



*Escuela Técnica Superior de Ingenieros de Caminos,  
Canales y Puertos.*  
**UNIVERSIDAD DE CANTABRIA**



# **Determinación de patrones espaciales de las comunidades de macroinvertebrados de la Península Ibérica utilizando la plataforma Iber-Rio**

Trabajo realizado por:  
***Miguel Aja Barquín***

Dirigido:  
***José Barquín Ortíz***

Titulación:  
**Máster Universitario en  
Gestión Integrada de Sistemas  
Hídricos**

Santander, Octubre de 2017

**TRABAJO FINAL DE MASTER**



## Tabla de contenido

1.	RESUMEN .....	3
2.	INTRODUCCIÓN .....	4
3.	OBJETIVOS .....	6
4.	DESCRIPCIÓN DATOS ENTRADA .....	7
4.1.	Área de estudio .....	7
4.2.	Red fluvial.....	11
4.3.	Variables dependientes.....	12
4.3.1.	Datos iniciales.....	12
4.3.2.	Correcciones de errores topológicos y asignación de variables ambientales a puntos de muestreo .....	13
4.3.3.	Depuración de los datos biológicos para la selección de puntos de muestreo ..	15
4.3.4.	Comunidades de macroinvertebrados.....	16
4.4.	Variables independientes.....	18
5.	METODOLOGÍA DE MODELADO .....	21
5.1.	Preparación de datos para el modelado .....	21
5.2.	Random Forest .....	22
6.	RESULTADOS .....	24
6.1.	Resultados de correlación .....	24
6.2.	Resultados del modelado.....	26
6.2.1.	Resultados numéricos .....	26
6.2.2.	Resultados de los valores de importancia con el IncNodePurity index y mapas de predicción.....	27
7.	DISCUSIÓN.....	41
7.1.	Selección de datos validos e índices .....	41
7.2.	Selección de variables .....	42
7.3.	Modelo .....	45
7.4.	Modelado: patrones espaciales y mapas de predicción. ....	46
8.	CONCLUSIONES .....	49
9.	VÍAS DE TRABAJO FUTURO.....	50
10.	BIBLIOGRAFÍA.....	51
11.	ANEXOS .....	57
	Anexo 1: Taxones de macroinvertebrados bentónicos presentes en los datos NABIA. ....	57
	Anexo 2. Variables independientes introducidas al modelo.....	59
	Anexo 3: Correlación de Spearman Rank entre los índices de calidad y las variables independientes. ....	64

Anexo 4: Tamaño de las muestras introducidas al modelo y estadísticas básicas. .... 70  
Anexo 5: Importancia de las variables según el IncNodePurity ..... 70

## 1. RESUMEN

La normativa relativa a las aguas en España viene definida por la Directiva Europea Marco del Agua (DMA). Esta exige la evaluación del estado ecológico de los ecosistemas acuáticos, y en entre otras, la evaluación de las comunidades de macroinvertebrados bentónicos fluviales. Para llevar una correcta gestión y conservación del medio ambiente se necesitan nuevas herramientas que permitan conocer los comportamientos de las comunidades de los ecosistemas acuáticos. De este modo el presente estudio pretende modelar los patrones espaciales de distribución de macroinvertebrados para toda la Península Ibérica, mediante modelo Random Forest (RF). Las variables independientes usadas fueron la red fluvial virtual Iber-Rio, desarrollada por el Instituto de Hidráulica Ambiental (IH), y sus variables descriptoras de la cuenca (geológicas, topográficas, usos de suelo, impactos antrópicos y climáticas) incluidas en esta red.

Para la creación de las comunidades sintéticas se crearon diferentes variables respuesta mediante la base de datos nacional de muestreo, NABIA. Las variables fueron los índices: Iberian Biological Monitoring Working Party (IBMWP); Iberian Average Score Per Taxon (IASPT); Efemerópteros, Plecópteros y Tricópteros con riqueza de taxones y abundancia relativa (EPT-R, EPT-A); Gastrópodos, Oligoquetos y Dípteros (GOLD) y la Riqueza de taxones. A las 2930 estaciones de muestreo se les asignó un tramo de la red Iber-Rio. Sobre este tramo de río se calcularon las diferentes métricas representativas de las comunidades de macroinvertebrados. Una vez unidos, mediante el método de Spearman Rank se calculó la correlación entre pares de variables. Seleccionando aquellas que presentasen una correlación mayor a 0,25, para introducirlas al RF y obtener los patrones espaciales de dichas comunidades y unos mapas de distribución de las comunidades en la Península Ibérica.

Las variables con mayor influencia sobre la distribución fueron la precipitación, los usos de suelo agrícolas y urbanos; la pendiente del río y del valle; así como la dureza de la roca en la cuenca y el porcentaje de sedimentos en la cuenca. Respecto a la información obtenida de los mapas se observan dos gradientes muy claros, una disminución de taxones de norte a sur, y también de las zonas de montaña hacia las partes bajas de los valles.

Estas conclusiones se validaron con la bibliografía existente cuyo resultado fue que el modelo ha predicho de manera notoria aquellos patrones que distribuyen las comunidades por la Península Ibérica. Mostrando a los modelos como una buena herramienta para la obtención de datos científicos y aplicable al mundo de la gestión de ecosistemas fluviales.

## 2. INTRODUCCIÓN

Una buena gestión de los ecosistemas acuáticos comienza aumentando el conocimiento del entorno en todas sus facetas. En este sentido, conocer con detalle la composición y estructura de las comunidades biológicas y los principales factores que determinan estos patrones es un aspecto clave para poder gestionar correctamente dichos ecosistemas.

Las aguas dulces son mucho más ricas por unidad de volumen de hábitat que las aguas saladas y, a pesar de suponer el 0,01% de la superficie total del planeta, poseen el 9,5% de especies reconocidas a nivel mundial (UNEP, 2002). De toda esa diversidad los macroinvertebrados es el grupo focal de este trabajo. En términos generales, definimos los macroinvertebrados como aquellos invertebrados que se pueden ver a simple vista o bien que son retenidos por una red de malla de aproximadamente 125  $\mu\text{m}$  (Rosenberg and Resh, 1993). Esta definición, aun siendo muy aceptada, es arbitraria. Por lo que podemos ser un poco más precisos definiendo los macroinvertebrados con base en la taxonomía. Este grupo tiene representantes en muchos filos de animales, entre ellos: Arthropoda, Mollusca, Annelida, Platyhelminthes, Nematoda y Nematomorpha, y para este trabajo aquellos situados en los lechos fluviales españoles (Oscoz, 2011).

Con los macroinvertebrados bentónicos de la Península Ibérica se pretende obtener aquellos patrones espaciales que determinan su distribución. Este trabajo ya se ha realizado en otros trabajos científicos similares con fauna (Alvarez-Cabria, 2010, 2017; Manolaki & Papastergiadou, 2013). Sin embargo, la novedad de este trabajo se encuentra en su escala, de la que no se han encontrado estudios previos, y si se han encontrado han sido en otras materias sin considerar la fauna presente (Snelder et al, 2011; 2013).

Para la consecución del presente se utilizará un modelo de regresión (en una de sus facetas), Random Forest (Breiman, 2001), el cual estima la importancia de las variables ambientales introducidas, manejando bien la interacción entre ellas, aunque sean un número elevado, e identificando los “límites” no lineales de las variables.

Para crear espacio virtual se van a utilizar las variables y la red fluvial virtual de la plataforma Iber-Rio. Plataforma que desarrolla el Instituto de Hidráulica Ambiental de Cantabria (IH), generando una red virtual de los ríos de la Península Ibérica con las características presentes que interaccionan con el agua. Dicha red se desarrolló a partir del proyecto MARCE dentro del Programa Nacional de Proyectos de Investigación Fundamental no orientada en el marco del Plan Nacional de I+D+i. 2008-2011. Llevado a cabo por el IH y el Centro de Estudios y

Experimentación de Obras Públicas (CEDEX). Exclusivamente dentro de las cuencas de la Confederación hidrográfica del Norte, Agencia Vasca del Agua, la Agencia Catalana del Agua y la Confederación Hidrográfica del Ebro.

### 3. OBJETIVOS

El objetivo principal del trabajo es la identificación de los patrones espaciales que determinan la composición de las comunidades de macroinvertebrados bentónicos en el territorio de la Península Ibérica a través de la plataforma Iber-Rio.

Dichos objetivos se alcanzarán a través de diferentes etapas:

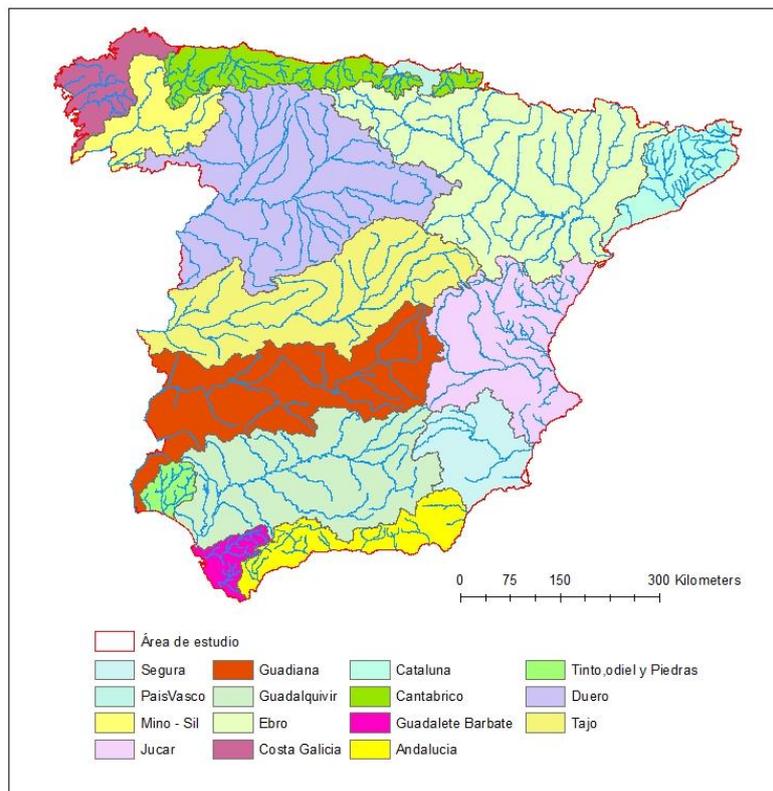
1. Elección de métricas que describan las comunidades de macroinvertebrados bentónicos.
2. Selección de los principales factores ambientales que puedan determinar los patrones espaciales de las comunidades de macroinvertebrados bentónicos.
3. Generación de un modelo de regresión que nos permita encontrar los patrones y factores más relevantes.
4. Evaluación de los resultados de dicho modelo mediante bibliografía.
5. Elaboración de mapas que reflejen los patrones de variabilidad espacial de las comunidades de macroinvertebrados bentónicos fluviales más importantes sobre toda la Península.

## 4. DESCRIPCIÓN DATOS ENTRADA

Para la elaboración de este estudio se ha de partir de una serie de datos, por un lado, aquellos datos de macroinvertebrados obtenidos de la base de datos NABIA, y por el otro, de las variables físicas desarrolladas por el Instituto de Hidráulica Ambiental de Cantabria, a partir de la red de Cuencas Virtuales Iber-Rio.

### 4.1. Área de estudio

Como se ha indicado al inicio, se pretende hacer un modelo de distribución para toda España. El área de estudio comprende todas las confederaciones hidrográficas del país (Figura 1), si bien es cierto que únicamente comprende el área peninsular, ocupando un territorio de 493770 Km<sup>2</sup>.



**Figura 1: Área de estudio y división administrativa hidrológica de España.**

España posee una orografía compleja, donde la situación de las cordilleras genera diferentes tipos de cauces, unos más extensos y caudalosos; otros con cuencas más reducidas y de corta distancia. En el tercio más septentrional se sitúa la Cordillera Cantábrica, uniéndose con los Pirineos por el este mediante una depresión del relieve en la parte oriental. Al estar cerca del

Mar Cantábrico, genera en su vertiente norte, un mosaico de ríos cortos, encajonados en la roca y caudalosos.

En la parte central de la Península, se ha formado una gran meseta con una altitud media de 660 msnm que permite un desarrollo fluvial diferente. Los ríos son redes amplias con caudales mucho mayores como lo es el río Duero, el Tago o el Ebro. Esta zona está dividida en tres partes, en el norte se sitúa el río Duero; en la parte media se encuentra el río Tago, y en la parte sur el río Guadiana, por el Sistema Central y los montes de Toledo respectivamente.

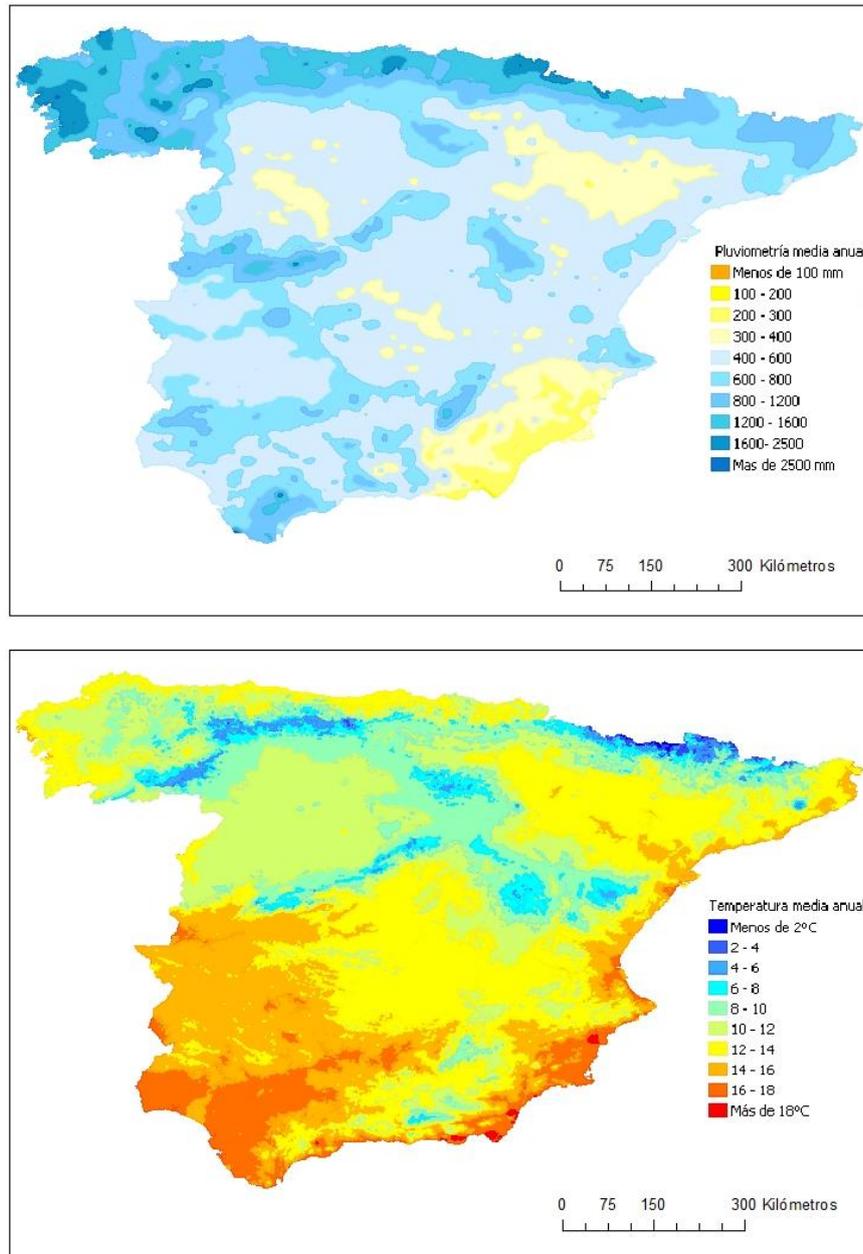
Al este de la Meseta Central se encuentra el Sistema Ibérico, que con los Pirineos al norte genera la depresión del río Ebro; con una orientación noroeste-sureste y picos de entre 1800 a 2200 msnm, llega a la zona mediterránea de Valencia suavizándose el relieve hasta llegar al mar. Por otro lado, Sierra Morena separa la Meseta Central de la Depresión bética. Se trata de una cadena montañosa de altura más bien baja (pico máximo 1324 msnm). Finalmente tenemos en la zona sureste de la Península Ibérica, donde el Sistema Bético, con sus 600 kilómetros de longitud, toca con sus extremos el Océano Atlántico y el Mar Mediterráneo. Es sin duda el mayor conjunto montañoso de España cuya cercanía al mar, crea ríos cortos, pero más largos que en el norte, debido al laberinto que representan sus picos.

A nivel geológico podemos dividir la Península en 5 zonas (IGME 2015). El primer gran bloque se encuentra en Cordillera Cantábrica; se puede diferenciar en oriental y occidental. En la primera parte nos encontramos dolomías, calizas y margas, areniscas y conglomerados, en cambio en la parte occidental presenta una mayor diversidad, con grandes bloques de granitoides, rocas básicas y ultrabásicas, junto a las cuarcitas, pizarras, areniscas y calizas. En las tres grandes depresiones de la parte central (Duero, Ebro y Bética) se encuentran dominadas por areniscas, conglomerados, arcillas, calizas y evaporitas, y en las zonas más cercanas a los cauces aparecen gravas, conglomerados, arenas y limos. Otro gran conjunto bastante heterogéneo es la parte oeste de España que contiene: granitoides, pizarras cuarcitas y conglomerados, con una predominancia de las areniscas y calizas en la parte más al suroeste. La última delimitación se encuentra en el Levante (zona mediterránea), donde encontramos calizas dolomías y margas, con areniscas.

El clima presente dentro de la zona de estudio se puede resumir en dos biorregiones: Templado y Mediterráneo. En toda la vertiente norte de la Cordillera Cantábrica y los Pirineos, se encuentran los bioclimas: templado oceánico, templado oceánico submediterráneo y templado hiperoceánico submediterráneo. El resto del territorio se encuentra dominado por: el bioclima mediterráneo pluviestacional, aunque en la parte interior encontramos otros como el

mediterráneo pluviestacional continental y mediterráneo oceánico estepario, mediterráneo xérico oceánico estepario, templados oceánico submediterráneo, templado oceánico estepario y templado xérico, además del mediterráneo xérico oceánico y mediterráneo desértico oceánico, presentes y dominantes en el sureste de España (Rivas-Martínez et al. 2004).

Para completar la descripción climática, en la figura 2 se muestra la pluviometría media anual (mm) y la temperatura media anual (°C) (IDEE MAPAMA). Las lluvias se concentran en la parte norte de la Península, debido a que los vientos del noroeste se encuentran cargados de humedad del océano y se detienen por la Cordillera Cantábrica, generando precipitaciones superiores a 1200 mm/año. También podemos observar que las acumulaciones de precipitación coinciden con las zonas montañosas, dibujando, a modo general, un plano de la orografía. El gradiente de distribución de precipitaciones es descendente a medida que avanzamos hacia el sur, con algunas zonas de precipitación escasa en la parte sureste del país (menores de 300 mm) propias de un clima mediterráneo. Por otro lado, tenemos la temperatura, este parámetro sufre un gradiente ascendente de norte al sur. En el norte se comprende como un clima oceánico con temperaturas suaves debido a la cercanía al mar y sometida a los vientos marinos del océano. El resto del país se encuentra dominado por el clima mediterráneo, con temperaturas medias superiores a 16 °C. Se puede dividir en dos partes, una más continental, situada en la parte central, donde las temperaturas son más extremas; con veranos calurosos e inviernos fríos y otra cercana al mar, con temperaturas elevadas y temperaturas mínimas suavizadas por la acción del mar.



**Figura 2: Mapa pluviométrico de España, con valores medios anuales (arriba). Mapa de temperaturas medias anuales de España (abajo). (Fuente: MAPAMA, IDEE, agricultura).**

El conjunto administrativo de los cursos fluviales está formado por 15 confederaciones hidrográficas. Siendo gestionados de diferentes maneras: aquellos que se encuentran situados entre dos o más Comunidades Autónomas se denominan intercomunitarios, y están gestionados de manera estatal y otros que se encuentran gestionados por la Comunidad Autónoma, dado que todo el territorio de la cuenca se encuentra dentro de la misma. Entre las primeras, de norte a sur, se encuentran: la Confederación Hidrográfica del Cantábrico; la Confederación del Miño-Sil; la Confederación Hidrográfica del Ebro y la Confederación del Duero. Las Demarcaciones

Hidrográficas del Tajo, el Guadiana, el Júcar, el Guadalquivir y el Segura. Por otro lado, encontraríamos a la Demarcación Costa Galicia; la Agencia Vasca del Agua; Agencia Catalana del Agua; las Cuencas Andaluzas Mediterráneas; la Demarcación Tinto, Odiel y Piedras, además de la de Guadalete y Barbate. A pesar de estas divisiones, todas tienen como marco legislativo la normativa española (LEY 10/2001, de 5 de julio, del Plan Hidrológico Nacional y su modificación LEY ORGÁNICA 2/2005), que se deriva de la Directiva Europea Marco del Agua (Directiva 2000/60/CE) y sus objetivos.

A nivel poblacional y usos de suelo, la zona de estudio es muy amplia y diversa, aunque tiene unos patrones muy claros: las densidades más altas (superiores a 100 habitantes/Km<sup>2</sup>) se presentan en la Comunidad de Madrid y en la costa, tanto mediterránea como atlántica. En cambio, las zonas de interior se encuentran con 50 habitantes Km<sup>2</sup> en los núcleos de mayor concentración (INE, 2011). España es un país eminentemente dedicado al sector servicios (65% de la población activa); con un tercio de la población dedicada al sector secundario y el resto al primario (5%). La agricultura, siendo el sector que menos población activa se encuentra trabajando en ella, ocupa el 50 % de la superficie del país, sin contar las grandes extensiones ganaderas del norte peninsular y las dehesas del suroeste (Agricultura en España, s.f.). Por otro lado, las grandes concentraciones industriales se sitúan en Madrid, Barcelona, País Vasco y Galicia.

#### 4.2. Red fluvial

El elemento principal sobre el que se fundamenta el proyecto es la Red Fluvial Iber-Rio, desarrollada por el IH Cantabria para toda la Península Ibérica.

