METODOLOGÍA

SELECCIÓN DE PREDICTORES AMBIENTALES PARA EL MODELADO DE LA DISTRIBUCIÓN DE ESPECIES EN MAXENT

SELECTION OF ENVIRONMENTAL PREDICTORS FOR SPECIES DISTRIBUTION MODELING IN MAXENT

Gustavo Cruz-Cárdenas[1*]; José Luis Villaseñor[2]; Lauro López-Mata[4]; Enrique Martínez-Meyer[3]; Enrique Ortiz[2].

[1]Instituto Politécnico Nacional, CIIDIR-IPN-Michoacán, COFAA, Justo Sierra 28. C. P. 59510. Jiquilpan, Michoacán, MÉXICO.

Correo-e: guscruz@ipn.mx Tel.: 353 53 3 00 83 (*Autor para correspondencia).

[2]Departamento de Botánica; [3]Depto. de Zoología, Instituto de Biología, Universidad Nacional Autónoma de México. Tercer Circuito s/n, Ciudad Universitaria, Delegación Coyoacán. C. P. 04510. México, D. F.

[4]Colegio de Postgraduados, campus Montecillo. Carretera México-Texcoco km 36.5. C. P. 56230. Texcoco, Estado de México, MÉXICO.

RESUMEN

Antes de realizar el modelado de la distribución potencial de una especie, se recomienda hacer una preselección de covariables pues la redundancia o variables irrelevantes pueden inducir sesgos en la mayoría de los modelos. En este estudio, se propuso un método automatizado para la selección *a priori* de covariables utilizadas en el modelado. Se emplearon cinco especies típicas de la flora mexicana (*Catopheria chiapensis, Liquidambar styraciflua, Quercus martinezii, Telanthopora grandifolia* y *Viburnum acutifolium*) y 56 covariables ambientales. Se generaron matrices de presencia-ausencia para cada especie y se analizaron empleando regresión logística; el modelo resultante de cada especie se evaluó mediante un remuestreo bootstrap. La distribución de las cinco especies se modeló usando el algoritmo de máxima entropía y con el empleo de tres conjuntos de covariables ambientales. La precisión de los modelos generados se evaluó con intervalos de confianza de cada curva característica operativa del receptor (COR). Los intervalos de confianza de las curvas COR resultantes no mostraron diferencia significativa ($P < 0.05$) entre los tres modelos predictivos generados; sin embargo, el modelo más parsimonioso se obtuvo con el método propuesto.

PALABRAS CLAVE:
Datos de sensores remotos, propiedades de suelos, selección automatizada de covariables.

ABSTRACT

Prior to conducting the modeling of the potential distribution of a species, it is advised to make a pre-selection of covariables because redundancy or irrelevant variables may induce errors in most modeling systems. In this study, we propose an automated method for *a priori* selection of covariables used in modeling. We used five typical species of the Mexican flora (*Catopheria chiapensis*, *Liquidambar styraciflua*, *Quercus martinezii*, *Telanthopora grandifolia* and *Viburnum acutifolium*) and 56 environmental covariables. Presence-absence matrices were generated for each species and were analyzed using logistic regression, and the resulting model of each species was evaluated via a bootstrap resampling. We modeled the distribution of five species using maximum entropy and employed three sets of environmental covariables. The precision of the models generated was evaluated with the confidence intervals for each receiver operating characteristic (ROC) curve. The confidence intervals of the resulting ROC curves showed no significant difference between ($P < 0.05$) the three predictive models generated; nevertheless, the most parsimonious model was obtained with the proposed method.

KEYWORDS: Remote sensing data, soil properties, automated selection of covariables.

RCHSCFA

Recibido: 17 de septiembre, 2013
Aceptado: 07 de mayo, 2014
doi: 10.5154/r.rchscfa.2013.09.034
http://www.chapingo.mx/revistas

