

Informe final* del Proyecto JM055
Modelos de distribución de especies de mamíferos y suculentas de la Faja Volcánica Transmexicana*

Responsable: Dra. Tania Escalante Espinosa
Institución: Universidad Nacional Autónoma de México
Facultad de Ciencias
Correo electrónico: tee@st.ib.unam.mx
Fecha de inicio: 15-Jun-2012
Fecha de término: 20-May-2015
Principales resultados: Informe final, Base de datos, Cartografía
Forma de citar el informe final y otros resultados:** Escalante, T. 2015. Modelos de distribución de especies de mamíferos y suculentas de la Faja Volcánica Transmexicana. Universidad Nacional Autónoma de México. Facultad de Ciencias. **Informe final SNIB-CONABIO, proyecto No. JM055.** México D. F.

Resumen:

La biodiversidad de la provincia de la Faja Volcánica Transmexicana exhibe patrones de endemismo que representan diferentes niveles de homología espacial, y los modelos de distribución de especies son la herramienta fundamental para el análisis descriptivo e integral de estos patrones. Los mamíferos y algunos grupos de suculentas tradicionalmente han sido utilizados para describir el endemismo, sin embargo, existe hoy en día una importante controversia sobre la naturalidad de la Faja Volcánica Transmexicana. Los datos disponibles sobre los mamíferos y las suculentas, además del amplio conocimiento que se tiene sobre sus historias de vida, los hacen taxones ideales para los estudios biogeográficos. Adicionalmente, el endemismo generalmente se encuentra asociado a algún grado de riesgo de las poblaciones de las especies por su misma restricción geográfica, por lo que los taxones endémicos se encuentran más vulnerables a efectos antropomórficos que los pueden llevar a situarse en categorías de riesgo como la NOM-059 o la lista roja de la IUCN.

En este proyecto nos proponemos modelar las distribuciones potenciales de 24 especies de plantas y 18 de mamíferos que en trabajos anteriores hemos identificado como endémicas de la Faja Volcánica Transmexicana, con miras a su aplicación en estrategias de conservación más eficientes. De las especies propuestas, 11 se encuentran categorizadas en la NOM-059-ECOL-2010.

Se propone generar modelos para esas especies a partir de bases de datos que hemos compilado y con el algoritmo MaxEnt, además de realizar una exhaustiva revisión bibliográfica de la historia de vida de los taxones para la selección adecuada de las variables a considerar en el esquema de modelado. Como resultados se propone generar modelos fiables de distribución de estas especies que puedan ser aplicados en biogeografía de la conservación.

-
- * El presente documento no necesariamente contiene los principales resultados del proyecto correspondiente o la descripción de los mismos. Los proyectos apoyados por la CONABIO así como información adicional sobre ellos, pueden consultarse en www.conabio.gob.mx
 - ** El usuario tiene la obligación, de conformidad con el artículo 57 de la LFDA, de citar a los autores de obras individuales, así como a los compiladores. De manera que deberán citarse todos los responsables de los proyectos, que proveyeron datos, así como a la CONABIO como depositaria, compiladora y proveedora de la información. En su caso, el usuario deberá obtener del proveedor la información complementaria sobre la autoría específica de los datos.



**MUSEO DE ZOOLOGÍA "ALFONSO L. HERRERA"
DEPARTAMENTO DE BIOLOGÍA EVOLUTIVA
FACULTAD DE CIENCIAS
UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
Apartado Postal 70-399, 04510 México, D.F., México
Tel. (52 55) 56 22 48 25**

INDICE GENERAL

I. RESUMEN EJECUTIVO

II. INTRODUCCIÓN

III. DESARROLLO DEL PROYECTO

III.1.- Base de datos y calidad de la información

III.2.- Criterio de Selección de región de referencia (*landscape* o M)

III.3.- Variables utilizadas en modelado

III.4.- Método de modelado

III.5. Parámetros utilizados en el modelado

III.6.- Evaluación del modelo

IV. CONCLUSIONES Y RECOMENDACIONES

V. REFERENCIAS

VI. ANEXO Resultados por especie: calidad de información y perfil ambiental

INFORME FINAL

PROYECTO JM055

MODELOS DE DISTRIBUCIÓN DE ESPECIES DE MAMÍFEROS Y SUCULENTAS DE LA FAJA VOLCÁNICA TRANSMEXICANA

DATOS GENERALES

Responsable: Dra. Tania Escalante Espinosa

Corresponsable: C. Dra. Niza Gámez Tamaríz

Participantes:

Biól. José Gerardo Rodríguez Tapia

Biól. Alberto Romo Galicia

Biól. José Ernesto Moreno Zúñiga

Biól. Isabel Adali Celis González

Instituciones participantes:

Facultad de Ciencias, Universidad Nacional Autónoma de México

Instituto de Ecología, Universidad Nacional Autónoma de México

Correo electrónico del responsable: tee@st.ib.unam.mx

I. RESUMEN EJECUTIVO

El objetivo de este proyecto fue "Generar los modelos de distribución potencial de tres especies de la familia Agavaceae, 21 especies de la familia Cactaceae y 18 especies de mamíferos, endémicos de la Faja Volcánica Transmexicana en México". Se comprometió la entrega de un modelo por especie, elaborados a partir de dos bases de datos (una para plantas y otra para mamíferos), las cuales contienen en total 7,093 registros. Se entregan 22 modelos de distribución para plantas y 17 para mamíferos (dos especies de plantas y una de mamíferos no fueron modeladas debido a un número insuficiente de registros para elaborar los modelos). Los objetivos se cumplieron totalmente.

