



IX Taller GBIF de Modelización de Nichos Ecológicos (sesión 2)

Aprendiendo a modelizar

Blas M. Benito

CONTENIDOS

- TEORÍA Y PRÁCTICA DE MÉTODOS DE MODELADO:
 - GLM
 - GAM
 - RANDOM FOREST
 - MAXENT
- EVALUACIÓN DE MODELOS
- APLICACIÓN DE “THRESHOLDS”
- PROYECCIÓN

SCRIPT DE R

Reinicia Rstudio y abre
2_modelos.R

CALIBRANDO MODELOS

MÉTODOS DE REGRESIÓN (GLM y GAM)

GENERALIZED LINEAR MODELS

- **Permiten modelar respuestas no lineales**
- Los residuos pueden seguir distintas distribuciones de probabilidad: normal, **binomial**, **Poisson**, binomial negativa, gamma

¿COMO FUNCIONA?

- Según complejidad de las curvas
 - Logística
 - Polinomio 2º, 3º, 4º, ... grado
- Según los datos de ausencia
 - Ausencia
 - Pseudo-ausencia
 - Background
- Según las interacciones entre variables
 - Sin interacción
 - Con interacción

DIBUJEMOS UN POCO PARA ENTENDERLO!

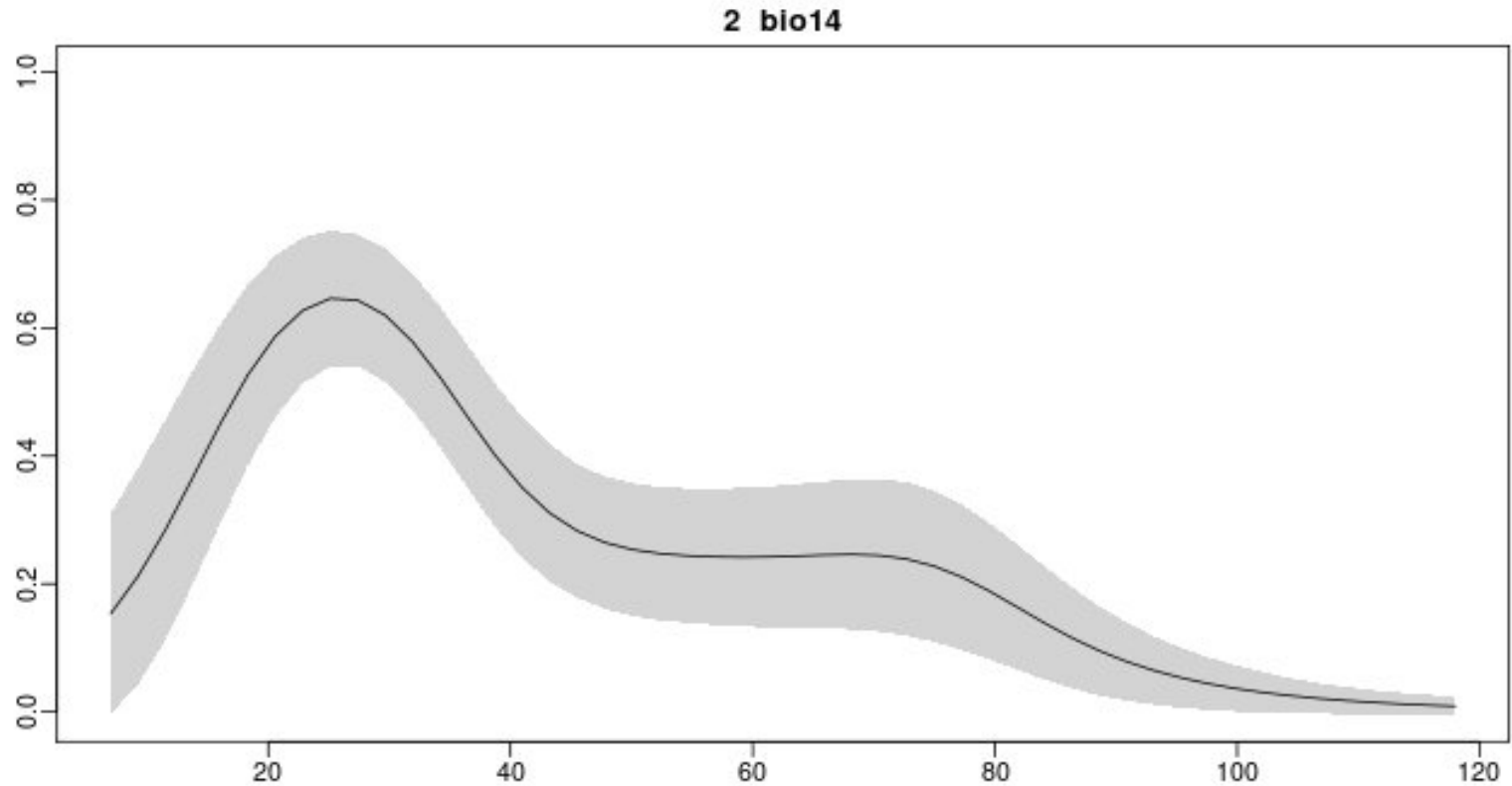
NÚMERO MÍNIMO DE PRESENCIAS

- Necesitamos al menos 5 presencias (y 5 ausencias, si el modelo es de presencia-ausencia) por cada variable.
- Necesitamos otras 5 presencias por cada término polinomial:
 - 1º grado: 5 presencias
 - 2º grado: 10 presencias
 - And so on.

GENERALIZED ADDITIVE MODELS (GAM)

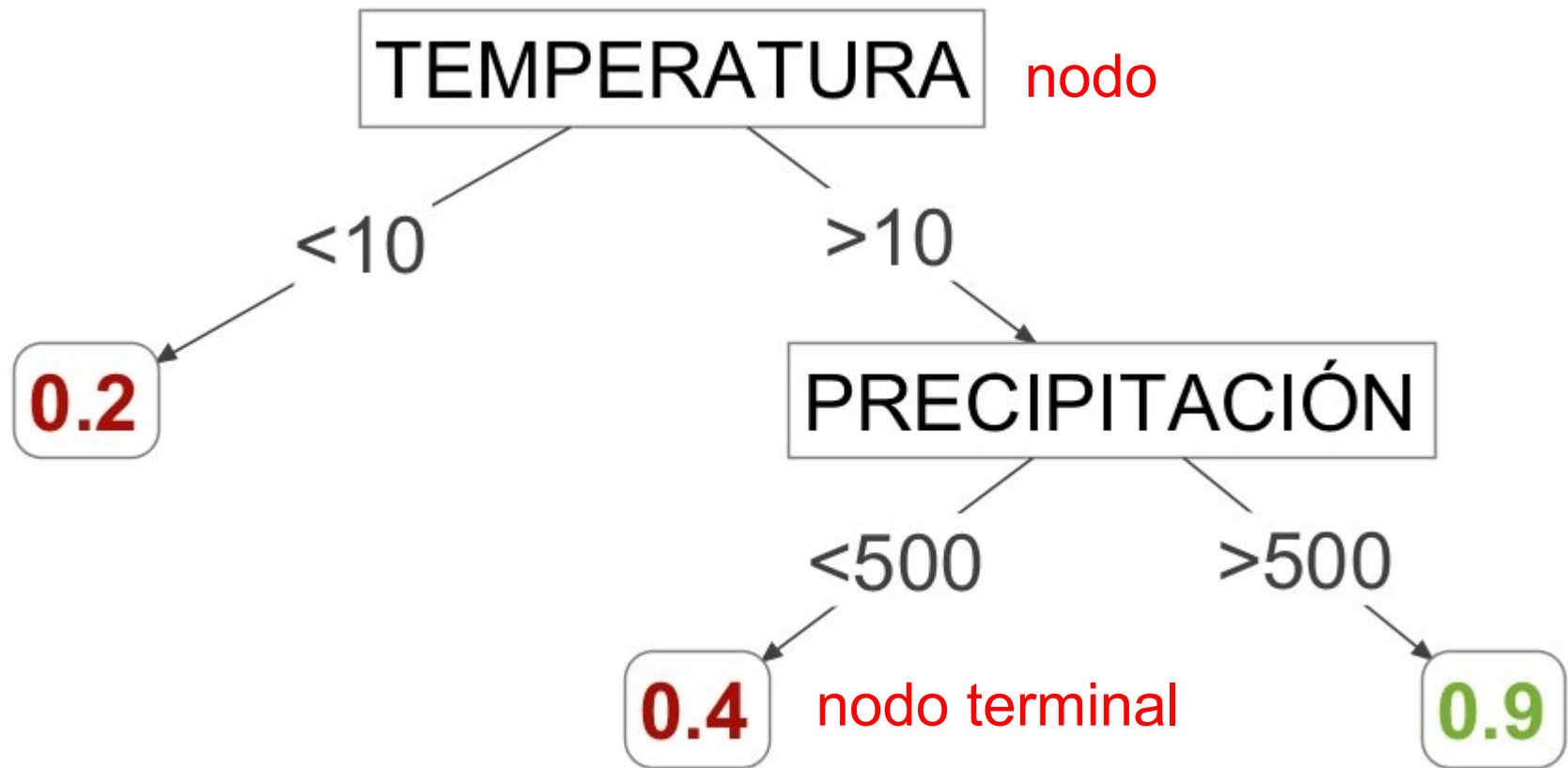
- Método de regresión no paramétrico
- Variables predictivas suavizadas (smoothing)
- Modelado de respuestas no lineales
- Requiere tamaños de muestra grandes (más que GLM)

GENERALIZED ADDITIVE MODELS

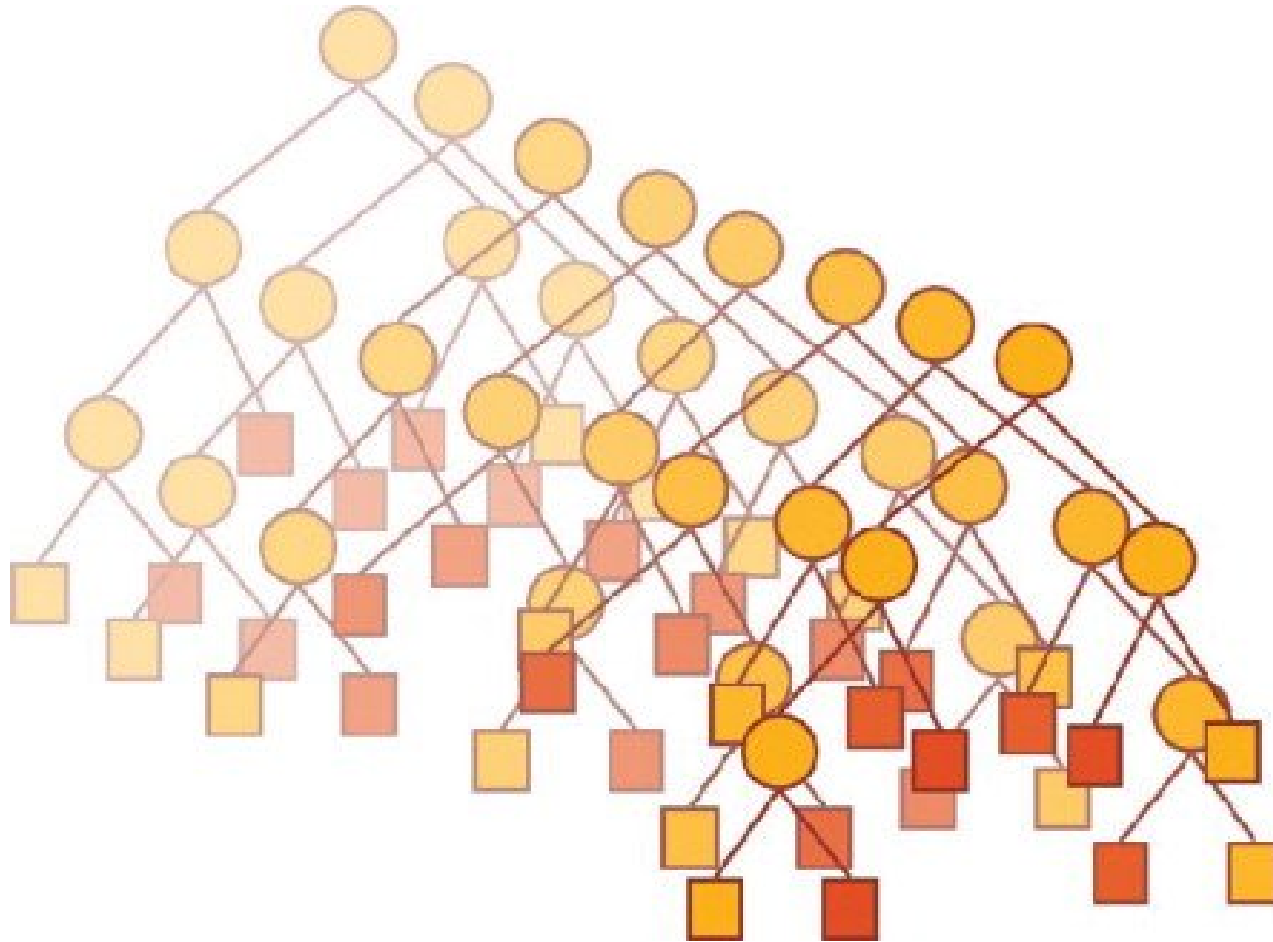


ÁRBOLES DE CLASIFICACIÓN Y REGRESIÓN

ÁRBOL DE REGRESIÓN



RANDOM FOREST



Fuente: Gedeck et al. 2010 Progress in Medicinal Chemistry

RANDOM FOREST

- Parámetros importantes:
 - **ntree**: número de árboles a calibrar
 - **mtry**: número de variables usadas en cada árbol
 - **nodesize**: número mínimo de casos en cada nodo terminal
 - **maxnode**: número máximo de nodos terminales

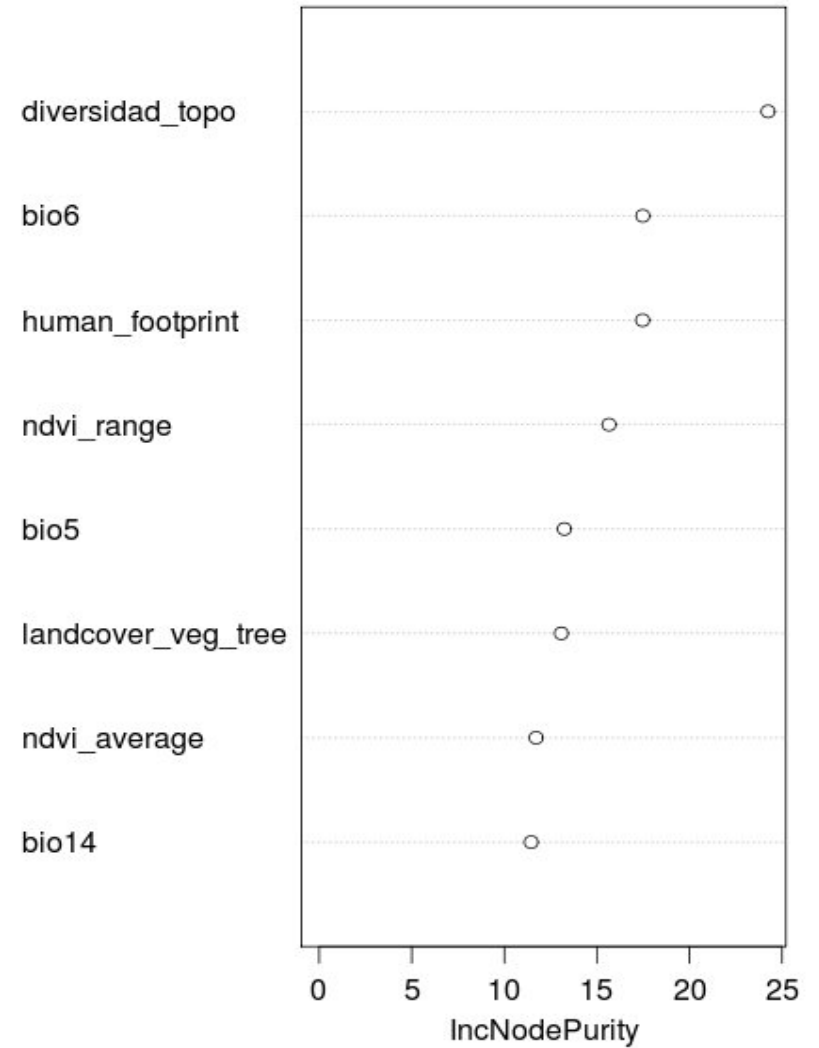
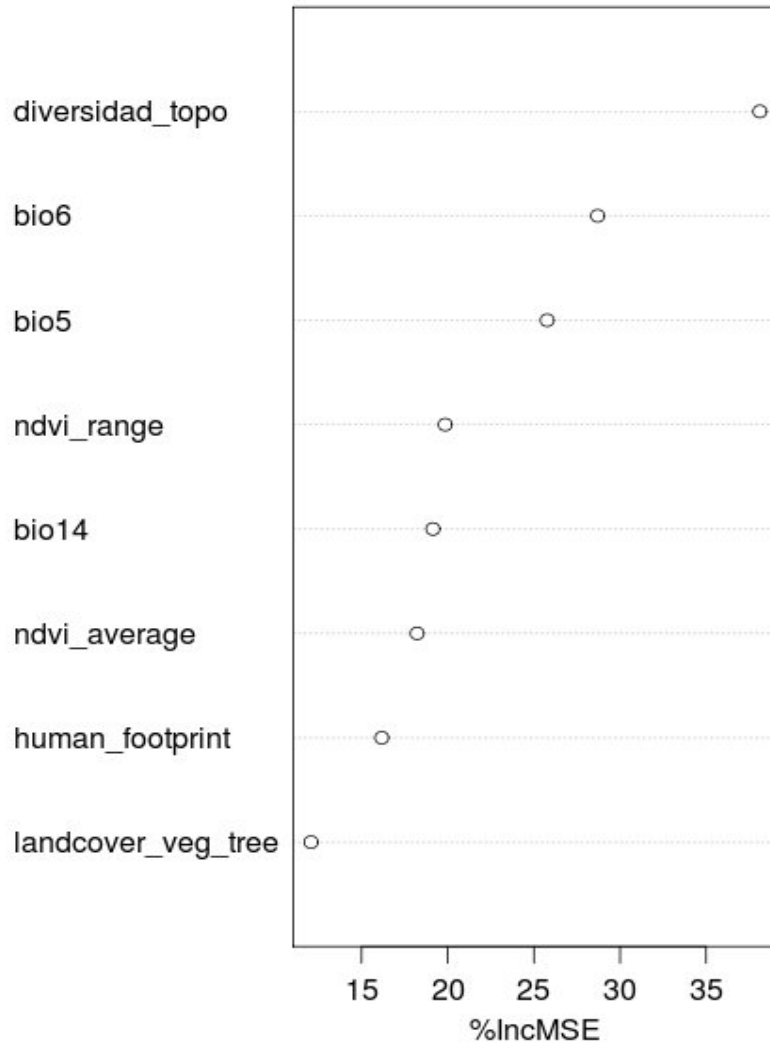
RANDOM FOREST

