1. Dependence plots and relative importance for the five most important variables in models that include Slope and VRM. Central Spanish Pyrenees

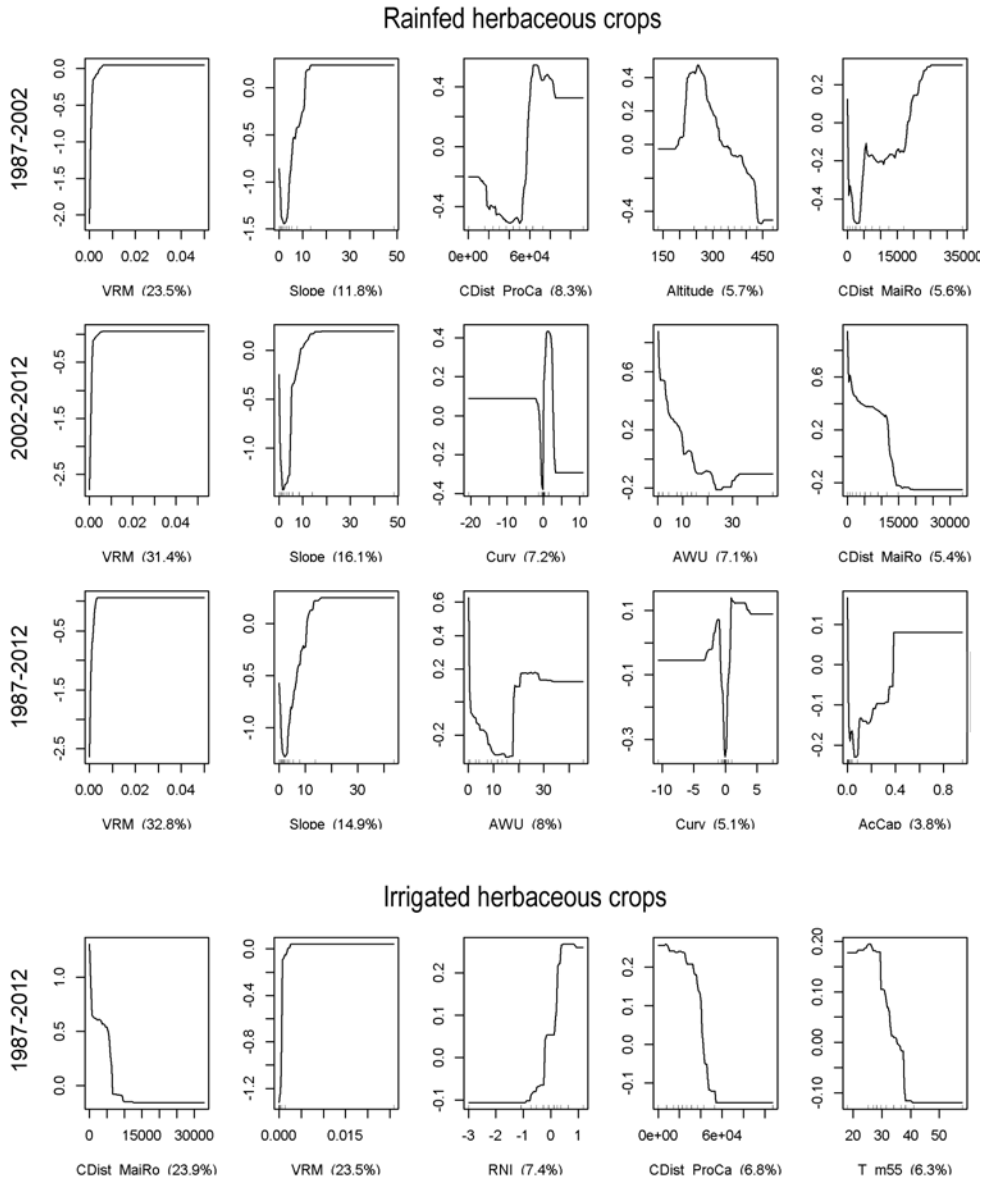


Figure A2. Dependence plots and relative importance for the five most important variables in models that include Slope and VRM. Central Ebro Basin

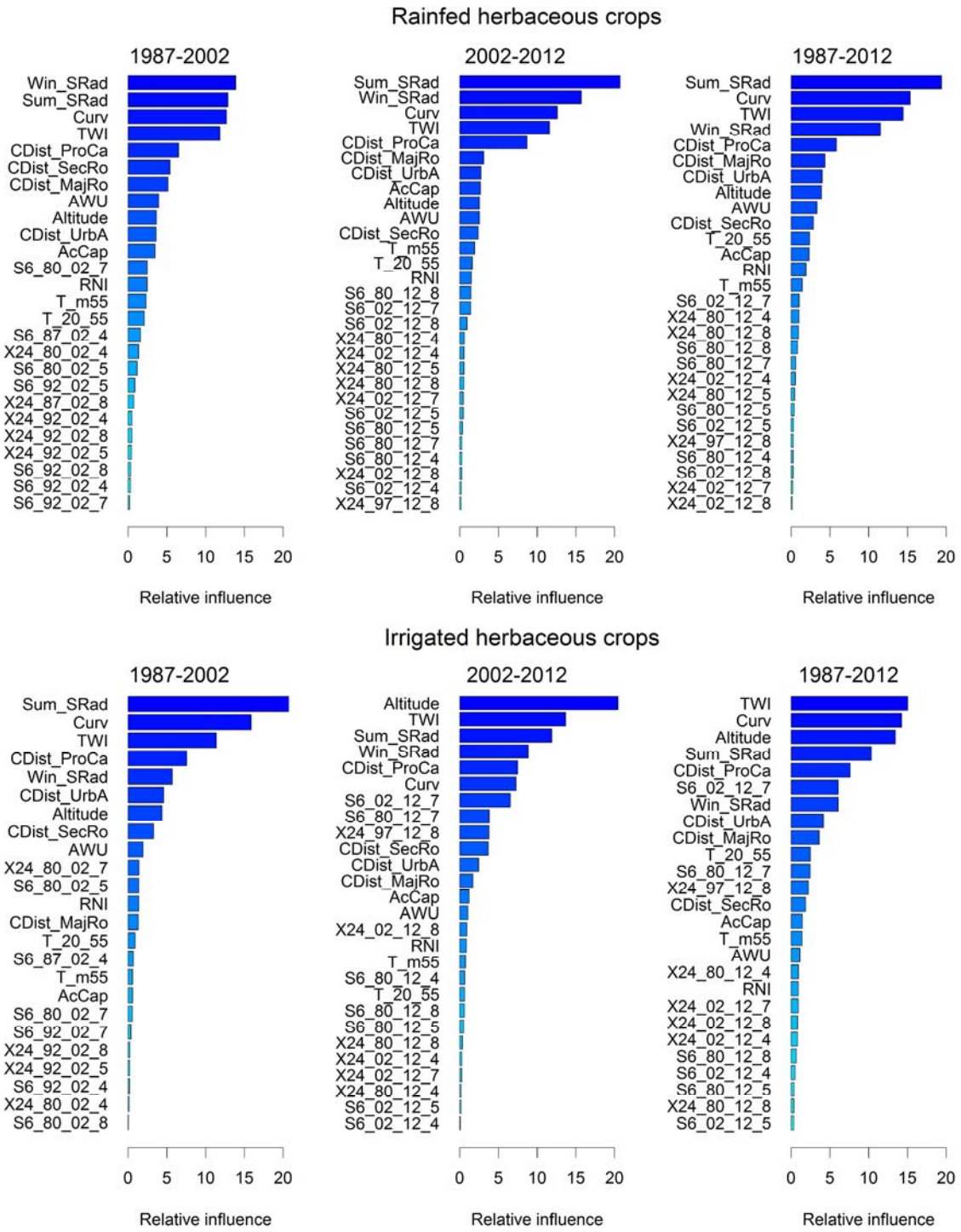


Figure A3. Relative importance of variables in models that do not include Slope and VRM. Central Spanish Pyrenees

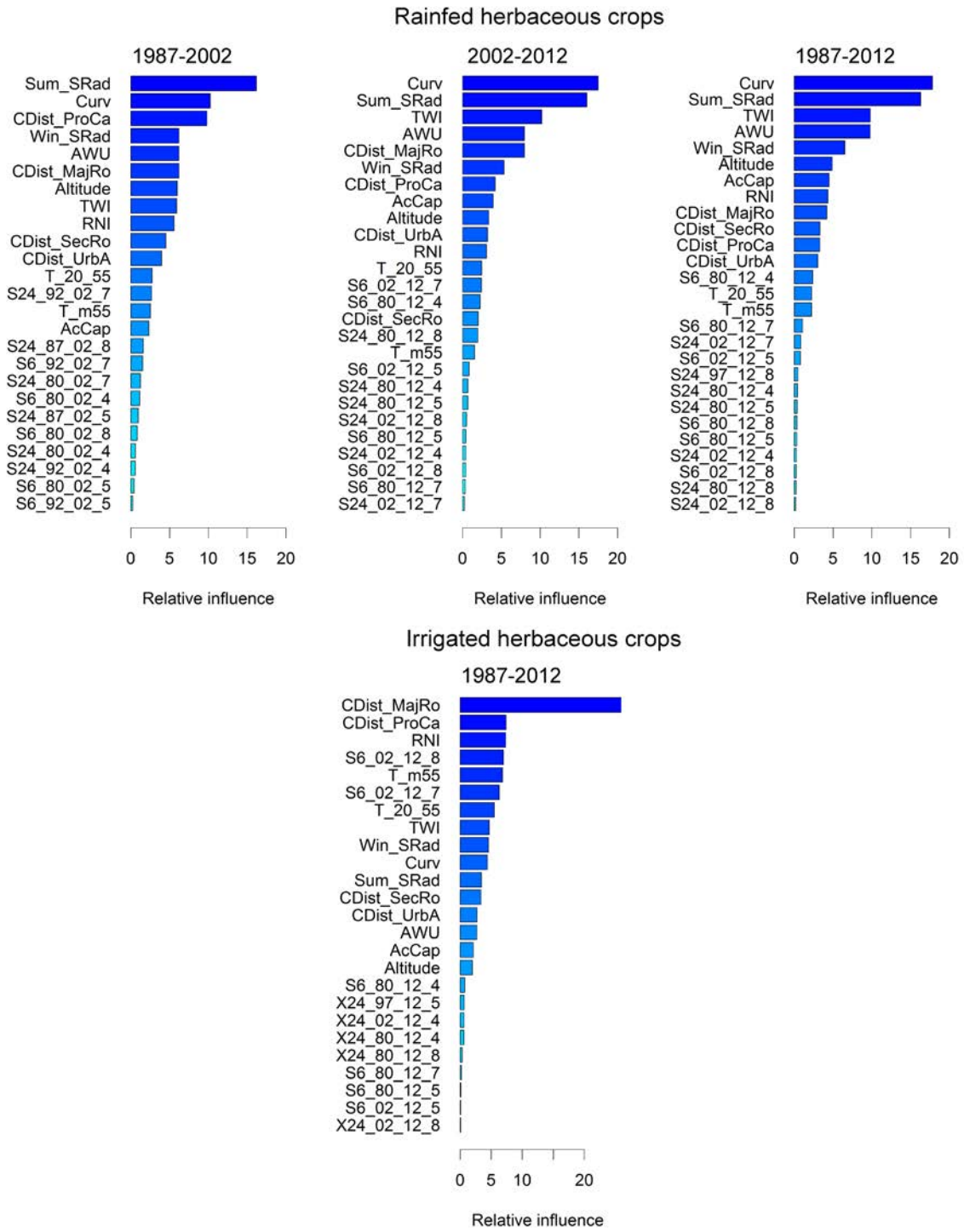


Figure A4. Relative importance of variables in models that do not include Slope and VRM. Central Ebro Basin.

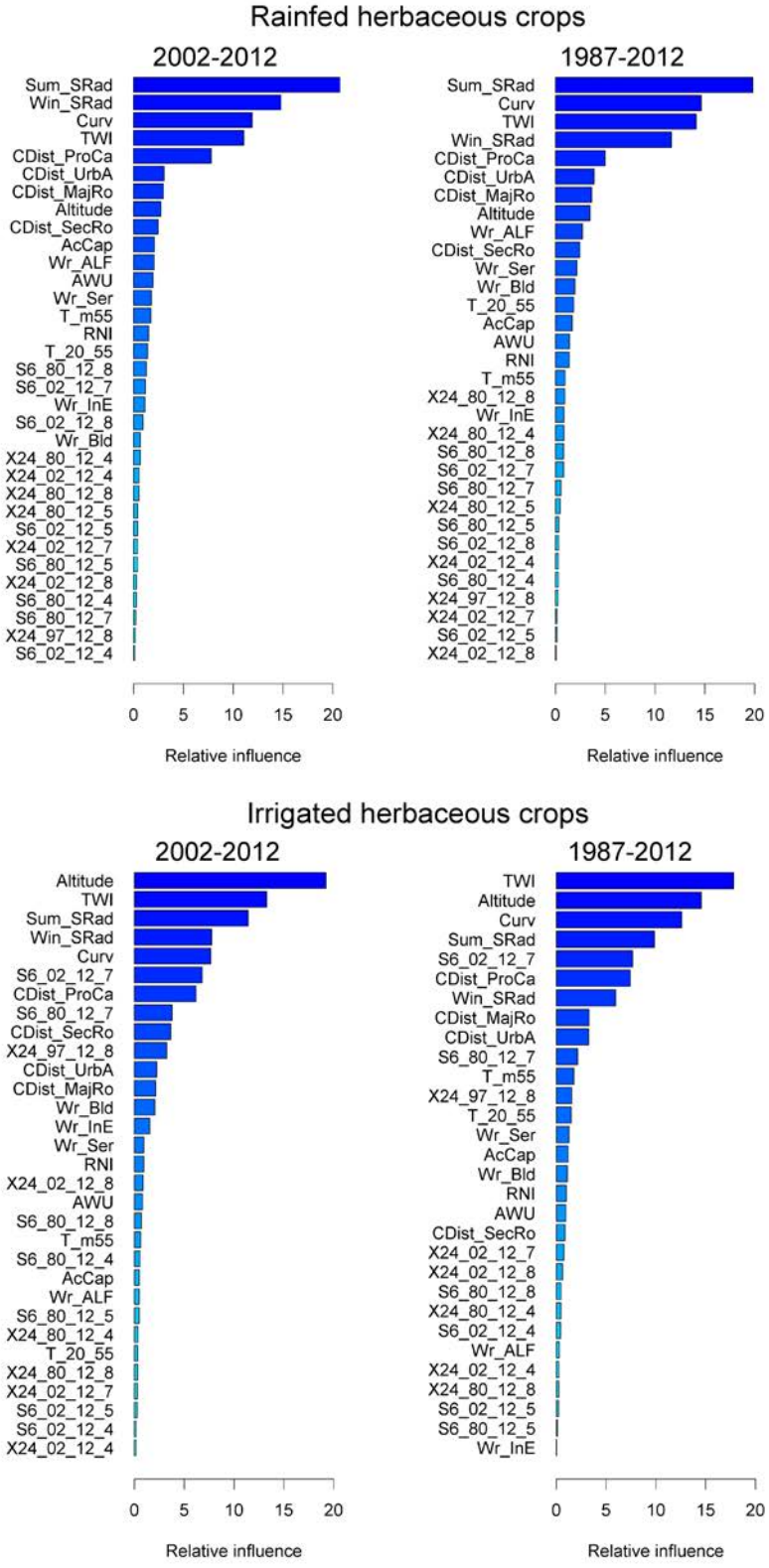


Figure A5. Relative importance of variables in models that include the variables *Wr_ALF*, *Wr_InE*, *Wr_Bld* and *Wr_Ser*. Central Spanish Pyrenees

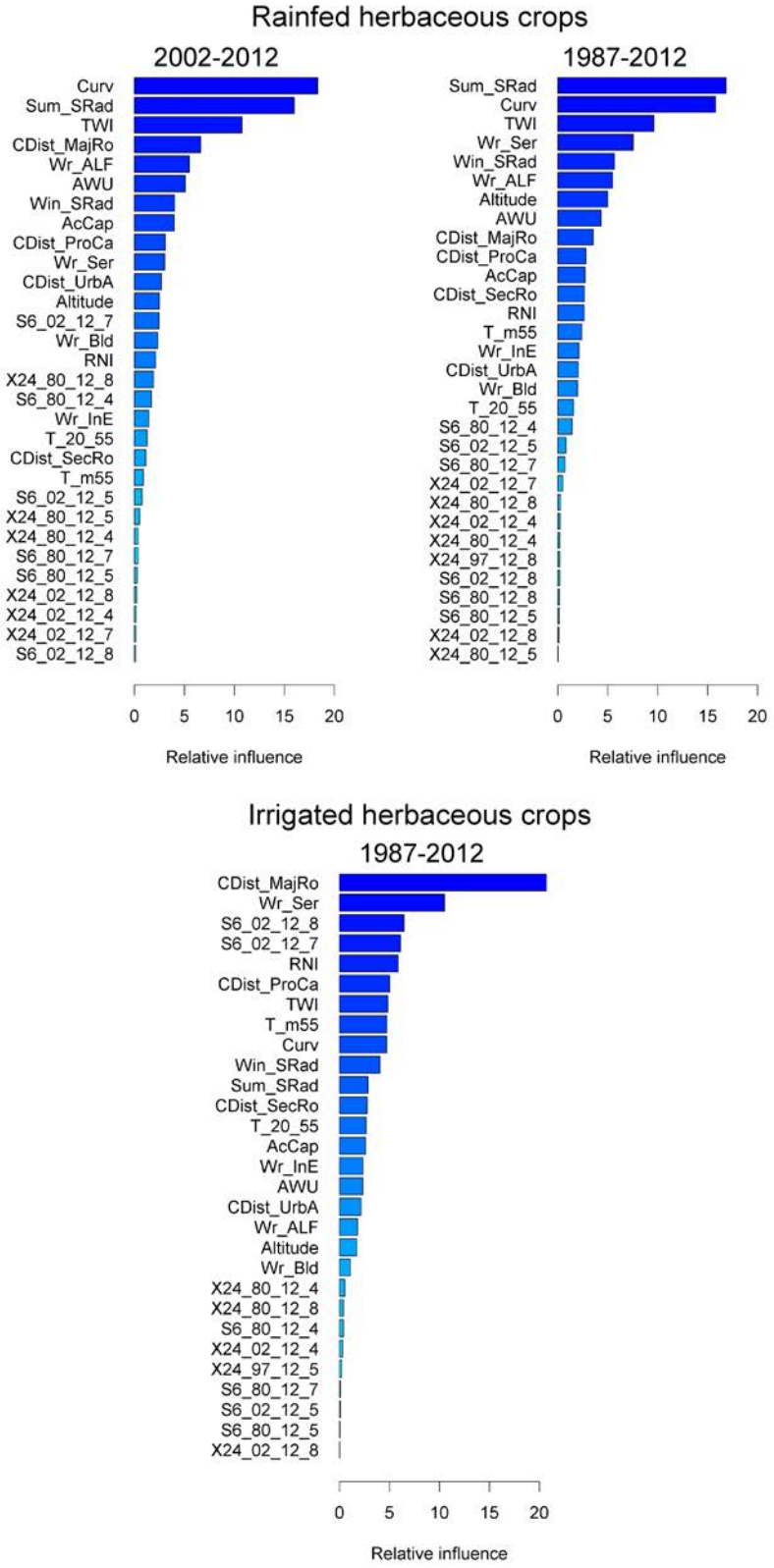


Figure A6. Relative importance of variables in models that include the variables *Wr_ALF*, *Wr_InE*, *Wr_Bld* and *Wr_Ser*. Central Ebro Basin

REFERENCES

Álvarez-Martínez, J. M., Stoorvogel, J. J., Suárez-Seoane, S., & de Luis Calabuig, E. (2010). Uncertainty analysis as a tool for refining land dynamics modelling on changing landscapes: a case study in a Spanish Natural Park. *Landscape Ecology*, *25*(9), 1385-1404.

