



Comparando las estimaciones de selección de hábitat mediante modelos de distribución de especies y *step selection functions*

Jorge F. S. Menezes^{1,2,*}

(1) Mitrani Department of Desert Ecology, Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, 84990, Midreshet Ben-Gurion, Israel.

(2) Center for Advanced Systems Understanding, Helmholtz Zentrum Dresden-Rossendorf, Untermarkt 20 D-02826, Görlitz, Sachsen, Germany.

* Autor de correspondencia: J.F.S. Menezes [jorgefernandosaraiva@gmail.com]

> Recibido el 14 de septiembre de 2022 - Aceptado el 04 de julio de 2023

Como citar: Saraiva de Menezes, J. F. 2023. Comparando las estimaciones de selección de hábitat mediante modelos de distribución de especies y *step selection functions*. *Ecosistemas* 32(2): 2455. <https://doi.org/10.7818/ECOS.2455>

Comparando las estimaciones de selección de hábitat mediante modelos de distribución de especies y *step selection functions*

Resumen: Recientemente, dos métodos de modelaje de selección de hábitat han ganado cada vez más relevancia en la literatura científica ecológica y de distribución de especies: las “step selection functions” (SSF) y el MaxEnt. A pesar de su semejanza, estos métodos raramente son usados en el mismo contexto. El primero es utilizado en modelos basados en datos de movimiento y el segundo en estudios de distribución de especies. Motivado por la dificultad de estimar modelos convergentes que tiene el SSF, he comparado la precisión de predicciones hechas por modelos MaxEnt en datos de movimiento. Como estudio de caso, utilicé datos de localizaciones de jaguares en cinco países de América Latina, y creé modelos de los dos tipos. Como variables de predicción utilicé datos climáticos y de uso del terreno, disponibles de imágenes de satélite. Comparé el rendimiento de ambos modelos mediante validación cruzada, y midiendo el área debajo de la curva (AUC) en el conjunto de datos de prueba. Los modelos de SSF presentaron una precisión media de 0.5482 ± 0.0217 en comparación con 0.7205 ± 0.0142 en los modelos MaxEnt equivalentes. Atribuyo estas diferencias, en parte, a la dificultad de los modelos SSF, y las regresiones logísticas condicionales, de converger en sus estimaciones. Por eso, yo recomiendo la utilización de modelos MaxEnt para actividades predictivas, como el diseño de reservas naturales o de corredores de fauna.

Palabras clave: América Latina; jaguares; modelaje de nicho; *resource selection function*; trayectoria

Comparing estimation of habitat selection between species distribution modelling and step selection functions

Abstract: Recently, two methods of habitat selection have gained more relevance in the ecological and animal distribution literature: step selection functions (SSF) and MaxEnt. Despite their similarity these models are hardly ever used in the same context. The former is usually associated with studies based in movement ecology, and the latter is connected to species distribution modeling. Motivated by the difficulty in estimating habitat preferences using SSF, I compared the accuracy of predictions from both models based on movement data. As a case study, I used jaguar movement data from 5 countries in Latin America and created SSF and MaxEnt models based on climatic data and land use available from WorldClim and satellite imagery. I compared the accuracy of both types of models using the “Area Under Curve” (AUC) metric, on a separate subset of data. SSF models presented an average AUC of 0.5482 ± 0.0217 in comparison with 0.7205 ± 0.0142 of their MaxEnt equivalents. I believe those differences are partially caused by the convergence difficulties of SSF and its use of conditional logistic regression. Consequently, I recommend the use of MaxEnt in predictive modelling, such as the ones needed in reserve and corridor design.

Keywords: Latin America; jaguars; niche modelling; resource selection function; trajectory

Introducción

La selección de hábitat es una de las áreas de estudio de más interés en ecología. Muchos estudios han investigado la selección de animales y otros organismos por cada tipo de hábitat, en taxones tan diversos como aves (Hildén 1965), mamíferos (Goulart et al. 2009), insectos (Sheldon y Haick 1981) y hasta protozoos (Svavars-son y Olafsdottir 1999). Así, no es sorprendente que también haya muchos estudios teóricos sobre selección de hábitat. Mientras que muchos de ellos se centran en prever la distribución de animales entre diferentes hábitats (Fretwell y Lucas 1969; Tregenza 1995; Cressman y Krivan 2006), también hay mucho interés en la mejora metodológica de las estimaciones de selección. Hay un cuerpo sig-

nificativo de estudios que desarrollan nuevos métodos y índices para estimar la selección (Manly 1974; Morris 1987, 2003; Brown 1988). Uno de los problemas centrales para cuantificar la selección de hábitat es el problema de la disponibilidad, dado que la selección y la presencia de un animal están condicionadas por los hábitats a los que tiene acceso (i.e. el nicho realizado, las condiciones donde el animal efectivamente ocurre; que es diferente del fundamental, las condiciones onde el animal podría ocurrir). Un jaguar puede ser encontrado con muy frecuencia en pastizales, aunque este no sea un ambiente adecuado para él. Eso puede ocurrir, por ejemplo, porque no haya hábitats más adecuados en la región. Por consiguiente, es necesario evaluar qué hábitats están disponibles para cada animal, esto es, cuál es el tamaño del área disponible para él.

Recientemente, una serie de estudios han propuesto el uso de datos de movimiento animal para estimar esta área de disponibilidad. Una de las primeras tentativas fue el uso de funciones de selección de recursos (en inglés, *resource selection functions*, Manly et al. 2002). En esta técnica se utilizan datos de la localización de los animales para estimar su área de campeo, asumiendo que esta área está disponible para el animal en todo momento. Un avance de esta técnica son las llamadas funciones de selección por pasos (*step selection functions*, SSF, Fortin et al. 2005). Este método propone que la selección de hábitat sea medida a cada paso, comparando el paso real con pasos potenciales. Estos pasos potenciales son muestreados de una distribución del tamaño de paso y una distribución de ángulos estimada a partir de la totalidad de pasos que ha realizado el animal. Así, estos pasos aleatorios se comparan con el real para estimar la probabilidad relativa de localizar al animal en un determinado lugar, basándose en las características ambientales en esta localización. Es importante mencionar que esta técnica compara la presencia del animal con localizaciones donde el animal no fue observado. No obstante, es posible que el animal haya visitado las áreas escogidas en el paso simulado entre las localizaciones reales. Por lo tanto, así como los métodos de distribución de especies (descritos a continuación), el SSF es un método basado en comparar presencias con muestras de *background* y en cuantificar preferencias relativas. El hecho de que esta técnica utilice información de movimiento, atrajo el interés de investigadores de ecología del movimiento, y muchos estudios la han utilizado (e.g. Coulon et al. 2008; Roever et al. 2010). Algunos estudios sugieren que esta técnica tiene un gran valor para la conservación de especies, pudiendo ser utilizada para el diseño de corredores de fauna y para calcular los periodos en que el conflicto entre humanos y animales es mayor (Thurfjell et al. 2014). Por tanto, queda claro que la SSF representa una herramienta de importancia para la ecología del movimiento y la conservación animal.

Esta herramienta puede permitir la incorporación de movimiento en los análisis, pero tal complejidad no es gratuita. El uso de SSF requiere seleccionar un número fijo de pasos aleatorios para compararlos con el paso verdadero. La selección de un número fijo de muestras en cada grupo crea un sesgo, ya que los métodos estadísticos tradicionales asumen que los datos han sido tomados de manera aleatoria, lo que impide conocer el número de muestras de cada grupo antes de la toma de datos. Ese problema también se encuentra con frecuencia en estudios de epidemiología, que habitualmente emparejan datos de pacientes infectados con los de un número fijo de controles que no presentan síntomas (Breslow y Day 1980; McCullagh y Nelder 1983). Como solución, se ha propuesto la utilización de regresiones logísticas condicionales. Estas incorporan el concepto de que las muestras están agrupadas en estratos, y consideran esa estratificación en el cálculo de las estimaciones (Breslow y Day 1980; McCullagh y Nelder 1983). La incorporación de estos estratos aumenta la precisión de las estimaciones, y este aumento es mayor cuando los estratos son pequeños (Pike et al. 1980). Sin embargo, el cálculo de la verosimilitud es poco eficiente con el uso de estratos. Hay estudios que muestran que si las diferencias entre estratos son pequeñas, o si estas pueden ser representadas en variables continuas, se puede obtener mayor confianza y menor varianza en las estimaciones con regresiones logísticas tradicionales (Hansson y Khamis 2008; Sjölander y Greenland 2013). Por consiguiente, el uso de regresiones logísticas condicionales, y del SSF, no es garantía de mejora en comparación con una simple regresión logística.