II. INTRODUCCIÓN

La provincia de la Faja Volcánica Transmexicana (FVT) ha sido estudiada tradicionalmente como un área con una alta diversidad de taxones exclusivos. Sin embargo, su reconocimiento como unidad biogeográfica natural se encuentra en discusión, por lo que recientemente se ha propuesto que debe revisarse su naturalidad. En los últimos años se han publicado algunos trabajos encaminados a la mejor comprensión de los patrones y procesos que operan en esta provincia, en particular los patrones de homopatría (endemismo). Sin embargo, se requiere el estudio de los patrones de distribución mediante la integración de las aproximaciones biogeográficas evolutiva y ecológica, por ello, los modelos de distribución de especies (MDS) son una herramienta esencial que permite la comprensión de los factores ambientales que permiten la sobrevivencia de los taxones en un área dada.

La FVT es un conjunto de cordilleras de volcanes alineados sobre una franja que cruza el territorio mexicano de oeste a este, desde Cabo Corrientes, Nayarit hasta la Sierra Chinconquiaco, Veracruz, que se localiza principalmente entre los 19 y 21° N y entre los 1500 y 3000 m de altitud. La faja de volcanes que conforman la FVT se ha formado durante los últimos 19 millones de años, a lo largo de cuatro grandes episodios de vulcanismo que afectaron de manera asincrónica el territorio, primero en el oeste y tardíamente en el este, sin embargo, los más prominentes se produjeron en los últimos tres millones de años. Esta asincronía geológica se ve igualmente reflejada en la distribución no homogénea de su riqueza y endemismos. Pese a que la FVT se reconoce como una unidad biogeográfica, su historia geológica y climática, y la distribución de sus especies, la hacen una de las provincias biogeográficas más heterogéneas y complejas del país. Dicha complejidad ha sido identificada por diversos autores, quienes han detectado porciones de su territorio con relaciones múltiples con otras provincias, elevado número de nodos biogeográficos, así como inconsistencias en los criterios de delimitación y límites geográficos.

Las especies que se incluyeron en este proyecto fueron: **Agavaceae:** *Furcraea martinezii*, *Furcraea guerrerensis*, y *Furcraea parmentieri*; **Cactaceae:** *Coryphantha ottonis*, *Coryphantha pycnantha*, *Echinocactus grusonii*, *Pachycereus grandis*, *Stenocactus obvallatus*, *Stenocactus wippermannii*, *Stenocereus beneckeii*, *Stenocereus*

queretaroensis, *Stenocereus stellatus*, *Opuntia atropes*, *Opuntia cantabrigiensis*, *Opuntia depressa*, *Opuntia fuliginosa*, *Opuntia joconostle*, *Opuntia pilifera*, *Mammillaria crinita*, *Mammillaria discolor*, *Mammillaria elongata*, *Mammillaria geminispina*, *Mammillaria haageana* y *Mammillaria nunezii*; **Mammalia:** *Cratogeomys gymnurus*, *Cratogeomys merriami* , *Cratogeomys tylorhinus*, *Liomys spectabilis*, *Megadontomys nelsoni*, *Microtus oaxacensis*, *Microtus quasiater*, *Nelsonia goldmani*, *Neotomodon alstoni*, *Peromyscus bullatus*, *Reithrodontomys chrysopsis*, *Reithrodontomys hirsutus*, *Reithrodontomys microdon*, *Romerolagus diazi*, *Sorex macrodon*, *Sorex ventralis*, *Spermophilus perotensis* y *Zygozomys trichopus*.

III. DESARROLLO DEL PROYECTO

III.1.- Base de datos y calidad de la información

Se revisó cada uno de los registros correspondientes a las distintas especies con la finalidad de identificar, corregir y clasificar registros con inconsistencias (nomenclaturales, de georreferenciación, redundantes, incompletos), etiquetando y corrigiendo aquellas inconsistencias que impliquen una fuente de incertidumbre para la base de datos.

Los datos finales son los siguientes:

Nombre de la entidad (tabla)	Número de campos*	No. de registros
BDmamíferos_JM055	38	4564
BDPlantas_JM055	38	2529

*Se entregan 32 campos obligatorios y seis campos adicionales (E_Coor, E_Geo1, E_Dupl, E_Geo2, E_Maha y Datos_Modelo), los cuales contienen información relacionada con la incertidumbre de los registros.

