

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE CIENCIAS BIOLÓGICAS



EVALUACIÓN DE LA DISTRIBUCIÓN POTENCIAL DE ESPECIES
EXÓTICAS DULCEACUÍCOLAS EN MÉXICO MEDIANTE
MODELADO DE NICHO ECOLÓGICO

POR

SERGIO ALBERTO LUNA PEÑA

COMO REQUISITO PARCIAL PARA OBTENER EL GRADO DE
DOCTORADO en Conservación, Fauna Silvestre y Sustentabilidad

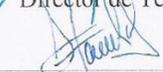
2024

EVALUACIÓN DE LA DISTRIBUCIÓN POTENCIAL DE ESPECIES
EXÓTICAS DULCEACUÍCOLAS EN MÉXICO MEDIANTE
MODELADO DE NICHOS ECOLÓGICOS

Comité de Tesis



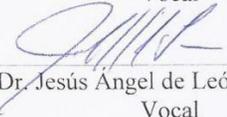
Dr. Roberto Eduardo Mendoza Alfaro
Director de Tesis



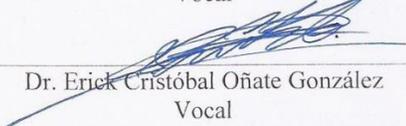
Dra. Susana Favela Lara
Secretario



Dr. Luis Alexander Peña Peniche
Vocal



Dr. Jesús Ángel de León González
Vocal



Dr. Erick Cristóbal Oñate González
Vocal



Dra. Katiushka Arcevalo
Subdirectora de Posgrado



SUBDIRECCIÓN
DE POSGRADO

EVALUACIÓN DE LA DISTRIBUCIÓN POTENCIAL DE ESPECIES
EXÓTICAS DULCEACUÍCOLAS EN MÉXICO MEDIANTE
MODELADO DE NICHOS ECOLÓGICOS

Dirección de Tesis



Dr. Roberto Eduardo Mendoza Alfaro
Director de Tesis

DERECHOS RESERVADOS ©

PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta Tesis está protegido, el uso de imágenes, fragmentos de videos, y demás material contenido que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde se obtuvo mencionando al autor o autores.

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico otorgado durante la realización de este proyecto de doctorado.

A la Facultad de Ciencias Biológicas, UANL, por haberme aceptado en el programa de posgrado.

A mi director de Tesis Dr. Roberto Mendoza Alfaro, quien ha sido una fuente de apoyo y guía desde mis primeros pasos en la licenciatura, pasando por la maestría, hasta este doctorado. Su conocimiento y dedicación no solo han marcado mi formación académica, sino también mi crecimiento personal. Sin su constante impulso y confianza en mis capacidades, este logro no habría sido posible.

A mi Comité de tesis: Dra. Susana Favela Lara, Dr. Luis Alexander Peña Peniche, Dr. Jesús Ángel de León González y Dr. Erick Cristóbal Oñate González, por su asesoría e invaluable aportaciones para la culminación de este trabajo de investigación.

A los profesores del Doctorado en Conservación, Fauna Silvestre y Sustentabilidad, así como a la Dra. Angela Cuervo Robayo, del Instituto de Biología, UNAM, quienes a través de sus clases contribuyeron al delineamiento y cumplimiento de los objetivos de este proyecto. También agradezco a mis amigos y compañeros del posgrado, quienes con su apoyo y compañía hicieron de este viaje académico una experiencia más gratificante.

DEDICATORIA

*“Nothing in life is to be feared, it is only to be understood.
Now is the time to understand more, so that we may fear less”
-Marie Curie*

A mis padres y a toda mi familia, gracias a todo el apoyo y amor que me han brindado durante este proceso.

A Isaura y Andrea, con su paciencia, comprensión y amor iluminan mi vida. Terminar este proyecto es solo el comienzo de otro aún más grande que ahora comparto con ustedes.

ÍNDICE

Contenido

RESUMEN	14
ABSTRACT	15
INTRODUCCIÓN	1
ANTECEDENTES	4
Concepto de nicho ecológico.....	4
Estimación del nicho ecológico	6
MaxEnt	7
Factores que influyen en el modelado.....	7
Área de calibración	8
Variables predictoras.....	10
Sesgo de muestreo.....	12
Evaluación del modelo	12
JUSTIFICACIÓN	14
HIPÓTESIS	15
OBJETIVO DEL TRABAJO	15
Objetivo general.....	15
Objetivos específicos	15
MATERIAL Y MÉTODOS	16
Simulación de especies virtuales	17
Modelos de nicho ecológico o distribución de especies	22
Filtración espacial de los registros de ocurrencia	23

Colinealidad de predictores	25
Área de calibración	25
Multiplicador de regularización (RM)	26
Modelos no sesgados.....	27
Evaluación del desempeño	27
Análisis estadístico.....	30
RESULTADOS	32
DISCUSIÓN.....	40
CONCLUSIONES	47
PERSPECTIVAS.....	48
BIBLIOGRAFÍA	50

INDICE DE FIGURAS

Figura	Titulo	Página
1	Diagrama BAM: una representación heurística de la interacción entre factores bióticos (B), abióticos (A) y de movimiento (M). Se representan cuatro áreas: el espacio geográfico (G), el área abióticamente adecuada (G_A), el área de distribución ocupada (G_O) y el área de distribución invadible (G_I). Se muestran datos de presencias (círculos sólidos) y ausencias (círculos abiertos) (tomado de Peterson et al. 2011).....	5
2	Esquema del proceso de simulación de área nativa e invasiva para una de las especies. A) Dimensiones del nicho [5 variables seleccionadas aleatoriamente]; B) Idoneidad del hábitat mediante la proyección del nicho al espacio geográfico (recuadro: transformación logística), C) Probabilidad de ocurrencia de la especie, D) Áreas accesibles potenciales en equilibrio ambiente (5 encontradas para la especie), E) Área nativa (verde) e invasiva (rojo) seleccionadas, y F) Sesgo de muestreo espacial en el rango nativo.....	21
3	Mapa que muestra los registros de ocurrencia utilizados para los modelos de distribución de especies (se muestran cuatro de las 5 réplicas utilizadas). Cada punto indica una ubicación de presencia en el área nativa, y las diferentes réplicas permiten evaluar la variabilidad espacial y el impacto del muestreo en los resultados del modelado.....	23
4	Mapa que muestra los resultados del filtrado espacial de registros de ocurrencia con distancias de separación de A) 0 km (sin filtración), B) 10 km y C) 20 km. Las diferentes distancias de filtrado permiten visualizar el impacto del proceso en la distribución espacial de los registros utilizados para los modelos de distribución de especies.....	24

5	Mapa que ilustra el efecto de la extensión del área de calibración en los modelos de distribución de especies. Se muestran áreas de calibración (área en color rojo) con radios de A) 25 km, B) 100 km, C) 500 km y D) 1000 km alrededor de los registros de ocurrencia.....	26
6	Evaluación del desempeño del modelo en términos de precisión funcional. La precisión funcional se calculó utilizando la correlación de rango de Spearman entre la probabilidad de ocurrencia real y la estimada por MaxEnt, aplicando la transformación log-log complementaria (cloglog). Aquí se muestra el modelo utilizando un área de calibración de 500 km, una filtración espacial de 10km, sin filtrar variables por colinearidad, y un multiplicador de regularización de 2 (la correlación de Spearman en este ejemplo es de 0.887).....	28
7	Mapa que muestra los puntos de presencia originales (A) y los dos tipos de partición de datos utilizados para la validación cruzada en los modelos de distribución de especies. En la partición aleatoria, los puntos se dividen en cuatro grupos representados por diferentes colores (B). En la partición por bloques geográficos, los datos se dividen en secciones espaciales, reduciendo la correlación espacial y favoreciendo la transferibilidad del modelo (C).....	29
8	Precisión funcional de los modelos por especie en el área nativa (400 modelos \times 5 réplicas). El eje Y representa la precisión funcional (medida como la correlación de rango de Spearman entre la probabilidad de ocurrencia real y la estimada por MaxEnt), y el eje X representa diferentes especies. Los puntos de datos se representan con cinco colores diferentes (uno para cada réplica). Las especies están ordenadas de mayor a menor en función del mejor modelo. La precisión del modelo no sesgado se denota para cada especie con un	33

- símbolo "+". La línea roja discontinua representa el valor esperado por azar.....
- 9 Porcentaje de varianza explicado por los cuatro factores analizados y la varianza residual (no explicada) sobre la precisión funcional en el área nativa. El eje X representa el porcentaje de varianza explicada, y el eje Y muestra los factores analizados. Los diagramas de caja resumen la distribución de estos porcentajes para todas las especies. Área: Área de calibración; RM: multiplicador de regularización; Cor: Colinealidad de predictores; Thin: Filtrado espacial de ocurrencias.... 34
- 10 Relación entre la precisión funcional y discriminatoria de los modelos en el área nativa para las 100 especies bajo diferentes extensiones de área de calibración y métodos de partición. El eje x representa la precisión funcional, dividida por la extensión del área de calibración, mientras que el eje y representa la precisión discriminatoria, dividida por el método de partición (aleatorio y bloques geográficos). Las líneas azules, obtenidas mediante regresión de Mínimos Cuadrados Ordinarios (OLS), proporcionan una representación visual de la correlación entre las dos métricas de precisión para cada subconjunto de puntos..... 35
- 11 Precisión funcional de los modelos por especie en el área invasiva (400 modelos \times 5 réplicas). El eje Y representa la precisión funcional (medida como la correlación de rango de Spearman entre la probabilidad de ocurrencia real y la estimada por MaxEnt), y el eje X representa diferentes especies. Los puntos de datos se representan con cinco colores diferentes (uno para cada réplica). Las especies están ordenadas de mayor a menor en función del mejor modelo. La precisión del modelo no sesgado se denota para cada especie con un símbolo "+". La línea roja discontinua representa el valor esperado por azar..... 36

- 12 Porcentaje de varianza explicado por los cuatro factores analizados y la varianza residual (no explicada) sobre la precisión funcional en el área invasiva. El eje X representa el porcentaje de varianza explicada, y el eje Y muestra los factores analizados. Los diagramas de caja resumen la distribución de estos porcentajes para todas las especies. Área: Área de calibración; RM: multiplicador de regularización; Cor: Colinealidad de predictores; Thin: Filtrado espacial de ocurrencias.... 38
- 13 Relación entre la precisión funcional y discriminatoria de los modelos en el área invasiva para las 100 especies bajo diferentes extensiones de área de calibración y métodos de partición. El eje x representa la precisión funcional, dividida por la extensión del área de calibración, mientras que el eje y representa la precisión discriminatoria, dividida por el método de partición (aleatorio y bloques geográficos). Las líneas azules, obtenidas mediante regresión de Mínimos Cuadrados Ordinarios (OLS), proporcionan una representación visual de la correlación entre las dos métricas de precisión para cada subconjunto de puntos..... 39

LISTA DE SIMBOLOS Y ABREVIATURAS

AS-ISK:	Aquatic Species Invasiveness Screening Kit (Kit de detección de invasividad de especies acuáticas)
AUC:	Area under the ROC curve (Área bajo la curva ROC)
BAM:	Biotic-Abiotic-Mobility framework (Enfoque Biótico, abiótico y de accesibilidad o movilidad)
CA:	Cellular automata (autómata celular)
cloglog:	Transformación log-log complementaria
ENFA:	Ecological Niche Factor Analysis (análisis de factor ecológico del nicho)
ENM:	Ecological niche model (modelo de nicho ecológico)
GARP:	Genetic Algorithm for Rule Set Production (Algoritmo Genético para la Producción de Conjuntos de Reglas)
G:	Espacio geográfico
G _A :	Área abióticamente adecuada
GBIF:	Global Biodiversity Information Facility (Red Mundial de Información sobre Biodiversidad)
G _O :	Área ocupada por la especie
G _I :	Área de distribución invadible
HII:	Human influence index (índice de influencia humana)
MaxEnt:	Maximum entropy (máxima entropía)
OLS:	Ordinary least squares (mínimos cuadrados ordinarios)
PCA:	Principal Component Analysis (Análisis de componentes principales)
RM:	REGularization multiplier (multiplicador de regularización)
ROR:	Relative occurrence probability (tasa de ocurrencia relativa)
SDM:	Species distribution model (modelo de distribución de especies)

RESUMEN

Las invasiones biológicas representan actualmente la segunda causa de pérdida de biodiversidad a nivel mundial, siendo los ambientes acuáticos especialmente susceptibles. Por lo cual es esencial la identificación de especies que presentan un alto riesgo de invasión para México. Una parte esencial en el análisis de riesgo de especies potencialmente invasoras es la estimación de áreas con mayor probabilidad de establecimiento. Dentro de este contexto, los modelos de nicho ecológico (ENM) o modelos de distribución de especies (SDM, por sus siglas en inglés) son herramientas ampliamente utilizadas en ecología y conservación, cuyo objetivo es predecir la distribución potencial de una especie en función de sus requisitos ambientales y datos de ocurrencia, representan una herramienta esencial para estimar la distribución potencial de las especies. Los ENM enfrentan muchos desafíos e incertidumbres que influyen en su precisión. Este estudio analiza la influencia de cuatro factores en la precisión de los SDM a través de simulaciones con especies virtuales y un caso de estudio con especies reales. Se generaron 100 especies virtuales utilizando variables bioclimáticas. Las probabilidades de ocurrencia se determinaron en función de la idoneidad ambiental, el sesgo de muestreo espacial y las áreas accesibles. Los SDM se construyeron utilizando MaxEnt, variando el tamaño del área de calibración, el filtrado espacial de los registros de ocurrencia, el tratamiento de la colinealidad de los predictores y el parámetro de regularización. El rendimiento de los modelos se evaluó en términos de precisión funcional (precisión del modelo real) y precisión de discriminación (capacidad del modelo para separar ocurrencias de sitios aleatorios), tanto en el área nativa como en el área invasiva. Los resultados muestran que la extensión del área de calibración fue el factor más influyente (explicando el 50% de la varianza en la precisión funcional en el área nativa y 18% en el área invasiva), mientras que el multiplicador de regularización, la colinealidad de los predictores y el adelgazamiento espacial tuvieron un impacto mínimo (menos del 5% de la varianza explicada en conjunto en ambas áreas). En general, las áreas de calibración más grandes condujeron a una mayor precisión funcional, aunque hay variación según la

especie. La correlación entre la precisión funcional y de discriminación fue relativamente baja, lo que indica que los modelos que tienen un buen rendimiento en una métrica pueden no sobresalir en la otra. En conclusión, esta investigación avanza en la discusión sobre la selección del área de calibración, proporcionando información sobre sus efectos sustanciales en la precisión del modelo. Nuestros hallazgos demuestran que el tamaño del área de calibración es uno de los factores más críticos que afectan la precisión de los modelos de distribución de especies invasoras, superando la influencia de otros factores. Estos conocimientos subrayan la importancia de seleccionar áreas de calibración apropiadas para mejorar las predicciones del modelo y garantizar aplicaciones más confiables de los mismos.

ABSTRACT

Biological invasions currently represent the second leading cause of biodiversity loss worldwide, with aquatic environments being particularly susceptible. Therefore, it is essential to identify species that pose a high risk of invasion to Mexico. An essential part of the risk analysis of potentially invasive species is the estimation of areas with a higher probability of establishment. In this context, ecological niche models (ENM) or species distribution models (SDM) are widely used tools in ecology and conservation, aimed at predicting the potential distribution of a species based on its environmental requirements and occurrence data. These models serve as an essential tool for estimating the potential distribution of species. ENM face many challenges and uncertainties that influence their accuracy. This study analyses the influence of four factors on the accuracy of SDM through simulations with virtual species and a case study of real species. Using bioclimatic variables, 100 virtual species were generated. Occurrence probabilities were determined based on environmental suitability, spatial sampling bias, and accessible areas. SDM were built using MaxEnt, varying size of calibration area, spatial filtering of occurrence records, predictor collinearity treatment, and regularization parameter. Model performance was assessed in terms of functional accuracy (true model accuracy) and discrimination

accuracy (model ability to separate occurrence from random sites), both in the native and invasive range. Results show that the extent of the calibration area was the most influential factor (explaining 50% of the variance in functional accuracy in the native range, and 18% in the invasive range), while regularization multiplier, predictor collinearity, and spatial thinning had minimal impact (less than 5% of explained variance combined for both areas). Overall, larger calibration areas generally led to higher functional accuracy, although it varies across species. The correlation between functional and discrimination accuracy was relatively low, indicating that models performing well in one metric may not excel in the other. In conclusion, this research advances the discussion on calibration area selection, providing insights on its substantial effects on model accuracy. Our findings demonstrate that the size of the calibration area is one of the most critical factors affecting the accuracy of models, surpassing the influence of other factors. These insights highlight the importance of select appropriate calibration areas to improve model predictions and ensure more reliable applications of the models.

INTRODUCCIÓN

Las invasiones biológicas representan actualmente la segunda causa de pérdida de biodiversidad a nivel mundial, solo superada por la destrucción del hábitat (Myers et al. 2000). El número de especies invasoras sigue incrementando (Mormul et al. 2022) y en particular los ambientes acuáticos, que albergan una gran diversidad de especies nativas, se han destacado por ser especialmente susceptibles al impacto generado por las especies invasoras, lo que ha resultado en una mayor tasa de extinción respecto a los ecosistemas terrestres (Rahel 2002; Leprieur et al. 2008).

Debido al impacto ecológico y económico ocasionado por las especies exóticas (Rico-Sánchez et al. 2021), conforme sea más rápida la respuesta ante una invasión, mayor será el costo-beneficio de las medidas implementadas (Pimentel et al. 2005; Brown y Daigneault, 2014). Por lo anterior la prevención de las invasiones biológicas representa una prioridad y un desafío importante en la conservación y manejo de los ecosistemas acuáticos, siendo esencial identificar aquellas especies que presentan el mayor riesgo para una región (Leung et al. 2012).

La identificación y priorización de zonas de alta probabilidad de establecimiento es un aspecto fundamental de la prevención, esto con el fin de validar y justificar programas de monitoreo eficaces, desarrollar programas de detección temprana, y predecir y evaluar los posibles escenarios de dispersión de aquellas especies que ya se encuentran introducidas o evaluar el efecto del cambio climático sobre invasiones actuales o futuras (Montecino et al. 2014; Masin et al. 2014).

Debido a lo anterior diversas herramientas de análisis de riesgo consideran aspectos acerca de la probabilidad de establecimiento y coincidencia climática. Por ejemplo, la herramienta *Harmonia*⁺ incluyen un módulo de *Establecimiento*, el cual evalúa la probabilidad de que la especie encuentre condiciones climáticas y hábitats adecuados para su supervivencia y reproducción en el área de estudio (D'hondt et al. 2015). En el caso de especies acuáticas, la herramienta de análisis *AS-ISK* (Aquatic Species Invasiveness Screening Kit) requiere un nivel de coincidencia climática en el área de distribución nativa de las especies evaluadas y el área de evaluación de riesgo (Copp et al. 2016).

