

Informe final* del Proyecto JM032
Distribución potencial en taxones hiperdiversos y poco conocidos (Insecta: Coleoptera: Scarabaeoidea): estimando la localización de nuevos taxones y el efecto de los cambios en los usos del suelo*

Responsable: Dr. Federico Escobar Sarria
Institución: Instituto de Ecología A. C.
Dirección: Carretera antigua a Coatepec # 351, El Haya, Xalapa, Ver, 91070 , México
Correo electrónico: federico.escobar@inecol.edu.mx
Teléfono/Fax: +52 228 8421800 ext. 4111
Fecha de inicio: Agosto 31, 2012.
Fecha de término: Enero 21, 2019.
Principales resultados: Base de datos, cartografía, informe final.
Forma de citar el informe final y otros resultados:** Lobo, J.M., Peña-Peniche, A., Escobar, F., Vaz de Mello, F., Halffter, G., Dellacasa, M. y J. Cabrero. 2019. Distribución potencial en taxones hiperdiversos y poco conocidos (Coleoptera: Insecta: Scarabaeoidea): estimando la localización de nuevos taxones y el efecto de los cambios en los usos del suelo. Instituto de Ecología A. C. **Informe final SNIB-CONABIO, Proyecto N° JM032.** Ciudad de México.

Resumen:

Se compilará y se georreferenciará en bases de datos toda la información disponible en colecciones y bibliografía acerca de las especies de escarabajos del estiércol presentes en México, pertenecientes a las familias Aphodiidae, Geotrupidae y Scarabaeidae. A partir de esta información se obtendrán datos básicos ecológicos y biogeográficos y se elaborarán mapas de idoneidad climáticas (distribuciones potenciales) para cada una de las especies.

-
- * 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.

INFORME FINAL – CONABIO JM032

Título del Proyecto

Distribución potencial en taxones hiperdiversos y poco conocidos (Insecta: Coleoptera: Scarabaeoidea): estimando la localización de nuevos taxones y el efecto de los cambios en los usos del suelo

Datos generales

Responsable: Federico Escobar Sarria
Instituto de Ecología, A. C, Xalapa, México
federico.escobar@inecol.mx

Corresponsable: Francisco Cabrero Sañudo
Universidad Complutense de Madrid, Madrid, España

Participantes: Jorge M. Lobo
Museo Nacional de Ciencias Naturales, Madrid, España

Marco Dellacasa
Museo di Storia Naturale e del Territorio dell'Università di Pisa,
Italia

Fernando Vaz de Mello
Universidade Federal de Mato Grosso, Mato Grosso, Brasil

Gonzalo Halffter
Instituto de Ecología, A.C, Xalapa, México

Equipo técnico: Alexander Peña (Técnico del proyecto)
Ángela Mejía (Técnico del proyecto)
Alfonso Aceves (Técnico del proyecto)
Fernando Escobar (Técnico Institucional, Inecol)

Xalapa, 15 de septiembre de 2018

Índice

Resumen ejecutivo	3
1. Introducción	4
2. Desarrollo del Proyecto	4
2.1. Modelos de distribución y patrones biogeográficos: bases conceptuales y protocolo seguido en la elaboración de los modelos	6
2.2. Discriminación y descripción de las principales áreas necesitadas de prospección para los Scarabeoidea coprófagos	14
3. Patrones geográficos	26
3.1. Bolboceratinae	
3.2. Geotrupinae	
3.3. Aphodiinae	
3.4. Scarabaeinae	
4. ¿Son fiables las predicciones realizadas?	35
5. Estimación de la vulnerabilidad ante los cambios de uso de suelo	38
6. Referencias	43

Resumen ejecutivo

El **objetivo general** de este proyecto fue proporcionar información exhaustiva acerca de la distribución geográfica de las especies de escarabajos del estiércol perteneciente a dos Familias: Scarabaeidae (subfamilias Scarabaeinae, Aphodiinae) y Geotrupidae (subfamilias Bolboceratinae, Geotrupinae) presentes en México y en territorios cercanos. Como **primer objetivo particular**, se conjuntó información proveniente de tres bases de datos (ScaMex: Scarabaeinae, AphMex: Aphodiinae y GeoMex: Geotrupidae), a través de la compilación de las referencias de distribución de las especies correspondientes, y datos de otras fuentes de información, como SNIB (Sistema Nacional de Información sobre Biodiversidad; CONABIO), GBIF (Global Biodiversity Information Facility) y PhanMex. Como **segundo objetivo particular**, se obtuvieron mapas de distribuciones potenciales para cada una de las especies, mediante *Multidimensional Enveloping*, un procedimiento simple de modelización de nicho ecológico recientemente consensuado entre distintos equipos de investigación. De acuerdo al criterio de los expertos que participaron en el proyecto, el total de especies de escarabajos del estiércol con nombre válido para México y áreas aledañas es de 473. La base de datos final (DarwinCore) quedó integrada por 507 especies y 52 subespecies (de las cuales 12 especies son consideradas introducidas). Se generaron mapas para 451 especies, a partir de la información de una base de datos integrada por 37 626 registros. Debido a que 170 de estos mapas fueron elaborados con menos de 3 registros, la CONABIO realizó la revisión de 281 mapas. Se cumplió prácticamente con la totalidad de los objetivos planteados por el proyecto y establecidos en el Convenio, excepto el *Atlas biogeográfico*. Sin embargo, se realizó un análisis biogeográfico de los principales patrones de distribución de las cuatro subfamilias en el país, el cual se presenta en este documento. También se realizó un primer análisis de las afectaciones del cambio de uso de suelo en los patrones de diversidad del grupo. Este estudio representa el primer esfuerzo para sintetizar información sobre la distribución geográfica de los escarabajos del estiércol, un grupo de insectos de gran importancia ecológica y funcional tanto en ambientes naturales como derivados de la actividad humana, en particular, en áreas de uso ganadero.

1. Introducción

Está globalmente aceptado que se está produciendo una rápida tasa de extinción a gran escala mundial para algunos componentes de la biodiversidad, con consecuencias impredecibles para los sistemas naturales y para las sociedades humanas (Millenium Ecosystem Assessment, 2005). Uno más de estos componentes es el de los escarabajos del estiércol.

La superfamilia Scarabaeoidea (Insecta: Coleoptera) contiene un gran número de especies especializadas en el consumo de excrementos de herbívoros (Scarabaeinae Aphodiinae, y Geotrupidae), comúnmente conocidos con el nombre de escarabajos coprófagos o del estiércol (Halffter y Edmonds, 1982). Estos insectos son de gran importancia ecológica, puesto que aumentan la permeabilidad del suelo y reciclan la materia orgánica, favoreciendo la productividad de los pastos (Bornemissza, 1976; Rougon et al., 1988). También son los principales controladores biológicos de insectos hematófagos y vectores de enfermedades del ganado (McQueen y Beirne, 1975; Waterhouse, 1974).

En México se cuenta con un número considerable de especies de escarabajos coprófagos como resultado de su localización geográfica entre la región Neártica y Neotropical, más conocida como Zona de Transición Mexicana (propuesta y descrita por Halffter, 1987). Según los últimos listados de especies para este grupo de insectos en México, este territorio podría estar acogiendo al menos 111 especies de Aphodiinae (Cabrero-Sañudo et al., 2007, 2010) y otras 110 especies del resto de subfamilias que integran este grupo (Eupariinae, Psammodiinae y Rhyariinae; datos no publicados de F.J. Cabrero-Sañudo), 45 especies de Geotrupidae (Trotta-Moreu et al., 2008) y más de 180 especies de Scarabaeinae (Morón, 2003).

Debido a que los escarabajos del estiércol se encuentran principalmente asociados tanto a áreas de bosque como pastizales, en el último medio siglo han sufrido una reducción en sus distribuciones geográficas y en sus poblaciones (Barbero et al., 1999; Carpaneto y Mazziota, 2007; Leclerk et al., 1980; Lobo, 2001; Lobo et al., 2001; Lumaret, 1990; Lumaret y Kirk, 1991; Romero-Samper y Lobo, 2006) a consecuencia de un cambio en las prácticas agropecuarias y pastorales tradicionales (Barbero et al., 1999; Hutton y Giller, 2003), el desarrollo urbano asociado a los cambios de uso (Lobo, 2001; Lobo et al., 1997, 2001), la contaminación química (Hole et al., 2005) y la depredación por otras especies (Gittings y Giller, 1999; Gittings et al., 1994; Horgan y Berrow, 2004). Por estas razones, los escarabajos del estiércol han sido utilizados como indicadores con interés en la conservación de la biodiversidad (Davis et al., 2004; McGeoch et al., 2002).

Por todo ello, este grupo de especies es especialmente idóneo para sacar provecho de la aplicación del procedimiento propuesto para estimar la distribución potencial. Dada su diversidad, este proyecto puede ser capaz de proporcionar datos de localización para nuevas especies susceptibles de validarse en campo, así como evaluar la disminución de la favorabilidad del territorio para estas especies que puedan estar afectando a la calidad edáfica y el uso ganadero del territorio.

Los principales aportes que se esperaban de este proyecto fueron: Bases de datos de escarabajos del estiércol mexicanos unificada en Darwin Core (con 29 campos de información); Compilación de las variables bioclimáticas a utilizar para modelizar el área de distribución potencial de cada una de las especies; Modelos de distribución potencial de 436 especies. Cartografía digital de los mapas de distribución de presencias observadas y distribución climática potencial para cada una de las especies. Metadato acompañando a cada mapa potencial. Adicionalmente, se realizó una enumeración y descripción de las principales áreas geográficas necesarias de ser exploradas en términos faunísticos para este grupo de insectos y se estimó la vulnerabilidad de cada una de las especies ante los cambios en los usos del suelo y la vulnerabilidad completa para el conjunto de especies.

Cabe resaltar que el *Atlas de Escarabajos del Estiércol* de México, comprometido como un producto del proyecto, aún se encuentra en proceso de elaboración. Sin embargo es este informe se ofrece un análisis comparativo de los patrones biogeográficos de las subfamilias, el cual es el insumo básico para la elaboración del Atlas. La idea en el mediano plazo es realizar una publicación de un libro profusamente ilustrado, que pueda ser accesado de forma digital, para los interesados y expertos científicos en los biología, ecología y conservación de este grupo de insectos.

2. Desarrollo del Proyecto

2.1. MODELOS DE DISTRIBUCIÓN Y PATRONES BIOGEOGRÁFICOS PRINCIPALES: Bases conceptuales y protocolo seguido en la elaboración de los modelos

Las modelizaciones de carácter correlacional para conocer la distribución actual de los organismos o extrapolar esta distribución bajo diferentes escenarios se han impuesto en la literatura, tanto debido a su facilidad operativa como a la carencia de datos fisiológicos, demográficos o ecológicos que permitan obtener predicciones basadas en resultados empíricos. Desde 1995 el número de trabajos publicados sobre este tema en revistas científicas indexadas ha crecido exponencialmente de modo que, actualmente, alrededor de un trabajo es publicado cada día (Hortal *et al.* 2012) ¿Hasta qué punto son fiables estos pronósticos? ¿Es posible minimizar o limitar los sesgos, los errores y las incertidumbres de estas predicciones? Si para la gran mayoría de la biodiversidad apenas poseemos información parcial sobre su distribución actual y/o pasada, y hay un gran desconocimiento sobre sus preferencias ambientales y tolerancias fisiológicas ¿Cómo debemos usar estos datos para ofrecer estimas razonables?

Sin ánimo de ser exhaustivo, diversos trabajos han constatado desde hace tiempo, que los datos disponibles sobre la distribución geográfica de la gran mayoría de los organismos están sesgados. Que cuando se consideran conjuntamente, estos mapas reflejan, básicamente, el esfuerzo de colecta y no la distribución de las especies, y que no permiten obtener una imagen fiel de las condiciones ambientales sobre las que pueden vivir la mayoría de las especies (p.ej., Dennis *et al.* 1999, Hortal *et al.* 2007 y 2008, Boakes *et al.* 2010, Rocchini *et al.* 2011, Ballesteros-Mejía *et al.* 2013, Yang *et al.* 2013 o Duputie *et al.* 2014). Las consecuencias de esta falta de calidad en los datos primarios son principalmente dos: una que atañe a lo que realmente muestran estas representaciones y otra que afecta a la gestión de los escasos recursos disponibles. Como se ha demostrado matemáticamente hace poco (Aarts *et al.* 2012) y ha sido ilustrado hace tiempo (Lobo 2008a), los resultados de estas predicciones no hacen sino mostrar la frecuencia de los propios puntos utilizados en el proceso de construcción del modelo, siendo altamente dependientes de la naturaleza y estructura de los datos de partida. Así, en la mayoría de las ocasiones, estamos prediciendo los propios datos incluidos en el modelo y las medidas de validación, inconsistentes y mal utilizadas, que pretenden avalar la consistencia de estas simulaciones (ver Lobo *et al.* 2008, Jiménez-Valverde 2012 y Hijmans 2012) ofrecen estimas de precisión también falseadas, al basarse en esos mismos datos sesgados. ¿Cómo vamos a

obtener valores de probabilidad que reflejen la frecuencia de aparición de una especie si los datos de partida reflejan la frecuencia de colecta? Desde el punto de vista de la gestión, el no reconocimiento de estas carencias en la información biológica, implica destinar fondos y recursos a la creación de simulaciones basadas en la aplicación de técnicas de parametrización complejas sobre datos incompletos y sesgados, en vez de facilitar y promover estructuras estandarizadas de toma de datos biológicos que nos puedan permitir realizar un seguimiento y unas simulaciones coherentes y eficaces.

La ausencia verdadera o altamente probable es un dato imprescindible para poder derivar la distribución de un organismo, pero no todas las ausencias poseen el mismo valor y son consecuencia de los mismos procesos (ver Lobo *et al.* 2010), siendo la falsa ausencia debida a los sesgos de colecta la que más inconvenientes causa a la hora de establecer modelos de distribución fiables. Las ausencias “contingentes”, que aparecen como consecuencia de la falta de equilibrio en las distribuciones (debido a la dispersión, factores históricos, interacciones bióticas, etc.), son reales y obligan a que los modelos tengan que incluir predictores capaces de dar cuenta de estos procesos (Sutherst 2014). Tener en cuenta la falta general de equilibrio de las distribuciones con el clima requiere considerar tanto datos de ausencia fiables como predictores capaces de dar cuenta de esas ausencias. No puede ser de otro modo. Pero también requiere poseer una muestra representativa de datos que refleje el gradiente ambiental que se pretende predecir, precaución a la hora de elegir las variables explicativas que acompañarán a las variables climáticas, así como limitar la extensión del área de estudio a la región en la que es menos probable que actúen estos limitantes de dispersión (Acevedo *et al.* 2012). La variedad de técnicas de modelización y simulación existentes permite obtener predicciones cabales, al igual que se hace en otras disciplinas como la medicina o las ciencias sociales, pero siempre y cuando la muestra sea una representación aleatoria y no sesgada de la población que se pretende conocer.

Si la presencia de una especie en una localidad es una evidencia empírica, su ausencia es, en parte, una cuestión probabilística. Una intensificación progresiva del esfuerzo de colecta permite asegurar, cada vez con mayor certidumbre, que la especie de interés no está presente, pero ello requiere esfuerzos repetidos y generosos. Desafortunadamente, los Atlas y las bases de datos de biodiversidad, como los que aquí hemos estudiado, no suelen ofrecer información sobre el esfuerzo de colecta desarrollado en cada una de las localidades, y las observaciones repetidas de una misma especie en una localidad se consideran redundantes y se omiten, generalmente, cuando sin embargo ofrecen una valiosa información sobre la frecuencia de aparición de una especie. Si una especie ha sido observada cien veces en una localidad es

factible asumir que la ausencia de una especie hermana en esa misma localidad sea probable, pero ¿qué decir cuando se posee un único dato de presencia? La base de datos sobre los Scarabaeoidea coprófagos que aquí hemos compilado y estudiado ha tratado de ser exhaustiva, a fin de recoger toda la información existente y, como hemos visto en la sección anterior, poder discriminar donde se encuentran las regiones exploradas y las inexploradas. Las localidades relativamente bien prospectadas en las que no se ha observado una especie podrían considerarse ausencias fiables y el número de registros existentes en ellas podría usarse como una covariable en los modelos, de modo que las predicciones pudiesen corregirse según el esfuerzo realizado en ellas. Desafortunadamente, como hemos visto anteriormente, la proporción y cobertura ambiental de las celdas con una resolución adecuada para este grupo de escarabajos y en general para los insectos, son tan reducidas que sería una temeridad elaborar modelos de distribución fiables teniendo en cuenta tanto los datos de presencia como los de ausencia. ¿Qué podemos predecir si no tenemos datos de ausencia fiables y cualquier localidad aparentemente no colonizada podría contener una presencia de la especie? ¿Cómo vamos a realizar predicciones fiables sobre la distribución de una especie si la información disponible refleja sesgos de muestreo desconocidos o no considerados?