Dicha red tiene su origen en el proyecto MARCE (PLAN NACIONAL I+D+I 2009-2011) donde el IH, en colaboración con el Centro de Estudios y Experimentación de Obras Públicas (CEDEX) generaron los cauces para un territorio que comprendía la Confederación Hidrográfica del Ebro (CHE); la Agencia Vasca del Agua (URA); la Confederación Hidrográfica del Cantábrico (CHC) y la Agencia Catalana del Agua (ACA). Posteriormente, se configuró la red para toda la Península, conformada por un total de 139119 tramos de río, que van desde 6,7 hasta los 1517 metros, cuya media se encuentra entorno a los 986,5 metros. Dichos tramos son estrictamente fluviales de aguas corrientes, eliminando las zonas estuarinas y embalsadas.

A través del software “NetStream” (Miller 2003; [www.netmaptools.org](http://www.netmaptools.org)) determinarían las vías de flujo, de un Modelo Digital de Elevaciones (DEM) de 30 metros de resolución espacial;

creando tramos a partir de las características de zona de descarga, gradiente de las laderas, curvatura en planta y criterios de longitud de cauces. Donde se pueden aplicar variables ambientales, con los diferentes roles e interacciones que nos permitirá inferir los principales impactos sobre los ecosistemas dulceacuícolas (Barquín et al, 2015).

### 4.3. Variables dependientes

#### 4.3.1. Datos iniciales

La base de datos utilizada es NABIA. Es una base de datos de la Dirección General del Agua, del Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente (MAPAMA). Esta base de datos se originó, en 2012 para el establecimiento de mecanismos de garantía y aseguramiento de calidad en todo el proceso de generación de datos, con coordinación y de manera homogénea dentro de las redes de seguimiento del estado de las masas de agua superficiales, emanado de la Directiva Marco del Agua (Directiva 2000/60/CE).

Esta base de datos posee 3370 estaciones de muestreo con valores relativos al periodo 2005-2015. De esta base de datos se han escogido aquellos relacionados con los macroinvertebrados bentónicos y, aunque no ninguna estación tenía el periodo de muestreos completo, no se ha hecho una preselección temporal, bajo el concepto de que los modelos espaciales recrean una imagen estática en el tiempo (Mateo et al., 2011).

De los datos que presentes se escogieron: el código del punto de muestreo, la fecha, el taxón (código), la categoría de la medición (abundancia, abundancia relativa, presencia/ausencia), el valor de medición, y el método de muestreo. Respecto a este último, se observaron los diversos protocolos de muestreo (8 en total) que implicarían la variedad de tipos de datos a procesar. Nos encontramos con:

**ML-Rv-I-2011 y ML-Rv-I-2013:** se trata de un método de muestreo cuantitativo que muestrea la composición y abundancia de la fauna bentónica de invertebrados, que se lleva a cabo desde septiembre de 2013. Las métricas principales para este método son Iberian Biological Monitoring Working Parti (IBMWP) (Alba- Tercedor & Sánchez-Ortega, 1988) y el Multimétrico específico del tipo (METI-2015), aunque también acepta otros.

**IBMWP – GUALDALMED** (Pablo Jáimez-Cuéllar et al., 2002): método desarrollado para el proyecto GUALDALMED, este permite hacer una valoración ecológica rápida, considerándose un

RBP (Rapid Bioassessment Protocol) (Barbour et al., 1999). En el caso que nos ocupa, puede ser cuantitativo o cualitativo, puesto que el índice de calificación es el IBMWP.

**IBMWP-C:** es un protocolo de carácter cualitativo que únicamente valora la presencia de las especies, a las que luego se les asigna un valor del IBMWP

**IBMWP-S:** es un protocolo de carácter semicuantitativo, donde el taxón toma un valor de una escala según la abundancia de este. Normalmente están divididos en 4 grupos.

**UNEEN 9391:1995:** Muestreo de macroinvertebrados en aguas profundas. Guía de utilización de aparatos de toma de muestra de colonización en sustratos artificiales, cualitativos y cuantitativos.

**DIPPING-CHE-ACA:** muestreo de aguas lenticas someras realizadas con un salabre en el seno de agua, por la Confederación Hidrográfica del Ebro y la Agencia Catalana del Agua.

**Litoral – ACA:** muestreo de aguas lenticas someras realizadas con un salabre en el seno de agua, en zonas de litoral por la agencia Catalana del Agua.

#### 4.3.2. Correcciones de errores topológicos y asignación de variables ambientales a puntos de muestreo

Para desarrollar el modelo se precisa de la introducción de los datos muestreados en los tramos de la Red Iber-Rio, de esta manera podremos relacionar las variables independientes (apartado 3.4) con las comunidades de macroinvertebrados del mismo segmento fluvial.

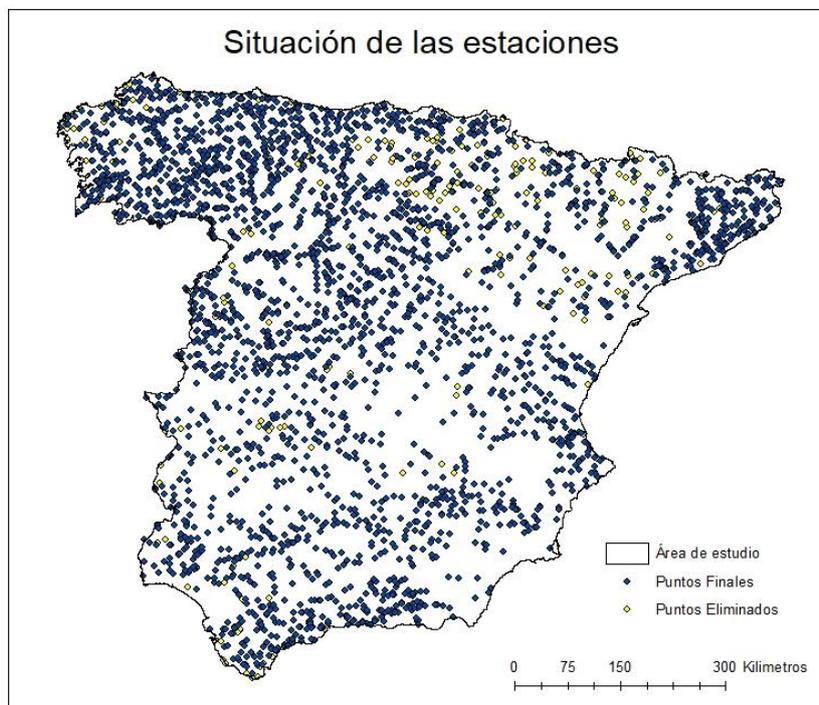
Para dicho cometido se precisa de la herramienta ArcMAP de ArcGIS, versión: 10.4.0.5524, donde se dibujan las 3370 estaciones de muestreo mediante sus coordenadas, junto con la Red Iber-Rio. En la primera fase se filtran aquellas estaciones que se encuentran a 20 metros del tramo del río, considerando que la relación estación-tramo es directa y no presenta errores; el resto de puntos de muestreo las asignaremos a la red tras un análisis minucioso de cada uno de ellos.

Con la ayuda de un mapa de los ríos de España (“Atlántico” y “Mediterráneo”), proporcionado por el CEDEX (“Clasificación Hidrográfica de los Ríos de España, Quintas, 2016) discretizaremos mejor la situación de los puntos de muestreo. Uno de los factores más relevantes de la Red fluvial de Iber-Rio, es la representación de los cauces principales, dejándose de lado aquellos que presentan caudales poco representativos. Algunos ríos de corta distancia situados en zonas

costeras no han sido representados, de modo que los puntos que no poseían un tramo de río asignable se han eliminado. Otras zonas conflictivas fueron los puntos situados en la alta montaña, donde no se representan con exactitud los órdenes más pequeños de los ríos. De este modo, el criterio de selección seguido ha sido la eliminación de aquellos puntos que se situasen a más de 100 metros del primer tramo aguas abajo del punto. Este problema de representación se ha producido en las zonas de desembocadura, encontrando puntos más lejanos a 100 metros que por precaución han sido eliminados.

Finalizado el proceso de corrección espacial, de la que se obtuvieron 3049 estaciones de muestreo, se procede a la asignación del código de identificación de cada tramo mediante software R versión 3.4.1 (R Core Team, 2017). De esta manera cada segmento de río tendrá una comunidad de macroinvertebrados procedentes de un muestreo, excepto 152 estaciones que se repartirán entre 76 tramos.

Como podemos ver en la figura 3, los puntos eliminados pertenecen en su mayoría a zonas de cabecera, sobre todo en la parte de la Cordillera Cantábrica y pirineos, en especial el río Ebro. De las zonas de costa cabe destacar la parte de Costa Galicia y la zona del Guadalquivir y Guadalete-Barbate.



**Figura 3: Distribución de los puntos antes y después de la selección.**

#### 4.3.3. Depuración de los datos biológicos para la selección de puntos de muestreo

Ante la heterogeneidad de los métodos de muestreo, se ha precisado una depuración de los datos biológicos iniciales. El primer factor de elección han sido los filos de macroinvertebrados bentónicos (*Arthropoda*, *Mollusca*, *Annelida*, *Platyhelminthes*, *Nematoda* y *Nematomorpha*) (Hanson, 2010). De esta manera se han eliminado aquellos datos no relativos al presente trabajo como fitobentos y especies de aguas saladas presentes en los resultados de los muestreos. Posteriormente se han filtrado de nuevo los datos, ante la presencia de taxones que no correspondían a especies estrictamente bentónicas, obteniendo 161 taxones con los que poder modelar.

Numerosas fuentes advierten de las dificultades de la identificación de los macroinvertebrados a nivel genérico o específico (Bonada et al., 2005) por lo que para evitar problemas se ha considerado detenernos en la Familia de las especies (Diez et al, 2004). No todas llegaban, sobre todo porque algunas identificaciones se detienen en el filo, como son los *Platyhelminthes*, otros en el Orden como los *Acari*, o en la Clase como los *Oligochaeta*. Esto en parte viene predefinido porque muchos de los muestreos poseen como objetivo medir la calidad de las aguas, siendo el índice IBMWP dominante, dejando de lado otros más complejos, como pudieran ser el índice de Shannon-Wiener (Shannon y Weaver, 1949) o el índice de Margalef (1969), donde se requiere un grado mayor de identificación de las especies. Por lo tanto, con la ayuda de la aplicación web ID TAX del MAPAMA, la guía "Identification guide of freshwater macroinvertebrates of Spain" (Oscoz, 2011), se han dejado todas aquellas especies que llegaban al taxón de Familia (148), luego a Clase (7), a Orden (4) y finalmente a filo (en aquellas cuya identificación permitían este grado de libertad, 2).

Una vez determinado los grados de identificación taxonómica, se observan diferentes errores que pueden provocar problemas a posteriori; en muchos casos se han encontrado celdas vacías o con numeraciones que no correspondían a las categorías de abundancia, abundancia relativa. La solución más apropiada ha sido cambiar su categoría a presencia/ausencia, debido a que muchos de los métodos de muestreo trabajan con los taxones de forma presencial a través del método visual, por lo que dicha información es válida y puede ser de utilidad para el modelo

El siguiente método de selección se llevó a cabo en dos fases; en primera instancia se eliminarían aquellos taxones con una abundancia relativa menor a 0,25 % en el 80 % de las estaciones, quitando 14 taxones. Una vez filtrados, se transformaban todos los datos a presencia/ausencia y se eliminaban aquellos taxones que tuvieran una presencia inferior al 0.25 % de los sitios,

eliminándose 15 taxones más. De esta manera se eliminaron aquellos grupos taxonómicos que por su baja frecuencia podían producir ruidos en los índices y las correlaciones con las variables independientes. Tras la selección quedaron 127 familias, 1 clases, 2 ordenes, 1 superorden y 1 filo ([Anexo 1](#)).

#### 4.3.4. Comunidades de macroinvertebrados

Para caracterizar las comunidades de macroinvertebrados bentónicos se van a utilizar una serie de variables derivadas de la composición y estructura de las mismas. Estas nos permitirán identificar diferentes patrones espaciales atendiendo a diferentes características de las comunidades de macroinvertebrados. En concreto se han escogido 6 variables para la caracterización de las comunidades biológicas:

- Iberian Biological Monitorin Working Party (IBMWP),
- Iberian Average Per Taxon (IASPT).
- Riqueza o número de taxones de “*Ephemeroptera, Plecoptera y Trichoptera*” (EPT-R).
- Abundancia relativa de “*Ephemeroptera, Plecoptera y Trichoptera*” (EPT-A).
- “*Gastropoda, Oligochaeta y Diptera*” (GOLD).
- Riqueza taxonómica o número de taxones (R).

Estas variables han sido seleccionadas porque combinan diferentes características de las comunidades de macroinvertebrados, aunque estén enfocadas en su mayoría al sector de la evaluación de la contaminación orgánica. A continuación, explicamos cada una de ellas y las características más destacables de las mismas.

Los EPT son insectos con una fase larvaria acuática. Los órdenes *Trichoptera* y *Plecoptera*, se desarrollan bien en aguas frías en altas altitudes, además este tipo de aguas, poseen un grado elevado de oxígeno disuelto, al que están muy ligados (*Plecoptera* y *Trichoptera*). Estos insectos tienen alas que desarrollan en su fase adulta, puesto que sufren derivas, dada la irregularidad de las aguas de alta montaña, ya sea por altas lluvias o por intermitencias del caudal, de modo que cuando son adultos, salen del agua y ascienden río arriba (Angelier E., 2002 pp: 27 - 38), no son muy buenos voladores, por lo que una barrera con árboles de ribera les ayuda en su ascenso. Otra característica por destacar es que en la mayoría de los casos son detritívoros, y los ciclos suelen estar muy ligados a la caída de la hoja en otoño.

**EPT-R:** Esta variable incluye el número de taxones o riqueza de tres órdenes, *Ephemeroptera*, *Plecoptera* y *Trichoptera*. Donde los valores más altos, significan una mayor riqueza de este grupo de estos órdenes.

**EPT-A:** Esta variable incluye la abundancia relativa de tres órdenes, *Ephemeroptera*, *Plecoptera* y *Trichoptera*. Donde los valores más altos, significan la población relativa de este grupo de estos órdenes respecto al total de la muestra.

**GOLD:** Esta variable se ha calculado sumando la abundancia relativa de Gasteropoda, Oligochaeta y Diptera. Estos taxones suelen encontrarse en las zonas lóaticas donde se retienen sedimentos más finos. Por regla general, en estas zonas se presentan como zonas con bajo oxígeno disuelto y de temperaturas mayores a las que podemos encontrar en las partes altas de la montaña. Son prácticamente recolectores de partículas de materia orgánica. Por lo que su presencia hace referencia a las partes bajas de las redes fluviales.

**IBMWP:** Esta variable es un índice de calidad del agua elaborado por Alba- Tercedor y Sanchez – Ortega (1988; Alba-Tercedor et al., 2004), se presenta como una adaptación del BMWP de Hellawell (1978), a la fauna bentónica de la Península Ibérica. Es un índice que trabaja a nivel de familia, asignándoles una puntuación, del 1 al 10 en función de la tolerancia o intolerancia de la especie a la contaminación orgánica. Entre los más intolerantes, a la contaminación se encuentran los órdenes de *Plecoptera*, *Ephemeroptera* y *Trichoptera*, seguidos de los *Odonata*. Los tres primeros, como se ha visto en el índice EPT, están “especializados” en ríos de montaña, en cambio los odonatos se presentan en zonas lóaticas, estos presentan las puntuaciones más altas. En el otro lado de la tabla tenemos los Oligoquetos y los Quironómidos, cuyas familias son totalmente acuáticas, y oportunistas, con una baja puntuación. Aunque la definición expuesta parezca muy taxativa, existen algunas familias de dípteros que presentan puntuaciones muy altas como Blephariceridae o Athericidae, que están presentes en zonas de aguas prístinas, siendo Dípteros un grupo con puntuaciones en general bajas.

**IASPT:** es una adaptación del ASPT británico. Su cálculo es la división del índice del IBMWP entre el número de taxones que componen ese valor. Ofreciendo una idea de la comunidad dominante en esa región y por ende de las condiciones para que se desarrolle ese tipo de comunidad.

**Riqueza de taxones:** es una variable básica a la hora de describir las comunidades biológicas. Para esta variable se han utilizado únicamente las familias presentes y aquellos taxones cuya determinación no se desarrolla hasta este punto (*Oligochaeta*, *Platyhelminthes*, etc.). En zonas con puntuaciones altas, se presentan aquellas familias tolerantes e intolerantes, es decir taxones que tienen unos requisitos tróficos, hidrológicos y de carácter fisicoquímico más restringido

están presentes. De esta manera buscaremos aquellos patrones más determinantes que promueven una comunidad amplia y más diversa (Mandaville, S.M., 2002; Plafkin et al., 1989).

#### 4.4. Variables independientes

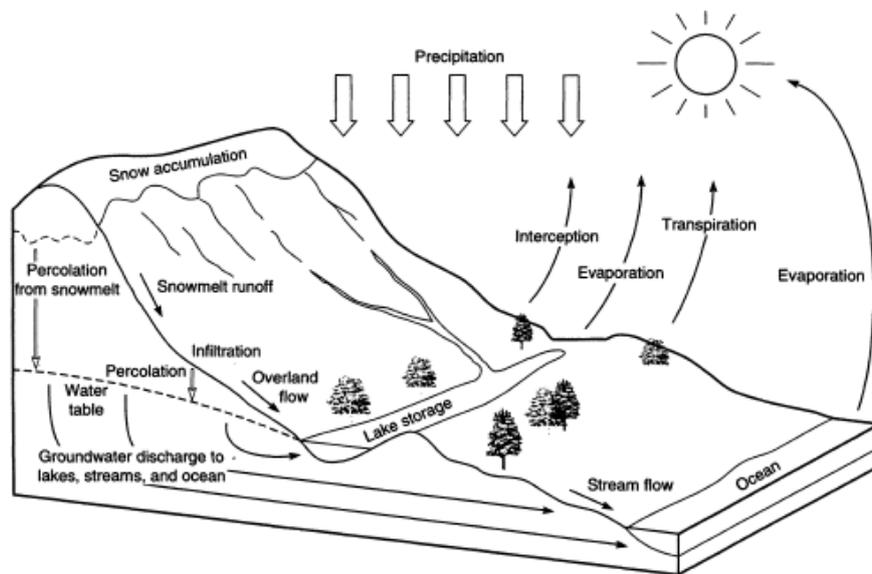
Las variables independientes con las que trabajaremos, procedentes de la plataforma Iber Rio ([IH.rivers](#)), son las condiciones del entorno que afectan a los ecosistemas fluviales (Peñas et al, 2014), introducidas en nuestra red fluvial virtual. La elección de variables predictoras es un paso importante debido a que de ello dependerán los resultados (Mateo et al., 2011). En este caso no se ha hecho una preselección, ya que, aun teniendo bibliografía para guiarse, al ser un modelo para toda España, podríamos eliminar alguna variable que fuese relevante sin saberlo.

Se han introducido 230 variables ([Anexo 2](#)) de 5 categorías: topográficas, climáticas, geológicas, usos de suelo y de impactos antrópicos. De esta manera reflejamos de manera general el contexto influyente en los sistemas fluviales. Aplicando en algunos casos diferentes ópticas, al tratarlos a nivel de cuenca mediante la media del parámetro aguas arriba del tramo (MN), a nivel parcial desde las vertientes exclusivas a ese tramo (LC), aquellas zonas situadas a 200 metros del cauce (BF) y del buffer, de 200 metros, concentrado aguas arriba (BC). A continuación, se describe la tipología de variables predictoras que se han utilizado en el estudio.

**Geología:** se han tratado varios factores, desde su composición (conglomerados, sedimentarias, volcánicas, pizarras...), un total de 11, hasta algunas de sus características como la dureza, la permeabilidad o la conductividad (a partir de 100 clases litológicas de datos originales). Estas se han creado a partir de los mapas litoestratigráficos y de permeabilidad 1:1.000.000 del Instituto de Geología y Minería de España (IGME), entre otros. Las medidas se estiman en tanto por uno. La relación con las aguas corrientes es muy compleja, ya que le ofrece al agua las propiedades de la roca que atraviesa de forma subterránea o cuando se diluye en el cauce (Hynes, 1975; Giller & Malmqvist, 1998, pp 4 - 6).

**Climatología:** se han introducido 4 tipos, entendiéndose como las más importantes en el ciclo hidrológico, a saber: la precipitación, la temperatura, la evapotranspiración y la evapotranspiración potencial. Las primeras poseen un efecto directo sobre los regímenes del caudal y comunidades, en cambio las segundas generan su principal actividad sobre la flora y el suelo, generando así un detrimento en los caudales cuando estas son elevadas. Estas han sido

elaboradas a partir de la información básica utilizada en el desarrollo del modelo SIMPA (Simulación Precipitación Aportación (Estrela and Quintana, 1996), desarrollado y empleado por el Centros de Estudios Hidrográficos del CEDEX. El modelo trabaja con información contenida en celdas de 1 Km<sup>2</sup> y estima la precipitación, la evapotranspiración potencias y temperatura mes a mes durante el periodo 1980 – 2006.



*Figura 4: Dibujo esquemático del ciclo hidrológico. (Dunne and Leopold, 1978).*

**Topográficas:** estas variables son descriptoras físicas de la cuenca y del cauce. Estos factores son muy importantes, debido a su intrínseca relación, entre otros, con la cantidad, velocidad del flujo y la adaptación que han adquirido algunos taxones (Angelier, 2002).

**Usos del suelo:** la actividad humana influye en muchos aspectos en la red fluvial (Álvarez-Cabria et al, 2016, 2017, Giller & Malmqvist, 1998, pp 215 - 244). Es por ello por lo que se han introducido los porcentajes de área ocupada por zonas urbanas, usos agrícolas, zonas de pastos y plantaciones. Además, se han querido representar otros usos de las extensiones ocupadas por bosques caducifolios, los bosques de coníferas, de matorral, zonas desnudas y zonas húmedas. Esta información se ha formado a través del Sistema de Información de Ocupación del Suelo en España (SIOSE), elaborado en el año 2005. Y mediante un software específico de la plataforma NETMAP se calculó el porcentaje de área en área.