Actualmente se dispone de una gran cantidad de información geográfica sobre la biodiversidad, así como de datos de carácter ambiental como temperatura y precipitación. La integración, análisis y comprensión de esta información sobre diversidad biológica, en conjunto con el concepto de nicho ecológico ha resultado en el desarrollo de distintos modelos de nicho ecológico (ecological niche model; ENM) o de distribución de especies (species distribution model; SDM) para inferir o explicar la distribución geográfica de las especies (Peterson et al. 2011; Hardisty y Roberts 2013).

Dentro de este contexto, MaxEnt es uno de los algoritmos de modelado de nicho ecológico más utilizados y ha demostrado un buen desempeño comparado con otros modelos (Wisz et al. 2008), representado una herramienta muy efectiva para estimar la distribución potencial de especies acuáticas invasoras (Reshetnikov y Ficetola 2011; Poulos et al. 2012; Merow et al. 2013). En esencia, MaxEnt determina la asociación entre los puntos de ocurrencia de una especie y la información de las variables ambientales asociadas a esos sitios para identificar patrones en la distribución potenciales, los cuales puede ser utilizados para determinar áreas ecológicamente similares a aquellas en donde habita una especie (Phillips et al. 2006).

Sin embargo, el proceso de modelado de la distribución de especies puede llegar a ser muy complejo y computacionalmente demandante, involucrando la colecta, selección y preparación de datos ambientales y de presencia de la especie, la selección y parametrización del algoritmo de modelado, y la evaluación, selección e interpretación de los resultados (de Souza-Muñoz et al. 2011; Cobos et al. 2019). Lo cual puede representar un desafío por el gran número de especies que se requiere evaluar en muchas ocasiones (Bomford 2008; D'hondt et al. 2015). Aunado a esto, no existe una medida de evaluación aceptada universalmente para su validación, lo cual representa uno de los principales desafíos en el campo (Wunderlich et al. 2019).

Debido a la necesidad de evaluar un gran número de especies exóticas, es necesaria la optimización de diversos aspectos del proceso de modelado. Con este propósito, el presente proyecto está centrado en una evaluación cuantitativa de la importancia de diversos factores involucrados en la generación y selección de los modelos de nicho de

especies exóticas acuáticas, enfocado en las condiciones ambientales relevantes para México.

ANTECEDENTES

Concepto de nicho ecológico

Las especies se encuentran distribuidas en patrones complejos que han resultado de factores ecológicos y evolutivos globales. La estimación y comprensión de la distribución de las especies es un aspecto fundamental para el entendimiento de los procesos que afectan la biodiversidad, así como su conservación (Elith y Leathwick 2007).

Un concepto de gran importancia en ecología y biogeografía es el de nicho ecológico, el cual ha demostrado ser difícil de definir (Hirzel y LeLay 2008). Por ejemplo, Grinnell (1917) describió el nicho como los requisitos climáticos y de hábitat necesarios para que una especie pueda existir. Posteriormente, Elton (1927) caracterizó el nicho como el rol funcional que un organismo desempeña dentro de una comunidad. Más adelante, Hutchinson (1957) definió el nicho fundamental de una especie como un hipervolumen compuesto por n variables ambientales, donde cada punto representa una condición ambiental que permite la existencia indefinida de la especie (Peterson et al. 2011).

Una terminología más adaptada al problema de estimar y comprender las áreas de distribución de las especies es el diagrama BAM (Figura 1), el cual representa subconjuntos del espacio geográfico (G) que presentan las condiciones bióticas (B), abióticas (A) y de accesibilidad (o movilidad, M) adecuadas para la especie. Dentro de este espacio geográfico existe el área ocupada por la especie (G_0), es decir el conjunto del espacio que contiene condiciones bióticas y abióticas adecuadas para la especie y que ha sido accesible en periodos de tiempo relevantes (Peterson et al. 2011).

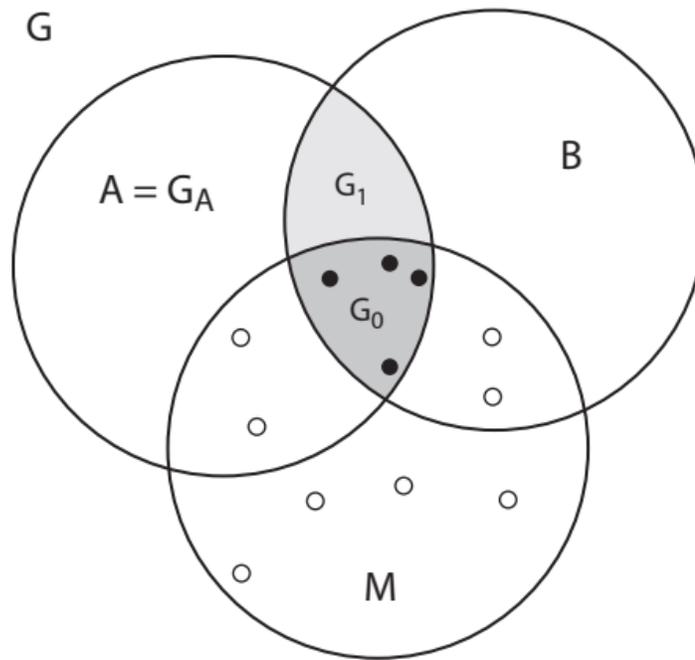


Figura 1. Diagrama BAM: una representación heurística de la interacción entre factores bióticos (B), abióticos (A) y de movimiento (M). Se representan cuatro áreas: el espacio geográfico (G), el área abióticamente adecuada (G_A), el área de distribución ocupada (G_0) y el área de distribución invadible (G_1). Se muestran datos de presencias (círculos sólidos) y ausencias (círculos abiertos) (tomado de Peterson et al. 2011).

Como consecuencia existe un espacio geográfico que presenta las condiciones bióticas y abióticas adecuadas para la supervivencia de una especie pero que no han sido accesibles a ella por cuestiones históricas, este espacio es considerado como área invadible (G_1) (Hirzel y LeLay 2008; Peterson et al. 2011).

Esto se ha mantenido constante a lo largo de miles de años. Sin embargo, gracias al transporte, comercio y modificaciones del hábitat muchas especies han logrado superar las barreras que limitaban su distribución y han logrado establecerse e invadir nuevas regiones. La identificación de estas áreas es un aspecto fundamental en el campo de las invasiones biológicas (Gallien et al. 2010; Blackburn et al. 2011; Briski et al. 2018)

Estimación del nicho ecológico

En literatura se han reportado diferentes aproximaciones para la estimación del nicho ecológico, siendo posible estudiarse de forma mecanística, mediante modelos fisiológicos basados en procesos o mediante la realización de experimentos de laboratorio para determinar los rangos de condiciones bajo las cuales una especie puede vivir (Porter et al. 2002; Buckley 2008). El principal inconveniente de estos enfoques es que requieren laboriosos experimentos, y por lo tanto no son fácilmente aplicables a un gran número de especies (Peterson et al. 2011). Otra aproximación se basa en métodos correlativos para estimar el nicho fundamental a partir de observaciones de la presencia de una especie en relación con la variación en las condiciones ambientales (Peterson and Soberon 2012; Sillero et al. 2021).

Estos modelos son herramientas ampliamente utilizadas en ecología y conservación para predecir la distribución potencial de una especie en función de sus requerimientos ambientales y datos de ocurrencia (Peterson y Soberón 2012). Ofrecen información valiosa para comprender patrones de biodiversidad (Weaver et al. 2006), identificar prioridades de conservación (Srinivasulu et al. 2021), evaluar los impactos del cambio climático (Brodie et al. 2022) y evaluar el riesgo de invasiones biológicas (Duque-Lazo et al. 2016). Sin embargo, los SDM también enfrentan varios desafíos e incertidumbres que pueden afectar su precisión (Sillero et al. 2021; Sillero y Barbosa 2021).

La factibilidad de estimar el nicho fundamental existente utilizando datos de distribución ha sido controvertida, ya que es probable que la distribución de una especie incluya interacciones bióticas y limitaciones de su dispersión (Pollock et al. 2014). Sin embargo, gracias a la cada vez mayor cantidad de información sobre biodiversidad, es posible realizar estimaciones para un gran número de especies, para lo cual han sido desarrollados diversos algoritmos (Peterson et al. 2011; Guisan et al. 2017).

Todos los algoritmos, en esencia, se construyen a partir de asociaciones entre datos de presencia en forma de coordenadas georreferenciadas de latitud y longitud y conjuntos de datos ambientales para localidades confirmadas de la especie para generar un modelo que

permita predecir la distribución geográfica potencial, el cual puede ser proyectado en un mapa de la región de estudio (Peterson et al. 2011; Guisan et al. 2017).

MaxEnt

Uno de los algoritmos de modelado de nicho ecológico más utilizado es MaxEnt (máximum entropy), el cual ha mostrado un buen desempeño comparado con otros modelos (Wisz et al. 2008), especialmente cuando se trabaja solo con datos de presencia. Por lo anterior, representa una herramienta muy efectiva para estimar la distribución potencial de especies acuáticas invasoras (Reshetnikov y Ficetola 2011; Poulos et al. 2012; Merow et al. 2013).

MaxEnt utiliza la información de un conjunto de variables predictoras ambientales (e.g. precipitación, temperatura) a partir de los sitios donde se ha registrado presencia de una especie, así como de un área definida por el usuario. A partir de esta información, MaxEnt extrae una muestra de puntos de fondo (background; donde la presencia de la especie es desconocida) que contrasta con las ubicaciones de presencia. Originalmente, MaxEnt fue utilizado para estimar la densidad de presencias en el paisaje, asumiendo un muestreo aleatorio en el área (Fithian y Hastie 2013). Sin embargo, el tamaño de la población generalmente se desconoce, por lo que solo es posible estimar una tasa de ocurrencia relativa (relative occurrence probability; ROR) (Phillips et al. 2006; Fithian y Hastie 2013; Merow et al. 2013).

MaxEnt predice la ROR en función de las variables predictoras ambientales en esa ubicación (Merow et al. 2013). Estas ROR toman la forma:

$$P^*(z(x_i)) = \exp(z(x_i)\lambda) / \sum_i \exp(z(x_i)\lambda)$$

donde \mathbf{z} es un vector de J variables ambientales en la ubicación x_i , y λ es un vector de coeficientes de regresión, con $\mathbf{z}(x_i) \lambda = z_1(x_i)*\lambda_1 + z_2(x_i)*\lambda_2 + \dots + z_J(x_i)*\lambda_J$.

Factores que influyen en el modelado

El proceso de modelado de nicho involucra la colecta, selección y preparación de datos ambientales y de presencia de la especie, la selección y parametrización del algoritmo de modelado, y la evaluación, selección e interpretación de los resultados. Durante este proceso es necesario tomar varias decisiones que pueden influir en el resultado. Por ejemplo, los registros de presencia de la especie pueden variar en calidad y cantidad. También es necesaria la identificación de los datos del entorno relevantes para la distribución de la especie, los cuales pueden no estar disponibles para las regiones a evaluar, la definición de algunos parámetros en el algoritmo que afectará la forma en que se procesan los datos, así como la selección de las métricas adecuadas para su evaluación (de Souza-Muñoz et al. 2011; Cobos et al. 2019; Sillero et al. 2021). Además, en el caso de las especies invasoras es fundamental la verificación de supuestos adicionales, ya que por ejemplo es posible que estas especies no se encuentren en equilibrio con su ambiente en el caso de etapas tempranas de una invasión (Gallien et al. 2010; 2012) o puede haber una disminución en el rendimiento del modelo paralelo a un cambio en la relación entre las especies y su entorno en un contexto temporal o espacial diferente, es decir, baja transferibilidad (Wenger y Olden 2012; Fernández y Hamilton 2015). A continuación, se describen algunos de los factores que influyen en el modelado que serán evaluados en el presente estudio.

Área de calibración

El área de calibración ideal es aquella donde la especie está en equilibrio ambiental (Guisan y Zimmermann 2000). Esto significa que la especie se ha adaptado a las condiciones ambientales actuales y, como resultado, su distribución refleja estas condiciones (Araújo y Pearson 2005; Barve et al. 2011). Bajo el enfoque Biotic-Abiotic-Mobility (BAM), la M cumple con este supuesto (Soberon y Peterson 2005; Barve et al. 2011); sin embargo, para la mayoría de las especies, su estimación es complicada (Acevedo et al. 2012).

La selección del área de calibración es especialmente importante para algunos algoritmos de presencia-entorno (presence-background), como Ecological Niche Factor Analysis

(ENFA; Hirzel et al. 2002), el Algoritmo Genético para la Producción de Conjuntos de Reglas (GARP; Stockwell y Noble 1992) y el algoritmo de Máxima Entropía (MaxEnt; Phillips et al. 2006; Merow et al. 2013), que se ven afectados por la selección del área de calibración. Es importante destacar que, aunque la selección del área de calibración es un desafío crítico para algunos algoritmos, no todos los métodos enfrentan este problema (Sillero et al. 2021). Por ejemplo, la distancia de Mahalanobis (Clark et al. 1993) y BIOCLIM (Booth et al. 2014) no requieren un área de calibración debido a que solo utilizan datos de ocurrencia.

Los métodos de presencia-entorno evalúan las condiciones ambientales en toda el área de estudio (entorno) y las comparan con las condiciones donde se encuentra la especie, según los registros de ocurrencia (Phillips et al. 2006; Merow et al. 2013). Debido a esto, la selección del área de calibración puede influir significativamente en la precisión del modelo, ya que diferentes áreas pueden tener condiciones ambientales y sesgos de muestreo distintos (Acevedo et al. 2012; Owens et al. 2013). Además, se ha demostrado que la extensión del área de calibración repercute en los análisis de desempeño posteriores del modelo. Por ejemplo, se ha encontrado una sobrevaloración del desempeño del modelo conforme aumenta esta área (Merow et al. 2013). Por otra parte, existe una mayor probabilidad de extrapolación del modelo conforme el área de calibración disminuye, debido a la menor cantidad de condiciones ambientales presentes (Thuiller et al. 2004).

A pesar de su importancia, aún no existe un consenso sobre cómo seleccionar un área de calibración adecuada, y en la literatura se han utilizado diferentes criterios y métodos (Machado-Stredel et al. 2021; Rotllan-Puig y Traveset 2021). Algunos enfoques incluyen seleccionar regiones políticas, polígonos o rectángulos alrededor de los registros de ocurrencia (Shcheglovitova y Anderson 2013), seleccionar regiones bioclimáticas o hidrogeográficas ocupadas por la especie (Espindola et al. 2019; Sillero et al. 2021), o generar buffers alrededor de los registros de ocurrencia (Zhu et al. 2014). En algunos estudios, el área de calibración utilizada no se menciona explícitamente (e.g., Masin et al. 2014). Estos diferentes métodos pueden resultar en regiones de tamaños variables, lo que puede tener un impacto significativo en las características posteriores de los modelos realizados (Thuiller et al. 2004; Yiwen et al. 2016).

Según varios estudios, la selección de áreas de calibración demasiado restringidas o amplias pueden afectar la precisión de las predicciones del modelo (VanDerWal et al. 2009; Acevedo et al. 2012). Otros estudios han encontrado que áreas de calibración más pequeñas pueden proporcionar una mayor precisión en el modelo, ya que mitigan el riesgo de sobreajuste a las condiciones cercanas a las localidades ocupadas o excluyen regiones con condiciones adecuadas donde la especie no se encuentra debido a limitaciones en la dispersión y/o interacciones bióticas (Anderson y Raza 2010). Además, la selección de un entorno a partir de áreas más grandes conduce a cambios en la importancia de las variables, lo que resulta en modelos cada vez más simplificados y dominados principalmente por unas pocas variables (VanDerWal et al. 2009).

Además, la relevancia de la selección del área de calibración en especies invasoras se extiende a otros aspectos de interés, como la evaluación de desfases de nicho (Broennimann et al. 2007). Cuando se establece cuidadosamente la extensión del análisis, basada en conjuntos análogos de condiciones ambientales, se ha demostrado que las evidencias de cambios de nicho desaparecen, sugiriendo que los informes previos de cambios de nicho son artefactos metodológicos o interpretativos, más que realidades biológicas (Qiao et al. 2017; Bates y Bertelsmeier 2021). Por lo tanto, el conservadurismo de nicho debe evaluarse dentro de espacios ambientales similares y accesibles para evitar extrapolaciones inadecuadas durante la transferencia de modelos. La similitud ambiental entre áreas nativas e invadidas es crucial para identificar con precisión los cambios de nicho y aplicar modelos de distribución de especies a estudios temporales de dinámica de nicho (Aravind et al. 2022; Lo Parrino et al. 2023).

Variables predictoras

Las variables ambientales se han clasificado de diversas formas, según sus relaciones con la distribución geográfica de las especies, la influencia de la especie sobre ellas, así como su correlación con otras variables. La cantidad y calidad de variables que comprenden las dimensiones del nicho ecológico puede entre especies, por lo que la importancia relativa

de determinadas variables ambientales puede variar según el contexto geográfico y biótico (Pulliam 2000; Peterson et al. 2011).

La selección de variables predictoras influye directamente sobre la capacidad predictiva de los modelos, así como su transferibilidad espacial y temporal (Austin 2002; Petitpierre et al. 2017). Por ejemplo, una baja cantidad o una selección no apropiada puede fallar en la estimación del nicho de la especie. Por otra parte, un número excesivo de variables puede producir modelos sobreajustados o pérdida de grados de libertad en los enfoques basados en regresión, y las distribuciones geográficas pueden estar poco representadas (Peterson y Nakazawa 2008; Peterson et al. 2011).

Otro aspecto que influye significativamente en el desempeño del modelo es la multicolinealidad entre las variables predictoras. La multicolinealidad ocurre cuando las variables predictoras están altamente correlacionadas entre sí. Esto puede llevar a un sobreajuste, donde el modelo se ajusta demasiado a los datos de entrenamiento, limitando su capacidad para predecir con precisión nuevos datos. Además, la multicolinealidad puede generar curvas de respuesta poco fiables, dificultando la determinación de la verdadera influencia de cada variable predictora, ya que sus efectos son más difíciles de determinar por sus correlaciones con otras variables (Dormann et al. 2013; De Marco y Nóbrega, 2018; Feng et al. 2019).

Sesgo de muestreo

Uno de los supuestos de MaxEnt es que los individuos han sido muestreados al azar en toda el área de calibración. En el estudio de especies exóticas usualmente esto no es posible, siendo necesario recurrir a la información recopilada en portales globales como Global Biodiversity Information Facility (GBIF). Estas bases de datos suelen mostrar un sesgo en la distribución espacial de los datos debido al esfuerzo desigual de muestreo, ya que regiones de fácil acceso se muestrean con mayor frecuencia. Lo anterior aunado a diferencias a nivel nacional en el financiamiento y el intercambio de datos dan lugar a enormes diferencias en la contribución almacenamiento y movilización de datos (Beck et al. 2014; Aiello-Llammens et al. 2015).