1. Por cada árbol:
 1. Selecciona n variables al azar
 2. Selecciona 60% de datos al azar
 3. Calibra un árbol de regresión
 4. Evalúa el árbol con el 40% de los datos no usados para calibrarlo
2. Una vez calibrados todos los árboles
 1. Calcula el resultado de un nuevo caso (celda) para cada uno de los árboles
 2. Calcula la moda del resultado de todos los árboles

RANDOM FOREST

- Ventajas
 - Muy potente
 - Puede manejar gran cantidad de datos
 - Analiza interacción de variables
- Inconvenientes
 - Potencial sobreajuste a los datos
 - El resultado es difícil de interpretar

IMPORTANCIA VARIABLES



ENSAMBLADO DE MODELOS

ENSAMBLADO

- JW Gibbs (1878): Muchas copias de un sistema consideradas simultáneamente. Cada copia representa un estado posible del sistema.
- JM Bates y CWJ Granger (1969): Un ensamblado tiene una probabilidad de error menor que cualquiera de sus constituyentes individuales.
- Araújo y New 2006: Al promediar varios modelos la señal objetivo emerge del ruido asociado a los errores e incertidumbres de los modelos individuales.

ENSAMBLADO

Las “copias” varían a lo largo de varios ejes

- Condiciones iniciales (tanto presencias como variables)
- Tipos de modelos
- Parámetros de los modelos

ENSAMBLADO

Espacio de
modelos
posibles



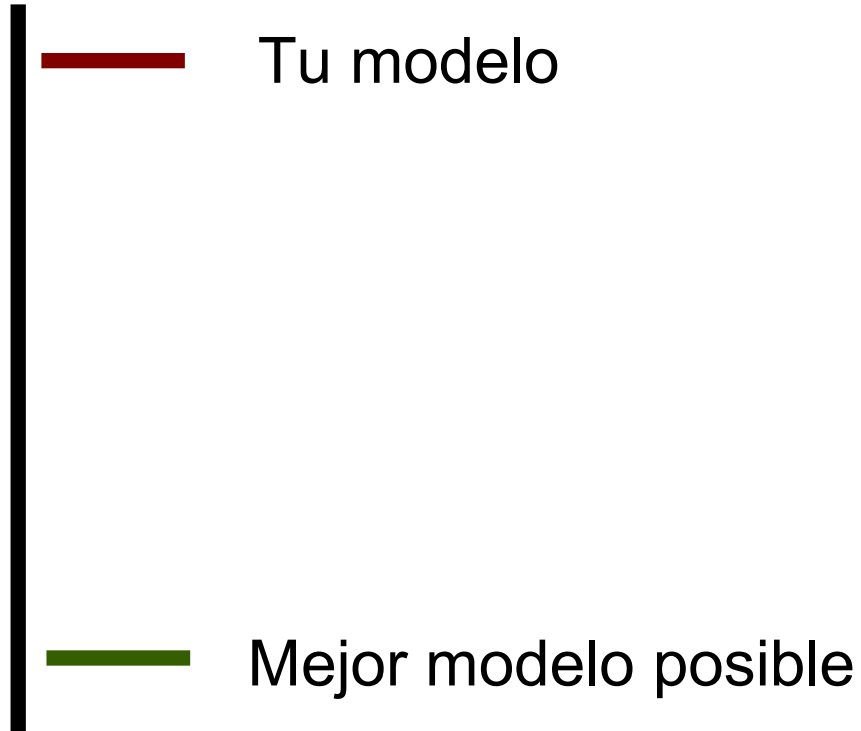
ENSAMBLADO

Espacio de
modelos
posibles

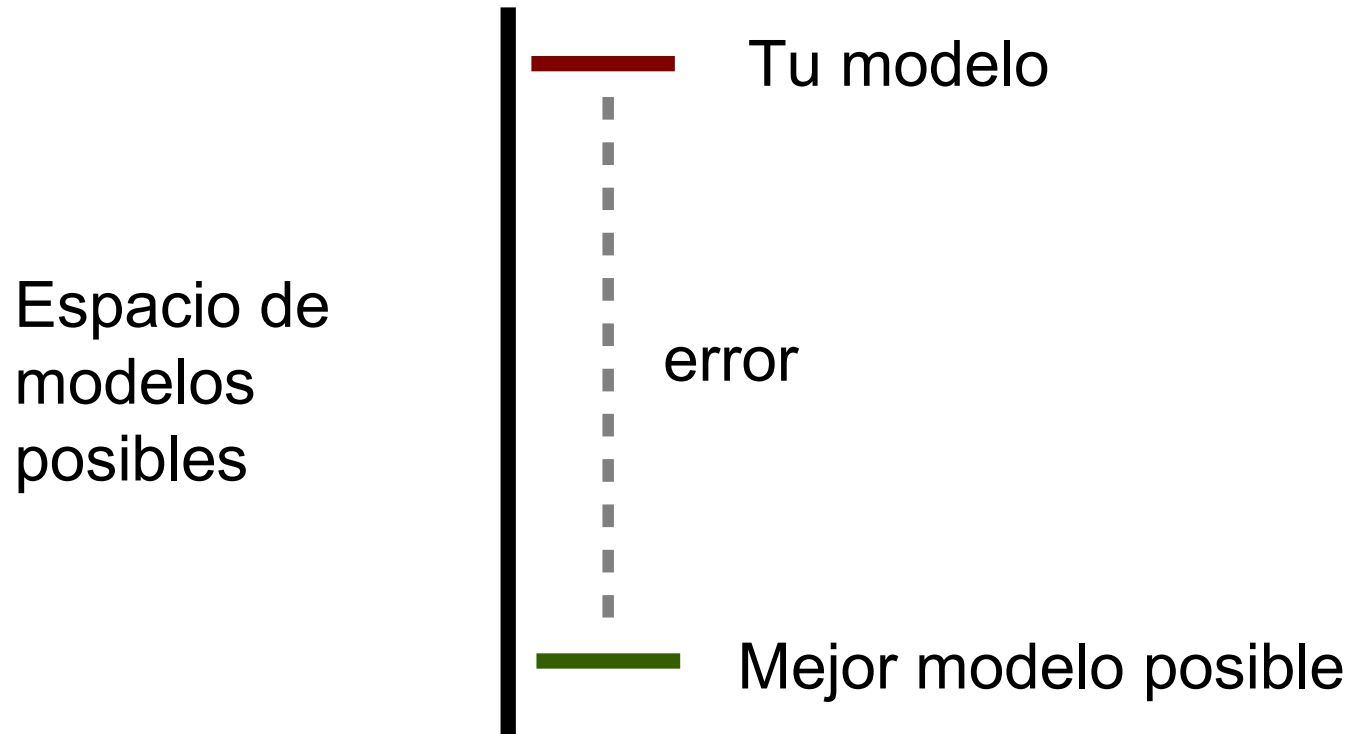


ENSAMBLADO

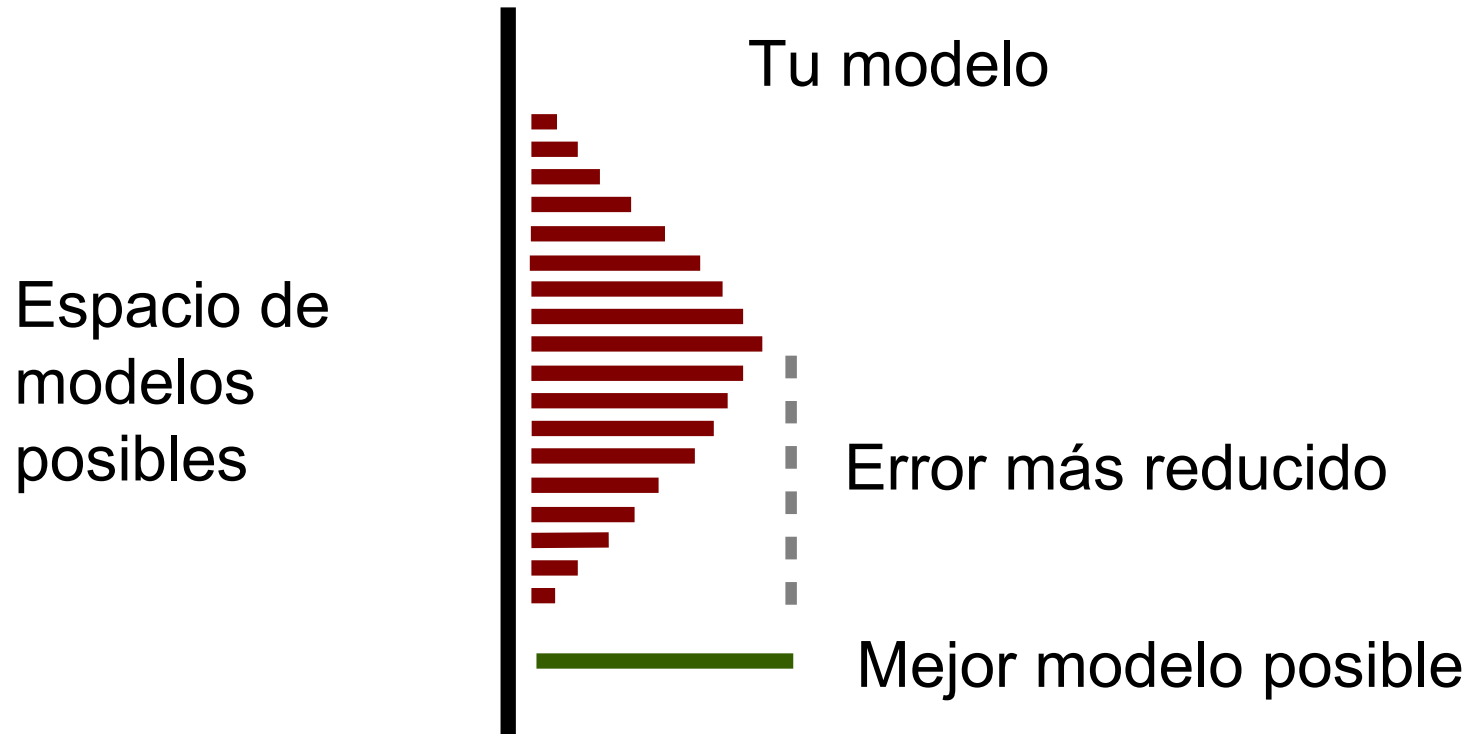
Espacio de
modelos
posibles



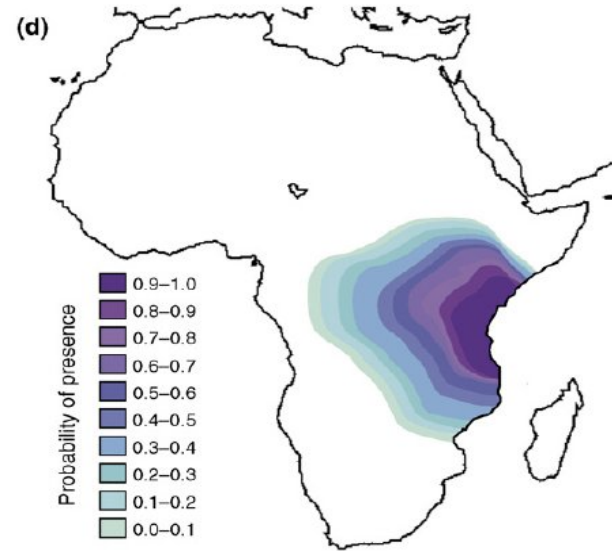
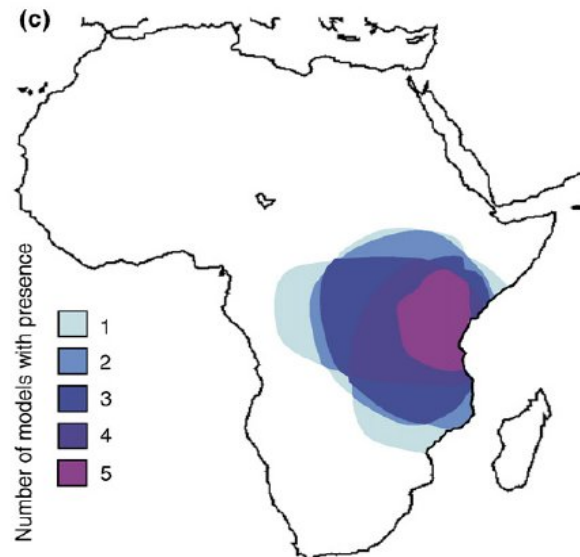
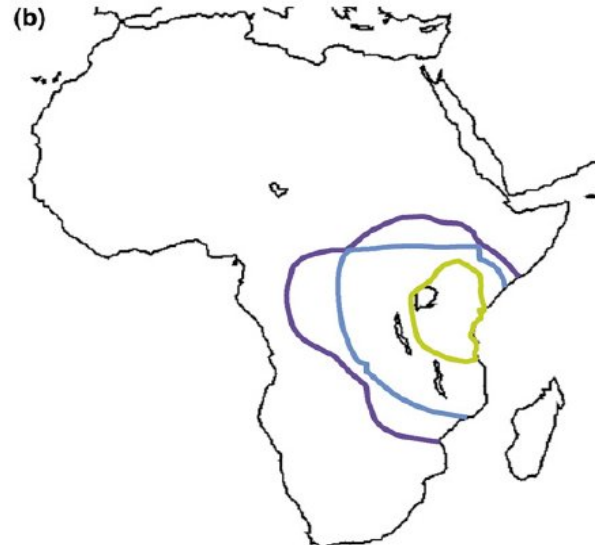
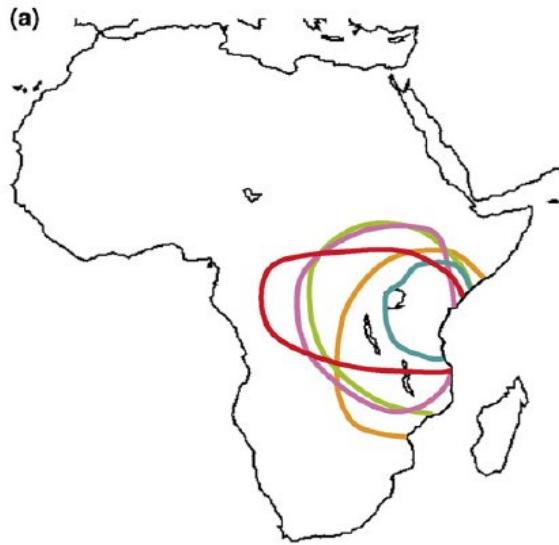
ENSAMBLADO



ENSAMBLADO



MÉTODOS DE ENSAMBLADO



MÉTODOS DE ENSAMBLADO

- Mediana **
- Media aritmética ** (buen método, Marmion 2009)
- Media ponderada según valores de AUC **
- Selección de modelos con mayor AUC
- PCA: primer componente refleja la tendencia general. Se seleccionan los modelos más relacionados con este componente, y se les calcula la mediana

Ojo con las escalas de valores de los modelos, deben ser todas iguales!