INTRODUCCIÓN

Los algoritmos utilizados para la inferencia de las distribuciones potenciales de las especies son herramientas precisas, rápidas y útiles en el campo de la biogeografía (Phillips, Anderson, & Schapire, 2006; Phillips & Dudik, 2008). El modelado de nicho ecológico se ha convertido en la principal herramienta para el modelado de distribución de especies. Se necesitan dos fuentes de entrada de datos para la generación de modelos de distribución basados en nichos, es decir, la presencia/ausencia de la especie o sólo la presencia de datos y un conjunto de variables ambientales relevantes. La pre-selección de covariables candidatas para el modelado de la distribución de las especies se debe hacer antes del modelado (Elith et al., 2011), por lo que es necesario realizar análisis para eliminar la autocorrelación entre covariables, y por lo tanto, evitar la inestabilidad en el modelado. Se han empleado diversas técnicas para este propósito, tales como el uso del coeficiente de correlación de Pearson (Kumar & Stohlgren, 2009), donde se evalúa la multicolinealidad, o la regresión múltiple, que utiliza el factor de inflación de la varianza (Der & Everitt, 2002). Otra alternativa es el uso de análisis de decisión del árbol, donde la parsimonia (el número mínimo de covariables con el mejor ajuste posible) es uno de los principales enfoques para evaluar la mejor clasificación. Por ejemplo, D'heygere, Goethals, y De Pauw (2003) aplicaron este análisis para reducir el conjunto original de covariables a 62.5 % sin afectar a la precisión del modelo predictivo. Las redes neuronales artificiales también se han utilizado para pre-seleccionar las covariables, en los que se aplican los mismos criterios como en el análisis de árbol de decisión para evaluar la parsimonia de los modelos predictivos. Con la aplicación de esta técnica, el conjunto original de covariables se redujo de 17 a 9 y la precisión del modelo predictivo se incrementó del 78 al 80 % (D'heygere, Goethals, & De Pauw, 2006). Como alternativa, Austin y Tu (2004a) proponen el uso de remuestreo bootstrap, donde se espera que las covariables verdaderamente independientes estén presentes en el mayor número de muestras bootstrap, mientras que las variables de ruido están presentes como predictores en un menor número de muestras bootstrap. La ventaja de la técnica de bootstrap sobre los demás es que se realiza de forma automática, permite la estimación de una función de distribución empírica a través del remuestreo de los datos observados, y el modelo más seleccionado no se ve afectado por autocorrelación (Austin & Tu , 2004a). La última afirmación se basa en el hecho de que, dadas dos variables de autocorrelación, X_1 y X_2, si X_2 se excluye del proceso de selección del modelo, entonces es posible que X_1 puede ser identificada como un predictor independiente en todas las muestras bootstrap.

El objetivo de este estudio fue evaluar y proponer una metodología para la selección de covariables con remuestreo bootstrap que se debe utilizar en el modelado de las distribuciones potenciales de las especies. Este método

INTRODUCTION

The algorithms used for the inference of the potential distributions of species are precise, rapid and useful tools in the field of biogeography (Phillips, Anderson, & Schapire, 2006; Phillips & Dudik, 2008). Ecological niche modeling has become the main tool for modeling species' distributions. Two input sources of data are needed for generating niche-based distribution models, namely species' presence/absence or presence only data and a set of relevant environmental variables. Pre-selection of candidate covariables for the modeling of the distributions of species should be done prior to modeling (Elith et al., 2011), so it is necessary to carry out analyses to eliminate autocorrelation between covariables and thus avoid instability in the modeling. Various techniques have been employed for this purpose, such as the use of the Pearson correlation coefficient (Kumar & Stohlgren, 2009), where multicollinearity is evaluated, or multiple regression, which uses the variance inflation factor (Der & Everitt, 2002). Another alternative is the use of decision tree analysis, where parsimony (the minimum number of covariables with the best possible fit) is one of the main approaches to evaluate the best classification. For example, D'heygere, Goethals, and De Pauw (2003) applied this analysis to reduce the original set of covariables to 62.5 % without affecting the precision of the predictive model. Artificial neuronal networks have also been used to pre-select covariables, in which the same criteria apply as in a decision tree analysis to evaluate the parsimony of the predictive models. With the application of this technique, the original set of covariables was reduced from 17 to 9 and the precision of the predictive model was increased from 78 to 80 % (D'heygere, Goethals, & De Pauw, 2006). Alternatively, Austin and Tu (2004a) propose the use of bootstrap resampling, where it is expected that truly independent covariables be present in the largest number of bootstrap samples, while noise variables be present as predictors in fewer bootstrap samples. The advantage of the bootstrap technique over others is that it is done automatically, it allows the estimation of a distribution function empirically through resampling of the observed data, and the best selected model is not affected by autocorrelation (Austin & Tu, 2004a). The latter statement is based on the fact that, given two autocorrelated variables, X_1 and X_2, if X_2 is excluded from the selection process of the model, then it is possible that X_1 can be identified as an independent predictor in all of the bootstrap samples.