Por lo anterior, uno de los desafíos más importantes en el proceso de modelado es la depuración de ocurrencias, que tiene como objetivo reducir el este sesgo espacial existente en los registros de ocurrencia de las especies. Sin embargo, la depuración también puede disminuir el número de puntos de datos disponibles para el modelado, lo que podría afectar negativamente el rendimiento y la precisión del modelo (Veloz 2009; Boria et al. 2014).

El filtrado espacial de los registros de ocurrencias es uno de los métodos para reducir el sesgo de muestreo, reduciendo la autocorrelación espacial que suele estar presente en los datos derivados de las colecciones biológicas (Taylor et al. 2020).

Evaluación del modelo

La validación de los modelos es un aspecto esencial, ya que nos permite evaluar su utilidad para la estimación del nicho de la especie. Con este objetivo han sido utilizadas diferentes métricas de acuerdo con el propósito de la evaluación. Una de las métricas más populares en la evaluación de modelos es el área bajo la curva ROC (AUC), el cual es un índice que permite conocer la capacidad predictiva entre puntos de presencia y ausencia de la especie. Sin embargo, una desventaja es que MaxEnt los puntos aleatorios de fondo son utilizados como ausencias, siendo que en realidad la presencia de la especie no es conocida en esos sitios (Peterson et al. 2008; Merow et al. 2013).

En el caso de las especies exóticas invasoras un aspecto adicional a considerar es la transferibilidad de los SDM (Yates et al. 2018; Liu et al. 2020a). A diferencia de las especies nativas, donde la interpolación dentro del área conocida es predominante, para las especies invasoras la extrapolación a nuevas áreas es igualmente importante (Heikkinen et al. 2012; Nguyen y Leung, 2022). Esto se debe a que las especies invasoras colonizan áreas fuera de su rango nativo, lo que requiere proyectar el modelo más allá de las condiciones ambientales conocidas. Además, en muchos casos se viola el supuesto de equilibrio ambiental, ya que las especies invasoras no siempre están en equilibrio con el ambiente en sus áreas de distribución invasivas, lo que introduce incertidumbres adicionales cuya consideración es necesaria en las predicciones (Václavík y Meentemeyer, 2012).

JUSTIFICACIÓN

Debido a la cantidad de factores que influyen en el proceso de modelado de nicho y la necesidad de evaluar un gran número de especies exóticas, es necesaria la evaluación de la importancia relativa de diversos factores que influyen en el nivel de precisión de los y selección de modelos de distribución de nicho ecológico para especies exóticas acuáticas en las condiciones ambientales relevantes para México. En esta investigación, abordamos esta cuestión mediante un enfoque de simulación. La simulación de especies virtuales nos permite simular especies con probabilidades de ocurrencia y respuestas ambientales conocidas, lo que permite aislar con precisión los efectos de los factores específicos (Meynard et al. 2019).

Se realizarán modelos para una selección de especies acuáticas, así como datos simulados computacionalmente. Se utilizará el algoritmo MaxEnt, registros de presencia disponibles en bases de datos globales y variables bioclimáticas, topográficas, litológicas y de influencia humana.

A partir de los registros de presencia y las variables ambientales predictoras se realizarán modelos de nicho para cada especie con los niveles de factores. Cada combinación de modelos será evaluada de acuerdo con su capacidad predictiva.

Este estudio permitirá determinar diversos aspectos que influyen en los modelos de nicho ecológico para especies exóticas acuáticas, particularmente en un contexto ambiental específico como el de México. La evaluación sistemática de los factores involucrados ayudará a entender cómo su selección afecta la precisión y aplicabilidad de las predicciones de estos modelos. Asimismo, se espera que los resultados arrojen luz sobre el grado de transferibilidad de los ENM, lo cual es fundamental para su utilización en la gestión de especies invasoras, facilitando la identificación de áreas de riesgo y la implementación de estrategias de conservación más efectivas.

HIPÓTESIS

La capacidad predictiva de los modelos de distribución de especies exóticas invasoras es influenciada, en orden decreciente de importancia, por la extensión del área de calibración, la filtración espacial de los registros de presencia, la colinearidad de las variables predictoras y por el valor de regularización.

OBJETIVO DEL TRABAJO

Objetivo general

-Evaluar la importancia relativa de diversos factores que influyen en el modelado de la distribución potencial de especies exóticas acuáticas en México.

Objetivos específicos

-Evaluar la importancia relativa de la extensión del área de calibración en el nivel de predicción de modelos de distribución potencial de especies exóticas acuáticas.

-Evaluar la importancia relativa de la extensión de la filtración espacial de los registros de presencia en el nivel de predicción de modelos de distribución potencial de especies exóticas acuáticas.

-Evaluar la importancia relativa de la colinearidad de las variables predictoras en el nivel de predicción de modelos de distribución potencial de especies exóticas acuáticas

-Evaluar la importancia relativa del valor de regularización predictoras en el nivel de predicción de modelos de distribución potencial de especies exóticas acuáticas

MATERIAL Y MÉTODOS

El tipo de estudio es computacionalmente intensivo y se encuentra centrado en el procesamiento de datos de biodiversidad para predecir la distribución potencial de especies exóticas acuáticas en México utilizando el algoritmo de modelado MaxEnt. El proyecto consideró la evaluación de distintos factores que influyen en el proceso de modelado. La influencia de los factores sobre el desempeño de los modelos se evaluó mediante su capacidad predictiva a través de la simulación de especies virtuales.

La simulación de especies virtuales, para las cuales se conoce con exactitud su relación ocurrencia-ambiente, es una alternativa cada vez más utilizada para comprender la influencia de diversos aspectos de la distribución de las especies sobre la evaluación y selección de los modelos, incluyendo aspectos de las invasiones biológicas (Meynard et al. 2019).

El uso de simulaciones desempeña un papel fundamental en la evaluación de la capacidad predictiva de los modelos antes de aplicarlos a especies reales. Estas simulaciones permiten crear un entorno controlado donde podemos manipular variables clave y comprender cómo afectan la precisión de nuestros modelos. Esto es esencial porque trabajar directamente con especies reales puede ser costoso, demorado y, en algunos casos, éticamente problemático. Al utilizar simulaciones, podemos generar datos sintéticos que reflejen la complejidad de la realidad y someter nuestros modelos a una amplia variedad de condiciones, lo que nos brinda información valiosa sobre su rendimiento bajo diversas circunstancias, en este caso la extrapolación de los modelos en el espacio (invasiones). Estos resultados obtenidos de simulaciones pueden luego guiar la selección de modelos y estrategias óptimas para la conservación de especies reales, ayudando así a tomar decisiones informadas y eficaces en la gestión de la biodiversidad y la preservación de ecosistemas (Zurell et al. 2010; Meynard et al. 2019).

Simulación de especies virtuales

Se simularon 100 especies virtuales utilizando el paquete *virtualespecies* en R (Leroy et al. 2016) y las 19 variables bioclimáticas de WorldClim 2.0 (Fick y Hijmans 2017; www.worldclim.org). Estas variables contienen información sobre la precipitación y temperatura del aire y abarcan factores climáticos críticos que influyen en la distribución de especies y son ampliamente utilizadas en estudios de SDM. Ya que no existen datos globales sobre la temperatura de cuerpos de agua continentales, la temperatura del aire se ha utilizado exitosamente ya que se encuentra directamente relacionada con la temperatura del agua (Morehouse et al. 2013).

Se utilizó una resolución de 2.5 minutos de arco para simular los nichos de las especies, ya que esta resolución ofrece un equilibrio entre el detalle espacial y la eficiencia computacional. La región considerada se restringió al continente americano debido a su amplia gama de hábitats y condiciones climáticas. Decidimos incluir variables que combinan precipitación y temperatura, a pesar de una tendencia a excluirlas de los estudios (Booth et al. 2022). Se ha demostrado que estas variables influyen considerablemente en la distribución de especies. Por ejemplo, la precipitación durante los meses más húmedos y más secos son factores clave que afectan la distribución global de plantas (Huang et al. 2021), mientras que la temperatura media del trimestre más seco ha sido identificada como la variable más influyente para explicar la distribución de peces dulceacuícolas en México (Castillo-Torres et al. 2017).

El concepto de nicho ecológico se basa en el principio de que cada especie tiene un conjunto óptimo de condiciones ambientales donde puede sobrevivir, crecer y reproducirse (Peterson y Soberón 2012). Para definir las preferencias ambientales de cada especie, se seleccionaron aleatoriamente de dos a cinco de las 19 variables bioclimáticas y luego se sometieron a un análisis de componentes principales (PCA) utilizando descomposición en valores singulares. Los dos componentes principales derivados se utilizaron para definir el nicho (Fig. 2A). Diferentes especies están influenciadas por distintos factores ambientales (Huang et al. 2021), y los factores primarios que determinan sus distribuciones pueden variar significativamente. La selección aleatoria entre 2 y 5 variables ayuda a simular esta variabilidad. Para los dos primeros componentes principales

fue definida una curva de respuesta gaussiana, con valores seleccionados aleatoriamente para la media (que representa los valores más adecuados para la especie) y la desviación estándar (que indica la tolerancia fisiológica de la especie).

La función de idoneidad inicial (proyección del nicho al espacio geográfico), que describe la relación entre la especie y el ambiente y puede tener cualquier rango de valores, se convirtió en una probabilidad de ocurrencia mediante una función logística (Fig. 2B y 2C). Después de esta transformación, la probabilidad de ocurrencia se encuentra acotada entre 0 y 1 (es decir, la probabilidad de que la especie esté presente en un sitio dado un conjunto específico de condiciones ambientales en ese sitio). Específicamente, utilizamos una transformación logística con β (punto de inflexión) fijo en 0.5 y α (pendiente de la curva) seleccionado aleatoriamente. Este método funciona como un equivalente a la aplicación de umbrales, pero ofrece una representación ecológica más realista al suavizar la transición entre condiciones adecuadas e inadecuadas y al simular procesos aleatorios que afectan las ocurrencias de las especies (Leroy et al. 2016; Meynard et al. 2019).

Un aspecto esencial, de acuerdo con la teoría del nicho, es la distinción entre el nicho fundamental, potencial y realizado. El nicho fundamental representa el rango completo de condiciones ambientales bajo las cuales una especie puede sobrevivir y reproducirse, mientras que el nicho potencial corresponde al subconjunto de esas condiciones que realmente existen en un momento dado (Jackson y Overpeck 2000). El nicho potencial incluye áreas que son ambientalmente adecuadas pero que pueden no estar ocupadas por la especie debido a factores adicionales como limitaciones en la dispersión y/o la influencia de interacciones biológicas (Barve et al. 2011; Soberón y Nakamura 2009). El nicho realizado es un subconjunto del nicho potencial después de considerar estos factores (Jiménez y Soberón 2022; Soberón y Arroyo-Peña 2017). La capacidad de dispersión, en particular, puede limitar la habilidad de una especie para explorar regiones favorables, un aspecto esencial para la simulación de áreas nativas e invasivas. Para tener en cuenta esto, se definió el área accesible de cada especie mediante la programación de un autómata celular (CA) para encontrar áreas potenciales donde la especie se encuentre en equilibrio ambiental. Para lograr esto, se asignó a cada especie una capacidad de dispersión aleatoria dentro del rango de 50 a 250 km. Esta capacidad de dispersión es fija para cada especie y

representa la distancia máxima que puede dispersar en cada iteración. Usando esta información, el área accesible se simuló seleccionando celdas mediante un ensayo de Bernoulli basado en su probabilidad de ocurrencia. El punto con la probabilidad de ocurrencia más alta en el paisaje sirvió como el punto de selección inicial, y se generó un buffer alrededor de él con un radio equivalente a la capacidad de dispersión asignada. Las celdas subsiguientes se seleccionaron dentro de este buffer mediante ensayos Bernoulli adicionales, generando un nuevo buffer. Este proceso se repitió hasta que no se pudieran agregar más celdas (equilibrio ambiental). El buffer resultante se consideró el área accesible para la especie, representando los sitios que ha explorado potencialmente. La función del CA se basa en la presencia de celdas con probabilidades de ocurrencia suficientemente altas. En regiones con altas probabilidades de ocurrencia, el buffer se expandirá, permitiendo a la especie explorar áreas adyacentes. Por el contrario, en regiones donde las probabilidades de ocurrencia son más bajas, se añaden menos celdas, y el crecimiento del buffer tiende a disminuir. Este proceso es análogo a las dinámicas ecológicas reales, donde las especies tienen más probabilidades de colonizar y establecerse en regiones que proporcionan condiciones óptimas para la supervivencia y reproducción. Este proceso se repitió hasta 5 veces (es decir se generaron hasta 5 áreas; Fig. 2D), una de las cuales fue seleccionada de forma aleatoria como el área nativa y otra como el área invasiva (Fig. 2E), asumiendo conservadurismo de nicho en este último caso (Liu et al. 2020b). Especies para las cuales solo se encontró un área fueron descartadas.

La magnitud del sesgo de muestreo espacial se simuló utilizando el conjunto de datos del Índice de Influencia Humana (HII; human influence index, WCS y CIESIN 2005), que representa la probabilidad de selección por celda. Estos valores se dividieron por el valor máximo, resultando en un rango de sesgo de muestreo entre 0 y 1 (Fig. 2F).

Para cada especie, se obtuvieron 100 registros de ocurrencia del área nativa al seleccionar aleatoriamente celdas (con reemplazo) mediante un ensayo de Bernoulli, con una probabilidad de éxito igual a:

$$p(x) = s(x)b(x)r(x)$$

donde,

$p(x)$ es la probabilidad de muestreo de la celda

$s(x)$ es la probabilidad de ocurrencia de la especie en esa celda

$b(x)$ es la fuerza relativa del sesgo espacial de muestreo, y

$r(x)$ es una variable binaria con valores 1 dentro del área nativa de la especie y 0 fuera (Warren et al. 2020).

Este proceso se repitió cinco veces para cada especie para tener en cuenta la estocasticidad en el proceso de muestreo.

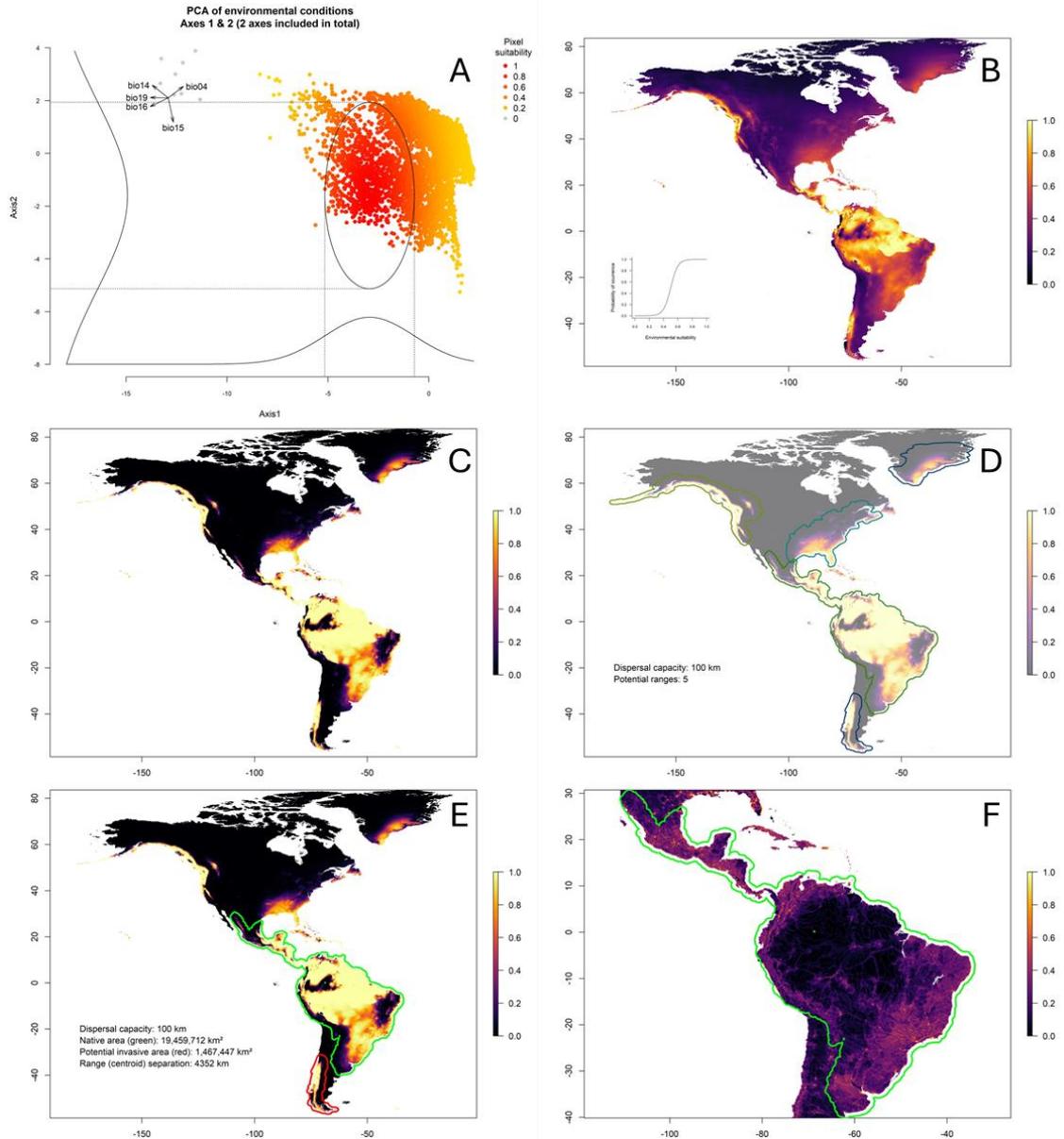


Figura 2. Esquema del proceso de simulación de área nativa e invasiva para una de las especies. A) Dimensiones del nicho [5 variables seleccionadas aleatoriamente]; B) Idoneidad del hábitat mediante la proyección del nicho al espacio geográfico (recuadro: transformación logística), C) Probabilidad de ocurrencia de la especie, D) Áreas accesibles potenciales en equilibrio ambiente (5 encontradas para la especie), E) Área nativa (verde) e invasiva (rojo) seleccionadas, y F) Sesgo de muestreo espacial en el rango nativo.