Mientras tanto, otro avance de la regresión logística ha aumentado su interés para la ecología. Los estudios de modelaje de distribución de especies utilizan una serie de técnicas de clasificación, siendo MaxEnt (Phillips et al. 2006, 2017; Elith et al. 2011) una de las más populares. Esta técnica tiene un objetivo muy semejante al de SSF: también estima la probabilidad de presencia, relativa a la probabilidad de presencia base del plano de fondo (*background*). Para ello, la técnica utiliza datos de presencia y localizaciones de

control sorteadas aleatoriamente del plano de fondo. La similitud entre MaxEnt y SSF es aparente. Incluso, la *resource selection function* es considerada una forma de modelaje de distribución de especies (Elith y Leathwick 2009). A pesar de no considerar información sobre pasos, MaxEnt tiene otras formas de mejorar su precisión. Por un lado, es una regresión logística tradicional, con un proceso de regularización añadido, donde el coeficiente de factores ambientales tienen su estimaciones reducidas y reduciendo sobreajuste (Elith et al. 2011). Además, puede incorporar variables no lineales con mayor facilidad. Por ejemplo, el programa original de cálculo de MaxEnt permite la inclusión de funciones cuadráticas, con quiebras, y límites (Phillips et al. 2017). Debido a ello, los modelos pueden representar una gran variedad de relaciones no lineales entre variables ambientales y las probabilidades. Esta técnica ha logrado un enorme éxito de utilización para estimar la distribución de especies, con más de 2000 modelos publicados (Bradie y Leung 2017) y junto con R, es una de las herramientas más utilizadas en modelaje de distribución (Ahmed et al. 2015).

Sin embargo, no hay, en principio, una razón por la que esta técnica no pueda funcionar también con datos de movimiento para estimar selección de hábitat. Los dos métodos tienen la misma estructura conceptual. Ambos comparan las localizaciones de animales con puntos muestreados aleatoriamente del área de estudio (*background*). Ambos los utilizan una variante de la regresión logística para medir cómo diferentes variables ambientales pueden aumentar o disminuir la probabilidad de que una determinada muestra sea una presencia o una de *background*. Por tanto, ambos los modelos estudian el mismo objeto matemático, lo que sucede es simplemente que lo interpretan manera diferente en los dos tipos de estudio. En MaxEnt y modelos de distribución de especies, se interpreta ese número como un indicativo de "calidades de hábitat". Mientras tanto, los estudios con SSF interpretan la misma cantidad como selección por el hábitat. Esa equivalencia es reforzada por algunos autores que han utilizado MaxEnt para inferir selección de hábitat (e.g. Francis et al. 2017; Taher et al. 2018). Considerando que la aplicación de SSF puede no ser la solución más eficaz, es posible que técnicas como MaxEnt logren ser mejores en estimar la selección de hábitat de los animales.

Para resolver esta duda, en este estudio he comparado la eficiencia de MaxEnt y de SSF en prever el uso de hábitat de jaguares por toda América Latina. Como métrica para la comparación, utilicé la validación cruzada, es decir, separé los datos en conjuntos de entrenamiento y prueba y comparé la capacidad de los dos modelos para separar entre presencias y *background* en el conjunto de prueba. Mi predicción es que la técnica de SSF sea la más precisa, debido a que utiliza más información (los ángulos y tamaños de paso) que la técnica MaxEnt.

Material y métodos

Para testar la precisión de las dos técnicas con datos de movimiento animal, utilicé una base de datos previamente publicada sobre jaguares (*Panthera onca*) de América Latina (Morato et al. 2018). Esta base de datos contiene información espacial de cerca de 117 animales con 134 690 localizaciones distribuidas entre 21 estudios de cinco países (Argentina, Brasil, Costa Rica, México y Paraguay). Como el proceso de prueba requiere la creación de muchos puntos de *background* para cada uno de presencia, necesité reducir el número de localizaciones, para que el análisis pudiese ser concluido en un plazo de tiempo adecuado. Para ello, utilicé cinco estudios del conjunto de datos, uno de cada país estudiado, que se identifican con los siguientes nombres en la base de datos: "Jaguar conservation in Argentina", "Jaguar Taima", "Costa Rica", "Greater Lacadona System" y "Dry Chaco" (ver mapa en la Fig. 1, y detalles en Fig. 2). Con estos cinco estudios, utilicé 30 individuos y 63 558 localizaciones, que habían sido obtenidas mediante collares GPS, con intervalos de 1 a 8 horas dependiendo del estudio.

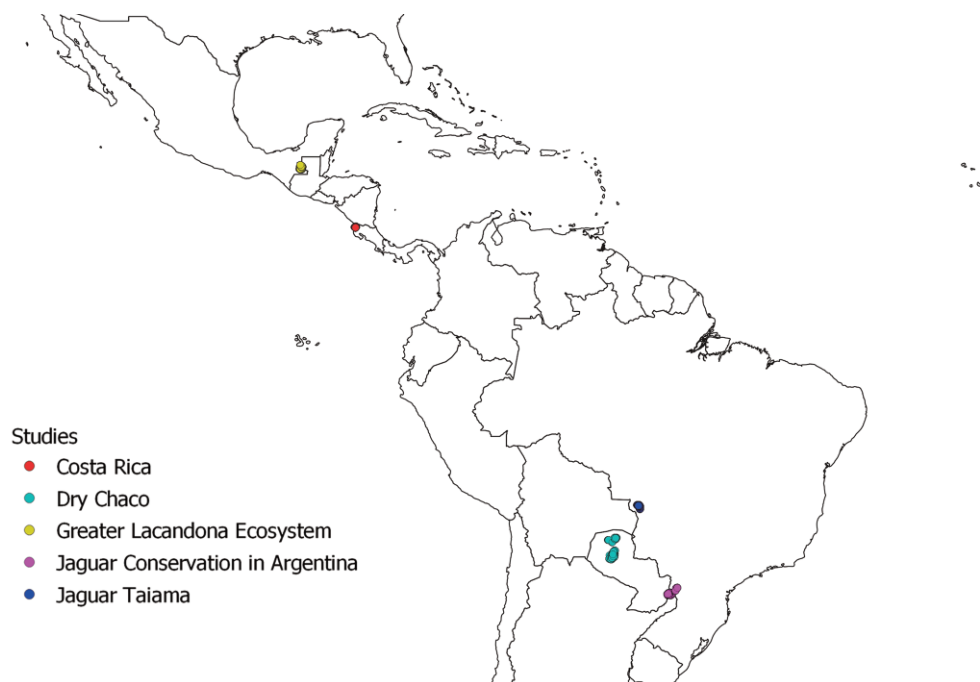


Figura 1. Mapa con las localizaciones de jaguares utilizadas. Las localizaciones pertenecen a cinco estudios que las agrupan geográficamente, representadas en el mapa con diferentes colores.

Figure 1. Map with the jaguar locations used. The locations belong to five studies that group them geographically, represented on the map with different colours.