The objective of this study was to evaluate and propose a methodology for the selection of covariables with bootstrap resampling that should be used in the modeling of the potential distributions of species. This method reduces model overfitting because the covariables are selected *a priori*, reduces autocorrelation because bootstrap resampling allows several combinations of data and covariables, and selects the most significant covariables for each particular species.

reduce el sobreajuste del modelo porque las covariables se seleccionan *a priori*, disminuye la autocorrelación ya que el remuestreo bootstrap permite varias combinaciones de datos y covariables, y selecciona las covariables más significativas para cada especie en particular.

MATERIALES Y MÉTODOS

Área de estudio

Catopheria chiapensis A. Gray ex Benth., *Liquidambar styraciflua* L., *Quercus martinezii* C.H. Müll., *Telanthopora grandifolia* (Less.) H. Rob. & Brettell, *Viburnum acutifolium* Benth., especies características del bosque húmedo de montaña (BHM) en México, fueron seleccionadas porque el BHM siendo el bioma con mayor riqueza de especies de plantas vasculares por unidad de superficie, aunque estas especies pueden estar presentes en otros tipos de vegetación, son características del BHM (Rzedowski, 1996; Villaseñor, 2010) (Cuadro 1). Los estados donde se presentan las cinco especies fueron 13, mientras que los estados con mayor número de registros (82) fueron Oaxaca y Veracruz, y los que tienen un solo registro fueron el Estado de México, Jalisco, Tabasco y San Luis Potosí (Figura 1). A pesar de esta gran riqueza el estar situado en una superficie relativamente pequeña de terreno montañoso, el bioma tiene una heterogeneidad elevada de ambientes (clima, suelos y altitud) y un alto grado de fragmentación del hábitat en toda su área (Ramírez-Marcial, González-Espinosa, & Williams-Linera, 2001; Vázquez-García, 1995). Este bioma incluye un amplio conjunto de asociaciones vegetativas heterogéneas, que varían geográficamente tanto en la composición y estructura de la flora, así como en el estado de conservación y el nivel de perturbación a los que están siendo o han sido sometidos (Challenger & Caballero, 1998; Luna-Vega, Alcantara-Ayala, Ruíz-Pérez, & Contreras-Medina, 2006; Ramírez-Marcial et al., 2001; Rzedowski, 1996).

Registros de presencia-ausencia de especies

Los lugares para la colección de cinco especies se obtuvieron mediante la georreferenciación de los ejemplares depositados en el Herbario Nacional de México (MEXU) del Instituto de Biología de la UNAM. En el Cuadro 1 se muestra una base de datos con 336 registros de cinco especies presentes en el BHM. El segundo grupo consistió en especies de biomas diferentes al BHM (100 registros de 25 especies), tales como el matorral desértico, y de sitios en los estados donde no se ha registrado el BHM.

Covariables ambientales

Se utilizaron 56 covariables en el análisis para evaluar cuáles están mejor asociadas con una distribución específica, todas ellas con una resolución de pixel de 1 km^2:

a) La información sobre el clima (26 covariables), que incluyen 19 covariables tomadas de WORLDCLIM

MATERIALS AND METHODS

Study area

Catopheria chiapensis A. Gray ex Benth., *Liquidambar styraciflua* L., *Quercus martinezii* C.H. Müll., *Telanthopora grandifolia* (Less.) H. Rob. & Brettell, *Viburnum acutifolium* Benth., species characteristic of the Cloud Forest (Bosque Húmedo de Montaña-HMF) in Mexico, were selected because the HMF being the biome with greatest richness of vascular plant species per unit of surface area although these species can be present in other vegetation types, they are characteristic of the HMF (Rzedowski, 1996; Villaseñor, 2010) (Table 1). The states where the five species occurs were 13, while the states with the highest number of records (82) were Oaxaca and Veracruz, and those with a single record were the states of Mexico, Jalisco, Tabasco and San Luis Potosí (Figure 1). Despite this great richness being located on a relatively small surface of mountainous terrain, the biome has an elevated heterogeneity of environments (climate, soils and altitude) and a high degree of fragmentation of the habitat in its entire area (Ramírez-Marcial, González-Espinosa, & Williams-Linera, 2001; Vázquez-García, 1995). This biome includes a broad set of heterogeneous vegetative associations, which vary geographically both in the composition and structure of the flora, as well as in the state of conservation and the level of perturbation to which they are being or have been subjected (Challenger & Caballero, 1998; Luna-Vega, Alcantara-Ayala, Ruíz-Pérez, & Contreras-Medina, 2006; Ramírez-Marcial et al., 2001; Rzedowski, 1996).