EVALUACIÓN DE MODELOS

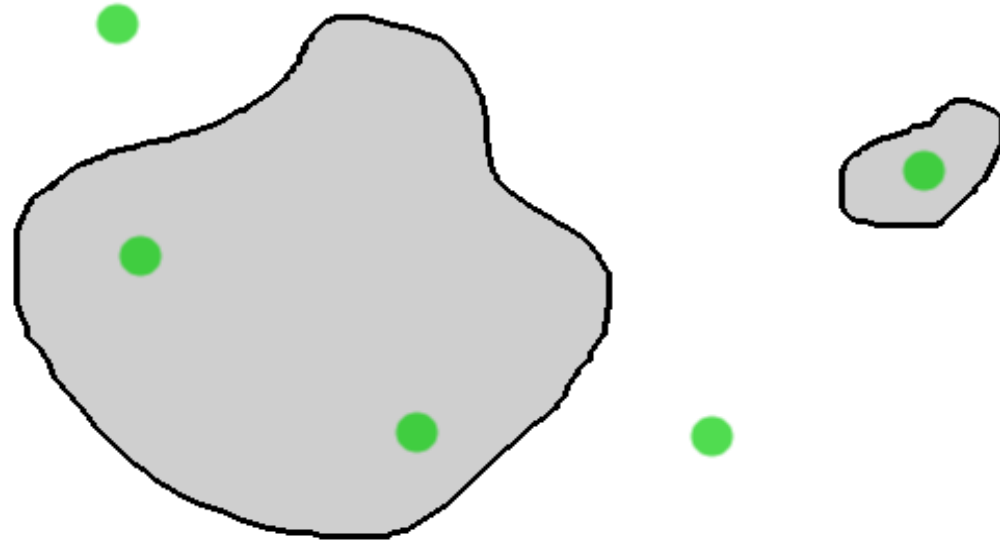
EVALUACIÓN

Artículo clave:

Fielding AH y Bell JF 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24(1), 38-49
(2856 citas en abril de 2014)

SOLO-PRESENCIA EN MODELOS BINARIOS

EVALUACIÓN



5 presencias

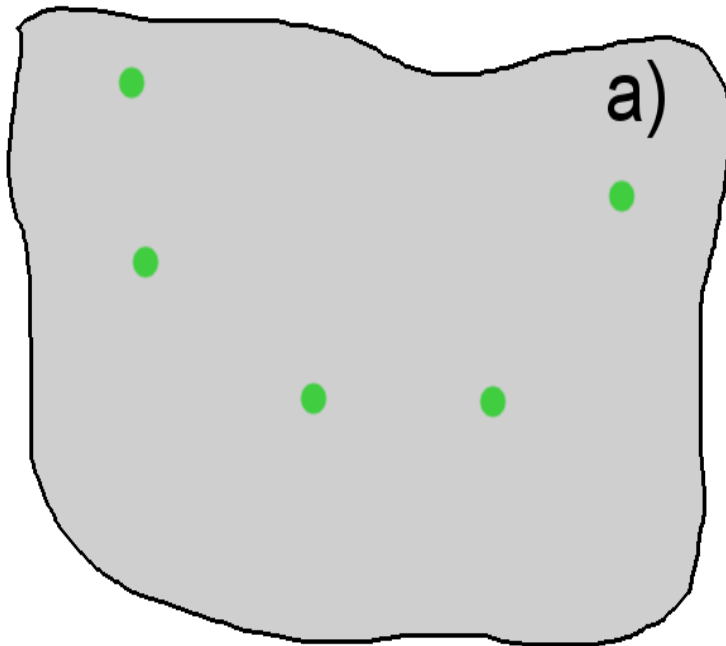
3 aciertos

Sensibilidad=0,6

2 errores de OMISIÓN

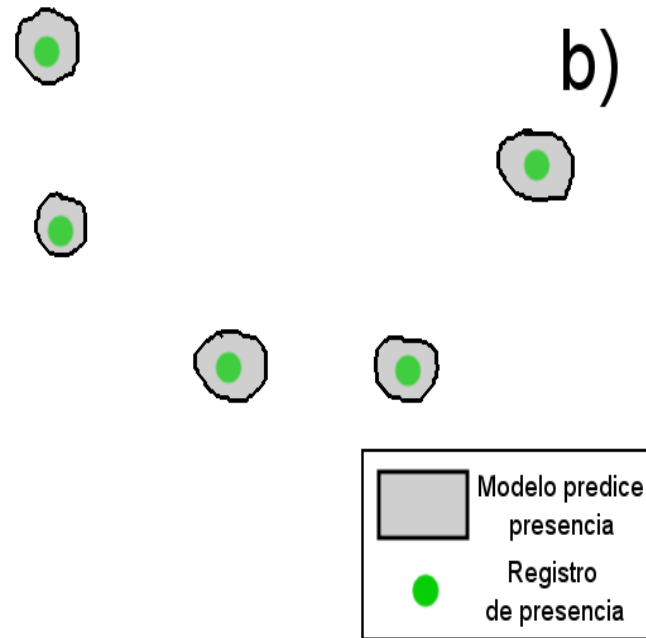


EVALUACIÓN



Sensibilidad=1

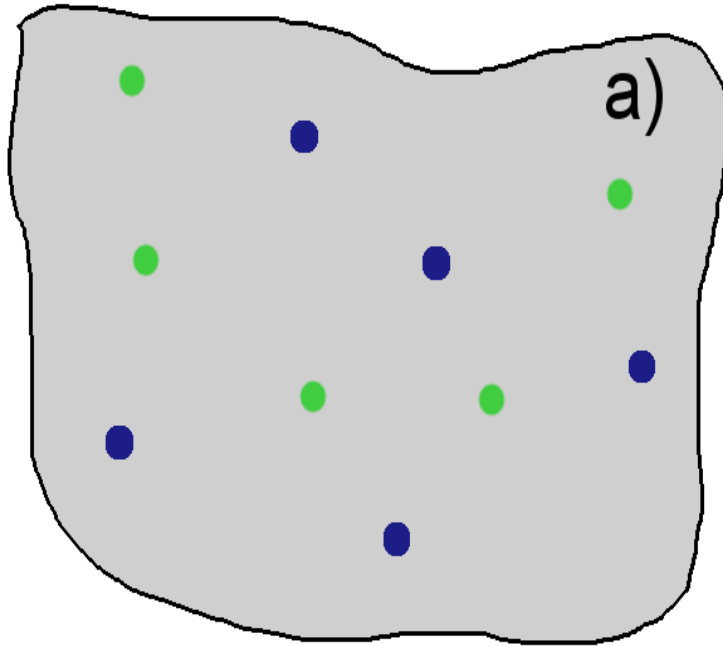
¿Error de comisión?



Sensibilidad=1

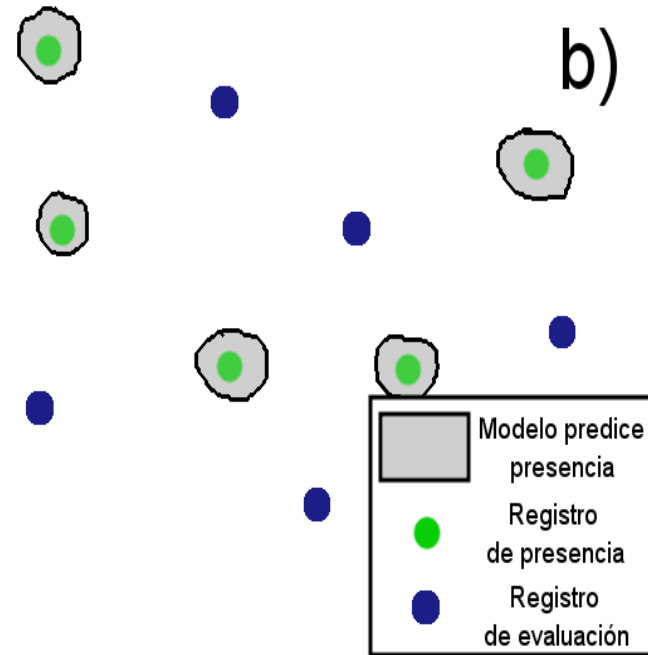
¿Sobreajuste?

EVALUACIÓN



Sensibilidad=1

¿?



Sensibilidad=0

¡Sobreajuste!

PRESENCIA – AUSENCIA EN MODELOS BINARIOS

MATRIZ DE CONFUSIÓN

A → presencias acertadas

D → ausencias acertadas

B → ausencias fallidas (falsos positivos o **error de comisión**)

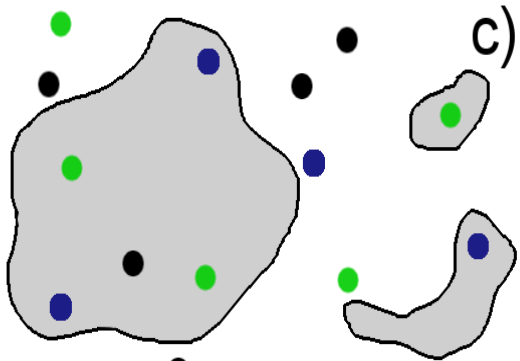
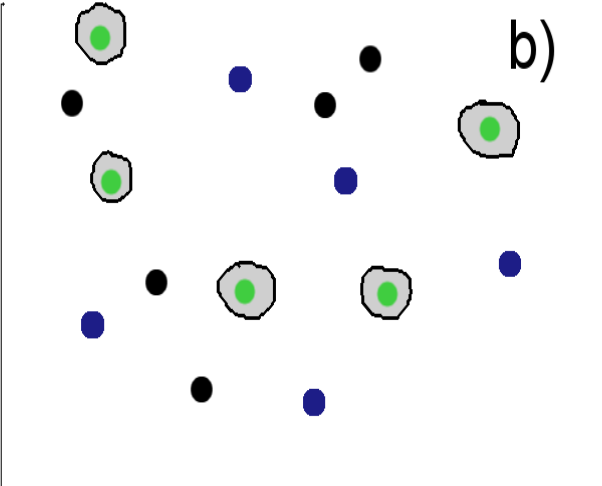
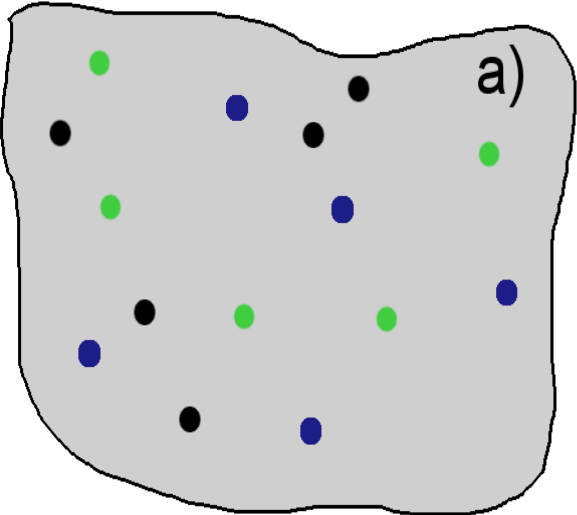
C → presencias fallidas (falsos negativos o **error de omisión**)





		Datos reales (registros de presencia y ausencia)	
		presencia	ausencia
Datos simulados (modelo de distribución)	presencia	A	B
	ausencia	C	D

$$\text{SENSIBILIDAD} = A/(A+C)$$

$$\text{ESPECIFICIDAD} = D/(B+D)$$

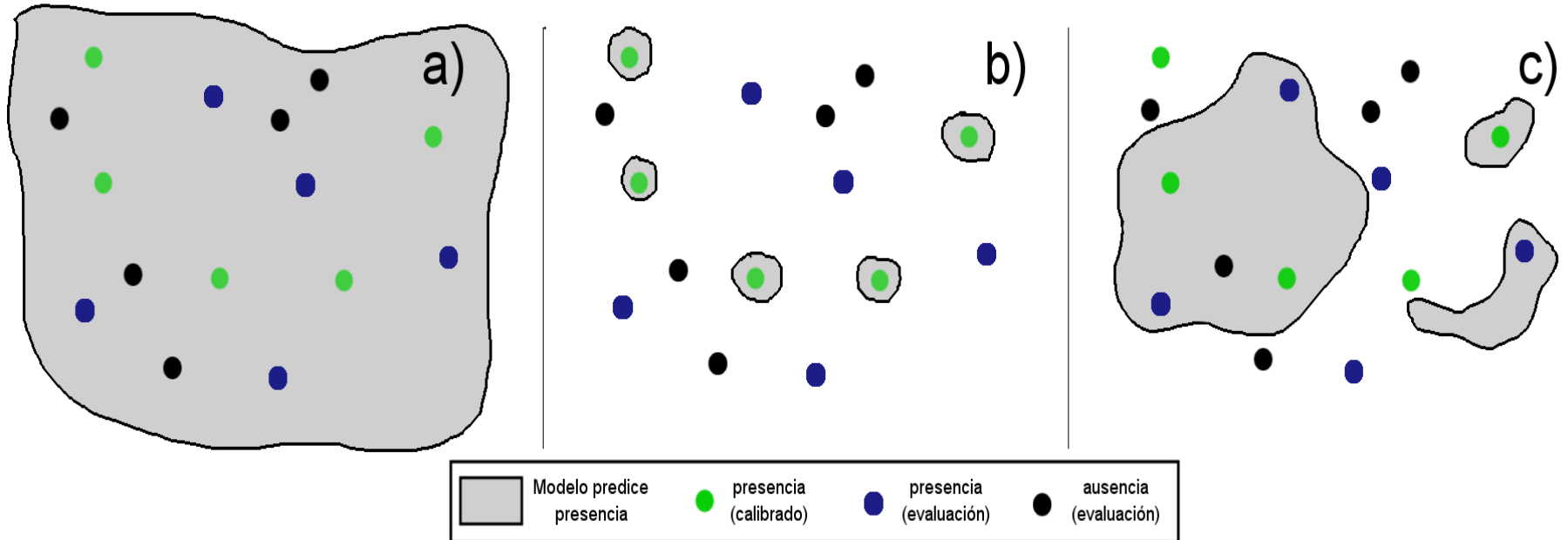
EVALUACIÓN



 Modelo predice presencia	 presencia (calibrado)	 presencia (evaluación)	 ausencia (evaluación)
--	---	--	---

		Datos reales (registros de presencia y ausencia)					
		modelo a		modelo b		modelo c	
		<u>pres.</u>	<u>aus.</u>	<u>pres.</u>	<u>aus.</u>	<u>pres.</u>	<u>aus.</u>
Datos simulados (modelo de distribución)	presencia	5	5	0	0	3	1
	ausencia	0	0	5	5	2	4