Baylis, K., Peplow, S., Rausser, G., & Simon, L. (2008). Agri-environmental policies in the EU and United States: A comparison. *Ecological Economics*, *65*(4), 753-764.

Beilin, R., Lindborg, R., Stenseke, M., Pereira, H. M., Llausàs, A., Slätmo, E., et al. (2014). Analysing how drivers of agricultural land abandonment affect biodiversity and cultural landscapes using case studies from Scandinavia, Iberia and Oceania. *Land Use Policy*, *36*, 60-72.

Benayas, J. R., Martins, A., Nicolau, J. M., & Schulz, J. J. (2007). Abandonment of agricultural land: an overview of drivers and consequences. *CAB reviews: Perspectives in agriculture, veterinary science, nutrition and natural resources*, *2*(57), 1-14.

Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Journal*, *24*(1), 43-69.

Bezák, P., & Halada, L. (2010). Sustainable management recommendations to reduce the loss of agricultural biodiversity in the mountain regions of NE Slovakia. *Mountain Research and Development*, *30*(3), 192-204.

Bielsa I., Pons X., Bunce B. (2005) Agricultural abandonment in the North Eastern Iberian Peninsula: the use of basic landscape metrics to support planning. *Journal of Environmental Planning and Management, 48* (1), 85-102.

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author. *Statistical science 16*(3), 199-231.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification and regression trees. Wadsworth & Brooks, Monterey.

Burrough, P.A., van Gaans, P.F. & MacMillan, R.A. (2000). High-resolution landform classification using fuzzy k-means. *Fuzzy Sets and Systems, 113*(1), 37-52.

Cánoves, G., Villarino, M., Priestley, G. K., & Blanco, A. (2004). Rural tourism in Spain: an analysis of recent evolution. *Geoforum*, *35*(6), 755-769.

Cohen, M., Varga, D., Vila, J., Barrassaud, E. (2011). A multi-scale and multi-disciplinary approach to monitor landscape dynamics: a case study in the Catalan pre-Pyrenees (Spain). *The Geographical Journal, 177*(1), 79-91.

Collantes, F., & Pinilla, V. (2004). Extreme depopulation in the Spanish rural mountain areas: a case study of Aragon in the nineteenth and twentieth centuries. *Rural History*, *15*(2), 149-166.

Comíns, J. S., & Jiménez, P. N. (1995). La extensificación de los secanos herbáceos españoles: Efectos territoriales y sociales de la nueva PAC De la Unión Europea. *Estudios Geograficos*, *56*(221), 717.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J. (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4, Geoscientific Model Development.

Craft, K. E., Mahmood, R., King, S. A., Goodrich, G., & Yan, J. (2013). Drought and corn in Kentucky. *Applied Geography*, *45*, 353-362.

Crase, B., Liedloff, A. C., & Wintle, B. A. (2012). A new method for dealing with residual spatial autocorrelation in species distribution models. *Ecography*, *35*(10), 879-888.

Dax, T. (Ed.). (1998). *Mountain farming and the environment: towards integration; perspectives for mountain policies in Central and Eastern Alps*. Bundesanst. für Bergbauernfragen.

Dax, T., Loibl, E., & Oedl-Wieser, T. (1995). Pluriactivity and rural development. Theoretical Framework. Forschungsbericht, (35).

Domingo-Marimon, C. (2016). Contributions to the knowledge of the multitemporal spatial patterns of the Iberian Peninsula droughts from a Geographic Information Science perspective. Ph.D. thesis, Universitat Autònoma de Barcelona.

Duarte, R., Pinilla, V., & Serrano, A. (2014). The water footprint of the Spanish agricultural sector: 1860–2010. *Ecological economics*, *108*, 200-207.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S.& Zimmermann, N.E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, *29*, 129–151.

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802-813.

Environmental Systems Research Institute (ESRI) (2012). ArcGIS Release 10.1. Redlands.

Farina, A. (2008). *Principles and methods in landscape ecology: towards a science of the landscape* (Vol. 3). Springer Science & Business Media.

Fawcett, T. (2004). ROC graphs: Notes and practical considerations for researchers. *Machine learning*, *31*(1), 1-38.

Fernández-Comuñas, J. and Arrojo, P. (2000). Biscarrués-Mallos de Riglos. Inundación o modernización. Zaragoza: Egido editorial.

Fielding, A. J. (1989). Migration and urbanization in Western Europe since 1950. *The Geographical Journal*, *155*(1), 60-69.

Freund, Y., & Schapire, R. E. (1996). Experiments with a new boosting algorithm. International Conference of Machine Learning. New Jersey, 22 January 1996.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics: 1189-1232.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*(4), 367-378.

Friedman, J. H., & Meulman, J. J. (2003). Multiple additive regression trees with application in epidemiology. *Statistics in Medicine*, *22*(9), 1365-1381.

Gallart, F., & Llorens, P. (2004). Observations on land cover changes and water resources in the headwaters of the Ebro catchment, Iberian Peninsula. *Physics and Chemistry of the Earth, Parts A/B/C*, *29*(11), 769-773.

García-Ruiz, J. M. (2010). The effects of land uses on soil erosion in Spain: a review. *Catena*, *81*(1), 1-11.

García-Ruiz, J. M. (1990). Geoecología de las Áreas de Montaña. Geoforma Ediciones, Logroño.

García-Ruiz, J. M., & Lasanta, T. (1993). Land-use conflicts as a result of land-use change in the Central Spanish Pyrenees: A review. *Mountain Research and Development*, 295-304.

Garcia-Ruiz, J. M., Lasanta, T., Ruiz-Flano, P., Ortigosa, L., White, S., González, C., & Martí, C. (1996). Land-use changes and sustainable development in mountain areas: a case study in the Spanish Pyrenees. *Landscape Ecology*, *11*(5), 267-277.

García-Ruiz, J. M., López-Moreno, J. I., Vicente-Serrano, S. M., Lasanta, T., & Beguería, S. (2011). Mediterranean water resources in a global change scenario. *Earth-Science Reviews*, *105*(3), 121-139.

García-Ruiz, J. M., López-Moreno, J. I., Vicente-Serrano, S. M., Lasanta, T., & Beguería, S. (2011). Mediterranean water resources in a global change scenario. *Earth-Science Reviews*, *105*(3), 121-139.

Gellrich, M., & Zimmermann, N. E. (2007). Investigating the regional-scale pattern of agricultural land abandonment in the Swiss mountains: a spatial statistical modelling approach. *Landscape and Urban Planning*, *79*(1), 65-76.

Gellrich, M., Baur, P., Koch, B., & Zimmermann, N. E. (2007). Agricultural land abandonment and natural forest re-growth in the Swiss mountains: a spatially explicit economic analysis. *Agriculture, Ecosystems & Environment*, *118*(1), 93-108.

Gerard, F., Petit, S., Smith, G., Thomson, A., Brown, N., Manchester, S., ... & Boltiziar, M. (2010). Land cover change in Europe between 1950 and 2000 determined employing aerial photography. *Progress in Physical Geography*, *34*(2), 183-205.

Halada, L., Evans, D., Romão, C., & Petersen, J. E. (2011). Which habitats of European importance depend on agricultural practices?. *Biodiversity and Conservation*, *20*(11), 2365-2378.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC curve). *Radiology, 143*(1), 29-36.

Hastie, T., Tibshirani, R.I. (2009). The elements of statistical learning: data mining, inference and prediction. Springer, Berlin

Hatna, E., & Bakker, M. M. (2011). Abandonment and expansion of arable land in Europe. *Ecosystems*, *14*(5), 720-731.

Herrero, J., & Snyder, R. L. (1997). Aridity and irrigation in Aragon, Spain. *Journal of arid environments*, *35*(3), 535-547.

IAEST. Instituto Aragonés de Estadística. Censos Agrarios. (2017). http://www.aragon.es/iaest

IDESCAT. Institut d'Estadística de Catalunya. Censos Agraris. (2017). https://www.idescat.cat/

IPCC. 2007. Climate change 2007: the physical science basis. Contribution of Working Group I to the forth assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge.

IPCC. 2013. Climate change 2013: the physical science basis. In: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge and New York.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. Springer, New York.

Jiang, L., Deng, X., & Seto, K. C. (2012). Multi-level modeling of urban expansion and cultivated land conversion for urban hotspot counties in China. *Landscape and Urban Planning, 108*(2), 131-139.

Kawakita, M., Minami, M., Eguchi, S., & Lennert-Cody, C. E. (2005). An introduction to the predictive technique AdaBoost with a comparison to generalized additive models. *Fisheries research*, *76*(3), 328-343.

Keenleyside, C., Tucker, G., & McConville, A. (2010). Farmland Abandonment in the EU: an Assessment of Trends and Prospects. *Institute for European Environmental Policy, London*.

Kopecký, M., & Čížková, Š. (2010). Using topographic wetness index in vegetation ecology: does the algorithm matter?. *Applied Vegetation Science, 13*(4), 450-459.

Kosmas, C., Kairis, O., Karavitis, C., Acikalin, S., Alcalá, M., Alfama, P., ... & Brito, J. (2015). An exploratory analysis of land abandonment drivers in areas prone to desertification. *Catena*, *128*, 252-261.

Lasanta, T., Arnáez, J., Pascual, N., Ruiz-Flaño, P., Errea, M. P., & Lana-Renault, N. (2017). Space–time process and drivers of land abandonment in Europe. *Catena*, *149*, 810-823.

Lasanta, T., Vicente-Serrano, S. M., & Cuadrat-Prats, J. M. (2005). Mountain Mediterranean landscape evolution caused by the abandonment of traditional primary activities: a study of the Spanish Central Pyrenees. *Applied Geography*, *25*(1), 47-65.

Leathwick, J. R., Elith, J., Francis, M. P., Hastie, T., & Taylor, P. (2006). Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology Progress Series*, *321*, 267-281.

Lewis, H. G., Brown, M., & Tatnall, A. R. L. (2000). Incorporating uncertainty in land cover classification from remote sensing imagery. *Advances in Space Research*, *26*(7), 1123-1126.

Lieskovský, J., Kenderessy, P., Špulerová, J., Lieskovský, T., Koleda, P., Kienast, F., & Gimmi, U. (2014). Factors affecting the persistence of traditional agricultural landscapes in Slovakia during the collectivization of agriculture. *Landscape ecology*, *29*(5), 867-877.

Liu, X., Wang, Y., Li, Y., Liu, F., Shen, J., Wang, J., Xiao, R. & Wu, J. (2017). Changes in arable land in response to township urbanization in a Chinese low hilly region: Scale effects and spatial interactions. *Applied Geography, 88*, 24-37.

Lloret, F., Calvo, E., Pons, X., & Díaz-Delgado, R. (2002). Wildfires and landscape patterns in the Eastern Iberian Peninsula. *Landscape Ecology*, *17*(8), 745-759.

López-Moreno, J. I., & García-Ruiz, J. M. (2007). Hydrological effects of reservoirs in the Central Spanish Pyrenees. *Robinson P, Jones JAA, Woo MK (eds)*, 103-114.

López-Moreno, J. I., Beguería, S., & García-Ruiz, J. M. (2006). Trends in high flows in the central Spanish Pyrenees: response to climatic factors or to land-use change?. *Hydrological Sciences Journal*, *51*(6), 1039-1050.

López-Moreno, J. I., Vicente-Serrano, S. M., Moran-Tejeda, E., Zabalza, J., Lorenzo-Lacruz, J., & García-Ruiz, J. M. (2011). Impact of climate evolution and land use changes on water yield in the Ebro basin. *Hydrology and Earth System Sciences*, *15*(1), 311.

López-Moreno, J. I., Vicente-Serrano, S. M., Zabalza, J., Beguería, S., Lorenzo-Lacruz, J., Azorin-Molina, C., & Morán-Tejeda, E. (2013). Hydrological response to climate variability at different time scales: A study in the Ebro basin. *Journal of hydrology*, *477*, 175-188.

MacDonald, D., Crabtree, J. R., Wiesinger, G., Dax, T., Stamou, N., Fleury, P., et al. (2000). Agricultural abandonment in mountain areas of Europe: environmental consequences and policy response. *Journal of environmental management*, *59*(1), 47-69.

Marín-Yaseli, M. L., & Lasanta, T. (2003). Competing for meadows: A case study on tourism and livestock farming in the Spanish Pyrenees. *Mountain Research and Development*, *23*(2), 169-176.

Melendez-Pastor, I., Hernández, E. I., Navarro-Pedreño, J., & Gómez, I. (2014). Socioeconomic factors influencing land cover changes in rural areas: The case of the Sierra de Albarracín (Spain). *Applied Geography, 52*, 34-45.

Metternicht, G., & Zinck, A. (2008). Remote sensing of soil salinization: Impact on land management. CRC Press.

Meza, L., & Albisu, L. M. (1996). Aspectos economicos de la retirada de tierras en Aragon. CSIC, Zaragoza, 1995, por Sebastian Llompart. *Agricultura y Sociedad*, 239-240.

Molinillo, M., Lasanta, T., & García-Ruiz, J. M. (1997). Managing mountainous degraded landscapes after farmland abandonment in the Central Spanish Pyrenees. *Environmental Management*, *21*(4), 587-598.

Moreira, F., Viedma, O., Arianoutsou, M., Curt, T., Koutsias, N., Rigolot, E., ... & Mouillot, F. (2011). Landscape–wildfire interactions in southern Europe: implications for landscape management. *Journal of environmental management*, *92*(10), 2389-2402.

Nainggolan, D., de Vente, J., Boix-Fayos, C., Termansen, M., Hubacek, K., & Reed, M. S. (2012). Afforestation, agricultural abandonment and intensification: competing trajectories in semi-arid Mediterranean agro-ecosystems. *Agriculture, ecosystems & environment, 159*, 90-104.

Navarro, L. M., & Pereira, H. M. (2015). Rewilding abandoned landscapes in Europe. In *Rewilding European Landscapes* (pp. 3-23). Springer International Publishing.

Nogués, J., Herrero, J., Rodríguez-Ochoa, R., & Boixadera, J. (2000). Land evaluation in a salt-affected irrigated district using an index of productive potential. *Environmental management, 25*(2), 143-152.

Noun Project (2017). Icons for everything. https://thenounproject.com/

O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity, 41*(5), 673-690.

Osawa, T., Kohyama, K., & Mitsuhashi, H. (2016). Multiple factors drive regional agricultural abandonment. *Science of the Total Environment, 542*, 478-483.

Pazúr, R., Lieskovský, J., Feranec, J., & Oťaheľ, J. (2014). Spatial determinants of abandonment of large-scale arable lands and managed grasslands in Slovakia during the periods of post-socialist transition and European Union accession. *Applied Geography*, *54*, 118-128.

Petrou, A., Pantziou, E. F., Dimara, E., & Skuras, D. (2007). Resources and activities complementarities: the role of business networks in the provision of integrated rural tourism. *Tourism Geographies, 9*(4), 421-440.

Petrou, A., Pantziou, E. F., Dimara, E., & Skuras, D. (2007). Resources and activities complementarities: the role of business networks in the provision of integrated rural tourism. *Tourism Geographies*, *9*(4), 421-440.

Pinilla, V. (2006). The development of irrigated agriculture in twentieth-century Spain: a case study of the Ebro basin. *Agricultural History Review*, *54*(1), 122-141.

Pointereau, P. (2008). Analysis of farmland abandonment and the extent and location of agricultural areas that are actually abandoned or are in risk to be abandoned. EUR-OP.

Pons X. (2006). MiraMon. Geographical Information System and Remote Sensing Software. Bellaterra: CREAF Research Center.

Pons, X., & Ninyerola, M. (2008). Mapping a topographic global solar radiation model implemented in a GIS and refined with ground data. *International Journal of Climatology, 28*(13), 1821-1834.

Pons, X., Serra, P., & Saurí, D. (2003). A rigorous protocol for post-classification land cover and land use change detection. Geoinformation for European-wide integration. Millpress Science, Rotterdam.

Pontius, R. G., & Parmentier, B. (2014). Recommendations for using the relative operating characteristic (ROC). *Landscape Ecology, 29*(3), 367-382.

Ramankutty, N., & Foley, J. A. (1999). Estimating historical changes in global land cover: Croplands from 1700 to 1992. *Global biogeochemical cycles*, *13*(4), 997-1027.

Ridgeway, G. (2007). Generalized Boosted Models: A guide to the gbm package. Update 1(1): 2007.

Ridgeway, G., & Ridgeway, M. G. (2004). The gbm package. R Foundation for Statistical Computing, Vienna, Austria 5(3).

Roura-Pascual, N., Pons, P., Etienne, M., & Lambert, B. (2005). Transformation of a rural landscape in the Eastern Pyrenees between 1953 and 2000. *Mountain Research and Development, 25*(3), 252-261.

Sang, N., Dramstad, W. E., & Bryn, A. (2014). Regionality in Norwegian farmland abandonment: Inferences from production data. *Applied Geography*, *55*, 238-247.

Sappington, J. M., Longshore, K. M., & Thompson, D. B. (2007). Quantifying landscape ruggedness for animal habitat analysis: a case study using bighorn sheep in the Mojave Desert. *Journal of Wildlife Management, 71*(5), 1419-1426.

Schröter, D., Cramer, W., Leemans, R., Prentice, I. C., Araújo, M. B., Arnell, N. W., et al. (2005). Ecosystem service supply and vulnerability to global change in Europe. *science*, *310*(5752), 1333-1337.

Serra, P., Pons, X., & Saurí, D. (2003). Post-classification change detection with data from different sensors: some accuracy considerations. *International Journal of Remote Sensing*, *24*(16), 3311-3340.

Serra, P., Pons, X., & Saurí, D. (2008). Land-cover and land-use change in a Mediterranean landscape: a spatial analysis of driving forces integrating biophysical and human factors. *Applied Geography 28*(3), 189-209.

Serra, P., Vera, A., Tulla, A. F., & Salvati, L. (2014). Beyond urban–rural dichotomy: Exploring socioeconomic and land-use processes of change in Spain (1991–2011). *Applied Geography, 55*, 71-81.

Sörensen, R., Zinko, U., & Seibert, J. (2006). On the calculation of the topographic wetness index: evaluation of different methods based on field observations. *Hydrology and Earth System Sciences Discussions 10*(1), 101-112.

Strijker, D. (2005). Marginal lands in Europe—causes of decline. *Basic and Applied Ecology*, *6*(2), 99-106.

Stoebner, T. J., & Lant, C. L. (2014). Geographic determinants of rural land covers and the agricultural margin in the Central United States. *Applied Geography*, 55, 138-154.

Tapia, R., Stein, A., & Bijker, W. (2005). Optimization of sampling schemes for vegetation mapping using fuzzy classification. *Remote Sensing of Environment, 99*(4), 425-433.

Tenerelli, P., & Carver, S. (2012). Multi-criteria, multi-objective and uncertainty analysis for agro-energy spatial modelling. *Applied Geography* 32(2), 724-736.

Tiessen, H., Salcedo, I. H., & Sampaio, E. V. S. B. (1992). Nutrient and soil organic matter dynamics under shifting cultivation in semi-arid northeastern Brazil. *Agriculture, Ecosystems & Environment*, *38*(3), 139-151.

