

MODELOS DE NICHOS Y DISTRIBUCIÓN DE ESPECIES

INTRODUCCIÓN

En los últimos años, la generalización en el uso de los Sistemas de Información Geográfica (GIS, por sus siglas en inglés) junto con el desarrollo de varias técnicas estadísticas, matemáticas e informáticas ha permitido un notable incremento en las herramientas disponibles para el análisis de los patrones de distribución de las especies. Los modelos de nicho ecológico (ENM, por *Ecological Niche Modelling*) consisten en modelar las condiciones ambientales que reflejen el nicho fundamental existente (NE) o el nicho realizado (NR) de una especie, a través de una representación matemática de las condiciones presentes en su rango de distribución. Los modelos de distribución de especies (SDM, por *Species Distribution Modelling*) consisten en proyectar el nicho modelado al espacio geográfico, para describir la distribución actual de una especie o predecir su distribución potencial. Estos modelos son representaciones cartográficas de la idoneidad (*suitability*) de un área espacial para la presencia de una especie. La idoneidad es una relación matemática entre la distribución real conocida y un conjunto de variables independientes que se usan como indicadoras.

El uso de los modelos de nicho ecológico y de distribución de especies se ha incrementado notablemente durante los últimos años, entre otras cosas por sus diversas aplicaciones, tales como estudios de la distribución de especies (actual y pasada), análisis de riesgo asociado a especies invasoras, estudios de posibles efectos del cambio climático y estudios de conservación de la biodiversidad (especies y/o hábitats amenazados, diseño y localización de reservas, reintroducción de especies amenazadas). A pesar de sus múltiples aplicaciones, estos modelos también tienen una serie de limitaciones, como por ejemplo que se asume que la distribución es estable en el tiempo y el espacio y que rara vez se tienen en cuenta las interacciones bióticas con otros organismos.

Otras de las ventajas de estos modelos es la gran diversidad de fuentes a partir de las cuales obtener los datos necesarios para generarlos. La construcción de estos modelos se realiza en una serie de pasos (Figura 1). En primer lugar, se reúnen los datos de entrada que son los registros conocidos de la presencia de una especie y una serie de variables que describen las condiciones ambientales de esos sitios y que posean una relación potencial con la distribución de la especie. Los registros de ocurrencia de la especie pueden obtenerse de muestreos a campo, publicaciones científicas, colecciones de museo y bases de datos de biodiversidad tales como GBIF (<https://www.gbif.org>). Estos datos son en su mayoría datos de sólo presencia, aunque también pueden usarse datos de presencia/ausencia o de abundancia. Las variables ambientales utilizadas comúnmente en los modelos son capas climáticas, edáficas, hidrológicas y topográficas, disponibles

libremente en numerosos sitios de internet (e.g.: <http://www.worldclim.org>¹), generalmente en formato ráster. Los archivos ráster son matrices conformadas por celdas o píxeles, cada uno de los cuales tiene un determinado valor correspondiente a cada área del terreno (e.g. valores de temperatura, precipitación o tipo de suelo). Luego, la información de los registros de ocurrencia y las variables ambientales se asocia matemática o estadísticamente mediante algoritmos de modelado específicos, generando una representación del nicho de la especie. Este resultado se proyecta nuevamente al espacio geográfico, generando mapas de una determinada área con valores que reflejan la similitud ambiental de cada punto del terreno con las zonas de presencia registrada de la especie, o la idoneidad (*suitability*) para la distribución de la especie. Una parte de los datos de ocurrencia de la especie pueden separarse y reservarse para la posterior evaluación del poder predictivo del modelo, aunque es preferible contar con una serie independiente de datos de ocurrencia para dicho fin.

El objetivo de este minicurso es introducir los principales conceptos de estos modelos y familiarizarse con el manejo de los datos y programas que se utilizan más comúnmente para su realización. Por último, analizar las posibles aplicaciones que tienen los modelos de nicho y distribución de moluscos, con especial énfasis en las especies continentales.

