

Universidad Pública de Navarra

Nafarroako Unibertsitate Publikoa

ESCUELA TECNICA SUPERIOR
DE INGENIEROS AGRONOMOS

*NEKAZARITZAKO INGENIARIEN
GOI MAILAKO ESKOLA TEKNIKOA*

*Análisis de la distribución espacial y selección del hábitat del Milano
Real (Milvus Milvus) en Navarra.*

presentado por

AROA CORTE GARCÍA

aurkeztua

MÁSTER EN SISTEMAS DE INFORMACIÓN GEOGRÁFICA Y TELEDETECCIÓN
MASTERRA INFORMAZIO SISTEMA GEOGRAFIKOETAN ETA TELEDETEKZIOAN

Noviembre, 2020 / 2020, Azaroa



Agradecimientos

Quiero comenzar por agradecer a Gabriel Berasategi de Ornitolan Servicios Ambientales por acercarme a una nueva disciplina, mostrándome su labor educativa y medioambiental, cuyo conocimiento y datos me permitieron iniciarme en el análisis del hábitat del Milano Real. De igual forma agradecer a Ignacio Dean y su hijo Alberto, de la Sociedad de Ciencias Naturales Gorosti, por su amabilidad y labor, así como por facilitarme los datos de Milano Real a escala regional. A mis compañeros y profesores del MUSIGT, especialmente a mis tutores Miguel Ángel y Manuel, cuya ayuda ha sido fundamental para orientar este trabajo.

Finalmente, a mi familia, pareja y amigos, sin vuestro apoyo desde Asturias durante todo este tiempo, no habría sido posible.



Resumen

Los métodos de análisis espacial han sufrido diversos cambios en las últimas décadas debido a las nuevas herramientas tecnológicas disponibles, a la disponibilidad de información libre, así como su incursión en diversos campos científicos. Este es el caso de la Biogeografía, cuyo incremento del interés por encontrar métodos cuantitativos que expliquen la presencia de especies y su distribución, se ha visto propiciado por el desarrollo de los Sistemas de Información Geográfica (SIG) y su implementación mediante algoritmos en las herramientas estadísticas (R).

Mediante este trabajo se expondrá una metodología de investigación aplicada para la generación de modelos de distribución de especies y selección de hábitat de las especies (HSM), ejemplificándolo en el caso práctico de una especie amenazada, el Milano Real (Milvus Milvus) en la región de Navarra.

Palabras clave: Milvus Milvus, Modelos de idoneidad de hábitat (HSM), Sistemas de Información Geográfica (SIG), Biomod2, Maxent

Abstract

Spatial analysis methods have undergone various changes in recent decades due to the new technological tools available, the availability of free information, as well as its foray into various scientific fields. This is the case of Biogeography, whose increased interest in finding quantitative methods that explain the presence of species and their distribution, has been facilitated by the development of Geographic Information Systems (GIS) and their implementation algorithms in statistical tools (R).

This work will set out an applied research methodology for the generation of species distribution models and species habitat selection (HSM), exemplifying it in the case study of a threatened species, the Milano Real (Milvus Milvus) in the Navarra region.

Keywords: Milvus Milvus, Habitat Suitability Models (HSMs), Geographic Information Systems (GIS), Biomod2, Maxent

Índice general

ÍNDICE DE FIGURAS	4
ÍNDICE DE TABLAS	5
RESUMEN SIGLAS Y ACRÓNIMOS.....	5
1. INTRODUCCIÓN	6
1.1 Nicho ecológico y modelos de distribución de especies	6
1.2 Milano Real.....	9
1.3 Objetivos	12
2. MATERIALES Y MÉTODOS	12
2.1 Materiales.....	12
2.1.1 Área de estudio.....	12
2.1.2 Datos de muestreo.....	14
2.1.3 Variables ambientales	15
2.1.3.1 Variables topográficas.....	15
2.1.3.2 Variables climáticas.....	16
2.1.3.3 Variables de proximidad.....	17
2.1.3.4 Variables relacionadas con la vegetación.....	22
2.1.3.5 Correlación de las variables	24
2.2 Métodos	24
2.2.1 Modelos de simulación	24
2.2.2 Formatos de los modelos de HSM	27
2.2.2.1 Formato del paquete <i>Biomod2</i>	27
2.2.2.2 Formato del programa <i>MaxEnt</i>	27
2.2.3 Métodos de evaluación del modelo	28
3. RESULTADOS	28
3.1 Ajuste de modelos	28
3.1.1 Calibración	28
3.1.2 Evaluación.....	32
3.1.2.1 Evaluación de los resultados de <i>Biomod2</i>	32
3.1.2.2 Evaluación de los resultados de <i>MaxEnt</i>	35
3.2 Análisis de resultados	38
3.2.1 Condicionantes de hábitat	38
3.2.2 Representación cartográfica del hábitat potencial.....	38
4. CONCLUSIONES	43
5. BIBLIOGRAFÍA	44
6. ANEXOS.....	46

ÍNDICE DE FIGURAS

- Figura 1.** Diagrama de BAM: biótico, abiótico y movimiento.
- Figura 2.** Base teórica de los modelos de predicción: presencia y variables ambientales predictoras
- Figura 3.** Fotografía de Milano Real adulto en pleno vuelo, próximo a las instalaciones de la UPNA.
- Figura 4.** Mapa de distribución del Milano Real actual.
- Figura 5.** Censo invernal de milano real en Navarra (SEO-BirdLife) promovido por Sociedad de Ciencias Naturales Gorosti - Gorosti Natur Zientzi Elkarte (Deán, 2017)
- Figura 6.** Avistamientos del Milano Real a nivel provincial y distribución de los nidos de Milano Real identificados en el área de estudio local (Navarra).
- Figura 7.** Extensión de las áreas de estudio: Navarra y Zona Media.
- Figura 8.** Mapas generados a partir del MDT 25: Mapa de Pendientes y Mapa de Orientaciones.
- Figura 9.** Mapas generados a partir de los datos de Meteo Navarra (Gobierno de Navarra). Ejemplo año 2018. Mapa de Temperatura Media (1) , Mapa de Precipitación Media (2) y Mapa de Precipitación Total (3).
- Figura 10.** Mapa de proximidad de la red hidrográfica y superficie de agua de Navarra.
- Figura 11.** Mapa de proximidad a edificaciones visibles de Navarra.
- Figura 12.** Mapa de proximidad a vías de comunicación y redes de transporte de Navarra.
- Figura 13.** Mapa de proximidad de explotaciones ganaderas de Navarra.
- Figura 14.** Mapa de proximidad de punto de vertidos residuales de Navarra.
- Figura 15.** Mapa de proximidad de vías pecuarias y pistas forestales de Navarra.
- Figura 16.** Mapa de Usos y Aprovechamientos 2012 de Navarra.
- Figura 17.** Modelo Digital de Alturas de la Vegetación (MDHV) del área de estudio.
- Figura 18.** Representación gráfica de la evaluación de los algoritmos empleados en el modelo del año 2015.
- Figura 19.** Calibración de los principales parámetros de MaxEnt(1 entrada de datos, 2 variables predictoras, 3 métodos, 4 elementos de salida, 5 rutas y 6 otros parámetros de configuración del modelo) .
- Figura 20.** Representación gráfica de los evaluadores del modelo y de la respuesta de las variables ambientales.
- Figura 21.** Comparativa de resultados de la evaluación de los métodos mediante los índices ROC y TSS.
- Figura 22.** Representación gráfica de la respuesta del modelo de año 2018 a las variables ambientales: a) GAM, b) GBM, c) GML y d) RF.
- Figura 23.** Modelo de ensamble proyectado a partir de los TSS mediante Biomod2
- Figura 24.** Comparativa de resultados de la tasa de omisión y área predictiva mediante MaxEnt.
- Figura 25.** Comparativa de resultados de la evaluación de los métodos mediante el método AUC.

Figura 26. Representación gráfica de la respuesta de las variables ambientales en los modelos 2012, 2013 y 2014 mediante MaxEnt.

Figura 27. Evaluación mediante curva ROC y representación gráfica de la respuesta de las variables ambientales de modelo de zona de nidificación mediante MaxEnt.

Figura 28. Modelos predictivos mediante árboles de decisión generados mediante paquete Biomod2, periodo 2012 -2019.

Figura 29. Modelos conjunto generados mediante paquete Biomod2, periodo 2012 -2019

Figura 30. Modelos generados mediante MaxEnt, periodo 2012 -2019.

Figura 31. Modelos generados para la zona de nidificación mediante MaxEnt y el paquete Biomod2, periodo 2012 -2019.

ÍNDICE DE TABLAS

Tabla 1. Resumen datos de observación del Milano Real entre 2012 – 2019 durante su etapa reproductiva.

Tabla 2. Tabla resumen datos de ancho teórico aplicado a la red hidrográfica de Navarra.

Tabla 3. Tabla resumen datos de ancho teórico aplicado a la red de carreteras y ferrocarril de Navarra.

Tabla 4. Matriz de correlación de las variables ambientales predictoras generada con el algoritmo “r.covar”.

Tabla 5. Resumen de los métodos más estandarizados actualmente.

Tabla 6. Resumen del aporte medio de las variables ambientales a cada modelo.

RESUMEN SIGLAS Y ACRÓNIMOS

- **CART:** Árboles de clasificación y regresión
- **GAM:** Modelo aditivo generalizado
- **GARP:** Algoritmos genéticos
- **GLM:** Métodos de regresión lineal generalizados
- **GBM :** Árboles de regresión y técnicas de remuestreo 'boosted regression trees'
- **HSM:** Modelos de idoneidad del hábitat
- **MAXENT:** Máxima entropía
- **MDE:** Modelos de distribución de especies
- **RF:** Random Forest
- **SIG:** Sistemas de Información Geográfica
- **QGIS :** Quantum Gis (software libre de Sistemas de Información Geográfica)

1. INTRODUCCIÓN

Las investigaciones ecológicas están constatando la amenaza de extinción de diversas especies animales y vegetales, debido fundamentalmente a factores antrópicos y climatológicos (Cano-Barbacid and Cano Sánchez, 2016). Debido a ello, se han generado nuevos métodos para analizar la evolución de la biodiversidad, especialmente su distribución espacial con el fin de propiciar su conservación, mediante la protección de sus hábitats así como la investigación de sus procesos reproductivos.

Entre las especies amenazadas, se encuentra el Milano Real (*Milvus Milvus*), cuya especie ha sufrido un importante declive poblacional a nivel mundial. Esta ave rapaz ha sido foco de interés a nivel nacional desde hace más de 20 años, ya que España se sitúa entre los países con mayores registros de observaciones, a consecuencia de las migraciones de poblaciones del centro y norte de Europa en periodo invernal, así como por la presencia de poblaciones residentes y reproductoras (Sunyer and Viñuela, 2000).

Ante la situación crítica que presenta esta especie desde este proyecto se planteará la problemática de cuantificar la relación espacio – especie, mediante un marco conceptual de la situación actual del ave rapaz, una metodología de generalización de modelos predictivos de su nicho ecológico y un análisis de distribución de la especie rapaz en el periodo temporal de cría (primavera – verano), para poder definir su marco geográfico y las variables que permiten su desarrollo.

1.1 Nicho ecológico y modelos de distribución de especies

Para que una especie se establezca en un determinado espacio geográfico (G), se precisa que se desarrolle la combinación de tres requisitos, ejemplificados gráficamente en el Diagrama de BAM (Soberón, J. y Peterson, 2005): capacidad de dispersión (dinámica poblacional), variables ambientales abióticas y entorno biótico (Figura 1).

La capacidad de dispersión, normalmente denominada con la letra M, muestra la posibilidad de colonización de las especies en un territorio o a la imposibilidad debido a factores externos (barreras físicas, migración forzosa, etc), desde un periodo histórico.

Las variables ambientales abióticas, designada en la comunidad científica como como A, son las condiciones propicias para el desarrollo del nicho ambiental, contexto que determinará la supervivencia y reproducción de la especie. Estas condiciones podrán ser evaluadas, ya que se tratan variables scenopoéticas (Hutchinson, 1957), variables favorables que interactúan con la especie pero difícilmente son afectadas por ella.

Y finalmente, el entorno biótico, hace referencia a las interacciones con otras especies, ya sea como fuente de alimento, foco de enfermedades o como posible depredador (Soberón, 2007). Estos factores están correlacionados con las variables scenopoéticas, favoreciendo su capacidad predictora. Requieren datos de muestreo contrastados, para determinar las poblaciones y sus efectos.

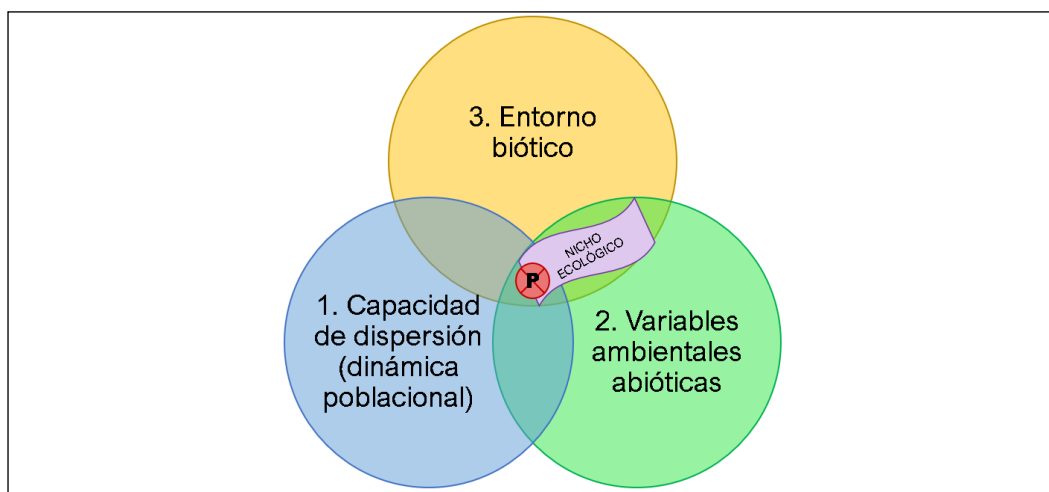


Figura 1. Diagrama de BAM: biótico, abiótico y movimiento.

Fuente: Elaboración propia tomando como base Soberón y Peterson (2005).

Estas condiciones determinarán la ocurrencia de las especies y su supervivencia, siendo de vital importancia las condiciones ambientales, las cuales definirán el nicho. La denominación del nicho y su alcance ha sido tema de debate entre los investigadores, siendo frecuentemente los más utilizados: nicho (Grinnell, 1917), nicho fundamental (Hutchinson, 1957) o nicho realizado (Soberón, 2007).

El nicho fundamental hace referencia al espacio cuyas condiciones ambientales tienen una influencia directa en la especie y en su crecimiento, sin tener en cuenta el condicionante biótico. Sin embargo, el concepto de nicho realizado se considera una definición más completa, ya que considera el nicho fundamental, pero limitándolo por las interrelaciones bióticas y su dispersión. A modo de simplificación, en este trabajo se hará referencia a este concepto como nicho o nicho ecológico, considerando como concepto más adecuado el nicho realizado.

Para poder determinar la idoneidad del hábitat o nicho ecológico de esta especie en la etapa reproductora, será preciso identificar los condicionantes que propician la selección del hábitat, mediante la aplicación de los modelos de distribución de especies.

Los modelos de distribución de especies (MDE o SDM) son modelos estadísticos que permiten estimar la ubicación geográfica de las especies, definiendo escenarios propicios para el asentamiento de especies, mediante datos de presencia y variables predictoras (Guisan y Thuiller, 2005), Figura 2.

Es fundamental mencionar en este apartado, que las predicciones de un modelo de nicho, calibrado con muestras de presencia únicamente, proporcionan una sobrepredicción del área ocupada. Este error es inevitable a no ser que se cuenten con registros verdaderos de ausencias, los cuales no constan en las bases de datos genéricas. Este sesgo se suple generalmente con el desarrollo de pseudoausencias o ausencias de trasfondo (background absences), definidos dentro del área de dispersión de la especie.



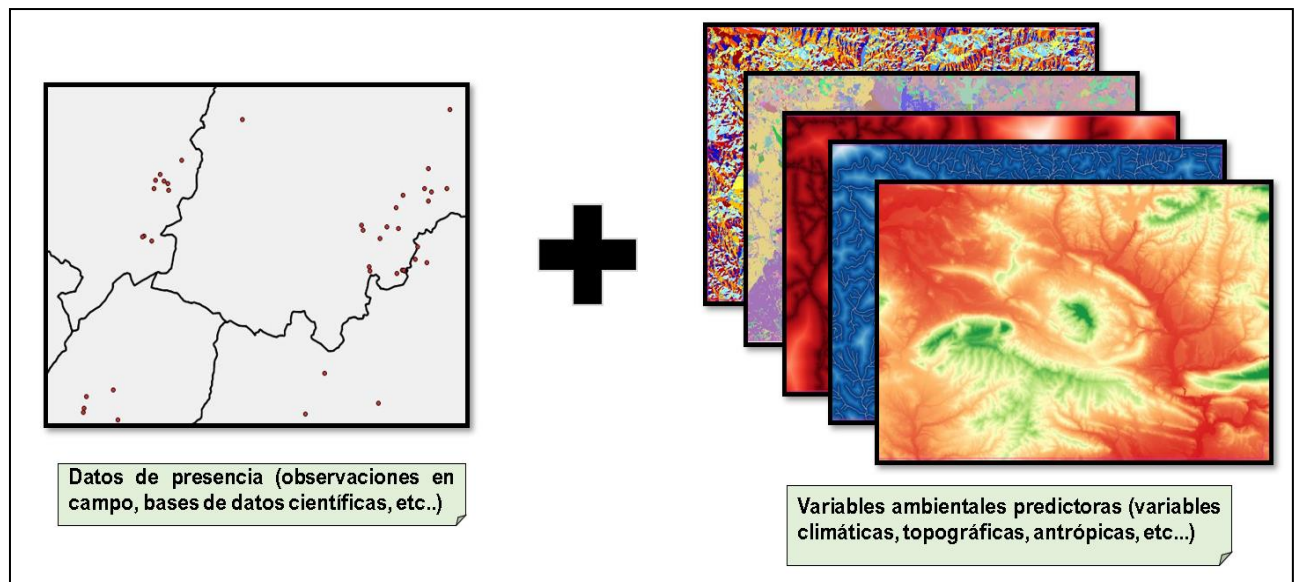


Figura 2. Base teórica de los modelos de predicción: presencia y variables ambientales predictoras.
Fuente: Elaboración propia

Aunque los modelos propuestos son similares, este concepto ha sido ampliamente debatido por los científicos desde los años 90, el cual se ha evolucionado en distintas variantes, atendiendo a su aplicación: modelos de distribución de especies (SDM), modelos de idoneidad del hábitat (HSM), modelos de distribución de hábitat (HDM), modelos de nicho ecológico (ENM) o funciones de selección de recursos (RSF) (Mateo et al., 2011). Debido a que este trabajo se centrará en los hábitats propicios para el crecimiento de las poblaciones del Milano Real en Navarra, se empleará el término modelo de idoneidad del hábitat (HSM).