Verificación nomenclatural: todos los registros fueron verificados con respecto a sus sinonimias y nombres válidos de acuerdo con los catálogos de referencia mencionados a continuación:

Mamíferos:

- Ramírez-Pulido J., J. Arroyo-Cabrales y A. Castro-Campillo. 2005. Estado actual y relación nomenclatural de los mamíferos terrestres de México. *Acta Zoológica Mexicana (nueva serie)* 91: 21-82.

Plantas:

- García-Mendoza, A. 2000. Revisión taxonómica de las especies arborescentes de *Furcraea* (Agavaceae) en México y Guatemala. *Boletín de la Sociedad Botánica de México* 66: 113-129.
- Guzmán, U., S. Arias y P. Dávila. 2003. *Catálogo de cactáceas mexicanas*. CONABIO, UNAM. México, D.F.

Depuración de datos: Una vez finalizado el proceso de verificación nomenclatural, se revisó cada uno de los registros correspondientes a las distintas especies con la finalidad de identificar, corregir y clasificar registros con inconsistencias (de georreferenciación,

redundantes, incompletos), dado que aumentan el grado de incertidumbre para la base de datos. De acuerdo con la fuente de incertidumbre asociada al registro, estos fueron categorizados de la siguiente forma (información contenida en los seis campos adicionales de cada registro):

- E_Coor.- registro sin georreferenciación.
- E_Dupl.- conjunto de registros del taxón asociados a un mismo pixel, por ende, considerados como redundantes para el modelo.
- E_Geo1.- registro con error de georreferencia tipo 1, fuera de límites geopolíticos de México.
- E_Geo2.- registro con error de georreferencia tipo 2, georreferencia fuera del área de distribución reportada para el taxón en información especializada.
- E_Maha.- registro con distancia de Mahalanobis mayor a dos desviaciones estándar, ubicación geográfica marginal o dudosa e información de la fecha/descripción de localidad/colecta/identificación de baja calidad).
- Datos_Modelo.- registro utilizado para la generación del modelo/validación.

Asignación de categoría a los registros

E_Coor: Con el fin de clasificar los registros de la categoría E_Coor, se identificaron aquellos que pese a tener información acerca de la localidad de recolecta no cuentan con datos de georreferenciación, por lo cual los campos decimalLongitude y decimalLatitude fueron codificados con el valor no disponible y asignados a la categoría E-Coor.

E_Dupl: Los datos de ocurrencia pueden presentar sesgo espacial cuando el taxón sólo se ha muestreado en un subconjunto no aleatoria de los sitios que pertenecen al área total donde ocurre, por ejemplo, en zonas de fácil acceso. Dicho sesgo en el muestreo afecta la calibración del modelo (Phillips *et al.*, 2009). Por otro lado, registros autocorrelacionados espacialmente debido a esfuerzo de muestreo diferenciado, pueden dar lugar a imprecisiones en calibrado, predicción y evaluación

del modelo, incluso si no hay ningún otro sesgo (Veloz, 2009). Con el fin de disminuir la incertidumbre asociada a los registros por causa del sesgo en el muestreo y realizar un adecuada exploración y calibrado de los datos, los registros con georreferencia fueron intersectados con las celdas del *landscape* (celdas de 30 segundo de arco longitud X 30 segundos de arco latitud) y en caso de contar con más de un registro del taxón por celda, se seleccionó a uno de ellos y el resto fueron asignados a la categoría E_Dupl.

E_Geo1: Todos los registros con coordenadas fuera de límites geopolíticos de México fueron asignados a la categoría E_Geo1.

E_Geo2: Los registros que hasta este punto no se encontraron asignados a ninguna de las siguientes categorías y campos respectivos: E_Coor, E_Dupl y E_Geo1, fueron cartografiados con el fin de identificar a aquellos que, de acuerdo con información de fuentes especializadas y/o asesoría de los especialistas, se localizan fuera de los límites geográficos reportados para la distribución de las especies. Dichos registros fueron asignados a la categoría E_Geo2.

E_Maha: En el caso de identificar registros para los cuales la información existente acerca de su distribución geográfica no es suficiente para considerarlos como registros erróneos, pero sí fueron considerados fuente potencial de incertidumbre, ya que presentan una ubicación geográfica marginal o dudosa. Para discriminar si estos registros debían ser incluidos o no en el modelo, fue utilizado el criterio de la distancia de Mahalanobis. Bajo esta categoría fueron asignados aquellos registros con distribución marginal o dudosa que además presentan una distancia ambiental multivariada (Mahalanobis) mayor a dos desviaciones estándar. Las distancias de Mahalanobis correspondientes a cada una de las especies se realizó a partir de los registros sin errores (con E_Coor, E_Dupl, E_Geo1, E_Geo2), y la selección de variables derivada del análisis de componentes principales y correlación de Spearman realizado para cada especie (ver más adelante).