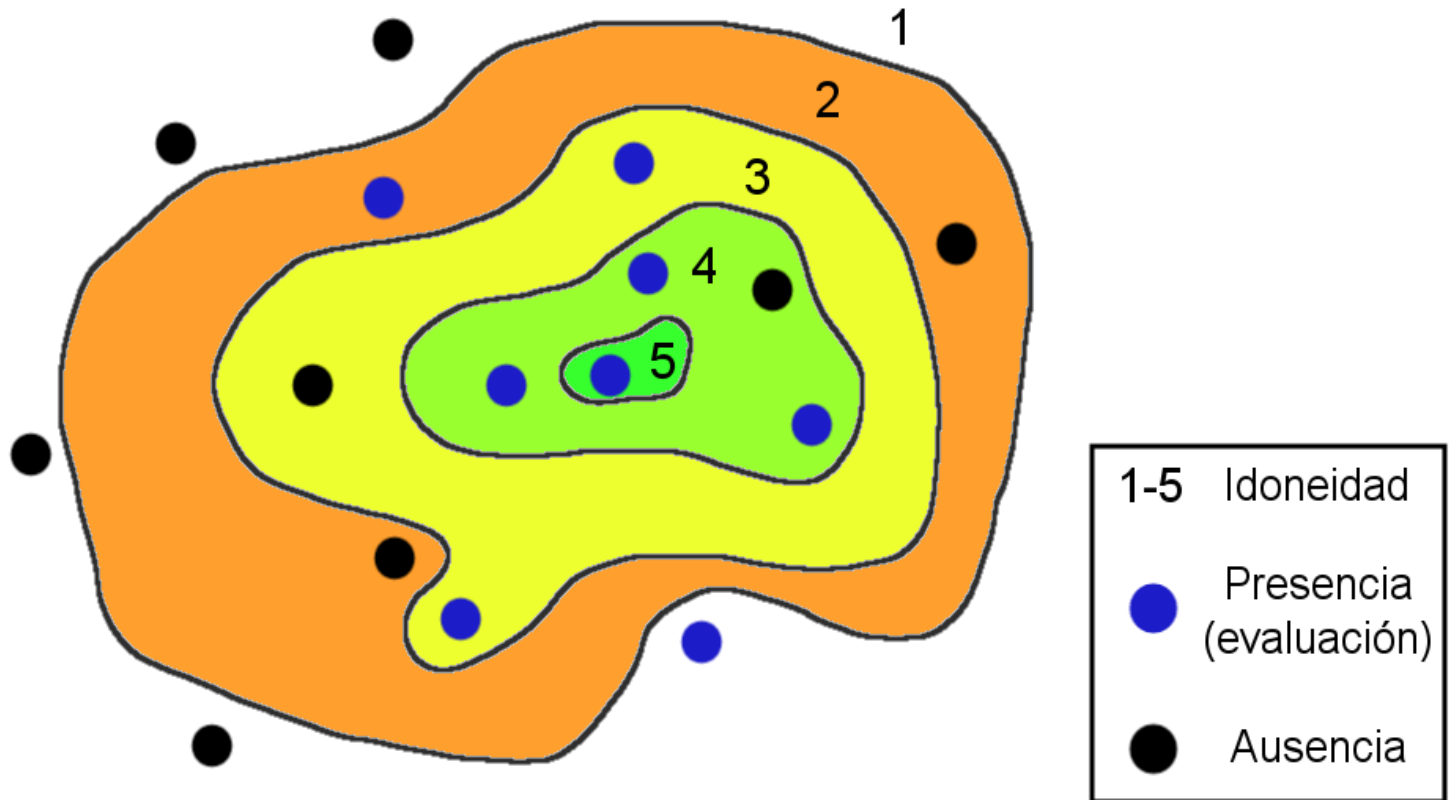
EVALUACIÓN



<u>modelo</u>	a	b	c
<u>sensibilidad</u>	1	0	0.6
<u>especificidad</u>	0	1	0.8

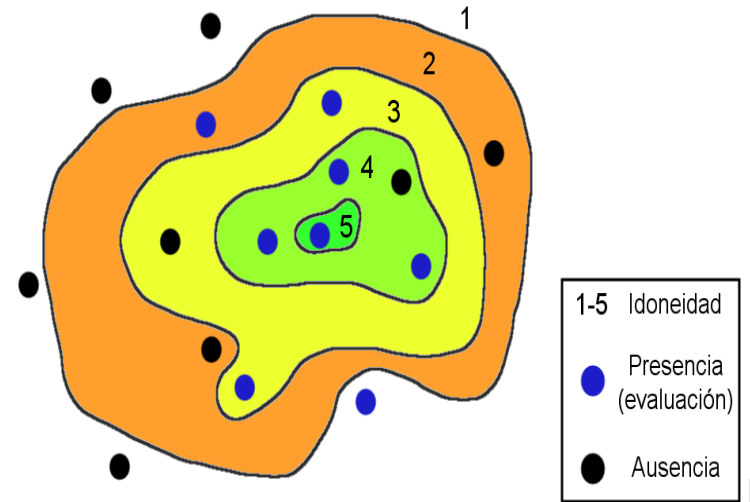
PRESENCIA – AUSENCIAS EN MODELO CONTINUO

CURVA ROC



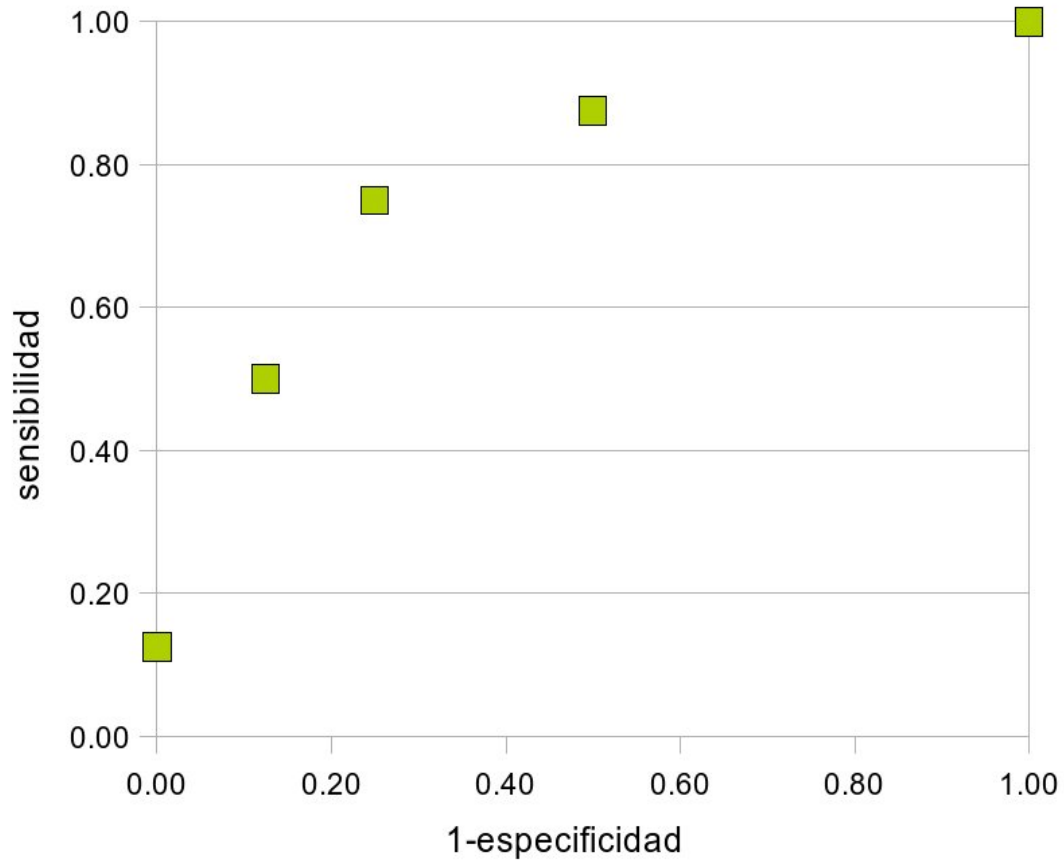
CURVA ROC

Observa que en lugar de la especificidad, usamos 1-especificidad

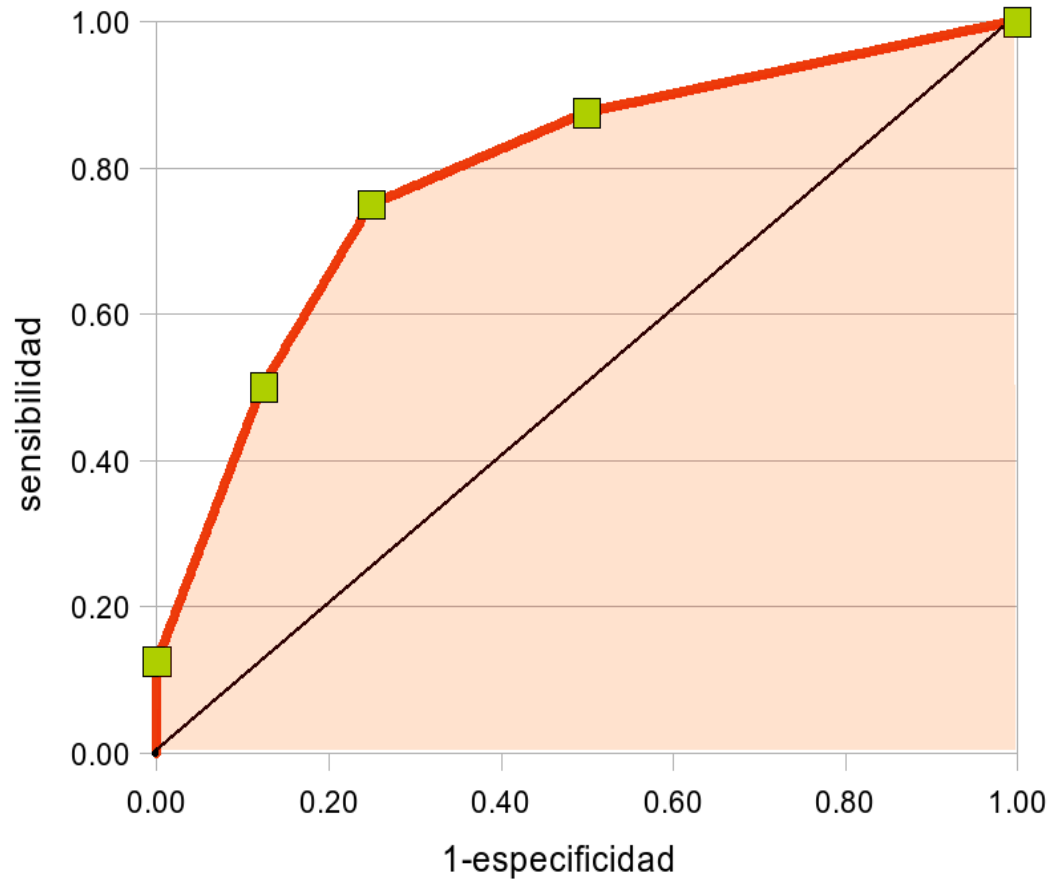


presencias acertadas	ausencias falladas	presencias falladas	ausencias acertadas	tamaño de muestra	UMBRAL CORTE	SENSIBILIDAD	1-ESPECIFICIDAD
A	B	C	D	N			
8	8	0	0	16	1	1.00	1.00
7	4	1	4	16	2	0.88	0.50
6	2	2	6	16	3	0.75	0.25
4	1	4	7	16	4	0.50	0.13
1	0	7	8	16	5	0.13	0.00

CURVA ROC



CURVA ROC



CURVA ROC

Area Under the Curve ROC (“Receiver Operating Characteristic”) -> probabilidad de que, seleccionando al azar una presencia y una ausencia, el modelo clasifique con un valor de idoneidad mayor a la presencia que a la ausencia.

Suponiendo $AUC = 0.74$, el modelo dará mayor valor de idoneidad a las presencias un 74% de las veces

PRESENCIA – ALEATORIOS EN MODELO CONTINUO

MATRIZ DE CONFUSIÓN MODIFICADA

A → presencias acertadas

D → ya no es un acierto

B → ya no es un error

C → presencias fallidas (falsos negativos o **error de omisión**)

		Datos reales (registros de presencia y puntos aleatorios)	
		presencia	aleatorio
Datos simulados (modelo de distribución)	presencia	A	B
	ausencia	C	D

CURVA ROC CON PUNTOS ALEATORIOS

- Cambia el significado: AUC es la probabilidad de que un punto de presencia seleccionado al azar tenga un valor de idoneidad más alto que el de un punto aleatorio seleccionado al azar.
- Pero ahora AUC siempre será menor que 1, porque siempre habrá puntos aleatorios sobre áreas de hábitat idóneo.



AUC: a misleading measure of the performance of predictive distribution models

Jorge M. Lobo^{1*}, Alberto Jiménez-Valverde¹ and Raimundo Real²

¹*Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales (CSIC), Madrid, Spain,* ²*Laboratorio de Biogeografía, Diversidad y Conservación, Departamento de Biología Animal, Facultad de Ciencias, Universidad de Málaga, Spain*

ABSTRACT

The area under the receiver operating characteristic (ROC) curve, known as the AUC, is currently considered to be the standard method to assess the accuracy of predictive distribution models. It avoids the supposed subjectivity in the threshold selection process, when continuous probability derived scores are converted to a binary presence–absence variable, by summarizing overall model performance over all possible thresholds. In this manuscript we review some of the features of this measure and bring into question its reliability as a comparative measure of accuracy between model results. We do not recommend using AUC for five reasons: (1) it ignores the predicted probability values and the goodness-of-fit of the model; (2) it summarises the test performance over regions of the ROC space in which one would rarely operate; (3) it weights omission and commission errors equally; (4) it does not give information about the spatial distribution of model errors; and, most importantly, (5) the total extent to which models are carried out highly influences the rate of well-predicted absences and the AUC scores.

Keywords

AUC, distribution models, ecological statistics, goodness-of-fit, model accuracy, ROC curve.

*Correspondence: Jorge M. Lobo,
Departamento de Biodiversidad y Biología
Evolutiva, Museo Nacional de Ciencias
Naturales (CSIC), Madrid, Spain.
E-mail: mcnj117@mncn.csic.es

ALGUNOS PROBLEMAS DE AUC

- Considera regiones del espacio ROC en los que no se trabaja, como los extremos de la curva, en los que las tasas de error son elevadas.
- Pondera por igual los errores de comisión y omisión.
- No informa de distribución espacial de los errores.
- Las áreas de trabajo amplias resultan en valores de AUC más altos.
- No pueden compararse modelos de distintas especies.

MÉTODOS DE EVALUACIÓN DE MODELOS

- Datos independientes: 1 solo valor de AUC por modelo
- Sin datos independientes: Cross validation
 - Data splitting: separas un conjunto de datos para calibrar el modelo, y otro para evaluar
 - K-fold: separación de los datos en n grupos. Calibras con $n-1$ y evalúas con 1.
 - Bootstrap: partición iterativa de los datos, para calibrar con unos, y evaluar con otros.
 - Leave-one-out: para muestras pequeñas.

AÚN ASÍ...

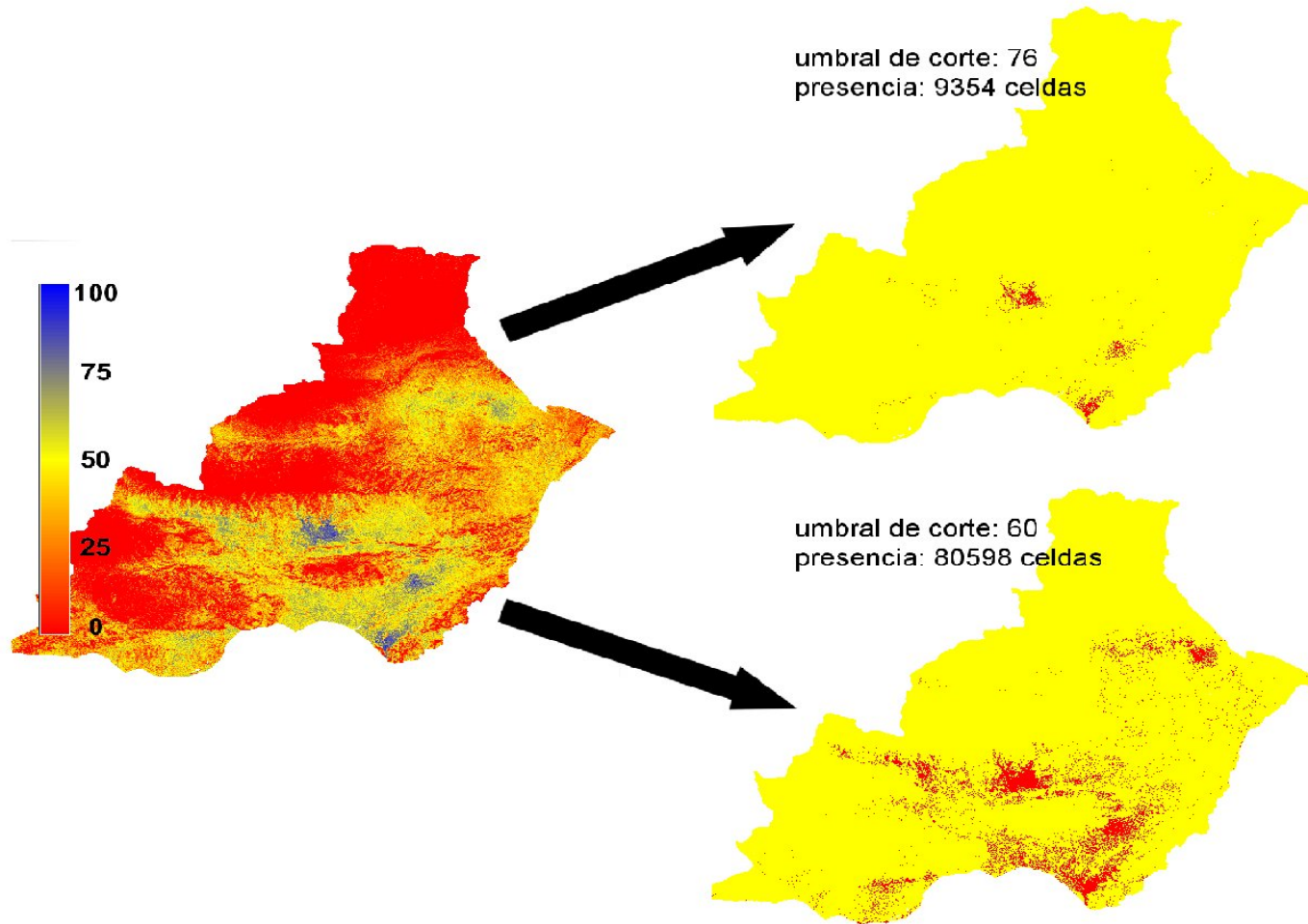
Es una buena herramienta para
comparar modelos para la misma
especie y área de trabajo

APLICACIÓN DE THRESHOLDS

DE CONTINUO A BINARIO

- Un mapa de valores binarios (0 o 1) es más fácil de comprender que uno continuo (0 a 1)
- Para transformar MDE continuos en binarios:
 - seleccionamos un valor de referencia: umbral (threshold)
 - asignamos valor 1 a todas las celdas por encima del umbral
 - asignamos valor 0 a todas las celdas por debajo del umbral

DE CONTINUO A BINARIO



DE CONTINUO A BINARIO

- ¿Cómo seleccionamos el umbral?...
 - Liu et al. 2005
 - Jiménez-Valverde y Lobo 2007
 - Freeman y Moisen 2008
 - Selección subjetiva
 - Selección “objetiva”

SELECCIÓN SUBJETIVA

“Elecciones arbitrarias sin base ecológica” (Osborne et al. 2001)

- Valores fijos: 0.5, 0.3, ...
- Porcentaje de comisión: 95%, 90%, ...