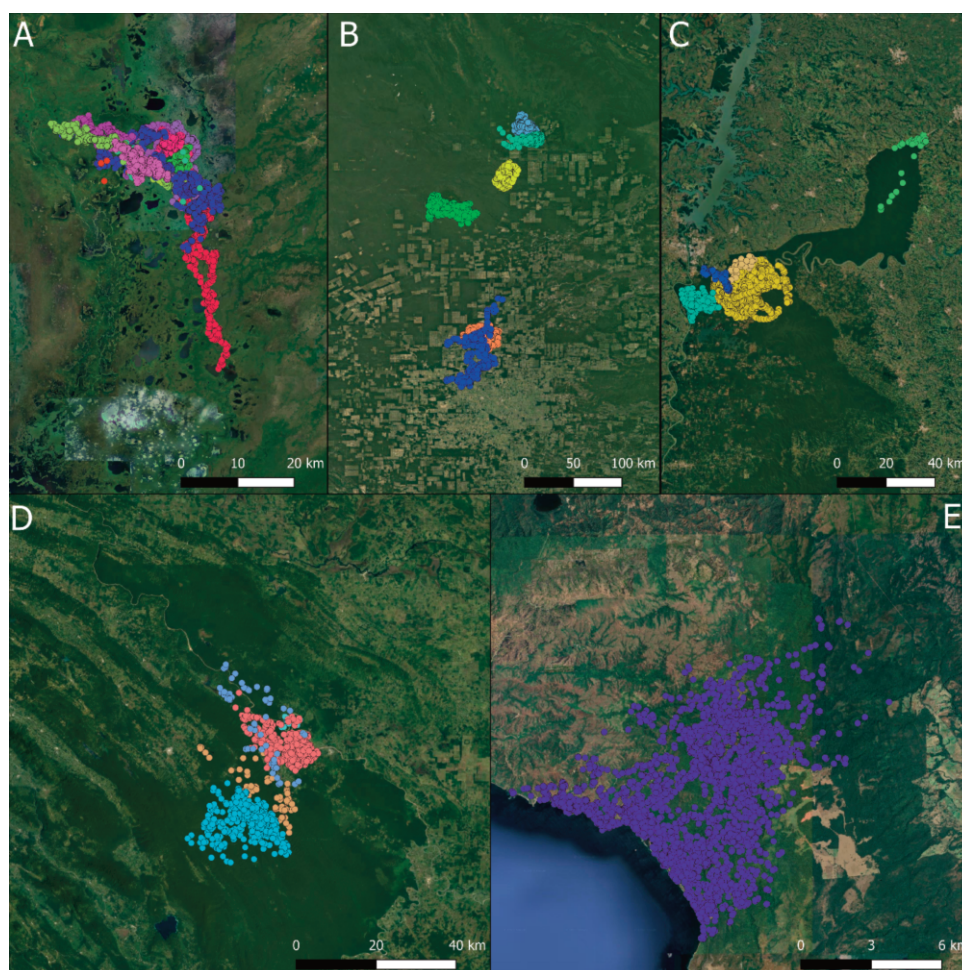


Figura 2. Distribución de las presencias en cada área de estudio. Imagen de satélite proporcionada por Google maps, través del plugin Web map services em QGIS. Todas las imágenes tienen el norte en la parte superior. Los estudios están dispuestos tal como sigue: A – Jaguar Taima, B- Dry Chaco, C- Jaguar Conservation in Argentina, D- Greater Lacadona Ecosystem, E- Costa Rica. Diferentes colores indican diferentes animales.

Figure 2. Distribution of occurrences in each study area. Satellite image provided by Google maps, through the Web map services plugin in QGIS. All images have north at the top. The surveys are arranged as follows: A - Jaguar Taima, B- Dry Chaco, C- Jaguar Conservation in Argentina, D- Greater Lacadona Ecosystem, E- Costa Rica. Different colours indicate different animals.

Una vez obtenidos los datos de presencia, necesité crear las correspondientes muestras de *background*. Para el análisis de SSF, ese proceso requirió múltiples pasos. Primero, interpolé las trayectorias de cada individuo, para asegurar que las localizaciones tienen exactamente los mismos intervalos de tiempo entre sí. Cuando una localización estaba muy lejos de las anteriores (más de 1.2 veces el intervalo de muestreo programado en el collar GPS), no realicé interpolación. Por ello, las trayectorias de los animales se transformaron en discontinuas, así que analicé cada sub-trayectoria separadamente. Después de la interpolación, retiré del análisis aquellas sub-trayectorias con menos de cinco localizaciones. Para cada una de las sub-trayectorias que superaron este criterio, calculé la distancia entre localizaciones (el tamaño de paso) y ajusté una distribución gamma a estas distancias. Una vez que tenía los parámetros de esta distribución, escogí para cada punto de presencia 15 puntos de *background*, respetando una distancia con la localización anterior sorteada de la distribución gamma, y un ángulo aleatorio sorteado de una distribución uniforme. Este valor parece demasiado bajo en comparación con el número de muestras de *background* del MaxEnt, pero se considera más que suficiente para datos de GPS (un único punto de *background* podría ser suficiente) (Clark et al. 2015; Francis et al. 2017). Además, vale la pena recordar que los datos de GPS son mucho más abundantes que en los estudios de modelaje tradicional.

Para el análisis de MaxEnt, el proceso de selección de puntos aleatorios fue mucho más simple. Así, escogí puntos de manera uniforme dentro de un área alrededor de las localizaciones (descrita en el párrafo siguiente), en una cantidad semejante a la del SSF (siete muestras de *background* por cada una de presencia). De nuevo, siete puntos pueden parecer un número muy pequeño, ya que los estudios de MaxEnt trabajan en números absolutos y de aproximadamente 1000 muestras de *background*. Sin embargo, yo mantuve el número en siete muestras de *background* por cada una presencia para mantener este análisis lo más próximo al análisis de SSF que fuera posible, y por limitaciones computacionales.

Además, utilicé dos áreas diferentes. Un área fue determinada como la distancia dada por el percentil 95% de la distribución gamma de los pasos de todos los animales: 2 km (después de redondear). Esta distancia fue escogida para que la comparación sea justa con el método de SSF. Sin embargo, se puede argumentar que esa distancia hace a MaxEnt dependiente de la información de SSF. Para evitar tal crítica también escogí una segunda área, usando 100 km alrededor de las localizaciones de cada animal, que sería más similar al área utilizada en modelos de distribución de especies. En ambos casos, los puntos de presencia también fueron añadidos al conjunto de muestras de *background*, conforme al procedimiento estándar en MaxEnt.

Una vez obtenidos los puntos de presencia y *background*, recopilé información sobre el clima y el medio ambiente para cada localización. Más específicamente, utilicé las variables climáticas del Worldclim v.2 (Fick y Hijmans 2017). Esta base contiene datos de pluviosidad, temperatura y diversas combinaciones de estas variables. Además, utilicé datos de uso del terreno obtenidos por satélite desde la plataforma Copernicus (Buchhorn et al. 2020), la distribución de población humana obtenida mediante GPWv4.11 (SEDAC-CIESIN-Columbia University 2016), la altitud de la base SRTM (Jarvis et al. 2008) y la distancia a los cuerpos de agua calculada a partir de un mapa de distribución de cuerpos de agua, con distancia máxima de 2250 m (Global Land Cover Facility 2022). Los últimos cuatro conjuntos de datos fueron obtenidos por medio de Google Earth Engine (Gorelick et al. 2017). Para las variables de Worldclim, realicé la extracción de datos para cada localización usando el paquete “terra” (Hijmans 2021) del lenguaje R (R Core Team 2019). Estas variables tenían resoluciones diferentes. Para las variables climáticas, la unidad eran píxeles de 1 km. Para la cobertura del terreno, la resolución era de 100 m, 927 m para pobla-

ción, y 30 m para la distancia al agua. Como separé los puntos de *background* de manera independiente, y utilicé a función *maxnet* en R, no necesite ajustar la resolución de las imágenes como es común cuando se utiliza el programa MaxEnt. Los valores de cada variable fueron obtenidos para cada punto, considerando el píxel del punto en que se encontraba.

Una vez obtenida la base de datos completa de las localizaciones y sus correspondientes variables ambientales, dividí las localizaciones en conjuntos de entrenamiento y prueba, con el conjunto de prueba conteniendo el 25% de los datos. Tanto para el SSF como para MaxEnt, la división fue hecha dentro de cada individuo. No obstante, en el caso del SSF, me aseguré de que ninguna sub-trayectoria estuviera en ambos grupos de entrenamiento y prueba. Por tanto, un cuarto de todas las muestras de sub-trayectorias fue colocada en el grupo de prueba.

Después de la división, creé un modelo para el SSF, que vino determinado principalmente por aquellas variables que podían ser estimadas. Como muchas de las variables climáticas estaban correlacionadas ($r > 0.8$), fue necesario eliminar algunas de ellas para que el modelo fuese estimable. En total, utilicé las siguientes variables: temperatura anual media, intervalo térmico medio diario (media de la temperatura máxima menos la mínima), precipitación media, precipitación en el mes más húmedo. Además, incluí las interacciones de segundo orden entre todas las variables, excepto del uso del terreno, ya que el modelo no convergía en una solución si incluía las interacciones entre esa variable y otras.

Para el análisis con MaxEnt utilicé dos modelos con formulaciones diferentes. En una primera formulación, utilicé la misma fórmula del modelo de SSF. Con eso, sería posible entender si la diferencia entre los modelos era causada por otro factor que no fuesen las variables, como la regularización o las diferencias en la forma de calcular la verosimilitud. En una segunda formulación, permití utilizar transformaciones en las variables. Más específicamente, permití la aplicación de transformaciones cuadráticas, con quiebras, y límites. Con eso, pude inferir si esas transformaciones causarían una mejoría o reducción de la precisión de los modelos en comparación con el análisis de SSF. Todos los dos modelos fueron calculados utilizando el paquete “maxnet” (Phillips 2021) en R 4.2.1 (R Core Team 2019).