Modelos de nicho ecológico o distribución de especies

La simulación de cada especie virtual tuvo como objetivo representar aspectos realistas que afectan a las especies reales, incluyendo la delimitación de una región invasiva. Al evaluar los modelos, el objetivo fue reflejar escenarios de modelado del mundo real, reconociendo que los parámetros exactos que influyen en las distribuciones de especies a menudo son desconocidos (Meynard et al. 2019). Por ejemplo, la inclusión del sesgo espacial en la simulación de la ocurrencia de cada especie virtual fue para simular las condiciones de muestreo desiguales debido a factores como la accesibilidad y el esfuerzo de muestreo. El proceso de depuración aplicado durante el modelado (ver más adelante) es uno de los métodos existentes destinados a reducir este sesgo (Aiello-Lammens et al. 2015). De manera similar, la selección de variables es un paso crítico en la modelización de distribuciones de especies (SDM), sin embargo, las verdaderas variables explicativas que limitan las distribuciones de cada especie rara vez se conocen (Inman et al. 2021). Al excluir variables basadas en colinealidad en lugar de utilizar los verdaderos factores influyentes, buscamos reflejar la incertidumbre inherente que existe en los escenarios de selección de variables del mundo real. Aunque simular todos los aspectos que influyen en las distribuciones de especies es conceptual y computacionalmente complejo, nuestro objetivo fue incluir algunos aspectos esenciales para proporcionar un marco de evaluación más realista.

Se construyeron modelos de distribución de especies para cada uno de los conjuntos de datos de ocurrencia (Fig. X) utilizando el algoritmo de Máxima Entropía MaxEnt v3.4.4 (Phillips et al. 2006) en el paquete *dismo* de R (Hijmans et al. 2017). Los diferentes niveles de los cuatro factores establecidos a continuación implicaron la evaluación de 2000 modelos por especie (400 modelos x 5 réplicas de ocurrencia). Las configuraciones restantes se establecieron en sus valores predeterminados. Los factores se describen en las siguientes secciones.

- Filtración espacial de los registros de ocurrencia: 0, 5, 10, 15 y 20 km
- Extensión del background: 25, 50, 100, 200, 300, 500, 700 y 1000 km
- Colinearidad de las variables predictoras: umbral de 0.7 y 1 (Pearson).

-Multiplicador de regularización β : 0.5, 1, 2, 3 y 5

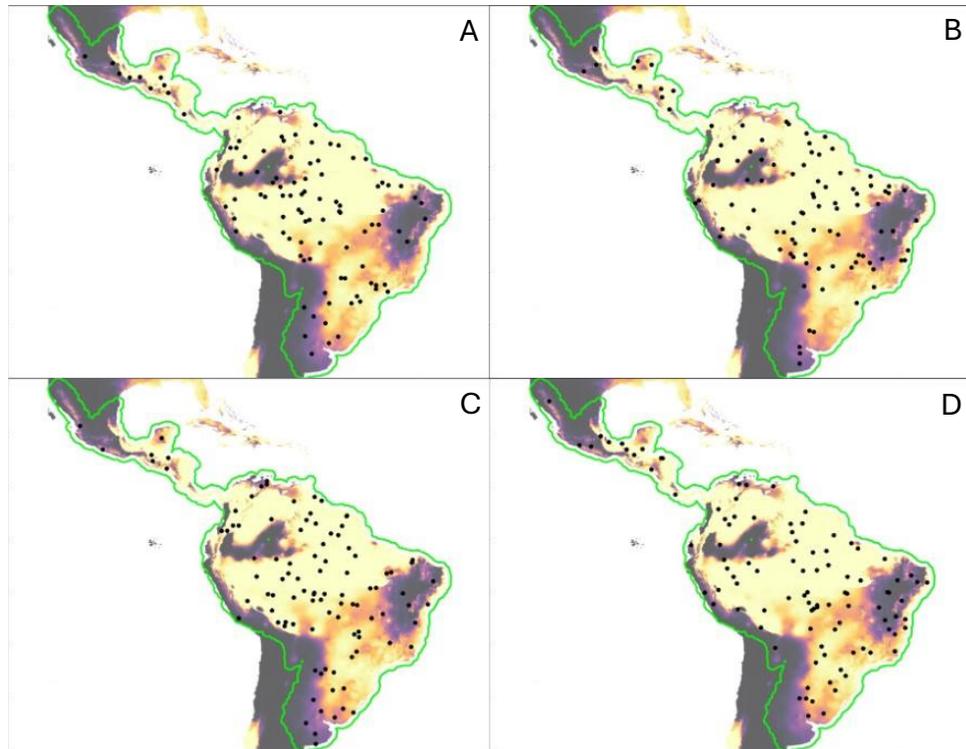


Fig. 3. Mapa que muestra los registros de ocurrencia utilizados para los modelos de distribución de especies (se muestran cuatro de las 5 réplicas utilizadas). Cada punto indica una ubicación de presencia en el área nativa, y las diferentes réplicas permiten evaluar la variabilidad espacial y el impacto del muestreo en los resultados del modelado.

Filtración espacial de los registros de ocurrencia

El filtrado espacial de registros de ocurrencia es uno de los principales métodos utilizados para reducir el sesgo de muestreo en conjuntos de datos derivados de colecciones biológicas (Taylor et al. 2020). Los registros de ocurrencia fueron filtrados utilizando el paquete *spThin*, el cual implementa un algoritmo aleatorio para mantener solo aquellos registros que se encuentran a una distancia mínima entre sí definida por el usuario (Fig. 4; Aiello-Lammens et al. 2015). Para esto se seleccionaron el conjunto de registros con el

mayor número de ocurrencias después de correr el algoritmo 10 veces con las siguientes distancias de filtrado: 0 (sin filtración), 5, 10, 15 y 20 km.

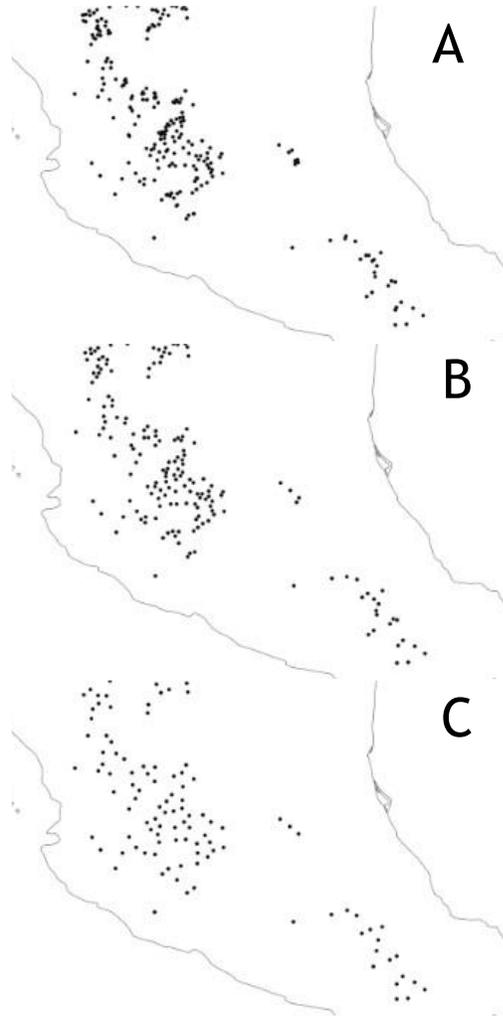


Fig. 4. Mapa que muestra los resultados del filtrado espacial de registros de ocurrencia con distancias de separación de A) 0 km (sin filtración), B) 10 km y C) 20 km. Las diferentes distancias de filtrado permiten visualizar el impacto del proceso en la distribución espacial de los registros utilizados para los modelos de distribución de especies.

Colinealidad de predictores

La selección de variables predictoras influye directamente sobre la capacidad predictiva de los modelos, así como su transferibilidad espacial y temporal (Peterson y Nakazawa, 2008; Petitpierre et al. 2017). Aún no existe un consenso sobre cómo se debe tratar la colinealidad de las variables predictoras en los modelos de nicho ecológico, dado que el uso de variables altamente correlacionadas puede influir en el rendimiento del modelo (Feng et al. 2019). Se comparó el efecto de la colinealidad de las predictoras calibrando modelos utilizando todas las variables disponibles o seleccionando variables con coeficientes de correlación de Pearson inferiores a 0.7 mediante la función *vifcor* del paquete *usdm* (Naimi et al. 2014).

Área de calibración

MaxEnt funciona con el principio de máxima entropía, por lo cual es necesario proporcionar datos ambientales aleatorios para su caracterización (Merow et al. 2013). La extensión de esta área influye directamente en el rango de condiciones ambientales disponibles para el entrenamiento del modelo, es de gran importancia considerar cuidadosamente el espacio geográfico en la calibración del modelo, ya que se ha demostrado que la extensión del área de calibración repercute en los análisis de desempeño posteriores del modelo (Thuiller et al. 2004). Dado que el área accesible para una especie suele ser desconocida, se evaluó el efecto de la extensión del área de calibración mediante la selección aleatoria de 10,000 celdas (por defecto) dentro de un radio de 25, 50, 100, 200, 300, 500, 700 y 1000 km alrededor de los registros de ocurrencia (Fig. 5). Se utilizaron todas las celdas disponibles para áreas de calibración con menos de 10,000 puntos de fondo.

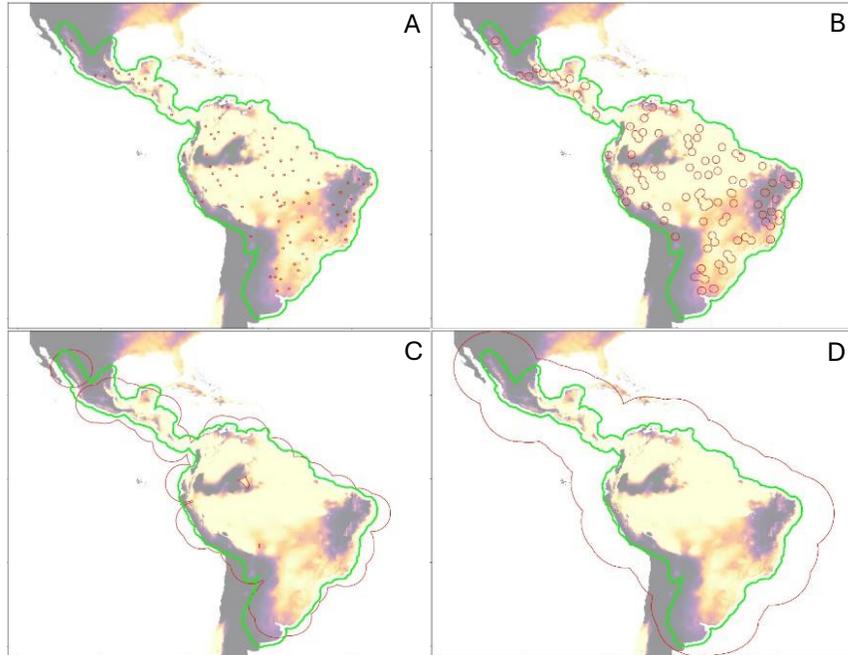


Fig. 5. Mapa que ilustra el efecto de la extensión del área de calibración en los modelos de distribución de especies. Se muestran áreas de calibración (área en color rojo) con radios de A) 25 km, B) 100 km, C) 500 km y D) 1000 km alrededor de los registros de ocurrencia.

Multiplicador de regularización (RM).

MaxEnt utiliza el principio de máxima entropía, que busca una función para cada variable que coincida con los datos empíricos y no sea informativa al máximo en otros lugares y tenga una media igual a la de los datos empíricos. No obstante, el cumplimiento estricto de este requisito puede dar lugar a modelos sobreajustados, es decir que se ajusten demasiado a los datos de entrenamiento. Para evitar esto, MaxEnt utiliza una regularización Lasso para restringir las estimaciones de los parámetros dentro de cierto intervalo alrededor de la media muestral en lugar de igualarla exactamente. Este sobreajuste puede reducirse especificando un valor de RM que penalice el uso de parámetros adicionales (Phillips et al. 2006; Warren y Seifert 2011). El efecto de este factor se evaluó mediante el uso de valores RM = 0.5, 1 (valor predeterminado), 2, 3 y 5.

Modelos no sesgados.

Además de evaluar las contribuciones relativas de los factores mencionados anteriormente, examinamos la precisión de un modelo para cada especie construido bajo un escenario en el que las características ecológicas de la especie son conocidas (denominado modelo no sesgado). Las consideraciones para construir este modelo incluyeron el uso de 1000 puntos de datos de ocurrencia (en lugar de 100) muestreados sin sesgo de muestreo espacial, utilizando el área accesible de la especie como área de calibración, empleando únicamente las variables ambientales que definen el nicho de la especie como predictores, y evaluando el valor del multiplicador de regularización (RM) con los mismos 5 valores previamente definidos.

Evaluación del desempeño

El desempeño del modelo se evaluó utilizando tanto la precisión funcional como la precisión discriminatoria (Warren et al. 2020). La precisión funcional (precisión real del modelo) se calculó a través de la correlación de rango de Spearman entre la probabilidad de ocurrencia real y la inferida por MaxEnt (con la transformación log-log complementaria (cloglog)) a lo largo del área nativa y el área invasiva para cada uno de los modelos (Fig. 6). Para áreas grandes, se seleccionaron 25,000 celdas al azar debido a restricciones computacionales. Ensayos preliminares mostraron que utilizar 25,000 celdas seleccionadas al azar producen resultados muy cercanos a los obtenidos utilizando toda el área, con una correlación de Pearson de 0.997 entre los valores de toda el área y los de las 25,000 celdas seleccionadas al azar.

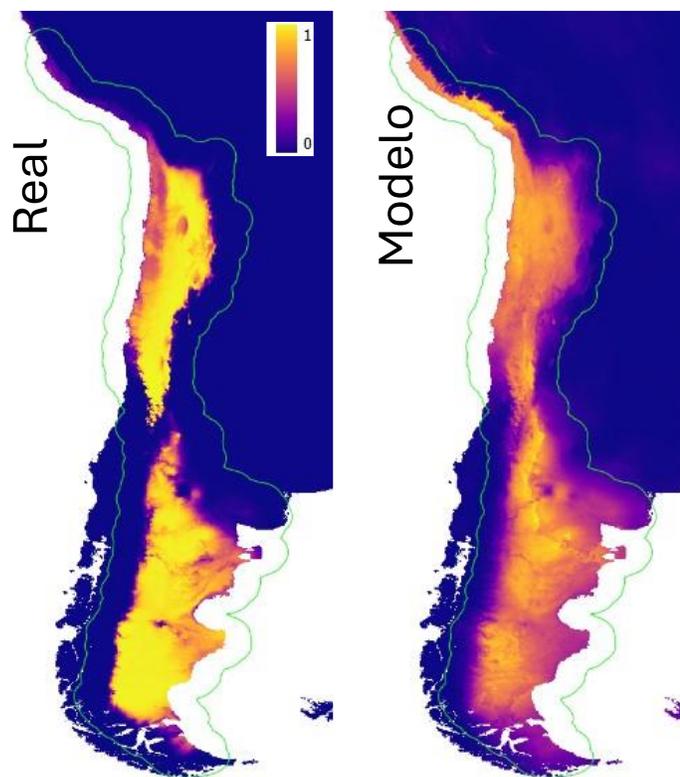


Fig. 6. Evaluación del desempeño del modelo en términos de precisión funcional. La precisión funcional se calculó utilizando la correlación de rango de Spearman entre la probabilidad de ocurrencia real y la estimada por MaxEnt, aplicando la transformación log-log complementaria (cloglog). Aquí se muestra el modelo utilizando un área de calibración de 500 km, una filtración espacial de 10km, sin filtrar variables por colinearidad, y un multiplicador de regularización de 2 (la correlación de Spearman en este ejemplo es de 0.887).

La precisión de discriminación se calculó utilizando validación cruzada, donde los datos se dividieron en cuatro grupos utilizando dos criterios: partición aleatoria y partición por bloques geográficos (Fig. 7), a través del paquete *ENMeval* en R (Muscarella et al. 2014; Kass et al. 2021). Se ha sugerido que este último método proporciona una mejor medida de la transferibilidad del modelo, debido a que se reduce la correlación espacial de los datos, permitiendo hacer uso de la heterogeneidad presente en ellos, lo cual es un aspecto crítico para evaluar si el modelo se utilizará de una forma predictiva fuera de las condiciones en las que es calibrado (Wenger y Olden, 2012; Warren et al. 2020). Esto es de gran importancia en el caso de especies exóticas y de estudios de cambio climático,

donde es necesario transferir el modelo en condiciones distintas a los datos de calibración (Muscarella et al. 2014).

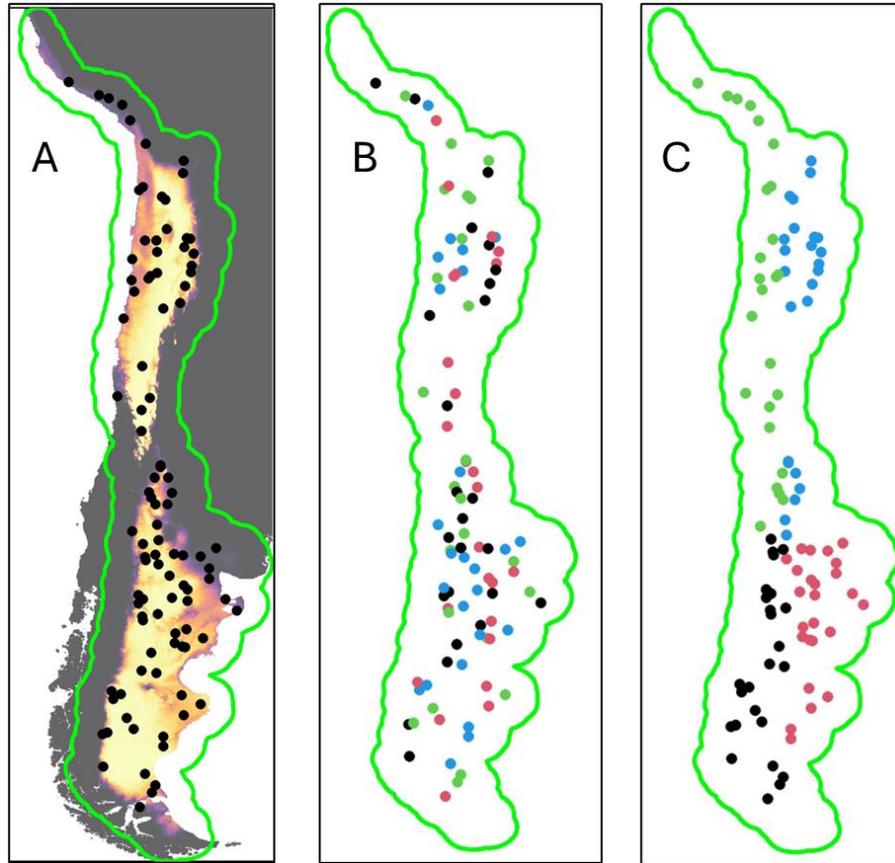


Figura 7. Mapa que muestra los puntos de presencia originales (A) y los dos tipos de partición de datos utilizados para la validación cruzada en los modelos de distribución de especies. En la partición aleatoria, los puntos se dividen en cuatro grupos representados por diferentes colores (B). En la partición por bloques geográficos, los datos se dividen en secciones espaciales, reduciendo la correlación espacial y favoreciendo la transferibilidad del modelo (C).