Presence-absence records of species

The locations for the collection of five species were obtained by georreferencing the specimens deposited in the National Herbarium of Mexico (MEXU) of the Institute of Biology of

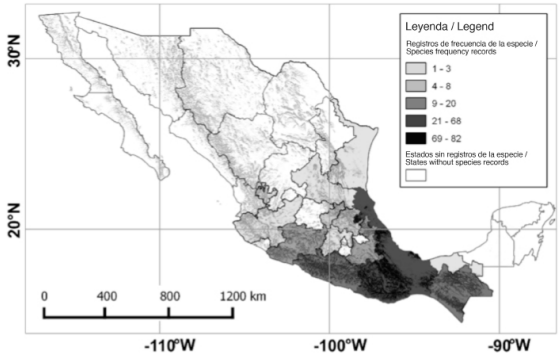

FIGURA 1. Frecuencia de registros de *Catopheria chiapensis, Liquidambar styraciflua, Quercus martinezii, Telanthopora grandifolia* y *Viburnum acutifolium* por estado.

FIGURE 1. Frequency of records of *Catopheria chiapensis, Liquidambar styraciflua, Quercus martinezii, Telanthopora grandifolia* and *Viburnum acutifolium* by state.

notas

CUADRO 1. Especies seleccionadas en el bosque húmedo de montaña para analizar el modelado de la distribución potencial de especies mediante la selección de covariables con remuestreo bootstrap.

Especie	Familia	Registros	Tipos de vegetación
Catopheria chiapensis A. Gray ex Benth.	Lamiaceae	28	BHM, BTEM, BTSE, BTHU
Liquidambar styraciflua L.	Altingiaceae	159	BHM, BTEM, BTHU
Quercus martinezii C. H. Müll.	Fagaceae	47	BHM, BTEM
Telanthopora grandifolia (Less.) H. Rob. & Brettell	Asteraceae	72	BHM, BTEM, BTSE, BTHU
Viburnum acutifolium Benth.	Adoxaceae	30	BHM

BHM = bosque húmedo montaña, BTEM = Bosque templado, BTSE = Bosque estacionalmente seco; BTHU = Bosque lluvioso tropical.

TABLE 1. Species in the cloud forest selected to analyze modeling of the potential distribution of species through the selection of covariables with bootstrap resampling.

Species	Family	Records	Vegetation types
Catopheria chiapensis A. Gray ex Benth.	Lamiaceae	28	HMF, BTEM, BTSE, BTHU
Liquidambar styraciflua L.	Altingiaceae	159	HMF, BTEM, BTHU
Quercus martinezii C. H. Müll.	Fagaceae	47	HMF, BTEM
Telanthopora grandifolia (Less.) H. Rob. & Brettell	Asteraceae	72	HMF, BTEM, BTSE, BTHU
Viburnum acutifolium Benth.	Adoxaceae	30	HMF

HMF = Cloud forest, BTEM = Temperate forest, BTSE = Seasonably dry forest; BTHU = Tropical rain forest.

(Hijmans, Cameron, Parra, Jones, & Jarvis, 2005), y cuatro derivados de los siguientes subgrupos: precipitación total en la temporada de lluvias del año (mayo a octubre), precipitación total en la época seca del año (noviembre a abril), temperatura media de la temporada de lluvias del año (mayo a octubre), y temperatura media de la estación seca del año (noviembre a abril). Las tres variables restantes son: tasa de evapotranspiración, evapotranspiración real anual (ETRA), evapotranspiración real de la temporada de lluvias del año (ETRAH, mayo a octubre) y evapotranspiración real de la época seca del año (ETRAS, noviembre a abril). La evapotranspiración se calculó basándose en el modelo de Turc (1954), como se muestra a continuación:

$$ETRA = P / [0.9 + (P/L)^2]^{½}$$

Donde:
ETRA = Evapotranspiración real anual (mm)
P = Precipitación anual total (mm)
T = Temperatura media anual (°C)
$L = 300 + 25T + 0.05T^3$