En total, ajusté cinco tipos de modelos: 1) SSF, 2) MaxEnt con 100 km de radio y sin variables extra, 3) MaxEnt con 100 km radio y con variables extra, 4) MaxEnt con radio de distribución gamma y variables extra, y 5) MaxEnt con radio de distribución gamma y sin variables extra. Cada uno de estos modelos fue aplicado a cada uno de los cinco estudios en la base de datos. Por tanto, cada uno de los 5 modelos fue comparado dentro de cada uno de los estudios, evitando así comparaciones entre ambientes muy diferentes o alejados, que podrían favorecer métodos como los de modelaje de distribución de especies. Los modelos de cada estudio fueron repetidos 10 veces, para eliminar la variación causada por la elección aleatoria de muestras de *background*, entrenamiento, y test. En total, tenemos 250 conjuntos de estimaciones (5 modelos x 5 estudios x 10 réplicas). En total, calculé 14 651 974 muestras de *background*. La cantidad de puntos de *background* para cada conjunto fue muy variable (Mínimo: 146, Mediana: 11 009, Máximo: 185 147). Así, los modelos fueron utilizados para predecir la presencia y *background* en el respectivo conjunto de prueba. Posteriormente, comparé su precisión usando la métrica del área debajo de curva (AUC, en las siglas en inglés), y haciendo curvas del índice kappa para diferentes probabilidades. Los valores de AUC fueron comparados usando una prueba de ANOVA, con contrastes para comparar SSF con cada uno de los modelos de MaxEnt.

Registro y disponibilidad de datos

Este estudio no tiene datos originales. Las fuentes de datos están disponibles en los métodos.

Resultados

Los resultados de AUC indican una diferencia muy grande entre los diferentes métodos, y muestran que el SSF fue el peor de los tipos de modelos analizados (Fig. 3). El mejor modelo fue el de MaxEnt con 100 km, y no se observaron grandes diferencias entre añadir las variables extra o no. Sin embargo, incluso cuando se compara dentro del área del SSF, se encuentra que los valores de

AUC son significativamente superiores para el MaxEnt que los valores de SSF (Fig. 4; Tabla 1, todos los p-valores menores que 0.001). La distribución de valores kappa indicó el mismo resultado, con valores muchos menores para SSF que para MaxEnt, independientemente del punto de corte (Fig. 5). Finalmente, un análisis espacial muestra las diferencias entre los dos tipos de modelos (Fig. 6). El modelo MaxEnt de 2 km indica, de manera sistemática, una probabilidad de presencia mayor en casi todas las localidades.

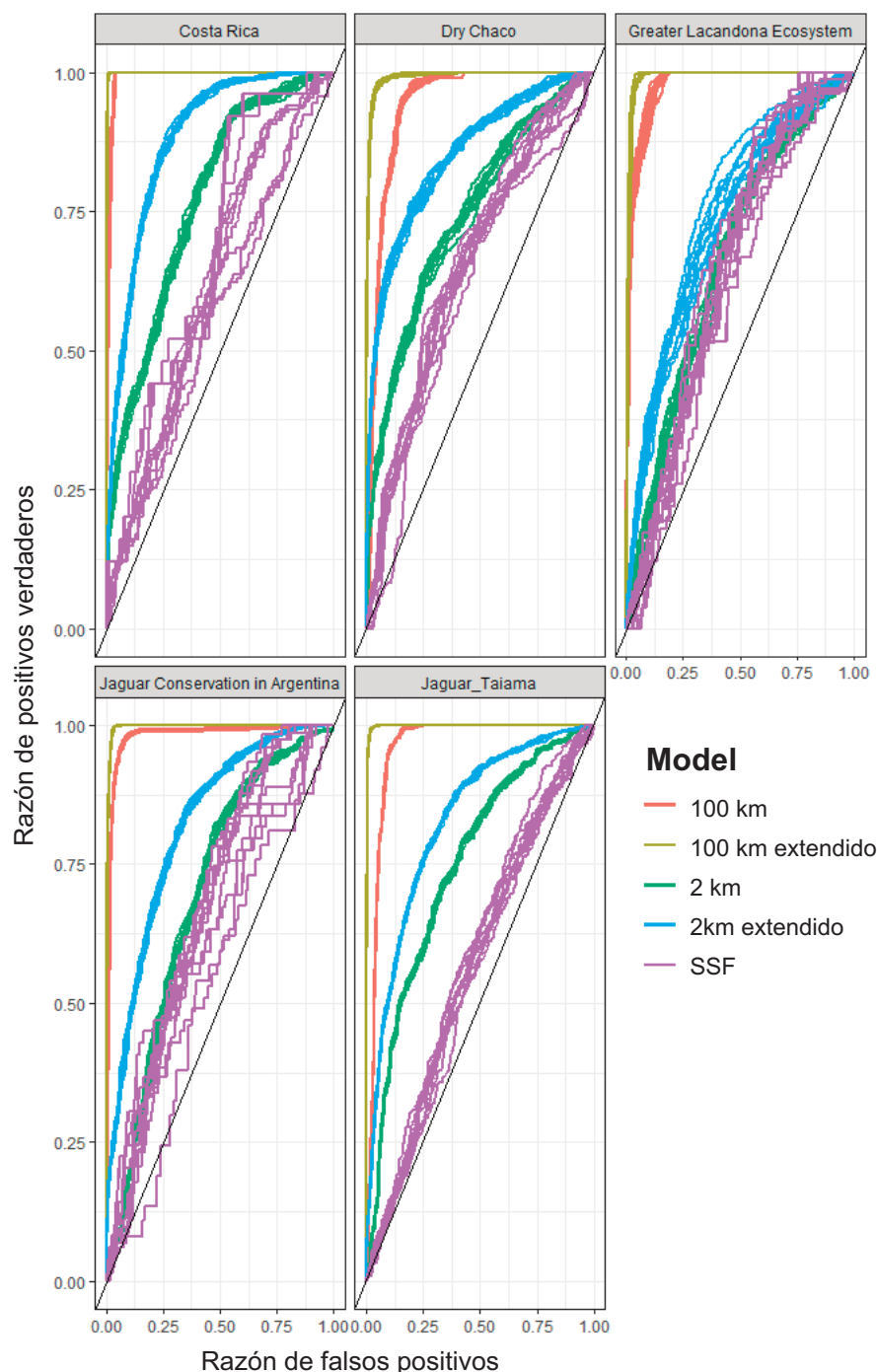


Figura 3. Curvas ROC (característica de operación del receptor) para cada uno de los modelos y estudios (5 modelos, 10 réplicas, en cada uno de los 5 estudios, con un total de 250 curvas). Estas curvas representan la capacidad del modelo de diferenciar entre presencias y background (puntos obtenidos del plano de fondo en MaxEnt). Los modelos producen probabilidades de presencia para cada localidad, lo que requiere el uso de un punto de corte. Cada punto en estas líneas representa la proporción de identificaciones correctas y de falso positivos para un determinado punto de corte. Naturalmente, si el punto de corte es muy bajo, habrá muchas identificaciones correctas y muchos falsos positivos. Cuanto más próxima está la curva de la esquina superior izquierda, mejor es el modelo. El AUC es una representación numérica de esta curvatura.

Figure 3. ROC (receiver operating characteristic) curves for each of the models and studies (5 models, 10 replicates, in each of the 5 studies, with a total of 250 curves). These curves represent the model's ability to differentiate between presences and background (points obtained from the background plane in MaxEnt). The models produce presence probabilities for each locality, which requires the use of a cut-off point. Each point on these lines represents the proportion of correct identifications and false positives for a given cut-off point. Naturally, if the cut-off point is too low, there will be many correct identifications and many false positives. The closer the curve is to the upper left corner, the better the model. The AUC is a numerical representation of this curvature.