La carencia de información fiable sobre las ausencias ha propiciado el uso de los denominados *presence-only-models* en los que, en vez de “verdaderas” ausencias se utilizan pseudoausencias, en la mayoría de las ocasiones seleccionadas al azar dentro del territorio considerado (*background absences*). Se trata del clásico procedimiento que compara las condiciones ambientales de los datos de presencia disponibles (uso) respecto a las existentes en todo el territorio considerado (disponibilidad). Son las denominadas *Resource Selection Functions* (Johnson 1980) introducidas en los años 80 para estimar la probabilidad de aparición de una especie en distintos tipos de hábitats. En la actualidad, buena parte de los programas más exitosos en la elaboración de modelos predictivos de distribución realizan esta comparación para ofrecer sus predicciones y la literatura está repleta de estudios que realizan esta práctica. Aunque se conocía que el efecto de la utilización de diversos tipos de pseudoausencias, más o menos alejadas ambiental o espacialmente del universo de las presencias, ejercía una importante influencia sobre las representaciones obtenidas (Chefaoui & Lobo 2008), se ha demostrado que este tipo de procedimientos no pueden estimar la probabilidad de aparición de una especie (Hastie & Fithiam 2013) y que sólo reflejan la intensidad de los propios datos de partida (Aarts *et al.* 2012). Cuando realizamos un modelo de este tipo, el resultado final se parece sospechosamente demasiado a los propios datos de partida y las evaluaciones con esos mismos

datos ofrecen poderes de discriminación inevitablemente altos que no evalúan verdaderamente la capacidad predictiva de los modelos.

Por todo ello, en este proyecto se han seguido los siguientes criterios básicos a la hora de elaborar representaciones sobre la **distribución probable** –que no potencial- de cada una de las especies:

- Utilizar solamente la información contenida en los datos de presencia. Las localidades en la que se ha observado una especie nos proporcionan una información valiosa sobre las condiciones en las que ésta puede habitar. Aunque sepamos que, en muchos casos, las colectas que hemos realizado están lejos de seguir los protocolos de un muestreo estandarizado clásico (por ejemplo, muestreos al azar o estratificados), y aunque sepamos también que los datos de presencia son, en ocasiones, insuficientes y sesgados, esta información es, para regiones y grupos de insectos hiperdiversos e insuficientemente prospectados, la única fuente de conocimiento empírico sobre las condiciones en las que una especie puede habitar. Hemos desechado, por tanto, el uso de pseudoausencias y la utilización de técnicas complejas de modelización con ellas.
- Siguiendo los anteriores razonamientos, los modelos o representaciones de las distribuciones de cada una de las especies se han realizado siguiendo un simple proceso de intersección generalizada de conjuntos o *Multidimensional Enveloping* (Jiménez-Valverde *et al.* 2011). Se trata, en esencia, de elegir una serie de variables ambientales predictoras, extraer sus valores para cada una de las localidades de presencias conocidas, y reconocer en el territorio deseado la ubicación de localidades con condiciones ambientales similares. Las predicciones así generadas no son, en realidad, escenarios que reflejen la probabilidad de aparición de la especie, sino **escenarios que representan las localidades con condiciones similares a las que se dan en aquellos lugares en donde la especie ha sido observada** (los datos de presencia ofrecen una representación parcial del “nicho” de una especie que después se traslada al espacio geográfico deseado).
- Cuando se representan geográficamente las condiciones ambientales en las que se ha observado una especie, se obtiene una inferencia de distribución potencial: una imagen provisional de las localidades “a priori” habitables, cuando el conjunto de factores contingentes que impide la colonización de ciertas regiones deja de actuar. Por ello y a fin de acercarnos a inferencias que represente la distribución realizada, o mejor, probable, necesitaremos circunscribir el área de estudio de cada especie a un

área de trabajo (*spatial background*), de modo que se minimice el posible efecto de los limitantes de dispersión y los factores históricos a la hora de generar distribuciones.

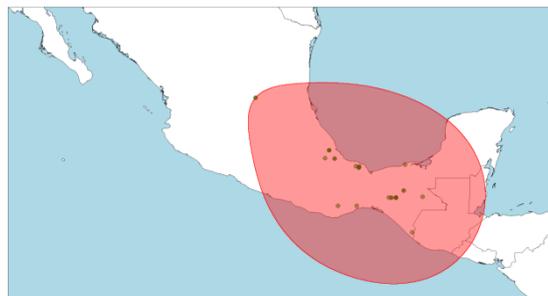
Los anteriores criterios han sido aplicados para ofrecer mapas de riesgo en el caso de especies invasoras (Aragón *et al.* 2010b), decidir la ubicación de nuevas áreas a prospectar (Sánchez-Fernández *et al.* 2011a), estimar el grado de equilibrio en las distribuciones (Sánchez-Fernández *et al.* 2012) o predecir los posibles efectos del cambio climático sobre las distribuciones (Aragón & Lobo 2012). Varias fuentes de incertidumbre afectan a los resultados de estas simulaciones, siendo uno de los principales la falta de datos geográficos capaces de maximizar la representación del “nicho” de las especies (Sánchez-Fernández *et al.* 2011b). En esta aproximación es clave incluir el mayor espectro de localidades con observaciones de presencia disponibles, tal vez incluso incorporando aquellas provenientes de los lugares ocupados en los que no hay crecimiento demográfico neto positivo. Se trata, básicamente, de identificar las condiciones en los límites ambientales y/o geográficos de la distribución, en vez de establecer curvas de respuesta basadas en unas frecuencias de observación a menudo influenciadas por los sesgos de colecta. En nuestro caso, siempre hemos incluido observaciones de las especies más allá del territorio mexicano cuando se encontraban disponibles.

Otra de las fuentes de incertidumbre en este tipo de modelos es debida a la identidad y el número de las variables explicativas que se utilicen (Beaumont *et al.* 2005). En nuestro caso, se han usado las 19 variables bioclimáticas descritas por Cuervo-Robayo *et al.* (2014), las cuales han sido recientemente interpoladas para México. Para identificar este número mínimo de variables se calculó el VIF (*Variance Inflation Factor*) que mide la multicolinealidad entre estas variables bioclimáticas eliminando aquellas cuyos valores están altamente correlacionados y son redundantes (VIF mayor de 30, en nuestro caso). Tras este primer proceso de selección, se procedió a escoger aquellas variables que parecen estar afectando más la distribución observada de cada especie mediante el índice de inestabilidad propuesto por Guisande *et al.* (2006), el cual se basa en el índice de fluctuación de Dubois (1973). En esencia, se trata de un índice que mide la diferencia en los valores de cada variable predictora entre las localidades en las que la especie ha sido observada (presencias) o no (ausencias inciertas). Una vez calculado este índice de inestabilidad, pueden ordenarse cada una de las variables predictoras según su contribución al valor del índice, de modo que aquellas variables que son capaces de discriminar mejor entre las “presencias observadas” y las “ausencias inciertas” serían las que “a priori” tendrían una mayor posibilidad de ser relevantes (una variable que

posee valores similares en las localidades en las que se ha observado una especie y en el resto de las localidades del territorio considerado, tiene pocas posibilidades de ser relevante). Este segundo proceso de selección se efectuó para cada una de las especies modelizadas, seleccionando aquellas variables que logran explicar el 90% del valor del índice de inestabilidad.

De las 436 especies consideradas, se realizaron modelos de distribución siguiendo el protocolo antes mencionado para las 282 que poseían 5 o más registros georreferenciados (detalles en el Cuadro 2). Para el resto de las especies, el área de distribución se estimó como el área circular aproximada de 20 km de radio alrededor de cada punto de observación. Todo el proceso de modelización para cada una de las 282 especies se realizó mediante el programa ModestR (García-Roselló *et al.* 2013; ver <http://www.ipez.es/ModestR/>) siguiendo los siguientes pasos:

- a. Para la elaboración de cada uno de los modelos se delimitó como región geográfica el área comprendida desde el sur de Estados Unidos, los países de México, Belice, Guatemala, Honduras, El Salvador, Nicaragua, Panamá y Costa Rica. Esta región fue delimitada por las coordenadas extremas: -74.9464 Este, -122.9933 Oeste, 37.8398 Norte y 6.391724 Sur. Las coberturas climáticas empleadas para la elaboración del modelo fueron las 19 variables del Cuervo-Robayo *et al.* (2014), las cuales representan condiciones climáticas actuales (~1950-2000).
- b. Definir un *spatial background* o área de trabajo individual para cada especie utilizando una función de densidad de las observaciones (*kernel*) cinco veces mayor que la desviación estándar de las distancias entre los puntos con observaciones. De este modo, tal y como muestra la imagen, no se extrapolan las predicciones fuera del ámbito espacial en el que ha sido observada la especie.



- c. Calcular el VIF predictoras de las variables dentro del *spatial background* anteriormente seleccionado, de modo que se eliminan aquellas variables redundantes de acuerdo a sus valores en el área considerada y no en el territorio con

menores posibilidades de estar habitado. Este proceso suele significar reducir, al menos, a la mitad el número de variables.

- d. Calcular el índice de inestabilidad de las variables anteriormente seleccionadas también dentro del *spatial background*, seleccionando aquellas que logran explicar el 90% del valor del índice. Esta segunda selección suele eliminar un tercio de las variables anteriormente elegidas.
- e. Construir con las variables finalmente seleccionadas una única variable utilizando coordenadas polares. Las coordenadas polares son un [sistema de coordenadas bidimensional](#) en el cual cada [punto](#) del plano se determina por una [distancia](#) al centro (el valor de la variable) y un [ángulo](#) (la identidad de la variable). Así, toda la información de las variables elegidas está comprendida en un par de valores asignados a cada localidad del mapa, los cuales se representan mediante un código de colores.
- f. Calcular los valores máximos y mínimos de la nueva variable construida mediante coordenadas polares para el conjunto de las presencias observadas, representando dentro del *spatial background* elegido todos los píxeles con valores comprendidos entre ambos extremos (o *Multidimensional Enveloping*), tal como muestra la imagen. Se trata de las zonas cercanas geográficamente a las localidades de presencia que, a la vez, poseerían condiciones climáticas similares a las de los lugares donde ha sido observada la especie.



- g. Eliminar todas las áreas de presencia probable que estén fuera del territorio mexicano, que no estén claramente conectadas con un área en la que haya una observación de presencia y/o que se encuentren en una cuenca hidrográfica sin presencias observadas (ver imagen). Se trata de minimizar el área climáticamente favorable para reducir el riesgo de errores de comisión, aunque según la exhaustividad del conocimiento faunístico de cada especie, algunas de estas áreas climáticamente favorables pudieran albergar poblaciones de la especie en cuestión.



Para medir el desempeño de los modelos obtenidos, se calculó el AUC, o área bajo la curva ROC, de cada predicción como medida de precisión de las predicciones siguiendo el método tradicionalmente usado y propuesto por Phillips (2017) para cuando no existen datos de ausencia fiables.

Se estableció un umbral para transformar valores continuos en binarios (presencia/ausencia). El umbral ideal es el que minimiza la diferencia entre sensibilidad (proporción de presencias correctamente predichas) y especificidad (proporción de ausencias correctamente predichas) (Liu et al., 2005; Jiménez-Valverde & Lobo, 2007). Sin embargo, como no se conocen las ausencias y la especificidad depende de la relación entre el área de estudio y el área de distribución de la especie (que tampoco se conoce), no puede calcularse la especificidad. En su lugar se ha propuesto sustituir la fracción de ausencias erróneamente predichas como presencias (1-especificidad) necesaria para el cálculo del AUC por la fracción del área de estudio (Phillips, 2017), de modo que se maximiza la obtención de predicciones capaces de predecir correctamente una alta proporción de las presencias en la menor área posible.

2.2. DISCRIMINACIÓN Y DESCRIPCIÓN DE LAS PRINCIPALES ÁREAS NECESITADAS DE PROSPECCIÓN PARA LOS SCARABAEOIDEA COPRÓFAGOS

La Base de datos DarwinCore comprende 37 626 registros pertenecientes a 507 especies y 52 subespecies: 416 taxones de la familia Scarabaeinae (382 especies y 34 subespecies, e incluye subfamilias Aphodiinae y Scarabaeinae), y 44 taxones de la familia Geotrupidae (382 especies y 34 subespecies). La base inicial quedó integrada por información disponible en seis fuentes distintas de datos (Cuadro 1). Algunos registros fueron eliminados durante el proceso de revisión que realizó la CONABIO. Con este proyecto se logró incrementar en un 50% el número de registros disponibles en el SNIB-CONABIO. Al listado total de especies habría que añadir 10 especies más mencionadas por los expertos taxónomos de los distintos grupos y que podrían estar presentes en territorio mexicano. Se ofreció finalmente información cartográfica para 451 especies, ya que no se posee información geográfica para México de las restantes especies, conociéndose su presencia en países limítrofes o cercanos (27 citadas del sur de EEUU), y 12 especies serían introducidas en territorio mexicano. Debido a que 170 de estos mapas fueron elaborados con menos de 3 registros, la CONABIO realizó la revisión de 281 mapas.

Cuadro 1. Relación de los registros aportados por cada una de las bases de datos y fuentes de información utilizadas para elaborar los modelos. *Corresponde a la base de datos elaborada a partir de un trabajo de tesis de maestría realizada por Viridiana Lizardo.

Fuente	Registros
CONABIO	19119
Gbif	9505
AphoMex	2763
GeoMex	1153
ScaMex	2789
PhanMex*	2645
Otros	146
Total	38120

El nivel general de conocimiento biogeográfico de las especies de Scarabaeoidea coprófagos puede considerarse adecuado (82 registros por especie, aproximadamente), si se considera que se trata de un grupo de insectos que, en ocasiones, posee una taxonomía difícil y todavía no resuelta. Sin embargo, este conocimiento biogeográfico resulta muy dispar entre especies y grandes grupos (Cuadro 2 y Fig. 1). Los Aphodiinae son, sin duda, el grupo menos

muestreado y con menor volumen de información; prácticamente la mitad de sus especies poseen menos de 5 registros y del 37% de ellas solo se conoce su presencia en una o dos localidades. El conocimiento sobre la distribución geográfica de los Scarabaeinae resulta algo más exhaustivo. Aunque un tercio de las especies reconocidas en México poseen 5 o menos registros, el alto número de registros por especie indica que este conocimiento es muy desigual entre los distintos géneros, siendo *Onthophagus* y *Cryptocanthon* los géneros con menor nivel de conocimiento. Por último, la familia menos diversificada en México (Geotrupidae) sería la que posee un nivel de conocimiento biogeográfico más completo (Cuadro 2).

Cuadro 2.- Datos básicos sobre el nivel de conocimiento de los tres principales grupos de Scarabaeoidea en México.

	Scarabaeinae	Aphodiinae	Geotrupidae
Nº especies consideradas	268	108	44
Especies con 5 o menos registros	86 (32%)	55 (51%)	7 (16%)
Nº medio registros por especie (\pm 95% IC)	114 \pm 24	28 \pm 11	46 \pm 22
Cuartil superior (registros)	149	25	54

Diversos métodos pueden permitirnos utilizar la información contenida en nuestras bases de datos de biodiversidad para discriminar las áreas bien e insuficientemente prospectadas, asumiendo que éstas reflejan, aunque sea parcialmente, el esfuerzo de colecta realizado. Las curvas de acumulación en las que se representa la creciente completitud en los inventarios según se incrementa el esfuerzo de colecta son una posibilidad (Chao & Jost 2012), como lo son diversos estimadores no-paramétricos (Hortal *et al.* 2006)

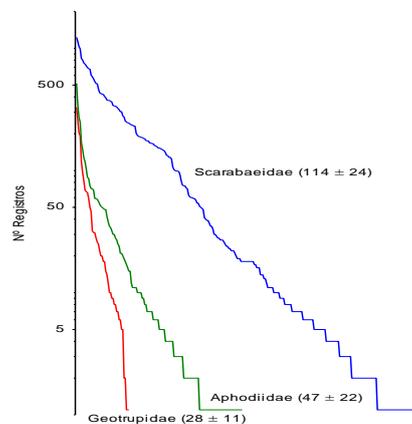


Figura 1.- Ordenación decreciente de las especies de Scarabaeoidea coprófagos pertenecientes a los tres grupos taxonómicos considerados en el proyecto según el número de registros que se

posee de cada una ellas. Los números entre paréntesis representan el promedio de registros por especie (\pm el intervalo de confianza al 95%).