b) Atributos topográficos (nueve covariables). El modelo digital de elevación (MDE) fue descargado desde el sitio web del USGS (Servicio Geológico de los Estados Unidos, 2010) y se utilizó para generar los siguientes ocho atributos topográficos con SAGA GIS (Cimmery, 2010): pendiente, orientación (de 0° a 359°), escorrentía, índice de convergencia, índice de humedad topográfica, índice de rugosidad del terreno (Riley, DeGloria, & Elliot, 1999), medida de rugosidad del vector (Sappington, Longshore, & Thompson, 2007) y calentamiento anisotrópico.

c) Propiedades del suelo (nueve covariables). Estas covariables fueron generadas por Cruz-Cárdenas et

UNAM. A database with 336 records of five species present in the HMF (Table 1). The second set consisted of species from biomes different from the HMF (100 records of 25 species), such as the scrubby desert, and from sites in states where HMF has not been recorded.

Environmental covariables

The analysis used 56 covariables, to evaluate which of them are better associated to their specific distribution all of them with a pixel resolution of 1 km^2:

a) Information on climate (26 covariables), including 19 covariables taken from WORLDCLIM (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005), and four derived from the following subsets: total precipitation in the rainy season of the year (May to October), total precipitation in the dry season of the year (November to April), average temperature of the rainy season of the year (May to October), and average temperature of the dry season of the year (November to April). The remaining three variables include rate of evapotranspiration, real annual evapotranspiration (ETRA), real evapotranspiration of the rainy season of the year (ETRAH, May to October) and real evapotranspiration of the dry season of the year (ETRAS, November to April). Evapotranspiration was calculated based on the model of Turc (1954), as follows:

$$ETRA = P / [0.9 + (P/L)^2]^{½}$$

Where:
ETRA = Real annual evapotranspiration (mm)
P = Total annual precipitation (mm)
T = Average annual temperature (°C)
$L = 300 + 25T + 0.05T^3$

notes

al. (2014), que incluyen: materia orgánica, relación de absorción de sodio, pH, conductividad eléctrica, carbono orgánico, K, Na, Mg y Ca.

d) Índices de vegetación (12 covariables). Los índices de vegetación medios mensuales fueron descargados para el año 2009 de la página web de la USGS (2010) del satélite MODIS. Son siete escenas que cubren el territorio de México y se encuentran en proyección sinusoidal. Se generó un mosaico con las siete escenas, las cuales se transforman a coordenadas geográficas para que sea compatible con las covariables previas. Este procedimiento permitió la generación de 12 índices de vegetación mensuales para todo el país.

Configuración de matrices de presencia-ausencia para la evaluación de modelos predictivos

Se generó una matriz de datos para cada una de las cinco especies analizadas desde el BHM. La matriz contiene 100 filas de ausencias más *N* filas que corresponden al número de registros de presencia. El número de columnas fue de 57, que incluye las 56 covariables ambientales más una de presencia (1) o ausencia (0). Elith et al. (2011) recomiendan datos de modelos de presencia y ausencia (cuando sea posible), ya que al usar ambos se proporciona más y mejor información sobre la prevalencia (es decir, las zonas donde la especie está presente) que solo con los datos de presencia.

Aplicación del método bootstrap

Se aplicó una regresión binomial para cada matriz de presencia-ausencia. Posteriormente, el siguiente proceso se aplicó 1000 veces (muestras bootstrap) para el modelo predictivo resultante de esta regresión, con este número de repeticiones es apropiado porque se obtienen modelos predictivos parsimoniosos (Austin & Tu, 2004b), utilizando la biblioteca bootStepAIC (Dimitris, 2009) del software R (R Development Core Team, 2010): 1) se simuló un nuevo conjunto de datos tomando una muestra con reemplazo de las filas de la matriz, 2) el modelo predictivo se ajustó utilizando el conjunto de datos en el paso 1, al ejecutar el algoritmo StepAIC (Venables & Ripley, 2010), que aplica un proceso «hacia atrás» (la eliminación se inicia al usar todas las covariables en el modelo y estos se excluyen de forma secuencial desde el modelo hasta que se cumpla el nivel de significación), con $P = 0.01$ para la eliminación de una covariable a partir del modelo. Los valores de salida de bootstrap que se usaron para la selección de covariables componen la matriz (cuántas veces se seleccionó la covariable de las 1000 muestras bootstrap para formar el modelo predictivo) y la matriz de significancia (de las veces que fue seleccionada, ¿cuántas veces fue significativa la variable?). Las dos matrices se multiplicaron entre sí y se dividieron entre 100, para ponderar las variables tomando en cuenta el porcentaje en su selección significativa por el modelo. Finalmente, el tercer cuartil de la matriz resultante se calculó con el objetivo de seleccionar aquellas covariables que fueron iguales o mayor a ese valor.