Dentro de este marco de operaciones críticas en el desarrollo de los modelos de idoneidad del hábitat (HSM) tienen vital importancia los datos de presencia, las variables predictoras y la selección del método (modelo estadístico). Existen multitud de métodos estadísticos que permiten generar predicciones de calidad y controlar su calidad, mediante la calibración y la evaluación de error, siendo el más empleado generalmente la regresión logística multivariable (Mateo et al., 2011). Debido a las limitaciones de este método paramétrico se han generado una serie de métodos, entre los que se encuentran:

- Métodos de regresión (GLM, generalized linear models)
- Redes neuronales.
- Algoritmos genéticos (GARP).
- Árboles de clasificación y regresión (CART)
- Multivariate adaptive regression splines (MARS).
- Máxima entropía (algoritmo MAXENT)

Para aplicar adecuadamente los modelos de idoneidad del hábitat, se requiere el seguimiento de una serie de pasos fundamentales, los cuales se analizarán con detalle en los apartados de materiales y metodología (Guisan, A., Thuiller, W., Zimmermann, 2017).

Al tratar este trabajo de estimar los factores más determinantes de presencia y generar una hipotética área potencial con datos de presencia adquiridos en campo de forma rigurosa, se han seleccionado métodos predicción basados en presencias como es el algoritmo MAXENT (máxima entropía) para compararlos con métodos estadísticos ensamblados como es BIOMOD2 (en los que figuraran métodos de regresión, algoritmos genéricos y árboles de clasificación), los cuales generan predicciones coherentes y de procesamiento ágil, incluso partiendo de muestras de pequeño tamaño (inferior a 50 muestras).

1.2 Milano Real

El Milano Real (*Milvus Milvus*, *Red Kite* en inglés y en euskera *Miru gorria*), Figura 3, es un ave rapaz de tamaño mediano, cuya altura media en etapa adulta ronda los 60 cm y una envergadura de 140 – 170 cm. Los ejemplares hembras son de mayor tamaño, con un peso medio de 890 – 1100g , mientras que los machos alcanzan los 760 – 920 g (Manzanares, 1991). Se caracteriza por su aspecto rojizo, con plumas orladas amarillentas y cola ahorquillada, así como por su vuelo lento y continuo, pudiendo recorrer largas distancias y que con frecuencia se puede observar en las inmediaciones de las urbes.



Figura 3. Fotografía de Milano Real adulto en pleno vuelo, próximo a las instalaciones de la UPNA.

Fuente: Fotografía cedida por Pablo A. Velásquez Franco.

En las últimas décadas su distribución mundial ha sufrido una notoria regresión debido fundamentalmente a causas antrópicas (destrucción de hábitats, impacto de la caza y pesticidas), quedando reducida a la región paleártica occidental concentrándose especialmente en la mitad meridional de Europa y en algunas islas Atlánticas, así como pequeñas poblaciones en el norte de África y Turquía (Cramp, S. and Simmons, 1980), Figura 4 .

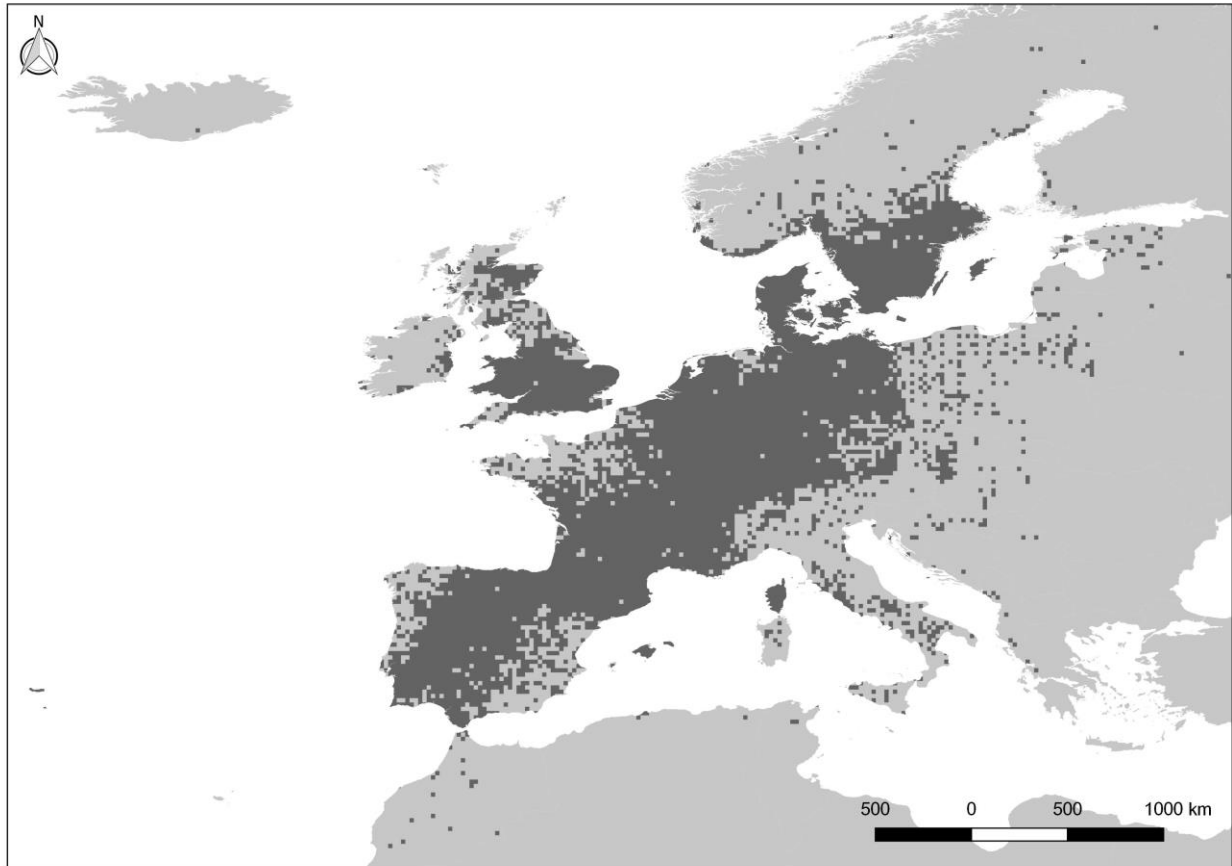


Figura 4. Mapa de distribución del Milano Real actual.
Fuente: GBIF.org (2019)

Concretamente, en la Península Ibérica, su presencia se observa principalmente en áreas de media montaña y penillanuras, destacando cuatro ámbitos: en áreas de media montaña del noreste de España (Álava, Navarra, Huesca y Zaragoza), el centro – oeste (Zamora, Salamanca y Cáceres), Sistema Central (Soria, Ávila; Segovia y Madrid) y en menor medida en el sureste de la península (Viñuela, 2004). Esta distribución denota cierta preferencia por los climas fríos o templados, así como las áreas de baja o media montaña, características a tomar en cuenta en las variables ambientales seleccionadas.

Está considerada como una especie migratoria parcial debido al desplazamiento de ejemplares el norte y centro de Europa durante el periodo invernal hacia latitudes más meridianas, como es el caso de la Península Ibérica, frente a los ejemplares sedentarios del mediterráneo y Reino Unido (Cramp, S. and Simmons, 1980). Debido a esta posición estratégica, España se ha convertido en una de las áreas de invernadas más importantes de la especie, datos contrastados gracias al censo nacional coordinado por SEO/BirdLife, desde 1994 (Viñuela et al 1999) y de los que ha sido pionera la Comunidad de Navarra, Sociedad de Ciencias Naturales Gorosti - Gorosti Natur Zientzi Elkarte (Figura 5).

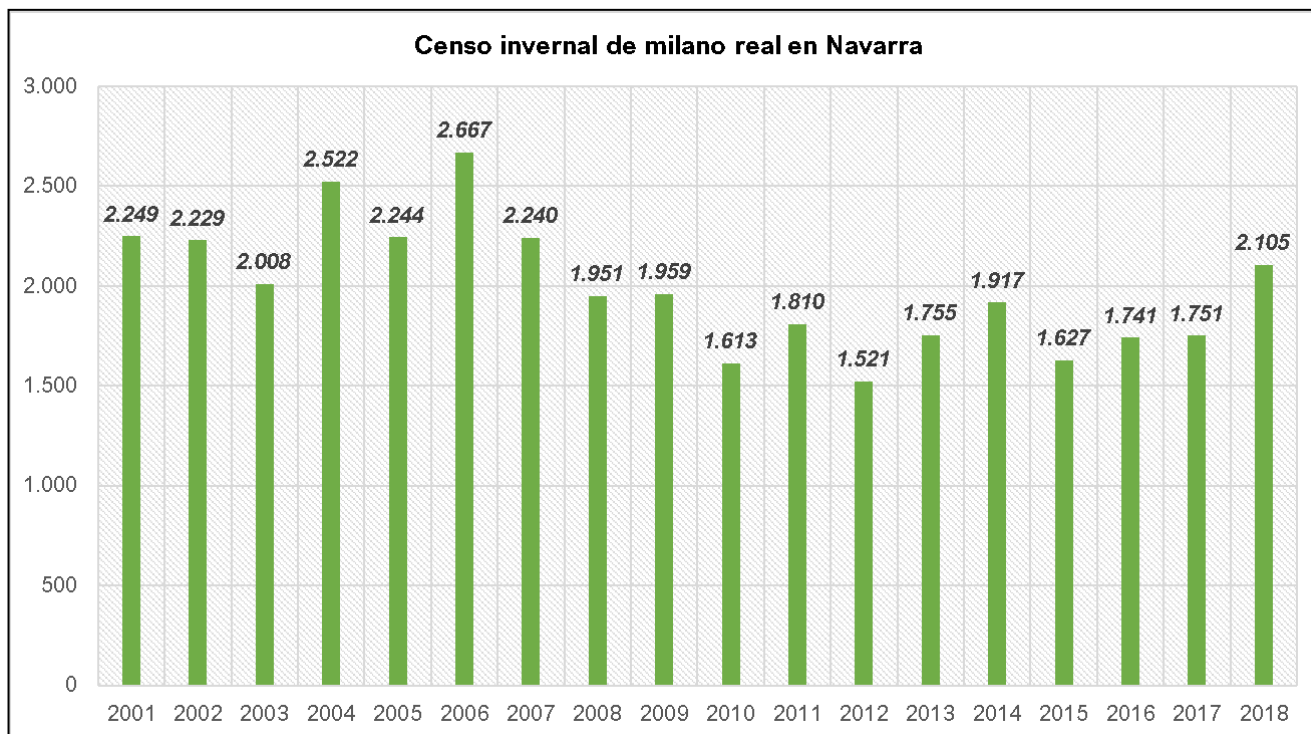


Figura 5. Censo invernial de milano real en Navarra (SEO-BirdLife) promovido por Sociedad de Ciencias Naturales Gorosti - Gorosti Natur Zientzi Elkarte. Fuente : Resumen datos del Censo invernial (SEO-BirdLife) (Deán, 2017)

Como se observa en la Figura 5, la población invernial en Navarra es estable, aunque ha sufrido fluctuaciones en las últimas décadas, probablemente causadas por las principales amenazas a las que se enfrenta esta especie (la caza ilegal y el uso de venenos en cebos, así como los casos de impactos en tendidos eléctricos y los atropellos por vehículos en las vías terrestres). Como consecuencia de estas amenazas, actualmente esta especie se encuentra catalogada a nivel nacional como “En peligro de Extinción” (Real Decreto 139/2011) y en el caso de la Comunidad Foral de Navarra está incluida en el Catálogo de Especies Amenazadas de Navarra (Decreto Foral 563/1995) como “Vulnerable”.

Su hábitat varía significativamente durante el año. En periodo reproductor, durante la primavera y verano, se sitúa en zonas forestales, así como en árboles aislados, tanto en especies de coníferas o caducifolios de tamaño medio o grande. La selección de estos hábitats de nidificación se repetirá año tras año, ubicando sus nidos en las cotas medias de los árboles u ocupando nidos de otras especies de rapaces, empleando ramas, hojas y lana de oveja. La puesta es de 2 o 3 huevos por pareja y su tiempo de incubación ronda los 28 – 32 días. Las parejas permanecen fieles a su territorio durante el periodo de reproducción, llegando a alejarse una distancia máxima contrastada al nido se estima en 15 – 20 km, según seguimiento de SEO Birds (Puente et al., 2018).

En la época de otoño e invierno se concentran en terrenos abiertos como pastizales y cultivos próximos a núcleos rurales generando dormideros comunales, generalmente en almedas o choperas.



Respecto a su alimentación es muy diversa: presas (mamíferos y aves de menor tamaño), reptiles procedentes de balsas de agua e incluso en el periodo invernal, se convierten en carroñeros y se les puede observar en vertederos y granjas, en busca de fuente de alimento potencial (Aldasoro, 1985).

1.3 Objetivos

El objetivo general de este trabajo es el análisis espacial de la distribución de la población de Milano Real en Navarra durante su etapa reproductora. Este objetivo principal se puede dividir en los siguientes objetivos específicos:

- Recopilación y procesamiento de datos de presencia de la especie obtenidos tanto en trabajo de campo como de bases de datos georreferenciadas.
- Desarrollo de variables ambientales predictoras de idoneidad de hábitat mediante SIG.
- Implementación de modelos predictivos de idoneidad del hábitat mediante métodos estadísticos (código R y Maxent), generando una hipótesis de hábitat propicio para la nidificación, mediante la respuesta a las variables predictivas.
- Generación de cartografía de hábitat potencial, que sirva como punto de partida para futuras políticas de conservación de hábitats del Milano Real.

2. MATERIALES Y MÉTODOS

En este apartado se presentarán de forma detallada el área de estudio y los métodos de análisis espacial empleados para cuantificar peso de la relación entre las variables y los nichos ecológicos. Para realizar este cálculo es preciso extraer las condiciones presentes en los espacios donde se registran ocurrencias y posteriormente relacionarlo con las variables ambientales disponibles.

2.1 Materiales

2.1.1 Área de estudio

El marco geográfico de este trabajo se sitúa en la Península Ibérica, en particular en la Comunidad Foral de Navarra, con una superficie total de 10 391 km². Este proyecto abarcará su análisis desde dos escalas geográficas distintas, ya que se han recopilado datos de observaciones, así como presencias de nidos activos de Milano Real: a nivel global del territorio foral, así como a nivel local, con área específica de 1324 km² situado en la Zona Media, ocupando las comarcas de: Pamplona, Prepirineo, Sangüesa y la Zona Media (Figura 6).

Esta área se articula fundamentalmente entorno a los valles formados por los ríos Besaire, Elorz, Irati, Areta, Sada y el río Aragón, generalmente en zonas de ribera y próximos a núcleos rurales y zonas de cultivos. Será en estas zonas de ribera donde se encuentren el grueso de nidos de esta especie, generalmente en chopos, así como en algunos árboles

dispersos, como pinos. La altitud media se encuentra entre 500- 600 m, llegando a 1100 m en zonas de pequeñas sierras, como la Sierra de Alaitz o la Sierra de Izco. La temperatura media anual es de 10 – 14°C y una pluviosidad en torno a los 700 - 1000, ya que comprende la zona climática denominada como Zona media. Otro rasgo que identifica esta zona de presencia de esta ave rapaz, es su proximidad a las vías de comunicación, debido en parte al mejor acceso a los nidos y su posible identificación, sesgo procedente del diseño de muestreo.



Figura 6. Avistamientos del Milano Real a nivel provincial y distribución de los nidos de Milano Real identificados en el área de estudio local (Navarra).

Fuente: Elaboración propia a partir de datos de presencia de SEO-BirdLife y Ornitolan S.L.

2.1.2 Datos de muestreo

Para realizar un modelado de idoneidad de hábitat (HSM, *habitat suitability mode*), se pueden emplear tres tipos de datos: ocurrencias (presencias), presencia – ausencia y abundancia de observaciones (muestreos aleatorios). En el área específica de estudio se han recopilado datos de ocurrencias in situ, a través de los trabajos de campo de la empresa Ornitolan S.L, mediante georreferenciación, con un margen de error aproximado inferior a los 3 m, debido al intento de no interferir en el desarrollo a causa de la sensibilidad de los nidos en la etapa reproductora. El análisis de esta zona concreta se debe a la experiencia en campañas anteriores, ya que el Milano Real regresa a su entorno de nidificación de años anteriores si la incubación se ha producido con éxito. Estos datos fueron recopilados entre los meses de abril y mayo de 2019 y se obtuvieron un total de 42 registros de presencia. El soporte de estos nidos se ha desarrollado predominantemente en Chopos, Fresnos y Robles

En el caso del estudio a nivel regional, los datos han sido obtenidos de los registros de SEO-BirdLife, facilitados gracias a la colaboración de Sociedad de Ciencias Naturales Gorosti, con 7.467 registros originales entre los periodos de 2012 a 2019. El método de muestreo empleado ha sido el de transectos. En este método el área de estudio en cuadrículas de 10 km, en las cuales se diseñan transectos lineales que se asignan a los diferentes colaboradores. Estos transectos lineales se desarrollarán en carreteras y caminos, recorridos tanto en vehículo como a pie, identificando y registrando tanto su posición como otros aspectos si es posible (sexo, edad estimada y actividad realizada durante la observación).

Estos datos han sido filtrados, seleccionando las observaciones comprendidas en la etapa reproductiva, de Abril a Agosto, que comprende tanto la etapa de puesta como la de nacimiento de los pollos. Los registros filtrados de conteo de Milano Real supera los 50 registros anuales y distribuidos en todo el territorio de Navarra (2621 registros totales).

Tabla 1. Resumen datos de observación del Milano Real entre 2012 – 2019 durante su etapa reproductiva. Fuente: Elaboración propia a partir de datos de presencia de SEO-BirdLife.

AÑO	AÑO	AÑO	AÑO	AÑO	AÑO	AÑO	AÑO
2012	2013	2014	2015	2016	2017	2018	2019
74	62	134	204	330	478	681	397

Como se puede observar en la Tabla 1, se ha producido un incremento de los datos de observación de Milano Real en la región, posiblemente propiciado por el incremento de medios y observadores, la mayor difusión informativa de los censos, la colaboración con otras entidades, así como a la concienciación medioambiental que ha favorecido la inscripción de voluntarios. Debido a que estos datos de observaciones se encontraban en formatos distintos, ha sido necesario su depuración y transformación de proyección, de WGS 84 a ETRS 89 UTM Zona 30N, proyección en la que se procesaran todos los datos de presencia, así como las variables ambientales predictoras.

2.1.3 Variables ambientales

Se obtendrán las variables ambientales predictivas a partir del procesamiento de capas de información procesadas mediante Sistemas de Información Geográfica de código libre Quantum Gis (QGIS) versión 2.18 (QGIS Development Team, 2019). Esta herramienta cuenta integra el software GRASS (GRASS Development Team, 2017), con el que se implementarán algunas de las metodologías clave de este proyecto. El formato de las variables será ráster (tiff), con resoluciones de 25 m y de 5m, dependiendo de la escala del área de estudio (provincial o local).

La información base para generar las variables ambientales se ha descargado del portal de la IDE, Sistema de Información Territorial de Navarra (IDENA)(Navarra and Trabajos Catastrales.,2020). IDENA es una organización corporativa dependiente Administración de la Comunidad Foral de Navarra (licencia CC BY 4.0) y del Centro de descargas del Centro Nacional de Información Geográfica (CNIG) (Instituto Geográfico Nacional (IGN),2020). Esta información ha sido procesada, generando las variables ambientales iniciales, las cuales posteriormente serán filtradas a partir de la matriz de correlación.