En estos análisis puede utilizarse el número de registros como un sustituto del esfuerzo de colecta (Lobo 2008b) con capacidad para discriminar las localidades con inventarios relativamente fiables, de aquellas otras todavía insuficientemente muestreadas en las que se deben realizar prospecciones adicionales (Hortal & Lobo 2005, Aranda *et al.* 2011). Con este propósito, se asignaron los registros disponibles de cada una de las especies a celdas de distinta resolución (cuadrículas de 1°, 30' y 15'), estimándose para cada celda el número de especies registradas, el número de registros existentes y la pendiente final de la curva de acumulación en el número de especies a medida que se incluyen registros en orden aleatorio (100 aleatorizaciones). Esta pendiente se puede considerar una estimación confiable de la tasa de incremento en la riqueza de especies con la adición de nueva información, cuando el proceso de inventario en cada celda se encuentra en su fase final. Se consideró que cuando esta tasa era igual o menor que 0.01 (una especie nueva añadida al inventario cada 100 nuevos registros) la celda puede considerarse bien muestreada (CBM). Para calcular estas pendientes se requirió que, al menos, hubiera el doble de registros que especies observadas y que la celda incluyera, como mínimo, 10 registros. Todos estos cálculos fueron realizados utilizando una aplicación en R (*KnobR*) concebida ex profeso para este cometido que, en nuestro caso, funciona a través de la GUI RWizard (ver <http://www.ipez.es/RWizard/>).

Los resultados de estos análisis se presentan en cuadro 2 y la figura 2. En ellos se puede observar que, a medida que aumenta la resolución disminuye el porcentaje de celdas que podrían considerarse bien muestreadas. Si alrededor de tres cuartas del territorio mexicano posee algún registro cuando se consideran celdas de un grado, solo una cuarta parte posee alguna información en el caso de celdas de 15'. Si se tiene en cuenta que estas celdas poseen una superficie aproximada de 900 km², un extenso territorio para un insecto, resulta evidente que gran parte del territorio carece de prospecciones. Una reducción similar puede apreciarse en el caso de las celdas con más de 5 registros (Cuadro 3) que, en la mayor de las resoluciones, apenas sobrepasan el 9% del total. Sin embargo, es el número de cuadrículas con inventarios fiables (CBMs) el que mejor ejemplifica la situación. En este caso, el porcentaje de celdas bien inventariadas apenas llega a suponer un 5% del total a la menor de resolución y un 0,9% a la mayor resolución. Si consideráramos una resolución adecuada para el tamaño corporal de un insecto y su capacidad de dispersión (celdas de 10 x 10 km, aproximadamente) el porcentaje de cuadrículas con inventarios fiables sería de un 0,3%, aproximadamente (datos no mostrados).

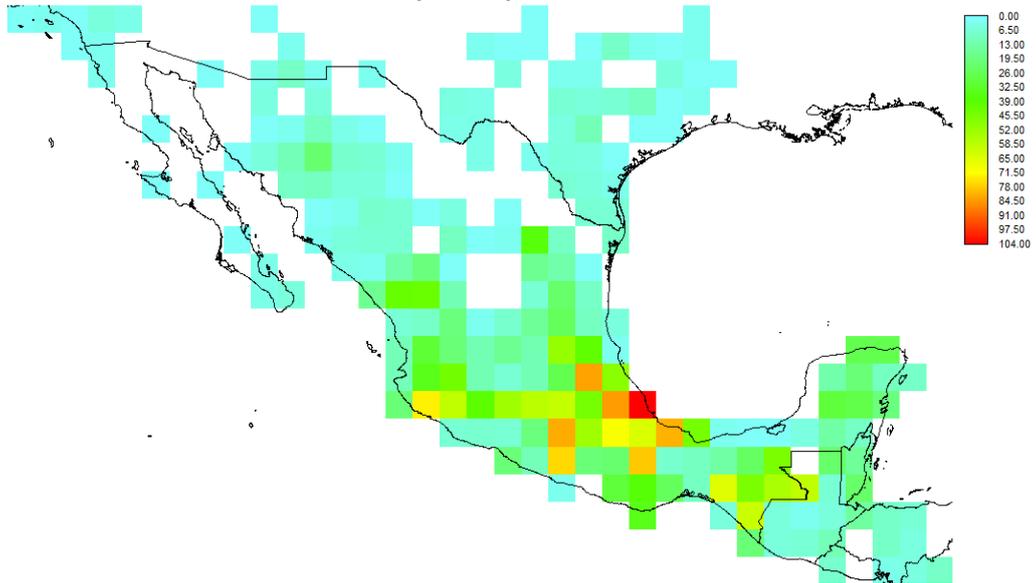
La precaria situación del conocimiento acerca de la diversidad de los Scarabaeoidea coprófagos mexicanos, se ejemplifica con claridad cuando se examina la correlación entre la riqueza de especies observada y la cantidad de registros (Fig. 3). Los valores de estas correlaciones son siempre altos y positivos (Cuadro 3), con relaciones que sugieren, claramente, que muchos de los lugares aparentemente pobres en especies podrían no serlo dado el esfuerzo de colecta desplegado en ellos (Fig. 3). Es lógico que los lugares ricos en especies posean un esfuerzo de colecta grande, pero no lo es tanto que la mayoría de las celdas posean bajos números de registros y especies.

Cuadro 3.- Datos básicos sobre el nivel de conocimiento de los tres principales grupos de Scarabaeoidea en México. r es el coeficiente de correlación de Pearson entre el número de registros y de especies observadas en cada una de las celdas (ver Fig. 2), mientras que CBMs es el número de celdas que podrían considerarse bien muestreadas (ver texto).

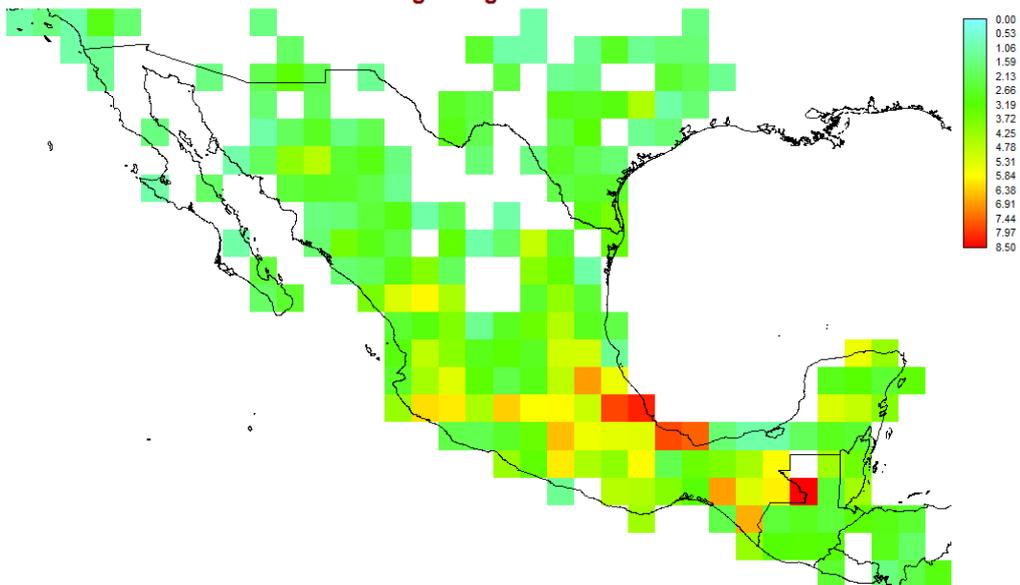
	Resolución		
	60'	30'	15'
Total de celdas	218	791	2998
Con registros	157 (72,0%)	333 (42,1%)	575 (10,2%)
Celdas con más de 5 registros	128 (58,7%)	237 (30,0%)	277 (9,2%)
r	0.881***	0.893***	0.893***
CBMs	10 (4,6%)	16 (2,0%)	28 (0,9%)

60'

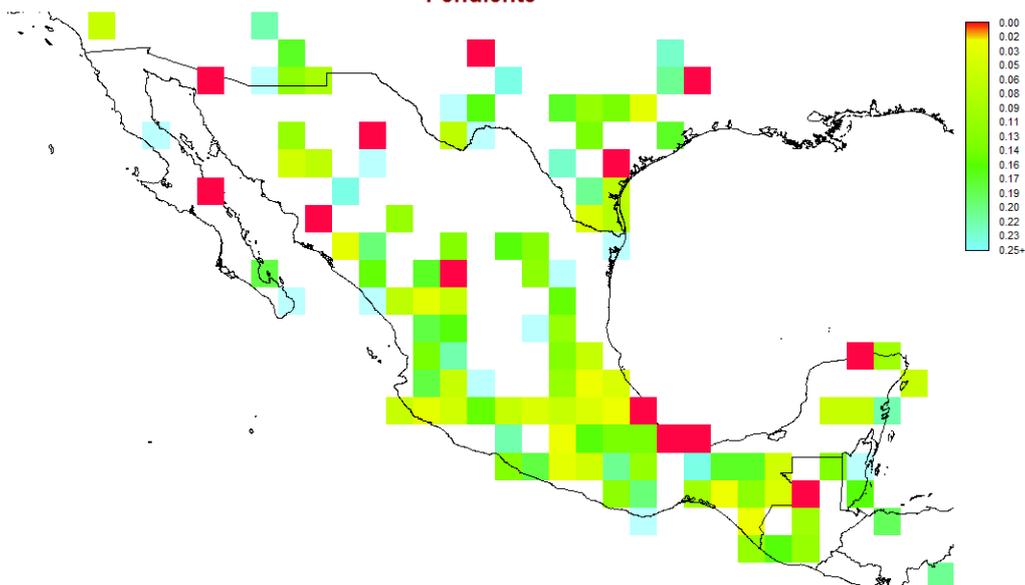
Riqueza especies



log n° registros

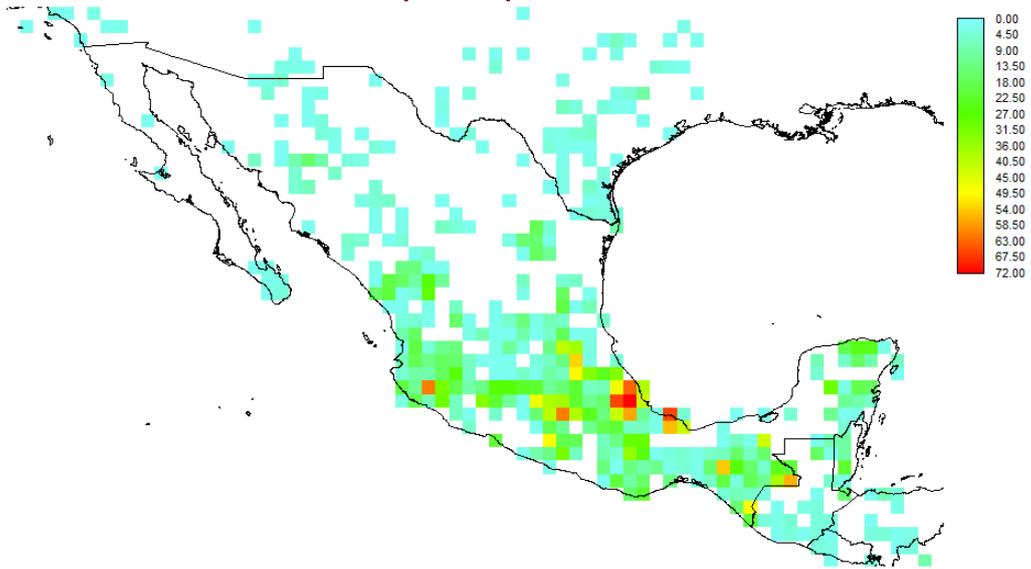


Pendiente

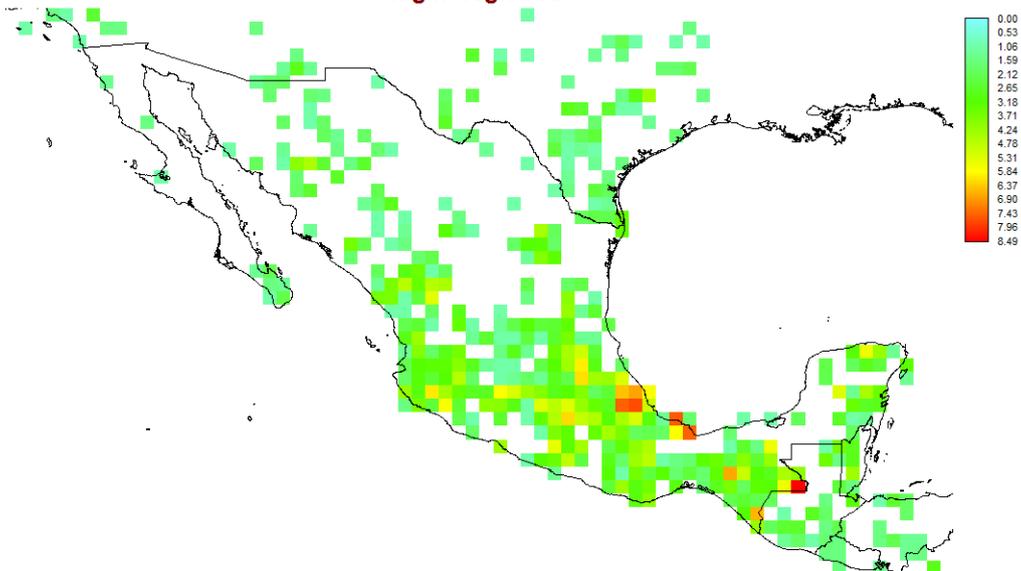


30'