El índice de Boyce se utilizó como métrica de evaluación (Hirzel et al. 2006), calculado para cada uno de los cuatro grupos, el valor promedio entre estos grupos se utilizó como la medida global de precisión discriminadora.

Se seleccionó el índice de Boyce como una métrica adicional para la capacidad de identificar los mejores modelos en función de la precisión de discriminación, particularmente cuando se consideran variaciones en los conjuntos de datos que pueden no estar perfectamente alineadas con la precisión real. Aunque conocemos la precisión real del modelo, el uso del índice de Boyce nos permite evaluar cómo se desempeñan los modelos en un contexto comparativo, ofreciendo una perspectiva adicional sobre el rendimiento del modelo en escenarios que simulan condiciones del mundo real.

Análisis estadístico

Para evaluar la significancia de los hallazgos no se utilizaron *p-values* debido a las críticas bien documentadas sobre su uso (Hurlbert et al. 2019) y su inapropiada utilidad en el contexto de estudios de simulaciones (White et al. 2014). En su lugar, el que se centra en la evaluación de las contribuciones relativas y las relaciones entre los factores examinados mediante modelos lineales, utilizando el método LMG implementado en el paquete *relaimpo*. Este método se basa en el R^2 secuencial y aborda la dependencia del orden de los regresores al promediar sobre estos ordenamientos utilizando promedios simples no ponderados (Grömping 2006). La variable de respuesta fue la precisión funcional de los SDM en el área nativa e invasiva, y las variables explicativas fueron la extensión del área de calibración, el filtrado espacial de los registros de ocurrencia, la colinealidad de los predictores y el parámetro de regularización β tratados como variables categóricas. Este enfoque nos permitió analizar la variabilidad y las contribuciones de diferentes factores que influyen en el rendimiento del modelo de manera más robusta, en lugar de depender de las pruebas tradicionales de hipótesis con *p-values* que son sensibles al tamaño de muestra.

Además, se evaluó la correlación entre la precisión funcional y la precisión discriminatoria utilizando la correlación de Pearson para evaluar la capacidad de la métrica de discriminación para seleccionar los modelos con mejor desempeño. Estos análisis se realizaron individualmente para cada especie y de forma independiente para el área nativa e invasiva, dado que no esperamos que todas las especies se vean afectadas por los factores

evaluados de la misma manera; algunas pueden experimentar un sesgo de muestreo más pronunciado, otras pueden tener áreas accesibles más pequeñas, etc. Esta elección metodológica está alineada con nuestro objetivo de proporcionar un análisis integral y matizado de los factores que influyen en los modelos de distribución de especies.

RESULTADOS

A continuación, se muestran los resultados obtenidos en el presente estudio. Con el fin de una mayor claridad se presenta primero lo obtenido en el área nativa y posteriormente los resultados del análisis del área invasiva.

Área nativa

En general, los modelos no sesgados mostraron una alta precisión dentro del área accesible (nativa) de las especies, con el 80% de estos modelos presentando valores de precisión funcional superiores a 0.9. La precisión funcional mediana de estos modelos fue alta, con una correlación de Spearman de 0.968, y el rango de puntuaciones de precisión varió entre 0.560 y 0.999 (Fig. 8). En contraste, el resto de los modelos demostraron un rendimiento más variable tanto dentro de las especies como entre ellas. La precisión funcional máxima entre las especies varió entre 0.178 y 0.996, con una mediana de 0.902. La precisión funcional mínima entre las especies varió entre -0.959 y 0.699, con una mediana de -0.256. Esta considerable variabilidad en la precisión también se reflejó en el rango de valores de precisión funcional dentro de las especies (es decir, la diferencia entre el valor máximo y el mínimo para cada especie), mostrando valores entre 0.291 y 1.87, con una mediana de 1.045. En general, estos hallazgos destacan el considerable rango en la precisión del modelo entre las especies.

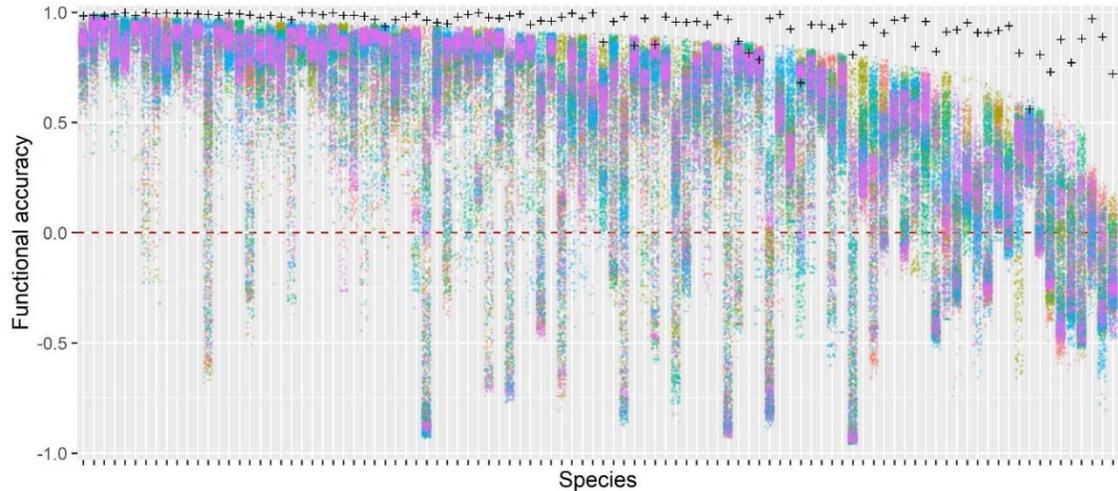


Figura 8. Precisión funcional de los modelos por especie en el área nativa (400 modelos \times 5 réplicas). El eje Y representa la precisión funcional (medida como la correlación de rango de Spearman entre la probabilidad de ocurrencia real y la estimada por MaxEnt), y el eje X representa diferentes especies. Los puntos de datos se representan con cinco colores diferentes (uno para cada réplica). Las especies están ordenadas de mayor a menor en función del mejor modelo. La precisión del modelo no sesgado se denota para cada especie con un símbolo "+". La línea roja discontinua representa el valor esperado por azar.

Entre las especies analizadas, 30 mostraron consistentemente modelos con una precisión funcional positiva, caracterizada por coeficientes de correlación de Spearman superiores a 0. Las otras 70 especies mostraron al menos un modelo con precisión funcional negativa. Además, 13 especies fueron particularmente notables, ya que la mayoría de sus modelos produjeron valores de precisión funcional negativa. Esta disparidad en la precisión del modelo subraya las respuestas diversas de las especies al área de calibración y otros factores, lo que conduce a variaciones en la precisión de los modelos generados.

Varianza explicada

El área de calibración resultó ser el factor más importante en términos de precisión real del modelo, con una mediana del 50.46% de varianza explicada y un rango del 2.49% al 92.99%. El valor del multiplicador de regularización (RM) jugó un papel menos destacado, con una mediana de varianza explicada del 3.65% (rango: 0.01% a 48.05%).

Mientras que la colinealidad de los predictores y el filtrado espacial mostraron un impacto insignificante en la precisión real del modelo, contribuyendo cada uno con una mediana de varianza explicada del 0.41% y 0.04%, respectivamente (Fig. 9).

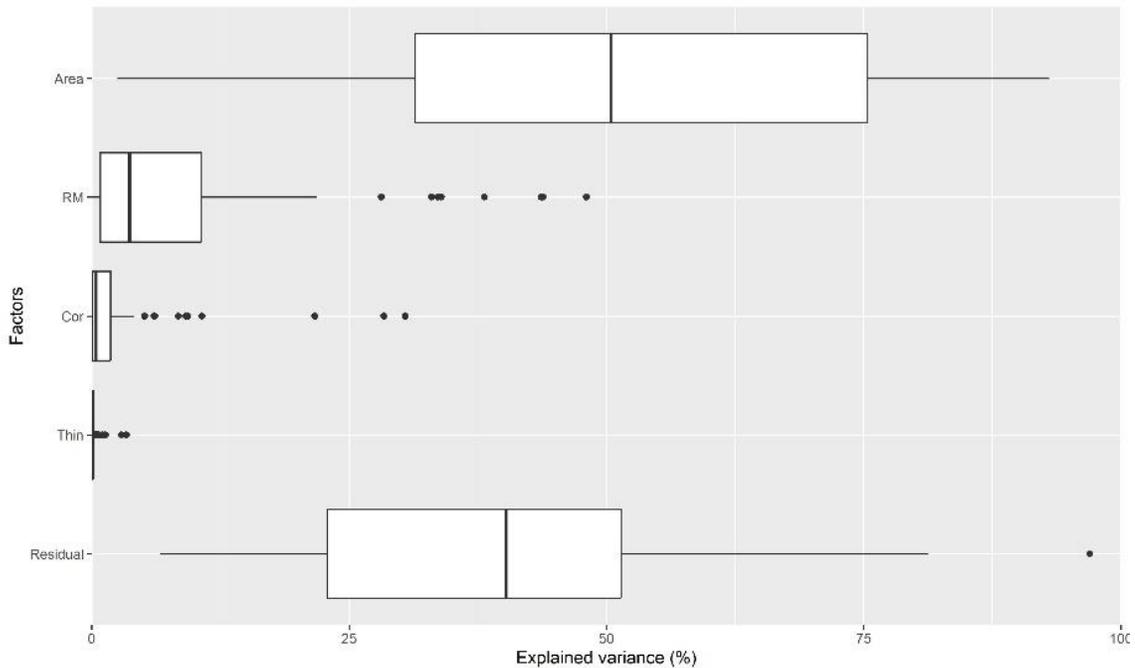


Figura 9. Porcentaje de varianza explicado por los cuatro factores analizados y la varianza residual (no explicada) sobre la precisión funcional en el área nativa. El eje X representa el porcentaje de varianza explicada, y el eje Y muestra los factores analizados. Los diagramas de caja resumen la distribución de estos porcentajes para todas las especies. Área: Área de calibración; RM: multiplicador de regularización; Cor: Colinealidad de predictores; Thin: Filtrado espacial de ocurrencias.

Correlación entre la precisión funcional y discriminatoria

Exploramos la correlación entre la precisión funcional y la precisión de discriminación para las 100 especies bajo dos escenarios diferentes de partición de datos (Fig. 10). Al considerar la partición de datos aleatoria, se observó una amplia gama de correlaciones. Los valores mínimo y máximo fueron -0.58 y 0.82, mientras que la mediana fue 0.46. De manera notable, 19 especies exhibieron correlaciones negativas.

Cuando la partición de datos se basó en bloques geográficos, las correlaciones también mostraron variabilidad, con un rango de -0.73 a 0.69, y una mediana de 0.32. Aquí, 25 especies mostraron correlaciones negativas entre la precisión funcional y la precisión de discriminación.

Sin embargo, un hallazgo notable surgió a pesar de la correlación relativamente baja entre la precisión funcional y la precisión de discriminación. Para 76 especies, los mejores modelos basados en la evaluación de partición aleatoria de datos mostraron precisiones funcionales superiores a 0.5. De manera paralela, para 71 especies, los mejores modelos seleccionados con la estrategia de partición de datos por bloques geográficos lograron precisiones funcionales superiores a 0.5. Esta observación destaca que, en ciertos casos, la elección de la estrategia de partición de datos permitió la selección de modelos con una precisión funcional razonablemente alta, incluso cuando su precisión de discriminación mostró una alineación limitada con la métrica de precisión funcional.

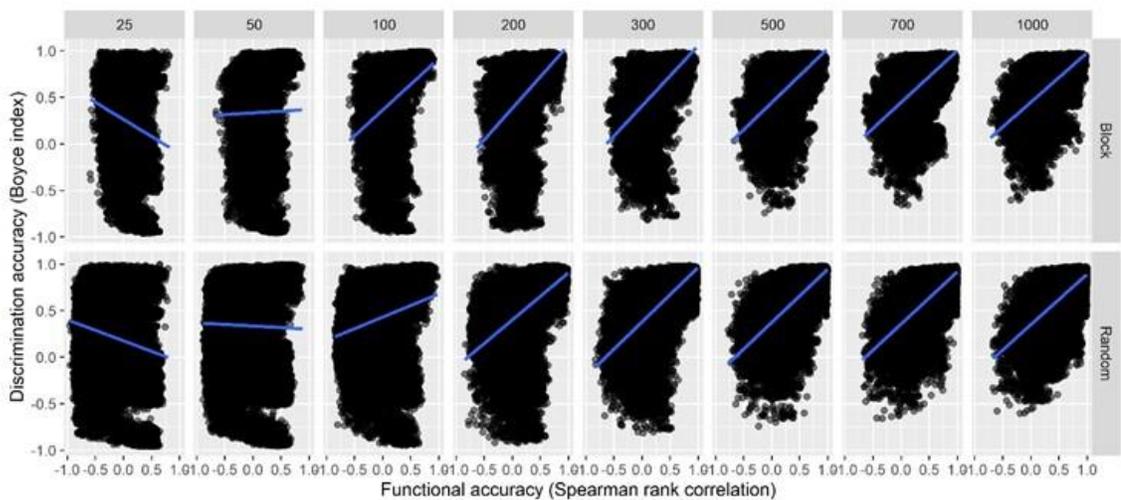


Figura 10. Relación entre la precisión funcional y discriminatoria de los modelos en el área nativa para las 100 especies bajo diferentes extensiones de área de calibración y métodos de partición. El eje x representa la precisión funcional, dividida por la extensión del área de calibración, mientras que el eje y representa la precisión discriminatoria, dividida por el método de partición (aleatorio y bloques geográficos). Las líneas azules, obtenidas mediante regresión de Mínimos Cuadrados Ordinarios (OLS), proporcionan una representación visual de la correlación entre las dos métricas de precisión para cada subconjunto de puntos.

Área invasiva

Los modelos no sesgados mostraron una precisión relativamente alta (aunque menor que en el área nativa) dentro del área invasiva de las especies, con el 54% de estos modelos presentando valores de precisión funcional superiores a 0.9. La precisión funcional mediana de estos modelos fue alta, con una correlación de Spearman de 0.912, y el rango de puntuaciones de precisión varió entre -0.2 y 0.997 (Fig. 11). El resto de los modelos también mostró un rendimiento variable tanto dentro de las especies como entre ellas. La precisión funcional máxima entre las especies varió entre 0.295 y 0.997, con una mediana de 0.833. La precisión funcional mínima entre las especies varió entre -0.96 y -0.12, con una mediana de -0.63. Esto también se reflejó en el rango de valores de precisión funcional dentro de las especies (la diferencia entre el valor máximo y el mínimo para cada especie), mostrando valores entre 0.603 y 1.88, con una mediana de 1.44.

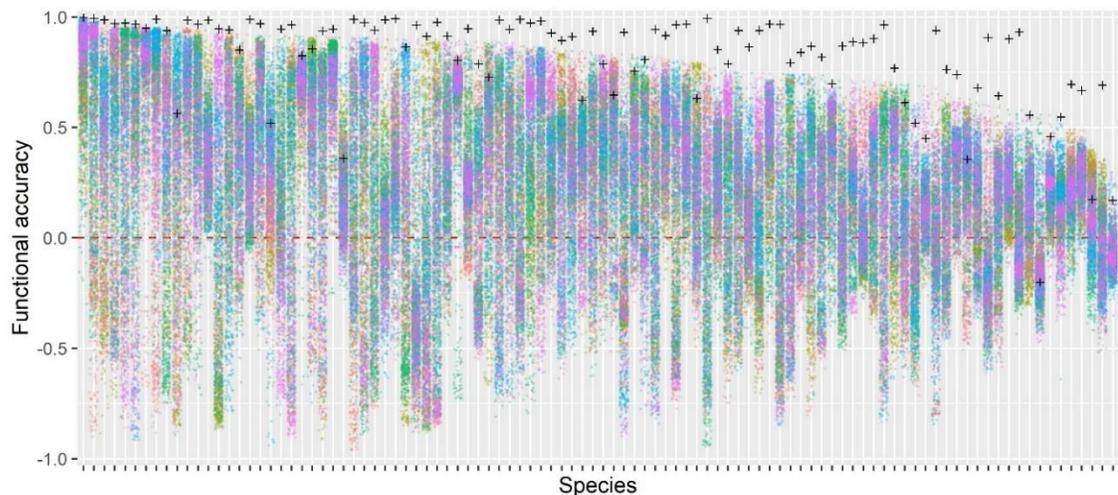


Figura 11. Precisión funcional de los modelos por especie en el área invasiva (400 modelos \times 5 réplicas). El eje Y representa la precisión funcional (medida como la correlación de rango de Spearman entre la probabilidad de ocurrencia real y la estimada por MaxEnt), y el eje X representa diferentes especies. Los puntos de datos se representan con cinco colores diferentes (uno para cada réplica). Las especies están ordenadas de mayor a menor en función del mejor modelo. La precisión del modelo no sesgado se denota para cada especie con un símbolo "+". La línea roja discontinua representa el valor esperado por azar.

A diferencia de los modelos en el área nativa, ninguna especie mostró consistentemente modelos con una precisión funcional positiva, caracterizada por coeficientes de correlación de Spearman superiores a 0. Mientras que solo en 13 especies se generaron una mayoría de modelos con precisión funcional negativa.

Varianza explicada

Al igual que en el área nativa, el área de calibración resultó ser el factor más importante en términos de precisión real del modelo, aunque disminuyó considerablemente en el valor con una mediana del 18.1% de varianza explicada (rango: 0.7% a 66.6%). El valor multiplicador de regularización (RM) jugó un papel de menor importancia, con una mediana de varianza explicada del 3.3% (rango: 0.01% a 33%). Mientras que la colinealidad de los predictores y el filtrado espacial mostraron un impacto insignificante en la precisión real del modelo, contribuyendo cada uno con una mediana de varianza explicada del 1.3% y 0.05%, respectivamente (Fig. 12).