**Impactos antropogénicos:** se han introducido 6 tipos de impactos que se muestran como alteraciones en el transcurso normal del río dada la implicación en los cambios de la composición

de las comunidades de macroinvertebrados, los cuales están bien documentadas (WARD & STANDFORD,1983; P. J. BOON, 1988; Craig & Kemper, 1985; Welch, 1992). Se han incorporado:

- **Alteraciones puntuales:** están referidas a vertidos urbanos o industriales al medio acuático de forma puntual. Todos ellos de manera superficial, directos o indirectos, cuya naturaleza es industrial y/o urbana con una población mayor o igual a 2000 habitantes, incluida el agua de escorrentía.
- **Alteración hidromorfológica:** representada por aquellos elementos, de origen antrópico, que producen un efecto de encauzamiento (canalizaciones, coberturas y protecciones), con una longitud mayor o igual a 100 metros.
- **Alteración hidrológica:** Viene determinada por las presas situadas en un embalse y/o con una altura mayor o igual la 10 m.
- **Alteración de la continuidad:** Se han considerado todos los aforos y/o azudes cuya altura es mayor a 0,5 m, por caracterizarse estos como obstáculo transversal al eje longitudinal del río y efecto barrera.

## 5. METODOLOGÍA DE MODELADO

### 5.1. Preparación de datos para el modelado

Una vez que tenemos los valores de la biota por tramos de río, se procede a calcular las variables explicativas de cada comunidad de cada tramo (tabla 1). Sin embargo, no se han podido hacer las mismas métricas en todos los tramos debido a que algunas estaciones solo presentan presencia/ausencia de las comunidades muestreadas. Para aprovechar todos los recursos disponibles y ofrecer mayor estabilidad al modelo, no se han homogeneizado el número de tramos

*Tabla 1 Índices con los correspondientes tramos y requisitos.*

Índice de calidad	Funcionamiento	Estaciones	Nº tramos
Iberian Biological Monitoring Working Party	Presencia	Presencia/Ausencia	2853
Iberian Average Score Per Taxon	Presencia	Presencia/Ausencia	2853
Ephemeroptera, Plecoptera y Trichoptera (Riqueza taxones)	Presencia	Presencia/Ausencia	2312
Riqueza total	Presencia	Presencia/Ausencia	2854
Ephemeroptera, Plecoptera y Trichoptera (Abundancia relativa)	Abundancia/Abundancia Total	Abundancia	2312
Gastropodo, Oligoqueto y Dipteros (Abundancia relativa)	Abundancia/Abundancia Total	Abundancia	2365

Por otro lado, para determinar qué factores son clave a la hora de explicar los patrones espaciales de las comunidades de macroinvertebrados, se tuvieron en cuenta todas las variables independientes explicadas. Sin embargo, no todas fueron incluidas en el modelo. Previamente se hizo una preselección a través del método de correlaciones de Spearman Rank entre pares de variables. De todos los valores obtenidos se escogieron aquellos que tenían una correlación entre índices y variables superior a  $|0,25|$ , con un  $p$ -valor  $< 0,05$ . Además, se tuvo en cuenta la relación que existía entre variables independientes, donde se desestimaron aquellas variables predictoras correlacionadas con un valor superior a  $|0,7|$ , para evitar la multicolinealidad. ([Anexo 3](#)). Aunque se han introducido en el proceso las variantes laterales de las variables, a izquierda o derecha (L\_ o R\_), no se han tenido en cuenta sus valores de correlación en el momento de introducir variables al modelo. Aunque se han utilizado a modo de referencia en casos donde los valores de correlación entre las variables más globales (MN, LC, BF y BC) eran similares.

## 5.2. Random Forest

La técnica de modelado seleccionada fue Random Forest (RF; Breiman, 2001). El RF es un modelo basado en la clasificación y los árboles de regresión. Esta técnica ha sido escogida porque ha demostrado ser muy consistente cuando existen gran cantidad de datos de entrenamiento (Caruana et al., 2008).

RF es una combinación de árboles predictores, donde cada árbol depende de los valores independientes aleatoriamente escogidos, con la misma distribución para todos los árboles en el bosque, para este proceso se han propuesto un total de 500 árboles (Alvarez-Cabria, 2017). Para este trabajo se ha escogido la diferencia de 75 % de los datos para el entrenamiento o calibración, y un 25 % para la evaluación del modelo, dado que es un valor intermedio entre los valores más comunes que se usan para modelar la distribución de especies. 70/30 (e.g. Marini et al. 2010) y 80/20 (e.g. Lomba et al., 2010). Para la generación de cada árbol, se escoge una porción de los datos dependientes (Bootstrap) y los une con un subset de variables para cada división. Como el proceso que se ha desarrollado es de regresión, dado que nuestras variables dependientes (variables biológicas) son continuas, se escoge aleatoriamente 1/3 del total de variables totales introducidas cada vez (Breiman, 2005) y se van generando resultados para posteriormente generar una media de todos ellos. Durante el proceso se ordenan las variables en función de su importancia, para posteriormente predecir un valor de la variable dependiente.

A modo de calibración de los datos de entrenamiento se utilizará el “R<sup>2</sup> training” (0-1), que valora el porcentaje de explicación por parte de las variables. Acompañado del Root Mean Square Error (RMSE), es decir la desviación estándar de los valores residuales, para medir cuanto de dispersos están los valores residuales.

$$RMSE = \left[ \sum_{t=1}^N \frac{(Y_t - \hat{Y}_t)^2}{n} \right]^{\frac{1}{2}}$$

Donde:

- $\Sigma$  = sumatorio
- $(Y_t - \hat{Y}_t)$  = diferencias entre el valor y la recta de regresión.
- N = tamaño de muestras

Como método de premeditación de datos, se ha introducido la función “Crossvalidation” que toma aleatoriamente valores del entrenamiento (10 grupos) y los lanza repetidamente (5 veces), de modo que el resultado que nos da es ya un promedio. El error del modelo (bosque) será proporcional a la fuerza que tengan los árboles generados.

La importancia relativa de las variables independientes viene definida por el índice IncNodePurity index. Este índice informa de la disminución total de las impurezas de los nudos en las divisiones de cada variable, promediada sobre todos los árboles. Siendo los valores más altos para las variables más importantes (Kuhn et al., 2008). La valoración se ha llevado a cabo filtrando aquellas variables con mayor puntuación, y fijando el umbral cuando la tendencia sufría un quiebro, asumiendo la idea de escoger más de una variable.

IncNodePurity index pone de manifiesto la relevancia de las variables que, posteriormente utilizarán para predecir en los puntos que se dejaron de la validación o testing (25% de los datos totales) y cuya valoración se llevará a cabo con el “R<sup>2</sup> testing” y el RMSD (Root Mean Square Deviation). Este último se presenta como la desviación estándar de las diferencias entre los valores predichos y los observados. Mediante la fórmula:

$$RMSD = \left[ \sum_{i=1}^N \frac{(z_{fi} - z_{oi})^2}{n} \right]^{\frac{1}{2}}$$

Donde:

- $\Sigma$  = sumatorio
- $(z_{fi} - z_{oi})$  = diferencias entre el valor y la recta de regresión.
- N = tamaño de muestras

## 6. RESULTADOS

### 6.1. Resultados de correlación

Lo primero que se ha de destacar es la ausencia de variables correlacionadas con la variable EPT-A, ninguna superó el límite de  $|0,25|$ , por lo que no se modeló, al considerar que las relaciones eran muy débiles. Otro aspecto reseñable es que de las 6 alteraciones antropogénicas (hidrológicas, de continuidad y vertidos), ninguna ha presentado una correlación considerable, aunque podemos observar una pequeña tendencia donde los daños aguas arriba afectan a la riqueza de especies negativamente (Riqueza, EPT...), a excepción de aquellas especies que de aguas lénticas (GOLD).

De las 28 variables que se introdujeron al modelo (tabla 3) se pueden clasificar en 4 apartados: topográficas (6), climáticas (8), geológicas (4), y usos del suelo (10). En las topográficas se encuentran el área de cuenca aguas arriba del tramo (AREA\_KM2), el gradiente del canal (GRADIENT), el grado de encajonamiento del cauce (VWI), la pendiente de las partes adyacentes (MnSlope) y su altura media (elevWing), así como el número de confluencias por área de cuenca (CONF\_DEN). Siendo las que mayor correlación tienen las pendientes, tanto del río como del cauce, seguidas por el área de cuenca. Respecto a las variables climáticas, podemos diferenciar 4 tipos, las precipitaciones (MN\_PREC; LC\_PREC), la temperatura (LC\_TEMP) y la evapotranspiración, real (MN\_EI; BF\_EI) y potencial (MN\_ETP; BF\_ETP; BC\_ETP). Y se sitúan de manera más generalizada, por cuencas aguas arriba y subcuenca del tramo, para las dos primeras, y más localizadas para la evapotranspiración real, por situarse en el buffer, y algo menos para la potencial, que se presenta como buffer acumulado. La composición de geológica destaca por el dominio de las variables que representan el porcentaje de rocas sedimentarias (MN\_sdim) y conglomerados (MN\_cong) a nivel de cuenca, susceptibles de la generación de sedimentos y ofrecer con una mayor cantidad de propiedades fisicoquímicas al agua. Y, por otro lado, la dureza de la roca en dos variantes, porcentaje de cuenca (MN\_HARD y subcuenca (LC\_HARD). Finalmente se encuentran los usos del suelo, que son los más numerosos, estos se componen de zonas urbanas (UHD), suelo agrícola (AGR), zonas boscosas caducifolias (BLF) y de matorral bajo, brezales, paramos, etc. (SSH) además de las zonas desnudas (DEN). Son también los más heterogéneos, porque se tratan desde diferentes ópticas, una más general desde la cuenca, donde están representados todos menos los bosques (MN\_UHD; MN\_AGR; MN\_SSH; MN\_DEN), otras más minuciosas como es el buffer de 200 metros del tramo, (BF\_BLF; BF\_AGR).

Y pasos intermedios, como las vertientes al tramo (LC\_BLF; LC\_SSH) y el buffer acumulado (BC\_BLF; BC\_UHD).

**Tabla 2: Variables independientes que se han introducido al modelo para cada índice de comunidad.**

<b>Variables</b>	<b>Riqueza</b>	<b>EPT-R</b>	<b>IBMWP</b>	<b>IASPT</b>	<b>GOLD</b>	<b>EPT -A</b>
AREA_KM2		x	x	x	x	
GRADIENT	x	x	x	x	x	
VWI		x	x		x	
MnSlope	x	x	x	x	x	
elevWing		x		x		
CONF_DEN		x		x		
MN_PREC		x		x	x	
LC_PREC	x		x			
LC_TEMP		x		x		
MN_EI				x		
BF_EI		x		x		
MN_ETP			x			
BF_ETP				x		
BC_ETP		x				
MN_cong	x	x	x	x		
MN_sdim	x	x	x	x		
MN_HARD		x	x	x		
LC_HARD	x	x	x	x	x	
MN_UHD		x	x	x	x	
BC_UHD		x	x	x	x	
MN_AGR	x	x	x	x	x	
BF_AGR		x	x	x		
LC_BLF		x	x	x		
BF_BLF			x			
BC_BLF		x		x		
MN_SSH	x	x	x	x	x	
LC_SSH				x	x	
MN_DEN				x		

## 6.2. Resultados del modelado

### 6.2.1. Resultados numéricos

Respecto al  $R^2$  del training, los valores son bajos respecto a la Riqueza y al GOLD, con 0,246 y 0,3 respectivamente. Siendo bastante mejores para el IBMWP (0,44), y bastante aceptables para el EPT (0,578) y IASPT (0,68).

**Tabla 3: Resultados obtenidos de la modelización de especies con Random Forest.**

Variable	Riqueza	EPT	IBMWP	IASPT	GOLD
Tamaño muestras	2142	1735	2141	2141	1775
training	0,246	0,578	0,44	0,68	0,3
RMSE	11.86	3.82	61.20	0.50	16.22
Porcentaje desvió del error	16.70	14.69	15.45	9.16	16.36
R2 testing	0.29	0.62	0.47	0.67	0.29
RMSD	11.65	3.70	61.06	0.52	15.53
Porcentaje desvió de los datos	16.41	14.24	15.42	9.52	15.66

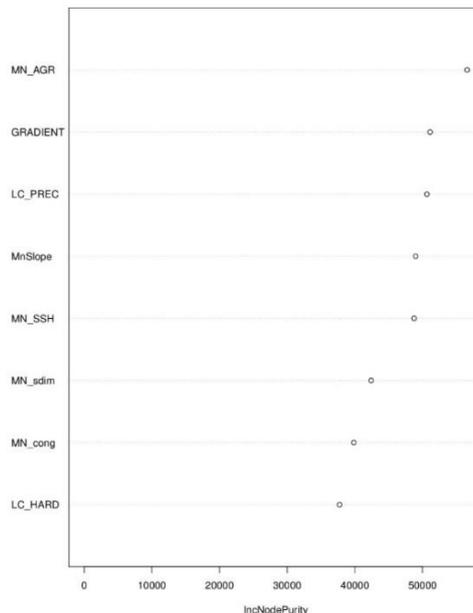
En referencia al RMSE, cuanto más pequeño sea mejor, pero siempre tiene que ir referenciado a la escala de datos de entrada ([Anexo 4](#)), y vemos que la riqueza es un tercio de la mediana. En el EPT, tenemos un valor de 3,82, que supone un 14% de desviación. El índice IBMWP tiene un valor de 61,20, que en un principio puede parecer alto pero que al referenciarlo en la escala se puede considerar una desviación del 16%. En los otros dos índices, IASPT y GOLD, se presentan con un error bastante bajo, ya que el primero se ha de referenciar respecto a un rango 5.46, y el GOLD, este con mayor error, de un 16 %.

Los resultados de la predicción se parecen bastante a los  $R^2$  del training. Siendo los valores más bajos para R y para el GOLD (0,29 para ambos). El IBMWP obtiene un valor de  $R^2$  predictivo de 0,47. Para el índice IASPT el  $R^2$  predictivo es de 0,68 y para el EPT es 0,62.

Los valores de RMSD nos indican como es la desviación entre los datos observados y los datos predichos, donde se han obtenido 11,65 para la Riqueza, que dentro de la escala presenta un 16% de desviación, para el EPT, con un 3,7, tiene una desviación del 14,24, para el IBMWP y el GOLD, 61,06 y 15,53, tienen un 15,5 % de desviación y finalmente para el IASPT, un 9,5% con un valor del RMSD de 0.52.

### 6.2.2. Resultados de los valores de importancia con el IncNodePurity index y mapas de predicción

Para la variable riqueza se obtuvieron los siguientes valores del IncNodePurity index (Ilustración 7, [Anexo 5](#)). Donde las máximas puntuaciones se obtuvieron en el porcentaje medio de las cuencas sobre el uso agrario (MN\_AGR), seguido por la pendiente del cauce (GRADIENT), la precipitación local (LC\_PREC), la pendiente media de la cuenca (MnSlope) y el porcentaje de suelo de matorral (MN\_SSH). En este primer ejemplo la tendencia cambiaba con la primera variable, pero al no ser tan clara como para dejar una sola variable se escogieron las otras 4. Desestimando la parte geológica por menor implicación con respecto a las otras.

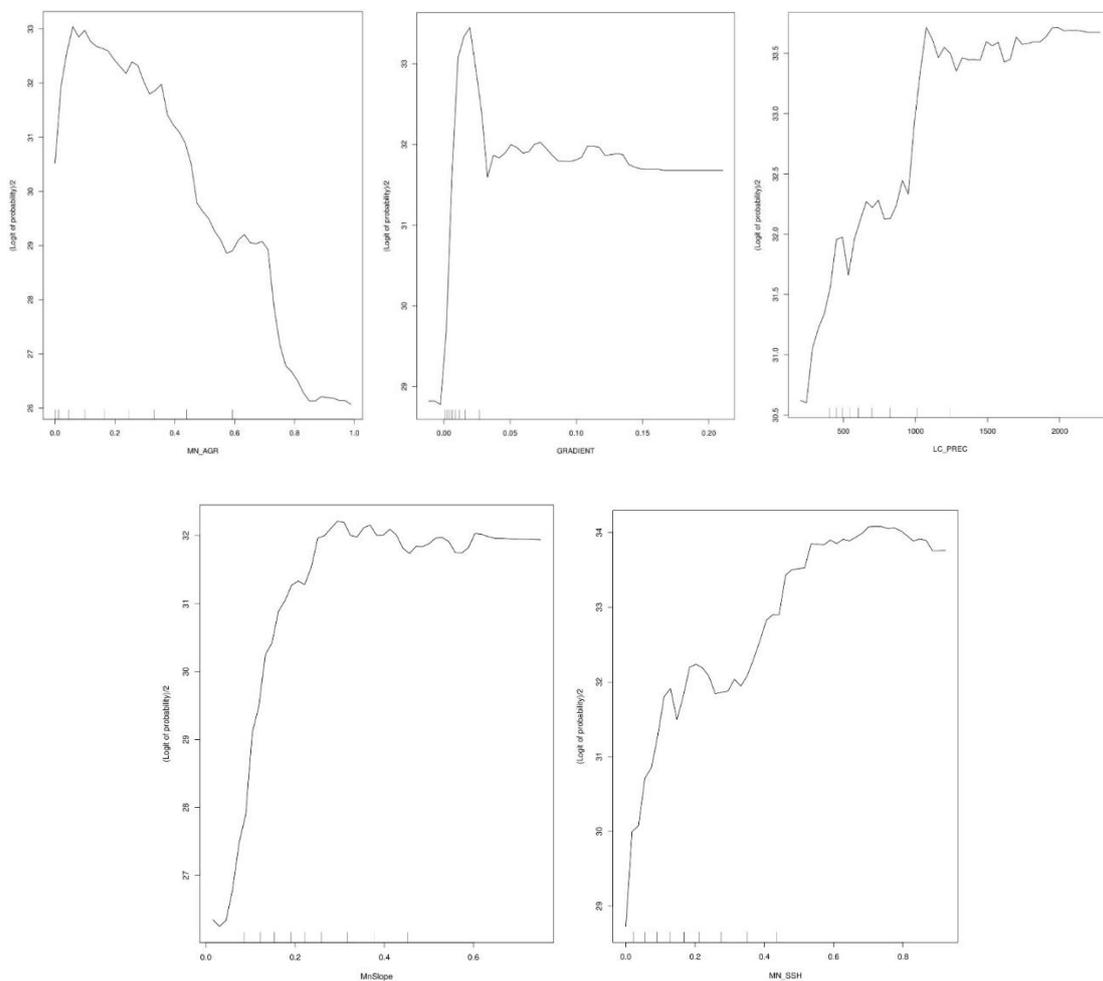


**Figura 5: Índice IncNodePurity de importancia de las variables para el índice de riqueza.**

A continuación, se va a mostrar los patrones más importantes que influyen sobre la variable riqueza de especies. La variable más importante según el IncNodePurity index es la agricultura, que aumenta la riqueza de taxones con una presencia menor al 10 % en la cuenca, una vez superado ese porcentaje, la riqueza disminuye, estabilizándose cerca del 60% y volviendo a descender a partir del 70% hasta valores mínimos.

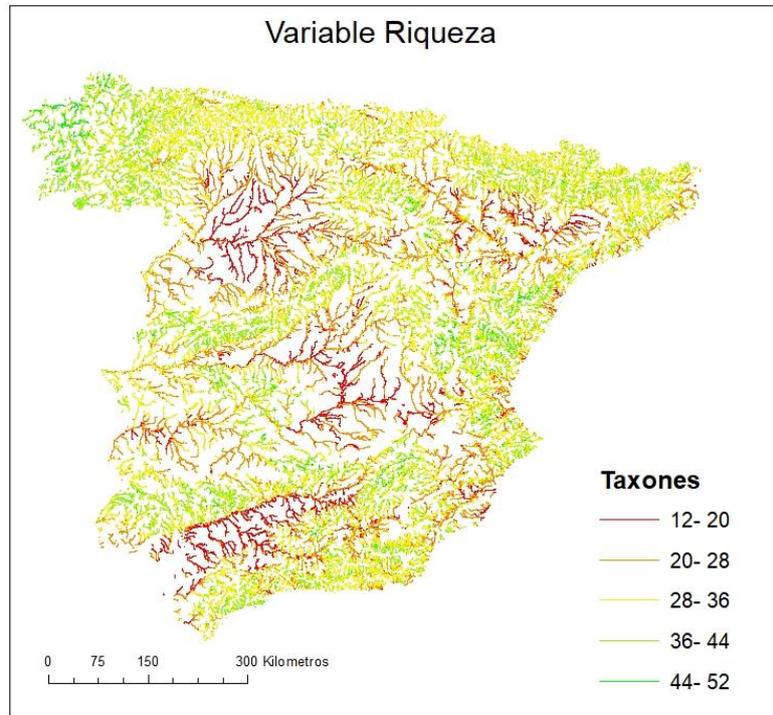
Respecto al gradiente del cauce, vemos que hay una relación directa al inicio, alcanzando el máximo hacia el 2,5% de pendiente, pero después disminuye estabilizándose, entorno al 4% de

pendiente. Las precipitaciones localizadas en las subcuencas juegan un papel importante aumentando el número de taxones a medida que aumentan las lluvias. El número máximo de diversidad se alcanza entorno a los 1100 – 1200 milímetros (mm) anuales, para después mantenerse en esos niveles de riqueza, aunque la precipitación se duplique. La pendiente media de la cuenca también ayuda a la aparición de especies, siendo un crecimiento exponencial hasta su máximo en 30 % y mantenerse a pesar del aumento de la inclinación. Finalmente, podemos ver que la presencia de un cierto arbolado como son los brezales, arbustos, matorrales afecta de manera positiva con un aumento casi vertical al inicio, para después estabilizarse, 0,13 – 0,4 y seguir creciendo hasta el 60% de inclinación.