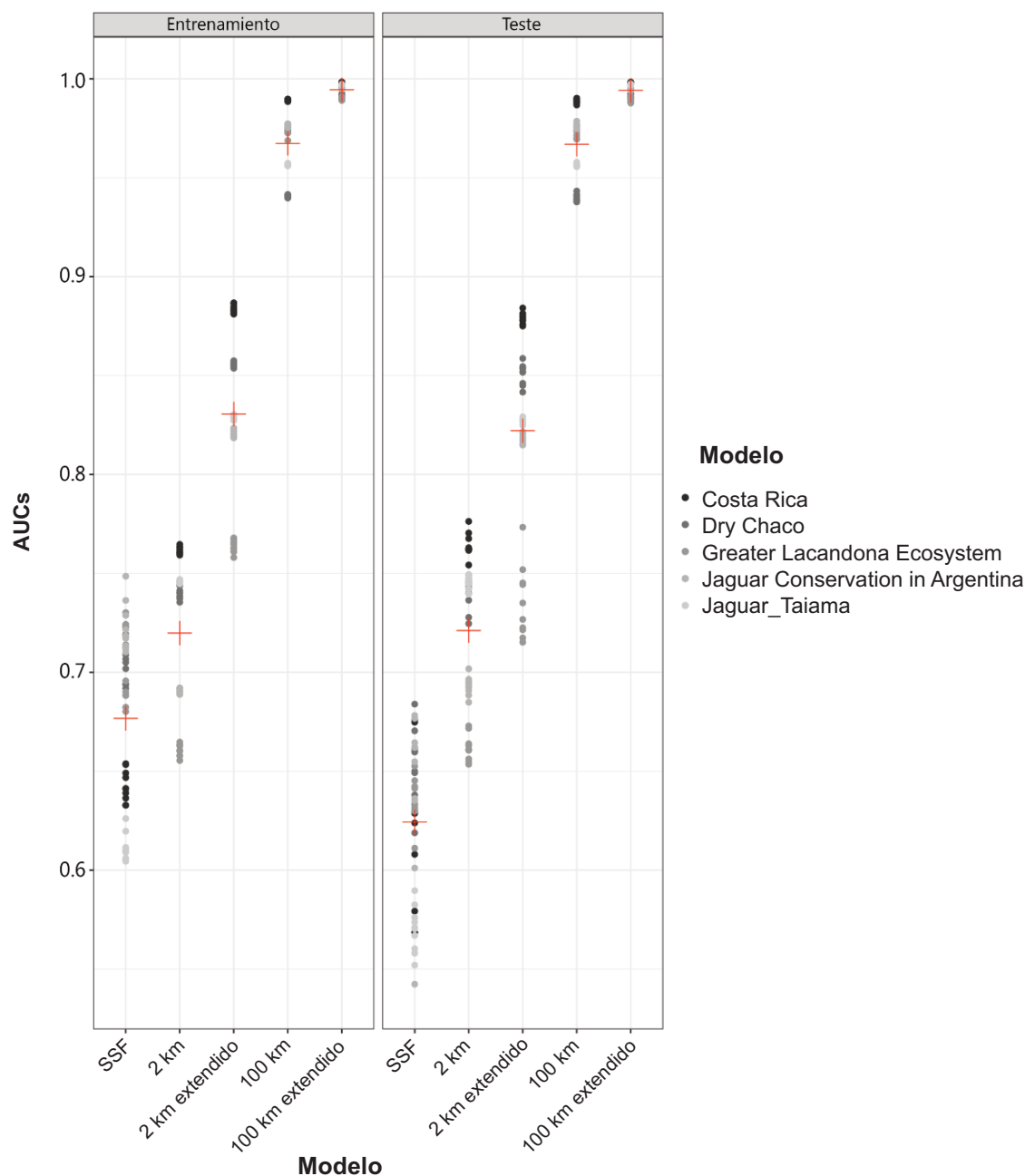


Figura 4. Grafico con la distribución de valores de AUC para cada tipo de modelo. Las cruces en rojo indican la media de cada modelo.

Figure 4. Graph showing the distribution of AUC values for each type of model. The red crosses indicate the mean of each model.

Tabla 1. Media y error estándar de AUC (área debajo de la curva) para cada uno de los dos modelos estudiados, para cada conjunto de datos. Nótese que el modelo de SSF tiene un rendimiento muy por debajo de los otros modelos, incluso cuando se compara con modelos MaxEnt que abarcan un área espacial semejante. La diferencia entre la media de la AUC de los modelos SSF es significativamente menor que los otros modelos, como se indica en los resultados de una prueba ANOVA.

Table 1. Mean and standard error of AUC (area under the curve) for each of the two models studied, for each dataset. Note that the SSF model performs well below the other models, even when compared to MaxEnt models covering a similar spatial area. The difference between the mean AUC of the SSF models is significantly smaller than the other models, as indicated by the results of an ANOVA test.

Estudio	SSF	MaxEnt 2km	MaxEnt 2km extendido	MaxEnt 100km	MaxEnt 100km extendido
Jaguar Conservation in Argentina	0.6413 ± 0.0132	0.6932 ± 0.0015	0.8179 ± 0.0008	0.9759 ± 0.0004	0.9949 ± 0.0001
Jaguar Taiama	0.5700 ± 0.0036	0.7460 ± 0.0010	0.8270 ± 0.0004	0.9564 ± 0.0002	0.9971 ± 0.0001
Costa Rica	0.6216 ± 0.0129	0.7641 ± 0.0019	0.8792 ± 0.0009	0.9886 ± 0.0003	0.9979 ± 0.0001
Greater Lacandona Ecosystem	0.6391 ± 0.0044	0.6618 ± 0.0021	0.7353 ± 0.0058	0.9741 ± 0.0010	0.9897 ± 0.0004
Dry Chaco	0.6495 ± 0.0063	0.7402 ± 0.0026	0.8511 ± 0.0016	0.9398 ± 0.0006	0.9913 ± 0.0003
Media	0.624246	0.7210	0.8221	0.9669	0.9941
Error estándar	0.0056	0.0054	0.0070	0.0024	0.00048
Error estándar (per-ANOVA)	0.0135	0.0068	0.0068	0.0068	0.0068
Estadístico T	129.09	14.15	28.92	50.10	54.08
Valor p	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