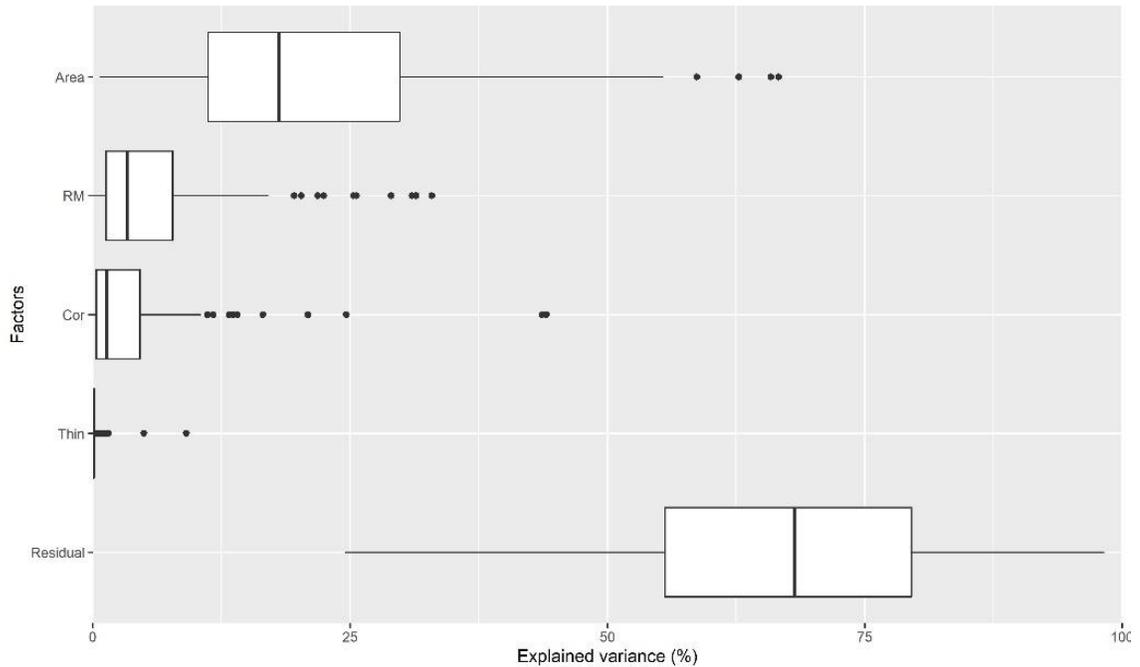


Figura 12. Porcentaje de varianza explicado por los cuatro factores analizados y la varianza residual (no explicada) sobre la precisión funcional en el área invasiva. El eje X representa el porcentaje de varianza explicada, y el eje Y muestra los factores analizados. Los diagramas de caja resumen la distribución de estos porcentajes para todas las especies. Área: Área de calibración; RM: multiplicador de regularización; Cor: Colinealidad de predictores; Thin: Filtrado espacial de ocurrencias.

Correlación entre la precisión funcional y discriminatoria

La correlación entre la precisión funcional en el área invasiva y la precisión de discriminación para las 100 especies bajo dos escenarios diferentes de partición de datos mostró resultados similares (Fig. 13). Al considerar la partición de datos aleatoria, se observó una amplia gama de correlaciones. Los valores mínimo y máximo fueron -0.68 y 0.78. La correlación mediana entre estas dos métricas fue 0.15. En total, 45 especies exhibieron correlaciones negativas.

Cuando la partición de datos se basó en bloques geográficos, las correlaciones también mostraron variabilidad, con un rango de -0.66 a 0.68, y una correlación mediana de 0.13. Aquí, 41 especies mostraron correlaciones negativas entre la precisión funcional y la precisión de discriminación.

Para 24 especies, los mejores modelos basados en la evaluación de partición aleatoria de datos mostraron precisiones funcionales superiores a 0.5. De manera paralela, para 26 especies, los mejores modelos seleccionados con la estrategia de partición de datos por bloques geográficos lograron precisiones funcionales superiores a 0.5. Esta observación destaca que, en ciertos casos, la elección de la estrategia de partición de datos permitió la selección de modelos con una precisión funcional razonablemente alta, incluso cuando su precisión de discriminación mostró una alineación limitada con la métrica de precisión funcional.

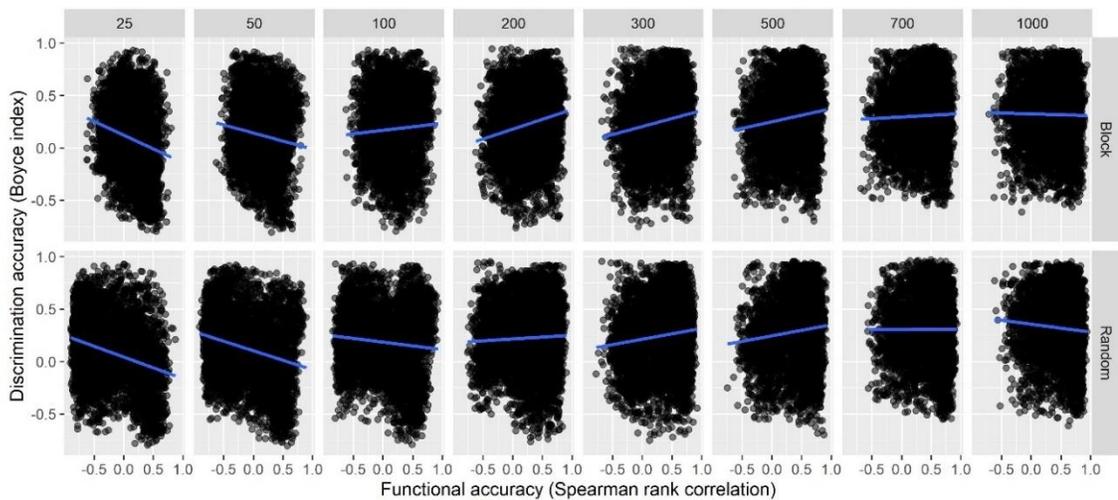


Figura 13. Relación entre la precisión funcional y discriminatoria de los modelos en el área invasiva para las 100 especies bajo diferentes extensiones de área de calibración y métodos de partición. El eje x representa la precisión funcional, dividida por la extensión del área de calibración, mientras que el eje y representa la precisión discriminatoria, dividida por el método de partición (aleatorio y bloques geográficos). Las líneas azules, obtenidas mediante regresión de Mínimos Cuadrados Ordinarios (OLS), proporcionan una representación visual de la correlación entre las dos métricas de precisión para cada subconjunto de puntos.

DISCUSIÓN

En esta investigación se realizó una evaluación más sistemática sobre cómo las diferentes extensiones del área de calibración afectan la precisión de los modelos de distribución de especies (SDM) en el área nativa e invasiva. La variabilidad observada en el rendimiento de los modelos inter e intraespecífica subraya la necesidad de enfoques personalizados, considerando las características particulares de cada especie. En este contexto, el factor más crítico que influye en la precisión de los modelos de distribución de especies resultó ser el tamaño del área de calibración. En general, los modelos calibrados con áreas más grandes tienden a mostrar una mayor precisión funcional que los modelos calibrados con áreas más pequeñas.

Aunque no es sencillo comparar diferentes algoritmos debido a su dependencia de distintos tipos de datos (por ejemplo, solo presencia vs. presencia-ausencia), metodologías estadísticas (por ejemplo, clasificación vs. regresión) o estrategias de evaluación (por ejemplo, ROC-AUC vs. Índice de Boyce), los resultados suelen compararse en la literatura (Bucklin et al. 2015; Valavi et al. 2022) y nuestros hallazgos principales están en línea con investigaciones previas. Por ejemplo, VanDerWal et al. (2009) utilizaron buffers con distancias crecientes que iban de 10 a 500 km alrededor de las ocurrencias de las especies para estudiar el impacto de diversas áreas de calibración trabajando con vertebrados de selvas tropicales australianas. Dichos autores encontraron un aumento rápido en la precisión a medida que el tamaño del área de calibración se expandía de 10 a 100 km (ROC-AUC > 0.93), con expansiones posteriores más allá de este umbral mostrando solo mejoras marginales (ROC-AUC > 0.99). Sin embargo, un inconveniente importante, fue la posible sobreestimación de la precisión del modelo cuando se evaluaba sobre una gran extensión geográfica (Lobo et al. 2008). VanDerWal et al. (2009) reconocieron este fenómeno y usaron un área de evaluación fija para calcular los valores de AUC en todas las especies. Sus hallazgos mostraron que la precisión en el área fija era máxima con un tamaño de fondo alrededor de 200 km, y disminuía gradualmente a medida que se generaban puntos a partir de regiones más grandes. Es importante destacar que el uso de datos simulados con probabilidades de ocurrencia conocidas nos permite evaluar directamente la precisión real de los modelos (precisión funcional) sin depender

únicamente de las métricas de discriminación, asegurando así que nuestros hallazgos no sean artefactos de estas métricas (debido a que la precisión funcional no es afectada por la extensión del área de calibración, a diferencia de la precisión discriminatoria).

De manera similar, Acevedo et al. (2012) realizaron un estudio utilizando datos de cuatro especies de ungulados en España para evaluar la precisión predictiva de los modelos de distribución de especies calibrados con diferentes extensiones de áreas de calibración. Sus resultados mostraron que, mientras que la precisión de calibración (evaluada con el estadístico de Miller) disminuyó con la expansión del área de calibración, la precisión de discriminación (evaluada con el ROC-AUC) aumentó. Este enfoque permitió a los autores del estudio generar modelos puramente ambientales para estimar la distribución potencial de las especies en el nuevo escenario.

Recientemente, Feng (2023), trabajando con 87 especies de colibríes, evaluó el efecto de una serie de buffers creados alrededor de las ocurrencias (desde 5 hasta 5000 km) como áreas de calibración. Los modelos calibrados con buffers espaciales se compararon con modelos calibrados con espacios geográficos considerados como áreas accesibles para las especies. Como resultado, la precisión de discriminación aumentó conforme al tamaño del área de calibración aumentaba, pero alcanzó un umbral de saturación específico para cada especie. Aunque el método de evaluación afectó este aspecto, típicamente se estimó que el buffer alcanzaba el umbral de saturación con una distancia menor de 200 km. De forma destacable, la precisión del modelo calibrado utilizando datos de las áreas accesibles para las especies fue comparable a la precisión de saturación de los modelos cuando se utilizaron buffers espaciales.

En el presente estudio, la comparación entre los modelos no sesgados dentro de las áreas accesibles de cada especie y el resto de los modelos resalta un desafío significativo: determinar el área accesible de una especie. Los modelos no sesgados demostraron consistentemente un rendimiento superior, logrando una alta precisión funcional en todas las especies. En contraste, el resto de los modelos mostraron una precisión más variable, con algunas especies mostrando incluso valores negativos de precisión funcional. Esta discrepancia subraya la importancia de considerar las características específicas de cada especie y seleccionar áreas de calibración apropiadas para asegurar predicciones precisas

del modelo. Los hallazgos sugieren que modelar dentro de las áreas accesibles de las especies puede mitigar sesgos y mejorar la fiabilidad del modelo, destacando los beneficios potenciales de adoptar enfoques no sesgados en los estudios de modelado de distribución de especies.

En contraste con nuestros hallazgos, Lobo y Tognelli (2011) reportaron resultados diferentes. Ellos investigaron los impactos del sesgo espacial de muestreo, así como el número y la ubicación de pseudo-ausencias en la precisión del modelo utilizando especies virtuales. Sus resultados indicaron que el número de pseudo-ausencias y la presencia de sesgo espacial en las localidades de muestreo, junto con su interacción, ejercieron una influencia sustancial en la precisión del modelo (interpretada aquí con ROC-AUC, aunque también evaluaron sensibilidad y especificidad). Como era de esperar, un mayor número de pseudo-ausencias junto con una ausencia de sesgo espacial produjeron modelos superiores. La ubicación de las pseudo-ausencias, ya sea distribuida a lo largo de toda el área de estudio o restringida a regiones fuera de los límites de tolerancia ambiental de la especie, tuvo un efecto relativamente menor en la precisión del modelo. Ellos reconocieron que esto podría atribuirse a la baja área relativa de ocurrencia de la especie (solo el 3.5% del área total de estudio habitada por la especie).

Al contrastar nuestro trabajo con el suyo, se destacan algunas diferencias. Los autores no consideraron explícitamente las capacidades de dispersión de las especies, lo que podría resultar en una sobreestimación del área accesible (Araújo y Pearson 2005). Además, utilizaron mapas binarios (presencia-ausencia definida por umbrales) en lugar de considerar la mayor precisión contenida en la probabilidad de ocurrencia (Leroy et al. 2016; Meynard et al. 2019). También, los modelos fueron calibrados utilizando las mismas variables bioclimáticas que se usaron para crear el nicho de las especies virtuales (una condición ideal, pero poco realista de los escenarios encontrados en el estudio de especies reales). Usaron un umbral nuevamente para la evaluación del modelo, restringiendo el uso de datos relativos a la probabilidad real de ocurrencia. Finalmente, simulaban una sola especie virtual, mientras que el presente estudio abarcó los resultados de 100 especies virtuales (en este trabajo, la contribución relativa menor para el área de calibración fue del 2.5% y la mayor alcanzó el 93%). Estas distinciones destacan las

complejidades involucradas en la modelización de distribuciones de especies y la importancia de considerar múltiples factores para mejorar la robustez y relevancia ecológica de dichos modelos.

Aumentar la extensión del área de calibración implica incorporar datos que están, en promedio, ambientalmente más distantes de registros de ocurrencia. En consecuencia, la precisión de la discriminación del modelo puede aumentar debido a la facilidad para parametrizar modelos con una buena capacidad de discriminación pero que contienen poca información útil (Barve et al. 2011; Acevedo et al. 2012). Esto podría ser el resultado de áreas de calibración más grandes que cubren lugares con condiciones ambientales apropiadas que están desocupadas debido a interacciones bióticas y/o restricciones de dispersión, lo que podría inducir a un sobreajuste a condiciones cercanas a las localidades ocupadas (Anderson y Raza 2010). Por otro lado, la importancia de factores a gran escala, como el clima, podría ser subestimada en áreas de calibración pequeñas (Barve et al. 2011; Acevedo et al. 2012).

Nuestros resultados coinciden con investigaciones anteriores que destacan el impacto del área de calibración en la precisión del modelo y muestran la complejidad involucrada en su selección (VanDerWal et al. 2009). El supuesto del equilibrio ambiental, según el cual la especie está adaptada a sus condiciones ambientales actuales, subraya la importancia de elegir un área de calibración que refleje con precisión estas condiciones (Araújo y Pearson 2005). No obstante, nuestros hallazgos, junto con investigaciones anteriores, indican que no hay un consenso sobre la delimitación de un área de calibración ideal (Rotllan-Puig y Traveset 2021; Machado-Stredel et al. 2021).

Dado que algunos métodos correlativos para estimar nichos ecológicos se basan en contrastar las características ambientales de los sitios de ocurrencia conocidos con las condiciones disponibles en el área de estudio, se vuelve imperativo delinear y comprender el rango potencial que la especie podría haber explorado. Esto es crucial ya que la ausencia de la especie fuera de esta área accesible puede atribuirse a factores no relacionados con influencias abióticas o bióticas. Por ejemplo, una especie puede estar ausente de regiones adecuadas únicamente debido a su incapacidad para alcanzar esos sitios (Anderson y Raza 2010; Barve et al. 2011).

La forma en que los SDM manejan la colinealidad entre las variables predictoras (Feng et al. 2019), el sesgo de muestreo (Ranc et al. 2017; Inman et al. 2021) y la complejidad del modelo son aspectos adicionales que podrían afectar la precisión del modelo (Merow et al. 2014). Sin embargo, cuando se consideran de manera multifactorial, nuestros resultados demuestran que estos factores tienen un impacto significativamente menor en comparación con la selección del área de calibración. En este sentido, Barbet-Massin et al. (2012) observaron que el impacto de diferentes decisiones metodológicas en la calidad del modelo es dependiente del método o algoritmo de modelado empleado. Para las técnicas de aprendizaje automático (por ejemplo, boosted regression trees y random forest), el número de pseudoausencias explicaba una mayor cantidad de desviación (entre el 20 y el 85%) en comparación con el método para seleccionar y ponderar pseudoausencias (menos del 15%). Por lo tanto, explorar una variedad de técnicas de modelado además de MaxEnt podría ayudar a comprender mejor sus respuestas diferenciales y sus implicaciones para las técnicas de SDM.

En cuanto a la correlación entre la precisión funcional y la precisión de discriminación, nuestros resultados muestran que esta correlación es relativamente baja. La precisión de discriminación basada en ambos esquemas de partición de datos resultó ser una medida poco fiable de la precisión funcional; sin embargo, contrariamente a nuestras expectativas, los modelos seleccionados con partición aleatoria demostraron una mejor correlación entre la precisión funcional y la de discriminación, aunque se esperaba una menor correlación en este escenario debido al menor grado de independencia entre los datos utilizados para la evaluación y la calibración. Estos hallazgos sugieren que la partición aleatoria de los datos podría ser más efectiva para seleccionar modelos que con una mejor predicción de la precisión de los valores de idoneidad, a pesar de las ventajas teóricas de la partición por bloques en la creación de conjuntos de evaluación geográficamente independientes. Una posible razón es que ciertos rangos de valores ambientales no se utilizan durante la calibración en el particionamiento por bloques y pueden estar geográficamente agrupados, a diferencia del particionamiento aleatorio, que evita esta estratificación (Kass et al. 2021). Esto contrasta con los hallazgos de Warren et al. (2020), quienes observaron una mejor correlación utilizando el particionamiento por bloques geográficos.

En el caso de la correlación entre la precisión funcional en el área invasiva y la precisión de discriminación (evaluada en el área nativa), los resultados revelan un comportamiento diferente. Específicamente, una mayor cantidad de especies mostraron correlaciones negativas, oscilando entre 19-25 especies en el área nativa, y 41-45 especies en el área invasiva. Sin embargo, cabe destacar que, en este aspecto, la partición por bloques resultó en un menor número de especies con correlaciones negativas, lo que sugiere que este método es más adecuado para evaluar la transferibilidad del modelo. Esto parece estar relacionado con la estructuración geográfica de los puntos de validación, que permite capturar mejor las condiciones ambientales que se encuentran fuera del área de calibración original. Este resultado resalta que, al estructurar geográficamente los puntos de validación, se obtiene una representación más precisa del proceso de extrapolación del modelo hacia nuevas áreas, donde las condiciones ambientales suelen diferir considerablemente de las observadas en el área de calibración (Low et al. 2020; Liu et al. 2022a).

La selección inadecuada del área de calibración tiene implicaciones significativas para las aplicaciones de los modelos de distribución de especies (Barve et al. 2011). Las consecuencias se extienden a aspectos críticos como la estimación inexacta de la extensión de ocurrencia y el área de ocupación, lo que puede llevar a establecer prioridades de conservación erróneas (VanDerWal et al. 2009), la incapacidad para generar hipótesis mecanísticas apropiadas sobre los parámetros que gobiernan las distribuciones de especies (VanDerWal et al. 2009), y la determinación de curvas de respuesta sesgadas (Thuiller et al. 2004). Por lo tanto, es esencial considerar cuidadosamente y elegir el área de calibración para garantizar la robustez en la aplicación de los modelos de distribución de especies.