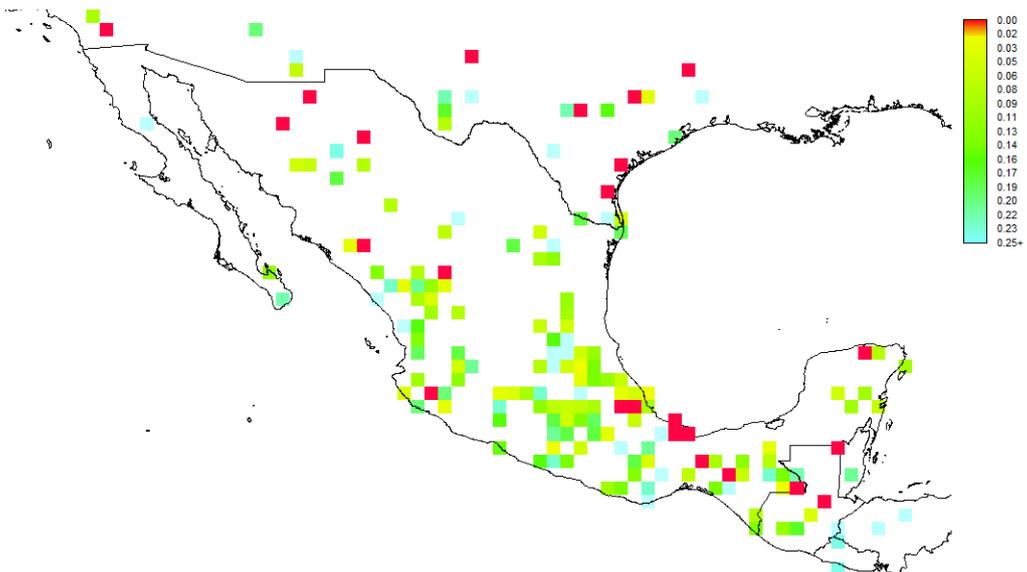
Riqueza especies



log n° registros



Pendiente



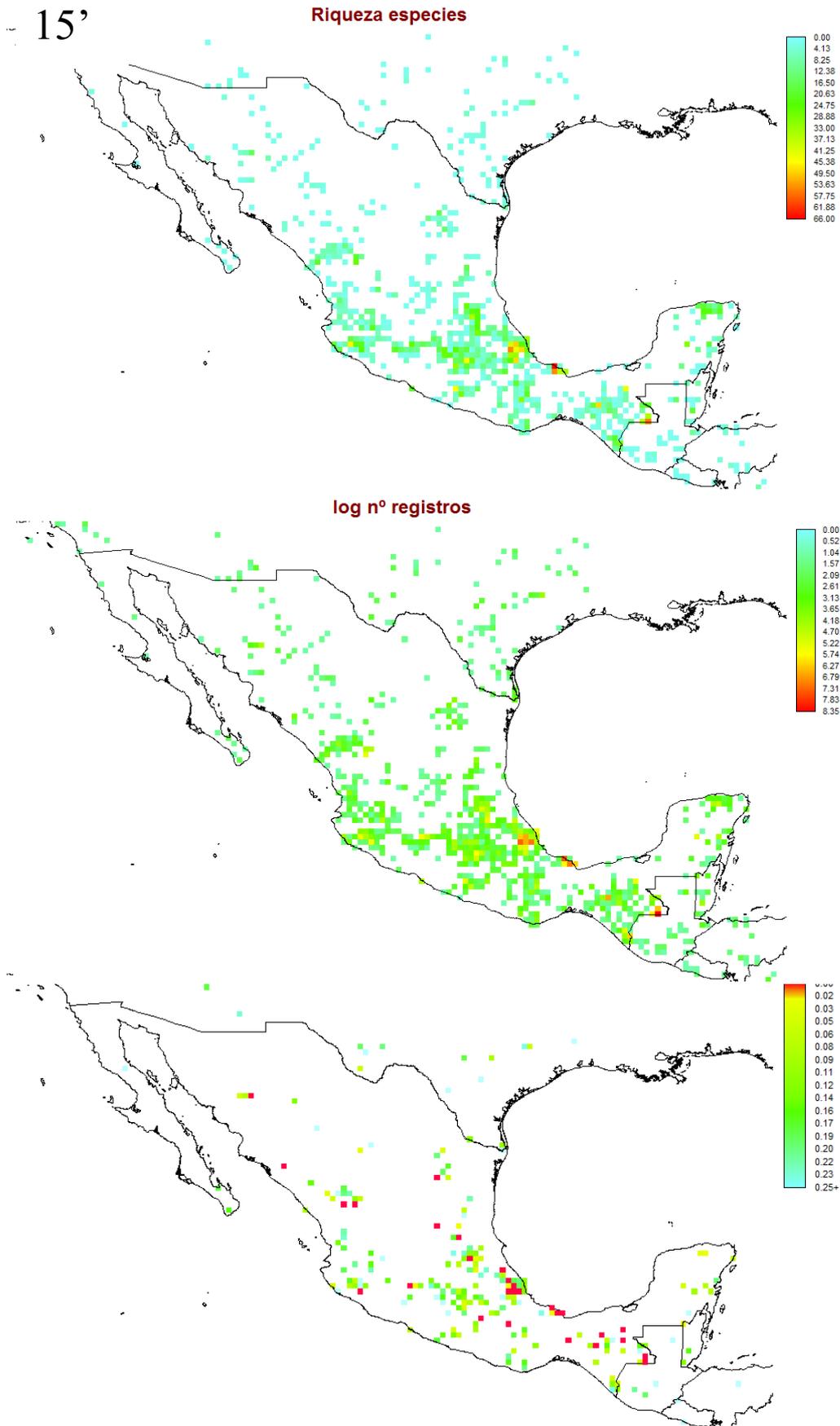


Figura 2.- Distribución de la riqueza observada de Scarabaeoidea coprófagos en México (S), el número de registros (R) y la pendiente de la curva acumulada de la relación S/R según tres diferentes resoluciones (pendientes menores de 0.01 en rojo, CBMs).

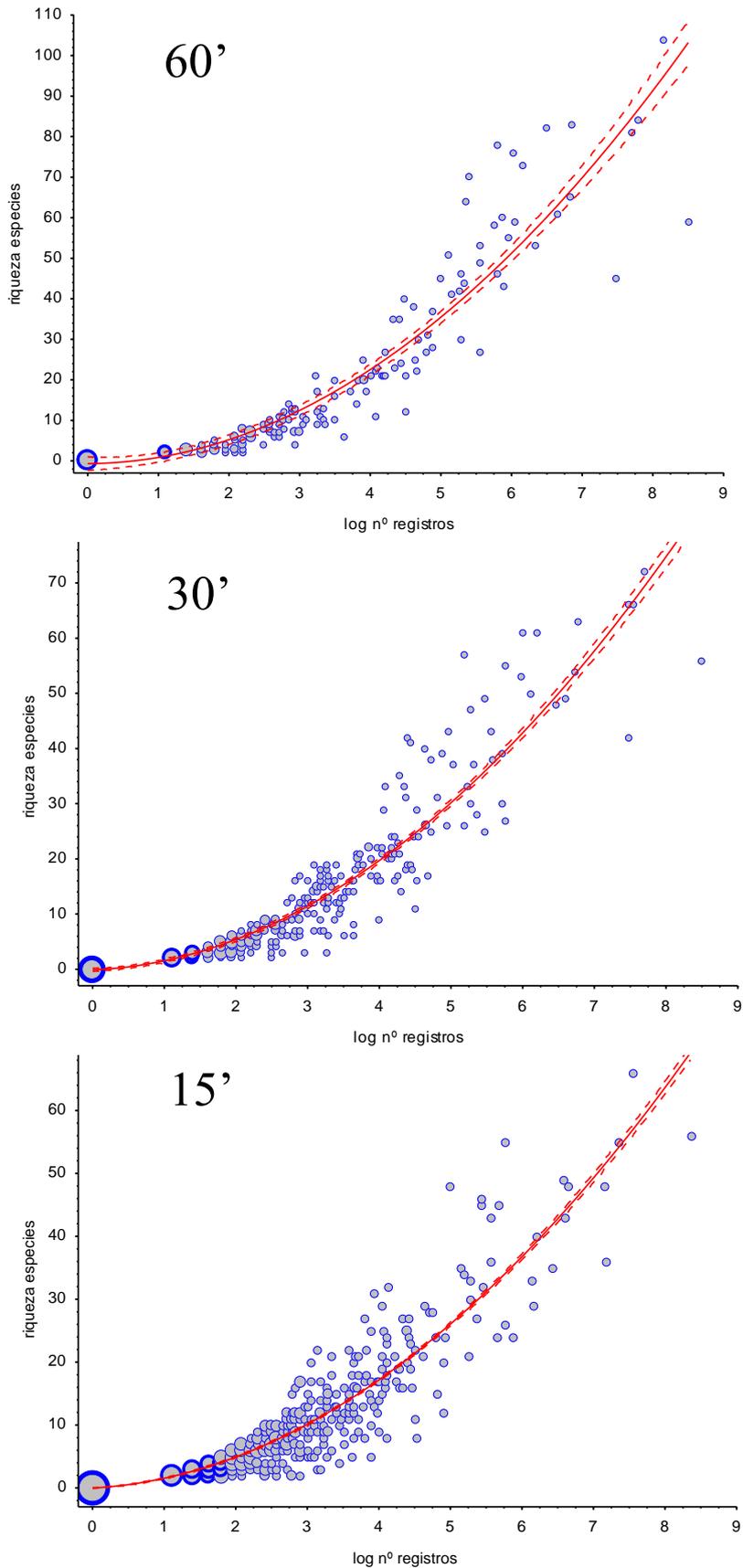


Figura 3.- Relación entre la riqueza y el logaritmo del número de registros de Scarabaeoidea en México según tres diferentes resoluciones. El tamaño de los círculos es proporcional al número de observaciones con el mismo valor. La línea roja representa el ajuste curvilíneo de la relación (función cuadrática).

Afortunadamente, la existencia de grandes carencias de información y el escaso número de áreas con inventarios fiables no ha impedido que exista un conjunto de localidades aparentemente bien muestreadas para el territorio mexicano. Si se examinan las CBMs a la mayor resolución estudiada (15'; Fig. 2) se observa que existen celdas con inventarios fiables por la Sierra Madre Oriental, la llanura costera del Golfo, sobre todo a la altura de Veracruz, la Sierra de Chiapas, el Eje Neovolcánico y la Sierra Madre Occidental. Sin embargo serían necesarios más y mejores inventarios en las penínsulas de California y Yucatán, algunas zonas del Altiplano y la Sierra Madre del Sur. Estas serían las regiones sobre las que habría que focalizar el esfuerzo futuro de colecta en este grupo de organismos. Si se examina esta cuestión desde un punto de vista climático (Fig. 4) se observa que las celdas con inventarios fiables, aunque escasas, representan un importante espectro de las condiciones climáticas existentes en México. Tres principales regiones climáticas estarían sub-representadas y deberían ser objeto de atención en un futuro próximo:

- condiciones templado frías (entre 10°C y 18°C de temperatura media anual) y precipitaciones moderadas (entre los 1000 y 1500 mm de precipitación anual) representadas geográficamente en localidades montañosas ubicadas en la Sierra Madre Occidental, el eje Neovolcánico, la Sierra Madre del Sur y La Sierra de Juárez (ver Fig. 4A).
- Condiciones cálidas (entre 26°C y 28°C) con bajas o moderadas precipitaciones (entre 500 mm y 2000 mm) representadas geográficamente por la depresión del Balsas, el istmo de Tehuantepec, las zonas costeras del Pacífico de Oaxaca y las de Golfo y Yucatán (ver Fig. 4B).
- Condiciones templado-cálidas (entre 16°C y 24°C) con muy bajas precipitaciones (menos de 500 mm). Se trata del área más extensa y con menor representación que se corresponde geográficamente con el Altiplano mexicano, la llanura costera del Pacífico Norte y la península de California (ver Fig. 4C).

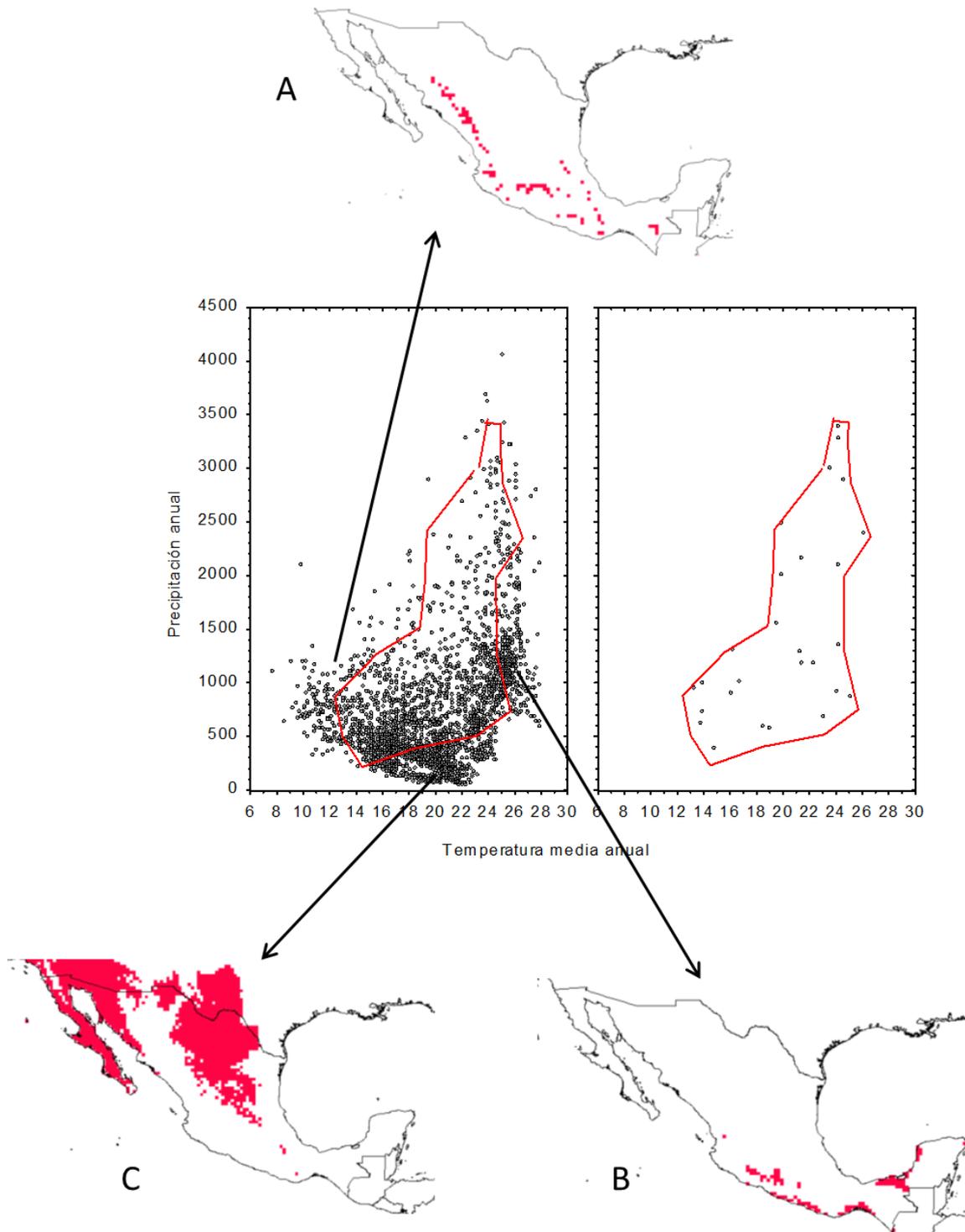


Figura 4.- Características climáticas de todas las celdas de 15' de México utilizando sus valores de temperatura media y precipitación anuales. A la derecha puede observarse la ubicación climática de las celdas con inventarios fiables (CBMs) previamente discriminadas. El polígono rojo representa el área climática abarcada por las CBMs al objeto de observar el espacio ambiental no adecuadamente representado. Los mapas representan las principales áreas con las condiciones climáticas fuera del polígono señalado.

Una manera diferente de confrontar la cuestión sobre adonde dirigir los futuros esfuerzos de colecta, podría ser examinar la localización de las CBMs con el mapa de riqueza total de todos los Scarabaeoidea coprófagos de México, según se deriva de la superposición de todos los modelos de distribución individuales (Fig. 5). En este caso, se observa que serían necesarios más inventarios a lo largo del Eje Neovolcánico, el sur del Altiplano o la Sierra de Juárez, y que los estados más necesitados de prospección serían Oaxaca, Puebla, Hidalgo, Chiapa, Jalisco, Michoacán y Guanajuato.

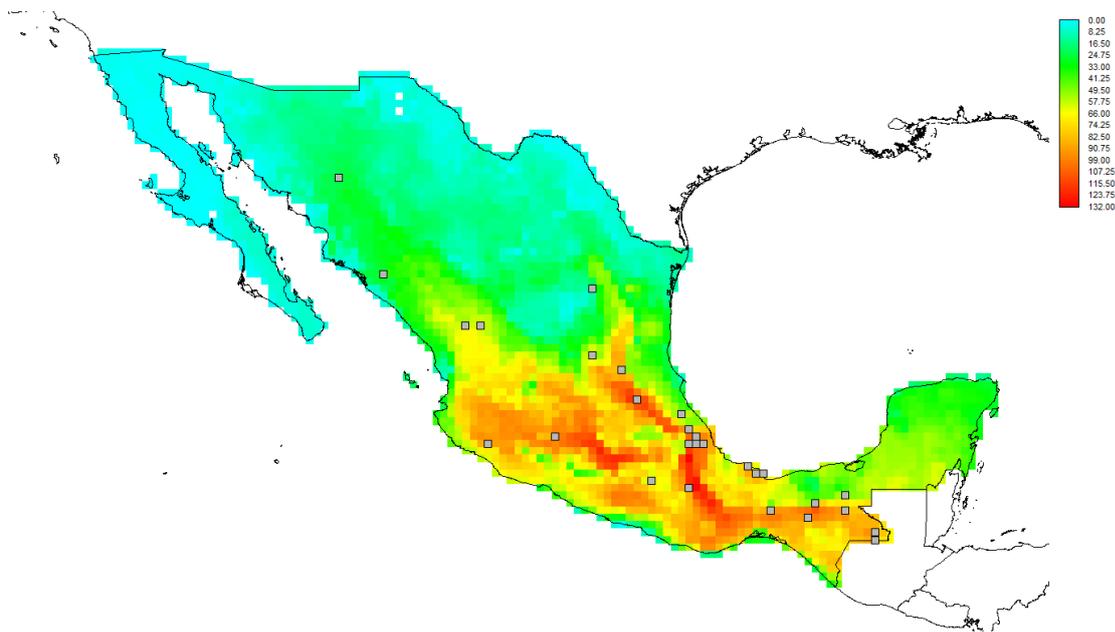


Figura 5.- Distribución de la riqueza total de todos los Scarabaeoidea coprófagos considerados en las celdas de 15' de México, resultado de la superposición de los modelos de distribución individuales, y localización de las celdas con inventarios fiables (CBMs; cuadrados grises).