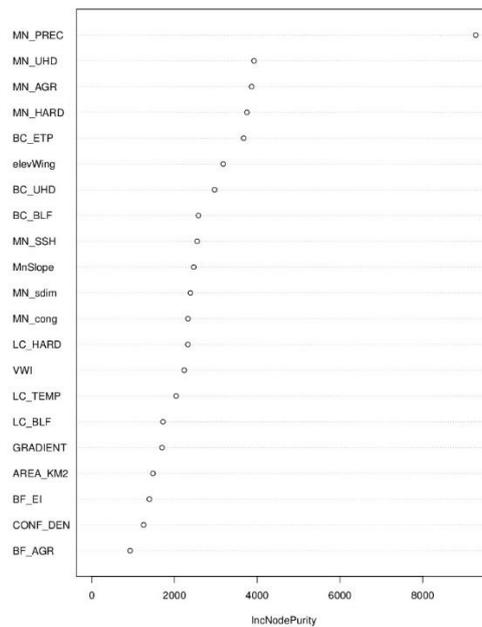
**Figura 6: Variables patrón introducidas al modelo a través del índice Riqueza de taxones. De izquierda a derecha y de arriba a abajo. Porcentaje media de agricultura en la cuenca (MN\_AGR), Gradiente medio del tramo (GRADIENT), precipitación media en la subcuenca (LC\_PREC), pendiente media de la subcuenca (MnSlope) y porcentaje medio de zonas arbustivas en la cuenca (MN\_SSH).**

El mapa de la Riqueza de taxones, como los mayores valores de riqueza se sitúan de manera generalizada en las partes altas de las montañas, destacando las cuencas del Miño – Sil y las Costa de Galicia. Otra observación es la baja concentración de taxones en los tramos bajos de los ríos de la meseta.



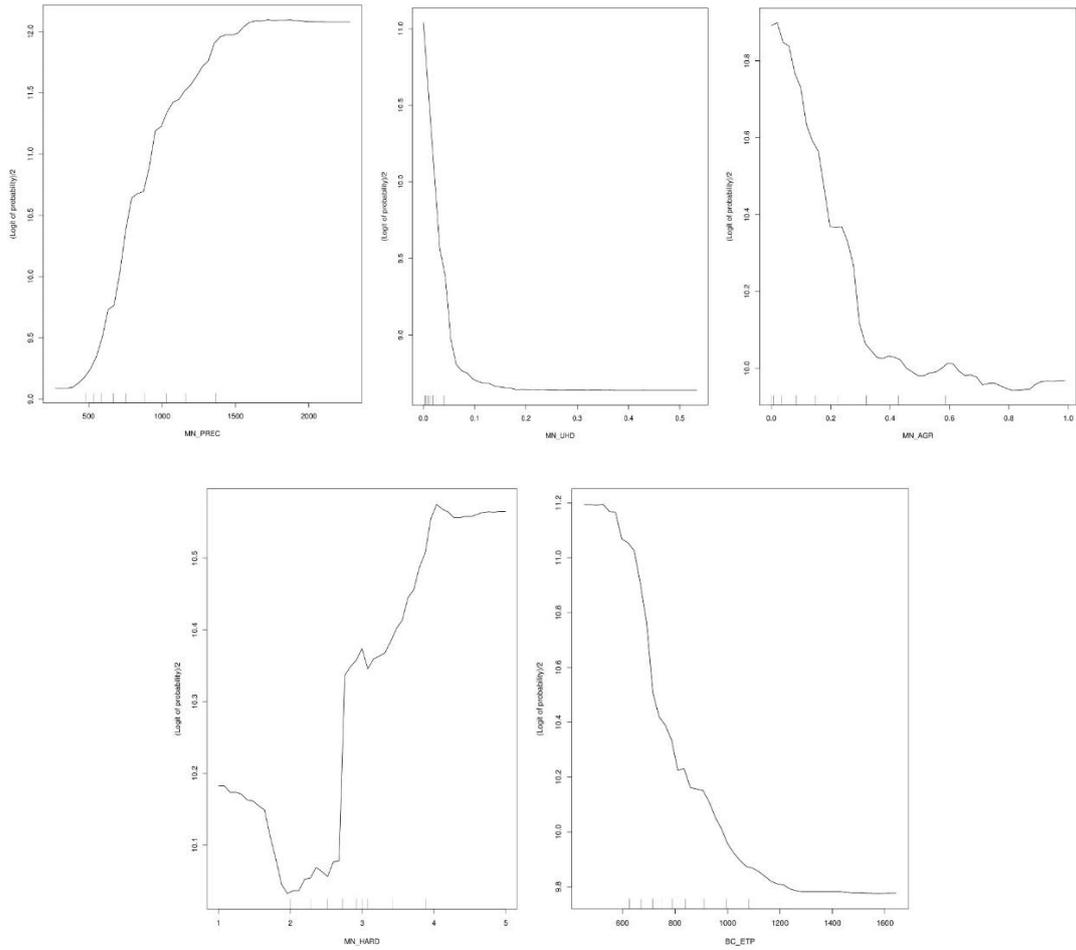
*Figura 7: Mapa de riqueza de taxones para la Península Ibérica.*

En el caso la riqueza de taxones EPT, las precipitaciones medias para toda la cuenca superior al tramo (MN\_PREC), suponen una diferencia más que apreciable, sin embargo, se ha querido completar el análisis, introduciendo las 4 siguientes variables (zonas urbanas, agricultura, dureza de la roca y evapotranspiración potencial respectivamente). Pues se entiende que estos ordenes no solamente se ven influenciados por la precipitación, aunque de los patrones sea la más predominante con más de 4000 puntos sobre la siguiente (figura 8).



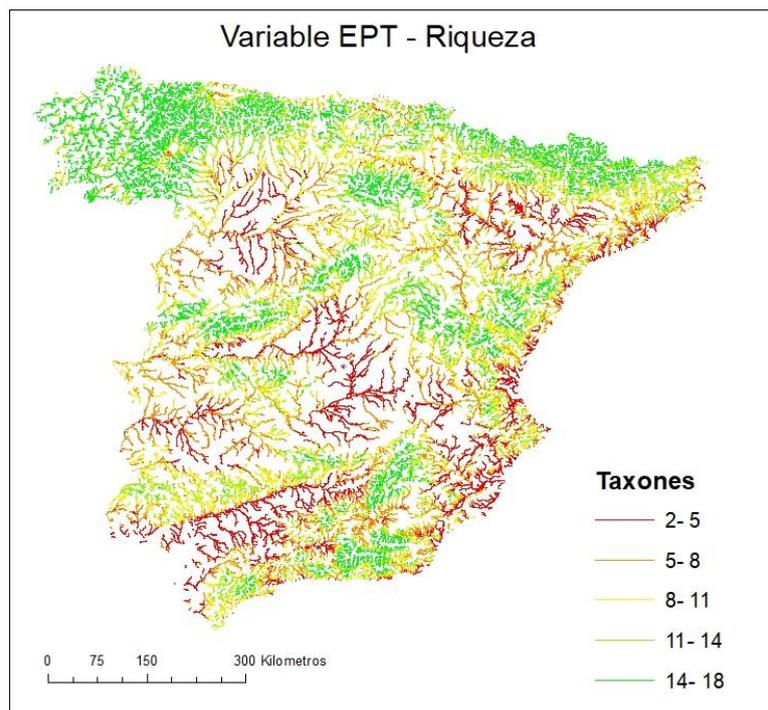
**Figura 8: Índice IncNodePurity de importancia de las variables para el índice de EPT.**

En las siguientes gráficas, observaremos cómo afectan las variables destacadas al aumento de la variable EPT-R. Comenzado por la más importante, como son las precipitaciones anuales; se vislumbra una relación exponencial entre el aumento de precipitación y el aumento de plecópteros, efemerópteros y tricópteros. Llegando al punto de inflexión cerca de los 1000 mm anuales, y al máximo cerca de los 1500 metros. Sin embargo, en la gráfica de la ocupación urbana del suelo, vemos que se produce un descenso casi vertical de la variable EPT-R ante porcentajes muy bajos de suelos destinados a este uso. Otro de los patrones espaciales que afectan de manera negativa a la distribución de esta comunidad es el porcentaje medio de uso agrícola en la cuenca, produciendo una caída brusca hasta el 30 % de ocupación agrícola en la cuenca y después descender lentamente desde valores bajos. Respecto a la dureza de la roca presente en la cuenca, el EPT-R requiere de unos valores de dureza bajos o tendiendo a altos (hay que recordar que la variable HARD esta ponderada). De esta manera, una dureza baja no permite el desarrollo de estas comunidades al máximo, encontrándose los registros más bajos en una resistencia media. A partir de este punto el aumento es prácticamente vertical hasta su punto máximo con una dureza de 4, en la cual la riqueza de EPT se estabiliza. Como ultimo patrón espacial destacable, se encuentra la evapotranspiración potencial del buffer acumulado; donde un aumento de dicha variable produce un efecto negativo sobre la riqueza taxonómica del EPT.



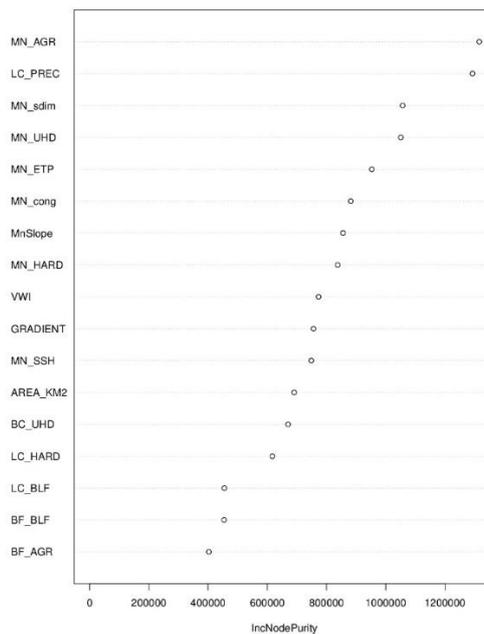
**Figura 9: Variables patrón introducidas al modelo a través del índice riqueza de taxones de la comunidad EPT. De izquierda a derecha y de arriba abajo. Precipitación media de la cuenca (MN\_PREC), porcentaje medio de la cuenca con uso urbano (MN\_UHD), porcentaje medio de la cuenca con uso agrícola (MN\_AGR), dureza media de la roca en la cuenca (MN\_HARD) y evapotranspiración potencial media en el buffer de 200 metros acumulado aguas arriba (BC\_ETP).**

El siguiente mapa es para la comunidad EPT, la cual se muestra más abundante en zonas de montaña y noroeste de España; con un claro gradiente decreciente de norte a sur, siendo las zonas de depresión de los valles y la costa los más afectados.



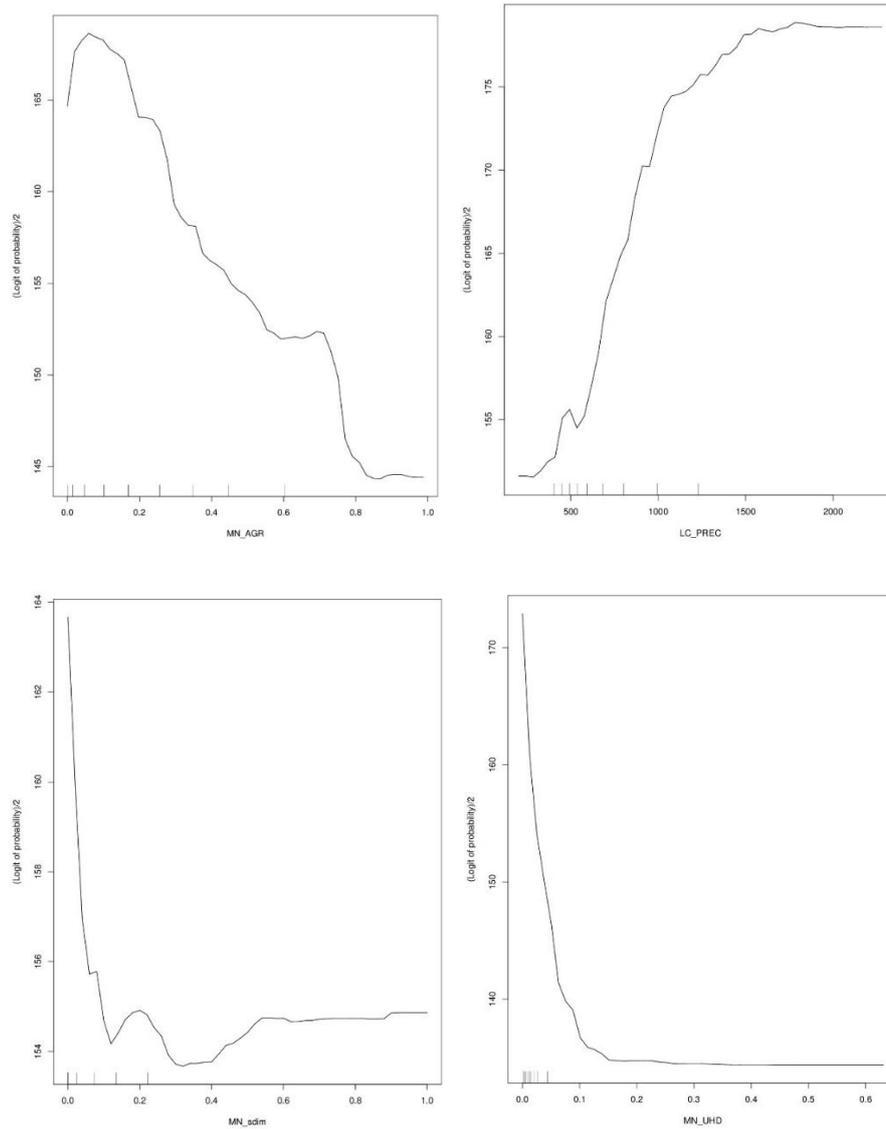
**Figura 10:** Mapa de riqueza de abundancia de taxones de efemerópteros, plecópteros y tricópteros para la Península Ibérica.

Para la variable IBMWP, se han obtenido 4 variables como las más importantes (Ilustración 13) siendo el uso del suelo agrícola (MN\_AGR) y la precipitación (LC\_PREC) como las variables que ejercen mayor influencia en la distribución de valores altos del IBMWP. Además, se han incorporado el porcentaje de rocas sedimentarias en la cuenca (MN\_sdim) y el uso de suelo urbano en la cuenca (MN\_UHD).



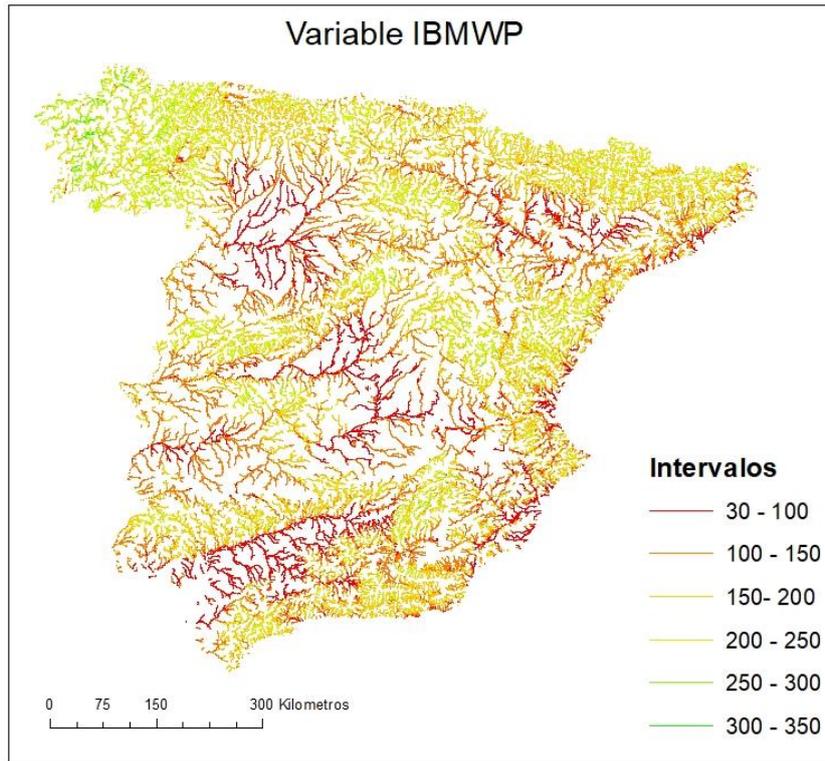
**Figura 11 Índice IncNodePurity de importancia de las variables para el índice IBMWP.**

Siguiendo el orden de importancia que nos ha dado el modelo, la ocupación agraria en la cuenca nos indica que un pequeño porcentaje de suelo ocupado por este sector es favorable y produce un crecimiento del número de taxones exponencial, pero seguidamente llega al pico máximo y genera una caída repentina, donde al llegar al 40 % de la cuenca produce un descenso a la mitad de la presencia, estabilizándose hasta el 70%, para generar un quiebro descendente en la gráfica. La siguiente variable nos explica que el aumento de la lluvia en la subcuenca provoca un aumento sobre la variable de forma exponencial, con un punto máximo en 1500 mm anuales. El tercer grafico hace referencia a los sedimentos en la cuenca, donde una presencia muy baja, alrededor del 5%, provoca una caída de taxones vertical; cerca del 50 % produce un pequeño aumento, en el que se estabiliza a pesar del aumento de roca sedimentaria en la cuenca. Respecto a las porciones de terreno dedicado a uso urbano produce una disminución de especies, alcanzando valores mínimos ante un 15 % de ocupación de la cuenca.



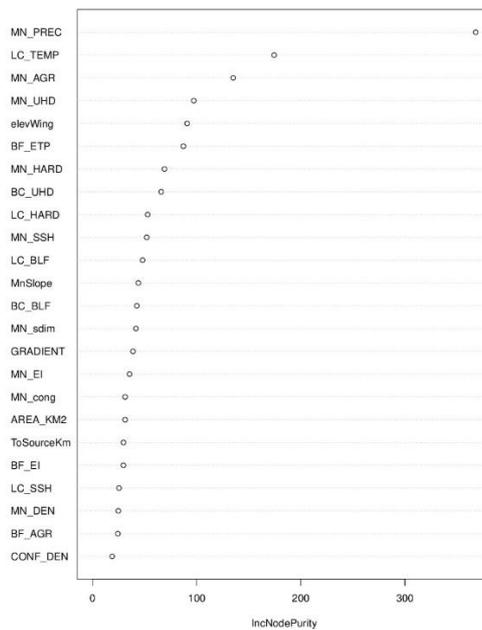
**Figura 12: Variables patrón introducidas al modelo a través del índice IBMWP. De izquierda a derecha y de arriba a abajo. Porcentaje medio de la cuenca dedicada al uso agrícola (MN\_AGR), precipitación anual media de la subcuenca (LC\_PREC), porcentaje medio roca sedimentaria de la cuenca (MN\_sdim) y porcentaje de cuenca destinado al uso urbano (MN\_UHD).**

Para las comunidades representadas por el índice IBMWP, tenemos con tendencias más claras y unos gradientes entre las partes altas y los valles aluviales. Además, se ve una decoloración entre un norte donde predomina el verde y el centro-sur, con zonas más amarillas y rojas.



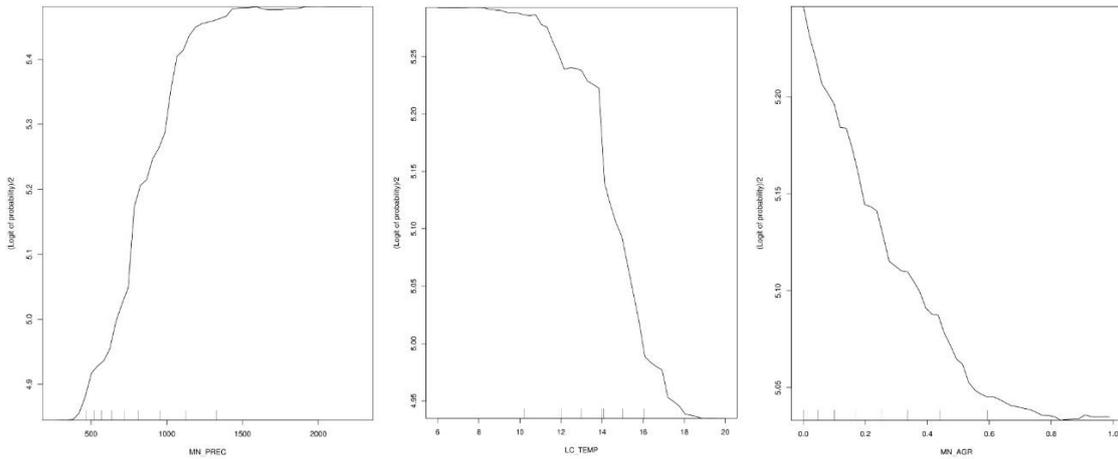
**Figura 13: Mapa de distribución de la variable IBMWP para la Península Ibérica.**

Para la variable IASPT la selección ha sido más fácil, puesto las tres variables más significativas son la precipitación en la cuenca (MN\_PREC), la temperatura (LC\_TEMP) y el uso agrícola de las tierras de la cuenca (MN\_AGR). Siendo de los índices que más variables tiene (24), se ha desmarcado de 21 de ellas con un rango de valores bastante homogéneo, entre 97,11 y 18,11.