Datos_Modelo: Bajo esta categoría fueron asignados todos aquellos registros que hasta este punto no se encontraron asignados a ninguna de las siguientes categorías: E_Coor, E_Dupl, E_Geo1, E_Geo2; así como aquellos registros que

presentan una distribución marginal o dudosa, distancia de Mahalanobis mayor a dos desviaciones estándar (E_{Maha}), pero no representan aumento en la incertidumbre, es decir, registros colectados hace más de 20 años con información de la localidad exacta insuficiente y en algunos casos sin información acerca del colector/identificador del ejemplar (ver sección 7, Resultados por especie).

La base de datos incluye un campo para cada uno de estos errores, codificando como "TRUE" en caso de que el registro represente un error y "FALSE" en caso de no ser considerado error; así como los que fueron empleados para la formulación del modelo (Datos_Modelo).

III.2.- Criterio de Selección de región de referencia (*landscape* o M)

En el presente proyecto los modelos se generaron con MaxEnt 3.3.3k (Phillips *et al.*, 2006), el cual genera el modelo de respuesta en función del ajuste estadístico entre dos funciones de densidad, la que describe los datos de presencia y la del *landscape*, y es precisamente dicho ajuste el que determina las características del modelo de respuesta. Si el *landscape* tiene una extensión geográfica que excede por mucho la distribución del taxón, es probable que la función de densidad que lo describe ambientalmente difiera mucho de la función de densidad que describe a los datos de presencia del taxón, lo que a su vez condiciona a que la función de densidad del modelo de respuesta sea poco adecuada y describa de manera ineficiente el comportamiento de los datos de presencia.

Aunado a lo anterior, en el taller de modelado de nicho que se nos impartió se mencionó que si los registros de presencia se encuentran ubicados de manera marginal en la nube de datos que describen el ambiente ecológico del *landscape*, la calidad del modelo de respuesta se ve comprometida.

Partiendo del planteamiento anterior, se exploró definir la M en función de las provincias fisiográficas asociadas a los puntos de presencia; sin embargo se observó que, bajo este criterio se corre el riesgo de que los registros de presencia del taxón queden en las zonas marginales del espacio ambiental que representa la M. Aunado a lo anterior se observó que las especies seleccionadas en el presente trabajo describen patrones de

distribución que se ajustan distintas porciones geográficas, razón por la cual se prefirió definir distintas M que describieran dichas regiones a partir de polígonos regulares.

Se emplearon nueve ajustes de *landscape* o M: **Balsas-Tehuacán**, coordenadas extremas -106.0006393, -95.00107886, 15.99988013, 21.99964038, utilizada en el caso de *Cratogeomys merriami*, *Furcraea guerrerensis*, *Liomys spectabilis*, *Mammillaria nunezii*, *Nelsonia goldmani* y *Stenocereus beneckeii*; **Balsas2**, coordenadas extremas -105.0006793, -95.00107886, 14.99992008, 24.99952051, utilizada en el caso de *Opuntia atropes* y *O. fuliginosa*; **FVT**, coordenadas extremas -105.5006593, -95.00107886, 17.49982019, 22.4996204, utilizada en el caso de *Cratogeomys gymnurus*, *C. tylosinus*, *Fourcraea parmentieri*, *Microtus quasiater*, *Neotomodon alstoni*, *Reithrodontomys chrysopsis*, *Romerolagus diazi*, *Stenocactus obvallatus*, *Zygozomys trichopus*; **FVT2**, coordenadas extremas -105.5006593, -95.00107886, 18.49978023, 22.99960042, utilizada en el caso de *Coryphanta ottonis*, *Echinocactus grusonii*, *Mammillaria elongata*, *Stenocereus queretaroensis*; **Centro-México**, coordenadas extremas -105.0006793, -95.00107886, 17.99980021, 24.99952051, utilizada en el caso de *Mammillaria crinita*, *M. geminispina*, *O. cantabrigiensis*, *O. joconostle*; **Pacífico centro**, coordenadas extremas -105.5006593, -95.00107886, 18.49978023, 22.99960042, utilizada en el caso de *Reithrodontomys hirsutus*; **Pacífico sur**, -106,-89, 14, 22, utilizada en el caso de *Reithrodontomys microtus*; **Tehuacán**, coordenadas extremas -100.0008791; -94.00111882, 15.49990011, 20.99968034, utilizada en el caso de *Mammillaria haageana*, *Microtus oaxacensis*, *O. depressa*, *O. pilifera*, *Pachycereus grandis*, *Sorex ventralis*, *Spermophilus perotensis*, *Stenocereus stellatus*; **Tehuacán2**, coordenadas extremas -101.0008391, -95.00107886, 15.49990011, 21.99964038, utilizada en el caso de *Coryphantha pycnantha*, *M. discolor*, *Peromyscus bullatus*, *Sorex macrodon*.

III.3.- Variables utilizadas en modelado

De acuerdo con diversos autores (Araújo y Guisan, 2006; Guisan *et al.*, 2006; Václavík *et al.*, 2011) una de las principales problemáticas ligada a la generación de modelos es la autocorrelación de las variables ambientales. Las diferencias en la respuesta o caracterización disminuye cuando el modelo se construye con menos dimensiones

ambientales, a lo cual Peterson (2011) agrega que en los casos en los que no se realiza una selección detallada de las variables, las conclusiones a las que se llega acerca de los cambios -geográficos o ecológicos- no necesariamente son incorrectas, pero las pruebas que sustentan dichas conclusiones son poco robustas.