b) Topographical attributes (nine covariables). The digital elevation model (DEM) was downloaded from the website of the USGS (United State Geological Survey, 2010) and was used to generate the following eight topographic attributes with SAGA GIS (Cimmery, 2010): slope, aspect (from 0° to 359°), runoff, index of convergence, index of topographic humidity, index of roughness of the terrain (Riley, DeGloria, & Elliot, 1999), measure of roughness of the vector (Sappington, Longshore, & Thompson, 2007) and anisotropic heating.

c) Soil properties (nine covariables). These covariables were generated by Cruz-Cárdenas et al. (2014), include organic matter, sodium absorption ratio, pH, electrical conductivity, organic carbon, K, Na, Mg and Ca.

d) Vegetation Indices (12 covariables). Monthly average vegetation indices were downloaded for the year 2009 from the MODIS satellite via the website of the USGS (2010). These remote sensing data are in sinusoidal projection and seven scenes cover the territory of Mexico. A mosaic was generated with the seven scenes, which was transformed into geographic coordinates to make it compatible with the previous covariables. This procedure allowed the generation of 12 monthly vegetation indices for the entire country.

Configuration of the matrices of presence-absence for evaluation of predictive models

A data matrix was generated for each of the five species analyzed from the HMF. The matrix contains 100 rows of absences plus *N* rows corresponding to the number of records of presence. The number of columns was 57, which includes the 56 environmental covariables plus one of presence (1) or absence (0). Elith et al. (2011) recommend modeling data of presence and absence (when possible) because using both provides more and better information on prevalence (i. e., areas where the species is predicted present) than with presence-only data.

Application of the bootstrap method

For each presence-absence matrix, a binomial regression was applied. Then, the following process was applied 1,000 times (bootstrap samples) to the predictive model resulting from this regression, with this number of replications is appropriate because parsimonious predictive models are obtained (Austin & Tu, 2004b), using the bootStepAIC library (Dimitris, 2009) of software R (R Development Core Team, 2010): 1) A new data set was simulated by taking a sample with replacement from the rows of the matrix, 2) the predictive model was fit using the data set from step 1, executing the StepAIC algorithm (Venables & Ripley, 2010), which applies a "backwards" process (elimination begins using all of the covariables in the model and these are sequentially eliminated from the model until the level of significance is met), with $P = 0.01$ for the elimination of a covariable from the model. The output values of the

notas

Modelado de distribución

Una vez que las covariables fueron seleccionadas con el método bootstrap, se realizó el modelado de la distribución potencial de la especie en Maxent (Phillips et al., 2006). La finalidad de Maxent es estimar la distribución de probabilidad del objetivo a través de la distribución de probabilidad de máxima entropía (cercana a la uniforme), delimitado por un conjunto de reglas para representar la información incompleta acerca de la distribución de destino. La información disponible acerca de la distribución de destino por sí misma se presenta como un conjunto de covariables conocidas como características y se esperan restricciones de cada característica que debe corresponder a sus valores medios de la muestra (la media muestral para un conjunto de puntos de muestreo se han extraído de la distribución destino). Ya que Maxent se utiliza para modelar la distribución de especies con sólo registros de presencia, los píxeles del área de estudio se han convertido en el espacio en el que se delimita la distribución de probabilidad, los píxeles para registrar las ocurrencias de especies conocidas son los puntos de muestreo, las características son las covariables de clima, topografía, suelos, vegetación y otras covariables ambientales. Maxent se configuró solo con los registros de presencia de la especie; los puntos totales de ocurrencia de cada especie se dividieron en: registros de entrenamiento (75 %) y validación (25 %). Se seleccionaron 10,000 puntos de trasfondo, la configuración predeterminada de los tipos de niveles de regularización y de características (Phillips & Dudik, 2008), y el formato de salida fue logístico. La misma configuración de Maxent se utilizó para el modelado con dos grupos de covariables adicionales, uno con todas las covariables (56) y otro con covariables bioclimáticas de wordclim (19) para comparar la diferencia entre los modelos resultantes. Los modelos logísticos finales se convirtieron en mapas binarios utilizando el valor de probabilidad de 10 % como umbral, para maximizar la sensibilidad y minimizar la especificidad (Phillips et al., 2006).