Construir buffers espaciales alrededor de las ocurrencias conocidas, reflejando el rango espacial potencial que una especie podría explorar, ofrece un método sencillo para delinear un área de calibración (Feng 2023). Este enfoque se alinea más estrechamente con las consideraciones teóricas sobre la movilidad de las especies (Holloway y Miller 2017), proporcionando una base más realista para los ejercicios de modelado de distribución de especies. En línea con consideraciones previas (VanDerWal et al. 2009;

Barve et al. 2011; Feng 2023), sería recomendable que los ejercicios de modelado de distribución de especies comiencen con análisis exploratorios del área de calibración, evaluando la extensión que puede ofrecer tanto los resultados más precisos como un ajuste biológica y ecológicamente significativo entre la ocurrencia de la especie y las variables predictoras.

En resumen, esta investigación avanza en la discusión sobre las áreas de calibración, destacando sus impactos matizados en la precisión de los modelos de distribución de especies (SDM). Reconociendo la complejidad de estas consideraciones, nuestros hallazgos contribuyen al perfeccionamiento continuo de las prácticas de SDM, enfatizando la necesidad de enfoques personalizados en diferentes contextos ecológicos. Este estudio ha demostrado que el área de calibración es uno de los factores más críticos que afecta la precisión funcional de los modelos de distribución de especies utilizando MaxEnt. En contraste, factores como el valor del parámetro de regularización y la colinealidad entre las variables predictoras tienen un impacto mucho menor.

CONCLUSIONES

Este trabajo contribuye al campo del modelado de nicho ecológico al proporcionar un análisis detallado de cómo la selección del área de calibración y otros factores influyen en la precisión de los modelos de distribución de especies exóticas invasoras. Se resalta la importancia de personalizar los enfoques para cada especie, destacando que una calibración adecuada puede mejorar sustancialmente la capacidad predictiva de estos modelos.

Al utilizar especies virtuales con respuestas ambientales y probabilidades de ocurrencia conocidas, este estudio proporciona información sobre cómo diferentes factores influyen en la precisión de los SDM y destaca la importancia de una cuidadosa selección del área de calibración y otros aspectos del proceso de modelado como el filtrado espacial de los registros de ocurrencia, la colinearidad de las variables predictoras y el valor de regularización utilizando el algoritmo MaxEnt, subrayando su relevancia para aplicaciones prácticas en la gestión de especies invasoras. Esto permite a los gestores de biodiversidad y conservacionistas aplicar modelos más robustos en la predicción de invasiones biológicas bajo escenarios complejos.

Finalmente, los hallazgos refuerzan la necesidad de seguir investigando cómo optimizar los ENM para escenarios específicos, como la extrapolación de condiciones ambientales y la transferibilidad de los modelos, contribuyendo a la predicción más efectiva de la distribución de especies invasoras en nuevas áreas.

PERSPECTIVAS

Una de las principales áreas de investigación futura podría enfocarse en mejorar la capacidad de transferir los modelos de nicho ecológico a nuevas áreas geográficas o bajo condiciones ambientales cambiantes, como el cambio climático. La extrapolación de los modelos a nuevas regiones es un reto significativo y se necesitan métodos más precisos para abordar esta incertidumbre.

Aunque este estudio se ha centrado en factores abióticos y de dispersión, las interacciones bióticas (competencia, depredación) también juegan un papel importante en la distribución de las especies invasoras. Incorporar estos factores en futuros modelos ofrecería una visión más completa y realista de las áreas de invasión potencial.

Otra perspectiva importante radica en la necesidad de expandir la resolución y la extensión geográfica de los datos utilizados en las simulaciones. Actualmente, el uso de datos a resoluciones limitadas puede generar modelos menos precisos y generalizados. Al emplear resoluciones espaciales más finas y datos que cubran una mayor extensión geográfica, especialmente en regiones submuestreadas, se podrían obtener predicciones más detalladas y representativas, mejorando la capacidad de los modelos para detectar patrones de distribución de especies invasoras.

En el aspecto del modelado, resulta clave explorar el uso de diferentes algoritmos para comparar su capacidad predictiva. Si bien MaxEnt ha mostrado ser eficiente, incluir otros algoritmos como Random Forest, Boosted Regression Trees o redes neuronales artificiales podría ofrecer una visión más amplia sobre qué técnicas son más adecuadas según el tipo de datos y las especies modeladas. Esta diversidad de enfoques podría mejorar la robustez de las predicciones al permitir comparaciones entre resultados y seleccionar el mejor enfoque para cada contexto.

Por último, en cuanto a la selección de los modelos, es necesario incorporar métricas de precisión de calibración, como el estadístico de Miller, que complementen las evaluaciones tradicionales de capacidad predictiva y actualmente son relativamente poco utilizadas. Estas métricas permitirían evaluar no solo qué modelos predicen mejor la distribución de las especies, sino también cuáles son más ajustados y eficientes desde una

perspectiva estadística, mejorando la selección de modelos y evitando sobreajustes o predicciones espurias.

BIBLIOGRAFÍA

- Acevedo P, Jiménez-Valverde A, Lobo JM, Real R. 2012. Delimiting the geographical background in species distribution modelling. *Journal of Biogeography* 39:1383–1390. <https://doi.org/10.1111/j.1365-2699.2012.02713.x>
- Aiello-Lammens ME, Boria RA, Radosavljevic A, Vilela B, Anderson RP. 2015. spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* 38:541–545. <https://doi.org/10.1111/ecog.01132>
- Anderson RP, Raza A. 2010. The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus *Nephelomys*) in Venezuela: effect of study region on models of distributions. *Journal of Biogeography* 37:1378–1393. <https://doi.org/10.1111/j.1365-2699.2010.02290.x>
- Araújo MB, Pearson RG. 2005. Equilibrium of species' distributions with climate. *Ecography* 28:693–695. <https://doi.org/10.1111/j.2005.0906-7590.04253.x>
- Aravind NA, Shaanker MU, Bhat HN, Charles B, Shaanker RU, Shah MA, Ravinkanth G. 2022. Niche shift in invasive species: is it a case of “home away from home” or finding a “new home”? *Biodiversity and Conservation*, 31(11), 2625-2638. <https://doi.org/10.1007/s10531-022-02447-0>
- Austin MP. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological modelling* 157(2-3):101-118. [https://doi.org/10.1016/S0304-3800\(02\)00205-3](https://doi.org/10.1016/S0304-3800(02)00205-3)
- Barbet-Massin M, Jiguet F, Albert CH, Thuiller W. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution* 3:327–338. <https://doi.org/10.1111/j.2041-210X.2011.00172.x>
- Barve N, Barve V, Jiménez-Valverde A, Lira-Noriega A, Maher SP, Peterson AT, Soberón J, Villalobos F. 2011. The crucial role of the accessible area in ecological

- niche modeling and species distribution modeling. *Ecological Modelling* 222:1810–1819. <https://doi.org/10.1016/j.ecolmodel.2011.02.011>
- Bates OK, Bertelsmeier C. 2021. Climatic niche shifts in introduced species. *Current Biology*, 31(19):R1252-R1266. <https://doi.org/10.1016/j.cub.2021.08.035>
- Beck J, Böller M, Erhardt A, Schwanghart W. 2014. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics* 19:10-15. <https://doi.org/10.1016/j.ecoinf.2013.11.002>
- Blackburn TM, Pyšek P, Bacher S, Carlton JT, Duncan RP, Jarošík V, John RU, Richardson DM. 2011. A proposed unified framework for biological invasions. *Trends in Ecology & Evolution* 26(7): 333-339. <https://doi.org/10.1016/j.tree.2011.03.023>
- Bomford M. 2008. Risk assessment models for establishment of exotic vertebrates in Australia and New Zealand. Canberra: Invasive Animals Cooperative Research Centre.
- Booth TH. 2022. Checking bioclimatic variables that combine temperature and precipitation data before their use in species distribution models. *Austral Ecology* 47:1506–1514. <https://doi.org/10.1111/aec.13234>
- Booth TH, Nix HA, Busby JR, Hutchinson MF. 2014. BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies. *Diversity and Distributions* 20:1-9. <https://doi.org/10.1111/ddi.12144>
- Boria RA, Olson LE, Goodman SM, Anderson RP. 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling* 275:73-77. <https://doi.org/10.1016/j.ecolmodel.2013.12.012>
- Bosso L, De Conno C, Russo D. 2017. Modelling the risk posed by the zebra mussel *Dreissena polymorpha*: Italy as a case study. *Environmental Management* 60(2):304-313. <https://doi.org/10.1007/s00267-017-0882-8>

- Briski E, Chan FT, Darling JA, Lauringson V, MacIsaac HJ, Zhan A, Bailey SA. 2018. Beyond propagule pressure: importance of selection during the transport stage of biological invasions. *Frontiers in Ecology and the Environment* 16(6):345-353. <https://doi.org/10.1002/fee.1820>
- Brodie S, Smith JA, Muhling BA, Barnett LAK, Carroll G, Fiedler P, Bograd SJ, Hazen EL, Jacox MG, Andrews KS, Barnes CL, Crozier LG, Fiechter J, Fredston A, Haltuch MA, Harvey CJ, Holmes E, Karp MA, Liu OR, Malick MJ, Pozo Buil M, Richerson K, Rooper CN, Samhuri J, Seary R, Selden RL, Thompson AR, Tommasi D, Ward EJ, Kaplan IC. 2022. Recommendations for quantifying and reducing uncertainty in climate projections of species distributions. *Global Change Biology* 28:6586–6601. <https://doi.org/10.1111/gcb.16371>
- Broennimann O, Treier UA, Müller-Schärer H, Thuiller W, Peterson AT, Guisan A. 2007. Evidence of climatic niche shift during biological invasion. *Ecology Letters* 10(8):701-709. <https://doi.org/10.1111/j.1461-0248.2007.01060.x>
- Brown P, Daigneault A. 2014. Cost–benefit analysis of managing the invasive African tulip tree (*Spathodea campanulata*) in the Pacific. *Environmental Science & Policy* 39:65-76. <https://doi.org/10.1016/j.envsci.2014.02.004>
- Bucklin DN, Basille M, Benscoter AM, Brandt LA, Mazzotti FJ, Romañach SS, Speroterra C, Watling JI. 2015. Comparing species distribution models constructed with different subsets of environmental predictors. *Diversity and distributions* 21(1):23-35. <https://doi.org/10.1111/ddi.12247>
- Buckley LB. 2008. Linking traits to energetics and population dynamics to predict lizard ranges in changing environments. *American Naturalist* 171:E1–E19. <https://doi.org/10.1086/523949>
- Castillo-Torres PA, Martínez-Meyer E, Córdova-Tapia F, Zambrano L. 2017. Potential distribution of native freshwater fish in Tabasco, Mexico. *Revista Mexicana de Biodiversidad* 88:415-424. <https://doi.org/10.1016/j.rmb.2017.03.001>

- Clark JD, Dunn JE, Smith KG. 1993. A multivariate model of female black bear habitat use for a Geographic Information System. *The Journal of Wildlife Management* 57(3):519-526. <https://doi.org/10.2307/3809276>
- Cobos ME, Peterson AT, Barve N, Osorio-Olvera L. 2019. *kuenm*: an R package for detailed development of ecological niche models using Maxent. *PeerJ* 7:e6281. <https://doi.org/10.7717/peerj.6281>
- Copp GH, Vilizzi L, Tidbury H, Stebbing PD, Tarkan AS, Miossec L, Gouilletquer P. 2016. Development of a generic decision-support tool for identifying potentially invasive aquatic taxa: AS-ISK. *Management of Biological Invasions* 7(4):343-350. <https://doi.org/10.3391/mbi.2016.7.4.04>
- D'hondt B, Vanderhoeven S, Roelandt S, Mayer F, Versteirt V, Adriaens T, Ducheyne E, San Martin G, Grégoire J-C, Stiers I, Quoilin S, Cigar J, Heughebaert A, Branquart E. 2015. *Harmonia*⁺ and *Pandora*⁺: risk screening tools for potentially invasive plants, animals and their pathogens. *Biological Invasions* 17(6):1869-1883. <https://doi.org/10.1007/s10530-015-0843-1>
- De Marco P Jr, Nóbrega CC. 2018. Evaluating collinearity effects on species distribution models: an approach based on virtual species simulation. *PLoS ONE* 13(9):e0202403. <https://doi.org/10.1371/journal.pone.0202403>
- de Souza-Muñoz ME, De Giovanni R, de Siqueira MF, Sutton T, Brewer P, Pereira RS, Lange Canhos DA, Perez Canhos V. 2011. openModeller: a generic approach to species' potential distribution modelling. *GeoInformatica* 15(1):111-135. <https://doi.org/10.1007/s10707-009-0090-7>
- Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, García Márquez JR, Gruber B, Lafourcade B, Leitão PJ, Münkemüller T, McClean C, Osborne PE, Reineking B, Schröder B, Skidmore AK, Zurell D, Lautenbach S. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36:27-46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>

- Duque-Lazo J, van Gils H, Groen TA, Navarro-Cerrillo RM. 2016. Transferability of species distribution models: the case of *Phytophthora cinnamomi* in Southwest Spain and Southwest Australia. *Ecological Modelling* 320:62–70. <https://doi.org/10.1016/j.ecolmodel.2015.09.019>
- Elith J, Leathwick JR. 2007. Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Diversity and Distributions* 13:265–275. <https://doi.org/10.1111/j.1472-4642.2007.00340.x>
- Elton CS. 1927. *Animal ecology*. Sidgwick and Jackson, London.
- Espindola S, Parra JL, Vázquez-Domínguez E. 2019. Fundamental niche unfilling and potential invasion risk of the slider turtle *Trachemys scripta*. *PeerJ* 7:e7923. <https://doi.org/10.7717/peerj.7923>
- Feng X. 2023. A test of species' mobility hypothesis in ecological niche modelling. *Journal of Biogeography* 50:1955–1966. <https://doi.org/10.1111/jbi.14704>
- Feng X, Park DS, Liang Y, Pandey R, Papeş M. 2019. Collinearity in ecological niche modeling: confusions and challenges. *Ecology and Evolution* 9:10365–10376. <https://doi.org/10.1002/ece3.5555>
- Fernández M, Hamilton H. 2015. Ecological niche transferability using invasive species as a case study. *PloS one* 10(3):e0119891. <https://doi.org/10.1371/journal.pone.0119891>
- Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37:4302–4315. <https://doi.org/10.1002/joc.5086>
- Fithian W, Hastie T. 2013. Finite-sample equivalence of several statistical models for presence-only data. *The Annals of Applied Statistics* 7(4):1917–1939. <https://doi.org/10.1214/13-AOAS667>

- Gallardo B, Aldridge DC. 2013a. Priority setting for invasive species management: risk assessment of Ponto-Caspian invasive species into Great Britain. *Ecological applications* 23(2):352-364. <https://doi.org/10.1890/12-1018.1>
- Gallardo B, Aldridge DC. 2013b. The ‘dirty dozen’: socio-economic factors amplify the invasion potential of 12 high-risk aquatic invasive species in Great Britain and Ireland. *Journal of Applied Ecology* 50(3):757-766. <https://doi.org/10.1111/1365-2664.12079>
- Gallardo B, Aldridge DC. 2015. Is Great Britain heading for a Ponto–Caspian invasional meltdown?. *Journal of Applied Ecology* 52(1):41-49. <http://doi.org/10.1111/1365-2664.12348>
- Gallien L, Münkemüller T, Albert CH, Boulangéat I, Thuiller W. 2010. Predicting potential distributions of invasive species: where to go from here?. *Diversity and Distributions* 16(3):331-342. <https://doi.org/10.1111/j.1472-4642.2010.00652.x>
- Gallien L, Douzet R, Pratte S, Zimmermann NE, Thuiller W. 2012. Invasive species distribution models—how violating the equilibrium assumption can create new insights. *Global Ecology and Biogeography* 21(11):1126-1136. <https://doi.org/10.1111/j.1466-8238.2012.00768.x>
- Grinnell J. 1917. The niche-relationships of the California Thrasher. *The Auk* 34(4):427–433. <https://doi.org/10.2307/4072271>
- Grömping U. 2006. Relative importance for linear regression in R: the package relaimpo. *Journal of Statistical Software* 17(1):1-27. <https://doi.org/10.18637/jss.v017.i01>
- Guisan A, Zimmermann E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135:147-86. [https://doi.org/10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9)
- Guisan A, Thuiller W, Zimmermann NE. 2017. *Habitat suitability and distribution models: with applications in R*. Cambridge University Press.
- Hardisty A, Roberts D. 2013. A decadal view of biodiversity informatics: challenges and priorities. *BMC Ecology* 13(1):16. <https://doi.org/10.1186/1472-6785-13-16>

- Hartmann J, Moosdorf N. 2012. The new global lithological map database GLiM: a representation of rock properties at the Earth Surface. *Geochemistry, Geophysics, Geosystems* 13(12): Q12004. <https://doi.org/10.1029/2012GC004370>
- Heikkinen RK, Marmion M, Luoto M. 2012. Does the interpolation accuracy of species distribution models come at the expense of transferability?. *Ecography* 35(3):276-288. <https://doi.org/10.1111/j.1600-0587.2011.06999.x>
- Hijmans RJ, Phillips S, Leathwick J, Elith J. 2017. Package ‘dismo’. *Circles*. 9(1):1-68.
- Hirzel AH, Hausser J, Chessel D, Perrin N. 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology* 83(7):2027-2036. [https://doi.org/10.1890/0012-9658\(2002\)083\[2027:ENFAHT\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2002)083[2027:ENFAHT]2.0.CO;2)
- Hirzel AH, Le Lay G, Helfer V, Randin C, Guisan A. 2006. Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling* 199:142–152. <https://doi.org/10.1016/j.ecolmodel.2006.05.017>
- Hirzel AH, Le Lay G. 2008. Habitat suitability modelling and niche theory. *Journal of Applied Ecology* 45:1372–1381. <https://doi.org/10.1111/j.1365-2664.2008.01524.x>
- Holloway P, Miller JA. 2017. A quantitative synthesis of the movement concepts used within species distribution modelling. *Ecological Modelling* 356:91–103. <https://doi.org/10.1016/j.ecolmodel.2017.04.005>
- Huang E, Chen Y, Fang M, Zheng Y, Yu S. 2021. Environmental drivers of plant distributions at global and regional scales. *Global Ecology and Biogeography* 30:697–709. <https://doi.org/10.1111/geb.13251>
- Hurlbert SH, Levine RA, Utts J. 2019. Coup de grâce for a tough old bull: “statistically significant” expires. *The American Statistician* 73:352–357. <https://doi.org/10.1080/00031305.2018.1543616>
- Hutchinson GE. 1957. Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology* 22:415–427.