Las extensiones de la región de Navarra y del área de la Zona Media se corresponde con las siguientes coordenadas, extensión que ha permitido el recorte de los mapas de variables generados:

NAVARRA

xMín, yMín: 541085,553921, 4640788,8006

xMáx, yMáx:685585,553921, 4796638,8006

ÁREA ESPECÍFICA (ZONA MEDIA)

xMín, yMín:607972,50012207,4713597,50012207

xMáx, yMáx:653852,50012207,4743337,50012207

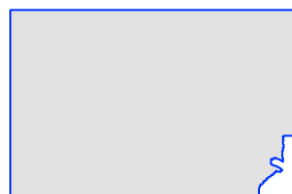


Figura 7. Extensión de las áreas de estudio: Navarra y Zona Media.

Fuente: Elaboración propia a partir de la capa Recinto de Navarra IDENA (SITNA – Gobierno de Navarra)

Respecto a la resolución espacial seleccionado a nivel de análisis espacial regional se empleará una celda de 25 m y a nivel de estudio local del área específica de la Zona Media de 5 m, debido a que se quiere analizar la influencia del tamaño de celda en el análisis espacial y al empleo de la variable “Altura de los árboles” de procedencia Lidar, del vuelo 2017.

2.1.3.1 Variables topográficas

Navarra se trata de una región con una orografía de contrastes, con dos zonas diferenciadas: la de montaña y la Ribera. La primera está formada por la región pirenaica y cantábrica,

abarcando cotas entre 600 m y 1400 m. La segunda está formada por suaves valles y grandes llanuras, destinadas principalmente a los cultivos. (Navarra, 2020) Ante este contraste de paisajes, se ha decidido incluir esta variable, ya que estas aves rapaces tienen preferencia por las llanuras y zonas de valles. Para obtener esta variable se ha partido del Modelo Digital Terrestre (MDT) con una malla de paso 5 y 25m, procedente del CNIG (MDT05 y MDT25) (Instituto Geográfico Nacional (IGN),2020). Fue necesaria la descarga de las hojas en formato ASCII que conforman el territorio foral, conformando un total de 36. Estas hojas independientes han sido fusionadas en un único ráster mediante la herramienta ráster de *Miscelánea: Combinar*. El MDT resultante (ver, p. ej. Figura 8 izquierda) se empleará para obtener otros productos como la pendiente (en grados) (Figura 8, centro) y de orientación (Figura 8, derecha). Para obtener estas variables se han aplicado las funciones de análisis de GDAL: *Pendiente* y *Orientación*, con la misma resolución espacial, dependiendo de su zona de estudio.

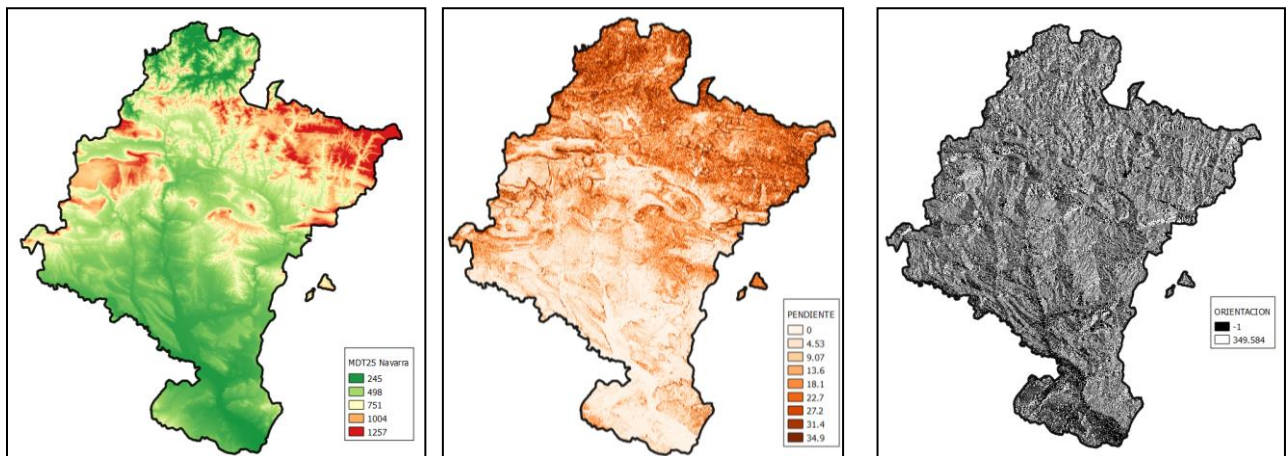


Figura 8. Mapas generados a partir del MDT 25 (1): Mapa de Pendientes (2) y Mapa de Orientaciones (3).

Fuente: Elaboración propia a partir de la capa MDT25 (CNIG)

2.1.3.2 Variables climáticas

La Comunidad Foral de Navarra se caracteriza por su diversidad climatológica, en la que se pueden clasificar 4 zonas muy diferenciadas: Zona Noroeste, Alpina, Zona media y la Ribera del Ebro. La Zona Noreste está influenciada por el Mar Cantábrico, con temperaturas medias suaves, en torno 8,5 - 14,5°C, y abundantes precipitaciones, entre 1100 y 2500 mm. La zona Alpina, se corresponde con los valles pirenaicos y prepirenaicos, territorio condicionado por su altitud y cuya influencia marítima ha perdido peso, convirtiéndose en una zona con un clima de transición, de temperaturas medias más bajas (entre 7 y 13°C) y una precipitación anual de 700 a 2200 mm. La Zona media, donde se sitúa la zona específica de estudio, se caracteriza por un clima mediterránea o cierta influencia atlántica, cuyas temperaturas medias serán más elevadas (11 – 14 °C) y cuya pluviosidad media anual se sitúa entre 450 y 1100 mm. Y finalmente la Ribera del Ebro, la zona de clima mediterráneo templado, donde se recogen los registros de más escasa precipitación de la



región (inferior a 500 mm), temperatura media templada (próxima a los 14°C) y con fuerte presencia de viento (cierzo) (Navarra,2020).

Debido a que el Milano Real busca zonas de temperaturas templadas y precipitaciones moderadas en su etapa reproductiva, se ha decidido generar capas de temperatura media y precipitación (tanto precipitación media como acumulada) de los meses que comprende esta etapa (Abril – Agosto), para cada año de análisis (2012 - 2019) (ver ejemplo del año 2018 en la Figura 9). Para ello, se han tomado los datos de los resúmenes climatológicos mensuales que se obtienen de las estaciones meteorológicas distribuidas por toda la geografía navarra (Navarra, 2020). Se ha obtenido las medias y se han generado mapas mediante el método de interpolación kriging ordinario, con un modelo de semivariograma esférico (Guerra et al., 2006) y empleando los parámetros definidos por defecto (herramienta de SAGA *Ordinary kriging*). Mediante este método geoestadístico se ha podido interpolar los valores de las estaciones, generando valores estimados donde se carecía de registros climáticos, Figura 9.

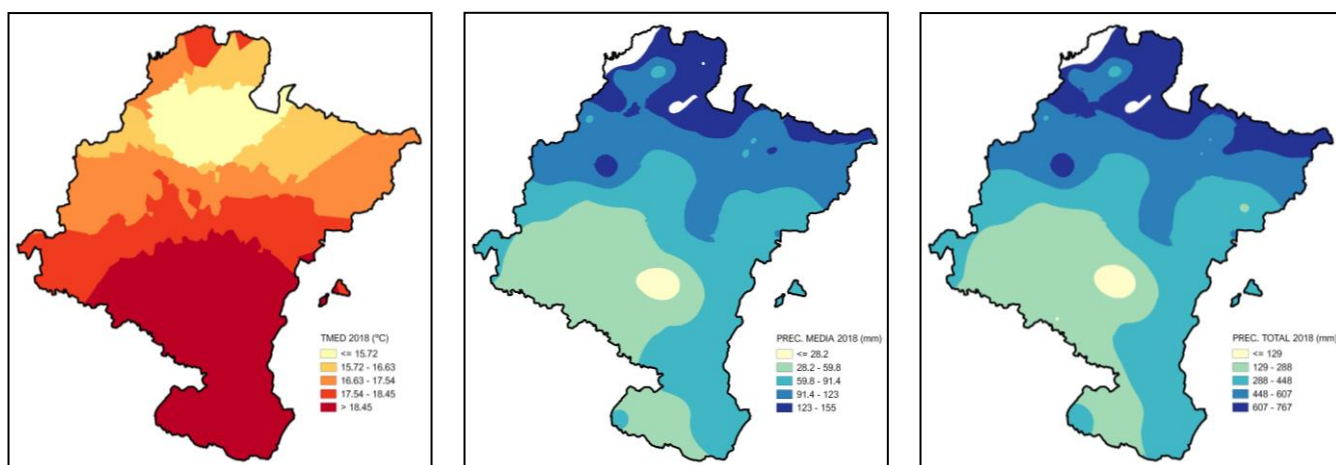


Figura 9. Mapas generados a partir de los datos de Meteo Navarra (Gobierno de Navarra). Ejemplo año 2018. Mapa de Temperatura Media (1) , Mapa de Precipitación Media (2) y Mapa de Precipitación Total (3).

Fuente: Elaboración propia a partir de los datos de resúmenes climatológicos mensuales de Meteo Navarra.

2.1.3.3 Variables de proximidad

En este apartado se tratarán las matrices de distancias, que han sido generadas mediante la herramienta de SAGA *Proximity Raster*. Para ello ha sido necesario realizar previamente la función GDAL de conversión *Rasterizar*, de las entidades que se desea procesar. Mediante este algoritmo se crean mapas de proximidad, los cuales indican la distancia desde el centro de cada pixel hasta el centro del pixel de destino (en este caso los elementos de las capas a analizar).

Dentro del conjunto de estas capas que han sido tratadas mediante estos tipos existen dos tipos según su procedencia: naturales y antrópicas. Dentro de las naturales, está ejemplificada en las hidrológicas, como fuente de alimento del ave rapaz. Se trataran tanto los cauces como las superficies de agua (marismas, embalses, lagos, etc...).

Para ello se han procesado las capas de IDENA: “Red hidrográfica. Ejes de los cauces principales de la red hidrográfica” (HIDROG_Lin_Hidrooje) y “Red hidrográfica. Superficies de agua” (HIDROG_Pol_SuperfiAgua) a escala 5000. La capa lineal de cauces principales viene jerarquizada en 7 tipologías: 1. Río principal, 2. Afluente principal, 3. Afluente secundario, 4. Afluente terciario, 5. Canales y acequias, 6. Canales y acequias subterráneos y 7. Otras corrientes y la capa poligonal de superficies de agua. Para este análisis, se ha decidido omitir las categorías de canales - acequias subterráneos y otras corrientes, ante la imposibilidad de acceso a ellas. El resultado es una capa con 5760 elementos.

Como se carece de una capa de recinto real del cauce de la red hidrográfica, se ha decidido generar un área de influencia mediante la herramienta de geoprocésamiento *Buffer de distancia fija*, a partir de la jerarquización de los cauces, otorgándole los valores reflejado en la Tabla 2, en base a la media aproximada de cauce de la región.

Tabla 2. Tabla resumen datos de ancho teórico aplicado a la red hidrográfica de Navarra. Fuente: Elaboración propia a partir de capa HIDROG_Lin_Hidrooje (SITNA – Gobierno de Navarra).

ANCHO TEÓRICO OTORGADO A LOS CAUCES DE LA RED HIDROGRÁFICA		
TIPOLOGIA DE CAUCE	ANCHO TOTAL	BUFFER
1. Río principal	10 m	5 m
2. Afluente principal	7 m	3,5 m
3. Afluente secundario	5 m	2,5 m
4. Afluente terciario	3 m	1,5 m
5. Canales y acequias	0,5 m	0,25 m

Una vez generado se unificará con la capa poligonal de superficies de agua mediante la herramienta *Unión*, la cual cuenta con 605 elementos y no precisa procesamiento previo, obteniendo una única capa, la cual será posteriormente procesada para generar su mapa ráster de proximidad (Figura 10).

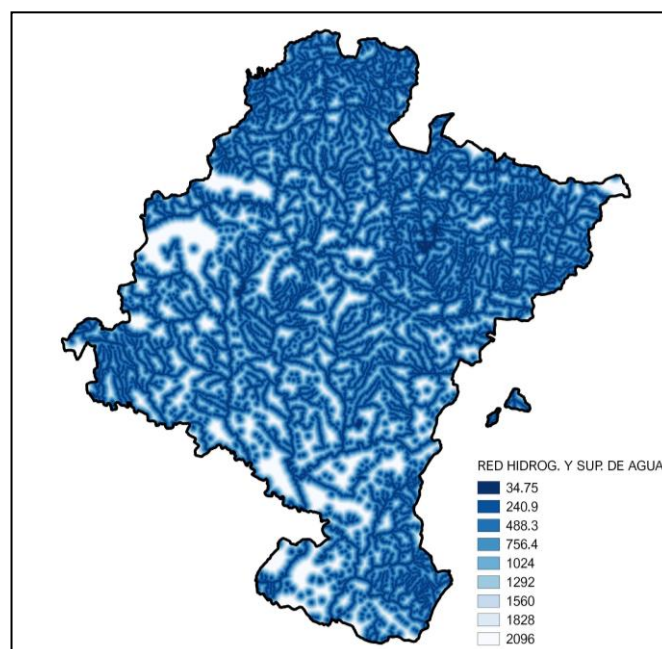


Figura 10. Mapa de proximidad de la red hidrográfica y superficie de agua de Navarra. Fuente: Elaboración propia a partir de los datos de IDENA (SITNA – Gobierno de Navarra).

Respecto a las variables de factor antrópico se encuentran la proximidad a edificaciones, vías de comunicación, explotaciones ganaderas y puntos de vertido de residuos. La información de edificaciones se extrae de la capa vectorial de “Edificaciones (visibles)” (DIRECC_Pol_Edifaltura) a escala 5000, la cual comprenden todas las edificaciones no subterráneas de Navarra. Al no necesitar ningún tipo de procesamiento previo, se generará directamente el mapa de proximidad a estas construcciones (Figura 11).

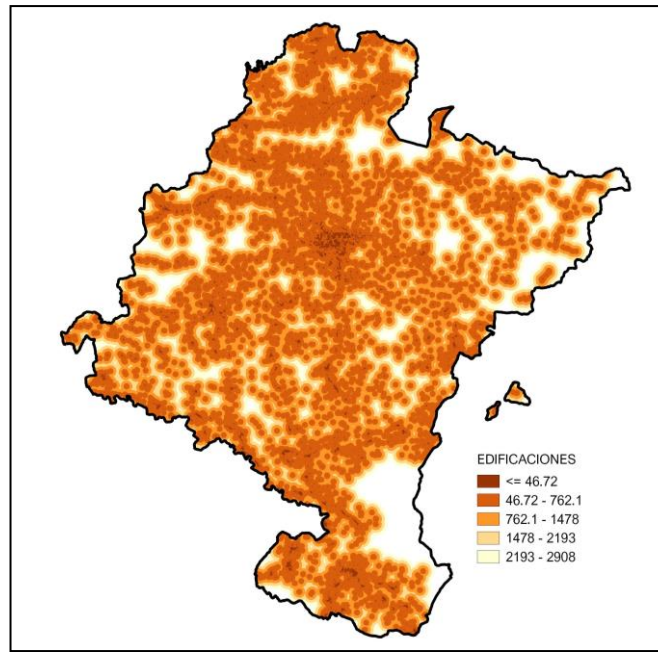


Figura 11. Mapa de proximidad a edificaciones visibles de Navarra.

Fuente: Elaboración propia a partir de los datos de IDENA (SITNA – Gobierno de Navarra).

Para realizar el análisis espacial de una de las amenazas principales, las vías de comunicación y redes de transporte se ha partido de las capas de IDENA: “Principales vías interurbanas” (INFRAE_Lin_EnlaceUrba), “Red viaria. Ejes de la red de carreteras” (INFRAE_Lin_CtraEje) y “Red de ferrocarril. Eje de la vía de la red ferroviaria” (INFRAE_Lin_FFCCViaTraz). Para las capas de vías interurbanas se generará un buffer de ancho fijo de 3 m (ya que la mayoría son vías dobles sentido), y en caso de la red ferroviaria y de las vías principales se aplicará un ancho según su clasificación (Tabla 3):

Tabla 3. Tabla resumen datos de ancho teórico aplicado a la red de carreteras y ferrocarril de Navarra. Fuente: Elaboración propia a partir de capas de Red viaria y Ferrocarril IDENA (SITNA – Gobierno de Navarra).

ANCHO TEÓRICO OTORGADO A LA RED DE CARRETERAS (CALZADA + ARCEN)		
TIPOLOGIA DE CAUCE	ANCHO TOTAL	BUFFER
1. Autopista y autovía	14 m	7 m
2. Interés de la Comunidad Foral y Interés General (Nacionales)	10 m	5 m
3. Locales y vías desdobladas	6 m	3 m
ANCHO TEÓRICO OTORGADO A LA RED FERROVIARIA (PLATAFORMA)		
TIPOLOGIA DE VÍA	ANCHO TOTAL	BUFFER
1. Única	5 m	2,5 m
2. Doble	10 m	5 m

Tras generar las áreas de influencia de estas redes, se unificarán las capas y finalmente se genera el mapa de proximidad de distancia euclidiana con la resolución adecuada a la zona de estudio (25 m y 5 m) Figura 12.

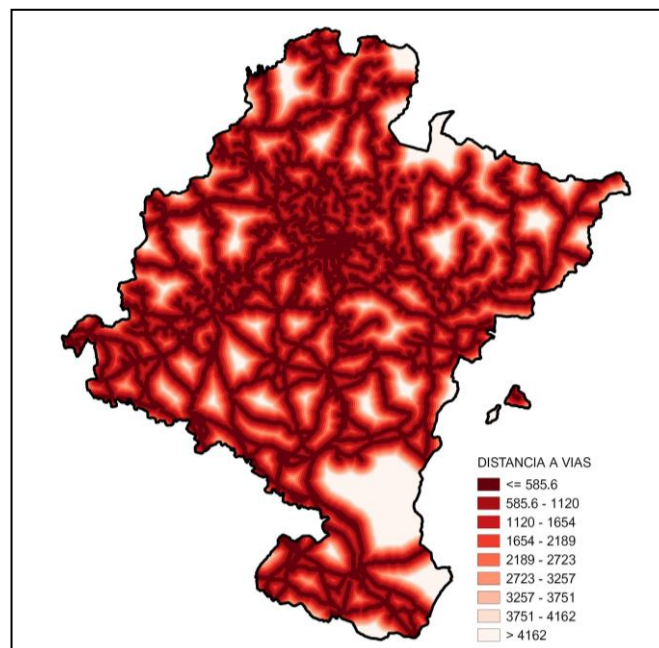


Figura 12. Mapa de proximidad a vías de comunicación y redes de transporte de Navarra. Fuente: Elaboración propia a partir de los datos de IDENA (SITNA – Gobierno de Navarra).

Otra de las variables susceptibles de demostrar la relación presencia y territorio, son las explotaciones ganaderas, ya que pueden ser fuente de alimento del Milano Real. Esta capa actualmente es de acceso libre, pero en el inicio del procesamiento de este trabajo era confidencial, pero gracias al Departamento de Desarrollo Rural, Medio Ambiente y Administración Local y a Gestión Ambiental de Navarra S.A se ha podido obtener una capa actualizada (última modificación 25/06/2019) con la información requerida: georreferenciación de la explotación y especie. En ella constan diferentes especies: aves, bovino, caprino, cerdos, caprino, etc., elementos susceptibles de ser fuente de alimento, a

los que se aplicará el algoritmo de SAGA *Proximity Raster* para generar el mapa de proximidad (Figura 13).