Análisis de modelos de distribución

Se realizó una evaluación inicial mediante la aplicación de una prueba binomial para cada modelo generado con la finalidad de determinar si el resultado era mejor que una distribución aleatoria. En una segunda prueba, se utilizó el análisis de la curva característica de operación del receptor (COR) para comparar los niveles de precisión de los tres modelos que se generan en Maxent: con las 56 covariables (Modelo 1), las 19 covariables Bioclim (Modelo 2), y la variable número determinado de covariables seleccionadas mediante el método bootstrap (Modelo 3). La curva COR evalúa la configuración de un modelo en términos de errores de omisión y comisión por medio de un solo número, mientras que el área bajo la curva (ABC) puede evaluar diferentes modelos (Fawcett, 2006). Se emplearon los intervalos de confianza bootstrap (BC_{α}) con 10,000 repeticiones para encontrar si las diferencias de los valores de ABC entre los tres modelos fueron significativos. Los intervalos de percentiles se definen

bootstrap that was used for the selection of covariables compose the matrix of covariables (how many times from the 1,000 bootstrap samples the covariable was selected to form the predictive model) and the matrix of significance (how many times the variable was significant from the times that it was selected). The two matrices were multiplied together and divided by 100, to weigh the variables taking into account that percentage in their significant selection by the model. Finally, the third quartile of the resulting matrix was calculated with the aim of selecting those covariables that were equal to or greater than this value.

Distribution modeling

Once the covariables were selected with the bootstrap method, modeling of the potential distribution of the species was carried out in Maxent (Phillips et al., 2006). The aim of Maxent estimate the probability distribution target through the probability distribution of maximum entropy (close uniform), delimited by a set of rules to represent incomplete information about the target distribution. Available information about the target distribution by itself is presented as a set of covariables known as features and the restrictions are expected from each features that should be matched their sample mean values (sample mean for one set of sampling points are drawn from the target distribution). As Maxent is used to model the specie distribution presence-only records, the pixels of the study area has become the space in which the probability distribution is delimited, pixels to record occurrences of known species are the sampling points, the features are the covariables of climate, topography, soils, vegetation and other environmental covariables. Maxent was configured using presence-only records of the species; the total points of occurrence of each species were split into training (75 %) and validation (25 %) records. We selected 10,000 background points and the default settings for the regularization level and feature types (Phillips & Dudik, 2008), and the output format was logistic. The same configuration of Maxent was used to model with two additional covariables, one with all the covariables (56) and another with wordclim bioclimatic covariables (19) to compare the difference between the resulting models. Final logistic models were converted into binary maps using the 10 % probability value as the threshold, to maximize sensitivity and minimize specificity (Phillips et al., 2006).

Analysis of distribution models

An initial evaluation was done by applying a binomial test to each model generated to determine whether the result was better than a random distribution. In a second test, the receiver operating characteristic (ROC) curve analysis was used to compare the levels of precision of the three models generated in Maxent: with the 56 covariables (Model 1), with the 19 Bioclim covariables (Model 2), and the variable number set of covariables selected using the bootstrap method (Model 3). The ROC curve evaluates the configuration of a model in terms of is omission and

notes

como la diferencia entre la mediana de $\hat{\theta}^{*b}$ del bootstrap, y la estimada a partir de la muestra original. El sesgo constante estimado se denota por $\hat{Z}_0$ y se define como:

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#\left(\hat{\theta}^{*b} < \hat{\theta}\right)}{B}\right)$$

Donde Φ^{-1} simboliza el inverso de distribución normal acumulativa y # "número de" medias. Entonces, un intervalo de confianza percentil de 100 (1-α) por ciento del sesgo corregido para θ se origina por:

$$(\ _1),\quad (\ _2)$$

Donde α_1 y α_2 son el cuantil revisado en el que se basa la valoración de los intervalos de confianza percentil. Estos cuantiles se definen como:

$$_1 = \Phi\ \ 2\ \hat{}_0 + \quad /2$$

y

$$_2 = \Phi\ \ 2\ \hat{}_0 + \quad 1- /2$$

Donde Φ es distribución normal estándar acumulativa. El método BC_α es un algoritmo automático para la producción de intervalos de confianza con alta precisión de una distribución bootstrap (DiCiccio & Efron, 1996).