SELECCIÓN OBJETIVA

“El umbral se selecciona para maximizar la concordancia entre la distribución observada y la modelada” (Liu et al 2005)

- maximización de Kappa (**no recomendado**)
- punto de curva ROC con pendiente = 1
- valor con igual sensibilidad y especificidad
- y muchos más en Freeman y Moisen 2008

Estos criterios requieren datos de presencia - ausencia!

'COSAS' IMPORTANTES

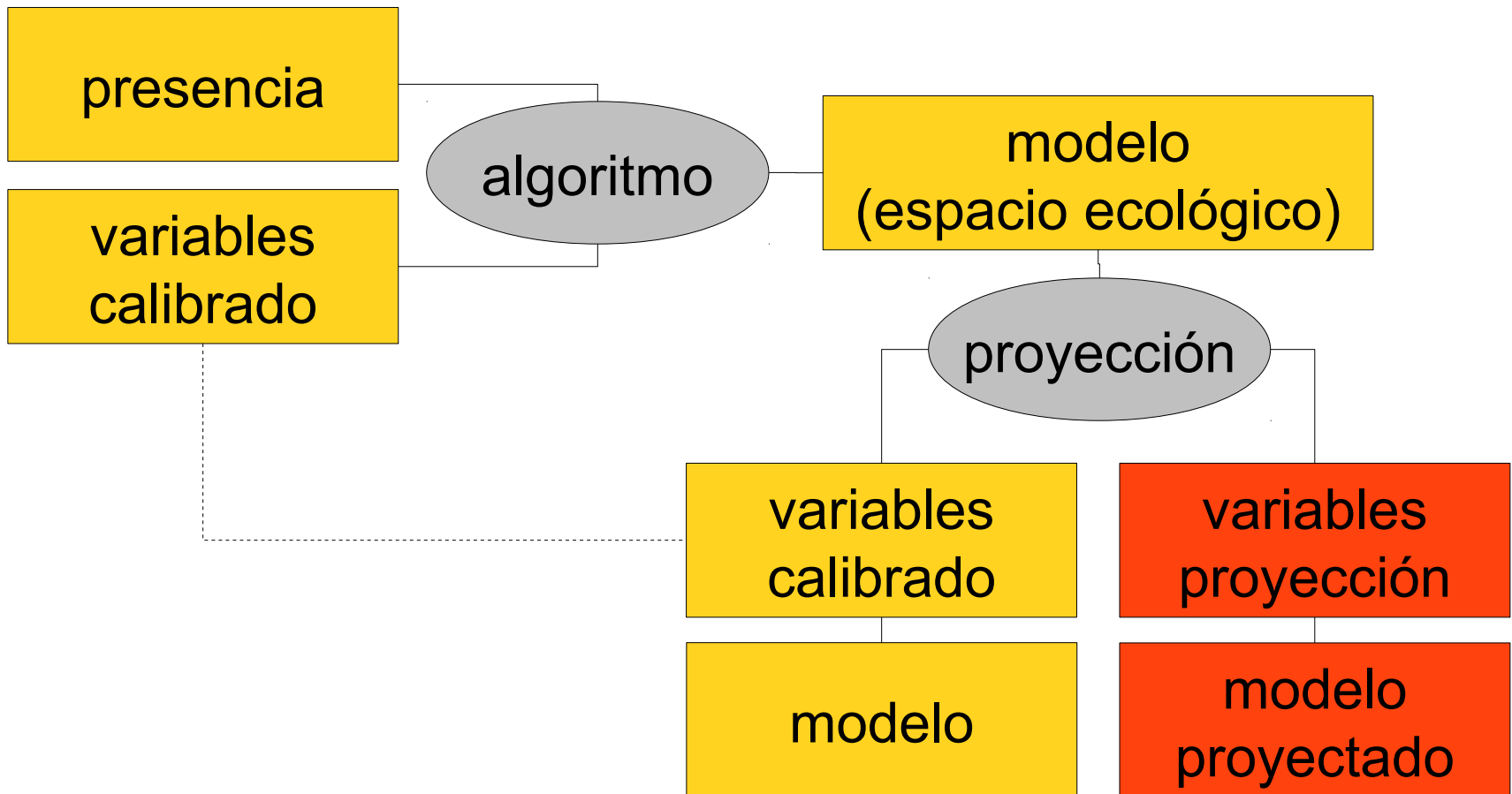
- La elección del umbral depende del objetivo del modelo, no hay una norma fija
- Los modelos de especies con pocas presencias o mal ajuste son muy sensibles a la elección del threshold
- No hay obligación de aplicar un threshold a un modelo, salvo que tu análisis concreto lo requiera
- Siempre es mejor usar la versión continua de un modelo

THRESHOLD CON R

- Con R podemos calcular las estadísticas de las presencias sobre el modelo para tomar decisiones
- También podemos ver el gráfico de densidad los valores de las presencias sobre el modelo con 'extract', 'density' y 'plot' para decidir manualmente un punto de corte
- La función 'evaluate' de dismo ofrece herramientas para calcular thresholds

PROYECCIÓN DE MODELOS DE DISTRIBUCIÓN EN EL ESPACIO Y EN EL TIEMPO

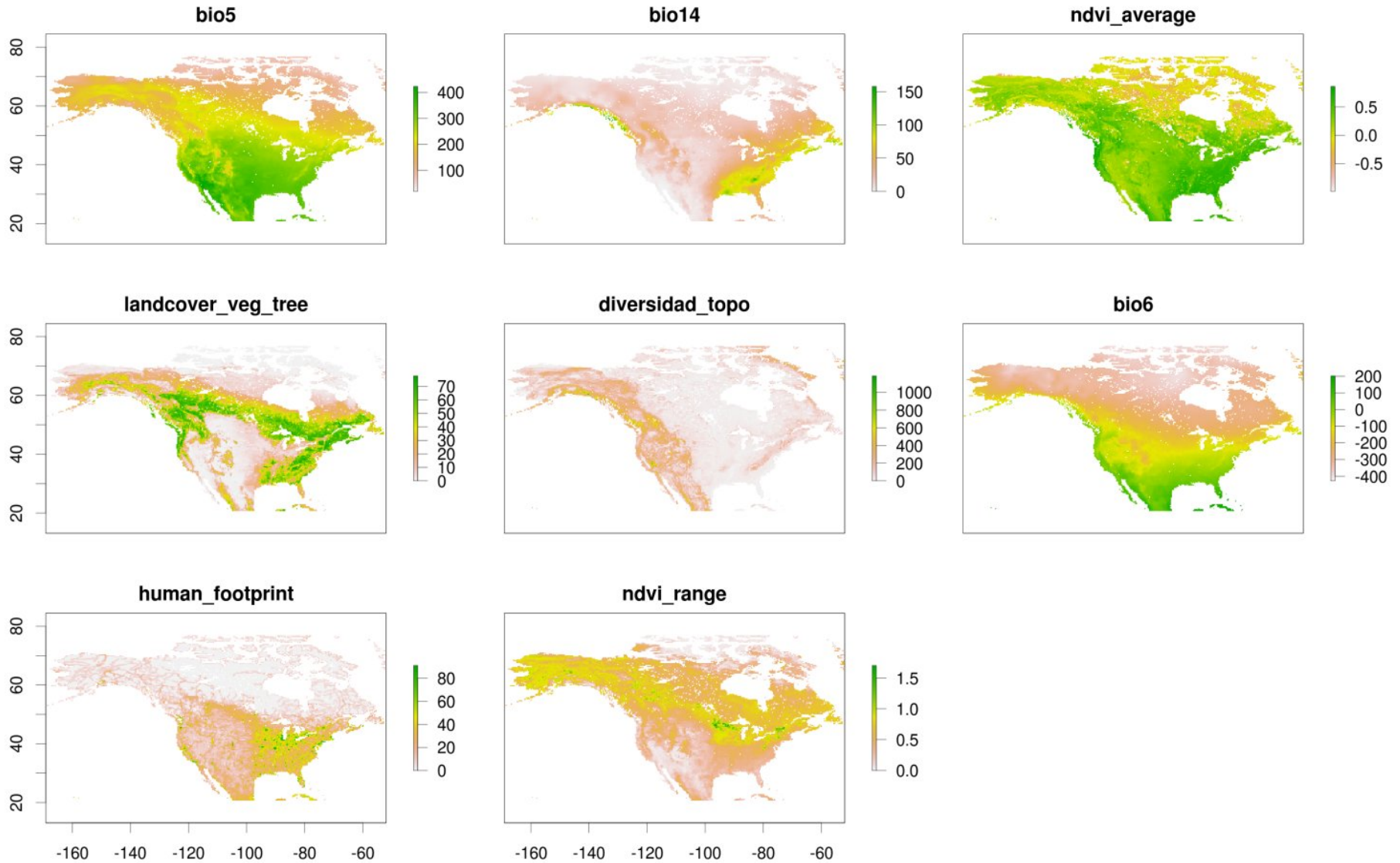
PROYECCIÓN DE MODELOS



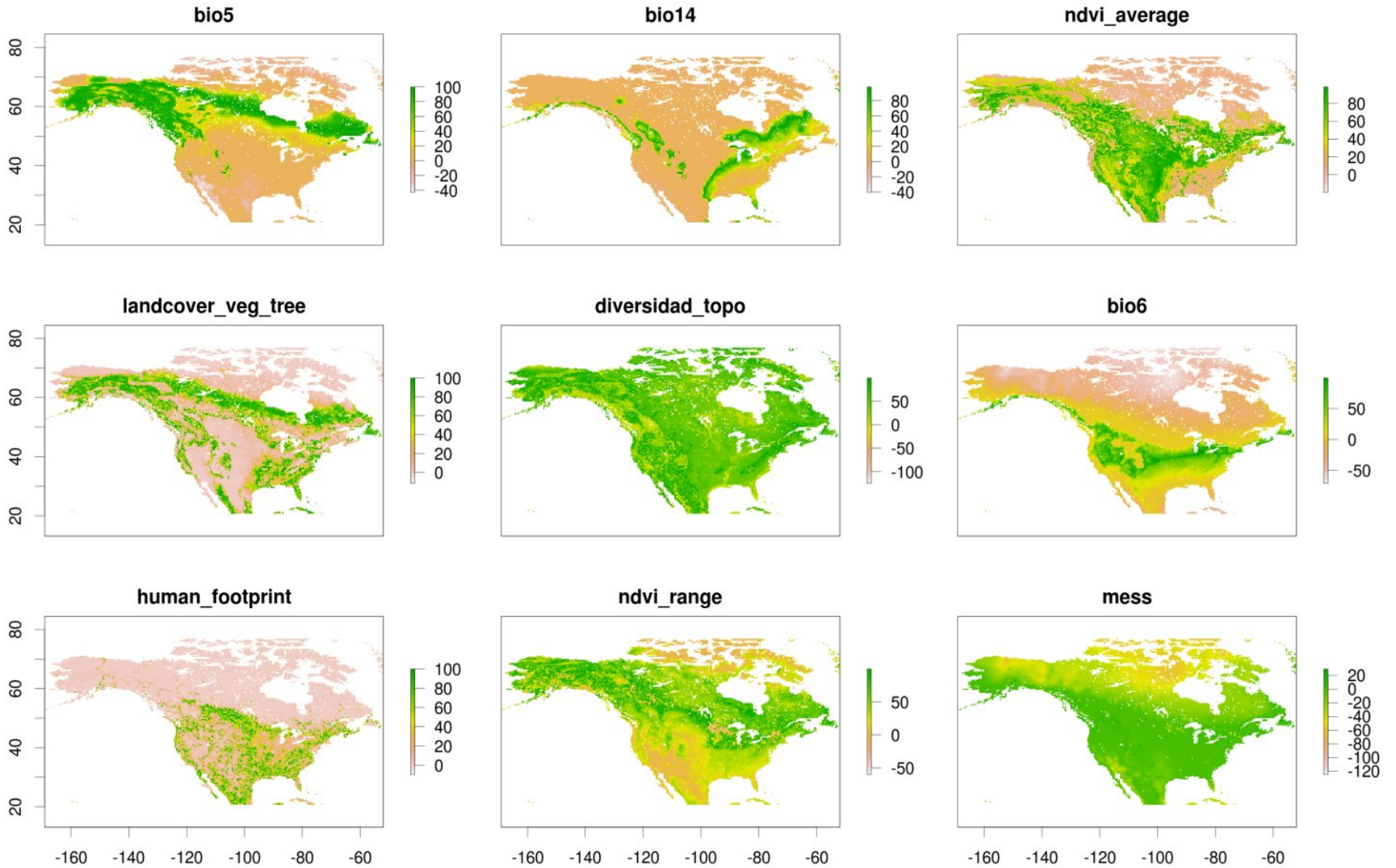
MULTIVARIATE ENVIRONMENTAL SIMILARITY SURFACES (MESS)

- Índice de similaridad entre el espacio ecológico que ocupan las presencias, y el espacio ecológico que hay en las variables de proyección.
- Cuanto más diferente sean, más estaremos extrapolando, ¡¡Y EXTRAPOLAR ES MALO!!
- Referencia: Elith J., Kearney M., & Phillips S. 2010. The art of modelling range shifting species. *Methods in Ecology and Evolution*, 1 :330-342.