Estructurar el conjunto adecuado de variables no es una cuestión simple, ya que si caracterizamos a una especie en función de una única dimensión ambiental o variable, esta no es suficiente para sumar la variación ecológica que dicha especie exhibe en la naturaleza. Por otro lado, si el modelo es generado a partir de demasiadas variables, entonces el modelo únicamente describe cada punto de ocurrencia con el cual fue generado, es decir tiene un alto grado de sobreajuste y pierde poder predictivo (Peterson *et al.*, 2007).

Por otra parte, al momento de hacer réplicas de los modelos y como resultado de procesos de aleatorización de los modeladores, se obtienen variaciones en los resultados. Dichas diferencias en la respuesta o caracterización disminuye cuando el modelo se construye con menos dimensiones ambientales (Peterson, 2011). En este sentido, una adecuada selección de variables deberá exhibir la mayor carga de información posible para explicar la varianza de los datos, implicar el menor número de dimensiones y limitar el patrón de correlación entre estas (valores menores al 0.85% de acuerdo con Elith *et al.*, 2011).

Para seleccionar con cuáles variables modelar, lo ideal es que las decisiones sean tomadas a partir de información acerca de la biología de las especies (monitoreo en campo o laboratorio de la respuesta fisiológica de sobrevivencia/adecuación de los individuos a la variación de los factores ambientales). Sin embargo, en la mayoría de los casos no se cuenta con información especializada para tomar este tipo de decisiones, por lo que en la búsqueda de mitigar el sobreajuste y la autocorrelación entre variables y reducir el número de dimensiones de los modelos resultantes, se han aplicado distintas estrategias:

- (1) generación de un conjunto reducido de dimensiones independientes a través del Análisis de Componentes Principales (ACP), el cual constituye la técnica base para el análisis multivariado de datos con el objeto de reducir

la cantidad de variables, manteniendo la máxima cantidad de información presente en una tabla de datos de variables cuantitativas;

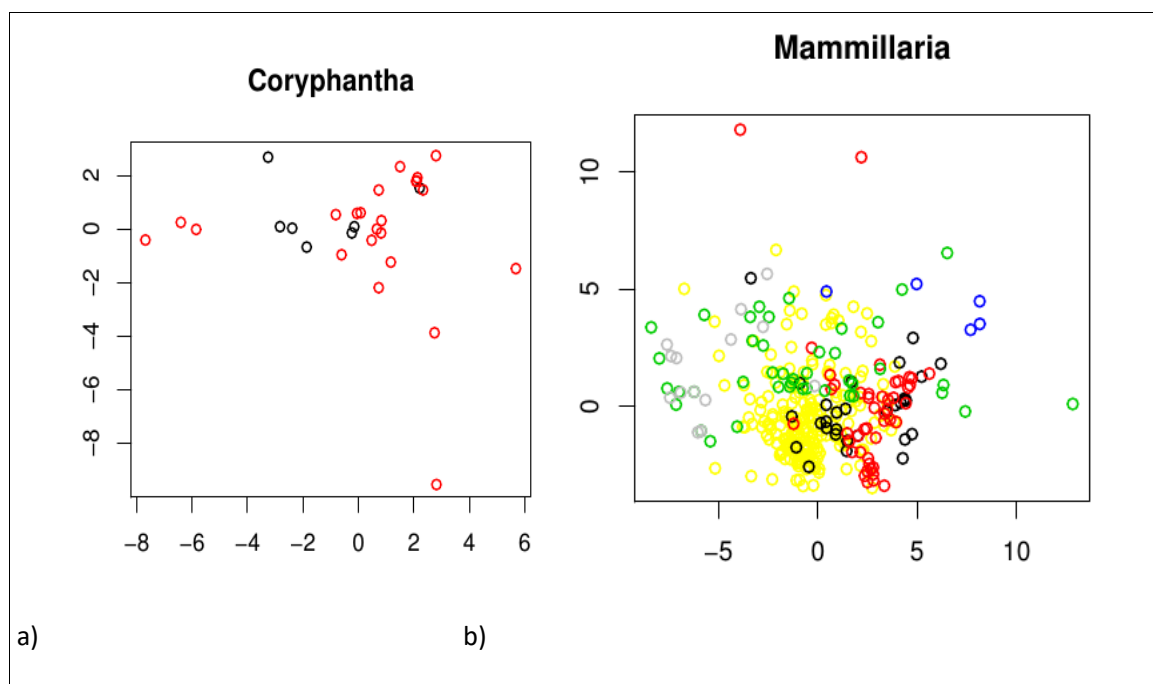
- (2) utilización de los resultados del ACP para la identificación de aquellas variables que más aporten a la distribución de la varianza que muestran los datos;
- (3) identificación y posterior eliminación de aquellas variables que presenten una alta correlación con otras variables, a través del análisis de las correlaciones entre las variables de acuerdo con los valores que presentan los registros de presencia o bien el espacio geográfico a reportado para las especies (*landscape*).