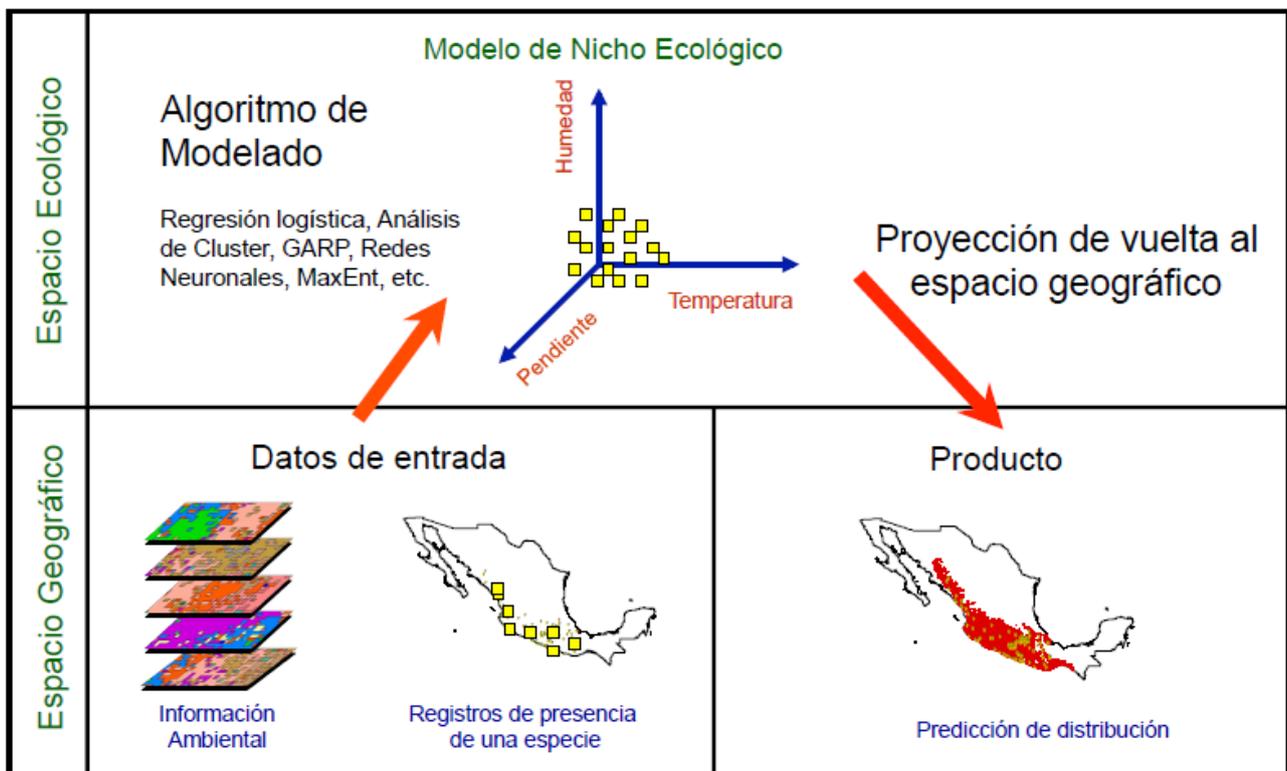


Figura 1. Proceso de modelado de nichos ecológicos y distribución de especies (de E. Martínez-Meyer).

¹ Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution-interpolated climate surfaces for global land areas. *International Journal of Climatology* 25:1965–1978.

ACTIVIDADES PRÁCTICAS

Existen diversas técnicas y algoritmos de modelación, aunque en este práctico veremos sólo dos:

- un algoritmo tipo "envoltura ambiental" muy fácil de interpretar, denominado BIOCLIM.
- un algoritmo de tipo "*machine learning*" que es uno de los más utilizado actualmente, denominado MaxEnt².

Vamos a crear modelos con datos de dos especies distintas:

- *Limnoperna fortunei*, el mejillón dorado, nativo de China pero que actualmente se encuentra en Argentina, Uruguay, Paraguay y Brasil.
- Una especie de molusco a elección. Los datos de ocurrencia que usaremos serán los que traiga cada uno o sino los buscaremos en alguna base de datos, como GBIF. Las capas ambientales también se pueden descargar o bien utilizar algunas de las provistas para el ejercicio de *L. fortunei*.

Limnoperna fortunei

Ejercicio 1:

En este ejercicio utilizaremos el algoritmo BIOCLIM, disponible en el programa DIVA-GIS (<http://www.diva-gis.org/download> para descargarlo).