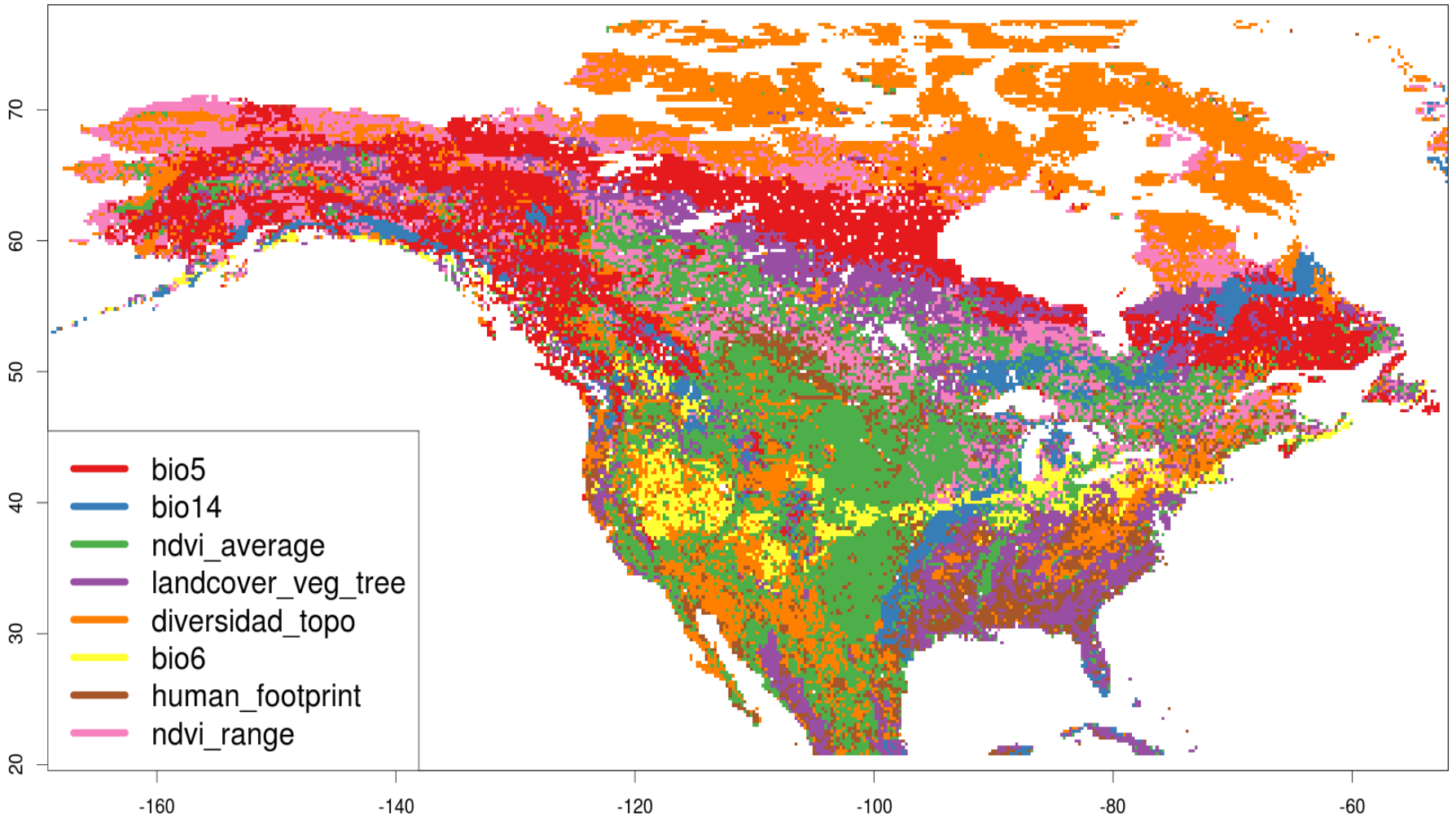
VARIABLES



MESS



MESS (máximo)



PROYECCIÓN EN EL ESPACIO

- Una región origen y una de destino
- Calibramos el modelo en la de origen
- Proyectamos el modelo en la de destino
- Necesitamos tener las mismas variables **CON LOS MISMOS NOMBRES** para ambas regiones, idealmente a la misma resolución.
- Uso más extendido: invasibilidad
- Problemas: a million...

PROYECCIÓN CON MAXENT (configuración)

- **Output directory:** taller2/resultados/
maxent/proyeccion_espacio
- **Projection layers directory:**
taller2/resultados/variables_norteamerica
- Desmarca **Auto features**
- Desmarca **Hinge features**
- Desmarca **Threshold features**
- Desmarca **Product features**

PROYECCIÓN EN EL TIEMPO

- Una región y tiempo de origen y un tiempo de destino (pasado o futuro)
- Calibramos el en origen y proyectamos en destino
- Mismos nombres de las variables
- Algunas variables no están disponibles para pasado o futuro (ndvi, human footprint, etc)
- Suelen hacerse solo con clima y topografía
- Uso más extendido: cambio climático, paleodistribuciones
- Problemas: a million...

FUENTES DE PALEOCLIMA

- **Último interglacial** (120 – 140 kyr BP) disponible en www.worldclim.org/past (resolución: 1 km)
- **Último máximo glacial** (21 kyr BP) según los modelos CCSM y MIROC, procedente de PMIP3 (pmip3.lsce.ipsl.fr) y disponible en www.worldclim.org/past (resolución: 5km)
- **Holoceno medio** (6 kyr BP) solo disponibles en PMIP3 (pmip3.lsce.ipsl.fr)
- **TraCE-21ka** (21 kyr BP → presente) disponible en www.cgd.ucar.edu/ccr/TraCE a resolución grosera (2°, solo para escalas continentales)

¿COMO PODEMOS EVALUAR ESTOS MODELOS?

- El AUC de un modelo actual no representa la capacidad predictiva del modelo en el pasado o el futuro
- Los modelos de paleodistribución de plantas se pueden evaluar con polen fósil y macrorrestos
- Los modelos de paleodistribución de animales se pueden evaluar con datos de registro fósil
- Los datos de evaluación y los modelos deben ser coetáneos.

ALGUNAS PREMISAS

- Una proyección NO representa la distribución futura o pasada (o en otro lugar) de una especie.
- Una proyección SOLO representa donde habrá condiciones ecológicas similares a aquellas en las que se ha observado la especie (¡siempre que el modelo no extrapole!) .
- Las proyecciones asumen que el nicho ecológico de las especies es constante.
- Los mapas climáticos del pasado o futuro son ESCENARIOS, no representan la realidad.

MAXENT

INTRODUCCIÓN

- OBJETIVO: hacer predicciones a partir de información incompleta
- Permite trabajar con un número bajo de registros de presencia
- Multiplataforma (requiere Java), código cerrado pero uso gratuito
- Ofrece muy buenos resultados
- Su interfaz gráfico es muy sencillo

RECURSOS WEB

- WEB

<http://www.cs.princeton.edu/~schapire/maxent/>

- Grupo de Google

<http://groups.google.com/group/Maxent?pli=1>

- Tutorial

<http://www.cs.princeton.edu/~schapire/maxent/tutorial/tutorial.doc>

<http://www.cs.princeton.edu/~schapire/maxent/tutorial/tutorial-data.zip>

PRINCIPIO DE MÁXIMA ENTROPÍA

Un modelo debe seleccionarse si es lo más parecido posible a una expectativa previa (prior expectation) siempre que sea consistente con los datos (Jaynes 2003)

En Maxent esta 'prior expectation' o hipótesis nula está predefinida: todas las celdas tienen la misma probabilidad de contener un individuo de la especie -> las especies usan el hábitat en función de su disponibilidad

CONCEPTOS CLAVE

- **Background** → muestreo aleatorio de celdas 'prior expectation'
- **Features** (features) → variables, transformaciones y combinaciones de las variables:
- **Regularización**: penalización LASSO para reducir sobreajuste
- **Gain** (penalized maximum likelihood function): gain más alto cuanto mejor diferencia el modelo entre presencia y background

ARTÍCULO CLAVE



EDITOR'S
CHOICE

Ecography 36: 1058–1069, 2013

doi: 10.1111/j.1600-0587.2013.07872.x

© 2013 The Authors. Ecography © 2013 Nordic Society Oikos

Subject Editor: Niklaus E. Zimmermann. Accepted 27 March 2013

A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter

Cory Merow, Matthew J. Smith and John A. Silander, Jr

C. Merow (cory.merow@gmail.com) and J. A. Silander, Jr, Univ. of Connecticut, Ecology and Evolutionary Biology, 75 North Eagleville Rd., Storrs, CT 06269, USA. – M. J. Smith and CM, Computational Ecology and Environmental Science Group, Computational Science Laboratory, Microsoft Research Ltd., 21 Station Road, Cambridge, CB1 2FB, UK.

BACKGROUND

- 'Prior expectation' por defecto → predicción del área de distribución más extensa consistente con los datos
- Cambiando la extensión del background, cambiamos la 'prior expectation', y por tanto, el resultado del modelo
- Recomendación: incluir en background las zonas que la especie puede alcanzar con su dispersión natural

FEATURES

- **Linear** → obliga a que el valor medio de una variable en la predicción coincida con su valor medio en las observaciones
- **Cuadratic** → como el anterior, con la varianza
- **Product** → equivalente a interacción en regresión
- **Threshold** → transforma predictor continuo en binario mediante una step function
- **Hinge** → Como el anterior, pero usa función lineal en lugar de una step function
- **Categorical** → transforma variable categórica en variables binarias (presencia-ausencia), una factor de la variable categórica

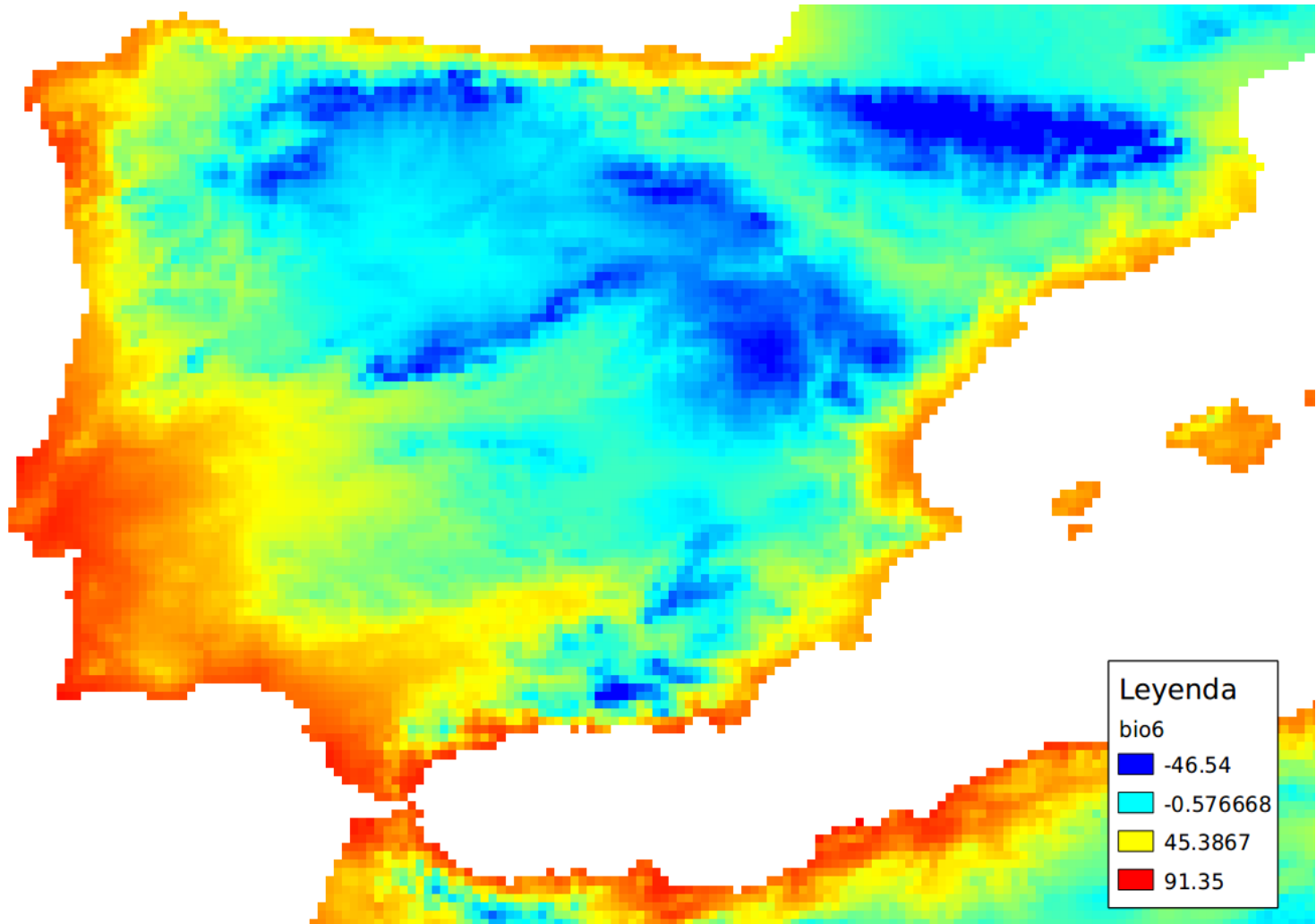
Los features que Maxent usa automáticamente dependen del número de presencias, pero se pueden modificar los valores

REGULARIZACIÓN

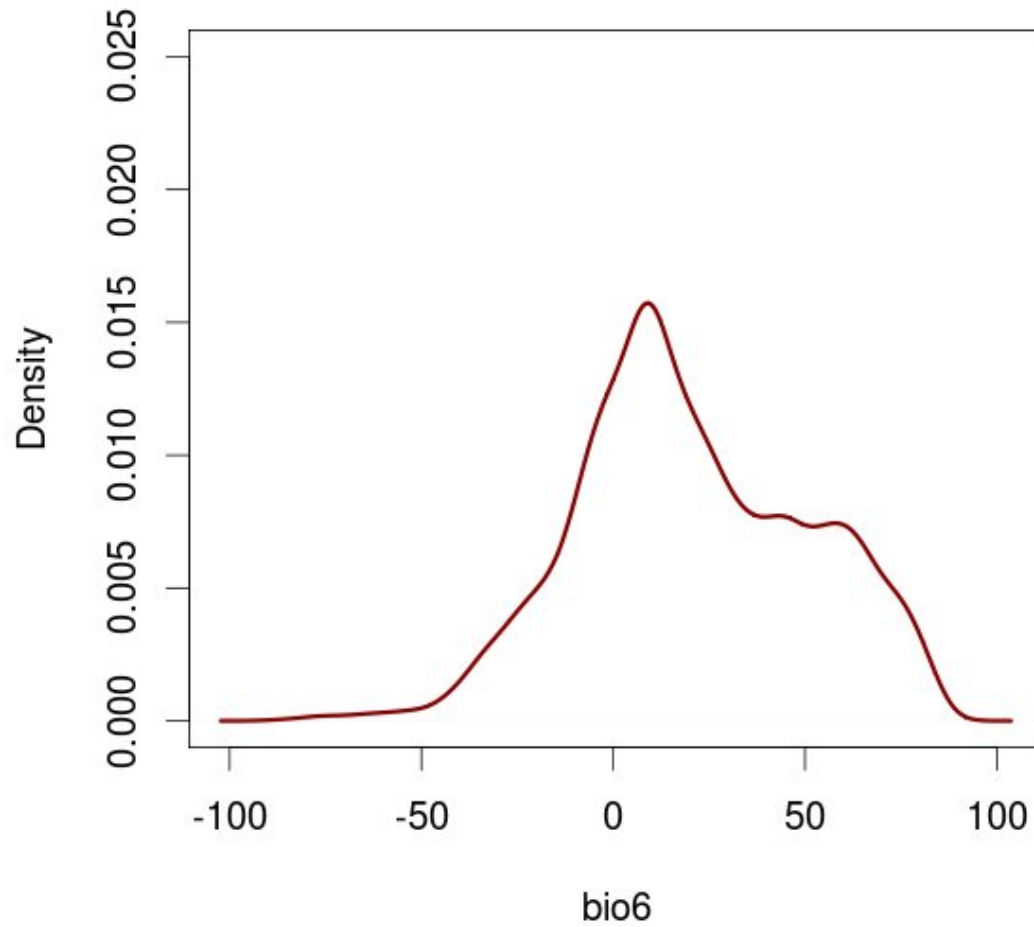
- Selecciona las 'features' que más contribuyen al ajuste del modelo, reduciendo el coeficiente de las demás a 0
- Evita un ajuste demasiado rígido a los datos (sobreajuste)
- Incrementar el factor de regularización disminuye la complejidad del modelo
- CONSEJO: probar distintos valores de regularización hasta obtener buenos resultados en validación con datos independientes