Podemos considerar, por tanto, que el conocimiento sobre la diversidad y distribución de estas especies apenas podría permitir la realización de algunos estudios biogeográficos y macroecológicos (regionalizaciones, relación con variables ambientales, etc.), utilizando para ello la información de aquellas celdas que tienen mayores probabilidades de ofrecer inventarios fiables. Lo más aconsejable sería planificar la colecta de especímenes y la priorización de futuros esfuerzos, sobre las áreas que aquí han sido detectadas como carentes de suficiente esfuerzo de colecta. Sería recomendable que, para ello, se efectuase un análisis más exhaustivo siguiendo protocolos estandarizados tendentes tanto a maximizar la cobertura ambiental no representada por las áreas “bien muestreadas”, como la representación de las especies de acuerdo a los resultados de los modelos de distribución. En todo caso, los presentes resultados ofrecen ya una visión que, aunque limitada, permite orientar la localización de futuras colectas

desde un punto de vista ambiental y geográfico. Inventarios relativamente fiables en las áreas detectadas como insuficientemente prospectadas, mejorarían la cobertura ambiental y faunística actualmente disponible, permitiendo la elaboración de estudios más precisos y modelos de distribución más fiables.



Malagoniella astyanax

3. Patrones biogeográficos generados

Bolboceratinae

Los modelos de distribución para las 25 especies Bolboceratinae ofrecen una distribución geográfica de la riqueza de especies y del tamaño promedio del área de distribución de las especies presentes en cada celda (Fig. 6). La mayor riqueza de especies se encontraría en las laderas orientales de la Sierra Madre Oriental y en los valles de la Sierra madre del sur y la Sierra de Juárez, así como en la confluencia entre la Sierra Madre del Sur y el eje Neovolcánico. La rareza geográfica, en cambio, es mayor en los confines septentrionales de las llanuras costeras del Atlántico y del Pacífico, así como en la península de California (Fig. 6).

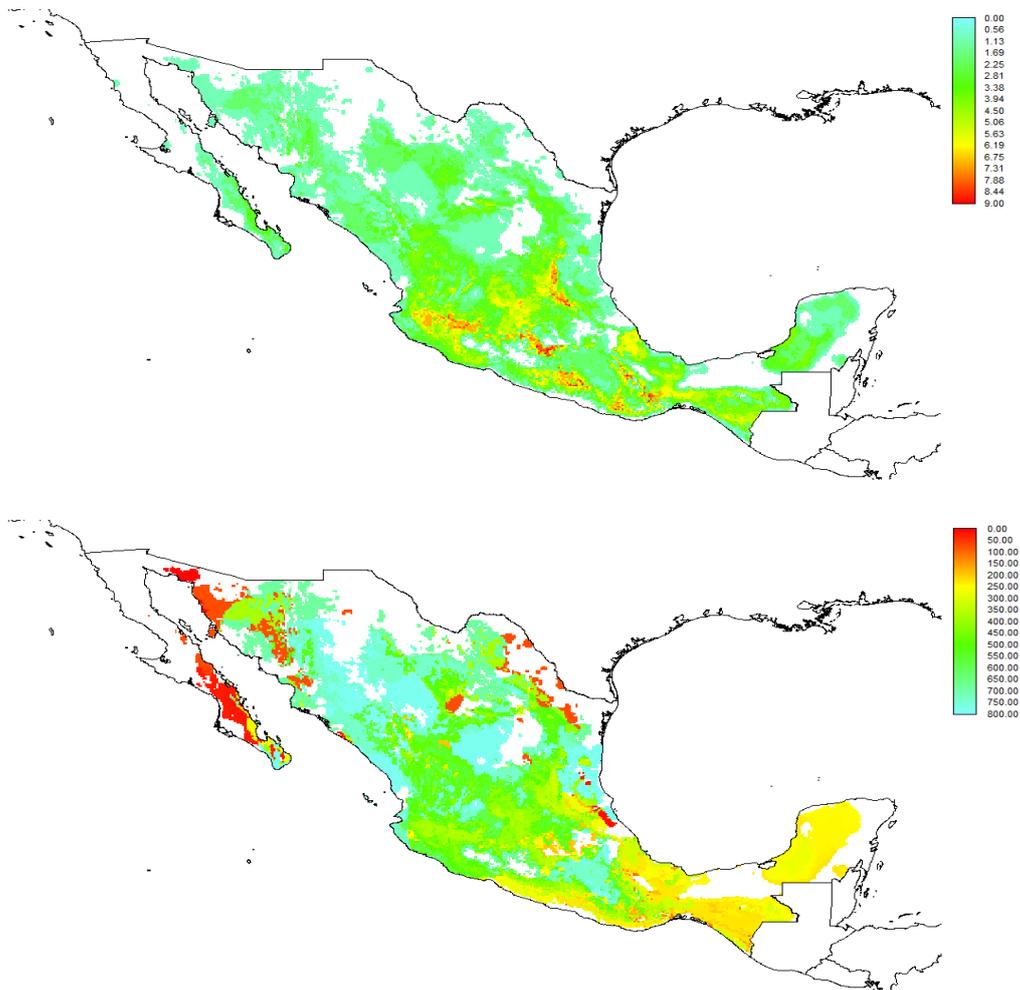


Figura 6.- Distribución de los valores de riqueza de especies (mapa superior) y del tamaño promedio del área de distribución (rareza geográfica; mapa inferior) de los Bolboceratinae en México en celdas de 2' de resolución. Téngase en cuenta que los colores del mapa del tamaño promedio de las áreas de distribución de las especies de cada celda esta invertidos para mejor representar la rareza geográfica.

La relación entre los valores de riqueza de especies y rareza geográfica permiten delimitar cuatro zonas o regiones diferentes según los cuartiles de ambas variables (Fig. 7). Las localidades con una riqueza y rareza elevadas (en rojo) constituirían lo que hemos denominado “áreas principales”, estando geográficamente limitadas a un lado y otro del istmo de Tehuantepec y a La Sierra Madre del Sur (Fig. 7). Esta área principal apenas significaría un 6% del total de la superficie mexicana y sus características ambientales (Cuadro 4) indican que se trata de territorios de altitud media-baja (≈ 600 m) con una temperatura media anual cercana a los 23°C y precipitaciones anuales en torno a los 1500 mm (Cuadro 4).

Cuadro 4.- Características ambientales de las celdas de 2' que constituyen las “áreas principales”. Es decir, aquellos territorios que poseen valores de riqueza superiores al cuartil superior de los datos de riqueza por celda y valores del tamaño promedio del área de distribución inferiores al cuartil inferior (rareza geográfica, en rojo en las figuras 6, 8, 10 y 12). El tamaño del área principal se refiere al porcentaje sobre el total del área mexicana. Q1 es el cuartil superior de los valores de riqueza de especies en las celdas de 2' con alguna presencia de especies. El porcentaje medio antropizado de las celdas de 2' está calculado sobre aquellas que constituyen el área principal.

	Bolboceratinae	Geotrupinae	Aphodiinae	Scarabaeinae
N° medio especies en celdas de 2'	1,68±0.02	1,02±0.01	4,78±0.05	18,20±0.18
N° especies máximo en celdas 2'	9	8	27	79
Q1	3	3	7	36
Tamaño del área principal	5,69%	8,03%	18,23%	15,47%
Altitud	620±23	1988±18	1904±11	1101±17
Temp media anual	23,2±0,1	16,1±0,1	16,8±0,1	21,05±0,1
Temp máxima mes más cálido	33.5±0,1	28,1±0,1	28,6±0,1	32,0±0,1
Temp mínima mes más frío	12,8±0,1	4,2±0,1	5,1±0,1	10,3±0,1
Prec anual	1464,3±22,7	872,7±8,1	873,9±7,6	1325,9±15,5
Prec mes más seco	0,2±0,1	0,0±0,0	0,02±0,01	0,9±0,1
Prec mes más húmedo	74,8±1,1	50,6±0,5	49,9±0,4	67,6±0,6
% medio antropizado	34,0±1,3	19,3±1,0	17,5±0,5	27,3±0,8

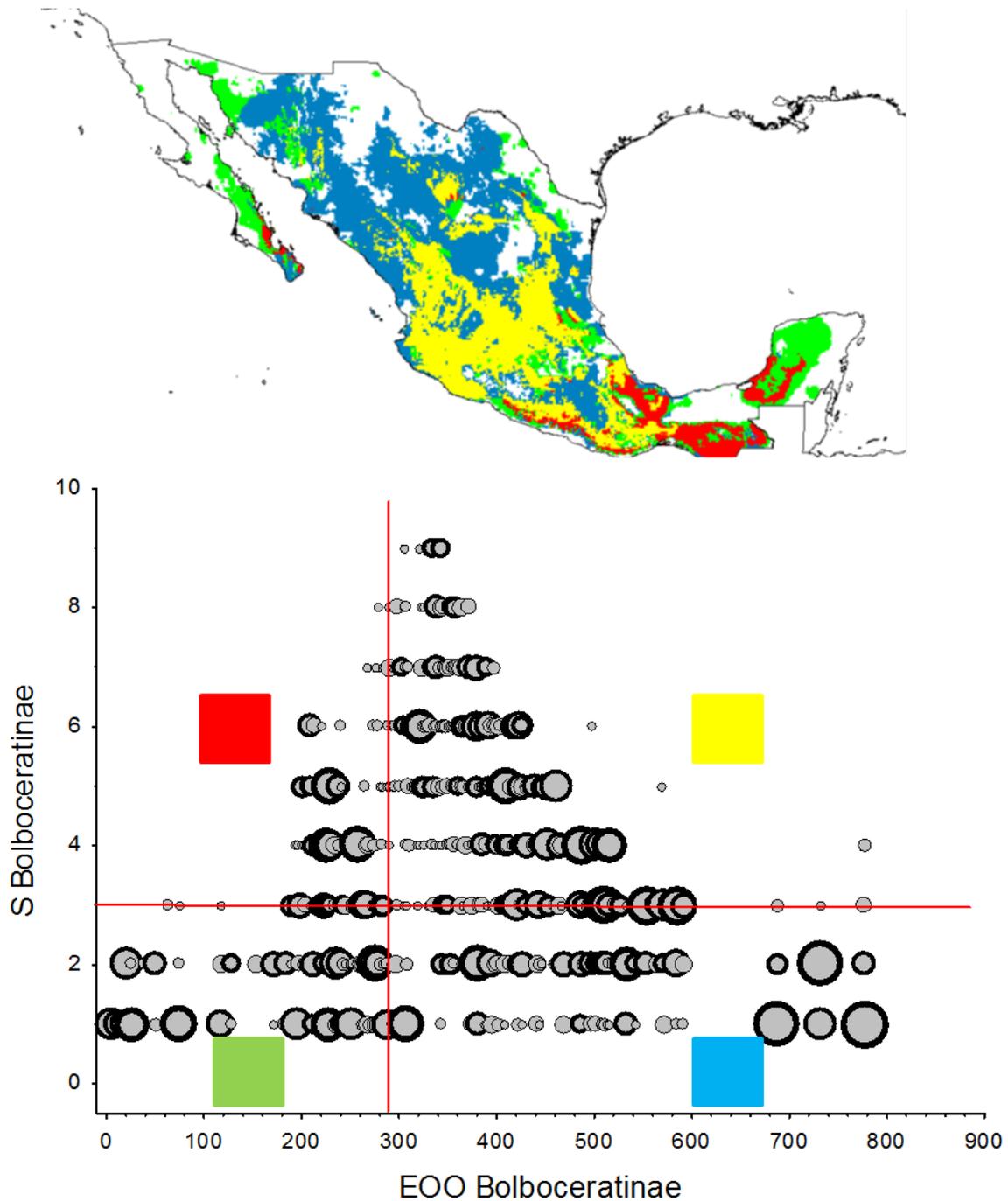


Figura 7.- Relación entre el tamaño promedio del área de distribución (EOO) y la riqueza de especies (S) de los Bolboceratinae en las celdas de 2' de resolución de México. El gráfico está dividido en cuatro cuadrantes según el cuartil superior de los datos de riqueza y el inferior de los datos de tamaño de área promedio (rareza geográfica). La ubicación geográfica de los datos pertenecientes a cada cuadrante se muestra en el mapa superior de acuerdo a los colores que los identifican.

Geotrupinae

Los modelos de distribución para las 19 especies de Geotrupinae ofrecen una distribución geográfica de la riqueza de especies de Geotrupinae y del tamaño promedio su área de distribución (Fig. 8) muy diferente de la subfamilia Bolboceranicae con la que esta filogenéticamente relacionada. La mayor riqueza estaría ubicada en los valles centrales del eje Neovolcánico y algunas zonas de la Sierra de Juárez, mientras que la rareza geográfica sería mayor en la fachada Atlántica meridional de la Sierra Madre Occidental, sur de la Sierra Madre del Sur y algunas áreas de Chiapas.

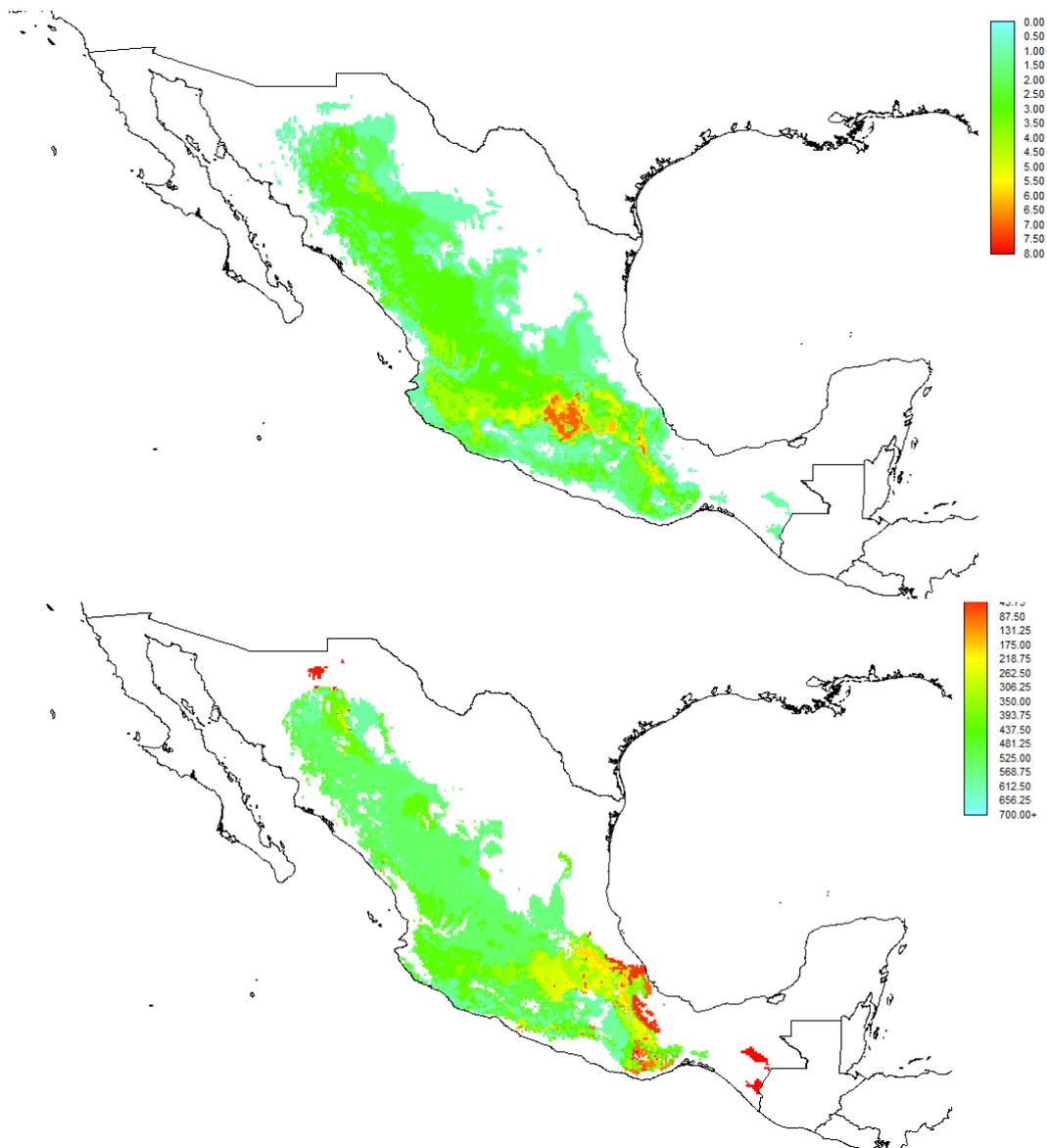


Figura 8.- Distribución de los valores de riqueza de especies (mapa superior) y del tamaño promedio del área de distribución (rareza geográfica; mapa inferior) de los Geotrupinae en México en celdas de 2' de resolución. Téngase en cuenta que los colores del mapa del tamaño promedio de las áreas de distribución de las especies de cada celda esta invertidos para mejor representar la rareza geográfica.