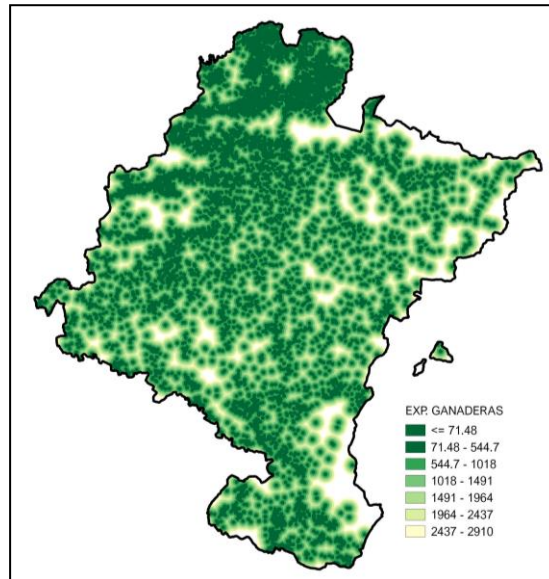


Figura 13. Mapa de proximidad de explotaciones ganaderas de Navarra.

Fuente: Elaboración propia a partir de los datos de IDENA (SITNA – Gobierno de Navarra).

Como se ha mencionado previamente, es común poder observar a esta especie en los vertederos buscando alimento, lo cual es susceptible de ser una variable explicativa de su presencia. Para ello se partirá de la capa de IDENA “Residuos y espacios degradados. Vertederos de residuos urbanos, inertes e industriales no peligrosos” (POLUCI_Sym_Vertederos), con las localizaciones de los puntos de vertidos controlados, entre los que se encuentran: residuos inertes, residuos urbanos e industriales no peligrosos. Esta capa no requiere ningún preprocesamiento por lo que directamente se le aplica el logaritmo de generación de mapa ráster de proximidad (Figura 14).

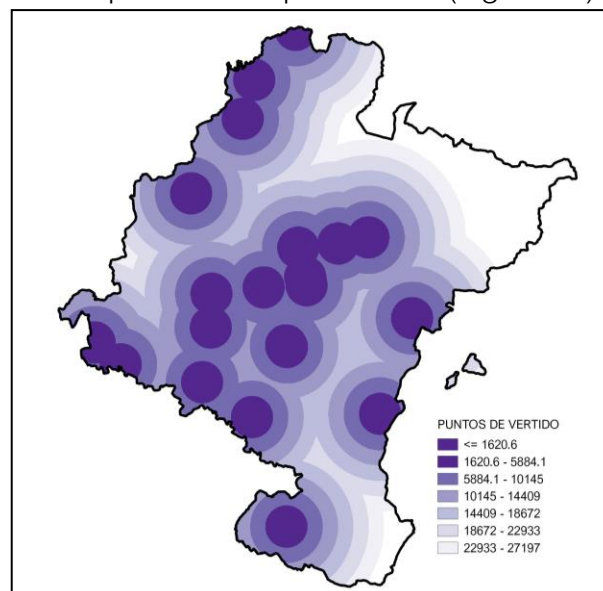


Figura 14. Mapa de proximidad de punto de vertidos residuales de Navarra.

Fuente: Elaboración propia a partir de los datos de IDENA (SITNA – Gobierno de Navarra).

Y para cerrar este apartado, se generarán mapas de proximidad a partir de las capas de IDENA “Vías pecuarias. Ejes de las vías pecuarias actualizadas” (GANADE_Lin_VPPAeje) y “Pistas forestales principales. Ejes” (FOREST_Lin_PistasForP), a las cuales se les aplicará un buffer fijo de 1,5 y 1 m respectivamente y posteriormente serán unificadas, para finalmente generar el mapa de proximidad (Figura 15).

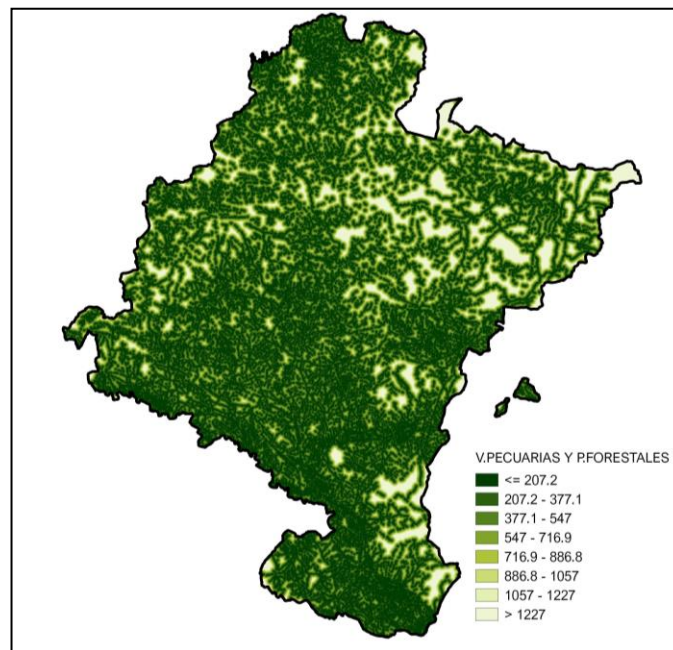


Figura 15. Mapa de proximidad de vías pecuarias y pistas forestales de Navarra.

Fuente: Elaboración propia a partir de los datos de IDENA (SITNA – Gobierno de Navarra).

2.1.3.4 Variables relacionadas con la vegetación

Relacionada con las fuentes de alimento, se encuentra el “Mapa de Usos y Aprovechamientos 2012” (OCUPAC_Pol_MCA_VE2012), capa obtenida de IDENA (SITNA – Gobierno de Navarra), la cual se encuentra repartida en 7 hojas: Noroccidental, Pirineos, Cuenca de Pamplona, Tierra de Estella, Navarra Media, Ribera Alta y Ribera Baja. Se trata de un mapa de usos del suelo producido en 1999, el cual ha sido actualizado en 2012, año del que consta la última actualización.

Vista la división de los cultivos en distintas capas, el primer proceso que se realizará en QGis será agrupar en un único archivo las capas mediante la herramienta *Unión*. Otro aspecto que se modificará será la creación de un nuevo campo de texto con el fin de agrupar en categorías los aprovechamientos, mediante la *Calculadora de Campo*, unificando el campo [GRUPO] y [COBERTURAP] y disolviéndolo posteriormente. Tras este procesamiento se procederá a rasterizar la capa vectorial mediante la herramienta GDAL de conversión *Rasterizar* (Figura 16).

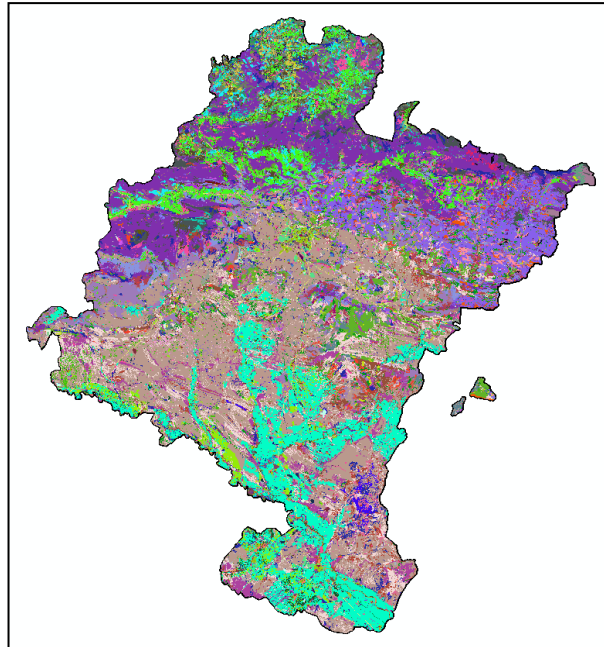


Figura 16. Mapa de Usos y Aprovechamientos 2012 de Navarra.

Fuente: Elaboración propia a partir de los datos de IDENA (SITNA – Gobierno de Navarra).

Y para concluir, se ha generado un Modelo Digital de Alturas de la Vegetación (MDHV), a partir de los datos Lidar 2017 (malla de 50 cm) procedentes de IDENA, calculando diferencia de superficie entre el Modelo Digital de Superficie (MDS) y el Modelo Digital de Terreno (MDT). Posteriormente se ha realizado un re-muestreo a la resolución empleada en el área de análisis (5m) Figura 17.



Figura 17. Modelo Digital de Alturas de la Vegetación (MDHV) del área de estudio.

Fuente: Elaboración propia a partir los datos Lidar 2017 de IDENA (SITNA – Gobierno de Navarra).

2.1.3.5 Correlación de las variables

Una vez generadas las variables ambientales iniciales, es necesario hacer una preselección de las mismas, mediante una matriz de correlación. Una matriz de correlación es matriz cuadrada constituida por los coeficientes de correlación de cada pareja de variables. Se tratará de evitar la correlación entre las variables predictivas, ya que generan errores en gran parte de los métodos de modelización. Además se tiene en cuenta la ratio recomendado entre variables y observaciones para el modelado de distribución es 1:10 (Guisan, A. and Zimmermann, 2000). Esto establece, por cada variable predictiva debe de haber al menos por 10 valores de presencia.

Tabla 4. Matriz de correlación de las variables ambientales predictoras generada con el algoritmo "r.covar". Fuente: Elaboración propia a partir de capas de Red viaria y Ferrocarril IDENA (SITNA – Gobierno de Navarra).

N= 16626114	MDT25	PEND25	ORIENT25	TMED25	PMED	PTOT	HIDRO25	EDIF25	REDVIA25	EXPGAN25	DISVERT	DISVIAECPF	USOS25
MDT25	1.000.000	0.471947	0.100281	-0.572112	0.465655	0.480507	0.093194	0.209398	0.103502	0.280729	0.545737	0.379301	0.229955
PEND25	0.471947	1.000.000	0.100204	-0.517938	0.559591	0.560971	-0.139074	0.079089	0.070974	0.070623	0.335332	0.250306	0.286576
ORIENT25	0.100281	0.100204	1.000.000	-0.103490	0.087186	0.084993	-0.031006	0.046535	0.034478	0.028593	0.042709	0.049718	0.052766
TMED25	-0.572112	-0.517938	-0.103490	1.000.000	-0.717780	-0.731234	0.168927	0.130152	0.129567	0.101103	-0.280364	-0.193329	-0.334111
PMED	0.465655	0.559591	0.087186	-0.717780	1.000.000	0.991635	-0.158233	0.029518	0.080942	-0.012920	0.400410	0.202144	0.308015
PTOT	0.480507	0.560971	0.084993	-0.731234	0.991635	1.000.000	-0.160629	0.027800	0.076168	-0.005850	0.406742	0.202143	0.309732
HIDRO25	0.093194	-0.139074	-0.031006	0.168927	-0.158233	-0.160629	1.000.000	0.062952	0.050555	0.099801	-0.081418	0.095237	-0.028437
EDIF25	0.209398	0.079089	0.046535	0.130152	0.029518	0.027800	0.062952	1.000.000	0.628749	0.602313	0.251650	0.292587	0.019170
REDVIA25	0.103502	0.070974	0.034478	0.129567	0.080942	0.076168	0.050555	0.628749	1.000.000	0.414541	0.203457	0.256279	0.043165
EXPGAN25	0.280729	0.070623	0.028593	0.101103	-0.012920	-0.005850	0.099801	0.602313	0.414541	1.000.000	0.260163	0.225509	-0.000989
DISVERT	0.545737	0.335332	0.042709	-0.280364	0.400410	0.406742	-0.081418	0.251650	0.203457	0.260163	1.000.000	0.321889	0.073903
DISVIAECPF	0.379301	0.250306	0.049718	-0.193329	0.202144	0.202143	0.095237	0.292587	0.256279	0.225509	0.321889	1.000.000	0.123761
USOS25	0.229955	0.286576	0.052766	-0.334111	0.308015	0.309732	-0.028437	0.019170	0.043165	-0.000989	0.073903	0.123761	1.000.000

Mediante QGIS se utilizará la función GRASS (GRASS Development Team, 2017) *r.covar*, la cual genera una matriz de correlación en formato html y txt. Como se puede observar en la Tabla 4, los resultados son favorables, existiendo una correlación muy elevada (superior al umbral de r 0,8 recomendado) únicamente entre las variables de precipitación (con un valor estimado de 0,991635). Se descartará una de ellas en el modelado para evitar problemas con la multicolinealidad (Precipitación Media). También existe una correlación moderada entre la capa de Edificaciones – Hidrología, Edificaciones – Red de Vías y Edificaciones – Explotaciones Ganaderas.

Todas las variables serán reproyectadas a coordenadas geográficas y exportadas en formato ráster ASCII, debido a las exigencias del paquete de análisis estadístico empleado en R.

2.2 Métodos

En este apartado se desarrollarán modelos que relacionen las variables predictoras disponibles con la presencia del Milano Real en el territorio de estudio.

2.2.1 Modelos de simulación

Existen numerosos métodos analíticos de distribución de especies, los cuales han sido integrados en diversas herramientas, tanto en Sistemas de Información Geográfica como puede ser el caso del incipiente plugging "QSDM" (<https://plugins.qgis.org/plugins/QSDM/>) o programas propios como MaxEnt (Phillips et al., 2006). Gran parte de los algoritmos que emplean estos métodos actualmente se encuentran integrados en paquetes de R, como



puede ser el caso de *sdm* (Naimi and Araujo, 2019), *biomod2* (Wilfried Thuiller, Damien Georges, Robin Engler, 2019), *dismo* (Wilfried Thuiller, Damien Georges, Robin Engler, 2019) o *SDMTools* (Vanderwal J., Falconi L., Januchowski S., 2012). En la Tabla 5 se presenta un resumen de los distintos métodos disponibles:

Tabla 5. Resumen de los métodos más estandarizados actualmente. Fuente: Elaboración propia a partir del ejemplo (Thuiller and Münkemüller, 2010) y R CRAN (<https://cran.r-project.org>)

HERRAMIENTA	AUTORES	MÉTODOS (*)
BIOCLIM	Busby (1991)	CE
BIOMAPPER	Hirzel et al. (2002)	ENFA
BIOMOD/ BIOMOD2	Thuiller (2003); Thuiller et al. (2009)	ANN, CE, CTA, GAM, GBN, GLM, MARS, MDA, ME y RF
DIVA	Hijmans et al. (2001)	CE
DOMAIN	Carpenter et al. (1993)	CE
GARP	Stockwell and Peters (1999)	GA
GDM	Ferrier et al. (2007)	GDM
GRASP	Lehmann et al. (2003)	GAM
MAXENT	Phillips et al. (2006)	ME
OPEN MODELLER	Sutton et al. (2007)	ANN, GARP, SVM, y CE
SDM	Naimi and Araujo (2019)	ANN, CE, CTA, GAM, GBN, GLM, MARS, MDA, ME y RF
SPECIES	Pearson et al. (2002)	ANN

* ANN: Redes neuronales artificiales; CE: Modelos climáticos de envoltura; CART: Árboles de clasificación y regresión; ENFA: Análisis factorial del nicho ecológico; GA: Algoritmo genérico (búsqueda); GAM: Modelos Aditivos Generalizados; GBN: Boosting machine; GDM, Modelo Generalizado de Disimilitud GLM: Modelos lineales generalizados; MARS: Splines de regresión adaptativa multivariante; MDA: Análisis discriminante; ME: Máxima entropía; SVM: Máquinas de vectores de soporte.

Ante la diversidad de métodos, es necesario establecer una serie de pasos clave para la generación de HSM, para establecer el marco de selección de modelo de análisis espacial:

PASOS FUNDAMENTALES PARA LA GENERACIÓN DE HSM

1. Formulación del modelo conceptual (definición del objeto de estudio)
2. Preparación de datos (diseño del muestreo, variables ambientales y escala del proyecto)
3. Selección del modelo y calibración (ajuste a partir del análisis de multicolinealidad y posible sobreajuste)
4. Evaluación del modelo (validación)
5. Reproducción de predicciones espaciales (mapeo de idoneidad del hábitat)

La identificación del método más adecuado es uno de los aspectos más problemáticos, ya que todos presentan fortalezas y debilidades, por lo que se ha decidido emplear tanto el modelo de máxima entropía MaxEnt como el paquete *biomod2*. Este último permite la

comparación simultánea y conjunta de distintos métodos para describir y modelar la interrelación entre especie y su entorno (Wilfried Thuiller, Damien Georges, Robin Engler, 2019).

Cabe mencionar que muchos de estos métodos generan modelados de distribución de especies a partir de datos únicamente de presencia (*Bioclim* o *Domain*), de datos presencia – ausencia (o blackground, como MaxEnt o CART) y otros mediante la combinación de diferentes modelos.

En el caso de los algoritmos predictivos aplicados en la herramienta *biomod2* son los siguientes (Thiller et al., 2019, Guisan et al., 2017):

- GLM: Modelos lineales generalizados (glm). Esta técnica lineal iterativa ponderada busca definir los parámetros que permiten relacionar mediante un predictor lineal la distribución de las variables dependientes con las variables independientes.
- GAM: Modelos Aditivos Generalizados (gam, gam o bam). Consiste en un método lineal que combina distintas funciones de suavizado paramétrico con el fin de ajustar las variables que no presentan una distribución normalizada.
- GBM: Árboles de regresión y técnicas de remuestreo 'boosted regression trees' (gbm). Consiste en una técnica de autoaprendizaje que combina diversos métodos de regresión y clasificación, tratando de generar un nuevo clasificador más robusto.
- CTA: Árboles de clasificación (rpart). Permite generar modelos de clasificación o regresión de estructura generalizada, realizando subdivisiones de los datos, a partir de una o varias variables según su error de clasificación.
- ANN: Redes neuronales artificiales (nnet). Consiste en un método computacional de redes de neuronas artificiales interconectadas, con el objeto de clasificar y predecir a través de sus conexiones.
- SRE: Modelos climáticos de envoltura (similar estructura que BIOCLIM). Estima una envoltura de rango a partir de los perceptibles extremos de cada variable.
- FDA: Análisis discriminante flexible (fda). Se trata de un método de regresión que permite el escalamiento óptimo de las categorías a partir de un conjunto de variables a partir de sus características.
- RF: Random Forest (randomForest). Es uno de los métodos de clasificación más preciso, basado en la combinación de árboles de clasificación (bootstrap).
- MARS: Splines de regresión adaptativa multivariante (earth). Es una técnica de regresión más flexible que GML, seleccionando variables entre cada par de nudos y evitando el sobreajuste.
- MAXENT.Phillips: Máxima Entropía. Es un método de inteligencia artificial, el cual se basa en máxima entropía (uniformidad), generando distribuciones en localizaciones lo más semejantes posibles de las presencias existentes.

En el caso del paquete descrito *Biomod2*, en este trabajo se ha decidido emplear 4 métodos, 3 lineales y 1 autoaprendizaje, en particular los siguientes métodos: GLM, GAM, GBM y RF siguiendo los resultados óptimos presentados en otros estudios similares (Guisan et al., 2017). Adicionalmente, como ya se ha mencionado anteriormente, se ha decidido analizar estas variables y presencias con MaxEnt, desde su interface propia, de

forma que se pueda controlar directamente los parámetros y la respuesta generada, siendo este uno de los modelos más estandarizados en el caso de modelado de datos únicamente de presencias (Joppa et al., 2013).