VARIABLE

Bio6 → temperatura media minima de enero

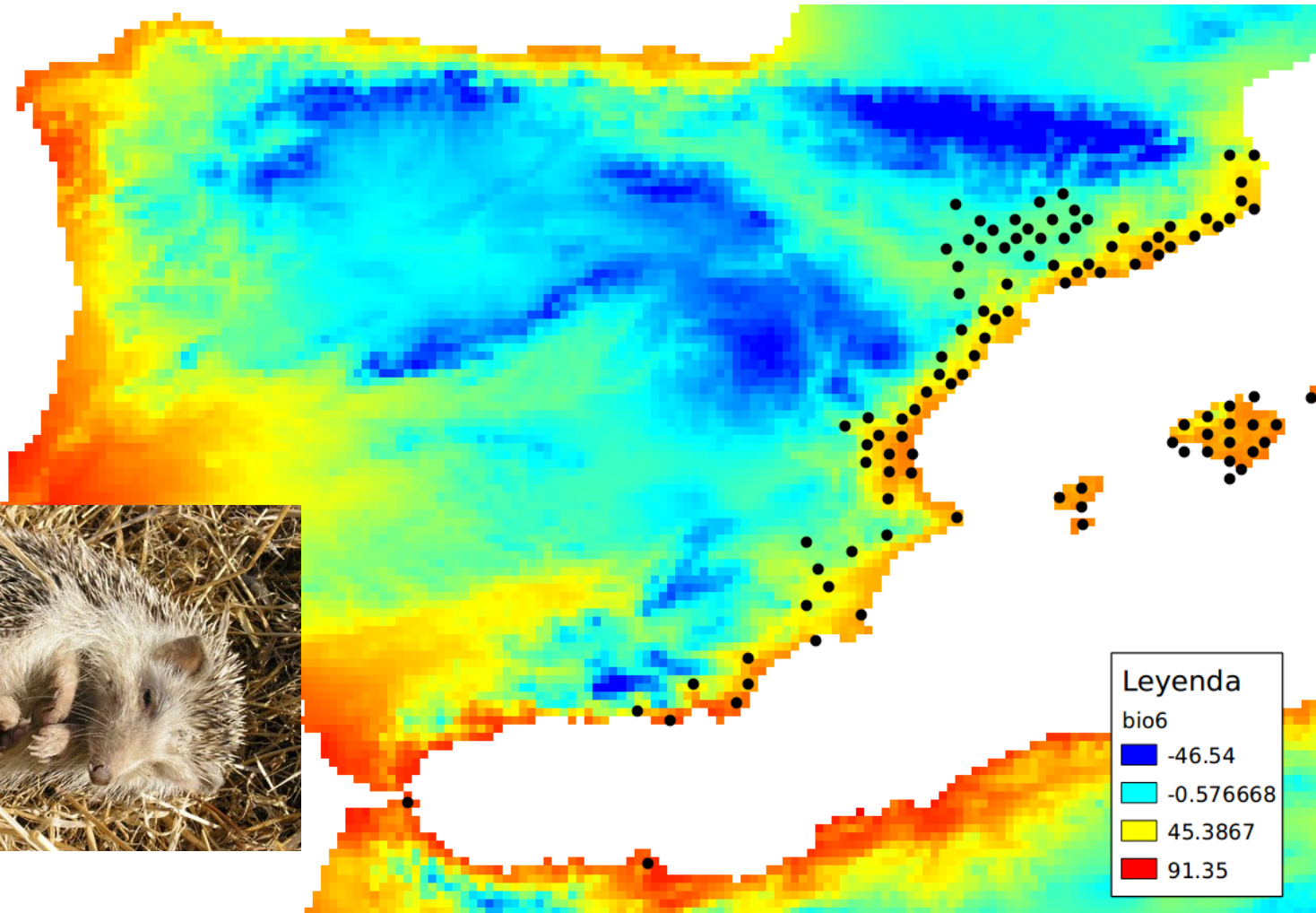


DENSIDAD DE LA VARIABLE

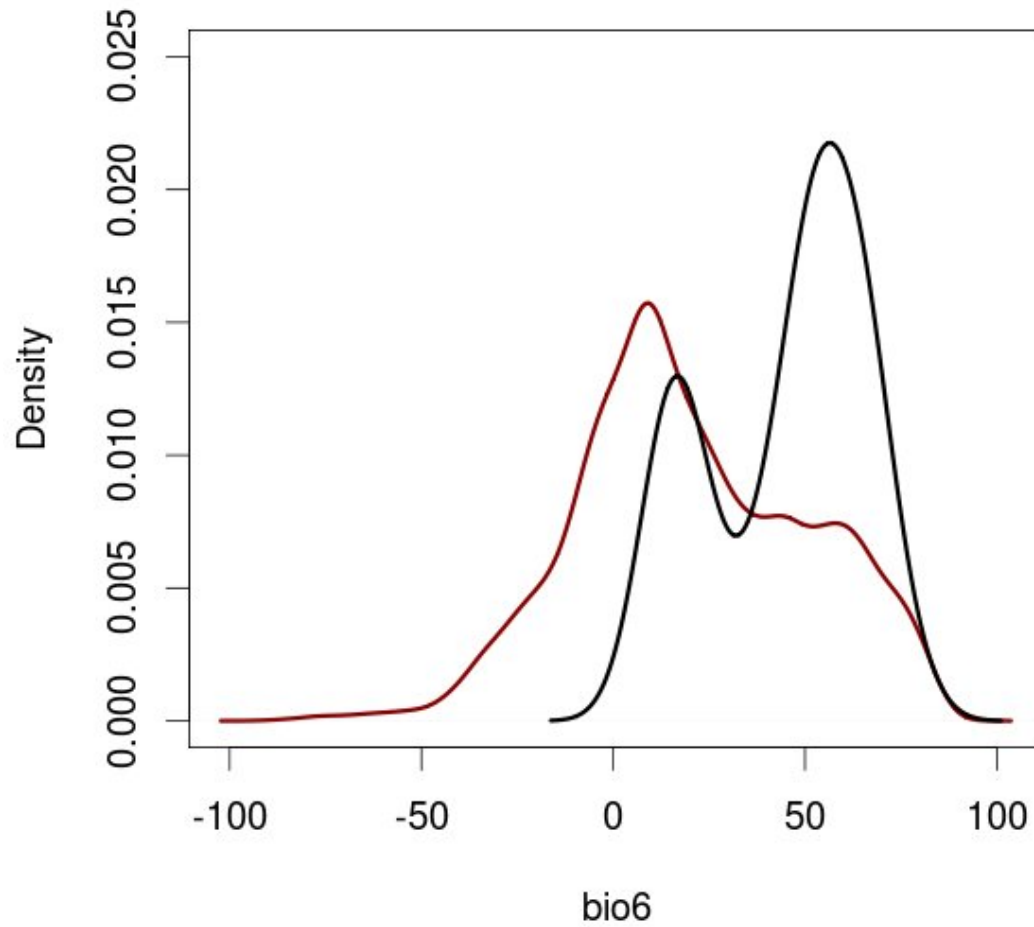


ESPECIE

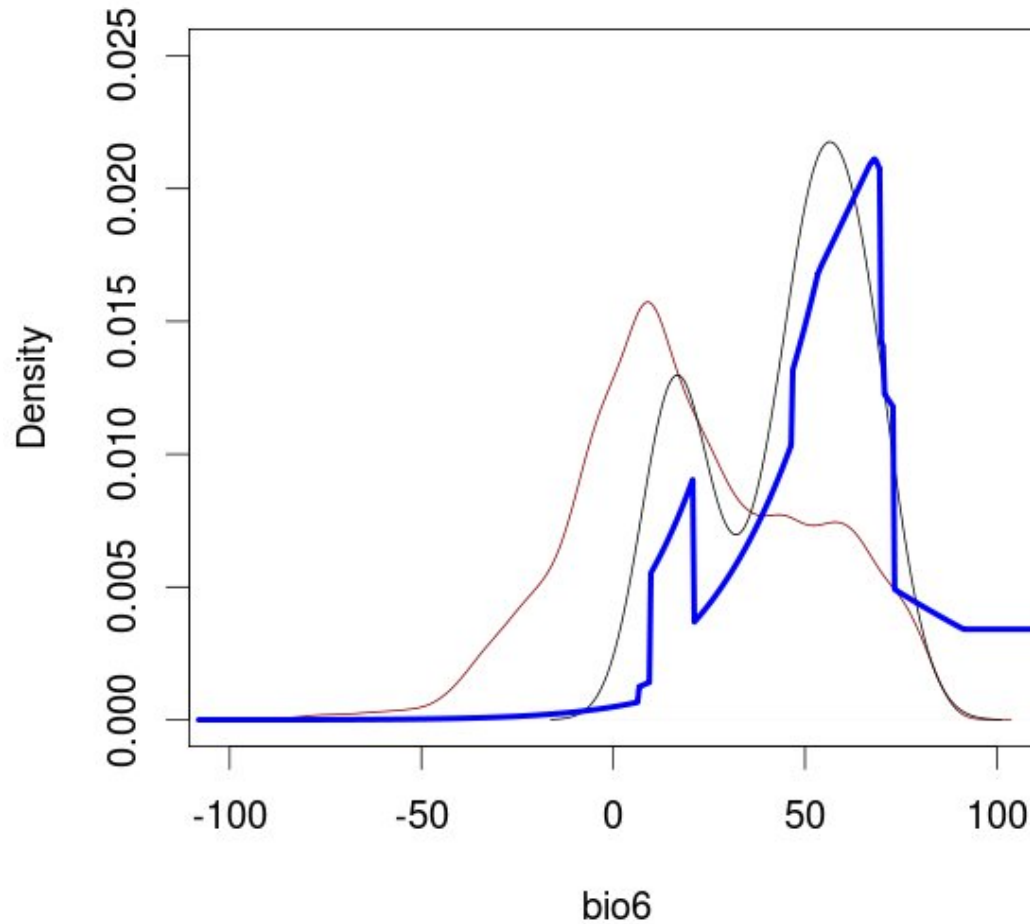
Atelerix algirus



DENSIDAD DE LA ESPECIE

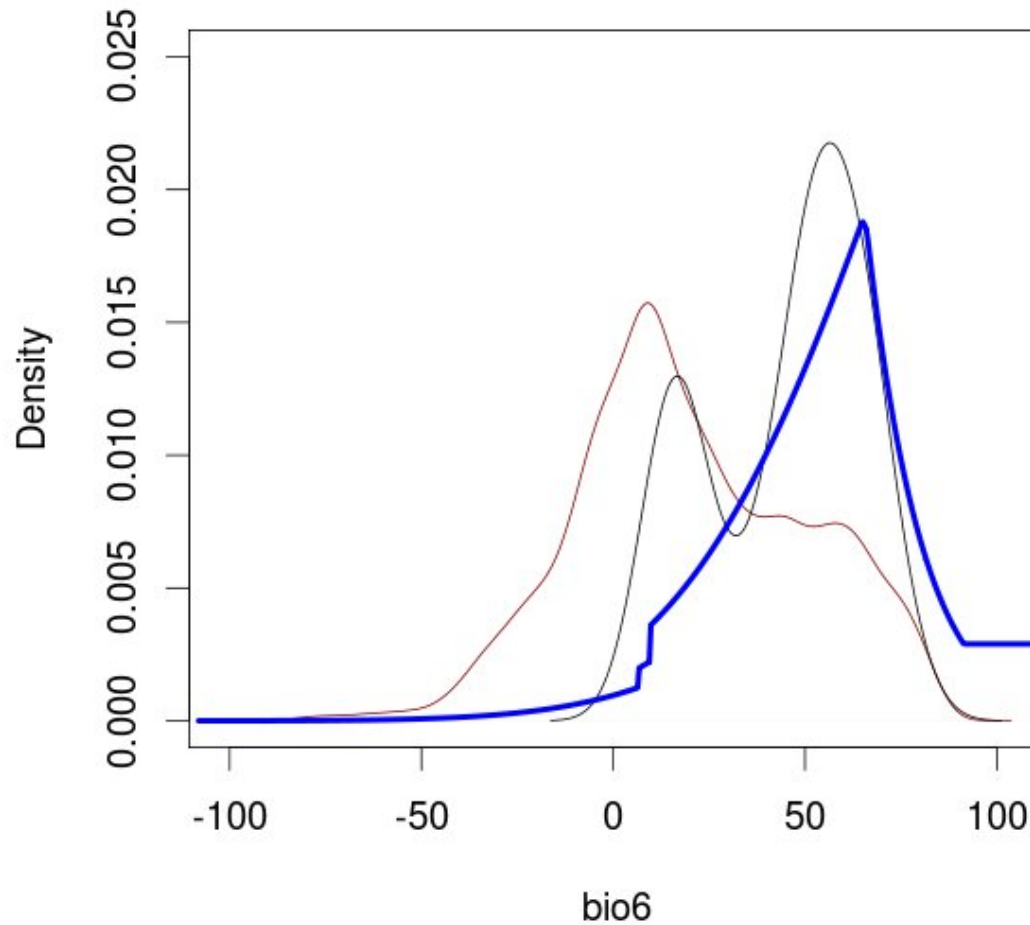


CURVA DE RESPUESTA DE MAXENT



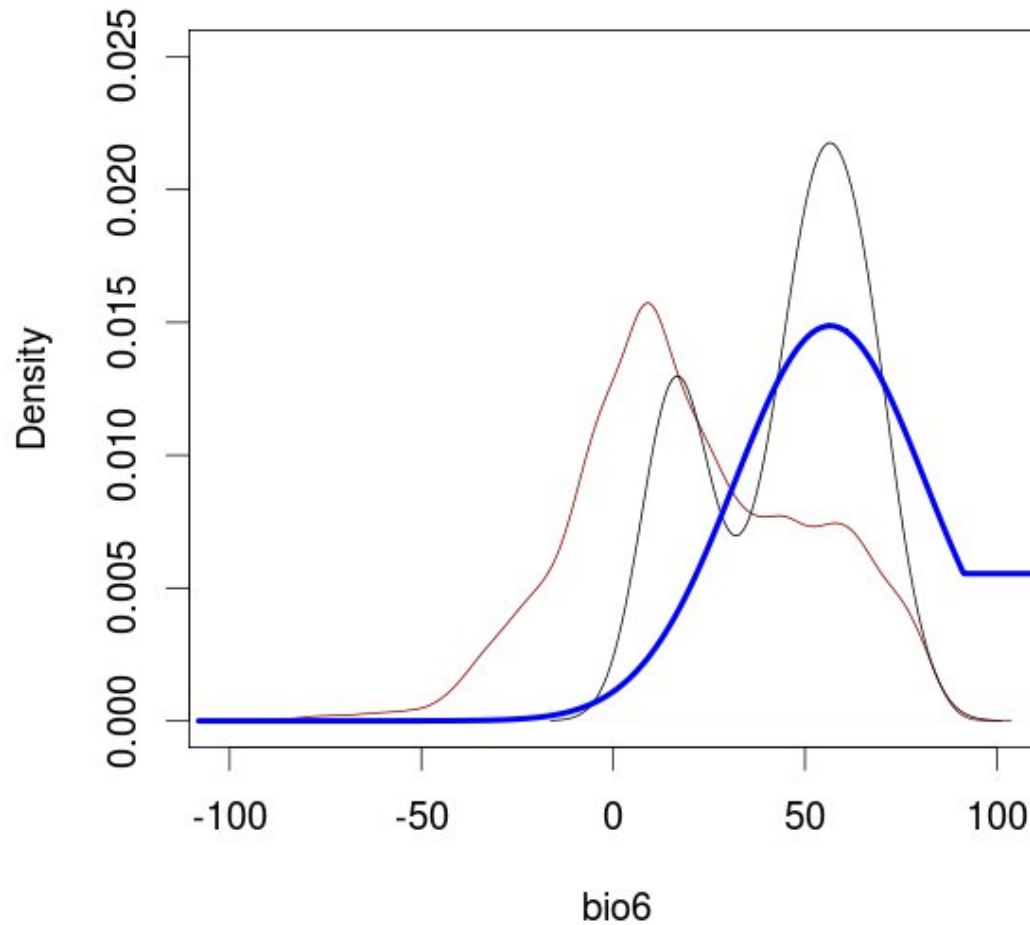
Con auto features activado

CURVA DE RESPUESTA DE MAXENT



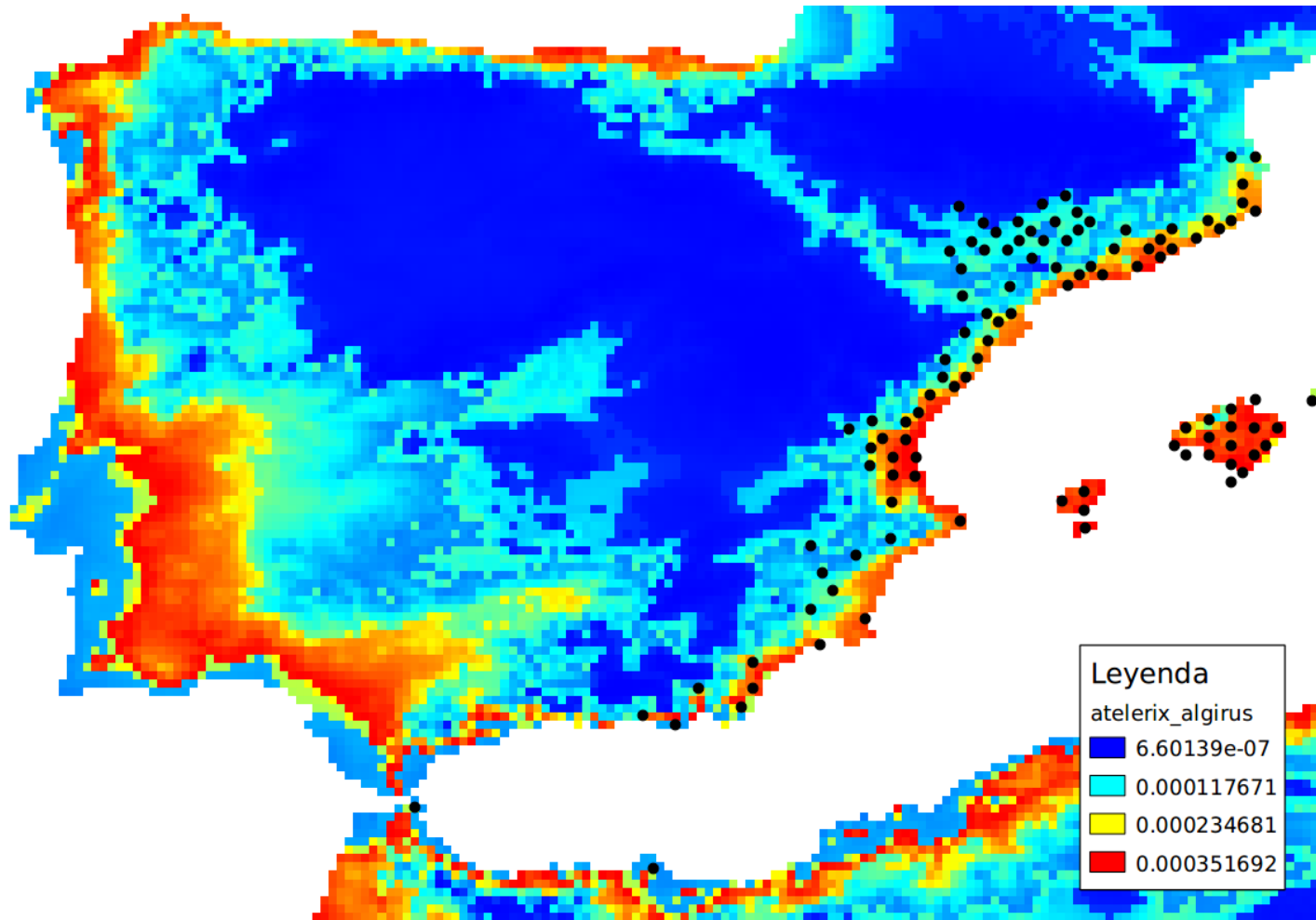
Con auto features y regularization multiplier = 3