Vicente-Serrano, S. M., Lopez-Moreno, J. I., Beguería, S., Lorenzo-Lacruz, J., Sanchez-Lorenzo, A., García-Ruiz, J. M., Azorín-Molina, C., Morán-Tejeda, E., Revuelto, J., Trigo, R., Coelho, F., Espejo, F. (2014). Evidence of increasing drought severity caused by temperature rise in southern Europe. *Environmental Research Letters 9*(4), 044001

Vicente-Serrano, S.M., Beguería, S. & López-Moreno, J.I. (2010). A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of Climate 23*(7), 1696-1718.

Vidal-Macua, J.J., Zabala, A., Ninyerola, M., & Pons, X. (2017). Developing spatially and thematically detailed backdated maps for land cover studies. *International Journal of Digital Earth 10*(2), 175-206.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). Mixed effects models and extensions in ecology with R. Spring Science and Business Media, New York.

Resumen de resultados

8. RESUMEN DE RESULTADOS

8.1. Validación de resultados

La validación de la clasificación de imágenes y de los modelos BRT confirma la coherencia y solidez de los resultados. Por un lado, el acierto global de las clasificaciones de cubiertas y usos del suelo es superior al 90%, y el acierto por categorías está por encima del 75% en la mayoría de los casos, especialmente en las cubiertas analizadas (cubiertas de vegetación y cultivos herbáceos). Tal y como se comentó en el apartado de metodología, tras la paralelización del clasificador kNN en la versión 8.0 del *software* MiraMon, la clasificación final se llevó a cabo mediante dicho algoritmo, reduciendo tiempos de procesado y manteniendo o aumentando ligeramente los porcentajes de acierto en prácticamente todos los casos (Tabla A1 del Anexo). En el Artículo 1, los resultados obtenidos se compararon además con los obtenidos mediante un clasificador *support vector machine* (SVM), en este caso del *software* Envi (2012), ya que es un clasificador muy utilizado y citado en la literatura internacional sobre clasificación de imágenes. La comparación de ambos clasificadores (método híbrido y SVM) demuestra que las diferencias en cuanto a porcentaje de acierto eran muy escasas. Por lo tanto, el kNN de MiraMon es una opción preferible ya que, además de estar paralelizado y mantener o mejorar porcentajes de acierto respecto al clasificador híbrido, permite clasificar, al igual que el clasificador IsoMM, aquellos píxeles que tienen valores NoData en alguna de sus variables; mediante un umbral se establece el máximo número de variables con valor NoData para que un pixel pueda ser clasificado usando el resto de variables. Este aspecto es fundamental cuando se cuenta con imágenes con presencia de nubes o zonas de montaña con sombras duras topográficas. Así mismo, los resultados se compararon con el clasificador supervisado de los *software* ArcGIS, MiraMon y Envi, aunque dicha información no se incluyó en el artículo dado que el clasificador SVM mostraba un rendimiento superior al clasificador supervisado (constituyendo así la mejor opción para establecer comparaciones con el modelo híbrido y el kNN), y por criterios de extensión de la revista.

De las cubiertas analizadas, las que presentan mayor confusión, (Tabla A1 del Anexo: resultados finales a partir del clasificador kNN), son las zonas de pradera (*grasslands*) y las de matorral (*shrublands*), y algunos casos de frondosas perennifolias (*broadleaf evergreen forests*) en las escenas norte. No obstante, se considera que el posterior filtrado realizado a las clasificaciones (descrito en detalle en el Artículo 2), basado en el índice de incertidumbre y el enmascaramiento de las zonas de frontera entre cubiertas, evita la inclusión de los píxeles peor clasificados en la posterior fase de análisis de cambios. Dicho filtrado permite reducir el posible "ruido" estadístico en la distribución de datos (en relación al valor de las variables en cada transición), favoreciendo así a la capacidad predictiva de los modelos BRT, aunque ya de por sí dichos modelos sean menos vulnerables a valores extremos (los denominados en la literatura estadística como *outliers* y *high-leverage points* (James et al. 2013).

En cuanto a la validación de los modelos BRT, tanto los referentes a las transiciones de vegetación como los referentes a abandono de cultivos, todos ellos presentan un rendimiento predictivo sólido, con un valor AUC de la curva ROC por encima del 0.70. En cuanto a los modelos de transición entre cubiertas de vegetación, todos están por encima de un valor AUC=0.85, excepto algunos modelos correspondientes al ámbito SE3 (escena 200034) del Artículo 2, con valores de AUC entre 0.80 y 0.85 (Tablas S2, S3 y S4 del Artículo 2). En lo que se refiere a los modelos de abandono de cultivos, la mayoría tienen valores AUC en torno a 0.75 o superior, quedando algunos modelos de abandono de cultivos de secano en la zona de Pirineos centrales con valores entre 0.70 y 0.75 (Tabla A1 del Artículo 3). En cuanto a los gráficos de

dependencia, en los que se visualizan las condiciones más o menos favorables para que tenga lugar un tipo de transición, los patrones encontrados en respuesta a ciertas variables han permitido comprobar la robustez de los modelos BRT. Es el caso de la respuesta a las variables topográficas, obteniéndose patrones acordes con la teoría ecológica. En cuanto a la dinámica de la vegetación, algunos de estos patrones en respuesta a las condiciones ambientales son los siguientes: que las transiciones a frondosas, especialmente las caducifolias, son más probables en las zonas con mayor pendiente, puesto que este grupo de vegetación tiene un sistema de raíces más potente y desarrollado (Blanco et al. 1997) (conviene recordar que las especies de frondosas dominantes en estos ámbitos son la encina (*Quercus ilex*), el quejigo (*Quercus faginea*) y el roble pubescente (*Quercus pubescens*)); que en las transiciones desde pasto o matorral, entendidas estas cubiertas como fases previas a estados de mayor madurez, la evaporación del agua y la disponibilidad hídrica son factores determinantes, y que, por lo tanto, todas las transiciones son vulnerables a valores altos de radiación solar y a valores bajos del índice topográfico de humedad (*Topographic Wetness Index*, TWI) (utilizado como *proxy* de la humedad del suelo); no obstante las coníferas (en este caso los pinos ya que el abeto casi no tiene representación como masas uniformes y extensas en las áreas de estudio), son más vulnerables a niveles bajos de radiación solar y más tolerantes a niveles altos, como especies heliófilas que son (Blanco et al 1997; Ameztegui & Coll 2011). En cuanto a los modelos de abandono de cultivos, algunos ejemplos que muestran la coherencia en relación a las variables topográficas son los siguientes: en todos los casos la probabilidad de abandono aumenta conforme incrementan la pendiente (variable que también puede utilizarse como *proxy* de la fertilidad del suelo (Yoshioka 2009)) y la rugosidad del terreno, indicando la dificultad que suponen estos limitantes de carácter físico para la mecanización y, en general, para la producción agrícola; las zonas planas de fondo de valle y las laderas suaves de orientación sur tienen menos probabilidad de ser abandonadas, hechos que se ponen en relación con la fertilidad y con cantidad de radiación solar incidente.

8.2. Factores determinantes de las transiciones forestales

En el Artículo 2 sobre la dinámica forestal en la franja mediterránea de la Península Ibérica, la altitud es la variable más importante en casi todos los modelos de transición, lo que ha permitido definir la distribución altitudinal de los grupos de vegetación e incluso aproximarse a la especie forestal en algunos casos. La radiación solar, la pendiente, la influencia de la topografía sobre la humedad del suelo y la tolerancia a la sequía se han mostrado como factores clave para la comparación entre tipos fisionómicos de vegetación. La sequía se ha mostrado como un factor clave en la sucesión y las transiciones entre tipos de vegetación: los patrones encontrados en respuesta a las variables de sequía han podido interpretarse de forma clara para determinar el efecto negativo o positivo de la ocurrencia de sequías. Los gráficos de dependencia en relación a las transiciones desde prado pueden consultarse en las Figuras A5-A10 del Anexo.

Todas las transiciones desde prado o matorral a formaciones forestales se han mostrado vulnerables a valores altos de radiación solar, indicando que la evaporación del agua es un factor de control en dichas etapas de crecimiento. Del mismo modo, la disponibilidad de humedad en el suelo también favorece la supervivencia de plantas jóvenes. No obstante, en estas fases iniciales, las coníferas son más tolerantes a niveles altos de radiación solar debido a, tal y como se ha comentado antes, su comportamiento heliófilo, y más tolerantes a niveles bajos de humedad en el suelo. A diferencia de otras variables relacionadas con la accesibilidad del paisaje (distancia a infraestructuras) y las estadísticas demográficas, en todas las transiciones que parten desde cubiertas de prado se ha encontrado un patrón común en relación a la carga

ganadera (*Liv_Units*), que indica que la probabilidad de la sucesión de prado a estados leñosos es mayor conforme la carga ganadera disminuye (Figuras A5-A10 del Anexo). Por lo tanto, podría deducirse que la retirada de la ganadería extensiva se traduce en una revegetación de los paisajes tradicionales de pastoreo en los tres ámbitos de estudio del Artículo 2.

En cuanto a la competencia entre formaciones forestales maduras, las frondosas caducifolias son el grupo de vegetación más dependiente en cuanto a condiciones topográficas. Estas especies son menos tolerantes a altos niveles de radiación solar y baja disponibilidad de humedad en el suelo que las frondosas perennifolias y coníferas. Las coníferas no mediterráneas, como *Pinus sylvestris*, se benefician de la retención de humedad del suelo y valores bajos de radiación solar en competencia con frondosas perennifolias, lo que se manifiesta claramente en un ambiente más seco como el ámbito NE2 (Artículo 2). En las transiciones entre coníferas mediterráneas y frondosas perennifolias no se han identificado preferencias definidas con respecto a la radiación solar. No obstante, los resultados sugieren que las frondosas perennifolias tienen una plasticidad mayor, ya que son más competitivas en un ámbito más xérico, con menor disponibilidad de agua y bajo condiciones de sequía (ámbito NE2, Artículo 2). Por otro lado, los resultados corroboran que en territorios con un contexto climático más cálido (ámbito SE3, Artículo 2), las frondosas tienden a ser mucho más competitivas con disponibilidad de humedad del suelo, lo que está en relación con la capacidad de sus raíces de llegar a penetrar hasta niveles de aguas freáticas. En general, las frondosas están más adaptadas a pendientes pronunciadas, mientras que las transiciones a coníferas tienden a dominar en pendientes inferiores a 25°.

Uno de los aspectos que merece una atención especial, son los patrones encontrados en respuesta a la ocurrencia de sequías. Los resultados indican que en los paisajes mixtos compuestos por frondosas y coníferas, las transiciones a frondosas mediterráneas (*Quercus ilex* y *Quercus coccifera*) y sub-mediterráneas (*Quercus pubescens*) son más probables en aquellas zonas más afectadas por la repetición de periodos secos, mientras que las transiciones a coníferas son más probables en aquellas zonas con menor ocurrencia de estos eventos. En referencia a la escala temporal del SPEI utilizado para calcular las variables de sequía, cabe destacar que en el caso de las transiciones a coníferas, la asociación negativa es con la repetición de valores del SPEI que indican sequía calculados en base a un escala temporal de 24 meses, mientras que la asociación positiva con las transiciones a frondosas es con la repetición de valores SPEI a una ventana temporal de 6 meses. No obstante, la mayor adaptabilidad de las frondosas a la sequía debe contextualizarse en relación a la interacción con otras variables. En efecto, si bien es cierto que en estos ambientes mixtos la ocurrencia de sequías favorece las transiciones a frondosas mediterráneas y sub-mediterráneas, esto es debido a que otros factores abióticos no están introduciendo un estrés añadido al de la sequía. Por ejemplo, en el caso de las transiciones desde matorral a frondosas, el contexto topográfico juega un papel pudiendo hacer más propicio superar este tipo anomalía climática: las transiciones están teniendo lugar en zonas con menor radiación solar, con mayor disponibilidad de humedad en el suelo y en zonas con menor intensidad del viento. En definitiva, estas transiciones a frondosas son más probables en zonas con mayor frecuencia de sequías debido a que estas especies pueden aprovechar mejor otras condiciones abióticas que las favorecen, es decir, se adaptan mejor a este tipo de perturbaciones de tipo climático. En este sentido, parece ser que la adaptación que supone la

esclerofilia[4] en las frondosas perennifolias (fundamentalmente *Quercus ilex* y *Quercus coccifera*) y el mayor desarrollo del sistema de raíces en las frondosas en general, hace a estas especies más resistentes a los periodos de sequía. Sin embargo, a partir de la investigación también se desprende que la prolongación de la sequía (valores de sequía del SPEI de 7-8 meses de duración) puede reducir la capacidad competitiva de las frondosas perennifolias en contextos topográficos menos favorables. Por otro lado, en los ámbitos septentrionales (NE1 y NE2), las transiciones de prado a matorral parecen estar influenciadas negativamente por la recurrencia de eventos de sequía a escala temporal de 24 meses (Figuras A6, A7, A8) y 7-8 meses de duración. Podría considerarse también que, en algún caso esta transición se vería favorecida por la prolongación de sequías a escala temporal de 6 meses y de 4 meses de duración (Figura A8, periodo 1987-2002 ámbito NE2). Esto podría indicar que el estrés producido por la recurrencia de sequías de 4 meses de duración no permite la sucesión a etapas de tipo forestal, siendo el matorral el estrato leñoso terminal, y que la repetición excesiva de sequías más largas (7-8 meses de duración) impide cualquier tipo de sucesión progresiva o desarrollo de plántulas.

Finalmente, del análisis de las tasas de transición (porcentaje de superficie que ha cambiado de un tipo de vegetación a otro entre dos fechas) se observa que la tasa de transición a frondosas (especialmente perennifolias) ha aumentado de un sub-periodo a otro a lo largo de 1987-2012 (Tablas S5, S6 y S7, Artículo 2) y, en cambio, la tasa de transición a coníferas ha disminuido a lo largo de dicho periodo. Por tanto, del paralelismo encontrado entre estas tendencias y la dinámica de las sequías (Figura 4) se deriva que el incremento de eventos secos en los últimos 30 años es un factor determinante importante en la expansión de las frondosas y en el declive observado de las coníferas.

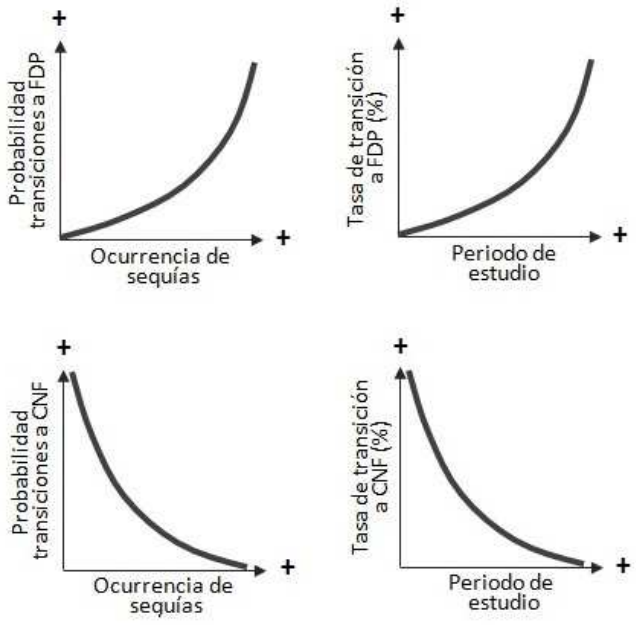


Figura 4. Síntesis de las tendencias encontradas en relación a la ocurrencia de sequías y a la tasa de transición de frondosas perennifolias (FDP) y coníferas (CNF) a lo largo del periodo de estudio.

[4] La esclerofilia rentabiliza la actividad fotosintética (Blanco *et al.* 1997), entre otras cosas mediante un recubrimiento más coriáceo de las hojas y una regulación eficaz de la apertura de los estomas según las horas de más o menos calor, reduciendo así las pérdidas de agua.

8.3. Factores determinantes del abandono de cultivos

En cuanto a los resultados obtenidos del análisis del abandono de cultivos (Artículo 3), los modelos confirman que las variables ambientales derivadas de la topografía representan los determinantes más importantes, excepto en el caso de los cultivos de regadío en la depresión central del Ebro, en los que la accesibilidad a carreteras principales y a capitales de provincia constituyen los factores que más influyen en el abandono. La rugosidad y la pendiente del terreno son los limitantes físicos más importantes para la producción agrícola desde el punto de vista de la accesibilidad y la mecanización de las labores. En el caso de la pendiente, algunas referencias del Artículo 2 también relacionan el incremento de ésta con la perdida de fertilidad del suelo, en el sentido de que las pendientes pronunciadas están más afectadas por la escorrentía superficial y la lixiviación de los nutrientes del suelo. En relación a la variabilidad entre solsticios del ángulo de incidencia del sol (relacionado con la posición del vector solar respecto de la pendiente y la orientación del terreno), los fondos de valle planos y las laderas suaves de orientación sur representan el óptimo de radiación solar recibida para el mantenimiento de la actividad agrícola. De la respuesta a la curvatura del terreno cabe destacar la mayor vulnerabilidad de los cultivos, tanto de secano como de regadío, a los terrenos cóncavos (más que a los convexos) en la depresión central del Ebro, situación que puede potenciar el mal drenaje del suelo y los conocidos problemas de salinización de esta zona (en relación a su litología y la propia salinidad del agua). Las variables que cuantifican la ocurrencia de sequías en los modelos de abandono de cultivos tienen en general una menor importancia relativa en comparación con los modelos de dinámica de la vegetación. Aún así, conviene recalcar que estas variables tienen mayor peso relativo en el abandono de cultivos de regadío que en los de secano. Los patrones encontrados reflejan con claridad el incremento de la tasa de abandono en las zonas irrigadas conforme aumenta el número de periodos secos, en este caso de 7-8 meses de duración.

La probabilidad de abandono en relación a la accesibilidad (distancia ponderada) a vías de comunicación, capitales de provincia y otras zonas urbanas presenta patrones polarizados, especialmente en el caso de los cultivos de secano. Uno de los ejemplos más representativos es la probabilidad de abandono de estos cultivos según la accesibilidad a zonas urbanas (capitales de provincia no incluidas) en el Pirineo: las localizaciones más propensas a ser abandonadas se localizan tanto en las zonas adyacentes a la zona urbanizada, indicando la posible presencia de nuevas oportunidades diferentes a los usos tradicionales, como en las zonas más alejadas de los núcleos urbanos, donde el abandono de la actividad se relaciona con el coste de desplazamiento. Este patrón también puede observarse en el valle central del Ebro, aunque en este caso en relación a la accesibilidad a las carreteras durante 1987-2002, donde el abandono de las zonas más aisladas se vería potenciado, además, por los programas de retirada de tierras habidos durante la primera década, tras el acceso de España a la Unión Europea (UE). Durante la década posterior (2002-2012) el patrón cambia en esta zona, y son fundamentalmente las localizaciones más cercanas a las vías de comunicación las más propensas a ser abandonadas. Esta situación podría relacionarse con la función desempeñada por estas infraestructuras en este contexto territorial durante el desarrollo industrial, y en relación a otros sectores como la construcción y servicios, durante los años posteriores al acceso a la UE. En este sentido, hay que tener en cuenta que, como se comentó en capítulos iniciales, la cuenca central del Ebro ha presentado condiciones más adecuadas para la implantación del sector industrial y el acceso a mercados externos (entre otras cosas por la mejor accesibilidad a los grandes centros urbanos). Por otro lado, los resultados indican que la actividad agrícola en el Pirineo es más dependiente de la accesibilidad a la red de carreteras secundarias (y otros núcleos urbanos en el caso de los cultivos de secano), manteniendo un patrón similar en los tres periodos analizados. Dicho patrón, y su estabilidad a lo largo del

periodo de estudio, reflejarían un contexto socioeconómico menos influenciado por los *inputs* del desarrollo económico que hubo hasta los primeros años de la década 2000-2010. En este contexto la producción agrícola sigue estando vinculada a un mercado interno o de escala local y, por tanto, el acceso a las carreteras secundarias juega un papel fundamental. Conviene recordar, además, que se trata de una sociedad más envejecida y que el elevado coste de desplazamiento puede ser un motivo importante para el abandono de las parcelas más remotas. En cambio, las localizaciones en torno a las carreteras primarias en el Pirineo son más propensas a ser abandonadas. Esto indicaría que la conexión a estas vías principales (que por sus características suponen un menor coste de desplazamiento) permitiría una reconversión de las economías basadas en actividades agrícolas a oportunidades más vinculadas con el entorno urbano o el acceso a mercados externos y que, por lo tanto, están más relacionadas con factores económicos a escala regional, nacional o internacional. La mayor probabilidad de abandono asociada al aislamiento respecto de las capitales de provincia también puede vincularse a tales factores, hecho potenciado además por el envejecimiento poblacional de las zonas más remotas y menos accesibles.