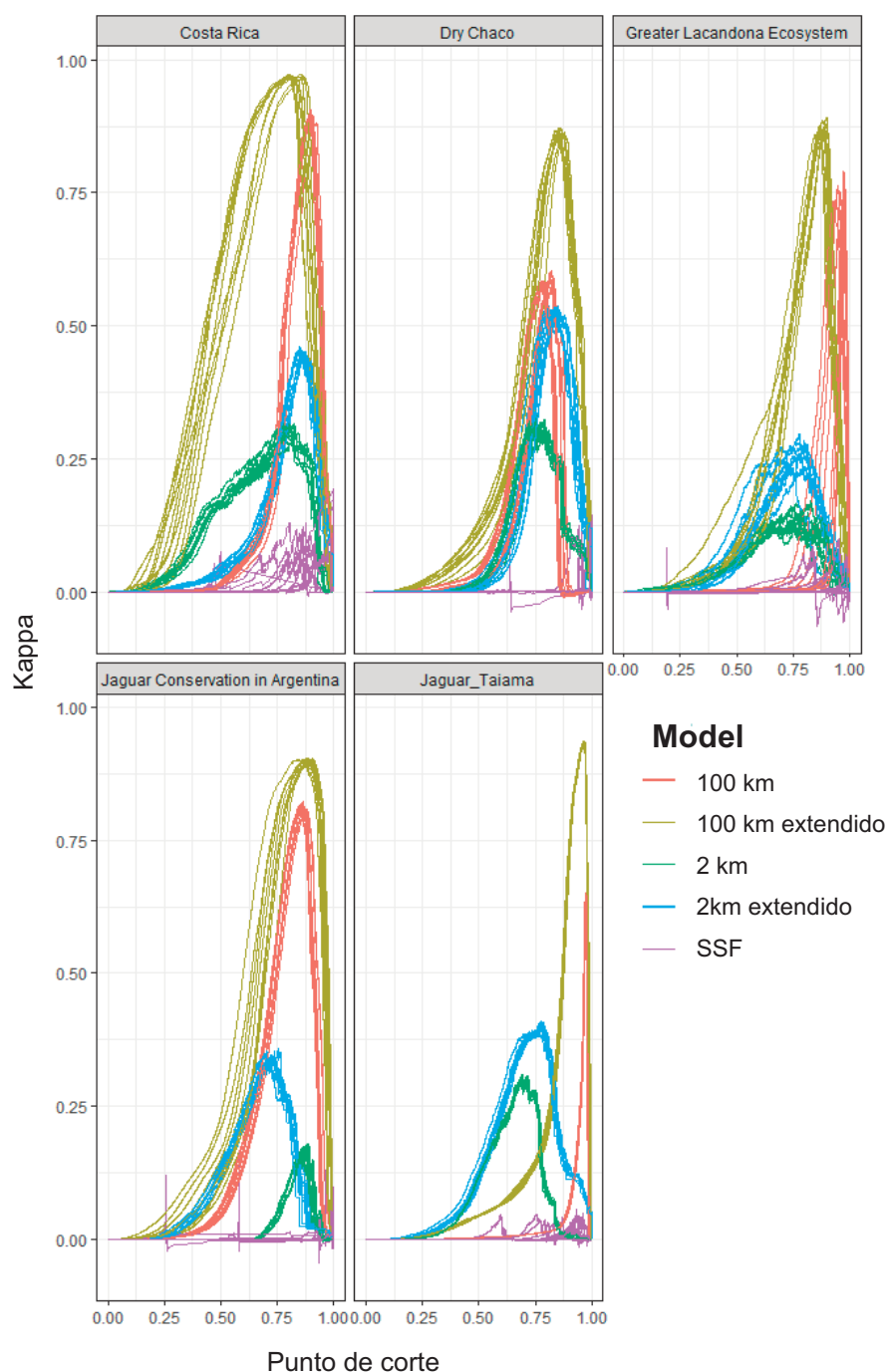


Figura 5. Distribución de valores de kappa dependiendo del valor del punto de corte. Cada línea representa una réplica (5 modelos, 10 réplicas, en cada uno de los 5 estudios). Los puntos de corte utilizados fueron escogidos por el proceso estándar de la función “evaluate” del paquete “dismo”, siendo diferentes para cada combinación de modelo y estudio. Para construir la figura, normalicé los puntos de corte utilizando la fórmula $(x - \min(x)) / (\max(x) - \min(x))$. Con eso, los valores de punto de corte vienen representados entre cero y uno. Nótese que el valor de kappa para los cinco modelos de SSF está muy debajo del valor de kappa de los modelos de MaxEnt.

Figure 5. Distribution of kappa values depending on the cut-off point value. Each line represents one replicate (5 models, 10 replicates, in each of the 5 studies). The cut-off points used were chosen by the standard process of the “evaluate” function of the “dismo” package, being different for each combination of model and study. To construct the figure, I normalised the cut-off points using the formula $(x - \min(x)) / (\max(x) - \min(x))$. With that, the cut-off point values are represented between zero and one. Note that the kappa value for the five SSF models is well below the kappa value of the MaxEnt models.

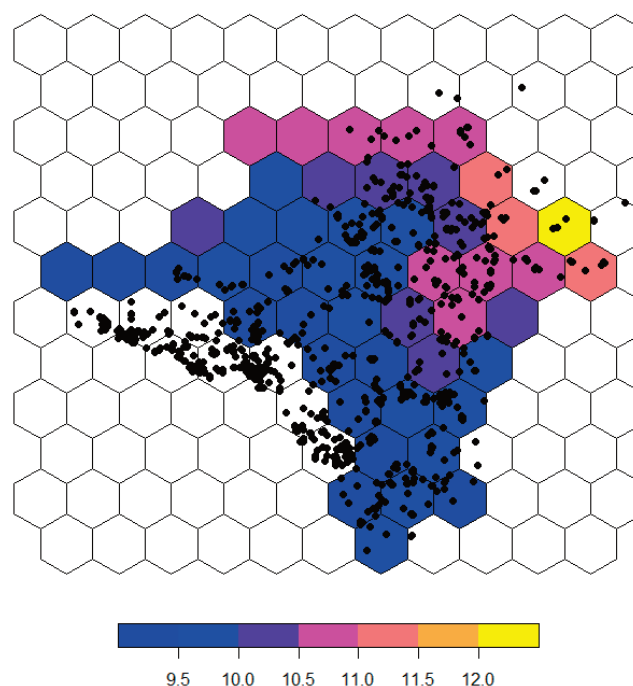


Figura 6. Medida espacial de la diferencia entre el modelo de SSF y MaxEnt 2 km sin variables extra, para el único jaguar del estudio de Costa Rica, y una de las réplicas. Nótese como la diferencia es positiva, indicando que las estimaciones de probabilidad de presencia son siempre mayores para MaxEnt. Nótese también que esa diferencia aumenta con una mayor distancia, en las áreas periféricas. Los valores de cada hexágono representan la media de la diferencia entre los valores predichos. Los valores no pudieron ser estimados para la esquina inferior izquierda de la figura debido a la proximidad con el océano, y porque algunas de las imágenes no contenían información para esos píxeles.

Figure 6. Spatial measure of the difference between the SSF model and MaxEnt 2 km with no extra variables, for the single jaguar in the Costa Rican study, and one of the replicates. Note how the difference is positive, indicating that estimates of probability of presence are always higher for MaxEnt. Note also that the difference increases with increasing distance, in outlying areas. The values in each hexagon represent the mean of the difference between the predicted values. Values could not be estimated for the lower left corner of the figure due to the proximity to the ocean, and because some of the images did not contain information for those pixels.

Discusión

Mis resultados indican que el método SSF es una mala alternativa para predecir la presencia o muestras de *background* de los animales, y, por lo tanto, sus trayectorias entre las diferentes localidades y su selección de hábitat. Existen dos explicaciones posibles para este resultado. Una primera explicación es que puede ser una consecuencia del uso de técnicas de regularización en MaxEnt que no están presentes en SSF. MaxEnt utiliza dos formas de regularización: lasso y ridge (Elith et al. 2011). Ambos métodos reducen la estimación de los coeficientes y, con eso, reducen la posibilidad de sobreajuste. Eso evitaría, teóricamente, que un modelo representase demasiado bien nuestra muestra, hasta el punto de volverse incapaz o ineficiente en prever la distribución de localizaciones fuera de la muestra. Aunque la regularización puede conllevar una influencia en el rendimiento de MaxEnt, no creo que este sea un factor determinante. Un estudio anterior ha comparado el efecto de modificar la intensidad de la regularización en la AUC del modelo, encontrando que eliminar toda la regularización resultaba en un cambio de AUC apenas próximo a 0.05 (Radosavljevic y Anderson 2014). Por consiguiente, parece que la regularización no es suficiente motivo para explicar una diferencia tan grande entre los dos métodos. Una segunda explicación es la forma en que los coeficientes son calculados en una regresión logística condicional, y, por lo tanto, en SSF. En este método, se suman las verosimilitudes de los puntos que pertenecen al mismo estrato, de forma que la verosimilitud de todo el modelo se calcula solamente después (Breslow y Day 1980). Así, si existen múltiples puntos por estrato, es posible que múltiples combinaciones de coeficientes generen la misma verosimilitud. Por ello, los modelos de regresión condicional logística tienen una notoria dificultad para converger, y frecuentemente contienen estimaciones alternativas. Algunos estudios utilizan menos variables para evitar problemas de convergencia (e.g. Coulon et al. 2008), o separan procesos de estimación para facilitar

el cálculo (Potts et al. 2014). Otros estudios sugieren, incluso, sustituir la regresión condicional por un proceso alternativo combinando modelos mixtos con un algoritmo de maximización de expectativas, solamente para evitar el uso de este método (Craiu et al. 2011; Muff et al. 2020). Por más que traté de asegurar que los modelos utilizados en este estudio convergieran, las estimaciones frecuentemente incluían valores muy altos, lo que puede sugerir que el modelo tiene dificultad de estimar valores realistas.

Cambiando el foco para los otros modelos, el AUC de los modelos de MaxEnt con 100 km es demasiado alto. Una exactitud de 99% no es común en datos ecológicos que frecuentemente tiene componentes aleatorios. Que los valores de AUC sean tan elevados podría indicar un sobreajuste en los individuos estudiados. Una explicación para ese sobreajuste es que las variables climáticas están correlacionadas espacialmente. Cuando se aumenta el área muestreada, se introduce más variación al conjunto de puntos de *background*, mientras que se mantiene la misma variación en el conjunto de presencia. Con eso, se vuelve más fácil distinguir entre *background* y presencias. Este proceso fue detectado originalmente por Lobo et al. (2008), cuando él mismo recomendó que no se utilicen los valores de AUC para medir la calidad de los modelos. Eso genera una paradoja interesante, en la que es más fácil encontrar animales cuando el área de búsqueda aumenta. En cualquier caso, ese aumento del AUC probablemente no represente una mejora de la precisión real, siendo un mero reflejo de que el problema de localización se vuelve trivial a partir de una determinada distancia (por ejemplo, es obvio que no hay jaguares en África). También se puede argumentar que estos grandes valores de AUC son resultado de mi utilización de variables ambientales que varían poco en escalas pequeñas, y que son más utilizadas en MaxEnt que en SSF. Para este argumento, hay dos contrarios. En primer lugar, los jaguares tienen un área de movimiento muy grande. Por lo tanto, es posible que los individuos experimenten variaciones en estos aspectos ambientales de amplia escala. En

segundo lugar, aunque las variables no sean ideales para el tamaño del área estudiada, esto no sería capaz de causar una diferencia entre las estimaciones de MaxEnt 2 km y SSF, ya que utilizan la misma área. En esta comparación, las únicas dos diferencias entre MaxEnt y SSF están en los detalles de cálculo (regresión logística condicional vs. regresión logística con regularizaciones) y en la forma de selección del *background* (con mayor probabilidad a cortas distancias o uniforme).