Las cuatro zonas delimitadas por los valores de riqueza de especies y rareza geográfica (Fig. 9) indican que el tamaño total de las áreas principales es ligeramente mayor que en el caso de los Bolboceratinae (8%; Cuadro 4), estando geográficamente limitadas al norte de los valles interiores de la Sierra Madre Occidental, el Eje Neovolcánico, la Sierra Madre del Sur y la Sierra de Juárez, siempre por encima del istmo de Tehuantepec (Fig. 9). El área principal para los Geotrupinae poseería una altitud elevada (≈ 2000 m), así como temperaturas y precipitaciones medias significativamente inferiores que las de su grupo hermano, los Bolboceratinae (Cuadro 4).

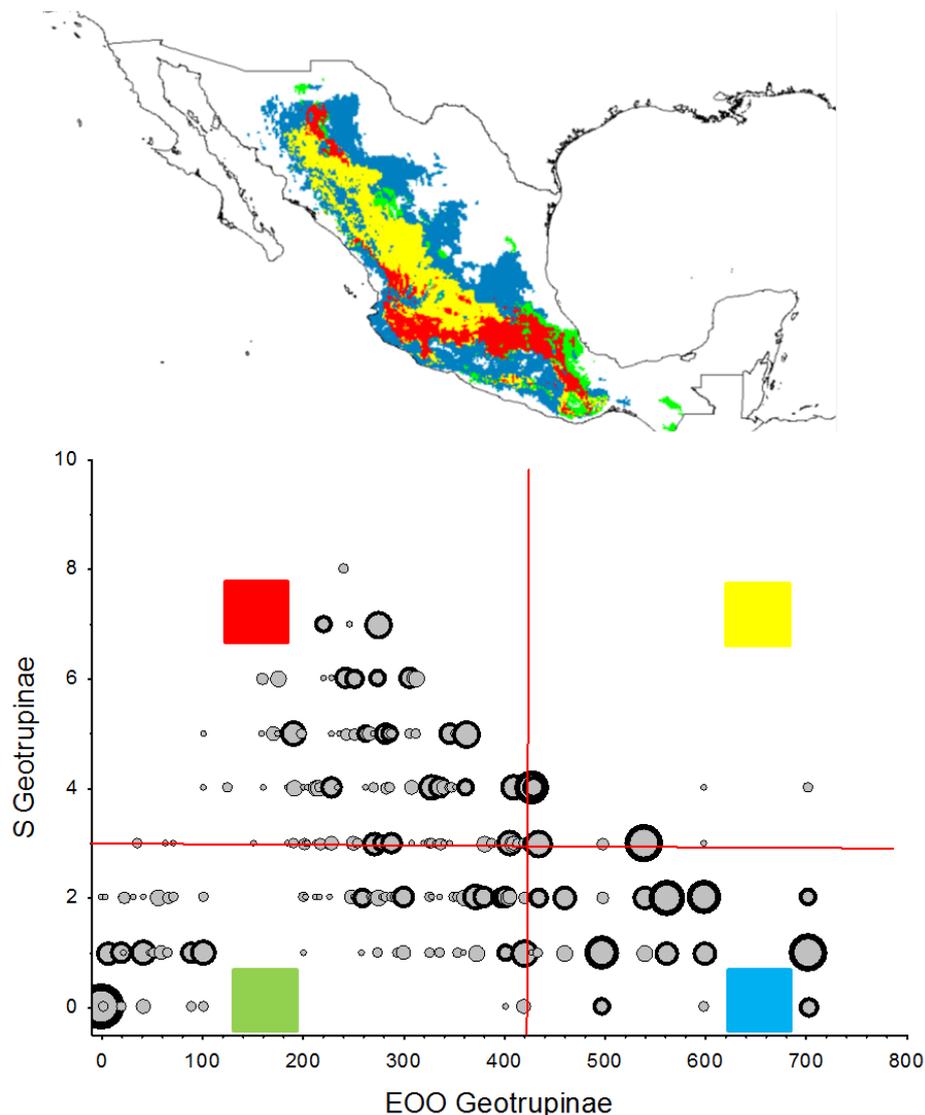


Figura 9.- Relación entre el tamaño promedio del área de distribución (EOO) y la riqueza de especies (S) de los Geotrupinae en las celdas de 2' de resolución de México. El grafico está dividido en cuatro cuadrantes según el cuartil superior de los datos de riqueza y el inferior de los datos de tamaño de área promedio (rareza geográfica). La ubicación geográfica de los datos pertenecientes a cada cuadrante se muestra en el mapa superior de acuerdo a los colores que los identifican.

Aphodiinae

Los modelos de distribución de las 108 especies de Aphodiinae nativas reconocidas para México ofrecen una distribución geográfica de la riqueza de especies de relativamente parecida a la de los Geotrupinae, estando la mayor riqueza de especies ceñida a la Altiplanicie mexicana, el Eje Neovolcánico y el sur de la Sierra Madre Oriental (Fig. 10). Las variables ambientales y la altitud promedio de las celdas de 2' poseen también valores similares a los de los Geotrupinae; altitudes medias de alrededor de 1900 m, temperaturas medias anuales cercanas a los 17°C y precipitaciones anuales en torno a los 870 mm (Cuadro 4).

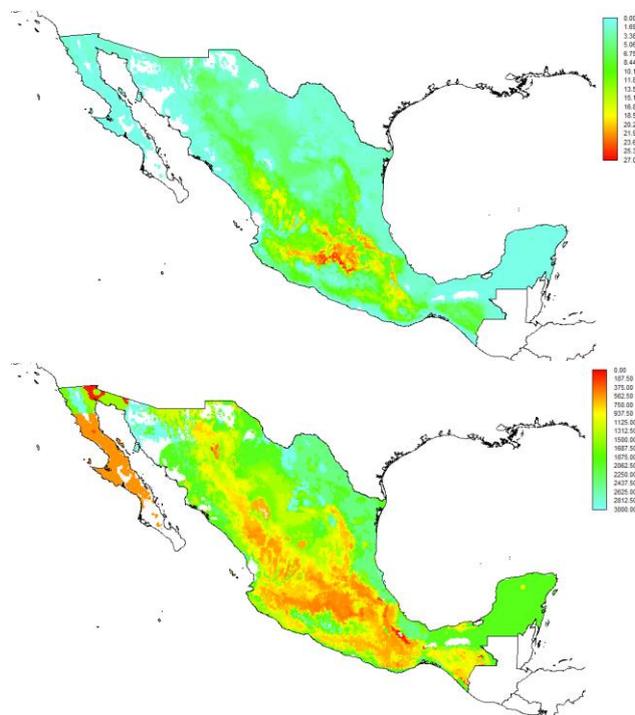


Figura 10.- Distribución de los valores de riqueza de especies (mapa superior) y del tamaño promedio del área de distribución (rareza geográfica; mapa inferior) de los Aphodiinae en México en celdas de 2' de resolución. Téngase en cuenta que los colores el mapa del tamaño promedio de las áreas de distribución de las especies de cada celda esta invertidos para mejor representar la rareza geográfica.

La variación geográfica de la rareza geográfica muestra un patrón mucho más repartido, de modo que, aunque la mayor rareza aparece en la vertiente atlántica de la Sierra de Juárez y en Sonora, podemos encontrar rarezas moderadamente altas por la Sierra Madre Occidental, la Sierra Madre Oriental, la Altiplanicie mexicana, el Eje Neovolcánico, la Sierra Madre del Sur o la península de California. Como consecuencia, el “área principal” de los Aphodiinae abarca una extensión de más del 18% del total de la superficie del territorio mexicano (Cuadro 4), estando ésta limitada a buena parte del territorio con altos valores de rareza antes mencionados (Fig. 11).

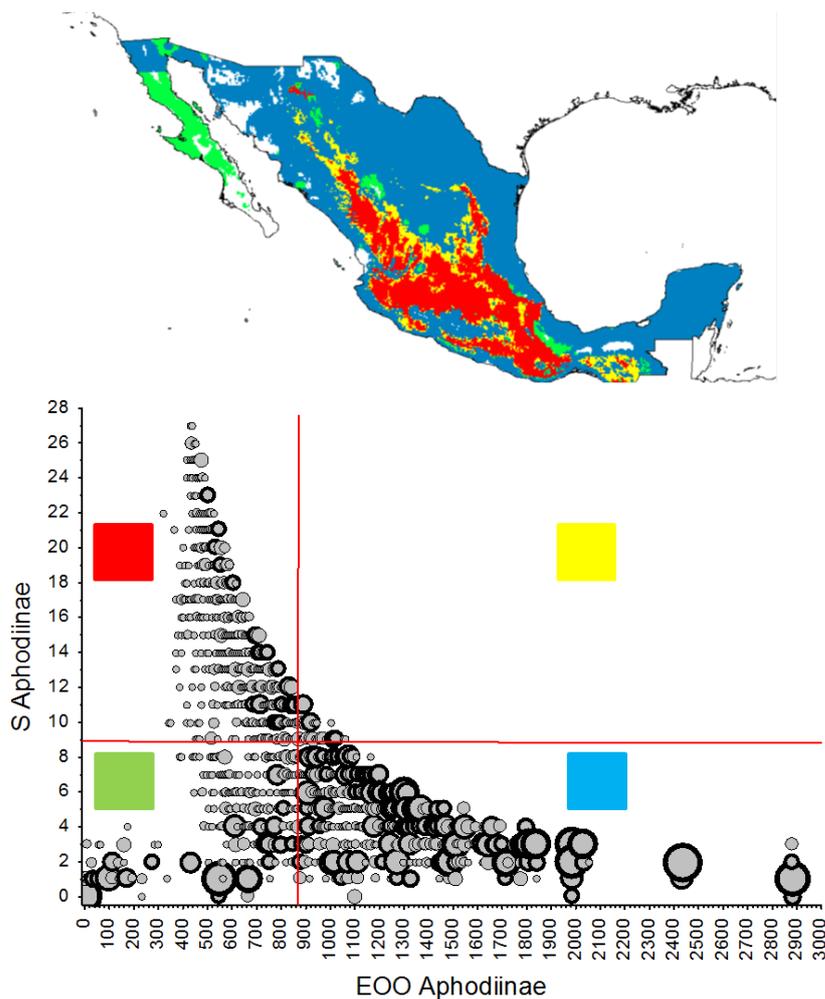


Figura 11.- Relación entre el tamaño promedio del área de distribución (EOO) y la riqueza de especies (S) de los Aphodiinae en las celdas de 2' de resolución de México. El grafico está dividido en cuatro cuadrantes según el cuartil superior de los datos de riqueza y el inferior de los datos de tamaño de área promedio (rareza geográfica). La ubicación geográfica de los datos pertenecientes a cada cuadrante se muestra en el mapa superior de acuerdo a los colores que los identifican.

Scarabaeinae

Según los modelos de distribución realizados para las 268 especies de Scarabaeinae nativas reconocidas para el territorio mexicano, las celdas con mayor riqueza de especies poseerían, en el caso de los Scarabaeinae, una distribución mucho más meridional que en todos los otros grupos estudiados, con grandes riquezas más allá del istmo de Tehuantepec. El propio istmo, la Sierra de Chiapas, la depresión del Balsas y la montaña Veracruzana serían las áreas más ricas en especies (Fig. 12), mientras que las mayores rarezas estarían localizadas en el noroeste árido y los valores de rareza intermedio-altos se repartirían por buena parte del territorio (Fig. 12). Debido a ello, las “áreas principales” abarcan casi un 16% del territorio mexicano y se reparten por casi toda la mitad sur de México (Fig. 13) en zonas de altitud moderada (alrededor de los 1100 m), con temperaturas medias anuales cercanas a los 21°C y precipitaciones anuales en torno a los 1300 mm (Cuadro 4).

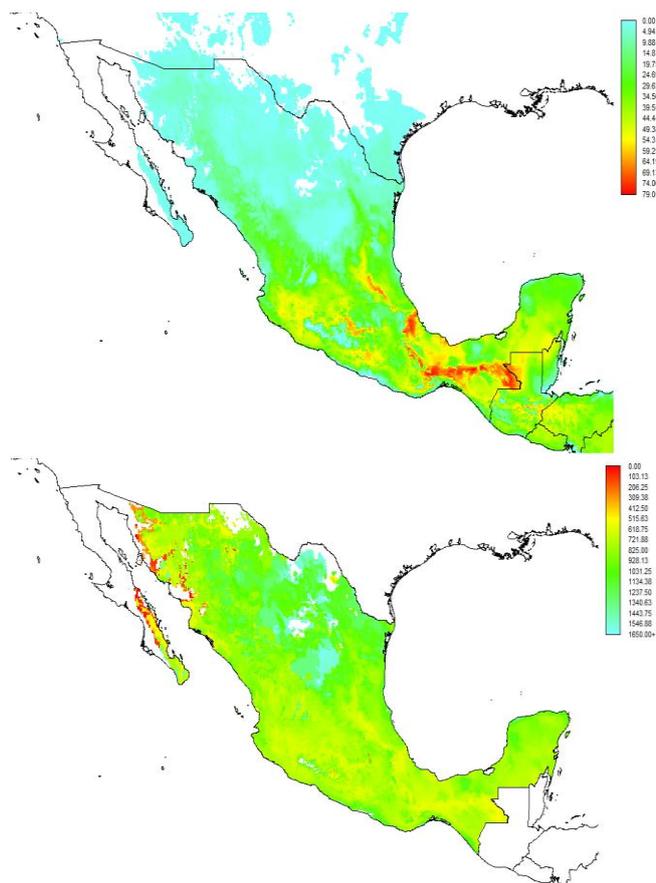


Figura 12.- Distribución de los valores de riqueza de especies (mapa superior) y del tamaño promedio del área de distribución (rareza geográfica; mapa inferior) de los Scarabaeinae en México en celdas de 2' de resolución. Téngase en cuenta que los colores el mapa del tamaño promedio de las áreas de distribución de las especies de cada celda esta invertidos para mejor representar la rareza geográfica.