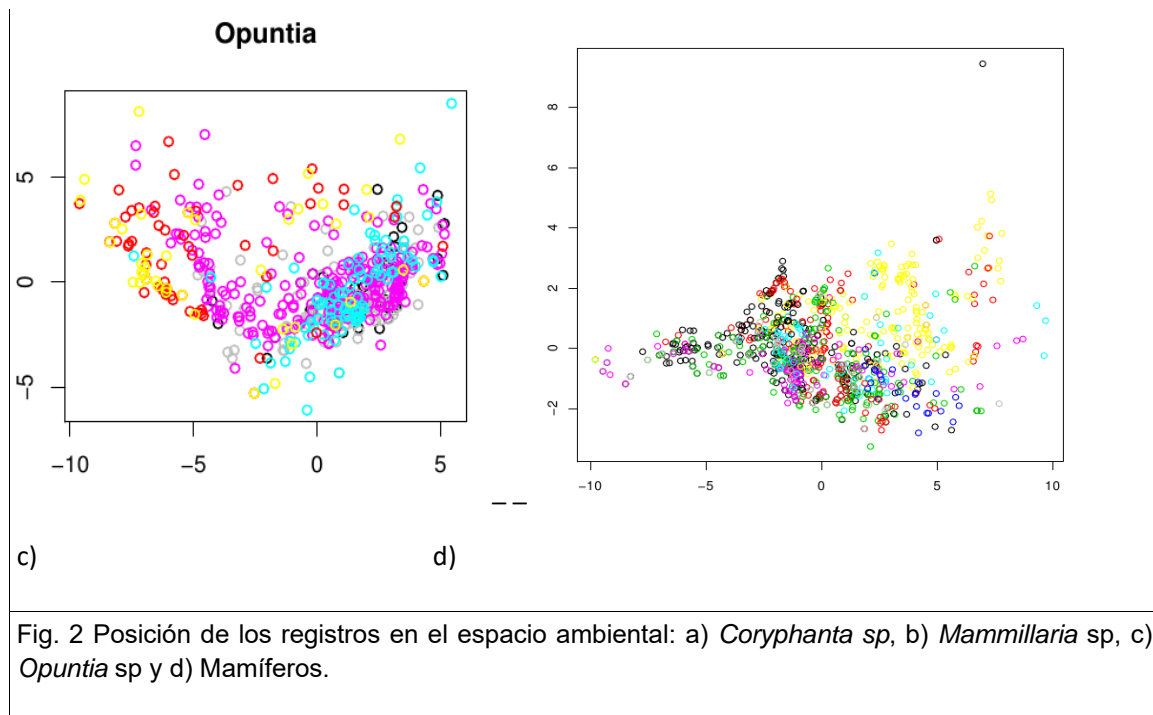
En el primer caso es difícil la interpretación de los componentes en términos biológicos, mientras que en los otros dos casos lo relevante es el grado de subjetividad, ya que puede darse el caso en que se supriman dimensiones o variables relevantes para la biología de las especies (Peterson, 2011), por lo que de tener que ser utilizada una estrategia preferimos el empleo de la segunda y tercera de manera integrada. Dicha estrategia permite que: mediante el ACP nos aproximemos al patrón de correlación que existe entre la presencia de las especie y las variables ambientales, es decir detectar aquellas variables relevantes para explicar la varianza total que muestran los datos, o bien, variables con valor de aporte a la varianza observada de los datos en extremo bajo; el análisis de correlación/ACP en conjunto faciliten detectar variables que, dado el alto grado de correlación con otras variables y su bajo aporte a la explicación de la varianza de los datos, sean elementos prescindibles al momento de formular un modelo de nicho con alto poder descriptivo/predictivo y lo más simple posible en términos de dimensiones.

Con base en el anterior argumento, para la gran mayoría de las especies se utilizó el análisis de correlación de Spearman en conjunto con el ACP para la selección de variables. En el caso de las especies pertenecientes a los géneros: *Coryphanta*, *Mammaillaria* y *Opuntia*, se observó que las especies que las integran describen una misma nube de puntos dentro del espacio ambiental (ver Fig. 2), por esa razón la selección de variables se realizó **en conjunto**; lo mismo ocurrió en el caso de las especies de mamíferos. En el caso de: *Echinocactus grusonii*, *Nelsonia goldmani*,

Mammillaria nunezii, *Reithrodontomys hirsutus* y *Sorex macrodon*, debido a que el número de registros fue insuficiente para ambos análisis, la selección de variables se hizo a partir del patrón de correlación de Spearman que mostraban las distintas variables en la *M/landscape* correspondiente.

La cartografía utilizada para todos los modelos correspondió a las 19 variables bioclimáticas de Tellez *et al.* (2011) y las variables topográficas de la USGS, ambas a una resolución de 30 segundos de arco.