Otro resultado de interés es sobre la importancia de las variables no lineales. Los modelos extendidos de 2 km tienen un AUC mejor que sus equivalentes lineales, mejorando en un 10%. Esta diferencia no es tan marcada como la comparación entre SSF y MaxEnt, pero tampoco es trivial. Este resultado está de acuerdo con un estudio anterior, que cuantificó la sensibilidad de MaxEnt a sus factores (Convertino et al. 2014). Este estudio encontró que los coeficientes de variables con quiebras tenían una gran importancia en el resultado final. Esto no es sorprendente, ya que esta variable permite la representación de una cantidad muy grande de curvas y relaciones similares a otros modelos flexibles, como los modelos aditivos generalizados (Elith et al. 2011).

Mis resultados me llevan a cuestionar cuándo puede ser útil la SSF. Tal vez el argumento más favorable radica en que la SSF no perdió utilidad porque SSF y MaxEnt estiman cosas diferentes, selección y ocurrencia, respectivamente. Sin embargo, no creo que este argumento tenga mucha validez. Es cierto que existe una diferencia entre ocurrencia y selección, específicamente cuando no podemos asumir que el animal tenga acceso a toda el área de modelaje. No obstante, incluso cuando restringimos el área de actuación del modelo de MaxEnt a la misma área del modelo de SSF, tenemos un resultado mejor con MaxEnt. En esta situación, en ambos modelos el área disponible para el animal es semejante, y ambos trabajan dentro de ella. Por tanto, el problema de ocurrencia y de selección se vuelve prácticamente igual en estos casos. La única diferencia es el proceso de selección de muestras, con SSF seleccionado puntos más lejos de las localidades con menos frecuencia que los modelos de MaxEnt. No discuto el hecho de que los animales tienden a dar pequeños pasos con más frecuencia que darían pasos mayores, dado que ya existen abundantes evidencias y discusiones al respecto, así como sobre la forma exacta de esta distribución (Viswanathan et al. 1996, 1999; Benhamou 2007, 2008; Plank y James 2008). Sin embargo, si esa es la única ventaja del modelo de SSF, se puede obtener el mismo resultado simplemente seleccionando muestras en un mapa de pasos derivado de un kernel con distribución gamma. Por lo tanto, no sería necesario utilizar la estructura de estratos de una regresión logística condicional, con todas sus complicaciones matemáticas y de ajuste.

Con base en la literatura de regresión logística condicional, se puede argumentar que SSF tienen valor si sus estratos controlan una gran cantidad de variación (Hansson y Khamis 2008). Esto refuerza la idea defendida por Merow et al. (2014) de que los modelos complejos (como SSF) deben ser utilizados cuando existen sesgos de muestreo muy grandes. En el caso de la SSF, el grupo control es el paso, o sea, el inverso de la autocorrelación espaciotemporal entre las muestras. Nuestros resultados sugieren que el rendimiento negativo de la SSF puede ser resultado de un control excesivo por esa variación, que genera un sobreajuste en los datos y baja generalización. No obstante, SSF puede ser un método más preciso en ambientes con correlación espaciotemporal negativa, como sería el caso de ambientes en los que las variables ambientales cambiasen bruscamente en el espacio de un paso a otro. O, de manera semejante, si la frecuencia de toma de datos fuera tan baja que el animal pudiera desplazarse para un ambiente muy diferente. Con el aumento de datos auto correlacionados, como es el caso de GPS con resolución fina, y usando técnicas que controlan la autocorrelación como el *autocorrelated kernel density estimation* parece que SSF corrige el problema opuesto al que se ha estudiado y enfrentado en datos modernos (revisión en Silva et al. 2022).

Aunque el uso de SSF sea ocasional en estudios predictivos, es necesario reconocer que mis resultados no discuten su uso para estudios de inferencia. Mi comparación fue hecha en el contexto de predicción, no de inferencia. Cuando un modelo es utilizado para predicción, el coeficiente de cada variable, su significación, y su diferencia están en segundo plano. Los modelos SSF, a pesar de todo, pueden ser los más adecuados en estudios inferenciales puramente ecológicos. En estos, el objetivo es entender qué ambientes son seleccionados significativamente (utilizados más que la media). Pero en estudios de conservación, donde el interés se encuentra en mapear corredores de fauna o para diseñar reservas naturales (ver ejemplos en Thurfjell et al. 2014), los modelos de distribución de especies parecen ser más eficientes.

Además, la comparación que hice utilizó una cantidad de puntos mayor que la utilizada habitualmente en estudios de SSF, y tal vez hasta mayor que en algunos estudios de distribución de especies. Por lo tanto, las diferencias en nuestros modelos pueden estar sobrestimadas. A primera vista también se puede argumentar que esta grande cantidad de puntos causaría un sesgo. Sin embargo, no creo que esto cause una diferencia cualitativa, ya que para que una muestra grande cause un sesgo, es necesario que el método de muestreo esté sesgado. El único sesgo en mi muestreo es el causado por la SSF, que favorece puntos de *background* próximos a los puntos originales. Como esto es parte inherente de la SSF, cualquier diferencia causada por lo mismo es un indicativo de que SSF no debería ser utilizada.

También es importante resaltar que mis resultados fueron corroborados con el uso de validación cruzada. Por tanto, los resultados encontrados en los cinco estudios de jaguares no son garantía que SSF siempre será peor que MaxEnt. Cualquier sesgo que hay en nuestros datos está presente tanto en los conjuntos de entrenamiento como en el conjunto de prueba, incluido los sesgos que puedan facilitar una métrica en comparación a otra. A pesar de esta limitación, creo que la validación cruzada sea la mejor técnica para realizar una comparación justa, aunque sea de generalización limitada. En cualquier caso, más estudios pueden indicar si el patrón encontrado aquí es universal o no.

En conclusión, este estudio muestra que a veces la solución más simple es la mejor. Añadir datos sobre movimiento animal simplemente porque están disponibles puede llegar a dificultar los objetivos del estudio, especialmente si este requiere la aplicación de métodos estadísticos complejos y con muchos requisitos. La información sobre movimiento claramente tiene su importancia y su valor, especialmente cuando se utiliza para diseñar un área de disponibilidad en una *resource selection function*. A pesar de eso, la información de movimiento, como casi todo en la vida, debería de utilizarse solamente cuando sea necesaria.

Agradecimientos

Estoy muy agradecido a Zaida Ortega por la ayuda con la traducción del texto de portugués a español. Como autor de este estudio, he sido financiado por una beca de postdoctorado de la fundación de ciencia de Israel (número 259-19). Ese trabajo también fue financiado por el Center of Advanced Systems Understanding (CASUS), que es financiado por el ministerio de educación y pesquisa alema (BMBF) y por el ministerio de ciencia, cultura, y turismo de la Sajonia (SMWK) con recursos basados en un presupuesto aprobado por el parlamento del estado de Sajonia.

Referencias

- Ahmed, S.E., McInerny, G., O'Hara, K., Harper, R., Salido, L., Emmott, S., Joppa, L.N. 2015. Scientists and software - surveying the species distribution modelling community. *Diversity and Distributions* 21: 258-267.
- Benhamou, S. 2007. How many animals really do the Levy Walk? *Ecology* 88: 1962-1969.
- Benhamou, S. 2008. How many animals really do the Lévy walk? Reply. *Ecology* 89: 2351-2352.