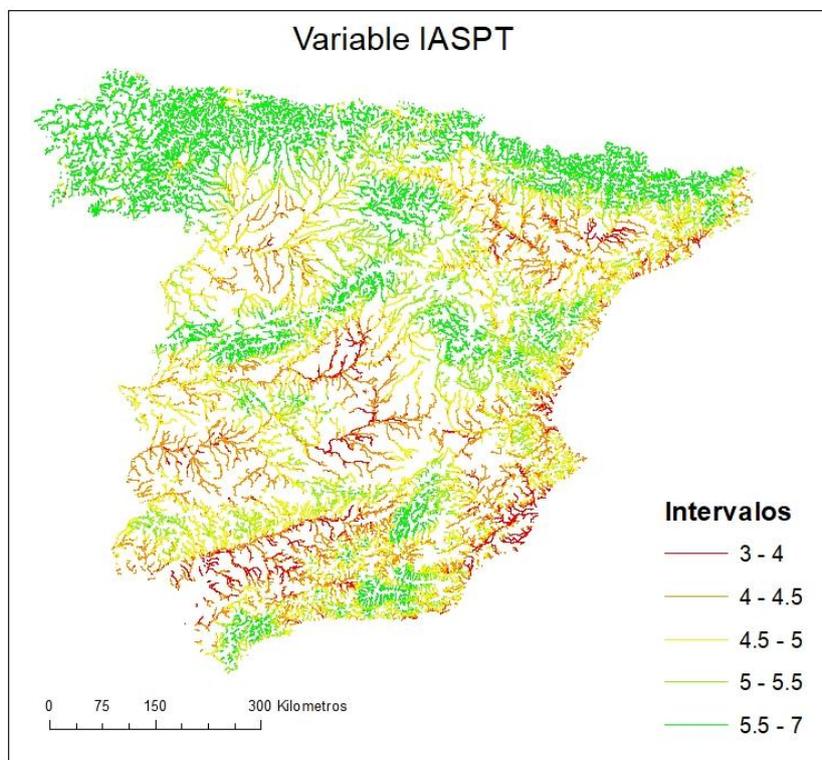
**Figura 14: Índice IncNodePurity de importancia de las variables para el índice IASPT.**

El índice IASPT se nutre del índice IBMWP, por lo que no es raro que con la precipitación anual de la cuenca ocurra lo mismo que en el anterior indicador. En este caso el crecimiento es más acusado que antes, puesto que el valor máximo se encuentra hacia los 1100 milímetros anuales. Como segundo componente tenemos la temperatura en la subcuenca; es la única variable que presenta un parámetro tan importante para la vida acuática. En este punto vemos que el aumento de la temperatura no es propicia para los taxones de mayor carga puntual. Desciende mucho entre los 10 y 12 grados centígrados, haciendo una caída suave hasta los 14 °C, y producir una caída brusca que no se detiene hasta antes de los 19 grados, donde llega a su mínimo. La agricultura media de la cuenca genera desde el primer momento un impacto negativo, disminuyendo gradualmente la presencia de taxones susceptibles. Aunque si nos fijamos en el eje de ordenadas y comparamos con los anteriores su efecto es latente pero menor.



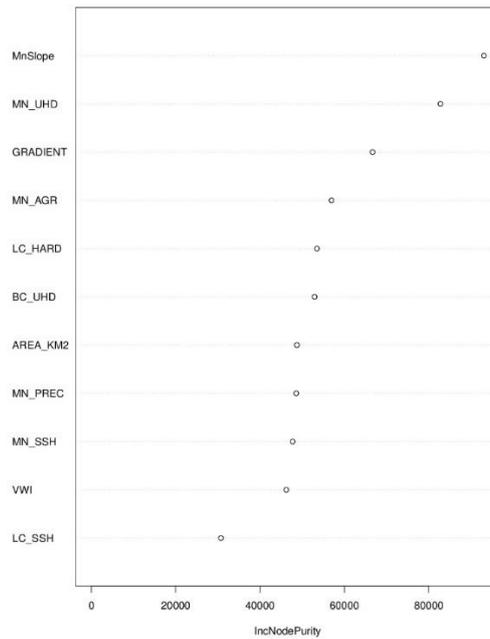
**Figura 15: Variables patrón introducidas al modelo a través de las comunidades del IASPT. De izquierda a derecha. Precipitación media anual de la cuenca (MN\_PREC), temperatura media anual en la subcuenca (LC\_TEMP) y el porcentaje de uso agrícola en la cuenca aguas arriba del tramo (MN\_AGR).**

En la figura 16, observamos un gradiente poblacional acusado; en el tercio norte se encuentran unas poblaciones diferentes a las situadas en el sur de la península. Al sur de la Cordillera Cantábrica, podemos definir otro gradiente bien definido entre las partes de la montaña y las partes bajas de los valles; sin embargo, en la vertiente norte de dicha cordillera no se encuentra tan bien diferenciada.



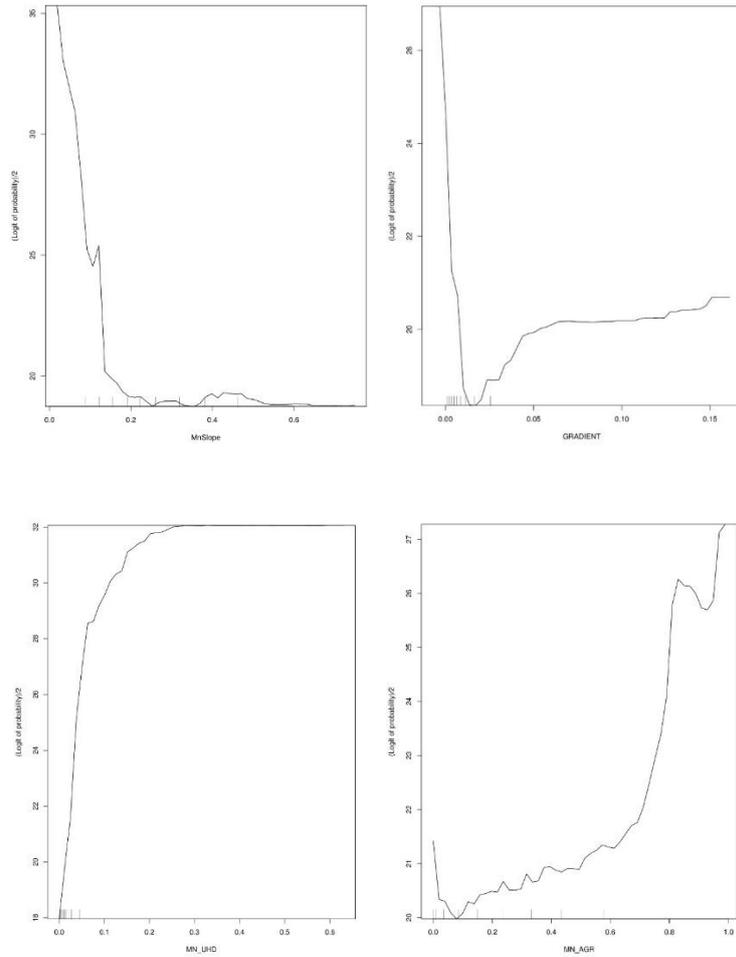
**Figura 16: Mapa de abundancia de las comunidades IASPT.**

Los patrones para la variable GOLD se encuentran definidos por la pendiente media de las laderas (MnSlope); la presencia de zonas urbanas (MN\_UHD); el gradiente del tramo fluvial (GRADIENT) y el porcentaje promediado de ocupación agrícola de la cuenca (MN\_AGR). Este último valor se ha seleccionado debido a la magnitud del eje de abscisas, la cual es muy grande, y las diferencias entre puntos se hacen mayores a pesar de la apariencia de proximidad (figura 17).



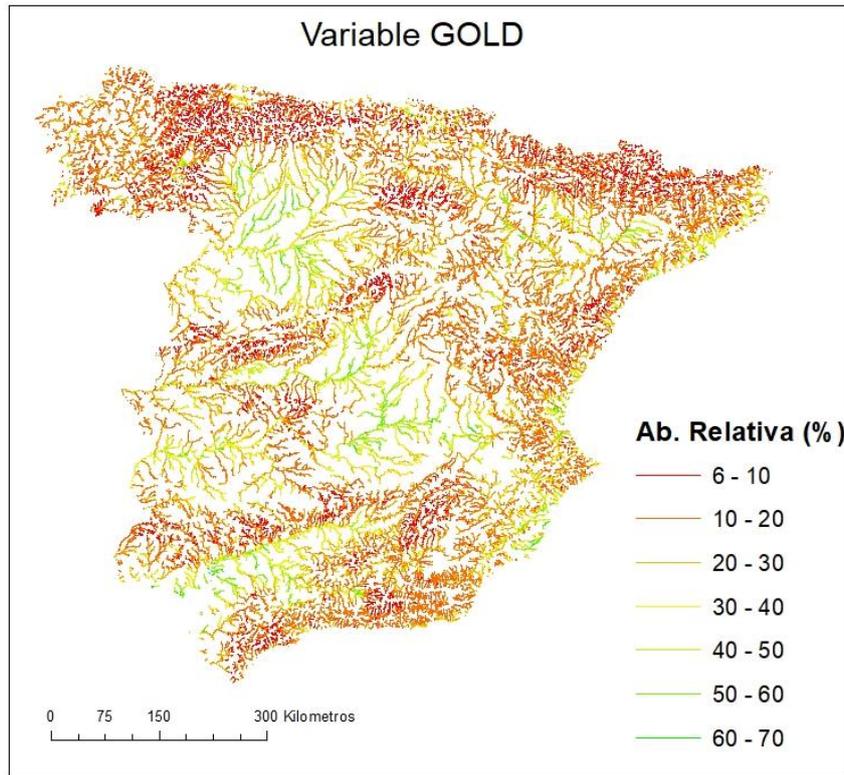
**Figura 17: Índice IncNodePurity de importancia de las variables para el índice GOLD.**

Esta variable es disonante del resto, puesto que indica aquellos taxones más eurioicos, conformado en su mayoría por no-insectos (excepto dípteros), por lo que su relación con el agua es total en sus fases vitales. El primer patrón se refleja en la pendiente media del valle, donde el aumento de esta variable topográfica hace disminuir la abundancia relativa de los GOLD en la población total. Para la pendiente del tramo fluvial encontramos los mismos patrones, donde un aumento de la inclinación provoca un descenso poblacional en la variable GOLD. El tercer factor en importancia en la distribución espacial es la ocupación del uso urbano, la cual produce un aumento vertical dentro de la población GOLD. Con respecto al uso de suelo agrícola se mantiene una relación directa positiva, con una inclinación moderada al inicio, hasta llegar al 60 % de la cuenca donde el crecimiento es exponencial.



**Figura 18: Variables patrón introducidas al modelo a través de las comunidades del GOLD. De izquierda a derecha y de arriba a abajo. Pendiente media de la cuenca (MnSlope), gradiente medio de inclinación del tramo (GRADIENT), porcentaje medio de uso urbano en la cuenca (MN\_UHD) y porcentaje medio de uso agrícola en la cuenca (MN\_AGR).**

El mapa de predicción para la comunidad Gold se distribuye de manera contraria al resto de mapas que hemos visto; la zona de montaña posee una población muy baja de estas, en cambio las zonas llanas ofrecen lugares con mejores condiciones para encontrarnos con estos taxones. Las zonas de costa también son prolíficas para desarrollarse, al menos en el levante español a diferencia del norte y noroeste de la Península.



*Figura 19: Mapa de abundancia relativa para la variable GOLD en la Península Ibérica.*

## 7. DISCUSIÓN

En el presente estudio hemos encontrado un claro patrón de variación espacial en las comunidades de macroinvertebrados bentónicos fluviales a lo largo de la Península Ibérica. Las comunidades de macroinvertebrados de la Península Ibérica se organizan atendiendo principalmente a características climáticas como la temperatura y la lluvia, al tamaño de las cuencas, y a la cantidad de uso de suelo agrícola aguas arriba de los puntos de muestreo.

### 7.1. Selección de datos validos e índices

La selección y búsqueda de anomalías en los datos es una fase crítica en cualquier modelado pues limita tu capacidad de predicción del modelo. En este proyecto era necesario, dada la variabilidad de los métodos y el periodo de realización de los muestreos. Esta heterogeneidad de datos ha influido en muchas de las decisiones respecto al análisis y tratamiento de las comunidades; detenerse en el nivel taxonómico de Familia, permitió utilizar la mayoría de los datos muestreados, al eliminar los posibles errores cometidos durante la identificación de los macroinvertebrados o durante los muestreos. Además, permitía la perspectiva de comunidad con una cierta exactitud, válida para mostrar características comunes a un nivel de discretización alto.

La selección de aquellas familias con una escasa representación fue por doble motivo, el primero que no confundiese a las métricas de las variables respuesta dando un peso a un taxón con muy poca relevancia poblacional; y la segunda razón estaba en el ruido que podía introducir al modelo por ser ambiciosos y no eliminar comunidades de bajísima presencia.

La selección de las variables de los macroinvertebrados bentónicos fue fruto de la observación de los datos refinados que obtuvimos. Se eligieron unos índices, que en principio son de calidad como el IBMWP y el IASPT, y otros algo más enfocados en comunidades como el EPT y el GOLD, para la representación de los macroinvertebrados. Lejos de presentar un interés por la métrica en sí, preferíamos observar la descripción de las comunidades que nos ofrecían; dado que estos índices pueden verse afectados por otras características que no sean originadas por la propia contaminación. Otra ventaja que nos conferían estas variables es la eliminación de anomalías que pudiesen seguir presentes en los datos tras la discretización, ya que el IBMWP y el IASPT por ejemplo, se rigen por una composición y estructura fija marcada por el protocolo de muestreo/valoración, aunque presentan un cierto sesgo hacia las especies estenoicas y pueden

dificultar en cierto modo la comprensión. Sin embargo, las otras tres variables pueden ayudar a discretizar mejor y valorar los patrones de manera más sencilla; la riqueza de taxones se centra en la heterogeneidad y estabilidad de los hábitats a través de la cantidad, más que en las propias comunidades, permitiendo conocer aquellos patrones más propicios para el desarrollo de la diversidad. Las variables GOLD y EPT-R describen unas clases u ordenes de comunidades concretas, que además de ser antagónicas en su comportamiento, ofrecen una perspectiva más detallada de los patrones por comportamiento de las familias. Como vemos, el uso de estas métricas puede ofrecer en su conjunto una perspectiva más completa de las comunidades bentónicas.

De otro modo, con una visión crítica al respecto, una opción para tener en cuenta es la repetición del modelado introduciendo otra variable respuesta elaborada con un NMDS (Non Metrical Multidimensional Scaling), he introducir los ejes como variable dependiente para una mejor comprensión de funcionamiento y composición más detallada

## 7.2. Selección de variables

Las variables independientes que se presentaron en el apartado 3.4, ofrecen una imagen completa de las cuencas desde una escala tan amplia como es la Península Ibérica. No se ha realizado una preselección por el hecho de no perder información que posiblemente fuera relevante para las diferentes variables respuesta. De manera que dicha selección se realizó de manera que fueran las propias correlaciones entre las variables de macroinvertebrados y las variables independientes las que mostraran su vínculo.

El valor de la correlación se estimó que fuera superior a  $|0,25|$ . El valor puede entenderse pequeño si no se tiene en cuenta la escala de trabajo, donde se mezclan muchos tipos de cuencas, y la heterogeneidad presente en los datos de muestreo. Observando el [anexo 3](#) con detalle, apreciamos que muy pocas las variables superaban el  $|0,4|$  en una correlación, de modo que para no dejar fuera variables importantes para la bibliografía (e.g.: Allan, 1995; Giller & Malmqvist, 1998; Angelier 2002), se propuso dicho valor. Aunque en la relación entre las variables independientes elegidas se propuso eliminar aquellas que fuesen fuertes ( $>|0,7|$ ) (Álvarez-Cabria, 2017). Es llamativo que el índice EPT con la abundancia relativa no tuvieran ninguna correlación superior al nivel propuesto, debido a esta razón se eliminó.

Sobre las propias variables cabe destacar que la precipitación está presente en todos los índices, unas veces a nivel de cuenca y otras de manera local, con una relación positiva excepto para el

índice GOLD, que como ya comentamos, son taxones de aguas lénticas y detritívoros, donde un aumento de precipitación puede conllevar un lavado de las partículas de materia orgánica. Las lluvias para las variables EPT y IASPT superan el 0,55, por lo que se estima que las puntuaciones más altas de ambos sean un reflejo de las comunidades lóxicas, principalmente formadas por taxones de insectos, que precisan de fuentes de agua constantes para mantener los caudales pequeños o de alta montaña con flujo. En cambio, en aquellos lugares donde llueve poco, las lluvias suelen ser torrenciales (e.g.: Mediterráneo costero), provocando una deriva muy fuerte con un gran arrastre de sedimentos, perjudicando a este tipo de comunidades. Según el IASPT e IBMWP, en las zonas costeras del mediterráneo se encuentran taxones no insectos; cuyo hábitat preferencial integra aguas más tranquilas, con sedimentos y una vegetación acuática que ralenticen las aguas y controle el ciclo del intercambio gaseoso.

Siendo a la inversa con la temperatura, donde se muestra como un factor negativo para todos, menos para el GOLD que no le afecta de manera tan directa. Como se ha indicado anteriormente los valores altos de las variables EPT-R, IBMWP y IASPT corresponden con familias que precisan de zonas de alta montaña donde las aguas tienden a estar más frías. Debemos tener en cuenta que la temperatura del agua está, por lo general, entorno a un grado de diferencia de la temperatura media anual del aire, por lo tanto, aquellos lugares dependientes del agua subterránea (manantiales y ríos de cabecera) variaran poco durante el año, siendo estas más frías (Ward & Dufford, 1979). En cambio, los tramos bajos del río son más anchos; no poseen una sombra suficiente; tienen una cuenca receptora mucho mayor, por lo que las aguas estarán más calientes, limitando la cantidad de oxígeno disuelto y cambiando la estrategia de adaptación a este medio.

Las variables geológicas se centran en la dureza y en el porcentaje de arenas y gravas presentes en la cuenca; donde los índices EPT, IBMWP y IASPT poseen una relación positiva con la primera y negativa con las segundas; de modo contrario a la comunidad bentónica GOLD. Valores de la variable HARD altos implican rocas duras, y valores bajos indican gravas y arenas. Por otro lado, atendiendo a los criterios de la Balanza de Lane, podemos concretar localizaciones en el perfil longitudinal del río, dado que en las partes altas de montaña el proceso de sedimentación es muy débil, siendo este mucho mayor en las zonas bajas, donde la corriente se ralentiza y las partículas caen al fondo. Concretando: para los efemerópteros, tricópteros y plecópteros precisan de aguas más vivas y con menor carga de sedimentos (de manera general, puesto que algunos taxones de plecópteros se encuentran en lagos) y los gasterópodos, oligoquetos y dípteros, presentan una apreciación por zonas de deposición de sedimentos, con aguas más tranquilas.

Los usos del suelo que presentan una relación con las variables se pueden clasificar en dos partes, unas más “humanas”, como es la urbana y la agrícola y, por otro lado, más natural o ecológica, aunque también estará afectada por el ser humano, como lo son los bosques de ribera, los matorrales y el suelo desnudo. El primer grupo constituye una alteración significativa del medio natural (estructura suelo, composición, etc.). Aquí no se han introducido parámetros más complejos que el porcentaje de cuenca ocupada, sin embargo, se perciben las afecciones a las comunidades de macroinvertebrados bentónicos la presencia de estos usos. En otros estudios llevados a cabo por Álvarez-Cabria et al. (2016; 2017), la agricultura y las zonas urbanas ejercían una presión sobre el medio desfavorable al aumentar los niveles de nitratos del agua y los fosfatos. Esto implica que aquellos individuos que precisen de una alta concentración de oxígeno se verán afectados (EPT, IBMWP, IASPT), tanto por reacciones químicas como por el aumento de fitobentos (eutrofización). Al disminuir las especies más sensibles, la riqueza disminuirá, quedando aquellos macroinvertebrados con una valencia ecológica mayor (GOLD). El segundo grupo de variables nos ofrecen una perspectiva interesante en la gestión del medio, donde la presencia del arbolado (bosques de ribera, matorral, etc.) implica una mayor protección de las familias de insectos bentónicos. Si bien son conocidos los efectos beneficiosos de los bosques como: fuentes de materia orgánica; creadores de heterogeneidad del medio; refugio de crecidas y filtro de sedimentos y de nutrientes, que permiten una mayor diversidad de taxones presentes en el tramo fluvial.

Finalmente se muestran las variables topográficas, las cuales muestran el tipo de cuenca y la morfología de los cauces. Estas se han dejado para el final, pues se ven influenciadas por las anteriores. Respecto al área de cuenca, se puede decir que cuanto mayor sea, mayor será el caudal e indica la distancia relativa a las zonas de desembocadura o al nacimiento del río. No es de extrañar que las variables, cuyas puntuaciones más altas representan a comunidades de aguas prístinas, tengan una correlación negativa con el tamaño de la cuenca, y aquellas que prefieren los cauces más amplios, zonas de sedimentación, temperaturas más altas, tengan una relación positiva. Un factor que discretiza la anatomía de los taxones es el gradiente del tramo; como ejemplo se expondrá la familia de dípteros: *Blephariceridae*, filtradores de alta montaña, que han desarrollado unas ventosas en la parte abdominal que le sirven para adherirse a las rocas de las torrenteras donde vive. Por otro lado, tenemos los oligoquetos. Estos son gusanos que no poseen patas ni estructuras fijadoras, por lo que prefieren las zonas de fondos arenosos, fangosos para vivir y protegerse. Estas zonas no se pueden mantener por mucho tiempo si existen fuertes pendientes donde el agua discurre rápido (trombas de agua o calados pequeños), por lo que indicarán zonas de acumulación de sedimentos, caudales lo más estables posibles, y

llanuras. La pendiente media de las cuencas y el grado de encajonamiento, indican algo parecido, ya que cuanto mayor sea la pendiente, antes llegará la precipitación de la ladera al cauce, produciendo alteraciones en el sustrato, limitando las zonas de protección (granulometría del sustrato).

### 7.3. Modelo

Como hemos visto en la tabla 4, los resultados numéricos del modelo no han sido tan buenos como cabría de esperar, aunque son bastante aceptables, es decir, en algunos índices no se han introducido las variables que expliquen adecuadamente la variabilidad que determinan las características de las comunidades, pero en otros han mostrado patrones muy claros. Los valores bajos de la Riqueza de taxones y el GOLD hacen pensar que se deben de introducir otras variables predictoras, como quizás las propiedades fisicoquímicas. Sin embargo, vemos como los valores de los ajustes de los valores con respecto la recta de regresión (RMSE), son bastante buenos. Si además tenemos en cuenta que existe una gran variabilidad en las estaciones de muestreo (personal, periodos de muestreo, diferentes tipologías de ríos, climas...), y permiten interpretar unos patrones tan claros, podemos concluir que el modelo ha sido consistente.

Los buenos resultados proceden de los índices IASPT y EPT-R, con un  $R^2$  mayor que 5,5, llegando al 0,67 en el primer caso, con unos valores de la desviación buenos. Esto implica que las variables introducidas explican una gran parte de la distribución de estas comunidades. Finalmente, el valor del IBMWP, no es del todo malo, porque casi llega al 50 % de los casos. Como la intención es utilizar las variables dependientes para comprender los patrones espaciales más relevantes de las comunidades, el resultado general puede entenderse como moderado - bueno, y se pueden sacar conclusiones válidas para el estudio.