Comportamiento ambiental de los registros: con el fin de describir el comportamiento geográfico/ambiental de los registros incluidos en el modelado de la distribución de los taxones, para cada especie se construyeron histogramas de los registros que muestran la distribución de los mismos respecto de los valores de cada variable; cartografía que muestra la distribución geográfica de los registros; mapeo de los registros en el espacio ambiental dado el ACP; y representaciones gráficas de las distancias ambientales (Mahalanobis), que muestran entre sí los datos (Anexo).

III.4.- Método de modelado

Todos los modelos fueron generados con el programa MaxEnt 3.3.3k (Phillips *et al.*, 2006). Este programa, es un modelador que trabaja bajo el principio de máxima entropía, y es implementado como un software que permite el modelado espacial de la distribución de taxones a partir de datos de presencia (Phillips *et al.*, 2006). Utiliza datos de

ocurrencias de especies y datos ambientales presentes y/o pasados, y genera modelos que describen la relación entre las variables ambientales y la presencia conocida de los taxones. Dichos modelos son proyectados en el espacio geográfico como modelos de distribución, lo cual a su vez permite obtener una estimación cuantitativa de la idoneidad de un área para el taxón modelado (Phillips *et al.*, 2006; Elith *et al.*, 2011). Elegimos Maxent, ya que se ha demostrado se uno de los más eficientes al momento de generar modelos de distribución de especies a partir de datos meramente geográficos (Elith *et al.*, 2006, 2011; Phillips *et al.*, 2006; Phillips y Dudik, 2008).

III.5. Parámetros utilizados en el modelado

Todos los modelos fueron realizados con el programa MaxEnt 3.3e (Phillips *et al.*, 2006), utilizando los siguientes criterios: Random Seed = on, Remove duplicate presence records = off; Random test percentage = 20, Regularization multiplier = 1 Replicates = 30, Replicate run type = Bootstrap, Extrapolate = off, Maximum iterations = 1000, y el resto pre-establecido. Los taxones con tamaño de muestra menor a 15 (*Coryphantha ottonis*, *Echinocactus grusonii*, *Nelsonia goldmani*, *Mammillaria nunezii* y *Reithrodontomys hirsutus*), el Random test percentage = 0 y Regularization multiplier = 1.2.

El parámetro “Regularization multiplier” controla el grado de ajuste de la función del modelo a la función que describen los datos de entrada. Cuando MaxEnt se corre con el criterio preestablecido, el modelo restringe el número de píxeles y el valor de probabilidad de estos, resultando un modelo con menor grado de sobrepredicción. Si este valor es aumentado, se le está indicando al modelador hacer un ajuste a los datos de presencia conocida más laxo.

Uno de los principales problemas con los que se enfrentan al momento de modelar la respuesta geográfica de los taxones, es el sesgo en el muestreo. Para el modelador el hecho de que un tipo de espacio ambiental domine en el comportamiento de los datos de presencia, es traducido en que ese tipo de ambiente es más adecuado para el establecimiento de las poblaciones; mientras que uno submuestreado, es interpretado

como un sitio con una magnitud de adecuación ambiental menor, en consecuencia la probabilidad de aquellos píxeles del modelo que sean próximos a este tipo de registros, será castigada.

Al observar la distribución de las especies con tamaños de muestra pequeños, sobre el espacio ambiental, se observa un alto grado de varianza y es difícil detectar un patrón. Este comportamiento pudiera estar relacionado en la mayoría de los casos a un alto grado de sesgo en el muestreo. A pesar de que en la gran mayoría de las especies el tamaño de muestra es adecuado, es altamente probable que en todos los casos se tengan altos grados de sesgo en el muestreo. Con el fin de disminuir el efecto del sesgo de muestreo sobre el rendimiento de los modelos, se decidió modificar el grado de ajuste de la función del modelo, respecto al predefinido por el programa.

III.6.- Evaluación del modelo

Todos los modelos se evaluaron mediante el método de la curva ROC parcial (Peterson *et al.*, 2008). A diferencia de la ROC, en la ROC parcial el eje de las abscisas no representa la proporción de errores de omisión, sino el área total predicha por el modelo como parte de los sitios adecuados para la presencia de la especie. Esta última, calculada en función del valor probabilístico umbral mínimo para el establecimiento no efímero de poblaciones de la especie. Dicho análisis fue realizado con el software Partial ROC (Barve, 2008). Para cada uno de los modelos se calculó el valor de AUC de la ROC Parcial, tomando como umbral de error el 10% de la muestra total empleada para modelar.

Para cada una de las especies, los modelos generados en MaxEnt fueron refinados con el umbral de probabilidad asociado al potencial error en la base de datos. Debido a que en todos los caso la base de datos fue construida a partir de consultas a otras bases de datos o bien ejemplares de colecciones científicas con datos de ejemplares colectados/observados en distintos tiempos, y con una calidad de georreferenciación y/o identificación no estandarizada, el error considerado en todos los casos fue el del 10 percentil. Para refinar los modelos con este criterio, se emplearon los valores de probabilidad asociados al 100% de los registros utilizados para la generación del modelo (entrenamiento + validación interna), a partir de los cuales se calculó la

probabilidad del 10 percentil. Los valores de los píxeles con probabilidades menores a la del 10 percentil fueron reasignados a NA.