- Inman R, Franklin J, Esque T, Nussear K. 2021. Comparing sample bias correction methods for Species Distribution Modeling using virtual species. *Ecosphere* 12(3): e03422. <https://doi.org/10.1002/ecs2.3422>
- Jiménez L, Soberón J. 2022. Estimating the fundamental niche: accounting for the uneven availability of existing climates in the calibration area. *Ecological Modelling* 464:109823. <https://doi.org/10.1016/j.ecolmodel.2021.109823>
- Jackson ST, Overpeck JT. 2000. Responses of plant populations and communities to environmental changes of the late Quaternary. *Paleobiology* 26(Suppl):194-220. <https://doi.org/10.1017/S0094837300026932>
- Kass JM, Muscarella R, Galante PJ, Bohl CL, Pinilla-Buitrago GE, Boria RA, Soley-Guardia M, Anderson RP. 2021. ENMeval 2.0: redesigned for customizable and reproducible modeling of species' niches and distributions. *Methods in Ecology and Evolution* 12:1602–1608. <https://doi.org/10.1111/2041-210X.13628>
- Leprieur F, Beauchard O, Blanchet S, Oberdorff T, Brosse S. 2008. Fish invasions in the world's river systems: when natural processes are blurred by human activities. *PLoS Biology* 6(2):0404-0410. <https://doi.org/10.1371/journal.pbio.0060028>
- Leroy B, Meynard CN, Bellard C, Courchamp F. 2016. virtualspecies, an R package to generate virtual species distributions. *Ecography* 39:599–607. <https://doi.org/10.1111/ecog.01388>
- Liu C, Wolter C, Xian W, Jeschke JM. 2020a. Species distribution models have limited spatial transferability for invasive species. *Ecology letters* 23(11):1682-1692. <https://doi.org/10.1111/ele.13577>
- Liu C, Wolter C, Xian W, Jeschke JM. 2020b. Most invasive species largely conserve their climatic niche. *Proceedings of the National Academy of Sciences* 117(38): 23643-23651. <https://doi.org/10.1073/pnas.2004289117>
- Liu C, Wolter C, Courchamp F, Roura-Pascual N, Jeschke JM. 2022. Biological invasions reveal how niche change affects the transferability of species distribution models. *Ecology* 103(8):e3719. <https://doi.org/10.1002/ecy.3719>.

- Leung B, Roura-Pascual N, Bacher S, Heikkilä J, Brotons L, Burgman MA, Dehnen-Schmutz K, Essl F, Hulme PE, Richardson DM, Sol D, Vilà M. 2012. TEASIng apart alien species risk assessments: a framework for best practices. *Ecology Letters* 15(12):1475-1493. <https://doi.org/10.1111/ele.12003>
- Lo Parrino E, Falaschi M, Manenti R, Ficetola GF. 2023. All that changes is not shift: methodological choices influence niche shift detection in freshwater invasive species. *Ecography* e06432. <https://doi.org/10.1111/ecog.06432>
- Lobo JM, Jiménez-Valverde A, Real R. 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* 17:145–151. <https://doi.org/10.1111/j.1466-8238.2007.00358.x>
- Lobo JM, Tognelli MF. 2011. Exploring the effects of quantity and location of pseudo-absences and sampling biases on the performance of distribution models with limited point occurrence data. *Journal for Nature Conservation* 19:1–7. <https://doi.org/10.1016/j.jnc.2010.03.002>
- Low, BW, Zeng Y, Tan HH, Yeo DCJ. Predictor complexity and feature selection affect maxent model transferability: evidence from global freshwater invasive species. *Diversity and Distributions* 27(3):497-511. <https://doi.org/10.1111/ddi.13211>
- Machado-Stredel F, Cobos ME, Peterson AT. 2021. A simulation-based method for selecting calibration areas for ecological niche models and species distribution models. *Frontiers in Biogeography* 13(4):e48814. <https://doi.org/10.21425/F5FBG48814>
- Masin S, Bonardi A, Padoa-Schioppa E, Bottoni L, Ficetola GF. 2014. Risk of invasion by frequently traded freshwater turtles. *Biological Invasions* 16:217-231. <https://doi.org/10.1007/s10530-013-0515-y>
- Merow C, Smith MJ, Silander JA. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* 36:1058–1069. <https://doi.org/10.1111/j.1600-0587.2013.07872.x>

- Merow C, Smith MJ, Edwards TC, Guisan A, McMahon SM, Normand S, Thuiller W, Wüest RO, Zimmermann NE, Elith J. 2014. What do we gain from simplicity versus complexity in Species Distribution Models? *Ecography* 37(12):1267-81. <https://doi.org/10.1111/ecog.00845>
- Meynard CN, Leroy B, Kaplan DM. 2019. Testing methods in species distribution modelling using virtual species: what have we learnt and what are we missing? *Ecography* 42:2021–2036. <https://doi.org/10.1111/ecog.04385>
- Montecino V, Molina X, Kumar S, Castillo ML, Bustamante RO. 2014. Niche dynamics and potential geographic distribution of *Didymosphenia geminata* (Lyngbye) M. Schmidt, an invasive freshwater diatom in Southern Chile. *Aquatic Invasions* 9(4): 507–519. <https://doi.org/10.3391/ai.2014.9.4.09>
- Morehouse RL, Papeş M, Tobler M. 2013. Predicting and mapping the potential distribution of the Painted devil crayfish, *Cambarus ludovicianus* Faxon (Decapoda: Cambaridae). *The Southwestern Naturalist* 58(4):435-439. <https://doi.org/10.1894/0038-4909-58.4.435>
- Mormul RP, Vieira DS, Bailly D, Fidanza K, da Silva VFB, da Graça WJ, Pontara V, Bueno ML, Thomaz SM, Mendes RS. 2022. Invasive alien species records are exponentially rising across the Earth. *Biological Invasions* 24(10):3249-3261. <https://doi.org/10.1007/s10530-022-02843-1>
- Muscarella R, Galante PJ, Soley-Guardia M, Boria RA, Kass JM, Uriarte M, Anderson RP. 2014. *ENMeval*: an R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution* 5(11):1198-1205. <https://doi.org/10.1111/2041-210X.12261>
- Myers JH, Simberloff D, Kuris AM, Carey JR. 2000. Eradication revisited: dealing with exotic species. *Trends in Ecology and Evolution* 15(8):316-320. [https://doi.org/10.1016/S0169-5347\(00\)01914-5](https://doi.org/10.1016/S0169-5347(00)01914-5)

- Naimi B, Hamm NAS, Groen TA, Skidmore AK, Toxopeus AG. 2014. Where is positional uncertainty a problem for species distribution modelling? *Ecography* 37:191–203. <https://doi.org/10.1111/j.1600-0587.2013.00205.x>
- Nguyen D, Leung B. 2022. How well do species distribution models predict occurrences in exotic ranges?. *Global Ecology and Biogeography* 31(6):1051-1065. <https://doi.org/10.1111/geb.13482>
- Owens HL, Campbell LP, Dornak LL, Saupe EE, Barve N, Soberón J, Ingenloff K, Lira-Noriega A, Hensz CM, Myers CE, Peterson AT. 2013. Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling* 263:10–18. <https://doi.org/10.1016/j.ecolmodel.2013.04.011>
- Papeş M, Havel JE, Vander Zanden MJ. 2016. Using maximum entropy to predict the potential distribution of an invasive freshwater snail. *Freshwater Biology* 61(4):457-471. <https://doi.org/10.1111/fwb.12719>
- Peterson AT, Nakazawa Y. 2008. Environmental data sets matter in ecological niche modelling: an example with *Solenopsis invicta* and *Solenopsis richteri*. *Global Ecology and Biogeography* 17(1):135-144. <https://doi.org/10.1111/j.1466-8238.2007.00347.x>
- Peterson AT, Papeş M, Soberón J. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling* 213(1):63-72. <https://doi.org/10.1016/j.ecolmodel.2007.11.008>
- Peterson AT, Soberón J, Pearson RG, Anderson RP, Martínez-Meyer E, Nakamura M, Araújo MB. 2011. *Ecological niches and geographic distributions*. Princeton University Press.
- Peterson AT, Soberón J. 2012. Species distribution modeling and ecological niche modeling: getting the concepts right. *Natureza & Conservação* 10:102–107. <http://dx.doi.org/10.4322/natcon.2012.019>
- Petitpierre B, Broennimann O, Kueffer C, Daehler C, Guisan A. 2017. Selecting predictors to maximize the transferability of species distribution models: lessons from cross-

- continental plant invasions. *Global Ecology and Biogeography* 26(3):275-287. <https://doi.org/10.1111/geb.12530>
- Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190:231–259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>
- Pimentel D, Zuniga R, Morrison D. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics* 52(3): 273-288. <https://doi.org/10.1016/j.ecolecon.2004.10.002>
- Pollock LJ, Tingley R, Morris WK, Golding N, O'Hara RB, Parris KM, Vesk PA, McCarthy MA. 2014. Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution* 5(5):397-406. <https://doi.org/10.1111/2041-210X.12180>
- Porter WP, Sabo JL, Tracy CR, Reichman OJ, Ramankutty N. 2002. Physiology on a landscape scale: plant-animal interactions. *Integrative and Comparative Biology* 42:431–453. <https://doi.org/10.1093/icb/42.3.431>
- Poulos HM, Chernoff B, Fuller PL, Butman D. 2012. Ensemble forecasting of potential habitat for three invasive fishes. *Aquatic Invasions* 7 (1): 59–72.
- Pulliam HR. 2000. On the relationship between niche and distribution. *Ecology letters* 3(4):349-361. <https://doi.org/10.1046/j.1461-0248.2000.00143.x>
- Qiao H, Escobar LE, Peterson AT. 2017. Accessible areas in ecological niche comparisons of invasive species: recognized but still overlooked. *Scientific Reports* 7(1):1213. <https://doi.org/10.1038/s41598-017-01313-2>
- Rahel FJ. 2002. Homogenization of freshwater faunas. *Annual Review of Ecology and Systematics* 33:291-315. <https://doi.org/10.1146/annurev.ecolsys.33.010802.150429>
- Ranc N, Santini L, Rondinini C, Boitani L, Poitevin F, Angerbjörn A, Maiorano L. 2017. Performance tradeoffs in target-group bias correction for species distribution models. *Ecography* 40(9):1076-87. <https://doi.org/10.1111/ecog.02414>

- Reshetnikov AN, Ficetola GF. 2011. Potential range of the invasive fish rotan (*Percocottus glenii*) in the Holarctic. *Biological Invasions* 13:2967-2980. <https://doi.org/10.1007/s10530-011-9982-1>
- Rico-Sánchez AE, Haubrock PJ, Cuthbert RN, Angulo E, Ballesteros-Mejia L, López-López E, Duboscq VG, Nuñez MA, Diagne C, Courchamp F. 2021. Economic costs of invasive alien species in Mexico. *NeoBiota* 67:459-483. <https://doi.org/10.3897/neobiota.67.63846>
- Rodríguez-Merino A, Fernández-Zamudio R, García-Murillo P. 2015. Determinación de las zonas con mayor riesgo de invasión por macrófitos acuáticos exóticos en La Península Ibérica. *Teledetección: Humedales Y Espacios Protegidos* 1:338-41.
- Rotllan-Puig X, Traveset A. 2021. Determining the minimal background area for species distribution models: MinBAR package. *Ecological Modelling* 439:109353. <https://doi.org/10.1016/j.ecolmodel.2020.109353>
- Shcheglovitova M, Anderson RP. 2013. Estimating optimal complexity for ecological niche models: a jackknife approach for species with small sample sizes. *Ecological Modelling* 269:9-17. <https://doi.org/10.1016/j.ecolmodel.2013.08.011>
- Sillero N, Barbosa AM. 2021. Common mistakes in ecological niche models. *International Journal of Geographical Information Science* 35(2):213-226. <https://doi.org/10.1080/13658816.2020.1798968>
- Sillero N, Arenas-Castro S, Enriquez-Urzelai U, Gomes Vale C, Sousa-Guedes D, Martínez-Freiría F, Real R, Márcia Barbosa A. 2021. Want to model a species niche? A step-by-step guideline on correlative ecological niche modelling. *Ecological Modelling* 456:109671. <https://doi.org/10.1016/j.ecolmodel.2021.109671>
- Soberon J, Peterson AT. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics* 2:1-10.
- Soberón J, Nakamura M. 2009. Niches and distributional areas: concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences* 106(2):19644-19650. <https://doi.org/10.1073/pnas.0901637106>

- Soberón J, Arroyo-Peña B. 2017. Are fundamental niches larger than the realized? Testing a 50-year-old prediction by Hutchinson. *PLoS ONE* 12(4):e0175138. <https://doi.org/10.1371/journal.pone.0175138>
- Srinivasulu A, Srinivasulu B, Srinivasulu C. 2021. Ecological niche modelling for the conservation of endemic threatened squamates (lizards and snakes) in the Western Ghats. *Global Ecology and Conservation* 28:e01700. <https://doi.org/10.1016/j.gecco.2021.e01700>
- Stockwell DRB, Noble IR. 1992. Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. *Mathematics and Computers in Simulation* 33:385-390. [https://doi.org/10.1016/0378-4754\(92\)90126-2](https://doi.org/10.1016/0378-4754(92)90126-2)
- Taylor AT, Hafen T, Holley CT, González A, Long JM. 2020. Spatial sampling bias and model complexity in stream-based species distribution models: a case study of Paddlefish (*Polyodon spathula*) in the Arkansas River basin, USA. *Ecology and Evolution* 10(2):705-717. <https://doi.org/10.1002/ece3.5913>
- Thuiller W, Brotons L, Araújo MB, Lavorel S. 2004. Effects of restricting environmental range of data to project current and future species distributions. *Ecography* 27:165–172. <https://doi.org/10.1111/j.0906-7590.2004.03673.x>
- Valavi R, Guillera-Arroita G, Lahoz-Monfort JJ, Elith J. 2022. Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. *Ecological Monographs* 92(1):e01486. <https://doi.org/10.1002/ecm.1486>
- VanDerWal J, Shoo LP, Graham C, Williams SE. 2009. Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? *Ecological Modelling* 220:589–594. <https://doi.org/10.1016/j.ecolmodel.2008.11.010>
- Varela S, Anderson RP, García-Valdés R, Fernández-González F. 2014. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography* 37:1084-1091. <https://doi.org/10.1111/j.1600-0587.2013.00441.x>

- Veloz SD. 2009. Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography* 36:2290-2299. <https://doi.org/10.1111/j.1365-2699.2009.02174.x>
- Václavík T, Meentemeyer RK. 2012. Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. *Diversity and Distributions* 18(1):73-83. <https://doi.org/10.1111/j.1472-4642.2011.00854.x>
- Warren DL, Seifert SN. 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications* 21:335–342. <https://doi.org/10.1890/10-1171.1>
- Warren DL, Matzke NJ, Iglesias TL. 2020. Evaluating presence-only species distribution models with discrimination accuracy is uninformative for many applications. *Journal of Biogeography* 47:167–180. <https://doi.org/10.1111/jbi.13705>
- Weaver KF, Anderson T, Guralnick R. 2006. Combining phylogenetic and ecological niche modeling approaches to determine distribution and historical biogeography of Black Hills Mountain snails (Oreohelicidae). *Diversity and Distributions* 12:756–766. <https://doi.org/10.1111/j.1472-4642.2006.00289.x>
- Wenger SJ, Olden JD. 2012. Assessing transferability of ecological models: an underappreciated aspect of statistical validation. *Methods in Ecology and Evolution* 3(2):260-267. <https://doi.org/10.1111/j.2041-210X.2011.00170.x>
- White JW, Rassweiler A, Samhouri JF, Stier AC, White C. 2014. Ecologists should not use statistical significance tests to interpret simulation model results. *Oikos* 123:385–388. <https://doi.org/10.1111/j.1600-0706.2013.01073.x>
- Whittier TR, Ringold PL, Herlihy AT, Pierson SM. 2008. A calcium-based invasion risk assessment for zebra and quagga mussels (*Dreissena* spp). *Frontiers in Ecology and the Environment* 6(4):180-184. <https://doi.org/10.1890/070073>
- Wildlife Conservation Society, and Center for International Earth Science Information Network. 2005. Last of the Wild Project, Version 2, 2005 (LWP-2): Global Human Influence Index (HII) Dataset (Geographic). Palisades, New York: NASA

Socioeconomic Data and Applications Center (SEDAC).
<https://doi.org/10.7927/H4BP00QC>

Wisz MS, Hijmans RJ, Li J, Peterson AT, Graham CH, Guisan A, NCEAS Predicting Species Distributions Working Group. 2008. Effects of sample size on the performance of species distribution models. *Diversity and Distributions* 14(5):763-773. <https://doi.org/10.1111/j.1472-4642.2008.00482.x>

Wunderlich RF, Lin YP, Anthony J, Petway JR. 2019. Two alternative evaluation metrics to replace the true skill statistic in the assessment of species distribution models. *Nature Conservation* 35:97-116.
<https://doi.org/10.3897/natureconservation.35.33918>

Yates KL, Bouchet PJ, Caley MJ, Mengersen K, Randin CF, Parnell S, Fielding AH, Bamford AJ, Ban S, Márcia Barbosa A, Dormann CF, Elith J, Embling CB, Ervin GN, Fisher R, Gould S, Graf RF, Gregr EJ, Halpin PN, Heikkinen RK, Heinänen S, Jones AR, Krishnakumar PK, Lauria V, Lozano-Montes H, Mannocci L, Mellin C, Mesgaran MB, Moreno-Amat E, Mormede S, Novaczek E, Opper S, Ortuño Crespo G, Peterson AT, Rapacciuolo G, Roberts JJ, Ross RE, Scales KL, Schoeman D, Snelgrove P, Sundblad G, Thuiller W, Torres LG, Verbruggen H, Wang L, Wenger S, Whittingham MJ, Zharikov Y, Zurrel D, Sequeira AMM. 2018. Outstanding challenges in the transferability of ecological models. *Trends in Ecology & Evolution* 33(10):790-802. <https://doi.org/10.1016/j.tree.2018.08.001>

Yiwen Z, Wei LB, Yeo DCJ. 2016. Novel methods to select environmental variables in MaxEnt: a case study using invasive crayfish. *Ecological Modelling* 341:5-13.
<https://doi.org/10.1016/j.ecolmodel.2016.09.019>

Zhu G-P, Rédei D, Kment P, Bu W-J. 2014. Effect of geographic background and equilibrium state on niche model transferability: predicting areas of invasion of *Leptoglossus occidentalis*. *Biological Invasions* 16(5):1069-1081.
<https://doi.org/10.1007/s10530-013-0559-z>

Zurrell D, Berger U, Cabral JS, Jeltsch F, Meynard CN, Münkemüller T, Nehrbaß N, Pagel J, Reineking B, Schröder B, Grimm V. 2010. The virtual ecologist approach:

simulating data and observers. *Oikos* 119(4):622-35. <https://doi.org/10.1111/j.1600-0706.2009.18284.x>.