Finalmente, del análisis de la probabilidad de abandono en relación a los sectores económicos, se desprende que la diversificación económica puede contribuir al mantenimiento de la actividad agrícola. Un ejemplo destacado es el caso de los cultivos de regadío en el Pirineo: mientras que el abandono de cultivos de secano tiende a aumentar conforme incrementa la participación de otros sectores (indicando una posible reconversión de la economía rural), los patrones en los cultivos de regadío parecen indicar una mayor probabilidad de abandono en los municipios menos diversificados. Los cultivos de regadío en esta zona están fundamentalmente relacionados con la ganadería y la industria de productos lácteos, los cuales sirven para abastecer el mercado interno, pero que a su vez son muy valorados por el turismo estacional. De esta forma, la coexistencia de esta actividad agrícola junto con el turismo de naturaleza, podría potenciar el turismo gastronómico, al mismo tiempo que se mantienen otro tipo de sectores como el de servicios y el de la construcción (ya sea como trabajos a tiempo parcial o completo).

Conclusiones y reflexiones

9. CONCLUSIONES Y REFLEXIONES FINALES

En la presente tesis se ha estudiado la dinámica de la vegetación y el abandono de cultivos en varias zonas de la Península Ibérica en el periodo 1987-2012, a partir del procesamiento de imágenes de satélite Landsat y la implementación del método *machine-learning Boosted Regression Trees* (BRT). El tratamiento de grandes volúmenes de datos provenientes de la información espectral de imágenes de satélite y otro tipo de variables, tales como un modelo digital de elevaciones de alta resolución para una gran superficie del territorio, ha podido gestionarse de manera eficaz mediante Tecnologías de la Información Geográfica (TIG) y tratamiento estadístico. Del conjunto de técnicas implementadas, los clasificadores estadísticos del SIG MiraMon y las aproximaciones *data-mining*[5] del *software* R, han tenido un rol central en el esquema metodológico. La paralelización de los algoritmos de clasificación en MiraMon ha permitido obtener resultados en tiempos relativamente cortos para un volumen elevado de datos de entrada, al mismo tiempo que se han podido generar numerosas baterías de pruebas gracias a la flexibilidad de ajuste de dichos algoritmos. Además, este *software* ha permitido concatenar y automatizar la mayoría del flujo de trabajo mediante ficheros de procesado por lotes (*batch files*), pudiendo integrar sus funciones (y procesos en bucle de éstas) con comandos basados en el lenguaje del entorno CMD del sistema operativo *Windows*. La gestión de la base de datos de zonas de cambio y variables explicativas mediante el *software* R ha permitido automatizar y sistematizar la extracción de datos para cada modelo y la evaluación de la consistencia estadística de cada subconjunto de datos. Los modelos BRT aplicados sobre esta plataforma han detectado patrones consistentes en la respuesta a las principales variables explicativas gracias a la sofisticación de su algoritmo, cuya eficacia predictiva se basa en la evaluación y combinación progresiva de miles de sub-modelos. El entorno R también ha permitido automatizar la producción de largas series de material gráfico que han ayudado a interpretar los numerosos resultados obtenidos. Y, en definitiva, en la presente tesis se pone de manifiesto el destacado papel de disciplinas como la teledetección, los SIG y las herramientas derivadas de la estadística computacional, en la comprensión de fenómenos a escala regional que, a su vez, contribuyen a la explicación de escenarios a escala continental y potencialmente global.

La fase de obtención de series temporales de mapas de usos y cubiertas del suelo ha sido resuelta de forma semi-automática gracias a la flexibilidad del algoritmo kNN. La implementación de dicho algoritmo permite identificar zonas cuya cubierta del suelo permanece estable entre dos fechas (las que serán utilizadas para entrenar y validar ambas fechas), por lo que se puede aplicar de manera retrospectiva o prospectiva a partir de la fecha de la que se obtiene la "muestra" estadística de referencia. Esto permite realizar estudios de dinámica temporal en cualquier otra región con información de referencia tan sólo en una de las fechas. Para la aplicación sistemática de esta metodología ha sido fundamental el SIOSE 2005, una base de datos continua para todo el territorio español con un importante nivel de desagregación espacial y temático. Si bien es cierto que las posteriores versiones del SIOSE hubieran permitido la obtención de áreas de entrenamiento y test en las correspondientes fechas, se ha preferido utilizar el kNN para la identificación de dichas áreas a partir del análisis prospectivo del 2005, dada la automatización del proceso y considerando su potencial aplicación en zonas en las que no exista cartografía de cubiertas con esta elevada resolución temporal.

En lo que a la dinámica de la vegetación se refiere, los resultados de este trabajo pueden ayudar a diseñar estrategias eficaces en la gestión de espacios forestales de la franja mediterránea de la Península Ibérica,

[5] https://en.wikipedia.org/wiki/Data_mining

así como de regiones con características ambientales similares. Se han identificado principios importantes para determinar las preferencias ecológicas de los grupos de especies forestales dominantes y de las etapas de sucesión más significativas en esta región. A diferencia de otras aproximaciones basadas en estudios locales más intensivos, el valor del trabajo presentado reside en buena parte en la robustez que, en cuanto a conclusiones, confiere haberlo desarrollado en base a grandes extensiones territoriales con datos disponibles de manera continua. La sostenibilidad ha de caracterizar también a los programas de gestión del medio natural, y tales principios pueden concebirse como una base para monitorizar la vulnerabilidad o eficacia de la vegetación, ya sea bajo las actuales condiciones del medio, o ante los posibles cambios derivados de la dinámica natural o de las actuaciones relacionadas con el ámbito socioeconómico (expansión urbana, planes de reforestación, etc). En la intersección de las actividades antrópicas con la estrategia de la vegetación natural, los patrones encontrados en respuesta a los factores clave pueden ayudar a mitigar posibles presiones o impactos negativos sobre la vegetación. Pero sin duda uno de los aspectos más destacados, en cuanto a competencias de gestión forestal, es el posible cambio en la extensión y proporción de unas especies respecto a otras ante la futura evolución de las condiciones ambientales en el ámbito mediterráneo. En este sentido, este trabajo y otros de ámbito europeo han mostrado la vulnerabilidad de coníferas como *Pinus sylvestris* ante los eventos de sequía; por lo que si tenemos en cuenta los modelos actuales que pronostican un aumento de la temperatura y de los periodos secos, en algunas zonas podría haber un cambio de especie dominante materializada en la expansión de las frondosas mediterráneas y sub-mediterráneas, más adaptadas a este tipo de perturbaciones. En consecuencia el rango latitudinal de la distribución de *P. sylvestris* en el continente europeo podría verse contraído en su límite meridional, teniendo en cuenta el incremento pronosticado de las condiciones secas en la región mediterránea. Las frondosas mediterráneas también podrían verse afectadas por esta tendencia climática si la frecuencia e intensidad de las sequías incrementa en el límite sur de su distribución en la Península Ibérica, donde el contexto climático es ya de por sí más cálido. No obstante, los resultados han mostrado que las frondosas mediterráneas están más adaptadas a este tipo de anomalías en el norte de la franja mediterránea de la Península. Por lo tanto, podría haber también un desplazamiento del rango altitudinal de estas especies, llegando a ocupar zonas de *P. sylvestris* cuyas condiciones bioclimáticas se verían alteradas por el cambio en el patrón temporal y espacial de las sequías.

La inclusión de tres contextos climáticos en el trabajo ha corroborado el desplazamiento de los pisos de vegetación como consecuencia de las condiciones más cálidas en el ámbito meridional de la Península (ámbito SE3, Artículo 2). En este ámbito de mayor termicidad, las frondosas mediterráneas como *Quercus ilex* tienden a ser mucho más competitivas que las coníferas mediterráneas (*Pinus halepensis*) y sub-mediterráneas (*Pinus nigra*) en condiciones de disponibilidad de humedad en el suelo. Por otro lado, las transiciones entre frondosas caducifolias y coníferas (en ambas direcciones) son menos frecuentes. Además, los resultados muestran que las áreas del norte (NE1 y NE2) se han visto más afectadas por eventos de sequía que el ámbito sur (en términos de importancia relativa de estas variables en los modelos BRT). Podría concluirse entonces que, tomando como referencia la influencia de las variables clave y los períodos estudiados, la dinámica de la vegetación en los ámbitos del norte se debe a la interacción de factores topoclimáticos y a las perturbaciones en el balance hídrico a lo largo del tiempo, mientras que en la dinámica de la vegetación del ámbito sur han influido más los factores derivados de la topografía a lo largo del periodo de estudio.

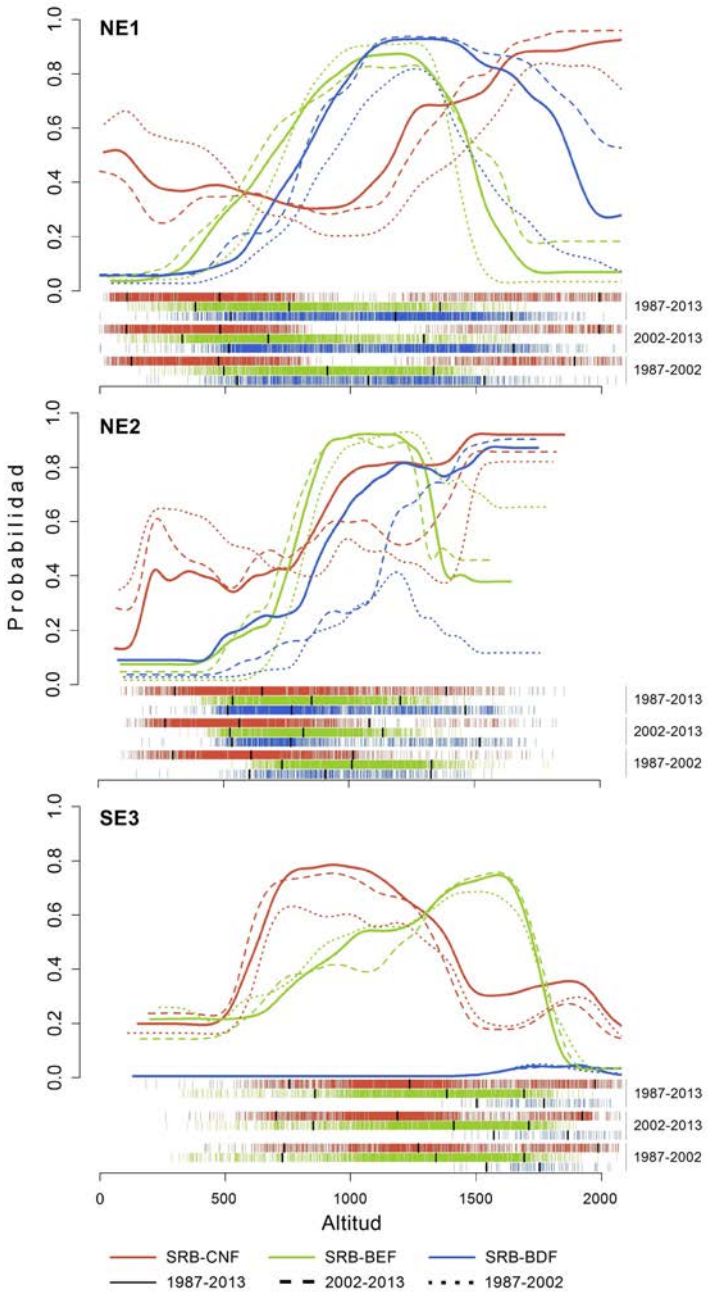


Figura 5. Gráficos de dependencia a la variable altitud. Transiciones: de matorral (SRB) a coníferas (CNF), frondosas perennifolias (BEF) y frondosas caducifolias (BDF). La densidad de los eventos de presencia de cambio está representada por barras verticales de color sobre el eje X. Las marcas verticales de color negro indican los percentiles 0.05, 0.50 y 0.95.

En el presente trabajo no se ha llevado a cabo un análisis pormenorizado del posible desplazamiento del rango altitudinal, según tipo de vegetación y ámbito de estudio, como consecuencia del incremento de las condiciones más secas a largo del tiempo. De todas formas, sirvan como ejemplo los resultados encontrados mostrados en la Figura 5 y Figura 6 para confirmar la importancia de esta cuestión, tanto como para dedicar un trabajo completo de investigación al respecto. Aún sin entrar en detalle, merece la pena observar que los gráficos de dependencia indican una mayor variabilidad en el rango altitudinal (sobre todo en las zonas más elevadas) de las transiciones de vegetación en los ámbitos del norte (NE1 y NE2) en

comparación con el ámbito del sur SE3. Por tanto quedaría por determinar las causas de tal variabilidad en el rango altitudinal, así como eventuales diferencias entre ámbitos, si se relacionan con la mayor presencia de anomalías climáticas en el norte de la franja mediterránea, si responden a factores antrópicos relacionados con el abandono actividades tradicionales (ganadería extensiva, aprovechamiento de recursos forestales, etc.), o a una combinación de ambos (teniendo en cuenta que del análisis se han descartado las zonas incendiadas y las de repoblación forestal). Otra cuestión que cabría abordar es un análisis paleogeográfico de los diferentes grupos de vegetación, para determinar qué grado de influencia ha tenido la distribución de los bosques en épocas pasadas en la configuración espacial de los bosques actuales.

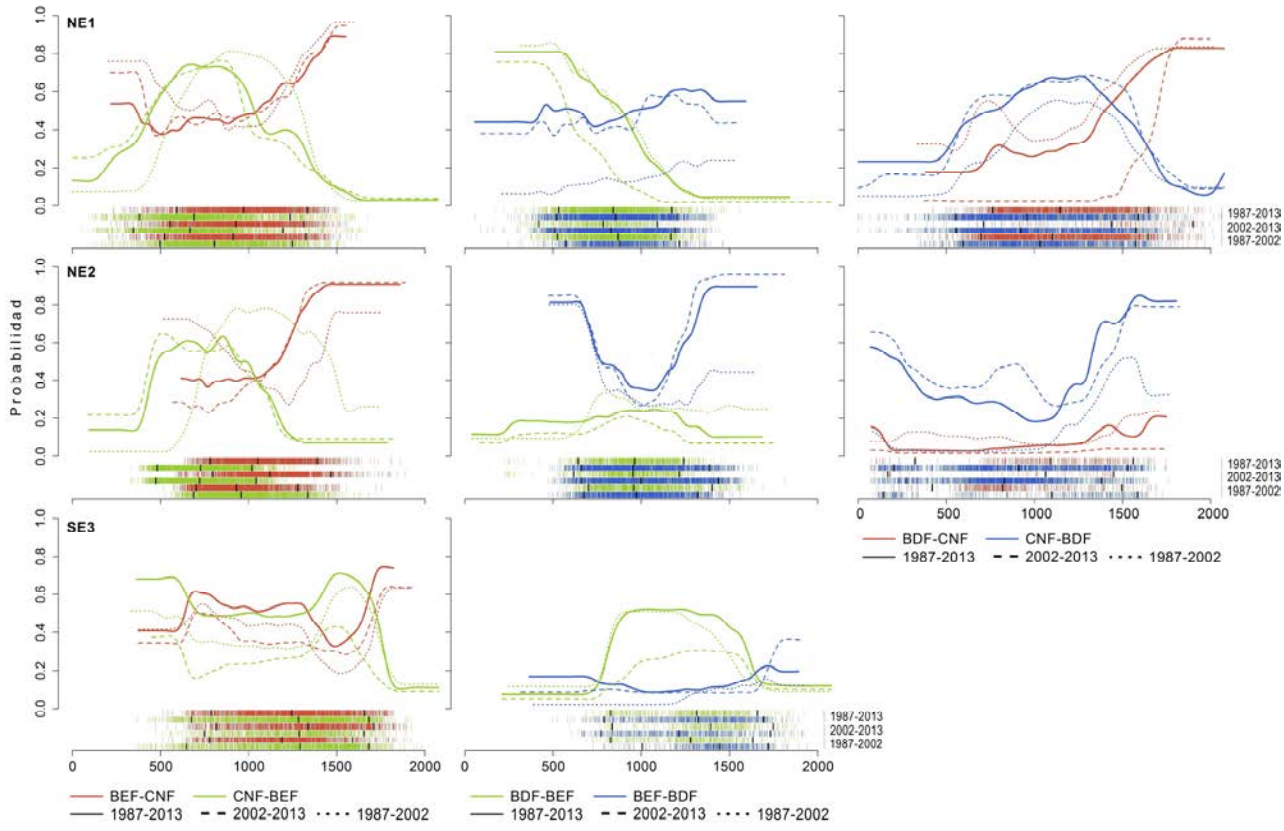


Figura 6. Gráficos de dependencia a la variable altitud. Transiciones entre: coníferas (CNF), frondosas perennifolias (BEF) y frondosas caducifolias (BDF). La densidad de los eventos de presencia de cambio está representada por barras verticales de color sobre el eje X. Las marcas verticales de color negro indican los percentiles 0.05, 0.50 y 0.95.

A excepción de los resultados relacionados con la carga ganadera, en el artículo sobre dinámica forestal no se han podido derivar interpretaciones robustas en relación a las estadísticas demográficas humanas o a la accesibilidad (distancia a infraestructuras). Por lo tanto se entiende que la complejidad en la interacción entre factores humanos y dinámica forestal no ha sido adecuadamente representada con los datos de este tipo, aunque se es consciente de que la gestión histórica del territorio ha tenido una fuerte influencia en la dinámica de la vegetación de la Península Ibérica. El abandono de actividades relacionadas con la gestión tradicional del bosque o, por ejemplo, la sustitución de la leña por combustibles fósiles, ha podido favorecer la expansión de las frondosas (cuya madera ha sido más aprovechada en este sentido). No obstante, la consistencia de la metodología implementada (dada la extensión y diversidad de los ámbitos y

la solidez de los modelos BRT) así como el paralelismo encontrado entre la dinámica de anomalías climáticas y tasas de transición, refuerza la idea de que la sequía ha tenido un papel muy importante en las transiciones de vegetación de esta región geográfica.