- 1) Dentro de la carpeta BIOCLIM se encuentran las siguientes carpetas:
 - a. Localidades: contiene un archivo de tipo *shapefile* denominado "*limnoperna.shp*", que contiene las coordenadas geográficas de registros del mejillón dorado en Sudamérica.
 - b. Variables ambientales: contiene una serie de archivos de tipo *raster* en formato *grid* ("*.grd*, *.gri*") que es el formato que maneja DIVA; cada uno describe una variable ambiental distinta.
- 2) Para agregar los registros del mejillón, simplemente seleccionamos el *shapefile* desde el botón que es un signo "+".
- 3) Antes de realizar el modelo con BIOCLIM, debemos "apilar" las variables ambientales mediante la función *stack*, disponible dentro de la opción *Stack* → *Make Stack*. Para poder hacer un *stack* todos los archivos *raster* deben tener la misma resolución (*i.e.* el mismo tamaño de pixel) y la misma extensión o *extent* (*i.e.* el mismo número total de píxeles de largo y de ancho).
- 4) Agregar todas las variables ambientales con el botón *Add Grid* y crear un archivo con el nombre "*limnoperna.grs*" en la carpeta BIOCLIM (este archivo será el *stack* que contiene todas las capas ambientales).

² Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190:231-259.

- 5) Con la capa de registros del mejillón seleccionada, ir a *Modelling* → *Bioclim* y seleccionar como variables ambientales al *stack* "limnoperna.grs". En la pestaña *Predict* seleccionaremos como *output* la opción *Bioclim*, asignándole al archivo de salida el nombre "bioclim.grd", y también la opción *Bioclim True/False*, asignándole al archivo de salida el nombre "binario.grd". Para obtener un mapa con la extensión de las capas ambientales se debe seleccionar *Maximum Extent* entre las opciones de *Output Grid Dimensions*.
- 6) Ambos modelos generados aparecerán en la vista principal de DIVA.
 - a. ¿Qué diferencia hay entre los dos modelos generados?
 - b. ¿Coinciden los sitios de ocurrencia con las regiones predichas como idóneas? ¿Hay zonas idóneas no habitadas?
 - c. Realizar varias corridas utilizando distinta cantidad de variables ambientales en cada una y comparar los resultados obtenidos.

Ejercicio 2:

En este ejercicio utilizaremos el algoritmo MaxEnt, para el cual utilizaremos el programa homónimo. A continuación, se detallan los pasos a seguir para la descarga y uso del programa.

- 1) Para descargar MaxEnt se debe ingresar al siguiente link: https://biodiversityinformatics.amnh.org/open_source/maxent/. Al solicitar la descarga se obtiene una carpeta denominada "maxent.zip", la cual contiene el programa.
- 2) Dentro de la carpeta MaxEnt se encuentran las siguientes carpetas:
 - a. Localidades: contiene dos archivos denominados "limnoperna_nativo.csv" y "limnoperna_sudamerica.csv"; MaxEnt utiliza archivos de tipo ".csv" (*comma separated values*) para especificar los sitios de ocurrencia. Abrir un archivo (se puede abrir con Excel o con bloc de notas) y observar que en la primera columna se especifica el nombre de la especie (en este caso hay sólo una, pero podría haber más), en la segunda figura la longitud y en la tercera la latitud. Éstas últimas deben ingresarse en grados con sus decimales y con signo negativo para los hemisferios Sur y Occidental y con signo positivo para los hemisferios Norte y Oriental.
 - b. Rango Nativo: contiene una serie de archivos de tipo ráster en formato ASCII (" .asc") que es el formato que utiliza MaxEnt; cada uno describe una variable ambiental distinta (detallada en el nombre de cada archivo). Todos los archivos de esta carpeta deben tener la misma resolución y la misma extensión o *extent*, que en este caso corresponde al rango nativo del mejillón dorado en el sudeste asiático.
 - c. Output: esta carpeta está vacía, pero debemos seleccionarla cuando corramos el programa para indicar dónde deben guardarse todos los archivos de salida generados. Vamos

a hacer tres corridas, por lo cual si se quiere guardar cada una en carpetas separadas se deben crear otras dos carpetas (por ejemplo, Output potencial y Output calentamiento) y seleccionarlas cuando se realicen los ejercicios 7 y 8. De lo contrario, MaxEnt preguntará si debe rehacer el análisis y reescribirá el contenido de esta carpeta.

d. Sudamérica: contiene las mismas capas ambientales de la carpeta Rango Nativo, pero su extensión comprende toda Sudamérica. Usaremos estos archivos para proyectar el modelo generado en el rango nativo del mejillón dorado a otro rango geográfico, y obtener así su distribución potencial en Sudamérica. Las variables utilizadas para realizar una proyección, ya sea en el espacio o en el tiempo, deben llamarse exactamente igual que las usadas para generar el modelo.