Los resultados obtenidos del testing, permiten validar los resultados del entrenamiento. Estos han sido similares a los obtenidos en la calibración, como regla generalizada vemos que el modelo ha aprendido del entrenamiento, con el caso más destacado del EPT llegando al 0,62, y el de la Riqueza, al 0,29. No es un aumento significativo, pero se puede decir que se mantiene estable. De esta manera se concluye con la misma opinión que en el entrenamiento, que las variables predictoras no son tan acertadas para explicar la variabilidad que poseen las comunidades a través de estos índices. Sin embargo, no se puede recaer toda la responsabilidad sobre las variables, aludiendo a que la heterogeneidad de los valores tomados en campo, no sean tan decisivos como se desearía.

#### 7.4. Modelado: patrones espaciales y mapas de predicción.

La escala de trabajo ha sido un factor muy importante y determinante en el estudio. Observándose en las relaciones tan bajas que se muestran entre variables y en su espectro de actuación, ya que la mayoría de los patrones se refieren a medias de la cuenca o subcuenca. Como se ha comentado anteriormente, existen tramos muy amplios que no recogen la heterogeneidad del entorno y precisa de valores promediados para explicar la distribución de las comunidades. Son 3 las variables independientes que poseen un foco de actuación más reducido: la evapotranspiración potencial del buffer de 200 metros; la pendiente del valle y el gradiente del tramo fluvial.

El primer patrón espacial hace referencia a una teoría ecológica llamada “la alteración intermedia” (Connell, 1978), donde se explica porque aumenta la diversidad taxonómica cuando se producen unas perturbaciones leves. Esta teoría se explica mediante la ocupación del suelo agrícola en la cuenca y el gradiente del tramo. Con la primera variable, se observa que el número de taxones es máximo cuando la cuenca posee un 10% del suelo con uso agrícola. Lo mismo ocurre con el gradiente del tramo, cuyo umbral se encuentra con una pendiente de 0,03, y pasado ese límite desciende la riqueza. En aquellas zonas donde predomina un patrón adquiere mayor densidad de población, pero la diversidad disminuye (Barquin & Death, 2004), y esto puede darse porque en las zonas más estables se produce una mayor competitividad por los recursos o la predación, generando dominancia de unas especies sobre otras. En el caso de los gasterópodos, oligoquetos y dípteros, aumentan en estas condiciones, debido a que los nutrientes que provienen de la actividad agrícola hacen crecer las algas o peryphiton que ralentizan las aguas, además de producirse un aumento de partículas en suspensión, traídas por el arado de las tierras y la escorrentía. En el caso de la riqueza en plecópteros, efemerópteros y tricópteros, descienden por este mismo motivo, al empeorar las condiciones del hábitat, bien sea por el consumo de oxígeno en la oxidación de los fertilizantes agrícolas, mayor turbidez, etc.

Otro de los usos del suelo que es determinante en la distribución espacial de los macroinvertebrados, es la presencia de suelos urbanos. Por norma general estos emiten sales, fosfatos (Álvarez-Cabria et al, 2016), aumentando la DBO y otros componentes como la impermeabilización del suelo, destrucción de hábitats ribereños etc. De esta manera, aquellos taxones que tengan una cierta sensibilidad podrán verse afectados (EPT), y otros como los quironómidos, gasterópodos con una valencia ecológica alta pueden permanecer en esos ámbitos. Además, el crecimiento de fitobentos permite aumentar la cantidad de alimento de taxones raspadores o colectores.

Respecto a las valencias ecológicas de las comunidades se puede observar como las comunidades GOLD son mucho más resistentes que otros, cuya sensibilidad les impide distribuirse por medios contaminados, permitiendo una predominancia de los taxones estenoicos.

A través de los patrones espaciales podemos zonificar cada tramo de río y la presencia de los macroinvertebrados presentes. En las partes altas se acumulan las precipitaciones, que junto con las laderas escarpadas y el gradiente del río, forman una lámina de agua con poco calado, con elevada velocidad, creando formaciones tipo pool-step. En estas zonas no se acumulan sedimentos ni materia orgánica particulada en grandes cantidades, por lo tanto, las especies presentes en este tipo de lugares deben poseer unas características fisiológicas para soportar las condiciones hidrológicas, bien sea por adaptación hidrodinámica o partes del cuerpo que permitan agarrarse, porque los refugios presentes se limitan a los bloques de roca. Otro de los factores limitantes y que precisan de adaptación es la captación del oxígeno. En las zonas de montaña la turbulencia controla el grado de oxígeno disuelto en el agua y para poder captarlo se precisan de unas branquias externas, porque algunos taxones no poseen la capacidad de generar movimientos de agua, la captación al producirse por difusión necesita del movimiento del agua (Allan, 2002). La alimentación también se ven condicionada, debido a ese arrastre de nutrientes, por lo que las fuentes de alimentación son externas al canal del río (hojas, madera) aunque también se produce la depredación. Generalmente con corrientes tan fuertes y turbulentas (números de Reynolds superior a 10000 y regímenes supercríticos) (Giller & Malmqvist, 1998), se producen las denominadas derivas, que consisten en el arrastre de los macroinvertebrados por la fuerza del agua. Si no existiesen métodos de recolonización de los tramos altos estos se quedarían desérticos, de modo que los taxones, como los EPT, han desarrollado alas en su fase adulta para remontar el río volando. Pero dichos taxones no son muy buenos voladores, por ello precisan de un bosque de ribera bien conservados a lo largo de la cuenca para protegerles de las ráfagas del viento y proveerles de alimento.

En el otro extremo se sitúan las partes bajas de los valles fluviales, donde la presencia de agua es continua y la velocidad de flujo es menor, en parte debido al gradiente del tramo fluvial, donde el proceso de sedimentación tiene mayor relevancia, aportando partículas más finas que en tramo superiores no se han podido retener porque la velocidad crítica de estas (Neill, 1968) se alcanza fácilmente. Al existir mayor cantidad de sedimento, se generan otro tipo de refugios, ideal para aquellos taxones que necesitan enterrarse. Estas zonas presentan flujos más lentos permiten establecerse a diversos tipos de fitobentos, implicando un aumento de la materia orgánica disuelta, ideal para aquellos taxones que son recolectores y raspadores. El aumento de

la materia orgánica produce un aumento de la demanda biológica de oxígeno (DBO) y la ausencia de movimiento de las aguas impiden a muchos EPT desarrollar su actividad en estas zonas (existen excepciones como los Baétidos). Relacionado con el oxígeno disuelto, se encuentra la temperatura, donde en dichas zonas son más elevadas que en las partes altas, conservando menor concentración de oxígeno y dependiendo en mayor grado de la actividad fotosintética de periphyton y otras especies vegetales de las aguas corrientes.

Los mapas desarrollados por el modelo permiten una mejor comprensión de lo escrito anteriormente, sintetizando conceptos de manera gráfica. La información más relevante que destacar son dos tipos de gradientes; uno que diferencia la composición de comunidades debidas al clima (temperatura y precipitaciones) entre las dos biorregiones, la atlántica y la mediterránea. Donde la primera posee una mayor diversidad, así como una mayor riqueza de taxones de los órdenes *plecoptera*, *efemeroptera* y *trichoptera*. Y el otro, se encuentra dentro de la parte mediterránea, donde las precipitaciones influyen en la riqueza taxonómica en las partes altas o de montaña y otras más limitadas, situadas en las depresiones fluviales, viéndose acentuadas por la presencia más acusada de suelos con usos agrícolas y concentraciones urbanas.

## 8. CONCLUSIONES

El presente estudio ha conseguido modelar, con una buena capacidad, los patrones espaciales de los macroinvertebrados bentónicos para toda la Península Ibérica. A una escala tan elevada se ha conseguido destacar aquellas variables más importantes que determinan la distribución de comunidades de macroinvertebrados.

Se ha puesto en valor la capacidad de los modelos para tratar grandes bases de datos y obtener de manera sintética los conceptos más relevantes para la elaboración de teorías fundamentadas a grandes escalas, así como de contrastar teorías ya desarrolladas, como las perturbaciones leves, que a una escala más local hubiese pasado desapercibida.

Las variables respuesta han conseguido discriminar, al menos, las distribuciones de las comunidades con comportamientos más polarizados, como lo son los EPT y los GOLD. Junto con la ayuda de la riqueza taxonómica, el IBMWP y el IASPT, se ha podido contrastar comportamientos y adaptaciones presentes en los taxones a través de la distribución de estos.

Los mapas desarrollados no se deben observar como índices de calidad, ya que han sido utilizados y modelados como descriptores de la comunidad. Estos nos han permitido observar otro tipo de gradientes presentes entre regiones tan distintas, siendo de ayuda al concentrar tanta información en una sola imagen.

Todos los resultados han sido discutidos y comparados con otros realizados en laboratorios o en cuencas más pequeñas, donde las variables independientes son más homogéneas y ofrecen visiones más regionales, además de poseer unas relaciones entre las variables independientes y la variable respuesta son más fuertes y el nivel de detalle más alto.

Dentro de los patrones especiales que se han mostrado en el estudio, cabe destacar los usos de suelo agrícolas (MN\_AGR) y urbanos (MN\_UHD) por su interacción directa con el ser humano. A través de estos valores y comprendiendo que la mayoría de las variables son de componente promediado a la cuenca, se pueden incorporar dichos patrones a la gestión de los sistemas hídricos, desarrollando estrategias de conservación de las aguas a través de, primero, un análisis más genérico, y posteriormente un análisis más pormenorizado de la estructura de la sociedad y la interacción con este medio.

## 9. VÍAS DE TRABAJO FUTURO

A través de este trabajo se plantean algunas líneas futuras, para la evolución de esta vía de conocimiento que, teniendo un pasado a escalas más modestas, no se han planteado tanto a este nivel y esa juventud es la que precisa de un ajuste y desarrollo en proyectos futuros.

- Introducir nuevos métodos descriptivos de las comunidades de macroinvertebrados bentónicos que permitan tener una visión más completa de todos los grupos, intentando recoger aquellos que permitan una visión más completa de los patrones bióticos (redes tróficas y competitividad). De esta manera se van a conseguir resultados más ajustados a la realidad.
- Desarrollar una red fluvial virtual más pormenorizada donde se recojan de manera más explícita la heterogeneidad del medio. Este punto relacionado con el anterior, permitirían compensar las deficiencias de ambos y obtener un ajuste a la realidad mayor.
- Realizar un análisis de campo de la comunidad tras la obtención de los resultados de patrones de cada taxón. De este modo se pueden ajustar los datos, estableciendo relaciones más concretas.
- Elaboración e introducción de variables nuevas, incluyendo aquellas procedentes de los análisis fisicoquímicos. Así quedarían aclarados ciertas sombras que producen algunas variables donde sus diferentes características implican un cambio en las estructuras comunitarias.
- Elaboración de datos de campo más homogéneos y de mayor detalle. Esto va a permitir eliminar la incertidumbre que pueda suscitar la toma de datos con métodos de muestreo variados y permitirá la elaboración o introducción de grupos taxonómicos con mayor grado descriptivo, estudiando las diferencias de comportamientos entre especies con características similares dentro de una escala nacional.

## 10. BIBLIOGRAFÍA

“Agricultura en España”, s.f. En Wikipedia. Recuperado el 7 sep 2017 de [https://es.wikipedia.org/wiki/Agricultura\\_en\\_Espa%C3%B1a#Superficies\\_y\\_producciones\\_agr.C3.ADcolas](https://es.wikipedia.org/wiki/Agricultura_en_Espa%C3%B1a#Superficies_y_producciones_agr.C3.ADcolas)

ALBA-TERCEDOR, J. & A. SANCHEZ-ORTEGA, 1988. “Un método rápido y simple para evaluar la calidad biológica de las aguas corrientes basado en el de Hellawell (1978).” *Limnética*, 4: 51-56.

ALBA-TERCEDOR, J., JÁIMEZ-CUÉLLAR, P., ÁLVAREZ, M., AVILÉS, J., BONADA, N., CASAS, J., MELLADO, A., ORTEGA, M., PARDO, I., PRAT, N., RIERADEVALL, M., ROBLES, S., SÁINZ-CANTERO, C.E., SÁNCHEZ-ORTEGA, A., SUÁREZ, M.L., TORO, M., VIDAL-ABARCA, M.R., VIVAS, S. & ZAMORA-MUÑOZ, C. 2002. Caracterización del estado ecológico de ríos mediterráneos ibéricos mediante el índice IBMWP (antes BMWP'). *Limnetica* 21(3-4): 175-185.

Álvarez-Cabria, M., Barquín, J., Juanes, J.A., 2011. Macroinvertebrate community dynamics in a temperate European Atlantic river. Do they conform to general ecological theory?. *Hydrobiologia* (2011) 658:277–291

Álvarez-Cabria, M., Barquín, J., Juanes, J.A., 2016. Modelling the spatial and seasonal variability of water quality for entire. *Science of The Total Environment*, Volumes 545–546, Pages 152-162

Álvarez-Cabria, M., González-Ferreras, A.M. Barquín, J., Juanes, J.A., 2017. Modelling macroinvertebrate and fish biotic indices: From reaches to entire river networks. *Science of the Total Environment* 577, 308–318.

Angelier E., 2002. *Ecología de las aguas corrientes*. Ed: Acribia. S.A. Zaragoza.

Araújo M.B. & Williams P.H., 2000. Selecting areas for species persistence using occurrence data. *Biological Conservation* 96: 331-345.

Balian, E. V., H. Segers, C. Leveque, and K. Martens. 2008. The freshwater animal diversity assessment: An overview of the results. *Hydrobiologia* 595:627-637.

Barquin et al, 2015. Coupling virtual watersheds with ecosystem services assessment: a 21st century platform to support river research and management. *WIREs Water* 2015, 2:609–621. doi: 10.1002/wat2.1106.

- Barbour, M. T., J. Gerritsen, B. D. Snyder & J. B. Stribling. 1999. Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish, 2nd ed. EPA 841-B-99-002. US EPA, Office of Water, Washington D.C., USA.
- Barquín, J. & Death, G., 2004. Patterns of invertebrate diversity in streams and freshwater springs in Northern Spain. Stuttgart, November 2004. Arch. Hydrobiol. 161 3 329–349.
- Bonada, N., Prat, N., Resh, V.H., Statzner, B., 2005. DEVELOPMENTS IN AQUATIC INSECT BIOMONITORING: A Comparative Analysis of Recent Approaches. Annu. Rev. Entomol. 2006. 51:495–523.
- Breiman, L. 2001. Random forests. Machine Learning 45:5-32.
- Breiman, 2005. Breiman and Cutler's Random Forests for Classification and Regression. Package 'randomForest'. Version 4.6-12.
- Busby, J. R. (1991) BIOCLIM – a bioclimatic analysis and prediction system. In Margules, C. R. & Austin, M. P. (Eds.) Nature Conservation: Cost Effective Biological Surveys and Data Analysis. East Melbourne, Australia: CSIRO, pp. 64–68.
- Carauna R., Karampatzakis N., Yessemañoma A., 2008. An Empirical Evaluation of Supervised Learning in High Dimensions. Helsinki, Finland. ISBN: 978-1-60558-205-4.
- Connell, J.H., 1978. Diversity in Tropical Rain Forests and Coral Reefs. Science, New Series, Vol. 199, No. 4335. (Mar. 24, 1978), pp. 1302-1310.
- Quintas, L. 2016. Clasificación hidrográfica de los ríos de España. CEDEX. Depósito Legal: M 26047-2016 ISSN: 0211-8203.
- Craig J.F. & Kemper J.B., 1985. Regulated Streams. En Regulated Streams. Advances in Ecology (pág. 427). Plenum Press: New York.
- Diaz R.J, Solan M., Valente R.M. 2004. (Es diaz et al...). A review of approaches for classifying benthic habitats and evaluating habitat quality. Journal of Environmental Management 73 (2004) 165–181.
- Dunne, T. and Leopold, L.B. (1978) Water in Environmental Planning, W.H. Freeman, San Francisco.
- ELITH J, JR LEATHWICK & T HASTIE (2008) A working guide to boosted regression trees. Journal of Animal Ecology 77: 802-813.

- Estrela T, Quintas L. 1996. El sistema integrado de modelización precipitación-aportación SIMPA. *Revista de Ingeniería Civil*. 104:43–52.
- Giller P.S. & Malmqvist B. (1998) *The biology of streams and rivers*. Oxford University Inc., New York.
- Gonzalez-Ferreras A.M., Barquín J., Peñas F.J. (2016). Integration of habitat models to predict fish distributions in several watersheds of Northern Spain. *J. Appl. Ichthyol.* 32 (2016), 204–216.
- González-Oreja J.A. De la Fuente A.A., Hernandez-Santín, Buzo-Franco S., Bonache- Regidor C., 2010. Evaluación de estimadores no paramétricos de la riqueza de especies. Un ejemplo con aves en áreas verdes de la ciudad de Puebla, México. *Animal Biodiversity and Conservation*, 2010. Vol.: 33, Num.: 1.
- Hanson P., Springer M., Ramirez A. 2010. Introducción a los grupos de macroinvertebrados acuáticos. *Rev. biol. trop* vol.58 suppl.4 San José Dec. 2010.
- Hastie, T. & R. Tibshirani. 1990. *Generalized additive models*. Chapman & Hall, London, 344 pp.
- Hynes, H.B., 1975. the stream and its valley. *Internationales Vereinigung für theoretische und angewandte Limnologie*, 19, 1-15.
- IDEE MAPAMA. [Servicios Web Mapas \(WMS\)](#). Estatales. “Pluviometría” y “Temperatura media anual”.
- IGME. 2015. Mapa de Litologías de España 1/1.000.000.
- INE, 2011. Evolución sectorial de la población ocupada en España. Ministerio de Fomento.
- Kuhn, S., Egert, B., Neumann, S., Steineck, C., 2008. Building blocks for automated elucidation of metabolites: machine learning methods for NMR prediction. *BMC Bioinf.* 9:400.
- Lomba, A., L. Pellissier, C. Randin, J. Vicente, F. Moreira, J. Honrado, and A. Guisan. 2010. Overcoming the rare species modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. *Biological Conservation* 143:2647-2657.
- Mandaville, S.M., 2002. *Benthic Macroinvertebrates in Freshwaters- Taxa Tolerance Values, Metrics, and Protocols*.

- Manolaki, P., Parastergiadou, E., 2013. The impact of environmental factors on the distribution pattern of aquatic macrophytes in a middle-sized Mediterranean stream. *Aquatic Botany* 104 (2013) 34–46.
- Mateo R.G., Felicísimo A.M., Muñoz J., 2011. Modelos de distribución de especies: Una revisión sintética. *Revista Chilena de Historia Natural* 84: 217-240, 2011.
- Margalef, R. (1969). El concepto de polución y sus indicadores biológicos. *Documentos de investigación Hidrológica. Suplemento de la revista Agua*, 7, 105–133.
- Marini, M. A., M. Barbet-Massin, J. Martinez, N. P. Prestes, and F. Jiguet. 2010. Applying ecological niche modelling to plan conservation actions for the Redspetacled Amazon (*Amazona pretrei*). *Biological Conservation* 143:102-112.
- McCullagh P., 1983. Generalized linear models. *European Journal of Operational Research* Volume 16, Issue 3, June 1984, Pages 285-292.
- METI-2015. Protocolo de cálculo del índice multimétrico específico del tipo de invertebrados bentónicos en ríos. Ministerio de Agricultura, Alimentación y Medio Ambiente (España). Secretaría General Técnica. NIPO: 280-15-123-1.
- Miller, D. 2003. *Programs for DEM Analysis*, Seattle, WA.
- Neill, C.R. (1968). A RE-EXAMINATION OF THE BEGINNING OF MOVEMENT FOR COARSE GRANULAR-BED MATERIALS. H. R. S. Wallingford.
- ORTEGA-HUERTA MA & AT PETERSON (2004) Modelling spatial patterns of biodiversity for conservation prioritization in North-eastern Mexico. *Diversity and Distributions* 10: 39-54.
- P. J. BOON, 1988. The impact of river regulation on invertebrate communities in the U.K. *Regulated rivers: research and management*, VOL. 2, 389-409
- Pattee, E., Lascombe, C. and Delolme, R. (1973) Effects of temperature on the distribution of turbellarian triclads, in *Effects of Temperature on Ectothermic Organisms*, (ed. W. Weiser), Springer-Verlag, New York, pp. 201-8.
- Peñas, F.J., Barquin J., Snelder T.H., Booker, D.J. and Alvarez C., 2014. The influence of methodological procedures on hydrological classification performance. *Hydrol. Earth Syst. Sci.*, 18, 3393–3409, 2014.
- Petts, G.E., 1984. *Impounded Rivers*, John Wiley, Chichester.

- Plafkin, J.L., Barbour, M.T., Porter, K.D., Gross, S.K., and Hughes, R.M.. 1989. Rapid Bioassessment Protocols for use in Streams and Rivers: Benthic Macroinvertebrates and Fish. Edición: 2. U.S. Environmental Protection Agency. EPA 440/4-89/001. 8 chapters, Appendices A-D.
- Phillips, S.J., Anderson, R.P., Schapire, R.E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190 (2006) 231–259.
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rivas-Martínez, S., A. Penas, and T. E. Díaz. 2004. Bioclimatic Map of Europe, Bioclimates. In: Cartographic Service. University of León, León, Spain.
- Rosenberg DM, Resh VH (1993) Introduction to freshwater biomonitoring and benthic macroinvertebrates. In: Rosenberg DM, Resh VH (eds) freshwater biomonitoring and benthic macroinvertebrates. Chapman & Hall, New York, pp 1–9.
- Shannon, C.E. and W. Weaver. 1949. *The Mathematical Theory of Communication*. University Illinois Press, Urbana, IL.
- SEOANE J, JH JUSTRIBÓ, F GARCÍA, J RETAMAR, F RABADÁN et al. (2006) Habitat-suitability modelling to assess the effects of land-use changes on Dupont's lark *Chersophilus duponti*: A case study in the Layna Important Bird Area. *Biological Conservation* 128: 241-252.
- Snelder, T. H, Lamouroux, N., Pella, H., 2011. Empirical modelling of large scale patterns in river bed surface grain size. *Geomorphology* 127 (2011) 189–197.
- Snelder T.H., Datry, T., Lamouroux N., Larned, S. T., Sauquet, E., Pella H., Catalogne C., 2013. Regionalization of patterns of flow intermittence from gauging station records. *Hydrology and Earth System Sciences.*, 17, 2685–2699, 2013.
- Thuiller, W., B. Lafourcade, R. Engler, and M. B. Araujo. 2009b. BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography* 32:369- 373.
- Urbina-Cardona J.N., Castro F., 2010. DISTRIBUCIÓN ACTUAL Y FUTURA DE ANFIBIOS Y REPTILES CON POTENCIAL INVASOR EN COLOMBIA: UNA APROXIMACION USANDO MODELOS DE NICHO ECOLÓGICO.