En cuanto al análisis del abandono agrícola, si bien los agentes externos de tipo socioeconómico (relacionados con procesos a escala nacional y/o internacional) son conocidos como sus principales promotores, la investigación presentada en esta tesis ha servido para concretar las condiciones inherentes al territorio que actúan como catalizadoras de este fenómeno. Tal y como se comentó en los resultados, y en paralelo con otras investigaciones al respecto (citadas en el Artículo 2), los principales determinantes del abandono se relacionan con factores ambientales que dificultan el desarrollo de la actividad agrícola.

Los modelos BRT también han permitido identificar otras relaciones entre variables ambientales y el abandono, que a pesar de tener un peso específico menor que el grupo anterior de variables, merece la pena destacar por la trascendencia de sus repercusiones en futuros escenarios. Los resultados sugieren que el incremento pronosticado del número e intensidad de las sequías podría conducir a un rendimiento más bajo de los cultivos de regadío, especialmente en la cuenca central del Ebro, donde se ha realizado una gran inversión para la intensificación agrícola a través de modernos sistemas de riego. A ello se suma que la revegetación natural del paisaje tras el abandono de actividades tradicionales en el Pirineo disminuye la escorrentía superficial del agua y, tal y como diversos trabajos han demostrado, esto a su vez reduce la cantidad de agua almacenada en embalses y, en general, el rendimiento hídrico de la cuenca. En su momento, los sistemas de nuevo regadío fueron diseñados para reaccionar a los periodos de sequía, pero la frecuencia e intensidad de estos eventos ha aumentado en las últimas tres décadas. De ahí que si a esta tendencia se suma la reducción del rendimiento hídrico de la cuenca y, como consecuencia, de los sistemas de abastecimiento de agua, las actuales políticas de gestión del territorio puedan quedar desactualizadas ante los de nuevos escenarios en cuanto al balance hídrico del territorio. Dada esta situación (agravada al mismo tiempo por el descenso en la rentabilidad de los cereales de regadío en el mercado internacional), las zonas de nuevo regadío de la cuenca del Ebro se han visto en la necesidad de aumentar sus requerimientos hídricos solicitando nuevas reservas de agua en la zona del Pirineo o el aumento de la capacidad de embalse de las ya existentes. Esto ha desembocado en algunos conflictos sociales, dado que numerosos municipios del territorio pirenaico reclaman un cambio de política con el fin de no verse afectados por las demandas de los municipios de la cuenca central. En los municipios de montaña, el fenómeno del éxodo rural fue muy acusado desde la segunda mitad del siglo XX. En cambio, durante las últimas décadas ha habido una estabilización e incluso ligera recuperación en términos de población y dinamismo económico, debido principalmente al auge del turismo y de otros sectores relacionados con éste. Actualmente, la principal fortaleza de alguno de estos municipios son las actividades de senderismo y los deportes de aventura relacionados con el medio fluvial. Es por ello que dichos municipios se oponen a actuaciones que puedan modificar el caudal de los ríos y el sistema fluvial actual. Además, existen otras cuestiones relacionadas con la conservación de la biodiversidad y el paisaje (natural y cultural) que se verían afectadas por tales actuaciones (Convenio Europeo del Paisaje). Por lo tanto, las estrategias de planificación y gestión del territorio deben diseñarse eficientemente con tal de reducir los posibles costos ambientales y sociales derivados de los nuevos escenarios en la dinámica ambiental y en la de los sistemas económicos. Dichas estrategias deberían basarse en enfoques de tipo multidimensional (en relación a las temáticas o sectores que pueden verse afectados) y multiescalar (por la influencia del contexto internacional en el ámbito local y porque las actuaciones locales, a escala municipal o comarcal, pueden tener repercusiones en el ámbito regional o nacional).

Las condiciones socioeconómicas que definen las zonas más propensas a ser abandonadas pueden interpretarse según diversas causas subyacentes, lo que en muchos casos ha quedado reflejado en patrones polarizados. Los resultados indican que la actividad agrícola en el Pirineo es más dependiente al acceso a los mercados internos y las redes de trabajo local, especialmente para los cultivos de secano. Sin embargo, otras oportunidades relacionadas con el sector secundario y terciario parecen promover el abandono en los municipios más accesibles (en términos del tipo de vía de comunicación). Esto se manifiesta claramente en la cuenca central del Ebro donde, además, hay un punto inflexión en el periodo de estudio que sugiere que durante 2002-2012 hubo un mayor número de hogares rurales que reajustaron su economía, adoptando roles menos tradicionales y más dependientes de las infraestructuras y redes urbanas. La mayor presencia de ayudas de la Política Agraria Común (PAC) durante el período 1987-2002 también ayudaría a explicar este cambio de patrón, especialmente en el caso de los cereales de invierno, tradicionalmente más dependientes de estas ayudas. Y, en definitiva, estos patrones indicarían una mayor influencia de la economía urbana y expansión industrial en la cuenca central del Ebro. Bajo estas circunstancias, la accesibilidad a vías de comunicación, que ha actuado como catalizadora del cambio de modelo en esta área geográfica, debería considerarse un factor estratégico en las políticas territoriales que tienen como objetivo mantener las funciones del medio rural.

Finalmente, el estudio señala que la diversificación económica podría contribuir al mantenimiento de la actividad agrícola y, en general, a integrar sectores profesionales en el contexto rural. Esta cuestión es especialmente relevante en una zona montañosa como el Pirineo, donde el aislamiento respecto de los principales centros económicos podría mitigarse mediante la convivencia de la actividad agrícola con otros sectores como el turismo y los servicios. Además, teniendo en cuenta el carácter de autosuficiencia que todavía mantiene la sociedad de este territorio (en referencia, por ejemplo, a los patrones encontrados en la accesibilidad a carreteras secundarias) debería promocionarse una política adecuada en infraestructuras, para mejorar así la vertebración territorial dentro de esta área geográfica. Dado que, a priori, el carácter autosuficiente no supondría una cualidad negativa, valdría la pena dedicar esfuerzos a identificar posibles desequilibrios entre municipios, manteniendo o actualizando la funcionalidad del sistema de comunicaciones en este sentido. La funcionalidad como objetivo debería ir acompañada de la sostenibilidad y respeto por los valores paisajísticos de estos territorios ya que, a su vez, este patrimonio es la fortaleza principal para la diversificación y apertura de su sistema socioeconómico.

No cabe duda de la relación existente entre la dinámica de la vegetación de los últimos 30 años y el proceso de abandono agrícola en el ámbito mediterráneo. El aislamiento en términos de accesibilidad ha hecho de las montañas mediterráneas espacios periféricos donde el proceso de abandono de tierras ha sido generalizado. Dicho fenómeno ha llevado consigo no sólo el cese de la actividad agrícola sino también el abandono de otras actividades relacionadas con la gestión y aprovechamiento de los bosques. Todo ello ha activado de nuevo los procesos de sucesión natural de la vegetación y la recuperación por parte de ésta de su área de distribución potencial. Esto ha derivado en cambios en la estructura y composición del paisaje que, tal y como ya se comentó, pueden considerarse positivos o negativos en función de cada contexto territorial, o en relación a un aspecto en concreto. Por ejemplo, el abandono de tierras y la consiguiente revegetación natural se ha relacionado con el incremento de incendios, principalmente por un aumento de la vegetación arbustiva, con un grado de inflamabilidad mayor, y en general de las áreas de bosque que han dejado de ser gestionadas o aprovechadas económicamente. Es razonable considerar entonces que la combinación de este proceso junto con el incremento de las sequías en las últimas décadas sea una de las

causas más importantes a la hora explicar el aumento en la frecuencia e intensidad de los incendios. No hay que olvidar, por otro lado, la inversión dedicada a la gestión activa del bosque, que en algunos casos es considerada como el determinante principal del incremento de incendios. De ahí la necesidad de una política sostenible y funcional del territorio, una política que promueva el sostenimiento del modelo socioeconómico rural, de sus particularidades y de las actividades integradas con el entorno, y que a su vez tenga la función de auto-custodia territorial. No obstante, no puede aplicarse un modelo general, ya que si bien es cierto que, por ejemplo, el mantenimiento de un sector agropecuario de intensidad moderada puede tener beneficios ambientales y sociales, en determinados casos el cese de las actividades agrarias también ha tenido un efecto positivo en el paisaje en términos de biodiversidad y conectividad.

La reactivación de la dinámica de la vegetación y la adaptación de ésta a la dinámica climática puede conllevar además una alteración de los servicios ambientales de los bosques, y en concreto los que tienen que ver con el ciclo del carbono. El cambio de especie dominante en algunas zonas podría suponer una modificación en la tasa de absorción del CO_2 dadas las nuevas condiciones ecofisiológicas que tendrían los bosques. Esta cuestión es importante ya que se ha de tener presente el balance entre emisiones de carácter antrópico y la capacidad de capturar CO_2 de las masas forestales. Así mismo, las nuevas condiciones en la estructura y composición forestal pueden comportar cambios de combustibilidad, influyendo así en el régimen de incendios.

Por otro lado, el incremento de las sequías y la reducción en el rendimiento hidrológico de las cuencas (por causas climáticas y por el cambio en el patrón de los usos del suelo) se han de considerar cuestiones clave en las políticas de intensificación agraria, y más concretamente en lo referente a la inversión en el sector de los cultivos de regadío. En este sentido han de valorarse bien los costes y beneficios de los proyectos, así como los riesgos e incertidumbres, más si cabe en un momento en que algunos productos han perdido su rentabilidad ante el contexto de globalización del mercado agrícola. Convendría plantear soluciones más sostenibles en relación al sistema de abastecimiento de agua, incentivando medidas de ahorro y eficiencia del sistema. No obstante, se ha de tener presente la debilidad que supone las economías locales sustentadas en un monocultivo, y que ante las amenazas que se derivan de los pronósticos climáticos y las fluctuaciones del mercado global de alimentos, deberían empezar a plantearse soluciones alternativas de mayor rentabilidad ambiental y socioeconómica.

Para terminar, me gustaría volver a insistir en la validez de la aproximación presentada en este trabajo como contribución al estudio de los sistemas terrestres. La combinación de la teledetección, los SIG y un modelo de cambios de cubiertas del suelo ha permitido definir con solidez patrones de comportamiento de la vegetación y del abandono agrícola a escala regional, con la potencialidad de poder ser definidos a escala continental. La metodología y recursos empleados están al alcance de todos los organismos encargados de una aplicación eficaz de las políticas territoriales.

Futuras líneas de trabajo

10. FUTURAS LÍNEAS DE TRABAJO

Se ha de continuar trabajando para ampliar el conocimiento sobre los factores impulsores de las dinámicas encontradas. En lo que respecta a la línea de investigación presentada en esta tesis, la intención es seguir avanzando en los siguientes aspectos:

- Integrar otras imágenes espectrales de resolución temporal y espacial similar o más detallada, como las proporcionadas por los satélites Sentinel-2, en el esquema de clasificación para valorar su influencia en los resultados.

- Comparar las clasificaciones actuales de las imágenes satelitales con las obtenidas utilizando algoritmos de reciente aplicación en la literatura científica, como por ejemplo *Random Forests*.

- Explorar procedimientos de validación de las clasificaciones temáticas que puedan diferenciar el porcentaje de acierto en base a tipologías de paisaje.

- Trabajar, en la medida de lo posible, con nuevas variables de tipo socioeconómico cuya representatividad espacial y temporal sea mayor que la información actual, a menudo sólo disponible a nivel municipal y en algunos casos con una serie temporal reducida.

- Incorporar nuevos cálculos en la cuantificación de la ocurrencia de variables para determinar con más precisión los umbrales de tolerancia de cada tipo de vegetación y etapa de sucesión.

- Aplicar esta metodología a otras zonas para determinar hasta qué punto los patrones observados son generalizables o particulares de cada contexto geográfico.

- Hacer un análisis de factores potencialmente explicativos de la variabilidad del rango altitudinal de la vegetación a lo largo del tiempo.

- Generar los correspondientes mapas de distribución potencial en relación a la probabilidad de cada transición de vegetación y la del abandono agrícola.

- En relación con el punto anterior, generar modelos que incorporen los posibles escenarios climáticos pronosticados para el futuro, o modelos que puedan cuantificar la influencia de planes estratégicos del ámbito socioeconómico.

- Modelizar los posibles cambios de patrón espacial y temporal en el régimen de incendios según la dinámica y distribución de la vegetación y/o el patrón espacial del abandono agrícola.

Referencias

11. REFERENCIAS

Álvarez-Martínez, J.M., Suárez-Seoane, S., Calabuig, E.D.L. (2011). Modelling the risk of land cover change from environmental and socio-economic drivers in heterogeneous and changing landscapes: The role of uncertainty. Landscape and Urban Planning 101(2), 108-119.

Allen, C. D., and Breshears, D. D. (1998). Drought-induced shift of a forest-woodland ecotone: rapid landscape response to climate variation. *Proceedings of the National Academy of Sciences*, 95(25), 14839-14842.

Alodos, C. L., Pueyo, Y., Barrantes, O., Escós, J., Giner, L., & Robles, A. B. (2004). Variations in landscape patterns and vegetation cover between 1957 and 1994 in a semiarid Mediterranean ecosystem. *Landscape ecology*, 19(5), 543-559.

Ameztegui, A., & Coll, L. (2011). Tree dynamics and co-existence in the montane–sub-alpine ecotone: the role of different light-induced strategies. *Journal of Vegetation Science,* 22(6): 1049-1061.

Andreu, L., Gutiérrez, E., Macias, M., Ribas, M., Bosch, O., & Camarero, J. J. (2007). Climate increases regional tree-growth variability in Iberian pine forests. *Global Change Biology*, 13(4): 804-815.

Arino, O., D. Gross, F. Ranera, L. Bourg, M. Leroy, P. Bicheron, J. et al. (2007). GlobCover: ESA Service for Global Land Cover from MERIS. Paper presented at the Geoscience and Remote Sensing Symposium, (IGARSS) 2412-2415. Barcelona, July 23-27.

Baker, C., Lawrence, R. L., Montagne, C., & Patten, D. (2007). Change detection of wetland ecosystems using Landsat imagery and change vector analysis. *Wetlands*, 27(3), 610-619.

Balzter, H. (2000). Markov chain models for vegetation dynamics. *Ecological modelling*, 126(2), 139-154.

Beilin, R., Lindborg, R., Stenseke, M., Pereira, H. M., Llausàs, A., Slätmo, E., et al. (2014). Analysing how drivers of agricultural land abandonment affect biodiversity and cultural landscapes using case studies from Scandinavia, Iberia and Oceania. *Land Use Policy*, 36, 60-72.

Benayas, J. R., Martins, A., Nicolau, J. M., & Schulz, J. J. (2007). Abandonment of agricultural land: an overview of drivers and consequences. CAB reviews: *Perspectives in agriculture, veterinary science, nutrition and natural resources*, 2(57), 1-14.

Bezák, P., & Halada, L. (2010). Sustainable management recommendations to reduce the loss of agricultural biodiversity in the mountain regions of NE Slovakia. *Mountain Research and Development*, 30(3), 192-204.

Blanco, E., Casado, M.A., Costa, M., Escribano, R., García, M., Génova, M. et al. (1997). Los Bosques Ibéricos, una interpretación geobotánica. Planeta, Barcelona.

Bürgi, M., Hersperger, A.M., Schneeberger, N. (2004). Driving forces of landscape change-current and new directions. *Landscape Ecology* 19(8), 857-868.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification and regression trees. Wadsworth & Brooks, Monterey.

Breiman, L. (1996). Bagging predictors. *Machine learning* 24(2): 123-140.

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author. *Statistical science,* 16(3), 199-231.

Camacho Olmedo, M. T., Paegelow, M., & Mas, J. F. (2013). Interest in intermediate soft-classified maps in land change model validation: suitability versus transition potential. *International Journal of Geographical Information Science*, 27(12), 2343-2361.

Camarero, J. J., Bigler, C., Linares, J. C., & Gil-Pelegrín, E. (2011). Synergistic effects of past historical logging and drought on the decline of Pyrenean silver fir forests. *Forest Ecology and Management* 262(5), 759-769.

Carnicer, J., Coll, M., Ninyerola, M., Pons, X., Sánchez, G. & Peñuelas, J. (2011). Widespread crown condition decline, food web disruption, and amplified tree mortality with increased climate change-type drought. *Proceedings of the National Academy of Sciences*, 108(4): 1474-1478.

Chen, G., Hay, G. J., Carvalho, L. M., & Wulder, M. A. (2012). Object-based change detection. *International Journal of Remote Sensing*, 33(14), 4434-4457.

Chuvieco, E. (1998). Fundamentos de teledetección espacial. Estudios Geográficos, 59, 135.

Chuvieco, E., & Salinero, E. C. (2002). Teledetección ambiental: la observación de la tierra desde el espacio.

Cánoves, G., Villarino, M., Priestley, G. K., & Blanco, A. (2004). Rural tourism in Spain: an analysis of recent evolution. *Geoforum*, 35(6), 755-769.

Collantes, F., & Pinilla, V. (2004). Extreme depopulation in the Spanish rural mountain areas: a case study of Aragon in the nineteenth and twentieth centuries. *Rural History*, 15(2), 149-166.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J. (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4, Geoscientific Model Development.

Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., & Lambin, E. (2004). Review ArticleDigital change detection methods in ecosystem monitoring: a review. *International journal of remote sensing*, 25(9), 1565-1596.

Crase, B., Liedloff, A. C., & Wintle, B. A. (2012). A new method for dealing with residual spatial autocorrelation in species distribution models. *Ecography*, 35(10), 879-888.

De la Riva, J., Pérez-Cabello, F., Lana-Renault, N., & Koutsias, N. (2004). Mapping wildfire occurrence at regional scale. *Remote Sensing of Environment*, 92(3), 363-369.

Domingo-Marimon, C. (2016). Contributions to the knowledge of the multitemporal spatial patterns of the Iberian Peninsula droughts from a Geographic Information Science perspective. Ph.D. thesis, Universitat Autònoma de Barcelona.

Duarte, R., Pinilla, V., & Serrano, A. (2014). The water footprint of the Spanish agricultural sector: 1860–2010. *Ecological economics*, 108, 200-207.

EEA (European Environment Agency). (2014). Terrestrial Habitat Mapping in Europe: An Overview. EEA Technical report No 1/2014. Joint MNHN-EEA report.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A. et al. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29, 129-151.

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802-813.

El Kenawy, A., López-Moreno, J. I., & Vicente-Serrano, S. M. (2012). Trend and variability of surface air temperature in northeastern Spain (1920-2006): linkage to atmospheric circulation. *Atmospheric Research*, 106: 159-180.

Exelis Visual Information Solutions (ENVI) (2010). Boulder, Colorado.

Environmental Systems Research Institute (ESRI) (2012). ArcGIS Release 10.1. Redlands.

Farina, A. (2008). Principles and methods in landscape ecology: towards a science of the landscape (Vol. 3). Springer Science & Business Media.

Fawcett, T. (2004). ROC graphs: Notes and practical considerations for researchers. *Machine learning*, 31(1), 1-38.

Florinsky, I. V., & Kuryakova, G. A. (1996). Influence of topography on some vegetation cover properties. *Catena* 27(2): 123-141.

Forests, ICP (2006). International co-operative programme on assessment and monitoring of air Pollution effects on forests. Institute for World Forestry. Federal Research Centre for Forestry and Forest Products (BFH), Hamburg, Germany.

Friedl, M. A., D. Sulla-Menashe, B. Tan, A. Schneider, N. Ramankutty, A. Sibley, and X. Huang. 2010. MODIS Collection 5 Global Land Cover: Algorithm Refinements and Characterization of New Datasets. *Remote Sensing of Environment* 114 (1), 168-182.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*: 1189-1232.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4), 367-378.