e. Rango Invadido: contiene las mismas capas ambientales, pero en este caso corresponden al rango invadido del mejillón dorado en Sudamérica. Para recortar estas capas se utilizó como máscara un *shapfile* (" .shp") que comprende los límites de las cuencas en las que actualmente se encuentra *Limnoperna fortunei*. Usaremos estos archivos para crear un modelo en el rango invadido de Sudamérica.

f. Calentamiento: contiene variables ambientales recortadas a la misma extensión que las de la carpeta Rango Invadido y proyectadas al año 2080, considerando el peor escenario de emisiones de gases de efecto invernadero (RCP 8.5). Estas variables fueron obtenidas de uno de los modelos climáticos globales disponibles (HadGEM2-ES). Usaremos estos archivos para proyectar en el tiempo el modelo generado en el rango invadido.

3) Para abrir el programa simplemente hay que hacer clic sobre el archivo "maxent.bat". En el recuadro *Samples* deben indicarse los puntos de ocurrencia, seleccionando con el botón *Browse* el archivo "limnoperna_nativo.csv" de la carpeta Localidades. En el recuadro *Environmental Layers* se deben indicar las variables ambientales; en este caso seleccionamos toda la carpeta Rango Nativo y deberían aparecer todas listadas en el recuadro. En este paso pueden seleccionarse las variables a utilizar y se debe indicar si son categóricas o continuas; nosotros tildaremos todas y marcaremos todas como continuas. También debemos seleccionar la carpeta Output en el recuadro *Output directory*.

4) Ingresando en la opción *Settings*, le indicaremos al programa que seleccione al azar un 75% de los datos de ocurrencia para generar el modelo y el 25% restante para testearlo, ingresando 25 en *Random test percentage*.

5) Seleccionaremos las opciones *Create response curves* y *Do jackknife test to measure variable importance*, que nos ayudarán a interpretar cuáles de las variables ambientales utilizadas contribuyeron en mayor medida al modelo. Las curvas de respuesta muestran cómo la idoneidad ambiental predicha por el modelo cambia a medida que varía cada variable,

manteniendo todas las demás variables en su valor promedio. El test *Jackknife* considera las variables que más aumentan la "ganancia" (*gain*) del modelo cuando son incluidas de manera individual y las que la reducen en mayor medida cuando son excluidas.

6) Finalmente presionamos el botón *Run* y esperamos unos minutos. Todas las salidas del programa se encontrarán en la carpeta *Output*; el archivo "*limnoperna_fortunei.html*" muestra los principales resultados del modelo y puede abrirse con cualquier navegador.

- a. ¿Coinciden los sitios de ocurrencia con las regiones de mayor *suitability*? ¿Hay zonas habitables no habitadas?
- b. ¿Por qué hay puntos blancos y violetas en el mapa?
- c. ¿Cuáles son las variables más relevantes para la distribución de esta especie?

7) A continuación, realizaremos una proyección a Sudamérica del modelo generado en el rango nativo del mejillón dorado. *MoxEnt* proyecta el modelo generado con las variables originales en un set diferente de variables ambientales, en este caso un rango geográfico diferente. Para hacer esta proyección volveremos a correr el programa con la misma configuración, pero seleccionaremos la carpeta *Sudamérica* en el recuadro *Projection layers directory/file*, también seleccionaremos la opción *Make pictures of predictions*.

- a. ¿Qué regiones de Sudamérica son potencialmente habitables según el modelo?
- b. Utilizando *QGIS* o *DIVA-GIS*, abrir el archivo "*Limnoperna_fortunei.asc*" generado por *MoxEnt* y también el archivo "*limnoperna_sudamerica.csv*", ubicado en la carpeta *Localidades*. Analizar si los sitios de ocurrencia en Sudamérica coinciden con los predichos por la distribución potencial generada. ¿Hay otras regiones "en riesgo" de ser invadidas según el modelo?

8) Por último, realizaremos un modelo utilizando los registros del rango invadido en Sudamérica y lo proyectaremos a un escenario futuro (2080) previendo variaciones en las condiciones climáticas. Para hacer esta proyección volveremos a correr el programa seleccionando en *Samples* el archivo "*limnoperna_sudamerica.csv*", la carpeta *Rango Invadido* en *Environmental Layers*, y la carpeta *Calentamiento* en el recuadro *Projection layers directory/file*, también seleccionaremos la opción *Make pictures of predictions*.

- a. ¿Qué sucederá, según el modelo, con la distribución del mejillón dorado en Sudamérica si se incrementa la temperatura en el área por el calentamiento global?