- Bradie, J., Leung, B. 2017. A quantitative synthesis of the importance of variables used in MaxEnt species distribution models. *Journal of Biogeography* 44: 1344-1361.
- Breslow, N.E., Day, N.E. 1980. *Statistical Methods in Cancer Research Volume I: The Analysis of Case-Control Studies*. Lyon, Francia.
- Brown, J.S. 1988. Patch use as an indicator of habitat preference, predation risk, and competition. *Behavioral Ecology and Sociobiology* 22: 37-47.
- Buchhorn, M., Lesiv, M., Tsendbazar, N.E., Herold, M., Bertels, L., Smets, B. 2020. Copernicus Global Land Cover Layers—Collection 2. *Remote Sensing* 12: 1044.
- Clark, J.D., Laufenberg, J.S., Davidson, M., Murrow, J.L. 2015. Connectivity among subpopulations of Louisiana black bears as estimated by a step selection function. *Journal of Wildlife Management* 79: 1347-1360.
- Convertino, M., Muñoz-Carpena, R., Chu-Agor, M.L., Kiker, G.A., Linkov, I. 2014. Untangling drivers of species distributions: Global sensitivity and uncertainty analyses of MaxEnt. *Environmental Modelling and Software* 51: 296-309.
- Coulon, A., Morellet, N., Goulard, M., Cargnelutti, B., Angibault, J.M., Hewison, A.J.M. 2008. Inferring the effects of landscape structure on roe deer (*Capreolus capreolus*) movements using a step selection function. *Landscape Ecology* 23: 603-614.
- Craiu, R. V., Duchesne, T., Fortin, D., Baillargeon, S. 2011. Conditional logistic regression with longitudinal follow-up and individual-level random coefficients: A stable and efficient two-step estimation method. *Journal of Computational and Graphical Statistics* 20: 767-784.
- Cressman, R., Křivan, V. 2006. Migration Dynamics for the Ideal Free Distribution. *The American naturalist* 168: 384-397.
- Eliith, J., Leathwick, J.R. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics* 40: 677-697.
- Eliith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* 17: 43-57.
- Fick, S.E., Hijmans, R.J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37: 4302-4315.
- Fortin, D., Beyer, H.L., Boyce, M.S., Smith, D.W., Duchesne, T., Mao, J.S. 2005. Wolves influence elk movements: behavior shapes a trophic cascade in Yellowstone national park. *Ecology* 86: 1320-1330.
- Francis, R.A., Taylor, J.D., Dibble, E., Strickland, B., Petro, V.M., Easterwood, C., Wang, G. 2017. Restricted cross-scale habitat selection by American beavers. *Current Zoology* 63: 703-710.
- Fretwell, S.D., Lucas, H. 1969. On territorial behavior and other factors influencing habitat distribution in birds I. theoretical development. *Acta biotheoretica* 19: 16-36.
- Global Land Cover Facility 2022. GLCF: Landsat Global Inland Water. Earth Engine Data Catalog. https://developers.google.com/earth-engine/datasets/catalog/GLCF_GLS_WATER
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* 202: 18-27.
- Goulart, F.V.B., Cáceres, N.C., Graipel, M.E., Tortato, M.A., Ghizoni, I.R., Oliveira-Santos, L.G.R. 2009. Habitat selection by large mammals in a southern Brazilian Atlantic Forest. *Mammalian Biology* 74: 182-190.
- Hansson, L., Khamis, H.J. 2008. Matched samples logistic regression in case-control studies with missing values: When to break the matches. *Statistical Methods in Medical Research* 17: 595-607.
- Hijmans, R.J. 2021. terra: Spatial Data Analysis. <https://cran.r-project.org/web/packages/terra/index.html>
- Hildén, O. 1965. Habitat selection in birds: A review. *Annales Zoologici Fennici* 2: 53-75.
- Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E. 2008. *Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database*: <https://srtm.csi.cgiar.org>.
- Lobo, J.M., Jiménez-Valverde, A., Real, R. 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* 17: 145-151.
- Manly, B.F.J. 1974. A Model for Certain Types of Selection Experiments. *Biometrics* 30: 281.
- Manly, B.F.J., McDonald, L.L., Thomas, D.L., McDonald, T.L., Erickson, W.P. 2002. *Resource Selection by Animals: Statistical Analysis and Design for Field Studies*. Kluwer Academic Publishers, Nueva York, Estados Unidos.
- McCullagh, P., Nelder, J.A. 1983. *Generalized Linear Models*. 2nd ed. Chapman & Hall, Londres, Reino Unido.
- Merow, C., Smith, M.J., Edwards, T.C., Guisan, A., McMahon, S.M., Normand, S., Thuiller, W., et al. 2014. What do we gain from simplicity versus complexity in species distribution models? *Ecography* 37: 1267-1281.
- Morato, R.G., Thompson, J.J., Paviolo, A., de La Torre, J.A., Lima, F., McBride, R.T., Paula, R.C., et al. 2018. Jaguar movement database: a GPS-based movement dataset of an apex predator in the Neotropics. *Ecology* 99: 1691.
- Morris, D.W. 1987. Tests of Density-Dependent Habitat Selection in a Patchy Environment. *Ecological Monographs* 57: 269-281.
- Morris, D.W. 2003. Toward an ecological synthesis: a case for habitat selection. *Oecologia* 136: 1-13.
- Muff, S., Signer, J., Fieberg, J. 2020. Accounting for individual-specific variation in habitat-selection studies: Efficient estimation of mixed-effects models using Bayesian or frequentist computation. *Journal of Animal Ecology* 89: 80-92.
- Phillips, S. 2021. maxnet: Fitting «Maxent» Species Distribution Models with «glmnet». <https://cran.r-project.org/web/packages/maxnet/maxnet.pdf>
- Phillips, S., Anderson, R., Schapire, R. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231-259.
- Phillips, S.J., Anderson, R.P., Dudík, M., Schapire, R.E., Blair, M.E. 2017. Opening the black box: an open-source release of Maxent. *Ecography* 40: 887-893.
- Pike, M.C., Hill, A.P., Smith, A.P.G. 1980. Bias and efficiency in logistic analyses of stratified case-control studies. *International Journal of Epidemiology* 9: 89-95.
- Plank, M.J., James, A. 2008. Optimal foraging: Lévy pattern or process? *Journal of the Royal Society Interface* 5: 1077-1086.
- Potts, J.R., Mokross, K., Stouffer, P.C., Lewis, M.A. 2014. Step selection techniques uncover the environmental predictors of space use patterns in flocks of Amazonian birds. *Ecology and Evolution* 4: 4578-4588.
- R Core Team. 2019. R: A language and environment for statistical computing. Viena, Austria. <https://www.R-project.org/>
- Radosavljevic, A., Anderson, R.P. 2014. Making better Maxent models of species distributions: Complexity, overfitting and evaluation. *Journal of Biogeography* 41: 629-643.
- Roeber, C.L., Boyce, M.S., Stenhouse, G.B. 2010. Grizzly bear movements relative to roads: Application of step selection functions. *Ecography* 33: 1113-1122.
- SEDAC-CIESIN-Columbia University 2016. *Gridded Population of the World, Version 4 (GPWv4): Population Density*. Disponible en: <https://sedac.ciesin.columbia.edu/data/collection/gpw-v4>
- Sheldon, A.L., Haick, R.A. 1981. Habitat selection and association of stream insects: a multivariate analysis. *Freshwater Biology* 11: 395-403.
- Silva, I., Fleming, C.H., Noonan, M.J., Alston, J., Foltá, C., Fagan, W.F., Calabrese, J.M. 2022. Autocorrelation-informed home range estimation: A review and practical guide. *Methods in Ecology and Evolution* 13: 534-544.
- Sjölander, A., Greenland, S. 2013. Ignoring the matching variables in cohort studies - When is it valid and why? *Statistics in Medicine* 32: 4696-4708.
- Svavarsson, J., Olafsdottir, S.H. 1999. Foraminiferan (Protozoa) epizoots on the anthuridean (Crustacea, Isopoda) calathura brachiata (Stimpson). *Journal of Natural History* 33: 1423-1432.
- Taher, T.M., Lihan, T., Mustapha, M.A., Nor, S.M. 2018. Habitat selection of *Tragulus napu* and *Tragulus javanicus* using MaxEnt analysis. *AIP Conference Proceedings* 1940: .
- Thurfjell, H., Ciuti, S., Boyce, M.S. 2014. Applications of step-selection functions in ecology and conservation. *Movement Ecology* 2: 26.
- Tregenza, T. 1995. Building on the ideal free distribution. *Advances in ecological research* 26: 253-302.
- Viswanathan, G.M., Afanasyev, V., Buldyrev, S. V., Murphy, E.J., Prince, P.A., Stanley, H.E. 1996. Lévy flight search patterns of wandering albatrosses. *Nature* 381: 413-415.
- Viswanathan, G.M., Buldyrev, S. V., Havlin, S., da Luz, M.G.E., Raposo, E.P., Stanley, H.E. 1999. Optimizing the success of random searches. *Nature* 401: 911-914.