2.2.2 Formatos de los modelos de HSM

2.2.2.1 Formato del paquete Biomod2

Biomod2 es paquete dedicado a técnicas de modelización en lenguaje estadístico R. Este software permite la ejecución de varios modelos, así como generar ensamblajes de modelos para reducir las incertidumbres. Este programa incluye distintos procedimientos de evaluación y la posibilidad de análisis de un número ilimitado de especies.

De igual modo que el caso de MaxEnt, para implementar los modelos se requiere dos tipos de variables: las variables de respuesta, ya sea de datos de presencia o ausencia y las variables explicativas, datos ambientales. Los datos de presencia pueden corresponder a un número ilimitado de especies, siendo la única limitación la capacidad operativa de la CPU. Estos datos deberán estar en formato csv, con la estructura muestra, longitud y latitud. La muestra se identificará de modo binario, correspondiendo el 1 y el 0 a la presencia y ausencia de la especie respectivamente. Se reserva el valor NA para las observaciones indeterminadas (generando un *SpatialPointData* Frame o *SpatialPoints* si se trabaja solo con presencias).

En el caso de las variables predictoras se presentará en modo raster ASCII, conservando todas las mismas propiedades de extensión y proyección. En este caso se agruparán todas las variables en un único conjunto (*expl.var*) para poder ser procesadas mediante la función de homogenización de datos de entrada al modelo, *BIOMOD_FormatingData*.

2.2.2.2 Formato del programa MaxEnt

MaxEnt es un programa de código abierto implementado mediante Java, cuya instalación se realiza a través de la aplicación descargada mediante el enlace facilitado por la Universidad de Princeton (EEUU) (<http://www.cs.princeton.edu/~schapire/maxent/>), entidad a la que pertenecen sus desarrolladores. Mediante el archivo ejecutable, se puede observar una interface intuitiva y con varias herramientas de configuración del modelo, para poder ajustarlo a nuestro estudio.

Para desarrollar el análisis mediante MaxEnt, se requieren 2 elementos: datos de presencia y variables predictoras ambientales. Estos datos de entrada tienen un protocolo fijado desde Maxent: los datos de entrada (*Samples*) deben estar configurados en un formato delimitado por comas .csv y estructurado en tres columnas: especie, longitud y latitud (coordenadas en formato de Grados decimales).

Respecto a las variables ambientales (*Environmental layers*), se recomienda el uso de archivos raster en ASCII (extensión .asc), las cuales deben tener la misma extensión y proyección, correspondiéndose con la proyección de datos de presencia. Una vez seleccionado el directorio de variables (*Browse*), cargará todas las variables almacenadas en él.

2.2.3 Métodos de evaluación del modelo

Una vez analizada la calidad de los datos de muestreo así como la evaluación de las variables predictivas evitando su colinealidad, se generarán los primeros modelados de HSM mediante los algoritmos mencionados. Esta respuesta deberá ser evaluada, buscando reducir al máximo el error generado. Por tanto, la evaluación del modelo se centrará en tasar la robustez del modelo generado, así como su capacidad de predecir sobre datos independientes y su realismo aplicable a la zona de estudio. De esta forma, se engloban los procedimientos de ensayos y validación, los cuales se evaluarán mediante los siguientes elementos:

- Método de validación cruzada: Es una técnica de remuestreo que permite garantizar la independencia de los datos de entrenamiento y de test. Existen dos variedades de validación cruzada: *Jackknife*, mediante el procedimiento de dejar fuera un objeto en cada iteración y la validación cruzada clásica *k-fold*, enfocado a grandes muestras de datos, ya que subdivide los datos en *k* valores (test) y el resto de entrenamiento.
- Matriz de confusión: *sensitivity*, *specificity*, CCR, Kappa. Mediante este análisis se podrán analizar las presencias y ausencias erróneas, tanto por omisión como por comisión.
- ROC (AUC). En relación al caso anterior se encuentra la curva ROC, que consiste en una representación gráfica de la relación entre sensibilidad y la especificidad de la clasificación. La calidad del modelo se observa en el área bajo la curva (AUC), cuyos valores próximos a 1 muestran un correcto ajuste del modelo y a 0,5 la dificultad de distinguir los valores reales de las ausencias, que equivaldría a valores generados al azar.

3. RESULTADOS

3.1 Ajuste de modelos

3.1.1 Calibración

Para generar modelos mediante el paquete de Biomod2 existen funciones claves, que definirán la calidad del modelo (estos aspectos pueden ser revisados en el anexo):

- *BIOMOD_FormatingData*: Mediante esta orden se reorganizarán los datos de entrada, mediante la estructura de Biomod2. Para ello es necesario transformar los datos de presencia a coordenadas geográficas y respetar la estructura de columnas con presencia-ausencia (valor 1 presencia y valor 0 ausencia), X ,Y. En este caso no hay datos de ausencia, por lo tanto, se generarán pseudo-ausencias para evitar el sobreajuste del modelo, con un total de 200 pseudo-ausencias generadas de modo aleatorio (*random*) en 2 repeticiones.
- *BIOMOD_ModelingOptions*: En esta parte del código se configurarán los argumentos de las técnicas empleadas en el modelado. Como se trata de un trabajo de acercamiento a técnicas de análisis espacial, se han configurado por defecto (Georges and Thuiller, 2013), cuyos aspectos más destacados son:

- Método GML: Método 'quadratic', interacción con las variables 0, método de evaluación 'AIC' y distribución de los errores en modo familia binomial.
 - Método GBM: Distribución 'bernoulli', número de árboles= 2500 e iteraciones 7
 - Método GAM: k =-1 y distribución de los errores en modo familia binomial.
 - Método RF: número de árboles 500 y sin máximo de nudos definidos.
- *BIOMOD_Modeling*: Función que permite la ejecución del modelo, su calibración y evaluación, empleando la técnica de la técnica de *k-fold*, con los datos de muestreo. En este caso se realizará una única evaluación con el 80% de los datos utilizados para calibrar y el resto para testeo. También se realizarán 3 permutaciones para estimar la importancia de las variables. Estos resultados serán evaluados y se mostrarán gráficamente.

Adicionalmente se mostrará la respuesta media de las variables a cada método, así como en una tabla de resultados.

- *BIOMOD_EnsembleModeling*: Mediante esta función se combinarán los modelos creando un conjunto de predicciones y se evalúa la precisión de los métodos mediante TSS (*True Skill Statistic*)(Peirce, 1884). Se filtrarán los valores inferiores a 0,8 y para ello se emplearán los métodos 'KAPPA', 'TSS' y 'ROC' (Figura 18). Como se puede observar en esta figura, se mostrará gráficamente la respuesta estadística de las diferentes técnicas de modelado, analizando su nivel de precisión predictiva, cuyos valores óptimos serán los más próximos a 1. En este caso, para el modelo del año 2015, los métodos empleados con mejores resultados son catalogados como *machine learning*, los árboles de decisión (cruz de tono morado) y GBM (cruz tono verde), seguido de los métodos lineales tradicionales (GLM y GAM).
- *BIOMOD_Projection*: Permite proyectar los modelos desarrollados, que se pueden aplicar adicionalmente en otras áreas y en otros contextos espacio temporales.
- *BIOMOD_EnsembleForecasting*: Para generar modelos individuales y modelos de conjunto de las proyecciones de las especies. En este proyecto se presentarán los resultados individuales de los distintos métodos, así como en conjunto, para poder si existe una aportación adicional de información al implementar varios métodos.

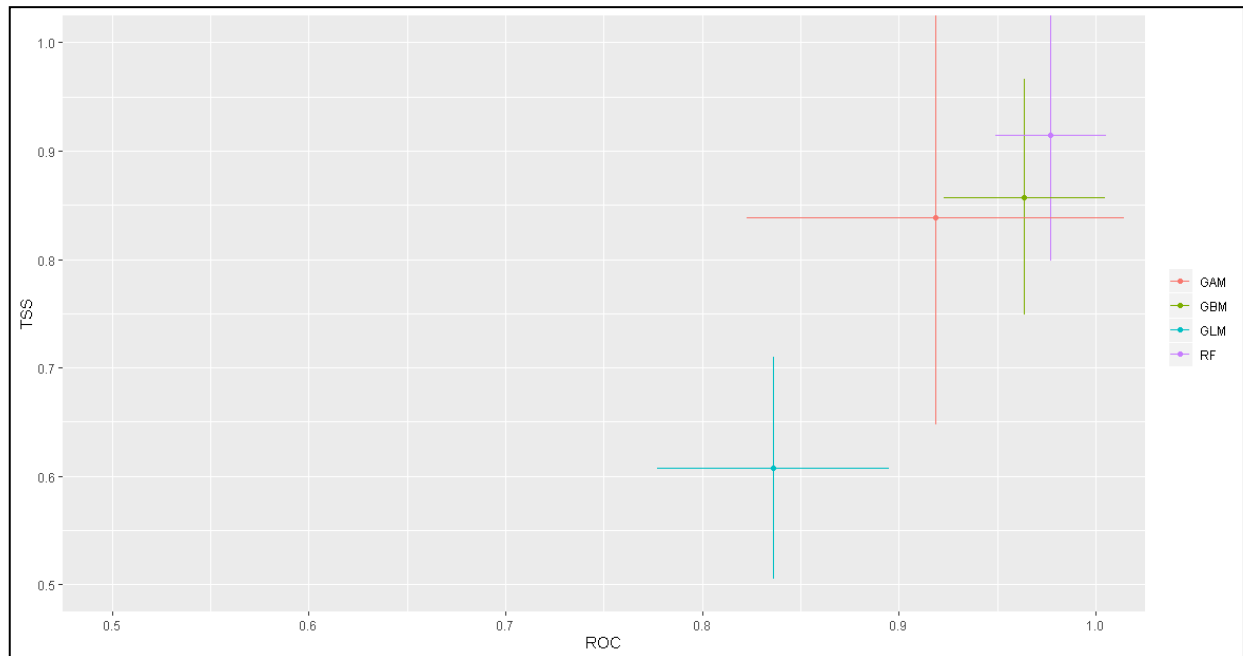


Figura 18. Representación gráfica de la evaluación de los algoritmos empleados en el modelo del año 2015. Fuente: Elaboración propia mediante la plataforma R-Studio aplicando el paquete Biomod2.

En el caso de MaxEnt, tras realizar varias pruebas siguiendo las recomendaciones presentadas por otros autores (Merow et al., 2013) entre sus distintos métodos (logístico o acumulativo, número de iteraciones y empleo de datos de testeo), se ha decidido que la mejor configuración sea aplicando un límite de 1000 iteraciones como máximo y 0,00001 de convergencia. Se emplearán datos de entrenamiento y test (ratio 80-20), los cuales serán testados con datos aleatorio. Respecto a las variables ambientales empleadas son las mismas que en el caso de Biomod2, que han sido evaluadas generando un AUC próximo a 0,95 en la mayoría de casos (Figura 19).

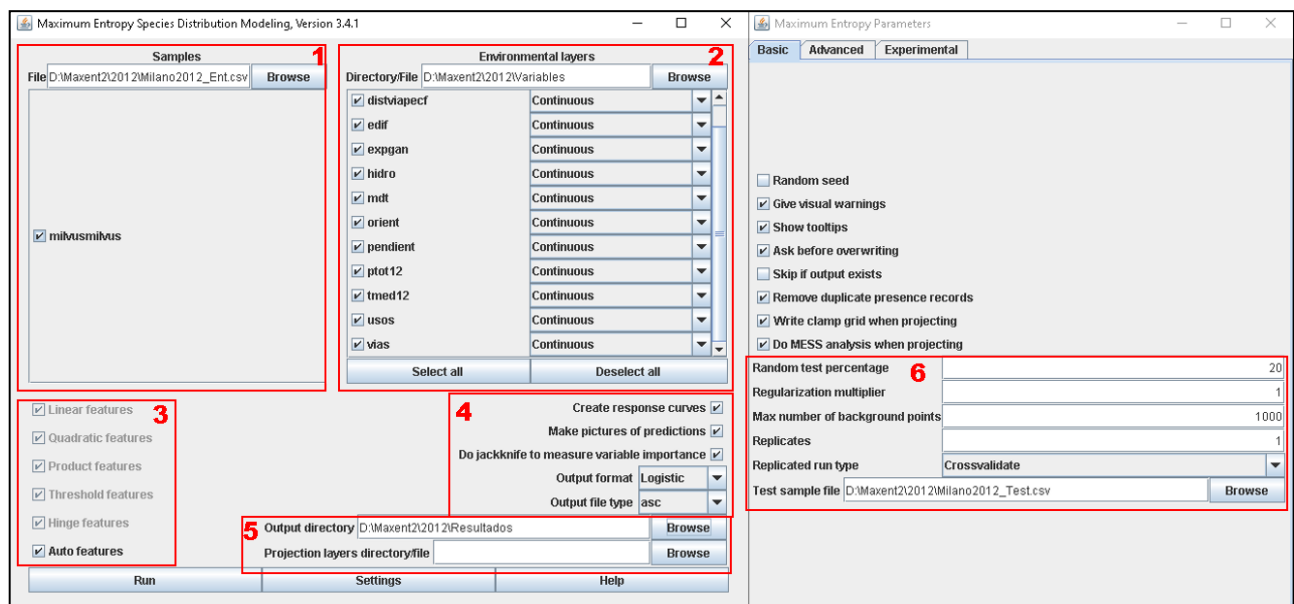


Figura 19. Calibración de los principales parámetros de MaxEnt(1 entrada de datos, 2 variables predictoras, 3 métodos, 4 elementos de salida, 5 rutas y 6 otros parámetros de configuración del modelo) . Fuente: Elaboración propia mediante la plataforma MaxEnt.



En la Figura 19, se puede observar la configuración aplicada en MaxEnt en detalle. En el elemento 1 se incluye los datos de entrada formato delimitado por comas .csv, con el formato especificado por MaxEnt (especie, X,Y). En el elemento 2, se ha incluido el directorio donde se engloban las variables ambientales que se desean analizar en formato ráster ASCII.

En el elemento 3 se incluyen los métodos empleados para transformar las variables combinando distintos métodos. En este caso se han incluido todos: 'linear', 'quadratic', 'product', 'threshold' y 'hinge'. El elemento 4 y 5 hace referencia al archivo de salida (ejemplo en la Figura 20), el cual contendrá la curva ROC y el resumen del *jackknife* con la importancia de las variables en el modelo. El resultado se exportará en un archivo .html en el directorio de salida especificado. Finalmente, en el elemento 6 se muestran otros aspectos básicos de la configuración del modelo, en que se especificará el máximo de iteraciones para generar el background, así como el porcentaje de testeo y el multiplicador de regulación. La configuración se ha ajustado mediante prueba y error, con el fin de lograr la mejor capacidad predictiva, generando distintos modelos con varios valores, descartando otras configuraciones debido al empeoramiento de resultados (valores de AUC rondando el 0,84). Respecto al método de evaluación se ha decidido emplear la validación cruzada, ya que con 'Bootstrap' no se ha observado variación significativa de los resultados. En los parámetros avanzados se ha deseleccionado la opción de 'do clamping', ya que generaba saltos muy bruscos que afectaban sustancialmente al modelo.

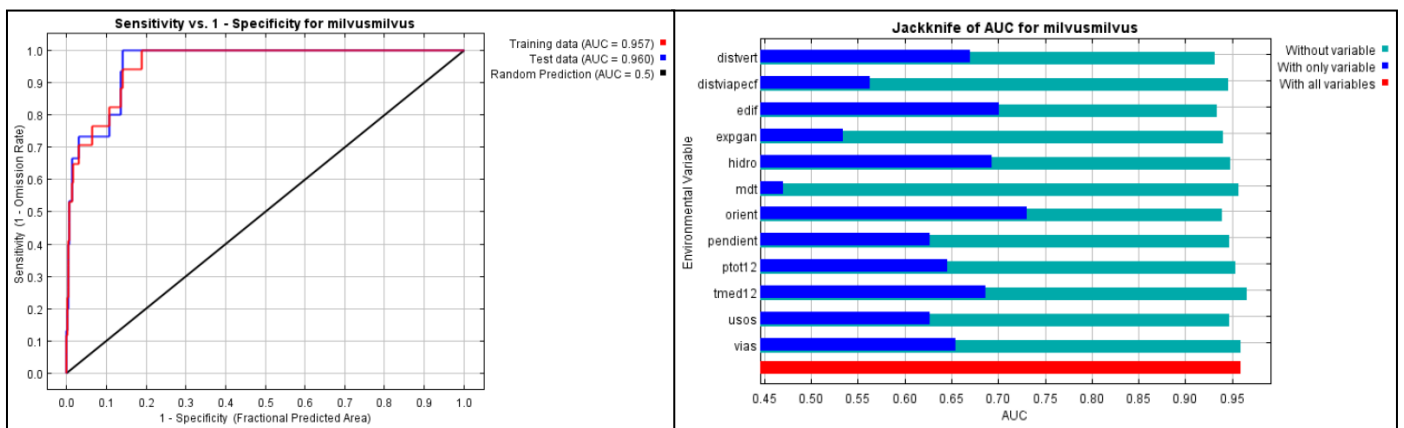


Figura 20. Representación gráfica de los evaluadores del modelo y de la respuesta de las variables ambientales. Fuente: Elaboración propia mediante la plataforma MaxEnt.

3.1.2 Evaluación

3.1.2.1 Evaluación de los resultados de Biomod2

La primera parte del análisis se centrará en el comportamiento del Milano Real en el periodo de reproducción a partir de los datos recogidos entre 2012 – 2019, los cuales se analizarán de forma global.

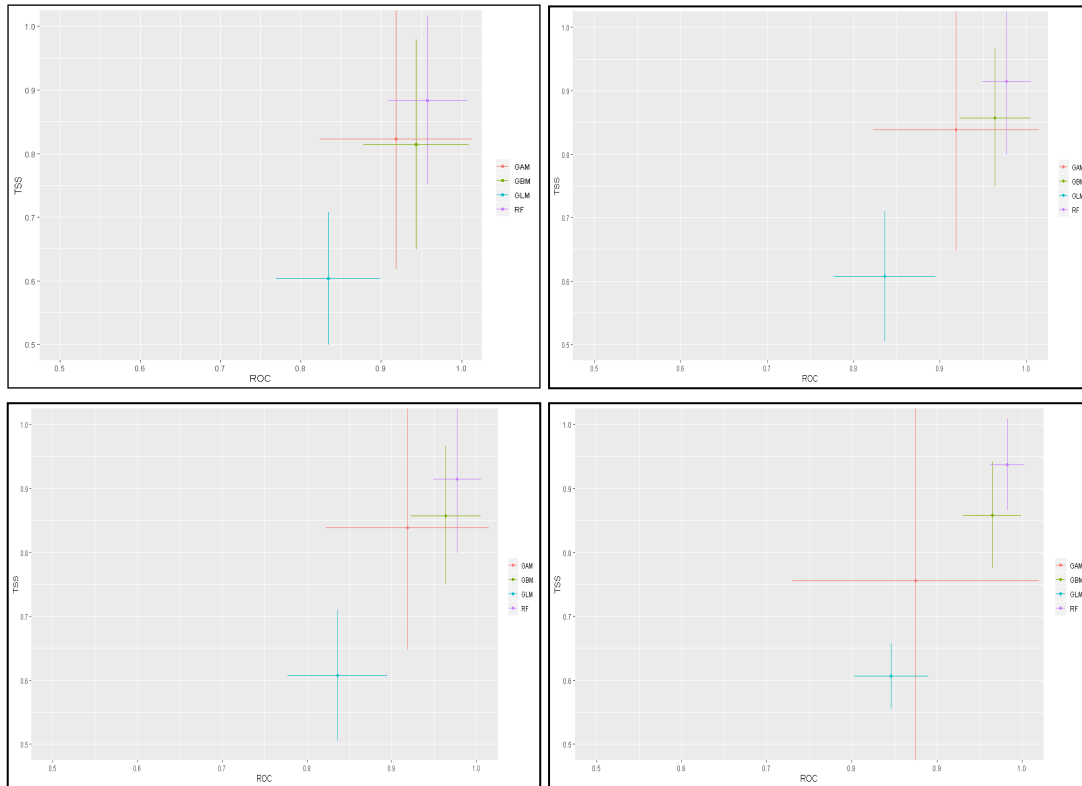


Figura 21. Comparativa de resultados de la evaluación de los métodos mediante los índices ROC y TSS.