Todos los procedimientos estadísticos, cartográficos y gráficos realizados durante la realización de este proyecto fueron realizados en R (R Development Core Team, 2008).

IV. CONCLUSIONES Y RECOMENDACIONES

De acuerdo con lo reportado en la literatura, consideramos que la metodología aquí empleada trata de resolver o aminorar gran parte de las fuentes de incertidumbre asociadas a la depuración de los datos y calibración del proceso de modelado. Sin embargo, sería interesante realizar una rigurosa cuantificación del rendimiento de distintas calibraciones, especialmente en la caso de la selección de variables.

Uno de los aspectos de mayor relevancia para lograr un modelado con mayor ajuste a la naturaleza de las especies, es el sesgo de muestreo en los datos. Pese a que dicha cuestión no fue planteada ni abordada, somos conscientes de que determina en gran medida el comportamiento de la distribución de la probabilidad en el universo de respuesta. Sin embargo, consideramos que dicha problemática tiene un efecto limitado cuando el modelo final es transformado a valores binarios (presencia/ausencia).

Consideramos que en la mayoría de los casos, el impacto del sesgo de muestreo sobre el rendimiento del modelo se encuentra directamente relacionado con el tamaño de la muestra, es decir, en especies con tamaños de muestra grande, una gran cantidad de registros se encuentran describiendo condiciones ambientales muy similares. En estos casos el modelador da un mayor peso a estas condiciones, asignando valores probabilísticos mucho mayores sitios ambientalmente similares y valores bajos a aquellos escasamente muestreados / ambientalmente raros.

V. REFERENCIAS

- Araujo, M.B. y A. Guisan. 2006. Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, 33: 1677-1688.
- Barve, N. 2008. Tool for partial-ROC. Biodiversity Institute, Lawrence KS. ver.10.
- Cervantes-Zamora Y., S.L. Cornejo-Olguín, R. Lucero-Márquez, J.M. Espinosa-Rodríguez, E. Miranda-Viquez y A. Pineda-Velásquez. 1990. Clasificación de regiones naturales de México, IV.10.2. *Atlas Nacional de México*, vol. II, escala 1 : 4 000 000. México.
- Elith, J., Graham, C. H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Mc Overton, J., Peterson, A.T., Phillips, S., Richardson, K., Scachetti-Pereira, R., Schapire, R., Soberón, J., Williams, S., Wisz, M. y Zimmerman, N. 2006. *Novel methods improve prediction of species' distributions from occurrence data. Ecography*, 29, 129-151.
- Elith, 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.
- Guisan, A., A. Lehmann, S. Ferrier, M. Austin, J.M.C. Overton, R. Aspinall y T. Hastie. 2006. Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology* 43: 386–392.
- Nakazawa, Y. 2013. Niche breadth, environmental landscape, and physical barriers: their importance as determinants of species distributions. *Biological Journal of the Linnean Society* 2: 241-250.
- Peterson, A.T., M. Papes y M. Eaton. 2007. Transferability, and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography* 30: 550–560.
- Peterson, A. T., M. Papes y J. Soberón. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling* 213: 63-72.

- Peterson T.A. 2011. Ecological niche conservatism: a time-structured review of evidence. *Journal of Biogeography* 38: 817–827.
- Phillips, S.J., R.P. Anderson y R.E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231–259.
- Phillips, S., & Dudík. M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31,161–175.
- Phillips, S.J., M. Dudík, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick y S. Ferrier. 2009. Sample selection bias and presence-only distribution models: implications for back-ground and pseudo-absence data. *Ecological Applications* 19,181–197.
- R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Ramírez-Pulido J., J. Arroyo-Cabrales y A. Castro-Campillo. 2005. Estado actual y relación nomenclatural de los mamíferos terrestres de México. *Acta Zoológica Mexicana (nueva serie)* 91: 21-82.
- Téllez Valdés, O., M. A. Hutchinson, H. A. Nix y P. Jones. 2011. Desarrollo de coberturas digitales climáticas para México. En: Cambio climático y biodiversidad en México. N. P. Pavón, C. Ballesteros Barrera and G. Sánchez Rojas (eds.) Colección Ciencia al Día. Universidad Autónoma del Estado de Hidalgo, Hidalgo, México
- Vaclavík T, J.A Kupfer y R. K. Meentemeyer. 2012. Accounting for multi-scale spatial autocorrelation improves performance of invasive species distribution modelling (iSDM). *Journal of Biogeography* 39: 42-55.
- Veloz, S. D. 2009. Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography* 12: 2290-2299.
- USGS. HYDRO1k, Geological Survey's (USGS) EROS Data Center, (DEM) of the world (GTOPO30), HYDRO1k.<http://eros.usgs.gov/>