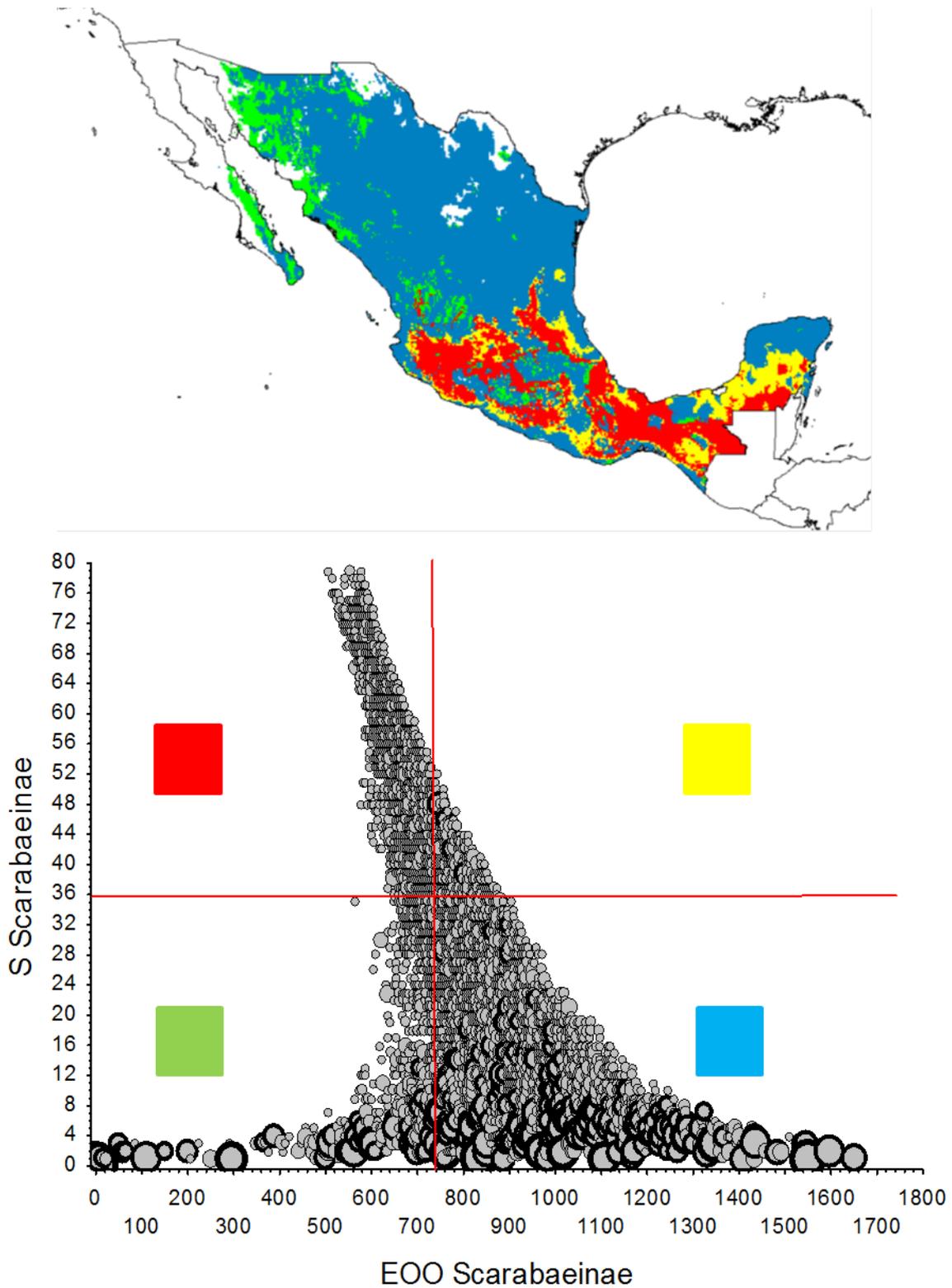


Figura 13.- Relación entre el tamaño promedio del área de distribución (EOO) y la riqueza de especies (S) de los Scarabaeinae en las celdas de 2' de resolución de México. El gráfico está dividido en cuatro cuadrantes según el cuartil superior de los datos de riqueza y el inferior de los datos de tamaño de área promedio (rareza geográfica). La ubicación geográfica de los datos

pertenecientes a cada cuadrante se muestra en el mapa superior de acuerdo a los colores que los identifican.

4. ¿Son fiables las predicciones realizadas?

El proceso de modelización anteriormente explicado impide que cualquier localidad de presencia no aparezca dentro de área predicha (no hay errores de omisión), de modo que los únicos errores que es posible cometer son los errores de comisión (incluir presencias en zonas en las que la especie no se encuentra presente). Sin embargo, como no se posee un conocimiento fiable sobre las ausencias (localidades en las que la especie no se encuentra presente tras la realización de un esfuerzo de colecta suficiente) o, en todo caso, como la cantidad de ausencias probables es muy escasa y sesgada como se demostró en la primera parte de este informe, es difícil realizar una evaluación de la importancia y magnitud de estos errores de comisión. Esta situación se afronta en los modelos mal denominados de “solo presencia” como MaxEnt, minimizando la “fractional predicted área” (el porcentaje del área total predicha como presencia) en vez de los errores de comisión. En otras palabras, se busca el área de menor tamaño que incluya el mayor número de puntos de presencia en un proceso que, inevitablemente, tiende a generar resultados sospechosamente parecidos a los datos de presencia utilizados para alimentar el modelo (Aarts *et al.*, 2012) y medidas de validación erróneas (Lobo *et al.*, 2008) ya que basta aumentar el área de análisis respecto al área del organismo para obtener bajos errores de comisión. Este proceso trata de evitar la generación de modelos que extrapolen la presencia de la especie más allá del territorio con observaciones y, en nuestro caso, esta posibilidad se restringe mediante la delimitación de un *spatial background* al inicio del proceso de modelización y también mediante la eliminación de aquellas áreas de presencia probable desconectadas o que se encuentran en una cuenca hidrográfica sin presencias observadas (ver apartado de construcción de los modelos).

De todo lo anterior se deduce que, sin remedio, los resultados de estos modelos solo pueden validarse recabando nueva información independiente, sobre todo cuando la escasez de información disponible impide dividir los datos existentes para efectuar la validación con una parte de ellos (validación cruzada o cross-validation), y/o cuando el sesgo en los datos de partida provoca que cualquier validación cruzada tienda a sobreestimar la precisión del modelo. Los modelos de distribución como los aquí realizados, sobre grupos híper-diversos y con muchas carencias y sesgos en los datos disponibles, deben considerarse hipótesis de distribución provisionales a validar y mejorar con la adquisición de nuevos datos faunísticos.

En nuestro caso hemos utilizado los modelos individuales obtenidos para cada una de las especies, al objeto de generar inventarios de las especies conocidas y predichas para los siete estados más al sur de México: Oaxaca, Chiapas, Quintana Roo, Campeche, Tabasco, Yucatán y Veracruz (Fig. 14). Sin ninguna duda, el estado con mejor conocimiento faunístico de esta región, y probablemente de todo México, es Veracruz, ya que en él hay más de 13 veces el número de registros que en el estado siguiente con más registros y casi 500 veces más registros que el estado menos prospectado. En el caso de Veracruz, los modelos incrementan en un 23% el ya de por sí gran número de especies conocidas (163 especies). Estas especies adicionales que podrían incluirse en un futuro en el catálogo del estado poseen, en su inmensa mayoría, distribuciones inferiores a los 2.500 km² (una celda de 50 x 50 km de lado). De este modo, puede decirse que los errores de comisión cometidos por los modelos realizados son, para este estado, moderados e implicarían a especies de distribución restringida o marginal según las condiciones ambientales del estado.

Sin embargo, el panorama es completamente diferente en el resto de los estados examinados. En los estados con un menor esfuerzo de colecta (Tabasco, Campeche y Quintana Roo) que, indudablemente, es insuficiente para establecer un catálogo fiable, los porcentajes de incremento en el número de especies oscilan desde el 780% y el 95%, siendo la cantidad de especies adicionales geográficamente raras mucho menor (26%, 6% y 7%). En el resto de los estados (Oaxaca y Chiapas), con un esfuerzo de colecta mayor, el porcentaje de incremento en el número de especies es también menor, poseyendo la mitad de las especies adicionales un área de distribución equivalente a una celda de 100 x 100 km. De este modo, el patrón de variación en la riqueza de especies mostraría una gradiente este-oeste en el que los estados de la península de Yucatán, menos heterogéneos ambiental y topográficamente, poseerían menores riquezas y aquellos otros occidentales o con mayor heterogeneidad serían más ricos en especies. En el caso especial de los Aphodiinae el número de especies adicionales que pueden encontrarse en un futuro será, probablemente, mucho mayor que predicho según los modelos de distribución. Se trata de especies que poseen un bajo número de registros y localidades (ver Cuadro 2), de modo que sus rangos predichos son inevitablemente menores. Muestreos futuros mostrarán si las especies que supuestamente poseen condiciones climáticas adecuadas en cada estado se encuentran efectivamente en ellos.

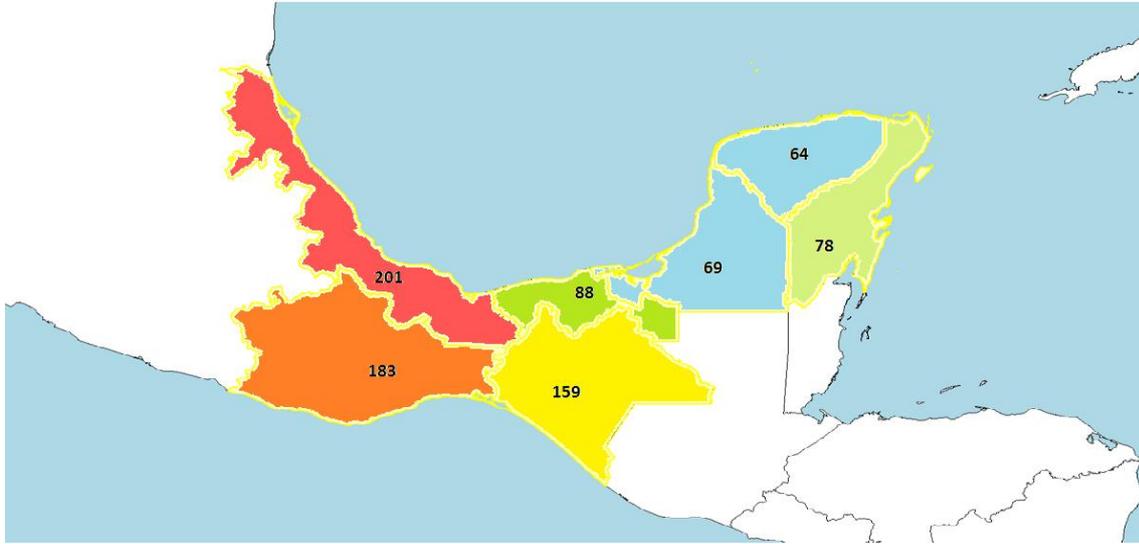


Figura 14.- Riqueza de Scarabaeoidea coprófagos predicha por los modelos para los siete estados más al sur de México.



Canthon indigaceus

5. Estimación de la vulnerabilidad ante los cambios de usos del suelo

En este apartado se ha tratado de estimar el efecto que las alteraciones antrópicas han podido producir sobre las distribuciones predichas de cada una de las especies. Asumiendo que los resultados de los modelos predictivos reflejan el área de distribución climáticamente favorable dentro del territorio accesible, hemos utilizado la reciente cartografía mundial de usos del suelo a 1 km² de resolución EarthEnv, proporcionada por Tuanmu & Jetz (2014), para estimar cual es la disminución probable en el área de distribución de cada una de las especies cuando se eliminan los usos del suelo antropizados (cultivated and managed vegetation + urban/built-up). EarthEnv es, en realidad, una cartografía basada en el consenso de cuatro distintas fuentes que miden la cobertura del suelo (land cover) mediante datos de sensores satelitales (ver detalles en Tuanmu & Jetz, 2014) y que proporciona valores que oscilan entre 0 y 100 reflejando el porcentaje de acuerdo entre las distintas fuentes de datos en la cobertura elegida. Puede decirse así, que los valores son una medida de las probabilidades de detectar el uso del suelo elegido en una celda de 1 km² (ver Fig. 15).

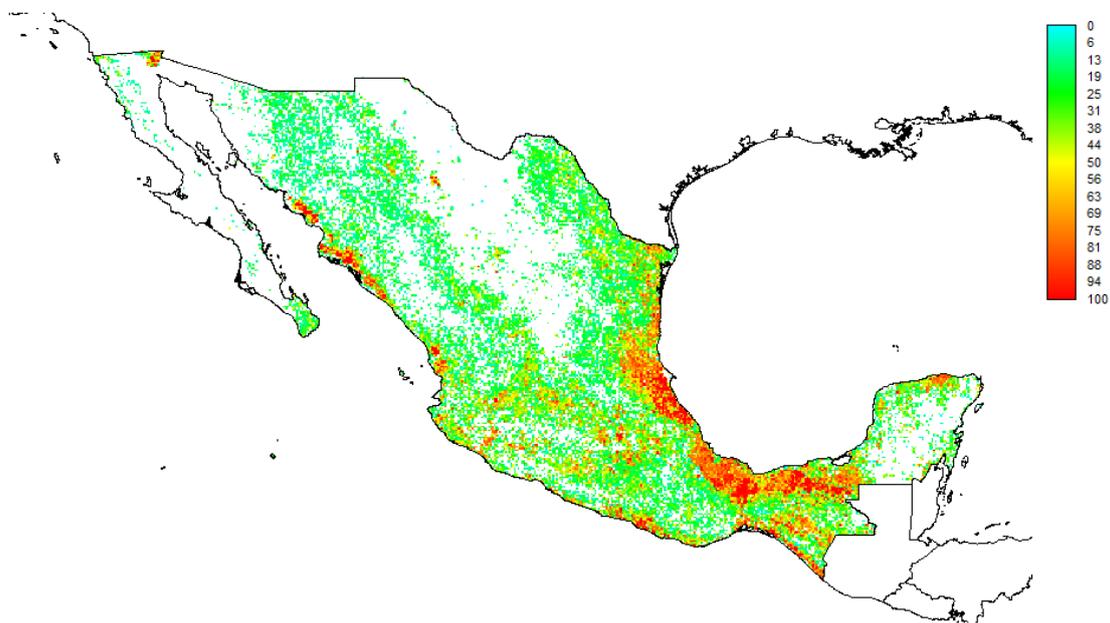


Figura 15.- Cobertura o probabilidad de presencia de suelos antropizados (cultivos, urbano y vegetación manejada) para México según Earth Env.

Como la cuestión clave a la hora de eliminar las áreas antropizadas de las zonas climáticamente favorables, es decidir cuál es el umbral en el grado o probabilidad de antropización por encima del cual no sería posible la persistencia de las especies, se ha estimado la disminución del área según 10 umbrales diferentes (1%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% y 100%). Es decir, en el caso del umbral 100%, se eliminan del área de distribución predicha todas aquellas celdas de 2' (aproximadamente, 4 x 4 km) que con certeza están antropizadas. En el caso del umbral inferior se eliminan todas las celdas con una cobertura antropizada igual o superior al 1%, considerándose que este porcentaje mínimo de antropización sería suficiente para impedir la persistencia de la especie.

Utilizando el umbral del 100% y eliminando las celdas de 2' de resolución con el total de su superficie antropizada, no desaparecería ninguna especie y solo tres perderían más de la mitad de su área de distribución: *Liothorax alternatus*, *Coelotrachelus venustus* y *C. ixtapalapaensis*. Las mismas especies perderían más de la mitad de su área de distribución, si se considerasen inadecuadas para la supervivencia aquellas celdas con más del 80% de su superficie antropizada. Quince especies (apenas un 3.6 % del total) perderían más de la mitad de su área de distribución, en el caso de que fuera suficiente una antropización de la mitad del territorio de cada celda para impedir la persistencia de las especies (Fig. 16). En realidad, incluso ante un escenario negativo en el que se asuma que bastaría un 20% de antropización para impedir la persistencia de las especies, solo desaparecería completamente el área de distribución de cuatro especies (las tres anteriormente mencionadas y *Onthophagus eulophus*, pero otras 191 especies (45% del total) habrían perdido, al menos, la mitad de su área de distribución (Fig. 16). Si consideramos que estas especies serían las que en mayor grado han sufrido disminuciones en su área de distribución, como consecuencia de las alteraciones humanas, y también las que serían más propensas a sufrir disminuciones futuras (en adelante, especies vulnerables), se observa que más de las tres cuartas partes de ellas son Scarabaeinae (152 de 195), que más de la mitad de las especies de esta subfamilia podrían considerarse vulnerables, y que el porcentaje de pérdida de área de distribución es significativamente mayor en esta subfamilia que en los otros grupos ($F_{2, 418} = 16.0$; $p < 0.0001$).