Vayssières M.P., Plant R.E., Allen-Diaz B.H., Classification trees: An alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science*, October 2000. Volume 11, Issue 5. Pages 679–694.

WARD, J.V. & J.A. STANDFORD, 1983. The intermediate disturbance hypothesis: and explanation for biotic diversity patterns in lotic ecosystems. En: *The dynamics of lotic ecosystems*. Fontaine, T.D. y S.M. (eds.). Ann Arbor Michigan

Worldwide Bioclimatic Classification System, 1996-2017, S.Rivas-Martinez & S.Rivas-Saenz, Phytosociological Research Center, Spain. <http://www.globalbioclimatics.org>

## 11. ANEXOS

Anexo 1: Taxones de macroinvertebrados bentónicos presentes en los datos NABIA.

Taxón	Nivel	Incorporación
Acariformes	Superorden	Apto
Aeshnidae	Familia	Apto
Ancylidae	Familia	Apto
Anthomyiidae	Familia	Apto
Aphelocheiridae	Familia	Apto
Arrenuridae	Familia	No apto
Asellidae	Familia	Apto
Athericidae	Familia	Apto
Baetidae	Familia	Apto
Beraeidae	Familia	Apto
Bithyniidae	Familia	Apto
Bivalvia	Clase	No apto
Blephariceridae	Familia	Apto
Brachycentridae	Familia	Apto
Branchiobdellidae	Familia	Apto
Caenidae	Familia	Apto
Calamoceratidae	Familia	Apto
Calopterygidae	Familia	Apto
Capniidae	Familia	Apto
Carychiidae	Familia	No apto
Ceratopogonidae	Familia	Apto
Chaoboridae	Familia	Apto
Chironomidae	Familia	Apto
Chloroperlidae	Familia	Apto
Chrysomelidae	Familia	Apto
Coenagrionidae	Familia	Apto
Coleoptera	Orden	No apto
Corbiculidae	Familia	Apto
Cordulegastridae	Familia	Apto
Corduliidae	Familia	Apto
Corixidae	Familia	Apto
Corophiidae	Familia	No apto
Crambidae	Familia	No apto
Culicidae	Familia	Apto
Curculionidae	Familia	Apto
Cylindrotomidae	Familia	No apto
Diptera	Orden	No apto

Taxón	Nivel	Incorporación
Dixidae	Familia	Apto
Dolichopodidae	Familia	Apto
Dreissenidae	Familia	Apto
Dryopidae	Familia	Apto
Dugesidae	Familia	Apto
Dytiscidae	Familia	Apto
Ecnomidae	Familia	Apto
Elmidae	Familia	Apto
Empididae	Familia	Apto
Ephemerellidae	Familia	Apto
Ephemeridae	Familia	Apto
Ephemeroptera	Orden	No apto
Ephydriidae	Familia	Apto
Erpobdellidae	Familia	Apto
Ferrissiidae	Familia	Apto
Gammaridae	Familia	Apto
Gastropoda	Clase	No apto
Gerridae	Familia	Apto
Glossiphoniidae	Familia	Apto
Glossosomatidae	Familia	Apto
Goeridae	Familia	Apto
Gomphidae	Familia	Apto
Gordiidae	Familia	No apto
Gyrinidae	Familia	Apto
Haemopidae	Familia	No apto
Haliplidae	Familia	Apto
Haplotaenidae	Familia	No apto
Hebridae	Familia	No apto
Helicopsychidae	Familia	Apto
Helophoridae	Familia	Apto
Heptageniidae	Familia	Apto
Hirudidae	Familia	Apto
Hirudinea	Clase	No apto
Hydraenidae	Familia	Apto
Hydrobiidae	Familia	Apto
Hydrochidae	Familia	Apto
Hydrodromidae	Familia	No apto

Taxón	Nivel	Incorporación
Hydrometridae	Familia	Apto
Hydrophilidae	Familia	Apto
<b>Hydropsychidae</b>	Familia	Apto
Hydroptilidae	Familia	Apto
Hydrosaphidae	Familia	Apto
Hygrobidae	Familia	Apto
Lepidostomatidae	Familia	Apto
Leptoceridae	Familia	Apto
Leptophlebiidae	Familia	Apto
Lestidae	Familia	Apto
Leuctridae	Familia	Apto
Libellulidae	Familia	Apto
Limnephilidae	Familia	Apto
Limoniidae	Familia	Apto
Lumbricidae	Familia	Apto
Lumbriculidae	Familia	Apto
Lymnaeidae	Familia	Apto
Melanopsidae	Familia	Apto
Mesoveliidae	Familia	Apto
Molannidae	Familia	Apto
Muscidae	Familia	Apto
Naucoridae	Familia	Apto
Nematoda	Filo	Apto
Nemouridae	Familia	Apto
Nepidae	Familia	No apto
Neritidae	Familia	Apto
Niphargidae	Familia	No apto
Noteridae	Familia	Apto
Notonectidae	Familia	Apto
Odonata	Orden	Apto
Odontoceridae	Familia	Apto
Oligochaeta	Clase	Apto
Oligoneuriidae	Familia	Apto
Osmylidae	Familia	Apto
Palaemonidae	Familia	Apto
Pediciidae	Familia	Apto
Perlidae	Familia	Apto
Perlodidae	Familia	Apto
Philopotamidae	Familia	Apto
Phryganeidae	Familia	Apto
Physidae	Familia	Apto
Piscicolidae	Familia	No apto

Taxón	Nivel	Incorporación
Planariidae	Familia	Apto
Planorbidae	Familia	Apto
Platycnemididae	Familia	Apto
Viviparidae	Familia	No apto
Platyhelminthes	Filo	No apto
Plecoptera	Orden	Apto
Pleidae	Familia	Apto
Polycentropodidae	Familia	Apto
Polymitarciidae	Familia	Apto
Potamanthidae	Familia	Apto
Prosopistomatidae	Familia	No apto
Psychodidae	Familia	Apto
Psychomyiidae	Familia	Apto
Ptychopteridae	Familia	Apto
Rhagionidae	Familia	Apto
Rhyacophilidae	Familia	Apto
Scathophagidae	Familia	Apto
Sciomyzidae	Familia	Apto
Scirtidae	Familia	Apto
Seriata	Orden	No apto
Sericostomatidae	Familia	Apto
Sialidae	Familia	Apto
Simuliidae	Familia	Apto
Siphonuridae	Familia	Apto
Sisyridae	Familia	Apto
Sphaeriidae	Familia	Apto
Sphaeromatidae	Familia	Apto
Stratiomyidae	Familia	Apto
Succineidae	Familia	Apto
Syrphidae	Familia	Apto
Tabanidae	Familia	Apto
Taeniopterygidae	Familia	Apto
Thaumaleidae	Familia	Apto
Thiaridae	Familia	Apto
Tipulidae	Familia	Apto
Trichoptera	Orden	No apto
Tubificidae	Familia	Apto
Uenoidae	Familia	Apto
Unionidae	Familia	No apto
Valvatidae	Familia	Apto
Veliidae	Familia	Apto

## Anexo 2. Variables independientes introducidas al modelo.

<b>Nombres</b>	<b>Descripción</b>	<b>Grupo</b>
AREA_KM2	Área total de la cuenca	Topográfico
ToOutletKm	Distancia desde el segmento a la desembocadura	Topográfico
LENGTH_M	Longitud del segmento	Topográfico
ELEV_M	Promedio de la elevacion de los pixels en el segmento	Topográfico
WIDTH	Anchura del rio	Topográfico
DEPTH	Profundidad del rio	Topográfico
GRADIENT	Pendiente media a lo largo del tramo (cambio vertical/longitud horizontal)	Topográfico
VWI	Indice de anchura del valle	Topográfico
MnSlope	Pendiente media del terreno	Topográfico
elevWing	Media de la elevacion de los pixels en la subcuenca	Topográfico
ToSourceKm	Distance from segment to river source.	Topográfico
SINUOSITY	Sinuosidad del segmento	Topográfico
CHAN_DEN	Densidad del drenaje: densidad de drenaje para canales con un area de contribución > 1.0 km2	Topográfico
CONF_DEN	Numero de confluencias en el area de la cuenca	Topográfico
AREA_ADJ	Area de la subcuenca	Topográfico
MN_PREC	Precipitación media anual en la cuenca	Climático
LC_PREC	Precipitación media anual en la subcuenca	Climático
BF_PREC	Precipitación media anual en un buffer de 200 m	Climático
BC_PREC	Precipitación media anual en un buffer de 200 m acumulado aguas arriba del tramo	Climático
MN_TEMP	Temperatura media anual en la cuenca	Climático
LC_TEMP	Temperatura media anual en la subcuenca	Climático
BF_TEMP	Temperatura media anual en un buffer de 200 m	Climático
BC_TEMP	Temperatura media anual en un buffer de 200 m acumulado aguas arriba del tramo	Climático
R_BC_TEMP	Temperatura media anual en un buffer de 200 m acumulado aguas arriba del tramo en la orilla derecha	Climático
L_BC_TEMP	Temperatura media anual en un buffer de 200 m acumulado aguas arriba del tramo en la orilla izquierda	Climático
MN_EI	Evapotranspiración real media anual en la cuenca	Climático
LC_EI	Evapotranspiración real media anual en la subcuenca	Climático

BF_EI	Evapotranspiración real media anual en un buffer de 200 m	Climático
BC_EI	Evapotranspiración real media anual en un buffer de 200 m acumulado aguas arriba del tramo	Climático
MN_ETP	Evapotranspiración potencial media anual en la cuenca	Climático
LC_ETP	Evapotranspiración potencial media anual en la subcuenca	Climático
BF_ETP	Evapotranspiración potencial media anual en un buffer de 200 m	Climático
BC_ETP	Evapotranspiración potencial media anual en un buffer de 200 m acumulado aguas arriba del tramo	Climático
MN_calc	Porcentaje de área ocupada por roca caliza en la cuenca	Climático
LC_calc	Porcentaje de área ocupada por roca caliza en la subcuenca	Geológico
BF_calc	Porcentaje de área ocupada por roca caliza en un buffer de 200 m	Geológico
BC_calc	Porcentaje de área ocupada por roca caliza en el buffer de 200 m acumulado aguas arriba del tramo	Geológico
MN_cong	Porcentaje de área ocupada por conglomerados en la cuenca	Geológico
LC_cong	Porcentaje de área ocupada por conglomerados en la subcuenca	Geológico
BF_cong	Porcentaje de área ocupada por conglomerados en un buffer de 200 m	Geológico
BC_cong	Porcentaje de área ocupada por conglomerados en el buffer de 200 m acumulado aguas arriba del tramo	Geológico
MN_snds	Porcentaje de área ocupada por areniscas en la cuenca	Geológico
LC_snds	Porcentaje de área ocupada por areniscas en la subcuenca	Geológico
BF_snds	Porcentaje de área ocupada por areniscas en un buffer de 200 m	Geológico
BC_snds	Porcentaje de área ocupada por areniscas en el buffer de 200 m acumulado aguas arriba del tramo	Geológico
MN_sdim	Porcentaje de área ocupada por acumulo sedimentos en la cuenca	Geológico
LC_sdim	Porcentaje de área ocupada por sedimentos en la subcuenca	Geológico
BF_sdim	Porcentaje de área ocupada por sedimentos en un buffer de 200 m	Geológico
BC_sdim	Porcentaje de área ocupada por sedimentos en el buffer de 200 m acumulado aguas arriba del tramo	Geológico
MN_shle	Porcentaje de área ocupada por esquistos en la cuenca	Geológico
LC_shle	Porcentaje de área ocupada por esquistos en la subcuenca	Geológico
BF_shle	Porcentaje de área ocupada por esquistos en un buffer de 200 m	Geológico
BC_shle	Porcentaje de área ocupada por esquistos en el buffer de 200 m acumulado aguas arriba del tramo	Geológico
MN_slic	Porcentaje de área ocupada por roca silíceas en la cuenca	Geológico

LC_slic	Porcentaje de área ocupada por roca silíceas en la subcuenca	Geológico
BF_slic	Porcentaje de área ocupada por roca silíceas en un buffer de 200 m	Geológico
BC_slic	Porcentaje de área ocupada por roca silíceas en el buffer de 200 m acumulado aguas arriba del tramo	Geológico
MN_slte	Porcentaje de área ocupada por pizarras en la cuenca	Geológico
LC_slte	Porcentaje de área ocupada por pizarras en la subcuenca	Geológico
BF_slte	Porcentaje de área ocupada por pizarras en un buffer de 200 m	Geológico
BC_slte	Porcentaje de área ocupada por pizarras en el buffer de 200 m acumulado aguas arriba del tramo	Geológico
MN_vlc	Porcentaje de área ocupada por rocas volcánicas en la cuenca	Geológico
LC_vlc	Porcentaje de área ocupada por rocas volcánicas en la subcuenca	Geológico
BF_vlc	Porcentaje de área ocupada por rocas volcánicas en un buffer de 200 m	Geológico
BC_vlc	Porcentaje de área ocupada por rocas volcánicas en el buffer de 200 m acumulado aguas arriba del tramo	Geológico
MN_watr	Porcentaje de área ocupada por hábitats acuáticos en la cuenca	Geológico
LC_watr	Porcentaje de área ocupada por hábitats acuáticos en la subcuenca	Geológico
BF_watr	Porcentaje de área ocupada por hábitats acuáticos en un buffer de 200 m	Geológico
BC_watr	Porcentaje de área ocupada por hábitats acuáticos en un buffer de 200 m acumulado aguas arriba del tramo	Geológico
MN_PERM	Permeabilidad media de la roca presente desde el tramo de río considerado hasta la parte más alta de la cuenca (valor de 1 a 5)	Geológico
LC_PERM	Permeabilidad media de la roca presente en la subcuenca (valor de 1 a 5)	Geológico
BF_PERM	Permeabilidad media de la roca presente en un buffer de 200m (valor de 1 a 5)	Geológico
BC_PERM	Permeabilidad media de la roca presente en un buffer de 200m acumulado aguas arriba del tramo (valor de 1 a 5)	Geológico
MN_COND	Conductividad media del agua conferida por el tipo de roca presente desde el tramo de río considerado hasta la parte más alta de la cuenca (valor de 1 a 5)	Geológico
LC_COND	Conductividad media del agua conferida por el tipo de roca presente en la subcuenca (valor de 1 a 5)	Geológico
BF_COND	Conductividad media del agua conferida por el tipo de roca presente en un buffer de 200m (valor de 1 a 5)	Geológico
BC_COND	Conductividad media del agua conferida por el tipo de roca presente en un buffer de 200m acumulado aguas arriba del tramo (valor de 1 a 5)	Geológico
MN_HARD	Dureza media de la roca presente desde el tramo de río considerado hasta la parte más alta de la cuenca (valor de 1 a 5)	Geológico

LC_HARD	Dureza media de la roca presente en la subcuenca (valor de 1 a 5)	Geológico
BF_HARD	Dureza media de la roca presente en un buffer de 200m (valor de 1 a 5)	Geológico
BC_HARD	Dureza media de la roca presente en un buffer de 200m acumulado aguas arriba del tramo (valor de 1 a 5)	Geológico
MN_UHD	Porcentaje de área ocupada por áreas urbanas en la cuenca	Uso de suelo
LC_UHD	Porcentaje de área ocupada por áreas urbanas en la subcuenca	Uso de suelo
BF_UHD	Porcentaje de área ocupada por áreas urbanas en un buffer de 200 m	Uso de suelo
BC_UHD	Porcentaje de área ocupada por áreas urbanas en un buffer de 200 m acumulado aguas arriba de tramo	Uso de suelo
MN_AGR	Porcentaje de área ocupada por usos agrícolas en la cuenca	Uso de suelo
LC_AGR	Porcentaje de área ocupada por usos agrícolas en la subcuenca	Uso de suelo
BF_AGR	Porcentaje de área ocupada por usos agrícolas en un buffer de 200 m	Uso de suelo
BC_AGR	Porcentaje de área ocupada por usos agrícolas en un buffer de 200 m acumulado aguas arriba de tramo	Uso de suelo
MN_PAS	Porcentaje de área ocupada por pastos en la cuenca	Uso de suelo
LC_PAS	Porcentaje de área ocupada por pastos en la subcuenca	Uso de suelo
BF_PAS	Porcentaje de área ocupada por pastos en un buffer de 200 m	Uso de suelo
BC_PAS	Porcentaje de área ocupada por pastos en un buffer de 200 m acumulado aguas arriba de tramo	Uso de suelo
MN_BLF	Porcentaje de área ocupada por bosque de frondosas en la cuenca	Uso de suelo
LC_BLF	Porcentaje de área ocupada por bosque de frondosas en la subcuenca	Uso de suelo
BF_BLF	Porcentaje de área ocupada por bosque de frondosas en un buffer de 200 m	Uso de suelo
BC_BLF	Porcentaje de área ocupada por bosque de frondosas en un buffer de 200 m acumulado aguas arriba del tramo	Uso de suelo
MN_CNF	Porcentaje de área ocupada por bosque de coníferas en la cuenca	Uso de suelo
LC_CNF	Porcentaje de área ocupada por bosque de coníferas en la subcuenca	Uso de suelo
BF_CNF	Porcentaje de área ocupada por bosque de coníferas en un buffer de 200 m	Uso de suelo
BC_CNF	Porcentaje de área ocupada por bosque de coníferas en un buffer de 200 m acumulado aguas arriba del tramo	Uso de suelo
MN_PLT	Porcentaje de área ocupada por plantaciones forestales en la cuenca	Uso de suelo
LC_PLT	Porcentaje de área ocupada por plantaciones forestales en la subcuenca	Uso de suelo
BF_PLT	Porcentaje de área ocupada por plantaciones forestales en un buffer de 200 m	Uso de suelo
BC_PLT	Porcentaje de área ocupada por plantaciones forestales en un buffer de 200 m acumulado aguas arriba del tramo	Uso de suelo
MN_SSH	Porcentaje de área ocupada por matorrales en la cuenca	Uso de suelo
LC_SSH	Porcentaje de área ocupada por matorrales en la subcuenca	Uso de suelo

BF_SSH	Porcentaje de área ocupada por matorrales en un buffer de 200 m	Uso de suelo
BC_SSH	Porcentaje de área ocupada por matorrales en un buffer de 200 m acumulado aguas arriba del tramo	Uso de suelo
MN_WAE	Porcentaje de área ocupada por hábitats acuáticos en la cuenca	Uso de suelo
LC_WAE	Porcentaje de área ocupada por hábitats acuáticos en la subcuenca	Uso de suelo
BF_WAE	Porcentaje de área ocupada por hábitats acuáticos en un buffer de 200 m	Uso de suelo
BC_WAE	Porcentaje de área ocupada por hábitats acuáticos en un buffer de 200 m acumulado aguas arriba del tramo	Uso de suelo
MN_DEN	Porcentaje de área ocupada por zonas sin vegetación en la cuenca	Uso de suelo
LC_DEN	Porcentaje de área ocupada por zonas sin vegetación en la subcuenca	Uso de suelo
BF_DEN	Porcentaje de área ocupada por zonas sin vegetación en un buffer de 200 m	Uso de suelo
BC_DEN	Porcentaje de área ocupada por zonas sin vegetación en un buffer de 200 m acumulado aguas arriba del tramo	Uso de suelo
Order_ST	Orden del rio en el tramo (Strahler)	Topográfico
Order_SH	Orden del rio en el tramo (Shreve)	Topográfico
Order_PU	Orden del rio en el tramo (PU)	Topográfico
A_DAB	Distancia al azud más cercano aguas abajo (m)	Presiones Antrópicas
A_DAR	Distancia al azud más cercano aguas arriba (m). Limitación de 5000 m	Presiones Antrópicas
P_DAB	Distancia a la presa más cercana aguas abajo (m)	Presiones Antrópicas
P_DAR	Distancia a la presa más cercana aguas arriba (m). Limitación de 5000 m	Presiones Antrópicas
V_DAB	Distancia al punto de vertido más cercano aguas abajo (m)	Presiones Antrópicas
V_DAR	Distancia al punto de vertido más cercano aguas arriba (m). Limitación de 5000 m	Presiones Antrópicas
CAN_DAB	Distancia a la canalización más cercana aguas abajo (m)	Presiones Antrópicas
CAN_DAR	Distancia a la canalización más cercana aguas arriba (m). Limitación de 5000 m	Presiones Antrópicas
COB_DAB	Distancia a la cobertura más cercana aguas abajo (m)	Presiones Antrópicas
COB_DAR	Distancia a la cobertura más cercana aguas arriba (m). Limitación de 5000 m	Presiones Antrópicas
PRO_DAB	Distancia a la protección más cercana aguas abajo (m)	Presiones Antrópicas
PRO_DAR	Distancia a la protección más cercana aguas arriba (m). Limitación de 5000 m	Presiones Antrópicas

Anexo 3: Correlación de Spearman Rank entre los índices de calidad y las variables independientes.