Friedman, J. H., & Meulman, J. J. (2003). Multiple additive regression trees with application in epidemiology. Statistics in Medicine, 22(9), 1365-1381.

Gallart, F., & Llorens, P. (2004). Observations on land cover changes and water resources in the headwaters of the Ebro catchment, Iberian Peninsula. *Physics and Chemistry of the Earth*, Parts A/B/C, 29(11), 769-773.

García-Ruiz, J. M. 1990. Geoecología de las Áreas de Montaña. Geoforma Ediciones, Logroño.

García-Ruiz, J. M., & Lasanta, T. (1993). Land-use conflicts as a result of land-use change in the Central Spanish Pyrenees: A review. *Mountain Research and Development*, 295-304.

Garcia-Ruiz, J. M., Lasanta, T., Ruiz-Flano, P., Ortigosa, L., White, S., González, C., & Martí, C. (1996). Land-use changes and sustainable development in mountain areas: a case study in the Spanish Pyrenees. *Landscape Ecology*, 11(5), 267-277.

García-Ruiz, J. M., López-Moreno, J. I., Vicente-Serrano, S. M., Lasanta, T., & Beguería, S. (2011). Mediterranean water resources in a global change scenario. *Earth-Science Reviews*, 105(3), 121-139.

García-Ruiz, J. M. (2010). The effects of land uses on soil erosion in Spain: a review. *Catena*, 81(1), 1-11.

Gerard, F., Petit, S., Smith, G., Thomson, A., Brown, N., et al. (2010). Land cover change in Europe between 1950 and 2000 determined employing aerial photography. *Progress in Physical Geography*, 34(2), 183-205.

Gehrig-Fasel, J., Guisan, A., & Zimmermann, N. E. (2007). Tree line shifts in the Swiss Alps: climate change or land abandonment*?. Journal of vegetation science,* 18(4): 571-582.

Halada, L., Evans, D., Romão, C., & Petersen, J. E. (2011). Which habitats of European importance depend on agricultural practices?. *Biodiversity and Conservation*, 20(11), 2365-2378.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC curve). *Radiology*, 143(1), 29-36.

Hastie, T., Tibshirani, R.I. (2009). The elements of statistical learning: data mining, inference and prediction. Springer, Berlin.

Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., Hobart, G. W., & Campbell, L. B. (2016). Mass data processing of time series Landsat imagery: pixels to data products for forest monitoring. *International Journal of Digital Earth* 9(11): 1035-1054.

Homer, C., C. Huang, L. Yang, B. Wylie, and Coan M. (2004). Development of a 2001 National Land-cover Database for the United States. *Photogrammetric Engineering & Remote Sensing* 70 (7), 829-840.

IAEST. Instituto Aragonés de Estadística. Censos Agrarios. (2017). http://www.aragon.es/iaest

IDESCAT. Institut d'Estadística de Catalunya. Censos Agraris. (2017). https://www.idescat.cat/

Inventario de Daños Forestales (IDF) (2016). Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente. Gobierno de España.

IPCC. (2007). Climate change 2007: the physical science basis. Contribution of Working Group I to the forth assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. Springer, New York.

Kauth, R. J., & Thomas, G. S. (1976). The tasselled cap-a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. In LARS Symposia (p. 159).

Kawakita, M., Minami, M., Eguchi, S., & Lennert-Cody, C. E. (2005). An introduction to the predictive technique AdaBoost with a comparison to generalized additive models. *Fisheries research*, 76(3), 328-343.

Kelly, A.E., Goulden, M.L. (2008). Rapid shifts in plant distribution with recent climate change. *Proceedings of the National Academy of Sciences* 105(33), 11823-11826.

Keogh, E., Chu, S., Hart, D., & Pazzani, M. (2001). An online algorithm for segmenting time series. In Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on (pp. 289-296). IEEE.

Kirches, G., O. Arino, M. Boettcher, S. Bontemps, C. Brockmann, O. Danne, P. Defourny, et al. (2013). CCI Land Cover Pre-processing. Challenges of Pre-processing for Land Cover Classification. Poster presented at ESA Living Planet Symposium. Edinburgh, September 9-13.

Kosmas, C., Kairis, O., Karavitis, C., Acikalin, S., Alcalá, M., Alfama, P. et al. (2015). An exploratory analysis of land abandonment drivers in areas prone to desertification. *Catena*, 128, 252-261.

Lambin, E. F., Geist, H. J., & Lepers, E. (2003). Dynamics of land-use and land-cover change in tropical regions. *Annual review of environment and resources*, 28(1), 205-241.

Lambin, E. F., & Strahler, A. H. (1994). Indicators of land-cover change for change-vector analysis in multitemporal space at coarse spatial scales. International Journal of Remote Sensing, 15(10), 2099-2119.

Lasanta-Martínez, T., Vicente-Serrano, S. M., & Cuadrat-Prats, J. M. 2005. Mountain Mediterranean landscape evolution caused by the abandonment of traditional primary activities: a study of the Spanish Central Pyrenees. *Applied Geography* 25(1), 47-65.

Leathwick, J. R., Elith, J., Francis, M. P., Hastie, T., & Taylor, P. (2006). Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology Progress Series*, 321, 267-281.

Lieskovský, J., Kenderessy, P., Špulerová, J., Lieskovský, T., Koleda, P., Kienast, F., & Gimmi, U. (2014). Factors affecting the persistence of traditional agricultural landscapes in Slovakia during the collectivization of agriculture. *Landscape ecology*, 29(5), 867-877.

Lloret, F., Calvo, E., Pons, X., & Díaz-Delgado, R. (2002). Wildfires and landscape patterns in the Eastern Iberian Peninsula. *Landscape Ecology*, 17(8), 745-759.

López-Moreno, J. I., Vicente-Serrano, S. M., Angulo-Martínez, M., Beguería, S., & Kenawy, A. (2010). Trends in daily precipitation on the northeastern Iberian Peninsula, 1955-2006. I*nternational journal of Climatology*, 30(7): 1026-1041.

López-Moreno, J. I., Vicente-Serrano, S. M., Moran-Tejeda, E., Zabalza, J., Lorenzo-Lacruz, J., & García-Ruiz, J. M. (2011). Impact of climate evolution and land use changes on water yield in the Ebro basin. *Hydrology and Earth System Sciences*, 15(1), 311.

Lu, D., Mausel, P., Brondizio, E., & Moran, E. (2004). Change detection techniques. *International journal of remote sensing*, 25(12), 2365-2401.

Lu, D., & Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International journal of Remote sensing*, 28(5), 823-870.

Martínez-Vilalta, J., Lloret, F. (2016). Drought-induced vegetation shifts in terrestrial ecosystems: The key role of regeneration dynamics. *Global and Planetary Change*, 144: 94-108.

Masalles, R., & Vigo, J. (1987). La successió a les terres mediterrànies: sèries de vegetació. Ecosistemes Terrestres. La resposta als incendis ia d'altres peretorbacions. Diputació de Barcelona, Barcelona, Spaña, 27-43.

Magliocca, N. R., Rudel, T. K., Verburg, P. H., McConnell, W. J., Mertz, O., Gerstner, K. et al. (2015). Synthesis in land change science: methodological patterns, challenges, and guidelines. *Regional Environmental Change*, 15(2), 211-226.

Magliocca, N. R., Van Vliet, J., Brown, C., Evans, T. P., Houet, T., Messerli, P. et al. (2015). From meta-studies to modeling: Using synthesis knowledge to build broadly applicable process-based land change models. *Environmental Modelling & Software*, 72, 10-20.

Marín-Yaseli, M. L., & Lasanta, T. (2003). Competing for meadows: A case study on tourism and livestock farming in the Spanish Pyrenees. *Mountain Research and Development*, 23(2), 169-176.

Moreira, F., Viedma, O., Arianoutsou, M., Curt, T., Koutsias, N., Rigolot, E. et al. (2011). Landscape-wildfire interactions in southern Europe: implications for landscape management. *Journal of environmental management*, 92(10), 2389-2402.

Mottet, A., Ladet, S., Coqué, N., Gibon, A. (2006). Agricultural land-use change and its drivers in mountain landscapes: A case study in the Pyrenees. *Agriculture, Ecosystems & Environment* 114(2), 296-310.

Mucher, C. A., K. T. Steinnocher, F. P. Kressler, and Heunks, C. (2000). Land Cover Characterization and Change Detection for Environmental Monitoring of Pan-Europe. *International Journal of Remote Sensing* 21 (6-7), 1159-1181.

Navarro, L. M., & Pereira, H. M. (2015). Rewilding abandoned landscapes in Europe. In Rewilding European Landscapes (pp. 3-23). Springer International Publishing.

Ninyerola, M., Pons, X., & Nolla, J. M. R. (2005). Atlas climático digital de la Península Ibérica: metodología y aplicaciones en bioclimatología y geobotánica. Universitat Autònoma de Barcelona, Departament de Biologia Animal, Biologia Vegetal i Ecologia (Unitat de Botánica).

Noun Project (2017). Icons for everything. https://thenounproject.com/.

Pasho, E., Camarero, J. J., de Luis, M., & Vicente-Serrano, S. M. 2011. Impacts of drought at different time scales on forest growth across a wide climatic gradient in north-eastern Spain. *Agricultural and Forest Meteorology*, 151(12): 1800-1811.

Pazúr, R., Lieskovský, J., Feranec, J., & Oťaheľ, J. (2014). Spatial determinants of abandonment of large-scale arable lands and managed grasslands in Slovakia during the periods of post-socialist transition and European Union accession. *Applied Geography*, 54, 118-128.

Pinilla, V. (2006). The development of irrigated agriculture in twentieth-century Spain: a case study of the Ebro basin. *Agricultural History Review*, 54(1), 122-141.

Pons, X., & Solé-Sugrañes, L. (1994). A simple radiometric correction model to improve automatic mapping of vegetation from multispectral satellite data. *Remote Sensing of Environment*, 48(2): 191-204.

Pons X. (2006). MiraMon. Geographical Information System and Remote Sensing Software. Bellaterra: CREAF Research Center.

Pons, X., & Arcalis, A. (2012). Diccionari terminològic de Teledetecció. Institut Cartogràfic de Catalunya.

Pons, X., M. Ninyerola, C. Cea, O. González-Guerrero, Serra P. et al (2014). It's Time for a Crisper Image of the Face of the Earth: Landsat and Climate Time Series for Massive Land Cover & Climate Change Mapping at Detailed Resolution. Poster presented at EGU General Assembly, Vienna, April 27-May 2. In EGU General Assembly Conference Abstracts 16: 13640.

Pontius, R. G., & Parmentier, B. (2014). Recommendations for using the relative operating characteristic (ROC). *Landscape Ecology*, 29(3), 367-382.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ridgeway, G., & Ridgeway, M. G. (2004). The gbm package. R Foundation for Statistical Computing, Vienna, Austria 5(3).

Sánchez-Salguero, R., Camarero, J.J., Gutiérrez, E., González-Rouco, F., Gazol, A. et al. (2016). Assessing forest vulnerability to climate warming using a process-based model of tree growth: bad prospects for rear-edges. *Global Change Biology*.

Schneeberger, N., Bürgi M., Hersperger, A.M., Ewald, K.C. (2007). Driving forces and rates of landscape change as a promising combination for landscape change research—an application on the northern fringe of the Swiss Alps. *Land Use Policy,* 24(2), 349-361.

Schuur, E. A. (2003). Productivity and global climate revisited: the sensitivity of tropical forest growth to precipitation. *Ecology,* 84(5), 1165-1170.

Serra-Díaz, J. M., Cristobal, J., & Ninyerola, M. 2011. A classification procedure for mapping topo-climatic conditions for strategic vegetation planning. *Environmental Modeling & Assessment*, 16(1): 77-89.

Serneels, S., & Lambin, E. F. (2001). Proximate causes of land-use change in Narok District, Kenya: a spatial statistical model. *Agriculture, Ecosystems & Environment*, 85(1), 65-81.

Serra, P., Pons, X., & Saurí, D. (2008). Land-cover and land-use change in a Mediterranean landscape: a spatial analysis of driving forces integrating biophysical and human factors. *Applied Geography* 28(3), 189-209.

Sluiter, R., & de Jong, S. M. (2007). Spatial patterns of Mediterranean land abandonment and related land cover transitions. *Landscape Ecology*, 22(4), 559-576.

Strahler, A., D. Muchoney, J. Borak, F. Gao, M. Friedl, S. Gopal, J. Hodges, et al. (1999). MODIS Land Cover Product, Algorithm Theoretical Basis Document (ATBD) Version 5.0. Center for Remote Sensing, Department of Geography, Boston University.

Verburg, P. H., Crossman, N., Ellis, E. C., Heinimann, A., Hostert, P., Mertz, O. et al. (2015). Land system science and sustainable development of the earth system: A global land project perspective. *Anthropocene*, 12, 29-41.

Vicente-Serrano, S. M., López-Moreno, J. I., Beguería, S., Lorenzo-Lacruz, J., Sanchez-Lorenzo et al. (2014). Evidence of increasing drought severity caused by temperature rise in southern Europe. Environmental Research Letters 9(4): 044001.

Vilà-Cabrera, A., Martínez-Vilalta, J., Galiano, L., & Retana, J. (2013). Patterns of forest decline and regeneration across Scots pine populations. *Ecosystems,* 16(2): 323-335.

Vogelmann, J. E., T. L. Sohl, P. V. Campbell, and Shaw, D. M. (1998). Regional Land Cover Characterization Using Landsat Thematic Mapper Data and Ancillary Data Sources. *Environmental Monitoring and Assessment* 51 (12), 415-428.

Walther, G. R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T. J. et al. (2002). Ecological responses to recent climate change. *Nature* ,416(6879), 389-395.

Wulder, M. A., J. C. White, S. N. Goward, J. G. Masek, J. R. Irons, M. Herold, W. B. et al. (2008). Landsat Continuity: Issues and Opportunities for Land Cover Monitoring. *Remote Sensing of Environment,* 112 (3): 955-969.

Yoshioka, P. M. (2009). Sediment transport and the distribution of shallow-water gorgonians. *Caribbean Journal of Science*, 45(2-3), 254-259.

ANEXO

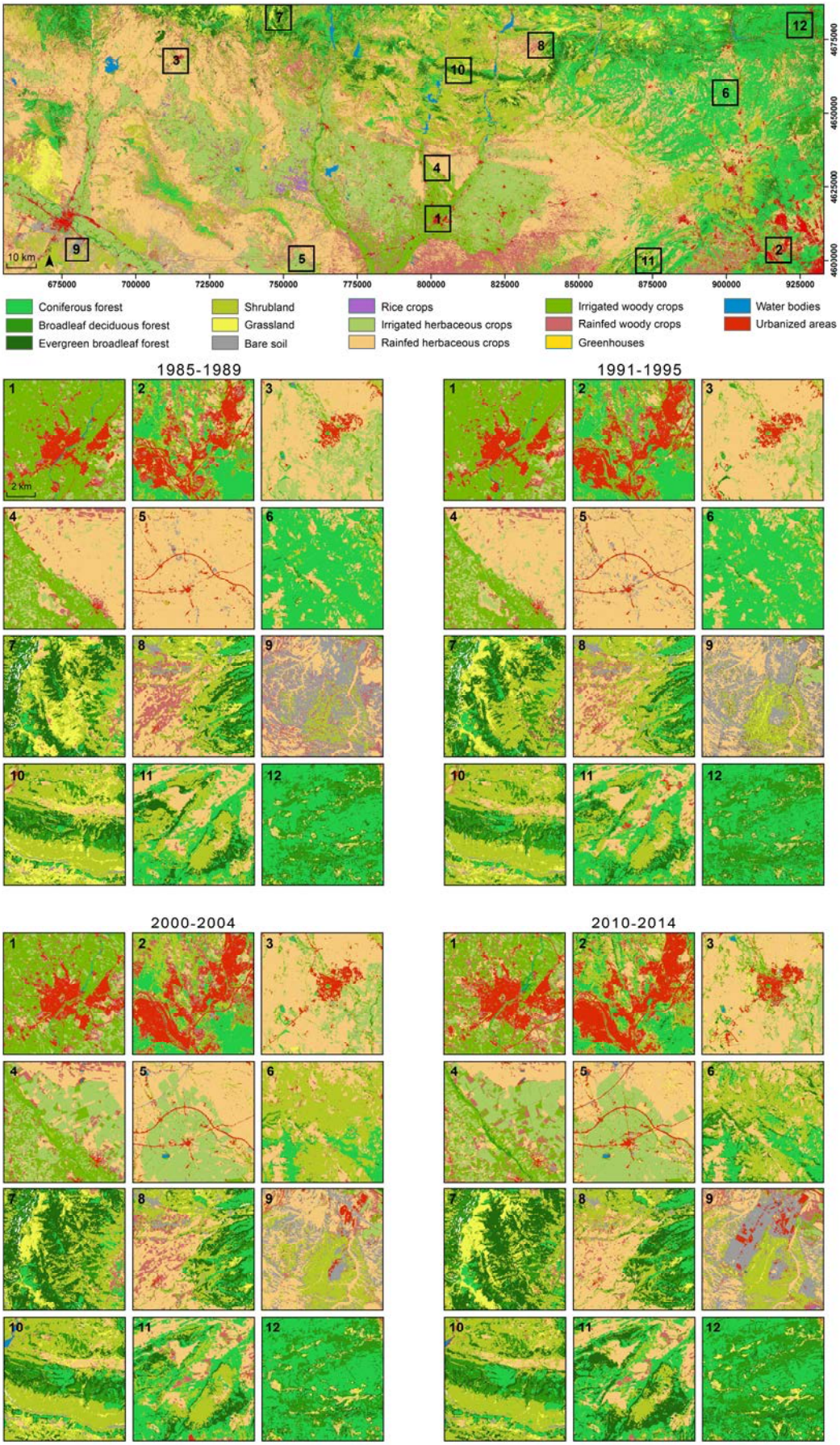


Figura A1. Ejemplo de clasificaciones para los quinquenios 1985-1989, 1991-1995, 2000-2004 y 2010-2014.