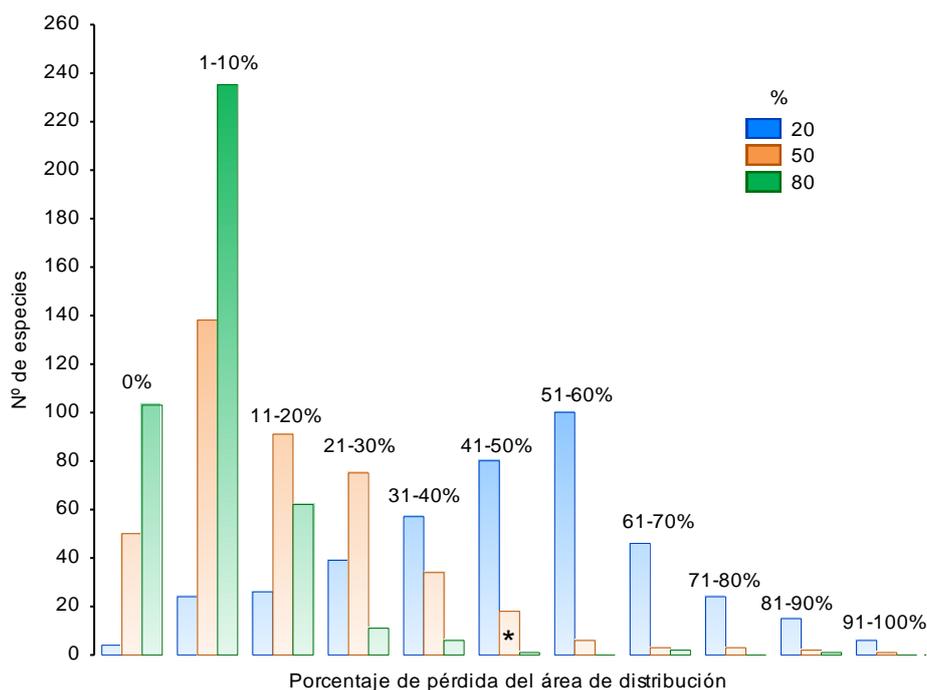
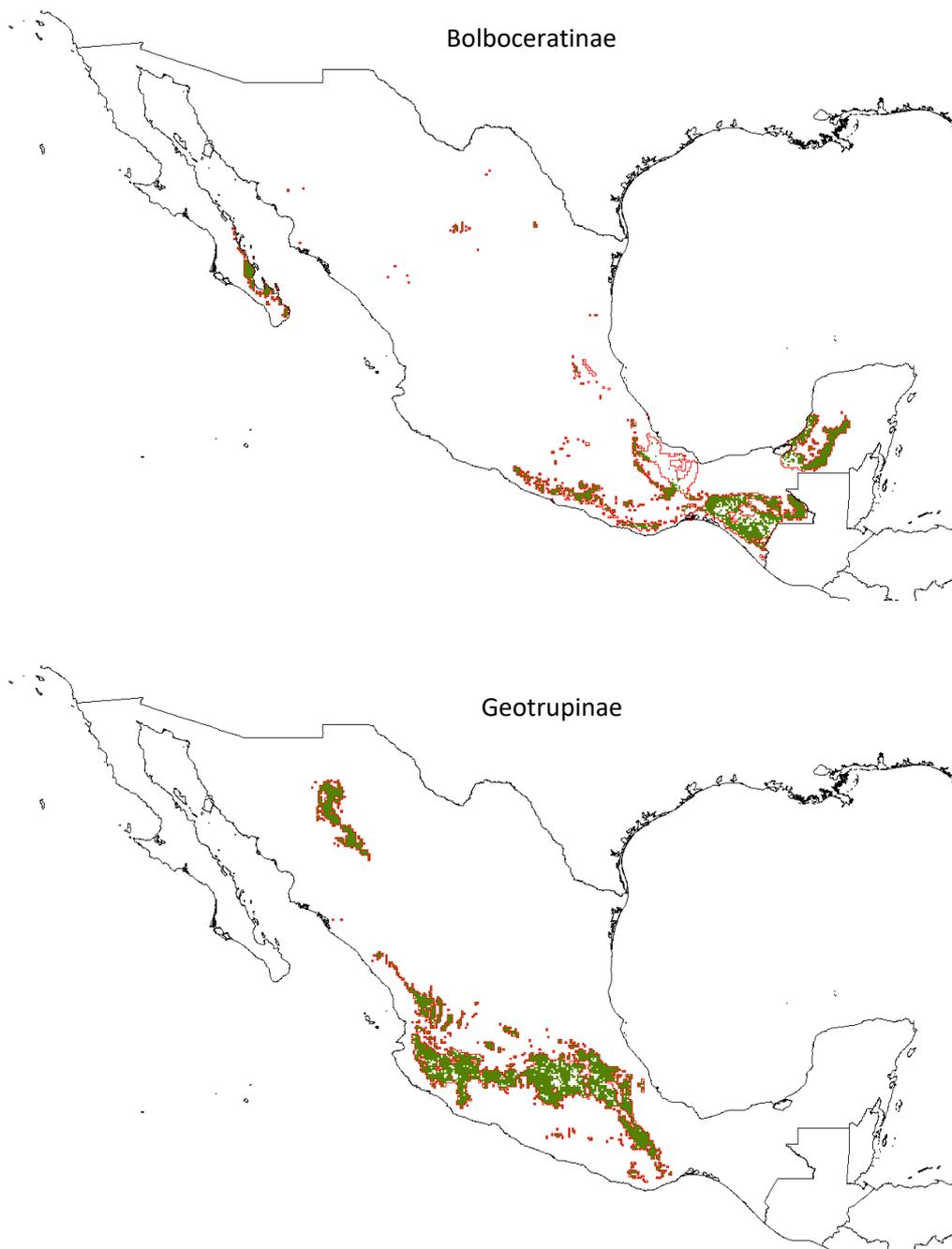


Figura 16.- Número de especies de Scarabaeoidea que perderían porcentajes crecientes de su área de distribución predicha (desde nada-0%- hasta prácticamente toda su área - 91-100%) según tres escenarios diferentes (20, 50 y 80 %) en los que bastaría ese porcentaje de antropización en una celda de 2' para que desapareciese la especie. Por ejemplo, si se considera que cuando la mitad de la superficie de una celda está antropizada no es posible la subsistencia de las especies (color naranja), apenas 15 especies perderían el 50% de su área de distribución potencial original (situación marcada con un asterisco).

Si consideramos las “áreas principales” de cada grupo taxonómico anteriormente estimadas en base a los datos de riqueza y rareza (ver Figs. 7, 9, 11 y 13), podemos observar que las mayores reducciones en su área se dan en el caso de los Bolboceratinae y, en menor medida, entre los Scarabaeinae. En el caso de los Bolboceratinae estas pérdidas de área están agrupadas principalmente en las zonas costeras atlánticas del sureste de la Sierra Madre Oriental, mientras que en los Scarabaeinae se sitúan en el istmo de Tehuantepec (Fig. 17). Tanto en Aphodiinae como en Geotrupinae la alteración antrópica del paisaje habría afectado en menor medida a las áreas de mayor importancia para la diversidad de estos grupos, no encontrándose un patrón de agregación o coincidencia entre las áreas antropizadas y las importantes desde el punto de vista de su diversidad.

En conclusión, la actual alteración del paisaje natural propiciada por el hombre habría disminuido la extensión y fragmentado el área de distribución original de estas especies, según las predicciones realizadas con nuestros modelos que buscan estimar el área climáticamente favorable y accesible. Estas pérdidas habrían sido provocadas por una coincidencia entre las áreas favorables para el desarrollo humano y las áreas favorables para la pervivencia de algunas especies. En todo caso, estas pérdidas variarían entre especies y grupos, pero parecen haber afectado en mayor grado al grupo de Scarabaeoidea coprófagos más diversificado en el territorio mexicano, los Scarabaeinae, y también a las áreas más diversas para los Bolboceratinae.



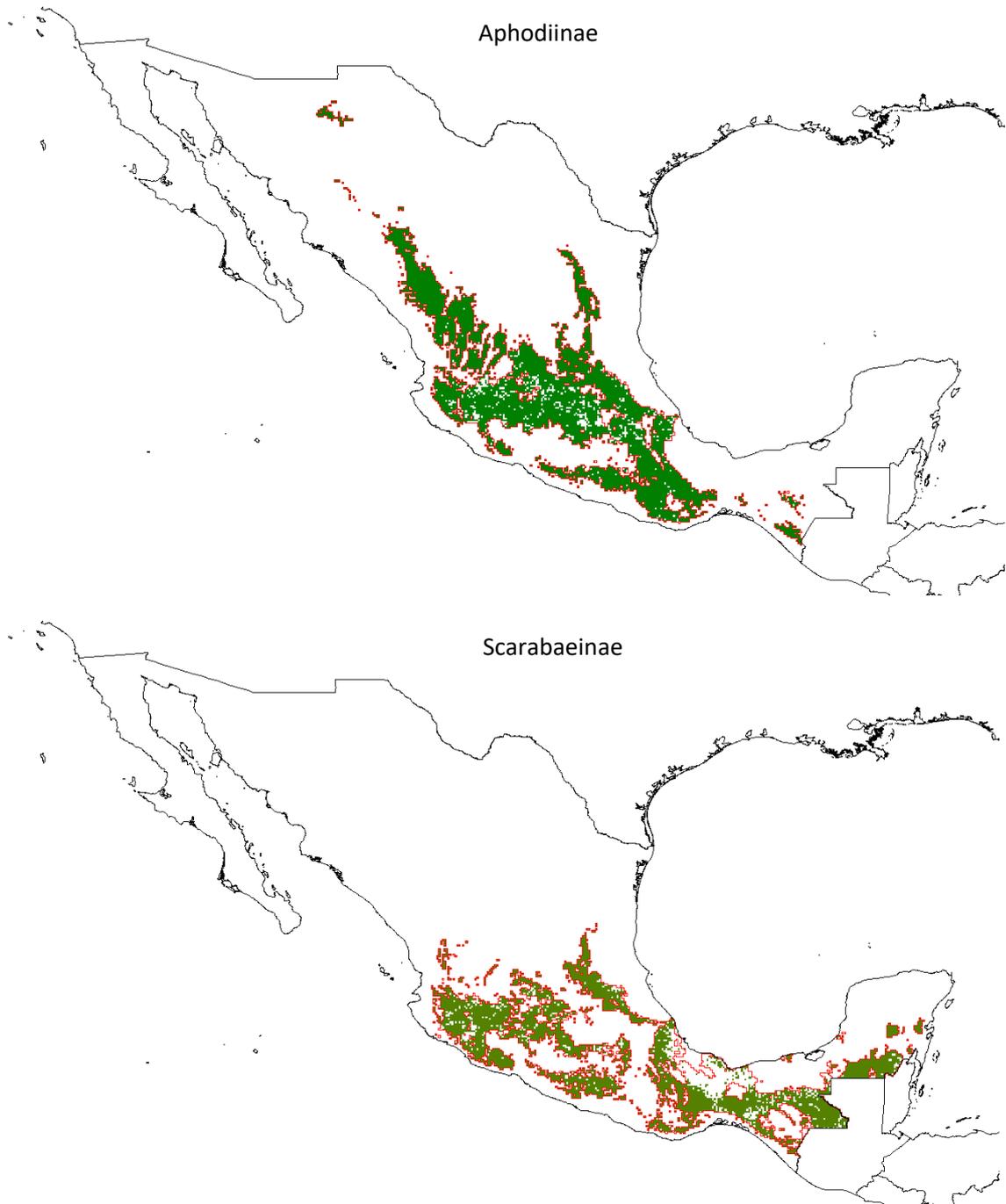


Figura 17.- Area principal de cada uno de los grupos tras eliminar aquellas celdas con al menos el 50% de su superficie antropizada (en verde). La línea roja marca la frontera de esa misma área principal (ver Figs. 7, 9, 11 y 13) para cada uno de los grupos sin eliminar el suelo antropizado. Recuérdese que el área principal está constituida por aquellos territorios que poseen valores de riqueza superiores al cuartil superior de los datos de riqueza por celda y valores del tamaño promedio del área de distribución inferiores al cuartil inferior.

6. Referencias

- Aarts G, Fieberg J, Matthiopoulos J (2012) Comparative interpretation of count, presence–absence and point methods for species distribution models. *Methods in Ecology and Evolution* 3:177–187
- Acevedo P, Real R (2012) Favorability: concept, distinctive characteristics and potential usefulness, *Naturwissenschaften* 99:515–522
- Acevedo P, Jiménez-Valverde A, Lobo JM, Real R (2012) Delimiting the geographical background in species distribution modelling. *Journal of Biogeography* 39:1383–1390
- Aranda SC, Gabriel R, Borges PAV, Brito de Azevedo E, Lobo JM (2011) Designing a survey protocol to overcome the Wallacean shortfall: a working guide using bryophyte distribution data on Terceira Island (Azores). *Bryologist* 114:611–624
- Ballesteros-Mejia L, Kitching IJ, Jetz W, Nagel P, Beck J (2013) Mapping the biodiversity of tropical insects: species richness and inventory completeness of African sphingid moths. *Global Ecology and Biogeography* 22:586–595
- Boakes EH, McGowan PJK, Fuller RA, Ding C, Clark NE, O'Connor K, Mace GM (2010) Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS Biology* 8:e1000385
- Chao A, Jost L (2012) Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology* 93:2533–2547
- Chefaoui RM, Lobo JM (2008) Assessing the effects of pseudoabsences on predictive distribution model performance. *Ecological Modelling* 210:478–486.
- Cuervo-Robayo, A.P. et al. 2014. An update of high-resolution monthly climate surfaces for Mexico. *International journal of Climatology* 34: 2427–2437.
- Dennis RLH, Sparks TH, Hardy PB (1999) Bias in butterfly distribution maps: the effects of sampling effort. *Journal of Insect Conservation* 3:33–42.
- Dubois, D.M. (1973) An index of fluctuations, Do, connected with diversity and stability of ecosystems: applications in the Lotka–Volterra model and in an experimental distribution of species. Rapport de sythèse III, Programme National sur l'environnement Physique et Biologique, Project Mer. Commission Interministérielle de la Politique Scientifique. Liège.
- Duputie A, Zimmermann NE, Chuine I (2014). Where are the wilds things? Why we need better data on species distribution. *Global Ecology and Biogeography* (en prensa)
- García-Roselló E, Guisande C, González-Acosta J, Heine J, Pelayo-Villamil P, Manjarras-Hernández A, Vaamonde A, Granado-Lozano C (2013) ModestR: a software tool for managing and analysing species distribution map databases. *Ecography* 36: 1202–1207.
- Guisande, C., Barreiro, A., Maneiro, I., Riveiro, I., Vergara, A.R., Vaamonde, A. (2006) Tratamiento de datos. Ediciones Díaz de Santos, 367 pp.
- Hastie T, Fithian W (2013) Inference from presence-only data; the ongoing controversy. *Ecography* 36:864–867
- Hijmans RJ (2012) Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology* 93:679–688
- Hortal J, Borges PAV, Gaspar C (2006) Evaluating the performance of species richness estimators: sensitivity to sample grain size. *Journal of Animal Ecology* 75:274–287
- Hortal J, Jiménez-Valverde A, Gómez JF, Lobo JM, Baselga A (2008) Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos* 117:847–858
- Hortal J, Lobo JM (2005) An ED-based protocol for optimal sampling of biodiversity. *Biodiversity and Conservation* 14:2913–2947

- Hortal J, Lobo JM, Jiménez-Valverde A (2007) Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife (Canary Islands). *Conservation Biology* 21:853-863.
- Hortal J, Lobo JM, Jiménez-Valverde A (2012) Basic questions in Biogeography and the (lack of) simplicity of species distributions: Putting Species Distribution Models in the right place. *Natureza & Conservação* 10:108-118
- Jiménez-Valverde A (2012) Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography* 21:498-507
- Johnson DH (1980) The comparison of usage and availability measurements for evaluating resource preference. *Ecology* 61:65-71
- Lobo JM (2008a) More complex distribution models or more representative data? *Biodiversity Informatics* 5:14-19.
- Lobo JM (2008b) Database records as a surrogate for sampling effort provides higher species richness estimations. *Biodiversity and Conservation* 17:873-881
- Lobo JM, Jiménez-Valverde A, Hortal J (2010) The uncertain nature of absences and their importance in species distribution modelling. *Ecography* 33:103-114
- Lobo JM, Jiménez-Valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* 17:145-151
- Rocchini D, Hortal J, Lengyel S, Lobo JM, Jiménez-Valverde A, Ricotta C, Bacaro G, Chiarucci A (2011) Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. *Progress in Physical Geography* 35:211-226
- Sutherst R (2014) Pest species distribution modelling: origins and lessons from history. *Biological Invasions* 16:239-256.
- Tuanmu MN, Jetz W (2014) A global 1-km consensus land-cover product for biodiversity and ecosystem modeling. *Global Ecology and Biogeography* 23: 1031-1045.
- Yang W, Ma K, Kreft H (2013) Geographical sampling bias in a large distributional database and its effects on species richness-environment models. *Journal of Biogeography* 40:1415-1426.