CURVA DE RESPUESTA DE MAXENT



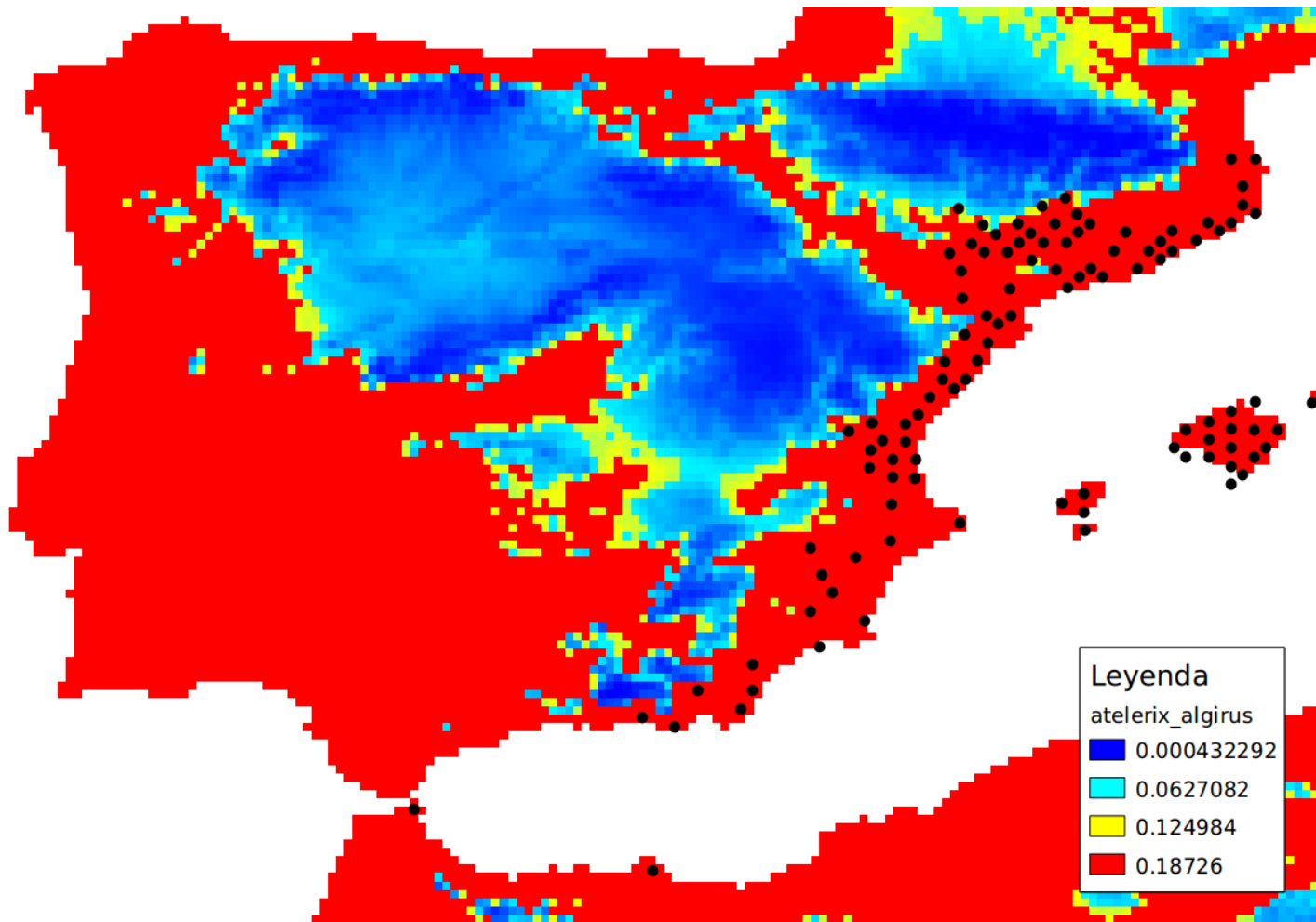
Sin hinge ni threshold

RAW OUTPUT



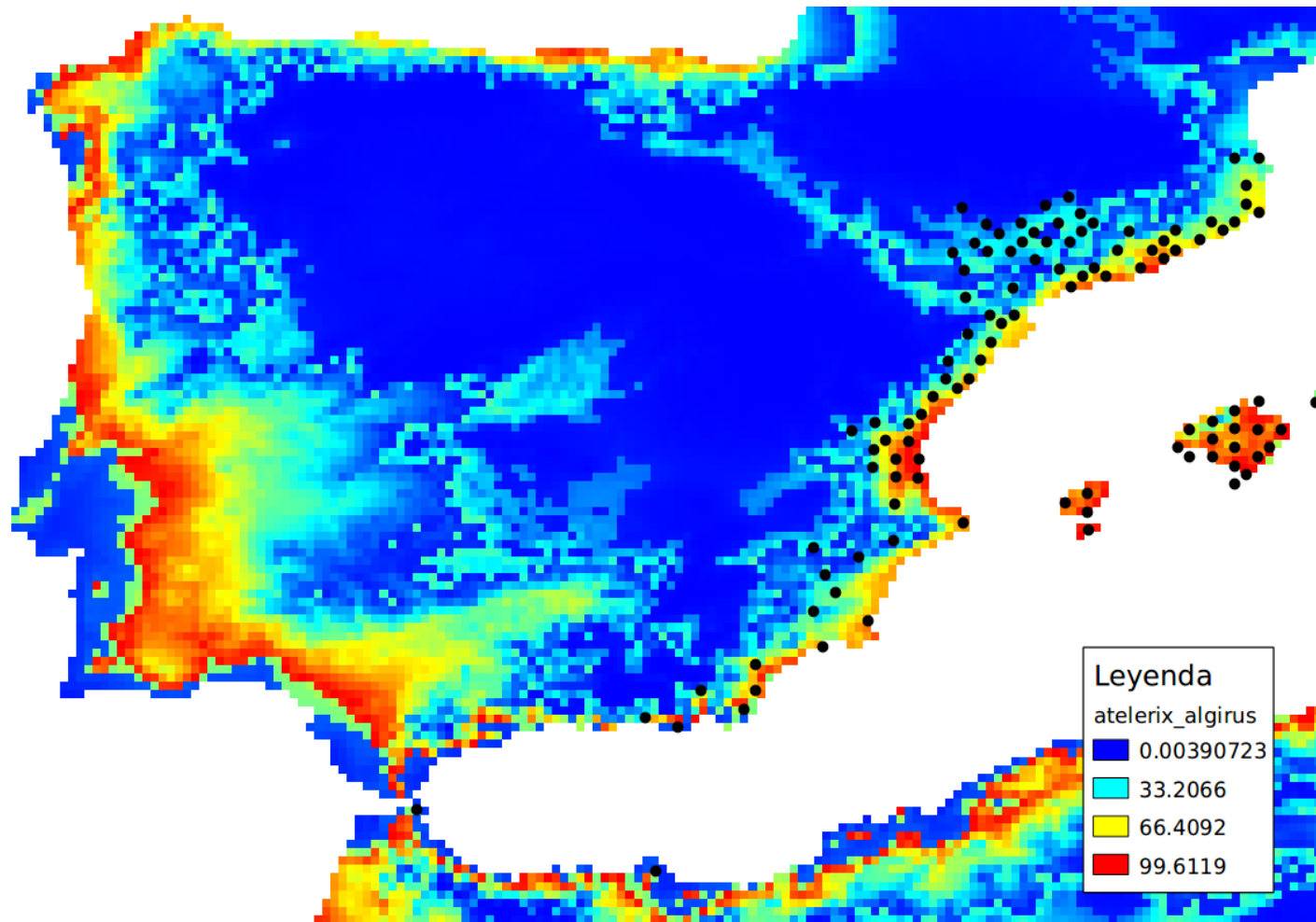
Se interpreta como 'tasa de ocurrencia relativa'

LOGISTIC OUTPUT



Transformación del raw output asumiendo prevalencia 0,5

CUMULATIVE OUTPUT



Se interpreta como 'error de omisión'

75 → el 75% de las presencias están por debajo de este valor

OUTPUT

- Usar 'raw output' cuando sea posible (no lleva transformaciones), ideal si lo escalas de 0 a 1.
- Usar 'cumulative output' cuando el objetivo es convertir el modelo en binario.
- No usar el 'logistic output' porque los valores de prevalencia no son realistas.

PRIMERA EJECUCIÓN (configuración)

- **Samples:** taller2 > resultados > presencia_maxent_entrenamiento.csv
- **Environmental layers:** taller2 > resultados > variables
 - bio14
 - bio5
 - Bio6
 - diversidad_topo
 - human_footprint
 - landcover_veg_tree
 - ndvi_average
 - ndvi_range

PRIMERA EJECUCIÓN (configuración)

- Marcamos los dos boxes:
 - Create response curves
 - Do Jackknife to measure variable importance
- Settings > Advanced > Adjust sample radius = -10
- Output directory:
“C:\taller2\resultados\maxent\ejecucion1”
- Run!
- **¡NO CIERRES MAXENT!**

PRIMERA EJECUCIÓN (resultados)

Vamos a la carpeta
C:\taller2\resultados\maxent\ejecucion1
y abrimos “**Ursus_arctos.html**”

PRIMERA EJECUCIÓN (resultados)

PRIMER CONJUNTO DE CURVAS

- Respuesta a cada variable de forma individual, dejando las demás como constantes usando su media.
- Mejor mirar aquí si la colinealidad es baja

SEGUNDO CONJUNTO DE CURVAS

- Cada curva es un modelo MaxEnt realizado solo con esa variable
- Mejor mirar aquí si la colinealidad es alta (no es nuestro caso)

PRIMERA EJECUCIÓN (resultados)

- Análisis de la contribución de las variables (difícil de interpretar si hay colinealidad)
 - **Percent contribution:** suma de la contribución de la variable al ajuste del modelo en cada iteración
 - **Permutation importance:** diferencia de ajuste entre los valores de los puntos en la variable y la permutación aleatoria de estos valores

PRIMERA EJECUCIÓN (resultados)

Jackknife

- **Verde:** ajuste del modelo con todas las variables excepto “esa”
- **Azul:** ajuste del modelo solo con “esa” variable
- **Rojo:** ajuste del modelo con todas las variables

PRIMERA EJECUCIÓN (resultados)

En la carpeta

C:\taller2\resultados\maxent\ejecucion1 le cambiamos el nombre al fichero

Ursus_arctos.asc a **maxent1.asc**, y movemos el fichero a C:\taller2\resultados\modelos

SEGUNDA EJECUCIÓN (configuración)

- Settings > Basic > Regularization multiplier=3
- Output directory=
C:\taller2\resultados\maxent\ejecucion2
- Run!

SEGUNDA EJECUCIÓN (resultados)

Comparamos las curvas de respuesta y el mapa con el modelo anterior. ¿Qué ha ocurrido?

SEGUNDA EJECUCIÓN (resultados)

En la carpeta

C:\taller2\resultados\maxent\ejecucion2 le cambiamos el nombre al fichero

Ursus_arctos.asc a **maxent2.asc**, y movemos el fichero a C:\taller2\resultados\modelos

TERCERA EJECUCIÓN (configuración)

- Settings > Basic > Regularization multiplier=1
- Desmarcamos:
 - Auto features
 - Hinge features
 - Threshold features
- Output directory=
C:\taller2\resultados\maxent\ejecucion3

TERCERA EJECUCIÓN (resultados)

- Abrimos
C:\taller2\resultados\maxent\ejecucion3\Ursus_arctos.html
- Veamos las curvas de respuesta: ¿qué creéis que ocurre?, ¿Tienen estas curvas más sentido que las anteriores?
- Veamos los tres mapas superpuestos...
- **¡Tus conocimiento y decisiones importan!**

TERCERA EJECUCIÓN (resultados)

En la carpeta

C:\taller2\resultados\maxent\ejecucion3 le
cambiamos el nombre al fichero

Ursus_arctos.asc a **maxent3.asc**, y movemos el
fichero a C:\taller2\resultados\modelos

CUARTA EJECUCIÓN (configuración)

- Output format: Cumulative
- Desmarcamos “Do jackknife to measure...”
- Desmarcamos “Create response curves”
- Output directory=
C:\taller2\resultados\maxent\ejecucion4
- Settings > Experimental > desmarca
“Logscale raw/cumulative pictures”

CUARTA EJECUCIÓN (resultados)

- Abrimos
C:\taller2\resultados\maxent\ejecucion4\Ursus_arctos.html
- Superponemos el nuevo mapa con los dos anteriores
- Fíjate en la correspondencia entre los mapas
- Las escalas de valores son diferentes
- Y ambos representan distintos conceptos

EVALUACIÓN CON MAXENT 1 (con fichero de evaluación)

- **Samples:** taller2 > resultados > presencia_maxent_entrenamiento.csv
- **Environmental layers:** taller2 > resultados > variables
 - bio14
 - bio5
 - Bio6
 - diversidad_topo
 - human_footprint
 - landcover_veg_tree
 - ndvi_average
 - ndvi_range

EVALUACIÓN CON MAXENT 1 (con fichero de evaluación)

- Output directory:
“**C:\taller2\resultados\maxent\evaluacion1**”
- Settings > Basic > Test sample file >
resultados/presencia_maxent_test.csv
- Settings > Advanced > Adjust sample radius =
-5
- Run!

EVALUACIÓN CON MAXENT 1 (con fichero de evaluación)

- Maxent muestra dos curvas ROC:
 - **Rojo**: datos de entrenamiento
 - **Azul**: datos de evaluación
- Los puntos de evaluación están en **violeta**.

EVALUACIÓN CON MAXENT 4 (bootstrap)

- Settings > Basic > marca Random seed
- Settings > Basic > Random test percentage = 40
- Settings > Basic > Replicates = 10
- Settings > Basic > Replicated run type = bootstrap
- Output directory:
“**C:\taller2\resultados\maxent\evaluacion4**”
- Run

EVALUACIÓN CON MAXENT 4 (bootstrap)

- Maxent entrena cada modelo con un conjunto distinto de datos en cada ocasión
- Resultados múltiples, el importante es el promedio
- La curva ROC es el promedio de las n ejecuciones tiene menor desviación estándar
- El modelo es la media de los 10 modelos (ensemble)

THRESHOLD CON MAXENT

- Output directory:
“**C:\taller2\resultados\maxent\threshold1**”
- Settings > Basic > Replicated run type = Crossvalidate
- Settings > Basic > Replicates = 1
- Settings > Basic > Random test percentage = 0
- Settings > Advanced - Apply threshold rule = 10 percentile training presence
- Run, y vamos a la carpeta de resultados
- Visualiza `Ursus_arctos_thresholded.asc` en Qgis

PROYECCIÓN CON MAXENT (configuración)

- **Samples:** taller2 > resultados > presencia_maxent.csv
- **Environmental layers:** taller2 > resultados > variables
 - bio14
 - bio5
 - Bio6
 - diversidad_topo
 - human_footprint
 - landcover_veg_tree
 - ndvi_average
 - ndvi_range

PROYECCIÓN CON MAXENT (configuración)

- Projection layers:
resultados/variables_norteamerica
- Settings > Advanced > marca **Extrapolate**
- Settings > Advanced > marca **Do clamping**
- Settings > Advanced > marca **Fade by clamping**
- Settings > Advanced > **Do MESS Analysis when projecting**
- **Run**



That's all Folks!