Como se puede observar en la Figura 21, en la que se muestra una comparativa entre los años 2013 – 2016 de los resultados de la evaluación. Estos resultados demuestran que por norma general el algoritmo Random Forest alcanza unos niveles superiores a 0,9 de ROC y TSS, convirtiéndose en el mejor algoritmo de selección de variables ambientales, mientras que el método GLM es el peor con valores próximos a 0,7. Otro rasgo significativo es la semejanza de valores obtenidos entre los métodos GAM y GBM, respuesta que se asocia al tipo de metodología que aplican ambos (regresión). Adicionalmente se ha obtenido el aporte medio de cada variable a cada método, siendo las variables de proximidad, las más significativas (edificaciones, hidrología, vías o explotaciones ganaderas).



Tabla 6. Resumen del aporte medio de las variables ambientales a cada modelo.

2012	GLM	GBM	RF	GAM
DISTVERT	0.01850	0.02425	0.00100	0.51075
EDIF	0.56575	0.32050	0.08625	0.39850
EXPGAN	0.32175	0.04000	0.00425	0.46200
HIDRO	0.04850	0.07000	0.01150	0.15375
MDT	0.00000	0.01675	0.00075	0.22900
ORIENT	0.11375	0.02525	0.00550	0.28525
PENDIENT	0.00000	0.02650	0.01150	0.14225
PTOT12	0.00000	0.01600	0.00300	0.44825
TMED12	0.22250	0.00525	0.00225	0.39125
USOS	0.07950	0.00575	0.00075	0.26800
DISTVIAPECF	0.13750	0.01125	0.00200	0.26150
VIAS	0.20825	0.05075	0.01475	0.23850
2013	GLM	GBM	RF	GAM
DISTVERT	0.16325	0.05100	0.01100	0.23600
EDIF	0.09225	0.03925	0.02625	0.22475
EXPGAN	0.22550	0.02975	0.01425	0.31400
HIDRO	0.17175	0.06675	0.01275	0.21650
MDT	0.19800	0.04475	0.02825	0.22175
ORIENT	0.11325	0.01050	0.00125	0.24025
PENDIENT	0.54675	0.21500	0.05825	0.51525
PTOT12	0.29700	0.00875	0.00225	0.34675
TMED12	0.11725	0.02250	0.00800	0.22900
USOS	0.36400	0.17500	0.05450	0.30575
DISTVIAPECF	0.06600	0.07350	0.02075	0.14575
VIAS	0.17425	0.01175	0.00550	0.27400
2014	GLM	GBM	RF	GAM
DISTVERT	0.00000	0.01050	0.01025	0.37325
EDIF	0.06075	0.07550	0.05700	0.36350
EXPGAN	0.09475	0.00600	0.00900	0.26225
HIDRO	0.08250	0.07700	0.05025	0.32225
MDT	0.00000	0.00825	0.02375	0.26200
ORIENT	0.01150	0.01050	0.00950	0.16200
PENDIENT	0.06750	0.04875	0.03475	0.29200
PTOT12	0.18575	0.07625	0.04850	0.16250
TMED12	0.14575	0.02925	0.03650	0.44150
USOS	0.35675	0.13275	0.08850	0.33025
DISTVIAPECF	0.00900	0.01375	0.01800	0.26800
VIAS	0.11600	0.04025	0.04200	0.38825

También son significativos los valores de los usos del suelo como fuente de alimento o los aspectos topográficos, como la pendiente. Estas respuestas han sido proyectadas adicionalmente, distinguiendo entre los distintos métodos, para facilitar su comprensión y

analizar cierta similitud en la respuesta estadística de las variables ambientales analizadas frente a los métodos seleccionados (Figura 22).

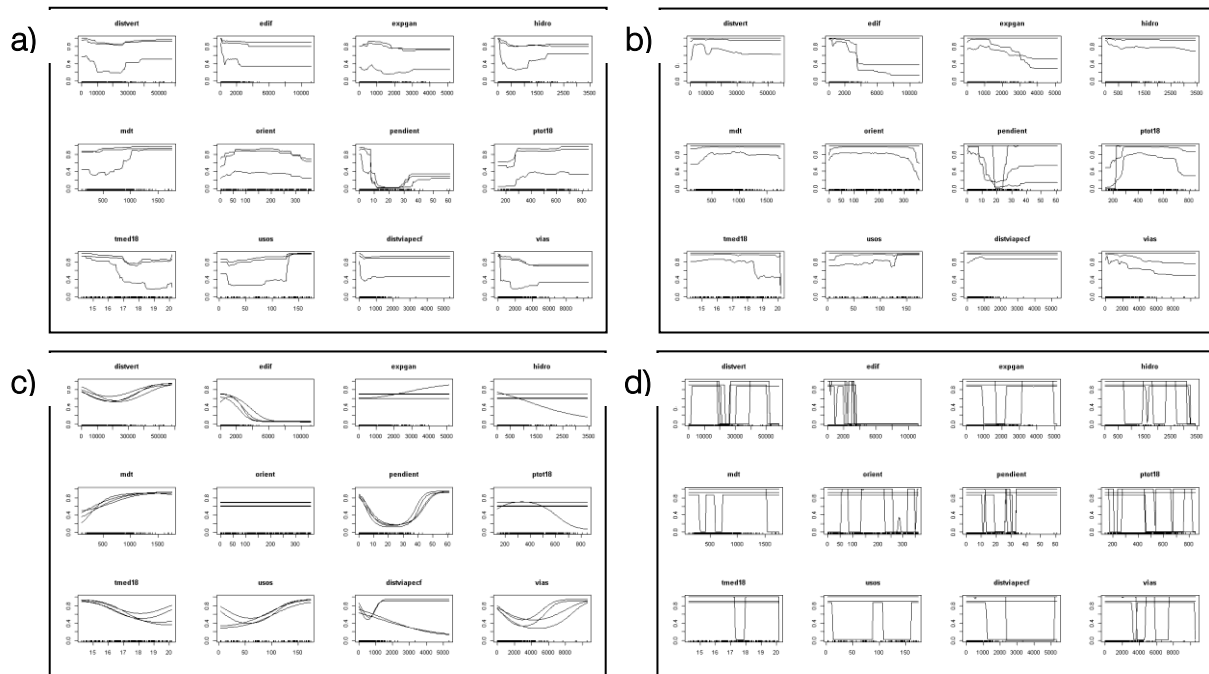


Figura 22. Representación gráfica de la respuesta del modelo de año 2018 a las variables ambientales: a) GAM, b) GBM, c) GML y d) RF.

Respecto al modelo del área de nidificación, el mejor método fue RF con un valor de TSS 0,85 y ROC 0,9. Respecto a GAM, no ha generado resultados, posiblemente debido a la escasa muestra de datos. En cuanto a las variables la más significativas por sus aportes medios han sido la distancia a la red hidrográfica, el MDT y el ráster de diferencia de alturas. Adicionalmente, se obtendrá una proyección del análisis del conjunto de métodos, atendiendo a los resultados de la evaluación mediante TSS, que generará una nueva simulación con los valores medios óptimos de los métodos mencionados previamente (Figura 23)

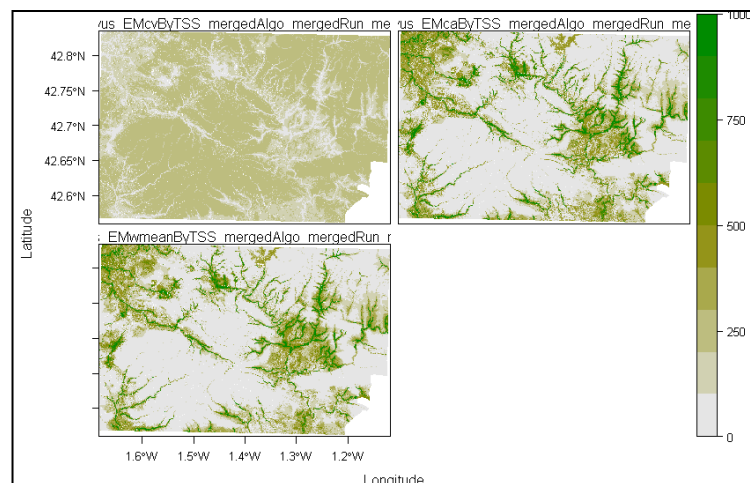


Figura 23. Modelo de ensamble proyectado a partir de los TSS mediante Biomod2

3.1.2.2 Evaluación de los resultados de MaxEnt

Mediante la interfaz intuitiva de MaxEnt y los numerosos archivos de resultados, se puede analizar detalladamente los resultados de la evaluación tanto del modelo configurado como de las variables predictivas aplicadas en el estudio.

Para comenzar se analizará la tasa de omisión de los mismos casos que en el análisis anterior (2013 – 2016, de izquierda a derecha y de arriba abajo en la Figura 24), para evaluar el rendimiento generado de la configuración del modelo predictivo. Este resultado se puede analizar tanto desde el archivo csv como del HTML generado en el modelo.

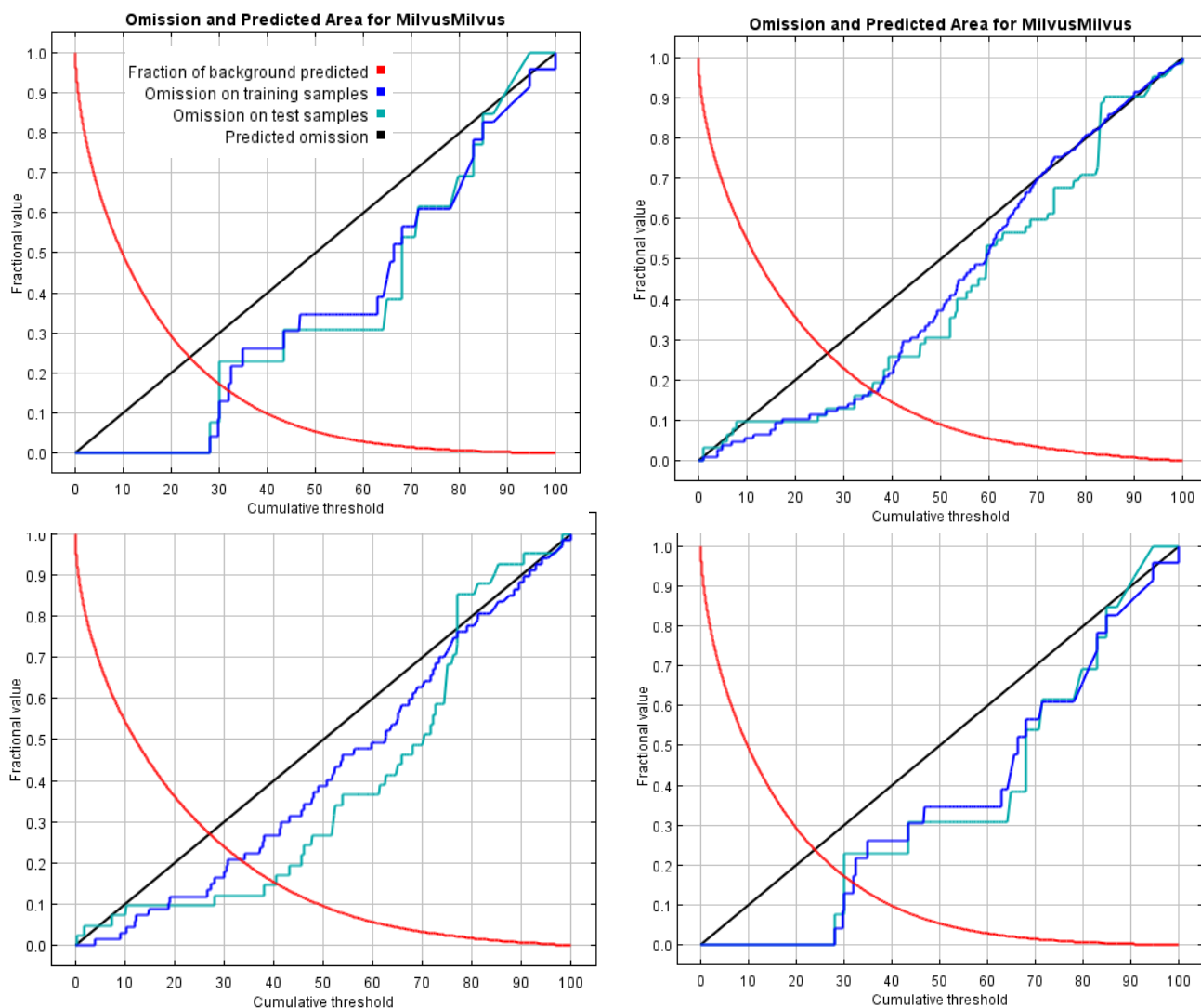


Figura 24. Comparativa de resultados de la tasa de omisión y área predictiva mediante MaxEnt

Como se observa en la Figura 24, los resultados de omisión en los modelos de 2012 y 2013, se sitúan en umbrales de baja probabilidad y se alejan de la línea de referencia. Esto podría indicar el sobreajuste del modelo. El sobreajuste podría deberse a la dependencia entre las observaciones y al reducido número de muestras.

Sin embargo, en el caso de 2013 y 2014, se aprecia unos valores más homogéneos, llegando a ser paralelas la línea de referencia y la representación de la tasa de omisión.

Empleando los mismos casos que en el análisis anterior, los modelos de MaxEnt de 2013 – 2016 arrojan unos valores AUC de 0,863 a 0,944, valores que se podrían calificar entre buenos y muy buenos según los rangos de Swets (Swets, 1988). Los años con mayor variación son 2014 y 2015, lo que podría deberse a un aumento significativo de registros, los cuales necesitarían un ajuste mayor de los parámetros (Figura 25).

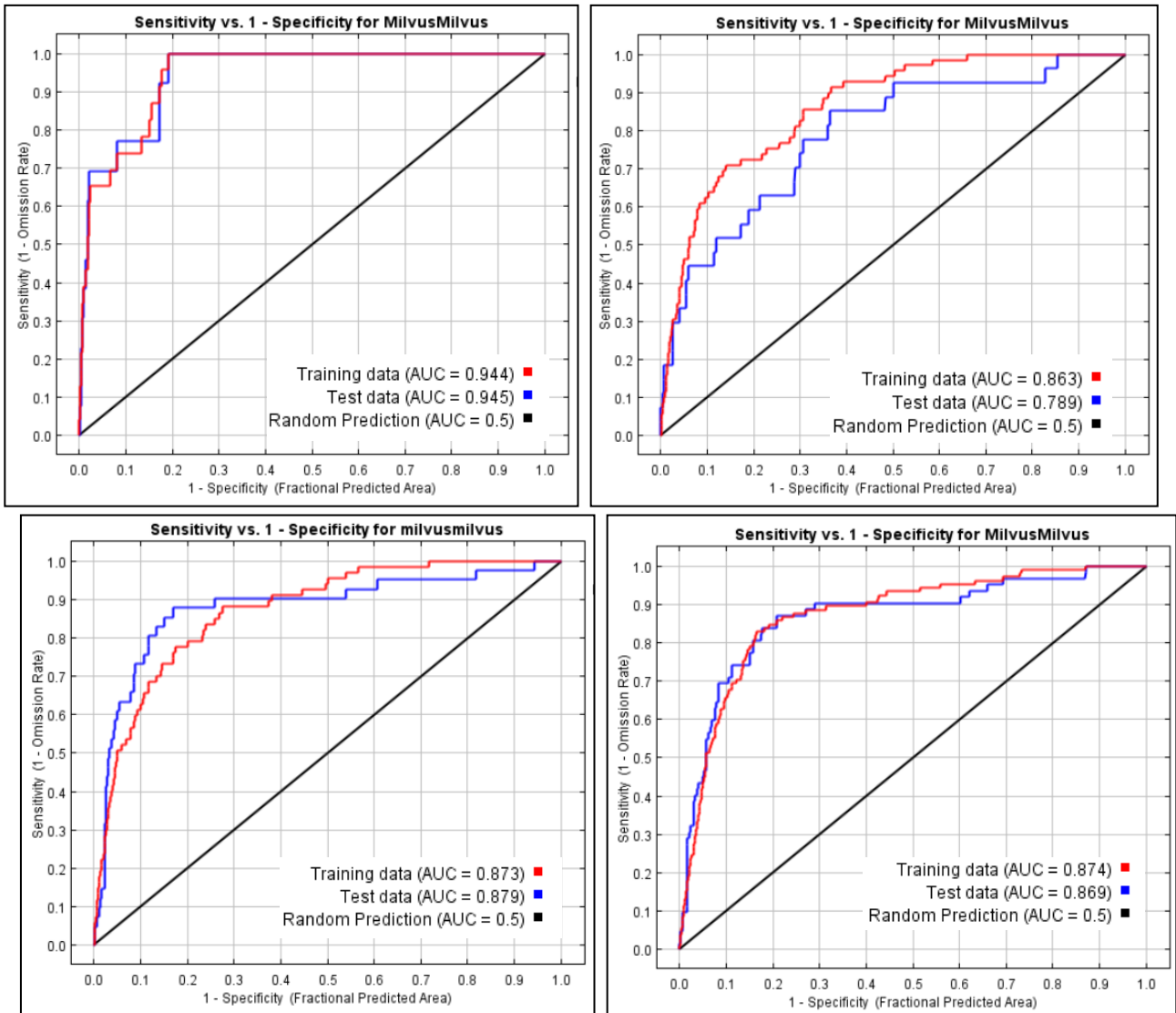


Figura 25. Comparativa de resultados de la evaluación de los métodos mediante el método AUC.

Respecto al peso de las variables ambientales en la distribución modelada, de nuevo se encuentran las variables de distancia a puntos de vertido, edificaciones, vías o hidrología entre las más importantes (valores superiores al 0,7). Las variables topográficas y climatológicas, así como los usos del suelo, también obtienen valores elevados mediante la evaluación ROC (Figura 26).