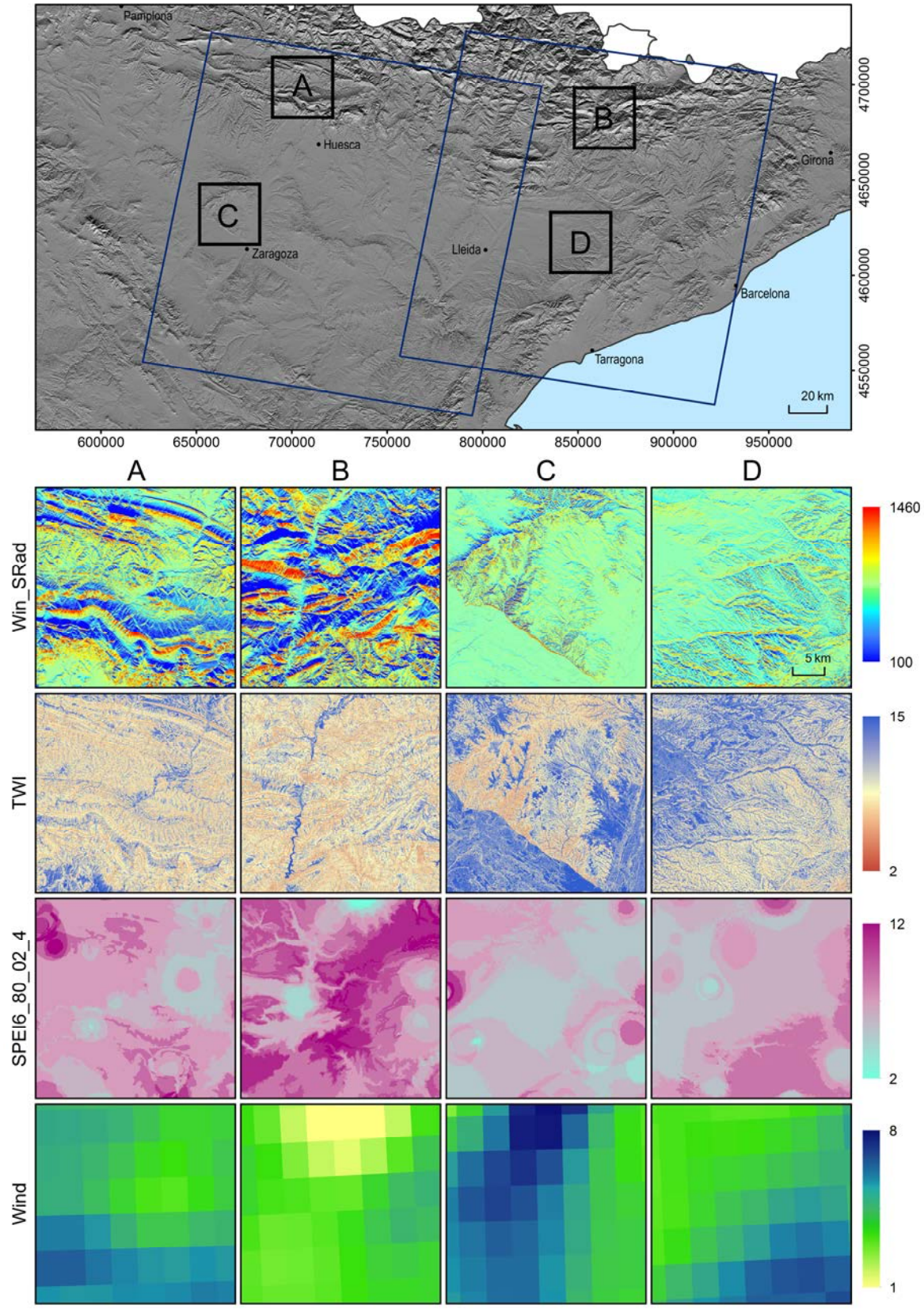


Figura A2. Ejemplos de variables explicativas en los ámbitos de estudio septentrionales. Radiación solar de invierno (10 kJ·m^{-2}·día^{-1}), índice topográfico de humedad (adimensional), número de sequías durante 1980-2002 (SPEI a escala temporal de 6 meses y 4 meses de duración) y viento (m·s^{-1}).

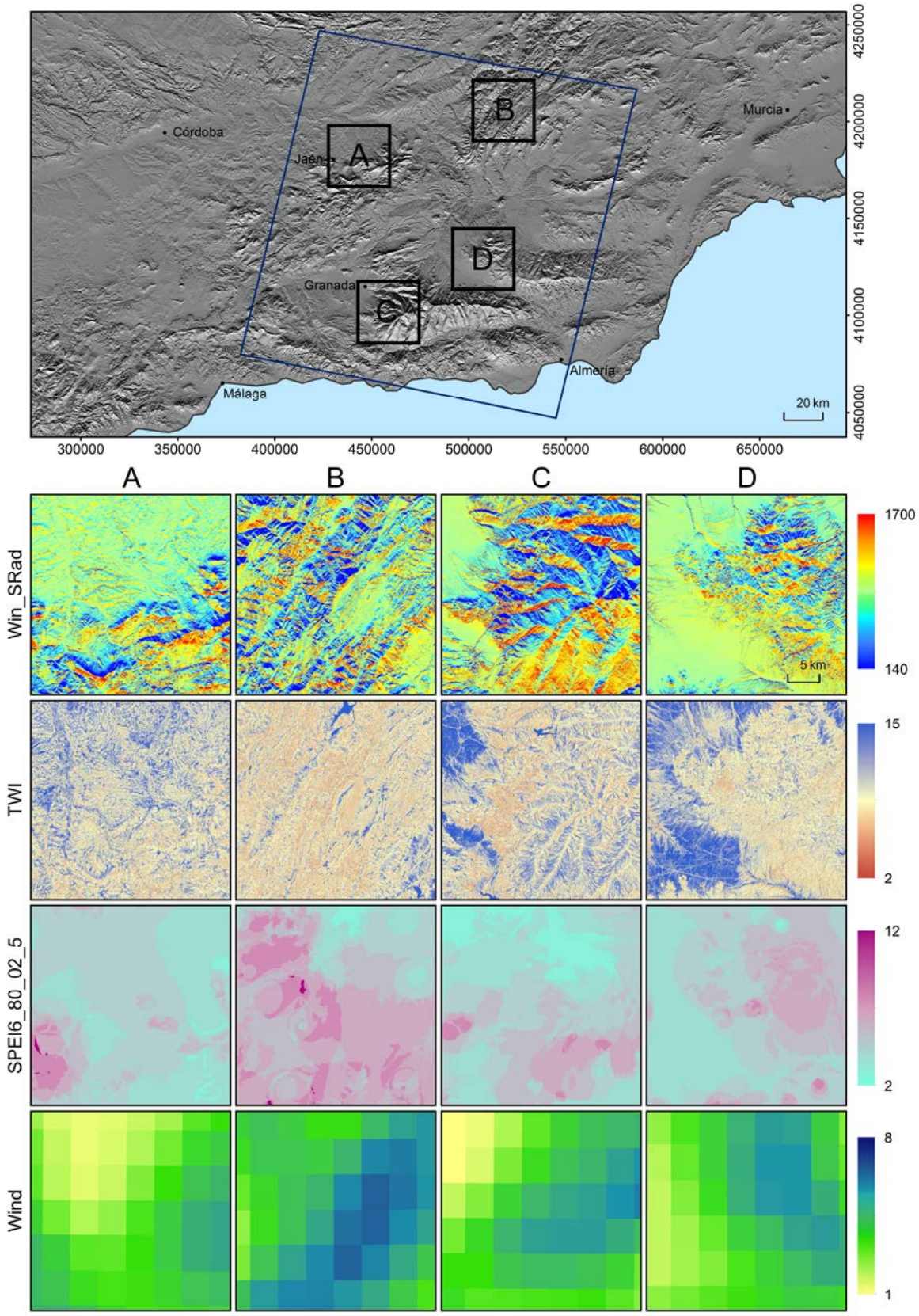


Figura A3. Ejemplos de variables explicativas en el ámbito de estudio meridional. Radiación solar de invierno (10 $kJ \cdot m^{-2} \cdot día^{-1}$), índice topográfico de humedad (adimensional), número de sequías durante 1980-2002 (SPEI a escala temporal de 6 meses y 5 meses de duración) y viento ($m \cdot s^{-1}$).

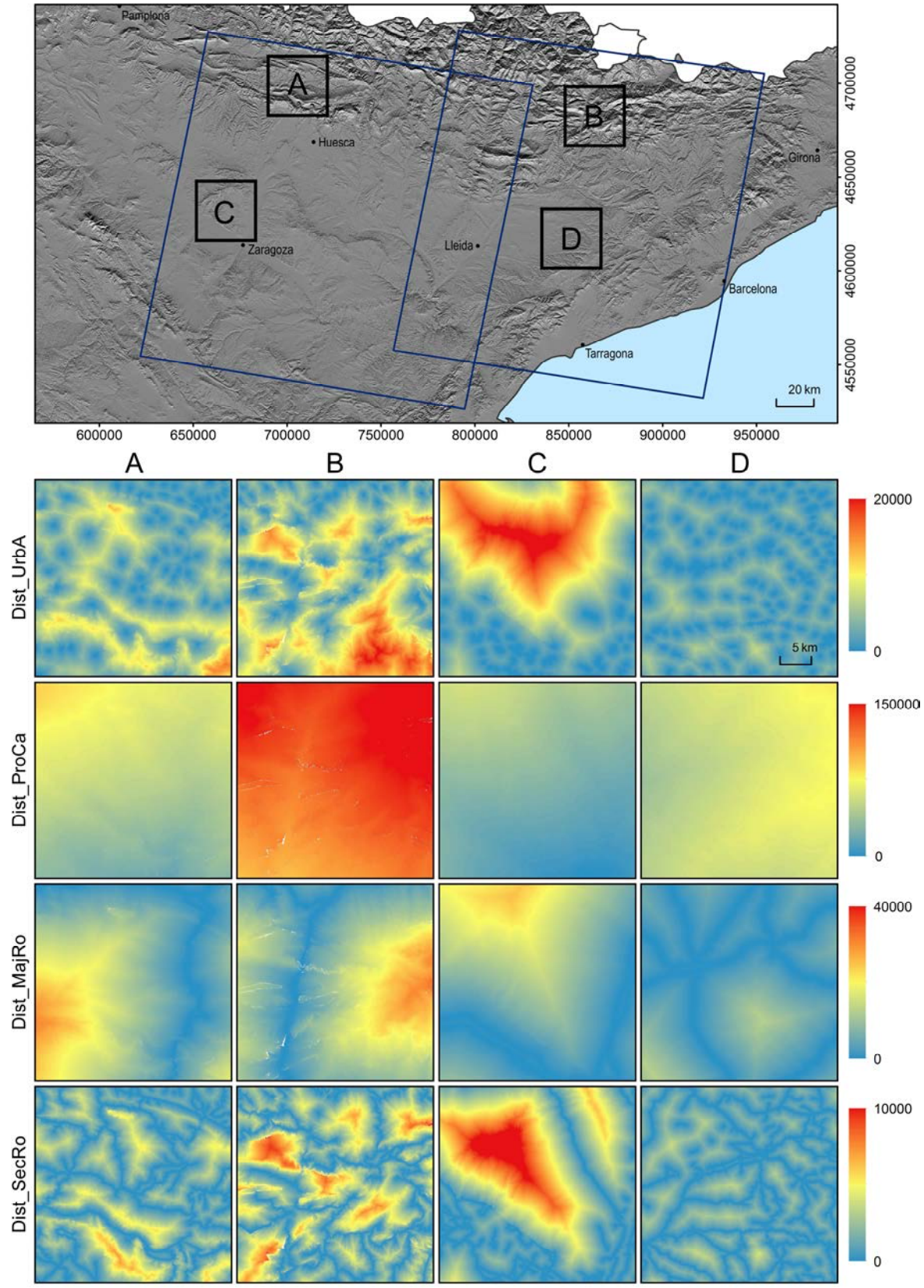


Figura A4. Ejemplos de variables de coste de desplazamiento (distancia ponderada por pendiente e impedancia al desplazamiento) en los ámbitos septentrionales. Distancia a núcleos urbanos (m), distancia a capitales de provincia (metros), distancia a carreteras primarias (m) y distancia a carreteras secundarias (m)

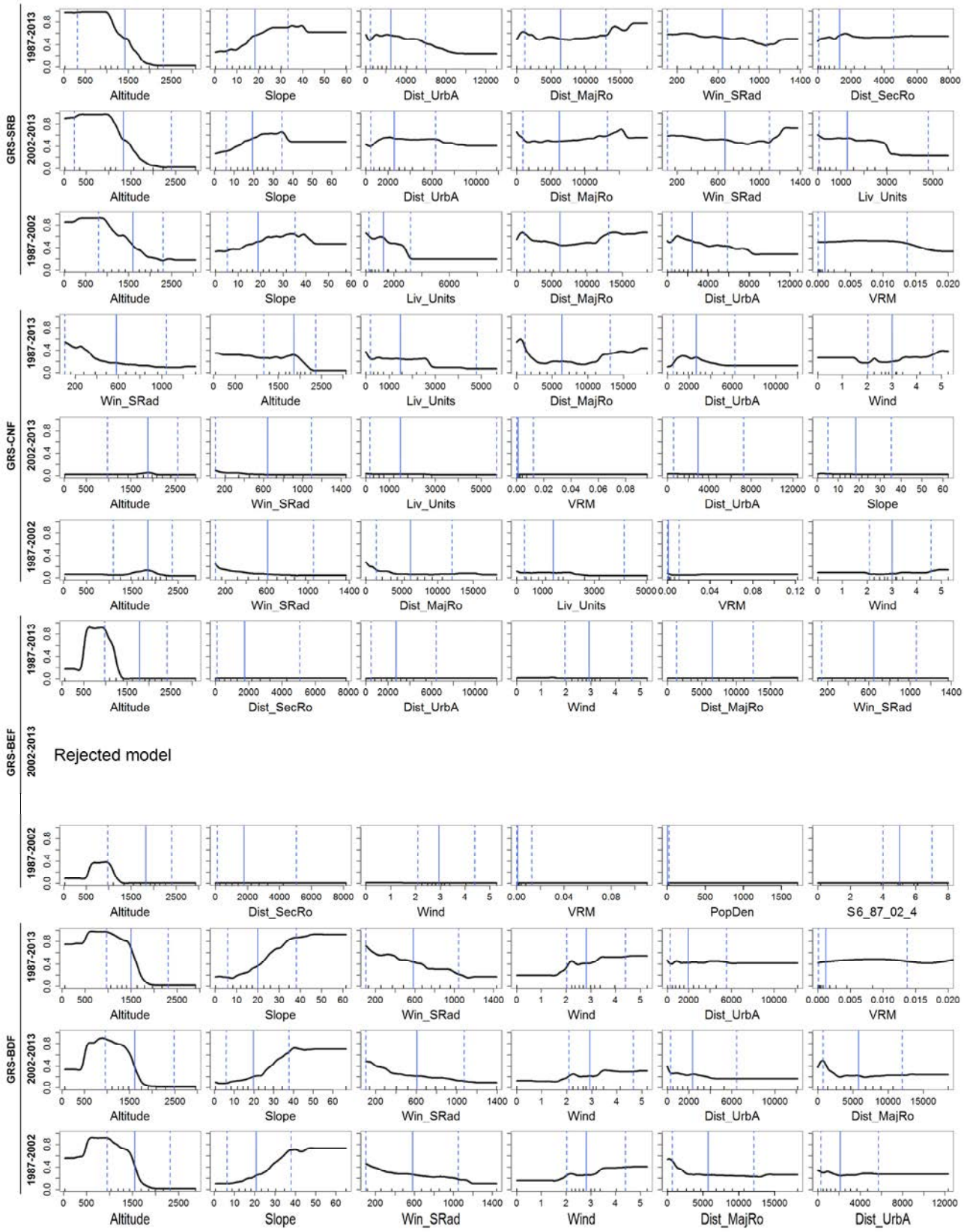


Figura A5. Gráficos de dependencia parcial para las seis primeras variables por orden de importancia. Ámbito NE1. Transiciones desde prado (GRS) a matorral (MTR), frondosas perennifolias (BEF) y frondosas caducifolias (BDF). Modelos BRT que incluyen la variable altitud. Eje Y: probabilidad. Líneas negras verticales sobre el eje X: deciles en la distribución de valores de la variable. Líneas azules verticales: percentiles 0.05, 0.50 y 0.95.

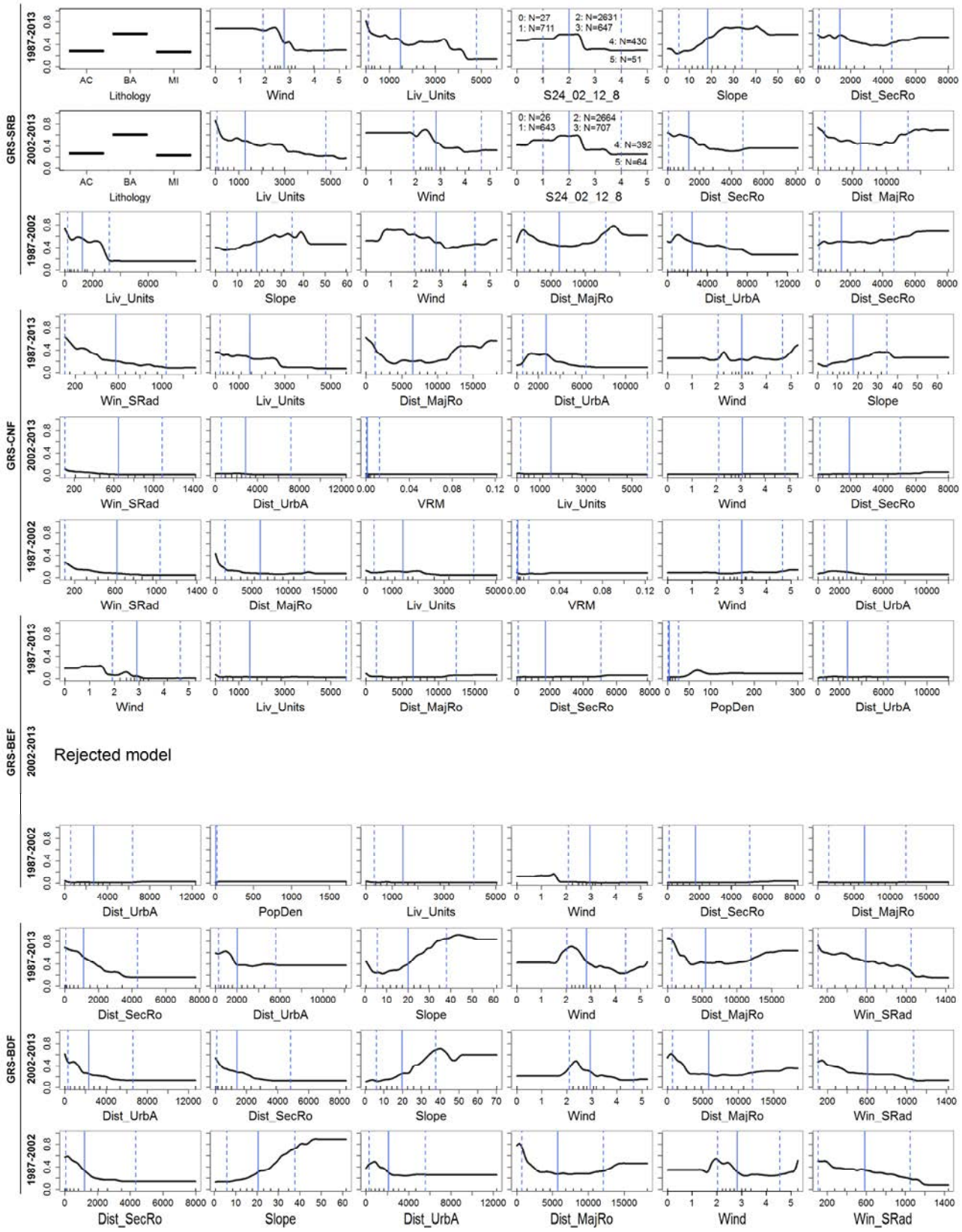


Figura A6. Gráficos de dependencia parcial para las seis primeras variables por orden de importancia. Ámbito NE1. Transiciones desde prado (GRS) a matorral (MTR), frondosas perennifolias (BEF) y frondosas caducifolias (BDF). Modelos BRT que no incluyen la variable altitud. Eje Y: probabilidad. Líneas negras verticales sobre el eje X: deciles en la distribución de valores de la variable. Líneas azules verticales: percentiles 0.05, 0.50 y 0.95.