Variables	Riqueza	EPT (Riqueza)	IBMWP	IASPT	GOLD	EPT (Ab. Relativa)
AREA_KM2	-0.216	-0.307	-0.291	-0.344	0.255	-0.026
ToOutletKm	-0.089	0.005	-0.069	0.022	0.016	0.122
LENGTH_M	-0.015	-0.033	-0.02	-0.032	0.02	-0.001
ELEV_M	0.058	0.206	0.133	0.27	-0.142	0.131
WIDTH	-0.161	-0.178	-0.198	-0.188	0.198	-0.018
DEPTH	-0.169	-0.198	-0.212	-0.211	0.207	-0.02
GRADIENT	0.283	0.342	0.346	0.351	-0.391	0.151
VWI	-0.247	-0.283	-0.278	-0.247	0.26	-0.122
MnSlope	0.263	0.39	0.35	0.42	-0.413	0.229
elevWing	0.108	0.281	0.2	0.346	-0.209	0.153
ToSourceKm	-0.171	-0.241	-0.231	-0.272	0.205	-0.001
SINUOSITY	0.048	0.01	0.04	0.011	0.016	0.008
CHAN_DEN	0.187	0.262	0.236	0.256	-0.274	0.187
CONF_DEN	0.194	0.267	0.241	0.256	-0.224	0.13
AREA_ADJ	-0.061	-0.052	-0.07	-0.062	0.071	-0.039
MN_PREC	0.263	0.558	0.419	0.661	-0.279	0.039
LC_PREC	0.293	0.51	0.429	0.601	-0.273	0.016
BF_PREC	0.292	0.501	0.424	0.589	-0.27	0.015
R_BF_PREC	0.291	0.499	0.423	0.587	-0.267	0.013
L_BF_PREC	0.291	0.501	0.424	0.59	-0.271	0.016
BC_PREC	0.262	0.547	0.413	0.645	-0.28	0.042
R_BC_PREC	0.263	0.547	0.414	0.645	-0.28	0.042
L_BC_PREC	0.262	0.546	0.413	0.645	-0.28	0.043
MN_TEMP	-0.056	-0.352	-0.199	-0.521	0.198	-0.145
LC_TEMP	-0.093	-0.4	-0.247	-0.564	0.185	-0.092
BF_TEMP	-0.094	-0.398	-0.248	-0.563	0.182	-0.09
R_BF_TEMP	-0.095	-0.399	-0.248	-0.563	0.182	-0.091
L_BF_TEMP	-0.095	-0.399	-0.249	-0.562	0.182	-0.09
BC_TEMP	-0.047	-0.344	-0.189	-0.513	0.188	-0.147
R_BC_TEMP	-0.047	-0.344	-0.189	-0.513	0.188	-0.147
L_BC_TEMP	-0.047	-0.344	-0.189	-0.513	0.188	-0.146
MN_EI	0.099	0.192	0.152	0.262	-0.02	-0.108
LC_EI	0.152	0.252	0.224	0.335	-0.085	-0.07
BF_EI	0.169	0.278	0.244	0.349	-0.108	-0.06
R_BF_EI	0.166	0.274	0.241	0.347	-0.108	-0.06
L_BF_EI	0.17	0.277	0.244	0.346	-0.107	-0.06
BC_EI	0.116	0.226	0.18	0.306	-0.05	-0.093
R_BC_EI	0.115	0.226	0.18	0.307	-0.05	-0.092
L_BC_EI	0.117	0.226	0.181	0.306	-0.05	-0.094

Variables	Riqueza	EPT (Riqueza)	IBMWP	IASPT	GOLD	EPT (Ab. Relativa)
MN_ETP	-0.186	-0.455	-0.319	-0.556	0.201	-0.018
LC_ETP	-0.18	-0.431	-0.312	-0.546	0.181	0.014
BF_ETP	-0.184	-0.428	-0.312	-0.533	0.189	0.013
R_BF_ETP	-0.183	-0.427	-0.311	-0.53	0.187	0.014
L_BF_ETP	-0.184	-0.427	-0.312	-0.533	0.19	0.013
BC_ETP	-0.186	-0.457	-0.319	-0.555	0.2	-0.012
R_BC_ETP	-0.185	-0.456	-0.318	-0.554	0.199	-0.012
L_BC_ETP	-0.186	-0.457	-0.32	-0.556	0.2	-0.011
MN_calc	0.014	-0.12	-0.019	-0.083	-0.067	0.047
LC_calc	0.129	0.048	0.138	0.099	-0.136	0.013
BF_calc	0.125	0.073	0.142	0.122	-0.14	0.01
R_BF_calc	0.124	0.073	0.14	0.118	-0.141	0.006
L_BF_calc	0.126	0.07	0.141	0.119	-0.137	0.005
BC_calc	0.021	-0.093	-0.006	-0.062	-0.07	0.044
R_BC_calc	0.022	-0.091	-0.005	-0.061	-0.071	0.044
L_BC_calc	0.022	-0.092	-0.005	-0.06	-0.069	0.046
MN_cong	-0.279	-0.401	-0.351	-0.391	0.189	0.053
LC_cong	-0.157	-0.259	-0.201	-0.235	0.085	0.063
BF_cong	-0.077	-0.155	-0.107	-0.145	0.013	0.07
R_BF_cong	-0.074	-0.155	-0.105	-0.149	0.011	0.063
L_BF_cong	-0.078	-0.147	-0.104	-0.133	0.009	0.072
BC_cong	-0.235	-0.35	-0.301	-0.343	0.137	0.072
R_BC_cong	-0.233	-0.349	-0.3	-0.344	0.135	0.075
L_BC_cong	-0.235	-0.347	-0.3	-0.341	0.135	0.072
MN_snd	0.032	0.169	0.064	0.134	-0.065	0.014
LC_snd	0.068	0.151	0.099	0.137	-0.092	0.012
BF_snd	0.077	0.144	0.106	0.136	-0.083	0.009
R_BF_snd	0.079	0.142	0.107	0.13	-0.08	0.008
L_BF_snd	0.077	0.145	0.105	0.134	-0.087	0.013
BC_snd	0.001	0.117	0.026	0.088	-0.044	0.017
R_BC_snd	0.001	0.116	0.025	0.087	-0.043	0.017
L_BC_snd	-0.001	0.114	0.023	0.085	-0.042	0.017
MN_sdim	-0.349	-0.457	-0.42	-0.412	0.365	-0.091
LC_sdim	-0.324	-0.385	-0.381	-0.353	0.335	-0.078
BF_sdim	-0.304	-0.346	-0.356	-0.328	0.314	-0.067
R_BF_sdim	-0.296	-0.34	-0.347	-0.321	0.314	-0.067
L_BF_sdim	-0.302	-0.347	-0.355	-0.33	0.311	-0.068
BC_sdim	-0.338	-0.436	-0.406	-0.396	0.357	-0.083
R_BC_sdim	-0.338	-0.435	-0.405	-0.395	0.356	-0.084
L_BC_sdim	-0.338	-0.435	-0.406	-0.398	0.356	-0.086
MN_shle	-0.014	0.012	-0.021	-0.027	-0.032	0.01
LC_shle	-0.01	NA	-0.017	-0.023	NA	NA
BF_shle	NA	NA	NA	NA	NA	NA

Variables	Riqueza	EPT (Riqueza)	IBMWP	IASPT	GOLD	EPT (Ab. Relativa)
R_BF_shle	NA	NA	NA	NA	NA	NA
L_BF_shle	NA	NA	NA	NA	NA	NA
BC_shle	-0.009	-0.012	-0.018	-0.029	0.008	0.018
R_BC_shle	-0.009	-0.012	-0.018	-0.029	0.008	0.018
L_BC_shle	-0.009	-0.012	-0.018	-0.029	0.008	0.018
MN_slic	0.025	0.208	0.082	0.198	-0.052	0.039
LC_slic	0.101	0.243	0.154	0.209	-0.14	0.058
BF_slic	0.125	0.234	0.174	0.214	-0.126	0.06
R_BF_slic	0.121	0.228	0.169	0.209	-0.119	0.061
L_BF_slic	0.127	0.234	0.175	0.213	-0.127	0.065
BC_slic	0.024	0.192	0.08	0.19	-0.039	0.041
R_BC_slic	0.023	0.191	0.079	0.19	-0.038	0.042
L_BC_slic	0.023	0.19	0.079	0.189	-0.04	0.043
MN_slte	0.053	0.058	0.041	-0.008	0.016	-0.025
LC_slte	0.075	0.061	0.07	0.032	-0.058	0.013
BF_slte	0.071	0.072	0.07	0.039	-0.065	0.023
R_BF_slte	0.072	0.07	0.069	0.035	-0.062	0.018
L_BF_slte	0.07	0.074	0.069	0.043	-0.067	0.028
BC_slte	0.042	0.046	0.027	-0.016	0.028	-0.029
R_BC_slte	0.041	0.047	0.027	-0.016	0.027	-0.03
L_BC_slte	0.041	0.046	0.027	-0.016	0.028	-0.029
MN_vlc	0.019	0.105	0.03	0.047	0.056	-0.107
LC_vlc	0.137	0.218	0.171	0.19	-0.018	-0.11
BF_vlc	0.142	0.202	0.173	0.19	-0.016	-0.102
R_BF_vlc	0.125	0.176	0.153	0.173	-0.009	-0.095
L_BF_vlc	0.133	0.181	0.159	0.172	-0.018	-0.105
BC_vlc	0.021	0.104	0.037	0.068	0.058	-0.112
R_BC_vlc	0.02	0.102	0.035	0.066	0.06	-0.111
L_BC_vlc	0.018	0.098	0.032	0.061	0.064	-0.118
MN_watr	-0.148	-0.179	-0.186	-0.184	0.11	0.064
LC_watr	-0.082	-0.093	-0.104	-0.11	0.031	0.013
BF_watr	-0.078	-0.099	-0.1	-0.104	0.044	0.002
R_BF_watr	-0.081	-0.099	-0.103	-0.104	0.041	-0.002
L_BF_watr	-0.074	-0.088	-0.094	-0.096	0.034	0.015
BC_watr	-0.144	-0.179	-0.186	-0.193	0.115	0.061
R_BC_watr	-0.145	-0.18	-0.187	-0.195	0.113	0.063
L_BC_watr	-0.144	-0.175	-0.185	-0.192	0.111	0.063
MN_PERM	-0.039	-0.089	-0.044	-0.034	-0.027	0.041
LC_PERM	-0.118	-0.121	-0.122	-0.082	0.04	0.012
BF_PERM	-0.143	-0.139	-0.152	-0.107	0.074	-0.003
R_BF_PERM	-0.144	-0.137	-0.151	-0.103	0.076	-0.005
L_BF_PERM	-0.142	-0.137	-0.152	-0.111	0.076	-0.004
BC_PERM	-0.081	-0.121	-0.09	-0.074	0.03	0.026

Variables	Riqueza	EPT (Riqueza)	IBMWP	IASPT	GOLD	EPT (Ab. Relativa)
R_BC_PERM	-0.083	-0.124	-0.092	-0.077	0.032	0.025
L_BC_PERM	-0.079	-0.118	-0.087	-0.071	0.028	0.027
MN_COND	-0.072	-0.246	-0.133	-0.232	0.03	0.019
LC_COND	-0.086	-0.221	-0.131	-0.192	0.019	0.023
BF_COND	-0.083	-0.207	-0.124	-0.175	0.022	0.011
R_BF_COND	-0.083	-0.207	-0.124	-0.176	0.02	0.008
L_BF_COND	-0.081	-0.201	-0.121	-0.171	0.019	0.011
BC_COND	-0.071	-0.232	-0.13	-0.226	0.018	0.025
R_BC_COND	-0.071	-0.233	-0.13	-0.227	0.018	0.024
L_BC_COND	-0.071	-0.231	-0.129	-0.224	0.017	0.026
MN_HARD	0.262	0.443	0.358	0.453	-0.247	0.021
LC_HARD	0.279	0.413	0.354	0.394	-0.29	0.03
BF_HARD	0.269	0.383	0.339	0.376	-0.277	0.026
R_BF_HARD	0.266	0.382	0.336	0.373	-0.276	0.026
L_BF_HARD	0.269	0.384	0.339	0.376	-0.277	0.027
BC_HARD	0.259	0.43	0.352	0.44	-0.251	0.032
R_BC_HARD	0.257	0.429	0.35	0.438	-0.248	0.031
L_BC_HARD	0.26	0.432	0.353	0.44	-0.253	0.034
MN_UHD	-0.266	-0.399	-0.346	-0.384	0.41	-0.187
LC_UHD	-0.198	-0.277	-0.245	-0.247	0.242	-0.088
BF_UHD	-0.147	-0.21	-0.176	-0.171	0.142	-0.048
R_BF_UHD	-0.113	-0.176	-0.143	-0.156	0.125	-0.052
L_BF_UHD	-0.157	-0.206	-0.183	-0.168	0.135	-0.033
BC_UHD	-0.236	-0.328	-0.286	-0.28	0.304	-0.077
R_BC_UHD	-0.228	-0.309	-0.276	-0.271	0.296	-0.068
L_BC_UHD	-0.23	-0.323	-0.281	-0.277	0.299	-0.078
MN_AGR	-0.274	-0.477	-0.401	-0.555	0.37	-0.14
LC_AGR	-0.271	-0.412	-0.368	-0.453	0.305	-0.073
BF_AGR	-0.191	-0.348	-0.276	-0.377	0.225	-0.062
R_BF_AGR	-0.175	-0.321	-0.254	-0.352	0.204	-0.053
L_BF_AGR	-0.172	-0.318	-0.251	-0.349	0.202	-0.057
BC_AGR	-0.236	-0.442	-0.353	-0.501	0.319	-0.117
R_BC_AGR	-0.231	-0.434	-0.346	-0.494	0.314	-0.116
L_BC_AGR	-0.234	-0.439	-0.35	-0.492	0.315	-0.116
MN_PAS	0.006	-0.093	-0.008	-0.042	-0.017	0.006
LC_PAS	0.04	-0.016	0.045	0.035	-0.06	-0.001
BF_PAS	0.02	0.004	0.035	0.052	-0.052	0.029
R_BF_PAS	0	-0.007	0.015	0.046	-0.045	0.021
L_BF_PAS	0.022	0.003	0.035	0.042	-0.044	0.025
BC_PAS	-0.039	-0.086	-0.038	-0.021	0.03	0.015
R_BC_PAS	-0.047	-0.09	-0.047	-0.032	0.033	0.013
L_BC_PAS	-0.037	-0.089	-0.036	-0.019	0.038	0.013
MN_BLF	0.138	0.334	0.222	0.374	-0.137	0.014

Variables	Riqueza	EPT (Riqueza)	IBMWP	IASPT	GOLD	EPT (Ab. Relativa)
LC_BLF	0.188	0.422	0.294	0.464	-0.198	0.036
BF_BLF	0.174	0.399	0.271	0.428	-0.153	0.017
R_BF_BLF	0.163	0.374	0.256	0.406	-0.143	0.015
L_BF_BLF	0.169	0.376	0.262	0.406	-0.139	0.004
BC_BLF	0.17	0.423	0.278	0.469	-0.164	0.002
R_BC_BLF	0.16	0.41	0.267	0.458	-0.151	0
L_BC_BLF	0.163	0.415	0.27	0.459	-0.159	-0.004
MN_CNF	0.11	0.025	0.08	-0.012	-0.034	0.021
LC_CNF	0.167	0.141	0.174	0.127	-0.095	0.004
BF_CNF	0.178	0.166	0.189	0.139	-0.119	-0.002
R_BF_CNF	0.178	0.139	0.183	0.115	-0.113	-0.018
L_BF_CNF	0.158	0.152	0.168	0.123	-0.1	-0.001
BC_CNF	0.089	0.017	0.061	-0.019	-0.034	0.027
R_BC_CNF	0.092	0.011	0.061	-0.029	-0.036	0.027
L_BC_CNF	0.083	0.012	0.053	-0.03	-0.031	0.037
MN_PLT	0.09	0.158	0.118	0.153	0.016	-0.056
LC_PLT	0.063	0.161	0.107	0.19	-0.011	-0.08
BF_PLT	0.027	0.116	0.062	0.144	0.001	-0.041
R_BF_PLT	0.012	0.089	0.043	0.122	0.006	-0.048
L_BF_PLT	0.019	0.104	0.051	0.126	-0.002	-0.026
BC_PLT	0.003	0.062	0.021	0.071	0.066	-0.074
R_BC_PLT	-0.009	0.048	0.006	0.058	0.069	-0.063
L_BC_PLT	-0.003	0.046	0.009	0.052	0.078	-0.072
MN_SSH	0.277	0.408	0.354	0.396	-0.291	0.102
LC_SSH	0.243	0.303	0.284	0.27	-0.256	0.082
BF_SSH	0.169	0.122	0.158	0.064	-0.172	0.085
R_BF_SSH	0.137	0.067	0.116	0.012	-0.134	0.071
L_BF_SSH	0.14	0.089	0.123	0.027	-0.15	0.086
BC_SSH	0.213	0.223	0.227	0.157	-0.221	0.101
R_BC_SSH	0.192	0.196	0.203	0.137	-0.209	0.099
L_BC_SSH	0.204	0.201	0.209	0.125	-0.199	0.099
MN_WAE	-0.183	-0.193	-0.225	-0.213	0.123	0.043
LC_WAE	-0.141	-0.113	-0.162	-0.121	0.072	0.033
BF_WAE	-0.154	-0.119	-0.173	-0.124	0.085	0.03
R_BF_WAE	-0.152	-0.115	-0.17	-0.12	0.087	0.025
L_BF_WAE	-0.152	-0.118	-0.17	-0.122	0.09	0.024
BC_WAE	-0.178	-0.177	-0.22	-0.216	0.147	0.034
R_BC_WAE	-0.17	-0.171	-0.212	-0.208	0.139	0.031
L_BC_WAE	-0.175	-0.172	-0.216	-0.213	0.144	0.033
MN_DEN	0.042	0.202	0.114	0.264	-0.157	0.15
LC_DEN	0.027	0.06	0.049	0.09	-0.06	0.039
BF_DEN	-0.021	-0.032	-0.027	-0.03	-0.003	0.017
R_BF_DEN	-0.033	-0.053	-0.041	-0.04	-0.001	0.012

Variables	Riqueza	EPT (Riqueza)	IBMWP	IASPT	GOLD	EPT (Ab. Relativa)
L_BF_DEN	-0.019	-0.042	-0.031	-0.048	0.003	0.021
BC_DEN	-0.045	0.017	-0.024	0.051	-0.056	0.11
R_BC_DEN	-0.064	0.001	-0.044	0.033	-0.044	0.099
L_BC_DEN	-0.043	0.003	-0.03	0.025	-0.041	0.104
Order_ST	-0.062	-0.06	-0.085	-0.093	0.041	0.074
Order_SH	-0.063	-0.076	-0.089	-0.102	0.04	0.086
Order_PU	-0.169	-0.237	-0.229	-0.269	0.201	0.002
A_DAB	-0.037	-0.044	-0.049	-0.091	0.018	0.083
A_DAR	-0.021	0.042	0.027	0.147	0.045	0.043
P_DAB	-0.064	-0.025	-0.052	-0.005	0.052	0.052
P_DAR	-0.022	-0.055	-0.034	-0.053	-0.007	0.004
V_DAB	0.017	0.016	0.011	-0.047	-0.023	0.078
V_DAR	-0.148	-0.173	-0.173	-0.144	0.161	0.007
CAN_DAB	0.047	0.025	0.034	-0.011	-0.093	0.077
CAN_DAR	-0.143	-0.128	-0.145	-0.099	0.181	-0.031
COB_DAB	-0.13	-0.013	-0.088	0.087	0.099	-0.032
COB_DAR	-0.055	-0.053	-0.05	-0.018	0.086	-0.062
PRO_DAB	0.012	0.055	0.026	0.058	-0.055	0.105
PRO_DAR	-0.079	-0.042	-0.061	0.025	0.065	0.065

Casillas verdes: valores superiores a  $|0,25|$ . Casillas amarillas: variables seleccionadas para introducir al modelo. Valores en rojo: aquellos con un p-valor  $< 0,05$ .

Anexo 4: Tamaño de las muestras introducidas al modelo y estadísticas básicas.

	Riqueza	EPT	IBMWP	IASPT	GOLD
Min	1	1	2	2	0.5
1st Qu	22	5	100	4.5	7.5
Median	30	10	151	5.2	12.82
Mean	31.44	10.32	159.5	5.162	19.67
3rd Qu	41	15	215	5.86	23.84
Max	72	27	398	7.46	99.66

Anexo 5: Importancia de las variables según el IncNodePurity

Riqueza	
Variable	Importancia
MN_AGR	56607.29
GRADIENT	51125.11
LC_PREC	50658.02
MnSlope	48992.33
MN_SSH	48759.99
MN_sdim	42411.23
MN_cong	39850.19
LC_HARD	37744.75

EPT	
Variable	Importancia
MN_PREC	9282.28
MN_UHD	3923.75
MN_AGR	3864.77
MN_HARD	3751.60
BC_ETP	3673.77
elevWing	3178.70
BC_UHD	2970.31
BC_BLF	2579.05
MN_SSH	2548.69
MnSlope	2465.27
MN_sdim	2386.09
MN_cong	2327.44
LC_HARD	2324.96
VWI	2236.80
LC_TEMP	2039.39
LC_BLF	1721.56
GRADIENT	1700.56
AREA_KM2	1480.66
BF_EI	1394.11
CONF_DEN	1254.84
BF_AGR	927.89

IBMWP	
Variable	Importancia
MN_AGR	1315922
LC_PREC	1293084
MN_sdim	1057074
MN_UHD	1051033
MN_ETP	953036.6
MN_cong	882045.1
MnSlope	856066.7
MN_HARD	837733.7
VWI	773170.5
GRADIENT	756136.3
MN_SSH	748840.5
AREA_KM2	690914.6
BC_UHD	670251.8
LC_HARD	617108.9
LC_BLF	455063.5
BF_BLF	454138.3
BF_AGR	402580.6

IASPT	
Variable	Importancia
MN_PREC	367.99
LC_TEMP	174.38
MN_AGR	135.00
MN_UHD	97.11
elevWing	90.75
BF_ETP	87.17
MN_HARD	68.99
BC_UHD	65.84
LC_HARD	52.84
MN_SSH	51.93
LC_BLF	47.98
MnSlope	43.92
BC_BLF	42.47
MN_sdim	41.62
GRADIENT	38.74
MN_EI	35.42
MN_cong	31.21
AREA_KM2	31.20
ToSourceKm	29.62
BF_EI	29.52
LC_SSH	25.35
MN_DEN	24.66
BF_AGR	24.19
CONF_DEN	18.71