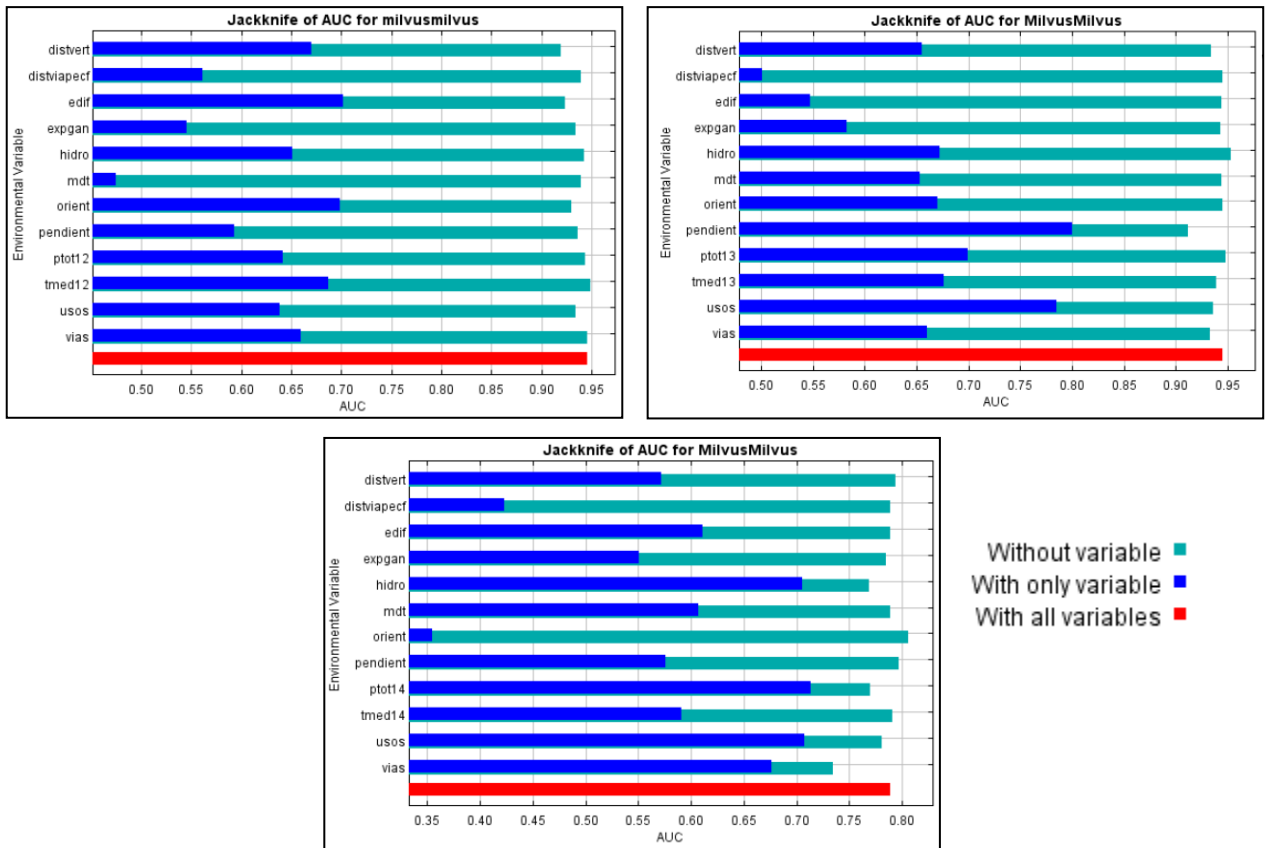


Figura 26. Representación gráfica de la respuesta de las variables ambientales en los modelos 2012, 2013 y 2014 mediante MaxEnt.

Y finalmente respecto a la zona de nidificación el valor obtenido en la curva ROC ha sido de 0,962 y cuyas variables de mayor importancia han sido las variables ambientales de proximidad, tanto a fuentes de alimento como puede ser los puntos de vertido, como hidrografía, además de las zonas urbanas (Figura 27). También tienen un elevado índice las variables topográficas, así como las enfocadas a la vegetación, tanto usos y aprovechamientos como la diferencia de alturas, obtenidas de Lidar.

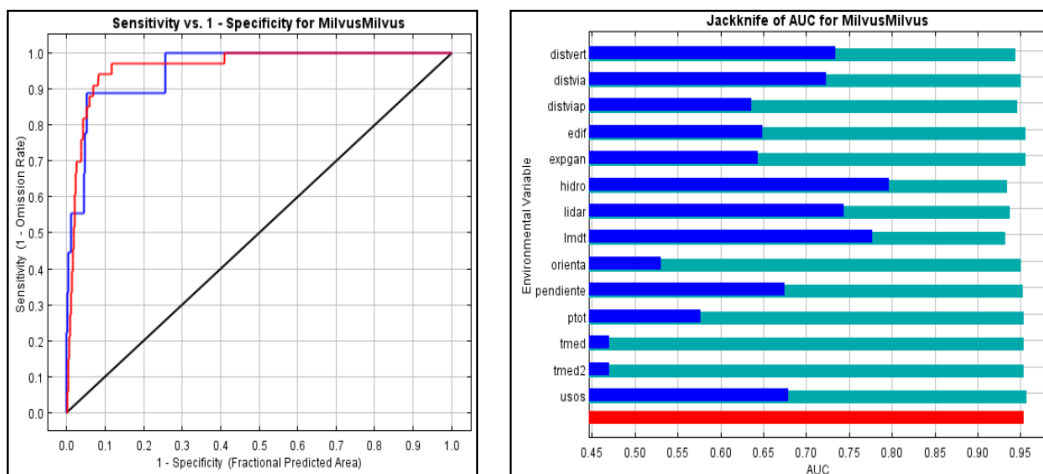


Figura 27. Evaluación mediante curva ROC y representación gráfica de la respuesta de las variables ambientales de modelo de zona de nidificación mediante MaxEnt.

3.2 Análisis de resultados

3.2.1 Condicionantes de hábitat

Una vez concluida la evaluación del modelado del nicho ecológico del Milano Real se han buscado evidencia de factores que favorecen el desarrollo de esta especie, derivadas de las predicciones de ocupación. Para ello se ha considerado, la incertidumbre de los modelos, ya sea por el sesgo de la muestra como en las superficies ambientales generadas. Los errores se han minimizado mediante el empleo de ensambles de algoritmos, así como diferente resolución.

Entre los factores más determinantes se encuentran los básicos para su supervivencia, proximidad a cursos fluviales o cultivos, así como puntos de vertidos donde pueden obtener fuente de alimento. Otros factores, aunque más moderados, son la orografía y el clima, ya que se desarrollan en zonas templadas y de elevación moderada. Aunque el factor más determinante de este nicho ecológico es la proximidad a vías de comunicación, resultado que es muy posible que se deba al método de muestreo mediante transectos lineales, así como el acceso propicio a los nidos de Milano Real.

3.2.2 Representación cartográfica del hábitat potencial

Mediante la representación cartográfica del hábitat potencial se puede atestiguar los condicionantes anteriormente descritos, generando áreas con alta probabilidad de ocurrencia en las proximidades de núcleos urbanos, así como en las vías de comunicación que articulan estos espacios. También son significativos los núcleos rurales, aunque con valores próximos a la probabilidad de ocurrencia al azar.

Debido a la amplitud de productos obtenidos mediante la modelización del hábitat potencial reproductor del Milano Real, se ha procedido a proyectar los resultados más significativos durante el periodo de análisis, seleccionando las mejores predicciones obtenidas mediante el paquete Biomod2, aplicando un único algoritmo (Figura 28) como mediante conjuntos (Figura 29) y comparándolos con los resultados obtenidos en MaxEnt (Figura 30).

Se puede observar concordancia entre los mejores resultados obtenidos mediante Biomod2 aplicando un único algoritmo (RF) y MaxEnt, así como la concentración de predicciones próximas al radio de los datos de observación. Sin embargo, al implementar *ensambles* en Biomod2, los resultados se dispersan, debido a la información adicional obtenida del conjunto de algoritmos aplicados (GAM, GBM, GML y RF), generando nuevas áreas susceptibles de colonización por la especie para su reproducción, así como la incertidumbre de posibles errores obtenidos de otros modelos menos adecuados, según los resultados de los parámetros de evaluación.

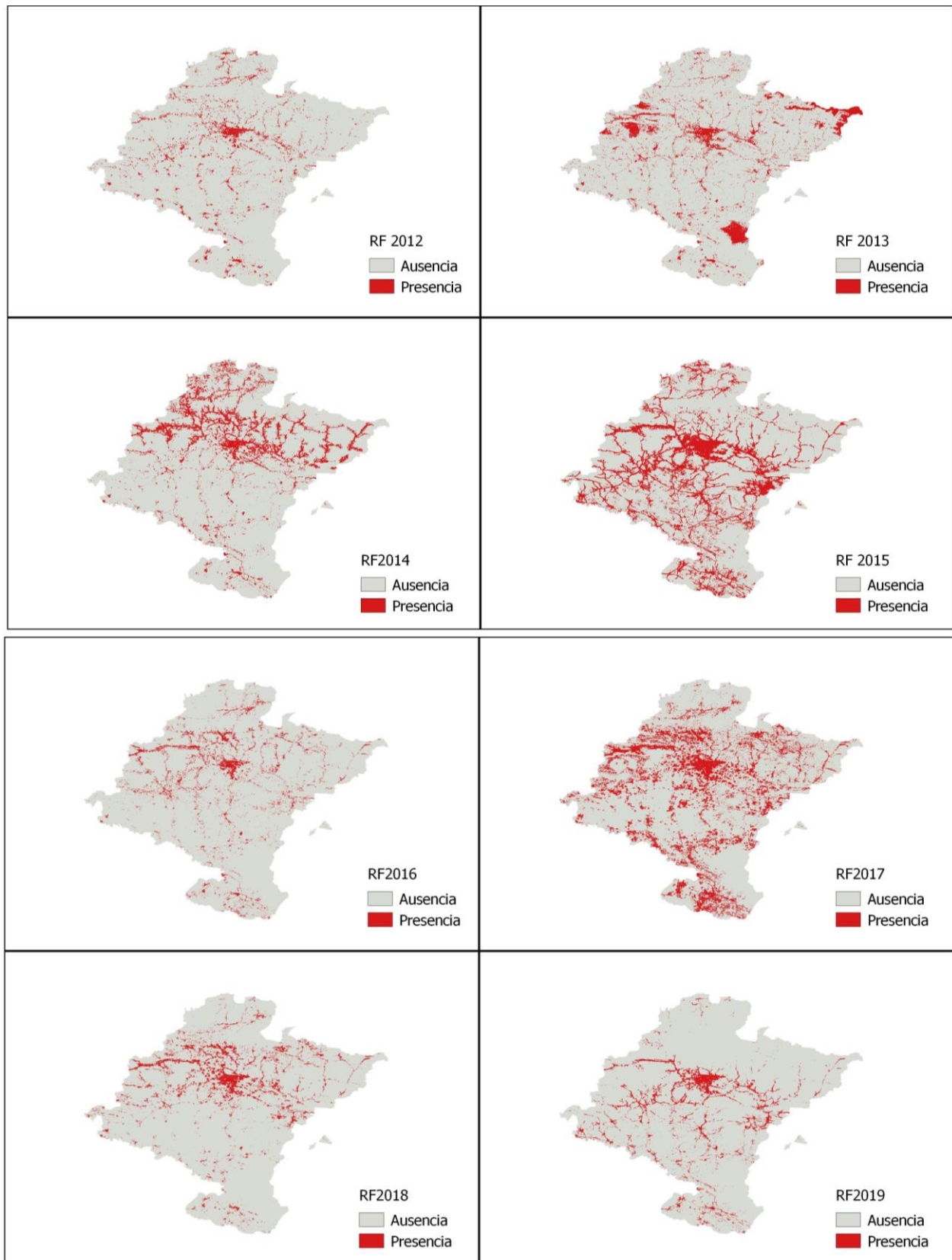


Figura 28. Modelos predictivos mediante árboles de decisión generados mediante paquete Biomod2, periodo 2012 -2019.

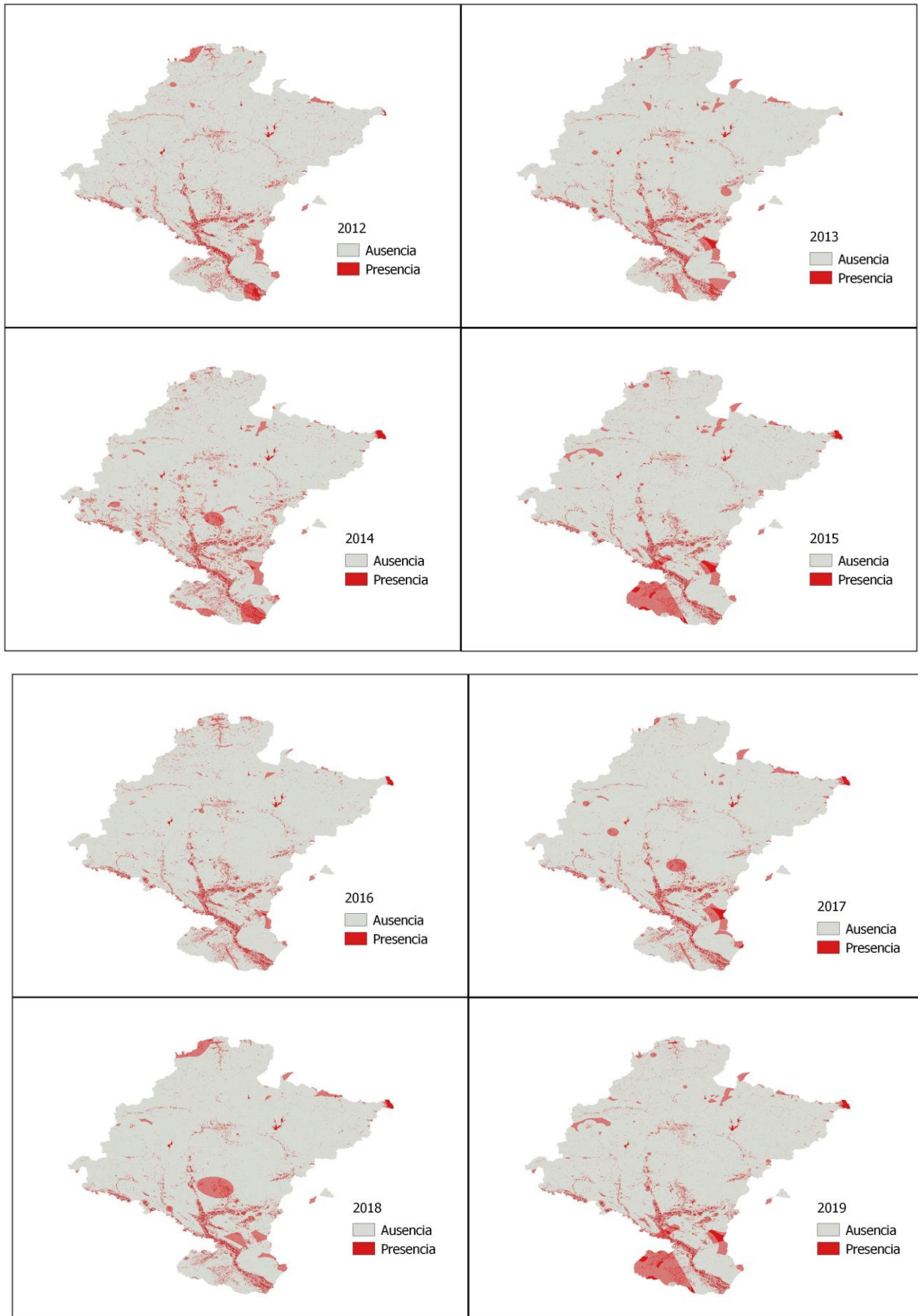


Figura 29. Modelos conjunto generados mediante paquete Biomod2, periodo 2012 -2019.

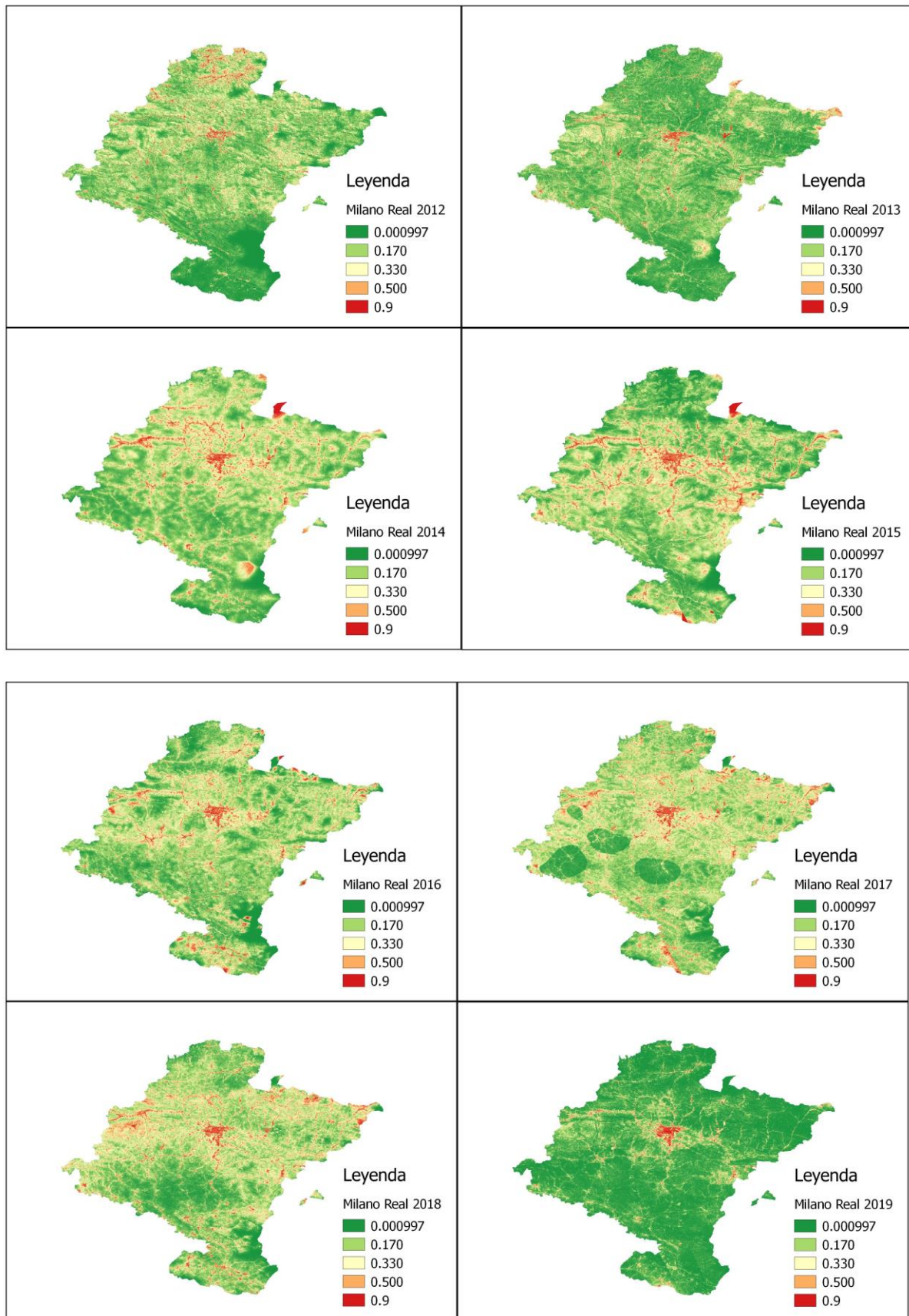


Figura 30. Modelos generados mediante MaxEnt, periodo 2012 -2019.

Otro aspecto significativo presente en todo el periodo de análisis, es el elevado potencial del área de la Cuenca de Pamplona. Esta zona se extiende, ramificándose a la Zona Media, especialmente en el periodo 2014 – 2018. La extensión de la zona coincide con periodos de mayor riqueza documental de presencias. De igual modo, los años con menor presencias registradas, presentan menor disgregación en el área potencia. Estas áreas se concentran, en el entorno de la Comarca de Pamplona y Pirineos. Estos fenómenos podrían deberse a los datos de partida.