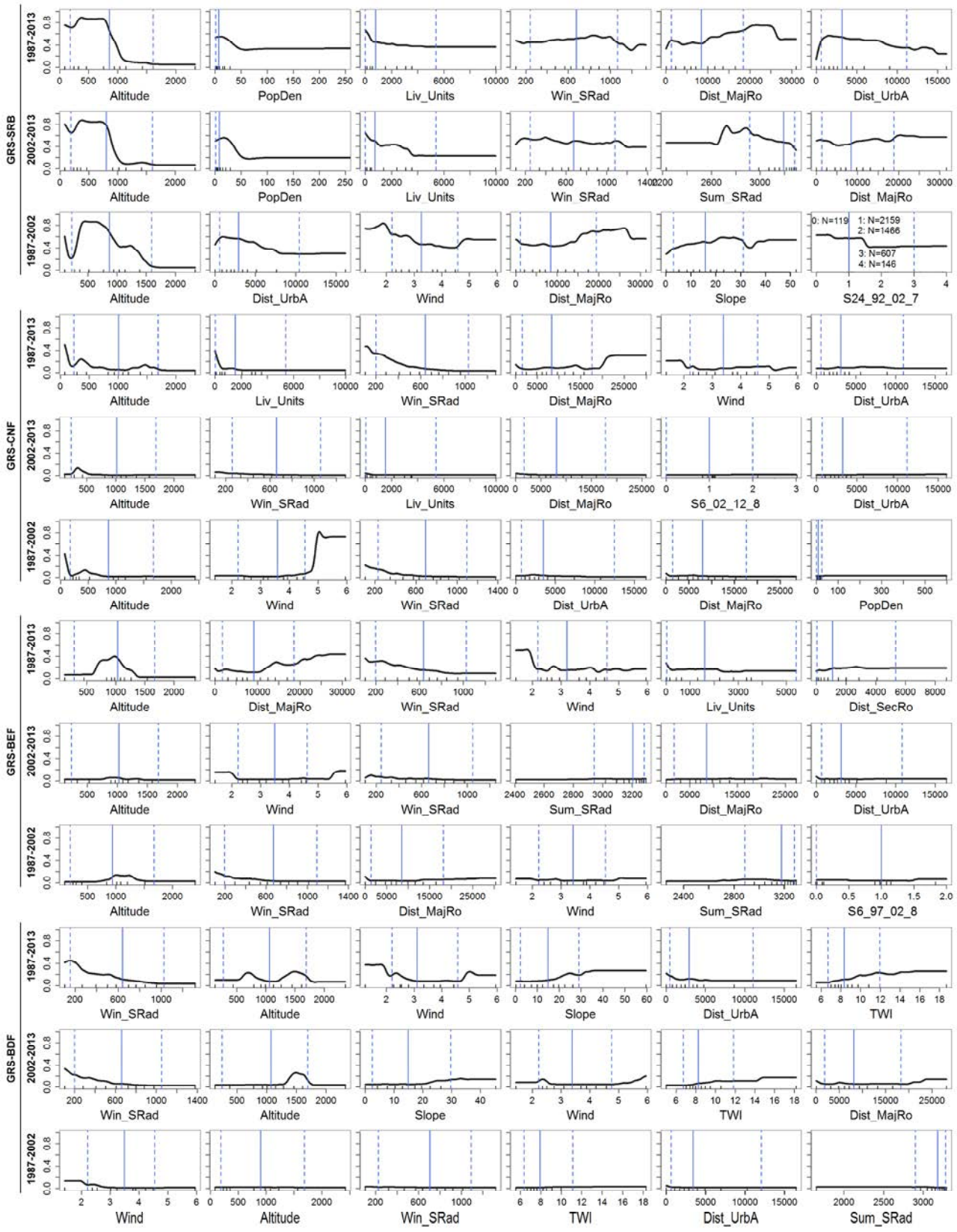


Figura A7. Gráficos de dependencia parcial para las seis primeras variables por orden de importancia. Ámbito NE2. Transiciones desde prado (GRS) a matorral (MTR), frondosas perennifolias (BEF) y frondosas caducifolias (BDF). Modelos BRT que incluyen la variable altitud. Eje Y: probabilidad. Líneas negras verticales sobre el eje X: deciles en la distribución de valores de la variable. Líneas azules verticales: percentiles 0.05, 0.50 y 0.95.

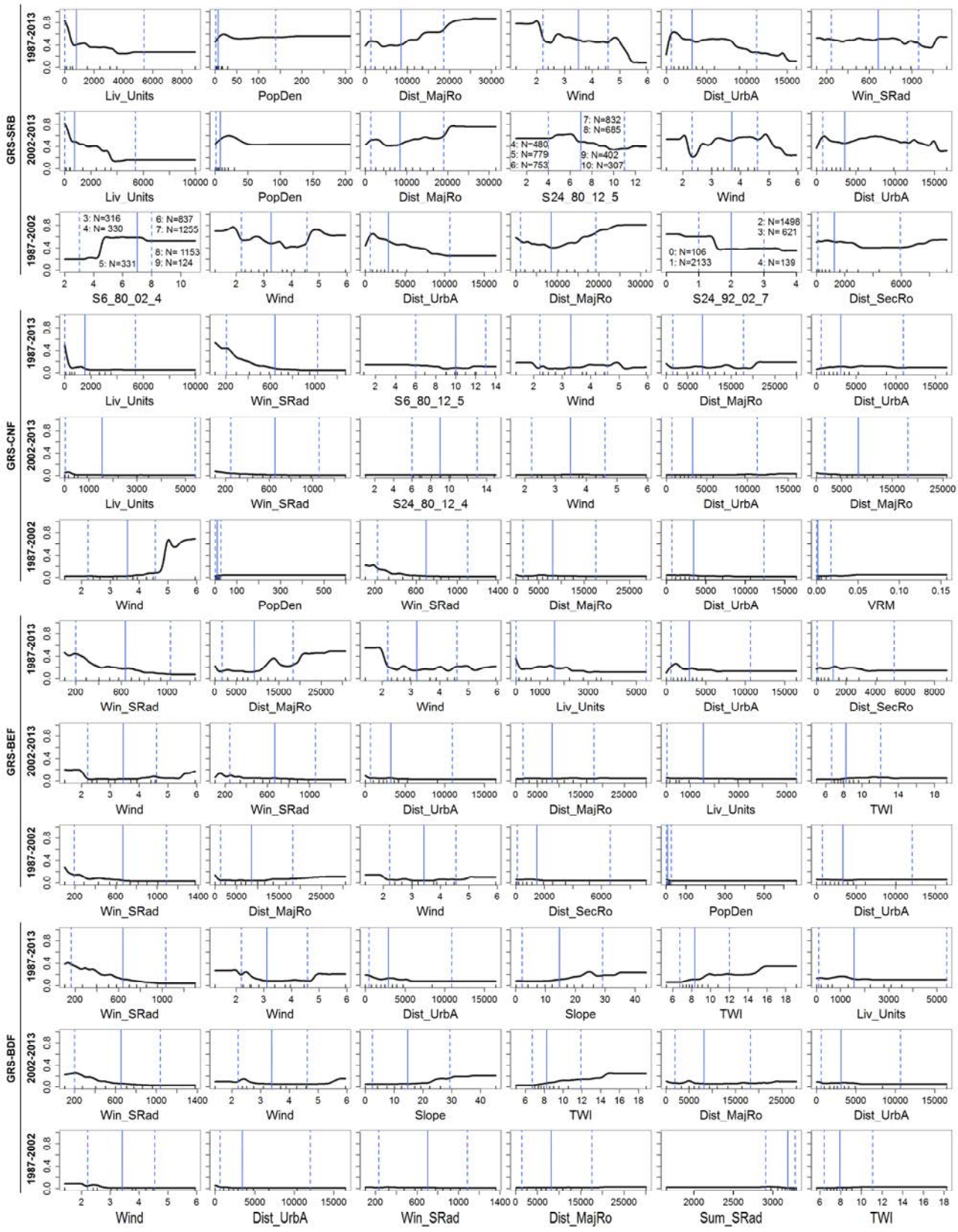


Figura A8. Gráficos de dependencia parcial para las seis primeras variables por orden de importancia. Ámbito NE2. Transiciones desde prado (GRS) a matorral (MTR), frondosas perennifolias (BEF) y frondosas caducifolias (BDF). Modelos BRT que no incluyen la variable altitud. Eje Y: probabilidad. Líneas negras verticales sobre el eje X: deciles en la distribución de valores de la variable. Líneas azules verticales: percentiles 0.05, 0.50 y 0.95.

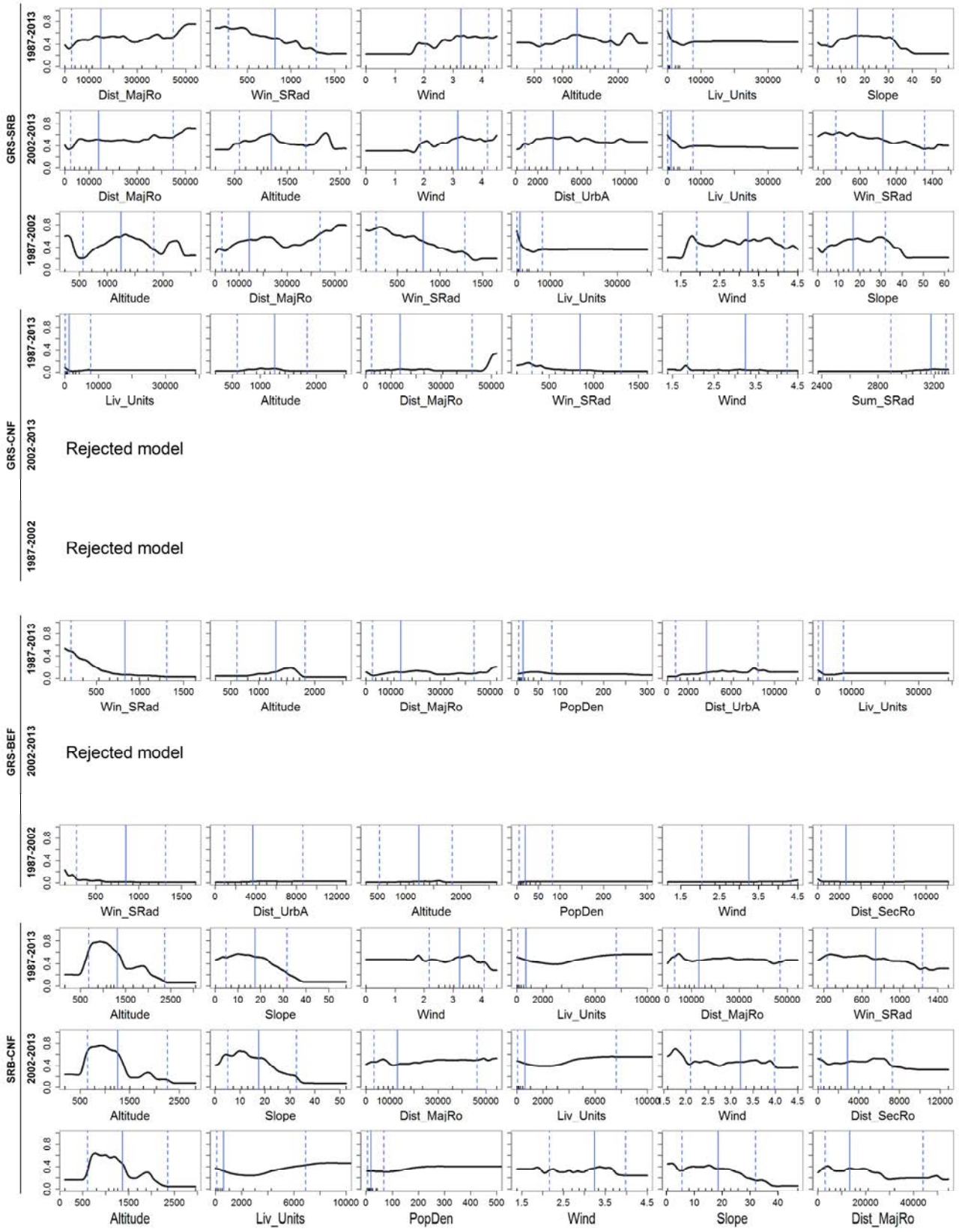


Figura A9. Gráficos de dependencia parcial para las seis primeras variables por orden de importancia. Ámbito SE3. Transiciones desde prado (GRS) a matorral (MTR), frondosas perennifolias (BEF) y frondosas caducifolias (BDF). Modelos BRT que incluyen la variable altitud. Eje Y: probabilidad. Líneas negras verticales sobre el eje X: deciles en la distribución de valores de la variable. Líneas azules verticales: percentiles 0.05, 0.50 y 0.95.

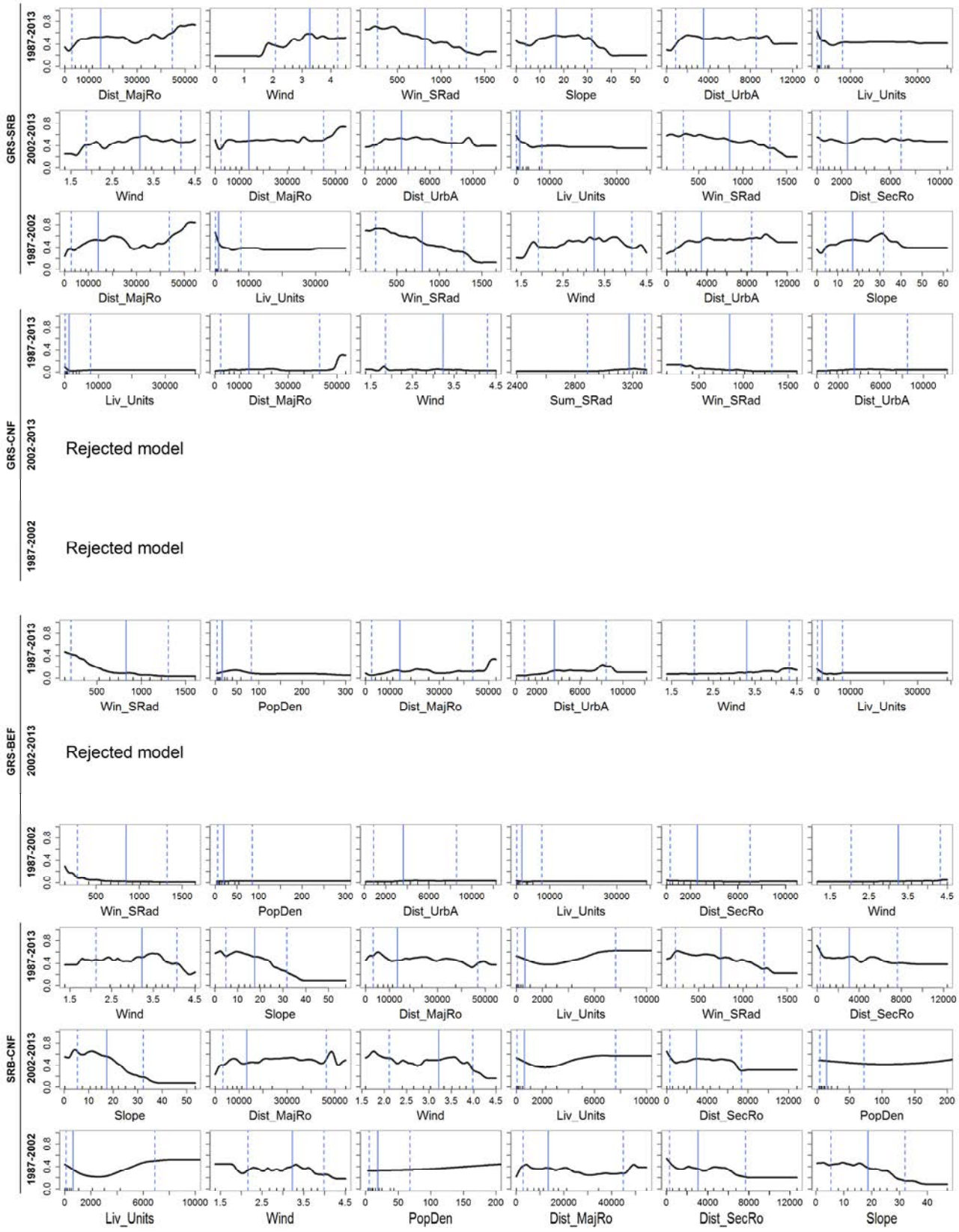


Figura A10. Gráficos de dependencia parcial para las seis primeras variables por orden de importancia. Ámbito SE3. Transiciones desde prado (GRS) a matorral (MTR), frondosas perennifolias (BEF) y frondosas caducifolias (BDF). Modelos BRT que no incluyen la variable altitud. Eje Y: probabilidad. Líneas negras verticales sobre el eje X: deciles en la distribución de valores de la variable. Líneas azules verticales: percentiles 0.05, 0.50 y 0.95.

	198030						**198031** (NE1)						**199030**						**199031** (NE2)						**200034** (SE3)					
	1987		**2002**		**2012**		**1987**		**2002**		**2012**		**1987**		**2002**		**2012**		**1987**		**2002**		**2012**		**1987**		**2002**		**2012**	
	PA	**UA**	**PA**	**UA**	**PA**	**UA**	**PA**	**UA**	**PA**	**UA**	**PA**	**UA**	**PA**	**UA**	**PA**	**UA**	**PA**	**UA**	**PA**	**UA**	**PA**	**UA**	**PA**	**UA**	**PA**	**UA**	**PA**	**UA**	**PA**	**UA**
Grasslands	95.1	91.6	65.9	85.9	73.7	93.5	96.0	82.7	95.2	79.4	95.9	77.7	98.5	95.8	70.0	89.6	96.5	93.8	92.1	78.3	91.1	69.3	95.5	68.1	97.4	82.5	97.3	92.7	96.6	90.4
Shrublands	97.3	87.9	77.9	73.3	97.1	98.5	92.1	96.2	85.8	93.4	91.6	97.3	94.0	93.0	62.0	65.7	95.3	91.5	90.2	86.4	83.9	81.3	89.8	96.0	95.6	93.8	91.7	91.6	92.4	89.8
Coniferous forests	98.0	99.8	91.5	88.2	98.6	99.0	97.1	98.8	96.2	97.9	99.0	98.5	98.5	98.7	86.6	89.2	98.6	99.5	95.3	99.5	97.1	99.1	99.0	98.8	97.8	99.6	98.6	92.9	94.5	98.9
Broadleaf evergreen forests	98.3	82.2	95.1	90.8	99.0	90.6	91.5	74.9	92.4	67.8	93.0	89.9	73.0	68.3	81.1	69.1	65.8	66.1	97.6	69.1	96.2	70.9	96.5	89.6	98.9	86.5	95.3	91.3	91.1	93.5
Broadleaf deciduous forets	99.1	94.8	75.2	73.4	99.4	97.2	98.8	95.3	98.2	95.8	99.0	97.8	95.5	95.6	86.3	83.6	98.6	98.3	95.1	93.0	95.2	84.2	96.7	91.3	98.0	86.6	96.8	95.3	97.0	86.9
Rainfed herbaceous crops	98.3	94.5	89.0	92.7	97.4	97.9	98.8	98.2	96.4	95.8	98.0	99.8	99.2	83.0	88.6	93.4	98.9	97.7	98.6	99.4	97.3	99.0	98.5	99.7	98.2	93.8	98.0	98.3	98.4	98.7
Irrigated herbaceous crops	96.1	97.6	74.2	72.3	96.7	96.9	92.8	97.5	85.0	95.6	95.4	98.4	87.4	98.7	69.3	71.8	86.5	89.4	93.1	96.2	87.4	94.2	94.9	99.3	98.4	94.3	97.4	96.2	98.4	99.6

Tabla A1. Porcentajes de acierto por categorías según exactitud de productor (PA) y exactitud de usuario (UA) de los mapas de usos y cubiertas del suelo obtenidos mediante el clasificador kNN.