Los modelos hipotéticos generados mediante *ensembles*, muestran importantes áreas potenciales en la zona de la Ribera o Pirineos, que no se corresponden con las áreas generadas por MaxEnt ni con la escasa expansión de los modelos óptimos obtenidos con los árboles de decisión, abriendo nuevas posibilidades de estudio.

Desde el punto de vista de la relación entre variables y áreas potenciales, se atestigua la relación entre la especie y el humano, como fuente de alimento con espacios como los puntos de vertido, así como el desarrollo del hábitat en espacios con alta densidad de población, o incluso proximidad a vías de comunicación.

En el caso del área específica estas observaciones se hacen más evidentes, ajustándose el modelo generado por MaxEnt a una distribución lineal en vías de comunicación y área de ribera frente al modelo generado con Biomod2, cuya distribución es más heterogénea. A esta resolución, se aprecia que existen amplios espacios catalogados como poco como propicios (ausencias), que concurren en ambos métodos, demostrando cierta eficiencia en los modelos generados.

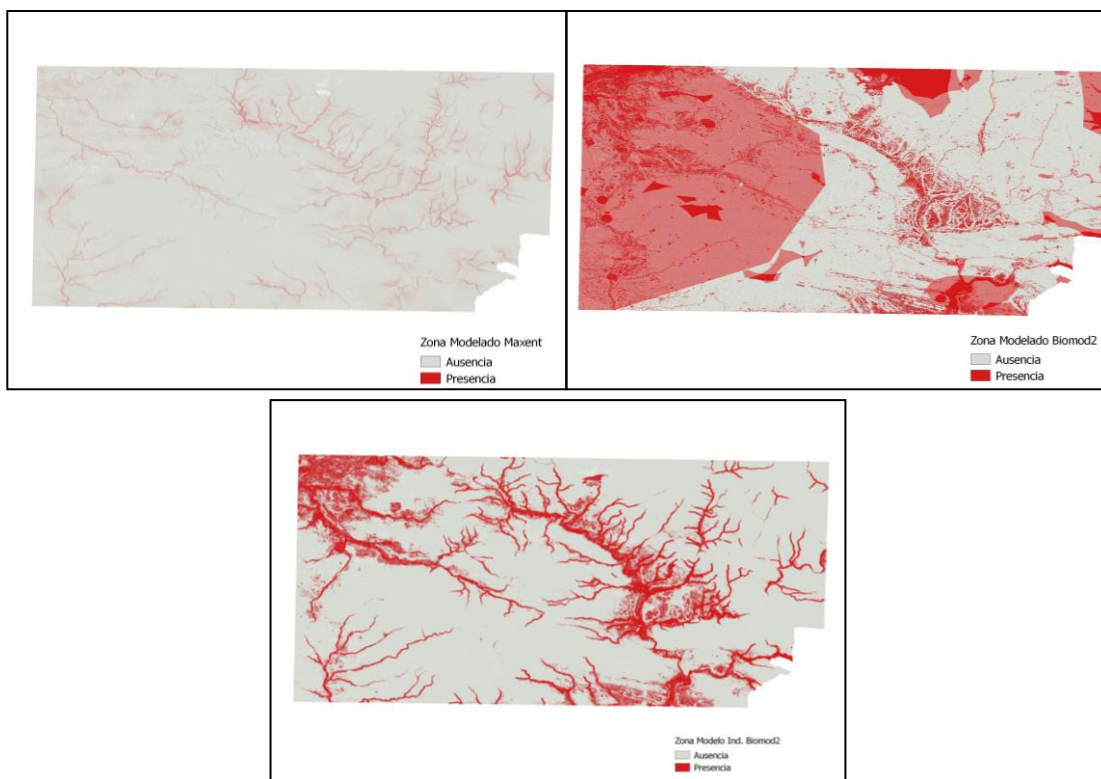


Figura 31. Modelos generados para la zona de nidificación mediante MaxEnt y el paquete Biomod2, periodo 2012 -2019.

4. CONCLUSIONES

Los modelos de selección de hábitat y modelado de nicho ecológico son herramientas especialmente útiles en la definición de las variables propicias para el desarrollo de organismo en un entorno considerado estable. Como se demuestra mediante este proyecto, los resultados del análisis estadístico son altamente sensibles, por lo que es indispensable una selección rigurosa de los datos de partida, los parámetros aplicables, la extensión de estudio y la escala aplicada de modelización, siendo preciso el seguimiento continuo para poder obtener resultados óptimos. En este proyecto, se han tratado de eliminar posibles sesgos en las observaciones, filtrando de datos, así como generando pseudoausencias aleatorias, que un posible condicionamiento geográfico de las presencias, así como realizar un seguimiento anual de las observaciones.

Cabe destacar la complejidad de la metodología empleada para el análisis del hábitat potencial de la especie, con gran diversidad de herramientas, que generan distintas respuestas dependiendo de los métodos estadísticos empleados. Estas técnicas de análisis que permitan generar hipótesis fundadas, mediante mecanismos de acceso libre y colectivo, siendo fundamentales en los estudios biogeográficos. Estas diferentes decisiones pueden generar modelos similares como es el caso del modelado con un único algoritmo óptimo mediante Biomod2 y MaxEnt, permitiendo generar un hábitat idóneo hipotético, como generar nuevas líneas de análisis como ocurre con los modelados de ensambles.

Y para concluir, desde este trabajo de aproximación al estudio de una especie tan representativa del territorio peninsular, se desea que pueda ser un punto de partida al estudio de nuevos espacios de muestreo, cuyas variables ambientales son similares y así poder generar un seguimiento reproductivo más aproximado a la realidad, para estimar la evolución de la especie y su interacción con sus factores abióticos y bióticos. De esta forma, se podría recuperar la especie en zonas ya desaparecida o en declive, así como redefinir hábitats de interés o estimar elementos perjudiciales para la especie, como puede ser el riesgo de electrocución o el índice de contaminación.

5. BIBLIOGRAFÍA

- Aldasoro, J.E., 1985. Navarra. Atlas de aves nidificantes. Pamplona.
- Cano-Barbacid, C., Cano Sánchez, J., 2016. Efectos del cambio climático sobre las aves. Cal. Meteorológico 2017 263–271.
- Cramp, S. and Simmons, K.E.L., 1980. The Birds of the Western Palearctic. Oxford.
- Deán, J.I., 2017. Censo de Milano Real invernante en España . Enero 2017 Recopilación de resultados.
- Georges, D., Thuiller, W., 2013. An example of species distribution modeling with biomod2 1–21.
- GRASS Development Team, 2017. Geographic Resources Analysis Support System (GRASS) Software, Version 7.2. Open Source Geospatial Foundation. Electronic document
- Grinnell, J., 1917. The niche-relationships of the California Trasher. Auk 34, 131–135.
- Guerra, F., Gómez, H., González, J., Zambrano, Z., 2006. Uso Actual de Métodos y técnicas para el Estudio de la Precipitación incluyendo Plataformas SIG Current use of Methods and Techniques in the Study of Precipitation Based on SIG Platforms.
- Guisan, A., Thuiller, W., Zimmermann, N.E., 2017. Habitat Suitability and Distribution Models: with Applications in R. Cambridge.
- Guisan, A. and Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Modell. 135, 147–186.
- Hutchinson, G.E., 1957. Population studies. Animal ecology and demography: concluding remarks. Cold Spring Harb. Symp. Quant. Biol. 22, 415–427.
- Instituto Geográfico Nacional (IGN), 2020. Centro Nacional de Información Geográfica (CNIG) [WWW Document]. URL <http://centrodedescargas.cnig.es/>
- Joppa, L.N., McInerney, G., Harper, R., Salido, L., Takeda, K., O'Hara, K., Gavaghan, D., Emmott, S., 2013. Troubling trends in scientific software use. Science (80-.). 340, 814–815. <https://doi.org/10.1126/science.1231535>
- Manzanares, A., 1991. Guía de las aves rapaces de España. Barcelona.
- Mateo, R.G., Felicísimo, Á.M., Muñoz, J., 2011. Modelos de distribución de especies: Una revisión sintética. Rev. Chil. Hist. Nat. 84, 217–240. <https://doi.org/10.4067/S0716-078X2011000200008>
- Merow, C., Smith, M.J., Silander, J.A., 2013. A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. Ecography (Cop.). 36, 1058–1069. <https://doi.org/10.1111/j.1600-0587.2013.07872.x>
- Naimi, B., Araujo, M.B., 2019. Package sdm: Species Distribution Modelling. R CRAN Proj. 1–10. <https://doi.org/10.1111/ecog.01881> >.License
- Navarra, G. de, Septiembre 2020. Meteo Navarra [WWW Document]. URL <http://meteo.navarra.es/>
- Navarra, G. de, Trabajos Catastrales, S., Septiembre 2020. Sistema de Información Territorial de Navarra (SITNA) [WWW Document]. URL <https://idena.navarra.es/>
- Peirce, C., 1884. The numerical measure of the success of predictions. Science (80-.). 4, 453–454.
- Phillips, S.B., Aneja, V.P., Kang, D., Arya, S.P., 2006. Modelling and analysis of the atmospheric nitrogen deposition in North Carolina. Int. J. Glob. Environ. Issues 6, 231–252. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>
- Puente, J. De, López, P., Vidal, J., Urios, V., 2018. Ecología espacial de las poblaciones españolas de milano real 1.
- QGIS Development Team, 2019. QGIS Geographic Information System.

- Soberón, J. y Peterson, A.T., 2005. Interpretation of models of fundamental ecological niches and species distributional areas. *Biodivers. Informatics* 2, 1–10.
- Soberón, J., 2007. Grinnellian and Eltonian niches and geographic distributions of species. *Ecol. Lett.* 10, 1115–1123.
- Sunyer, C., Viñuela, J., 2000. Conservación del Milano Real.
- Swets, J., 1988. Measuring the Accuracy of Diagnostic Systems. *Science* (80-.). 1285–1293.
- Thuiller, W., Münkemüller, T., 2010. Habitat suitability modelling, in: A.P. Møller, W. Fiedler, P.B. (Ed.), *Effects of Climate Change on Birds*. Oxford University Press, Oxford, pp. 77–85. <https://doi.org/10.1006/jema.2002.0546>
- Vanderwal J., Falconi L., Januchowski S., S.L.& S.C., 2012. Package ‘SDMTools’. *Species Distribution Modelling Tools: Tools for processing data associated with species distribution modelling exercises*. R-package version 1.1.12.
- Viñuela, J., 2004. Milano Real *Milvus milvus*. *Libr. Rojo las Aves España* 452.
- Wilfried Thuiller, Damien Georges, Robin Engler, F.B., 2019. Package “biomod2” Title Ensemble Platform for Species Distribution Modeling.

6. ANEXOS

SCRIPT R MEDIANTE BIOMOD2

```
#####INSTALAR LIBRERIAS NECESARIAS#####
library(biomod2)
library(ggplot2)
library(gridExtra)
library(raster)
library(dismo)

#####DEFINICION DEL ESCRITORIO#####
setwd("C:/2018")

#####1.CARGAR DATOS DE PRESENCIA#####
datos <- read.table("./Milano2018_presen.csv",header=TRUE,sep=";", dec=",")#Datos de
  presencia
plot(datos[,c("X","Y")])
datos
myRespName <- 'MilvusMilvus' #seleccion nombre especie
myResp <- as.numeric(datos[,myRespName]) #seleccion de las presencias
myRespXY <- datos[, c("X", "Y")] #cordenadas xy de los datos de milano real

#####2.CARGAR VARIABLES Y TRANSFORMACION A GRD#####
#cARGAR ARCHIVOS ASCII
disvert<-raster("./distvert.asc",header=TRUE)
edif<-raster("./edif.asc",header=TRUE)
expgan<-raster("./expgan.asc",header=TRUE)
hidro<-raster("./hidro.asc",header=TRUE)
mtn<-raster("./mdt.asc",header=TRUE)
orient<-raster("./orient.asc",header=TRUE)
pend<-raster("./pendient.asc",header=TRUE)
ptot18<-raster("./ptot18.asc",header=TRUE)
tmed18<-raster("./tmed18.asc",header=TRUE)
usos<-raster("./usos.asc",header=TRUE)
viapecpf<-raster("./distviapecf.asc",header=TRUE)
vias<-raster("./vias.asc",header=TRUE)

#TRANSFORMACION A .GRD (EMPLEADO EN BIOMOD2)
writeRaster(disvert,"./Variables/disvert.grd",overwrite=TRUE)
writeRaster(edif,"./Variables/edif.grd",overwrite=TRUE)
writeRaster(expgan,"./Variables/expgan.grd",overwrite=TRUE)
writeRaster(hidro,"./Variables/hidro.grd",overwrite=TRUE)
writeRaster(mtn,"./Variables/mtn.grd",overwrite=TRUE)
```

```
writeRaster(orient,"./Variables/orient.grd",overwrite=TRUE)
writeRaster(pend,"./Variables/pend.grd",overwrite=TRUE)
writeRaster(ptot18,"./Variables/ptot18.grd",overwrite=TRUE)
writeRaster(tmed18,"./Variables/tmed18.grd",overwrite=TRUE)
writeRaster(usos,"./Variables/usos.grd",overwrite=TRUE)
writeRaster(viapecpf,"./Variables/viapecpf.grd",overwrite=TRUE)
writeRaster(vias,"./Variables/vias.grd",overwrite=TRUE)

#ENSAMBLADO EN UN ARCHIVO
myExpl<-
  stack("./Variables/disvert.grd","./Variables/edif.grd","./Variables/expgan.grd","./Variables/hidro.grd",

  "./Variables/mtn.grd","./Variables/orient.grd","./Variables/pend.grd","./Variables/ptot18.grd",
  ,

  "./Variables/tmed18.grd","./Variables/usos.grd","./Variables/viapecpf.grd","./Variables/vias.grd")

####3.CONSTRUIR MODELADO CON BIOMOD2####
#TRANSFORMACION A FORMATO DATOS ENTRADA BIOMOD2 Y GENERACION DE AUSENCIAS
Biomod <- BIOMOD_FormatingData(resp.var = myResp, #presencias
  expl.var = myExpl, #variables
  resp.xy = myRespXY, #coordenadas
  resp.name = myRespName, #nombre especie
  PA.nb.rep = 2,
  PA.nb.absences = 100,
  PA.strategy = 'random') #repeticiones de ausencia

Biomod #resumen
plot(Biomod)
#OPCIONES
milano_opc <- BIOMOD_ModelingOptions() #por defecto
#milano_opc <- BIOMOD_ModelingOptions(GLM = list( type = 'quadratic',
  #interaction.level = 1 ),
  #GBM = list( n.trees = 1000 ),
  # GAM = list( algo = 'GAM_mgcv' ) )

milano_modelo <- BIOMOD_Modeling( data = Biomod,
```

```
models = c("GLM", "GBM", "RF", "GAM"),
models.options = milano_opc,
NbRunEval = 1,
DataSplit = 80,
Yweights=NULL,
VarImport=3,
models.eval.meth = c('TSS','ROC'),
SaveObj = TRUE,
rescal.all.models = TRUE)

milano_modelo #resumen

####4.EVALUACION MODELADO CON BIOMOD2####
## Conseguir los errores generados del modelo
milano_modelo_ev <- get_evaluations(milano_modelo) #de todos los modelos

## Graficas de los errores y de las iteraciones realizadas
dim(milano_modelo_ev)
dimnames(milano_modelo_ev)
milano_modelo_ev["TSS","Testing.data","RF",,]
milano_modelo_ev["ROC","Testing.data",,,]

models_scores_graph(milano_modelo, by = "models" , metrics = c("ROC","TSS"),
  xlim = c(0.5,1), ylim = c(0.5,1))
models_scores_graph(milano_modelo, by = "cv_run" , metrics = c("ROC","TSS"),
  xlim = c(0.5,1), ylim = c(0.5,1))
models_scores_graph(milano_modelo, by = "data_set" , metrics = c("ROC","TSS"),
  xlim = c(0.5,1), ylim = c(0.5,1))
(milano_mod_var_import <- get_variables_importance(milano_modelo))

## Media de la importancia de la variable por algoritmo
apply(milano_mod_var_import, c(1,2), mean)
milano_glm <- BIOMOD_LoadModels(milano_modelo, models='GLM')
milano_gbm <- BIOMOD_LoadModels(milano_modelo, models='GBM')
milano_rf <- BIOMOD_LoadModels(milano_modelo, models='RF')
milano_gam <- BIOMOD_LoadModels(milano_modelo, models='GAM')

#Generacion de evaluacion
glm_eval_strip <- biomod2:::response.plot2(
  models = milano_glm,
  Data = get_formal_data(milano_modelo,'expl.var'),
  show.variables= get_formal_data(milano_modelo,'expl.var.names'),
  do.bivariate = FALSE,
```

```
fixed.var.metric = 'median',
legend = FALSE,
display_title = FALSE,
data_species = get_formal_data(milano_modelo,'resp.var'))
gbm_eval_strip <- biomod2::response.plot2(
  models = milano_gbm,
  Data = get_formal_data(milano_modelo,'expl.var'),
  show.variables= get_formal_data(milano_modelo,'expl.var.names'),
  do.bivariate = FALSE,
  fixed.var.metric = 'median',
  legend = FALSE,
  display_title = FALSE,
  data_species = get_formal_data(milano_modelo,'resp.var'))
rf_eval_strip <- biomod2::response.plot2(
  models = milano_rf,
  Data = get_formal_data(milano_modelo,'expl.var'),
  show.variables= get_formal_data(milano_modelo,'expl.var.names'),
  do.bivariate = FALSE,
  fixed.var.metric = 'median',
  legend = FALSE,
  display_title = FALSE,
  data_species = get_formal_data(milano_modelo,'resp.var'))
gam_eval_strip <- biomod2::response.plot2(
  models = milano_gam,
  Data = get_formal_data(milano_modelo,'expl.var'),
  show.variables= get_formal_data(milano_modelo,'expl.var.names'),
  do.bivariate = FALSE,
  fixed.var.metric = 'median',
  legend = FALSE,
  display_title = FALSE,
  data_species = get_formal_data(milano_modelo,'resp.var'))

milano_ensemble_modelos <- BIOMOD_EnsembleModeling( milano_modelo,
                                                    modeling.output =
                                                    em.by = 'all',
                                                    eval.metric = 'TSS',
                                                    eval.metric.quality.threshold = 0.8,
                                                    models.eval.meth = c('KAPPA','TSS','ROC'),
                                                    prob.mean = FALSE,
                                                    prob.cv = TRUE,
                                                    committee.averaging = TRUE,
                                                    prob.mean.weight = TRUE,
```



```
VarIimport = 0 )
(milnano_ensemble_models_scores <- get_evaluations(milano_ensemble_modelos))

### 5. PROYECCION DEL MODELADO ###
milano_modelo_proj_current <- BIOMOD_Projection( modeling.output = milano_modelo,
                                                new.env = myExpl,
                                                proj.name = "current",
                                                binary.meth = "TSS",
                                                output.format = ".grd",
                                                do.stack = FALSE )
plot(milano_modelo_proj_current)

milano_ensemble_models_proj_current <- BIOMOD_EnsembleForecasting(
  projection.output = milano_modelo_proj_current,
  EM.output = milano_ensemble_modelos)
plot(milano_ensemble_models_proj_current)
```