Por último, se evaluó la incertidumbre de la distribución potencial de las especies mediante una validación cruzada 10-fold usando Maxent (Elith & Leathwick, 2009). Dicha validación consistió en dividir los datos sobre la presencia de especies en 10 subconjuntos. Uno de los subconjuntos se utilizó como validación de datos y los nueve restantes como datos de entrenamiento en el modelado. El proceso de validación cruzada se repite durante 10 iteraciones, con cada uno de los posibles subconjuntos de datos de validación. Los resultados de la validación cruzada son grids de promedio, la mediana y la desviación estándar de las iteraciones. El grid de la desviación estándar se utilizó para determinar la incertidumbre de los modelos de las especies.

RESULTADOS Y DISCUSIÓN

En el Cuadro 2 se muestran las covariables seleccionadas para cada especie. El número de covariables oscila de 13 a 14. Las covariables con la frecuencia más alta para este conjunto de especies fueron: rango promedio de temperaturas diarias (bio2), precipitación anual (bio12), y evapotranspiración real anual (ETRA), se presentan en cuatro de las cinco especies. Los resultados concuerdan con los de Bruijzeel, Waterloo, Proctor, Kuiters & Kotterink (1993), quienes también señalan como característica distintiva del BHM las lluvias abundantes, así como una tasa baja de pérdida de la evapotranspiración. Vogelman (1973) sugirió que las temperaturas moderadas y la humedad atmosférica alta son los principales determinantes ambientales de la vegetación en el BHM. Por otra parte, la covariable más importante

commission errors by means of a single number, while the area under the curve (AUC) can evaluate different models (Fawcett, 2006). Bootstrap confidence intervals were used (BC_α) with 10,000 repetitions to discover if the differences of AUC values between the three models were significant. The percentile intervals are defined as the difference among the median of $\hat{\theta}^{*b}$ from the bootstrap, and estimated from the original sample. The constant bias estimated is denoted by $\hat{Z}_0$ and is defined as:

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#\left(\hat{\theta}^{*b} < \hat{\theta}\right)}{B}\right)$$

Where Φ^{-1} does it symbolize the inverse accumulative normal distribution and # "number of" means. Then, a 100(1-α) percent interval of the percentile bias-corrected confidence for θ is given by:

$$(\ _1),\quad (\ _2)$$

Where α_1 and α_2 are the revised quantile on which the valuation of confidence intervals of the percentile are based. These quantile are defined as:

$$_1 = \Phi\ \ 2\ \hat{}_0 + \quad /2$$

and

$$_2 = \Phi\ \ 2\ \hat{}_0 + \quad 1- /2$$

Where Φ is the cumulative normal standard distribution. The BC_α method is an automatic algorithm for producing confidence intervals with high precision from a bootstrap distribution (DiCiccio & Efron, 1996).

Finally, the uncertainty of the potential distribution of the species was evaluated by 10-fold cross-validation with Maxent (Elith & Leathwick, 2009). Such a validation consisted on dividing the species occurrence data in 10 subsets. One of the subsets was used as data validation and the remaining nine as data training in the modeling. The process of cross-validation is repeated during 10 iterations, with each of the possible subsets of test data. The results of cross-validation are the average, median and standard deviation of the grid iterations. The grid standard deviation was used to determine the uncertainty of the models of the species.

RESULTS AND DISCUSSION

The covariables selected for each species are shown in Table 2. The number of covariables varied from 13 to 14. Covariables with the highest frequency for this set of species were the average range of daily temperatures (bio2), the annual precipitation (bio12), and the real annual evapotranspiration (ETRA), being present in four out of the five species. The results agree with those by Bruijzeel, Waterloo, Proctor, Kuiters, and Kotterink (1993) who also point out as characteristic distinctive of the

notas

CUADRO 2. Covariables ambientales relevantes seleccionadas en el análisis bootstrap para cada especie. Entre paréntesis se presentan los valores de porcentaje de participación en el modelo.

Covariable	*Catopheria chiapensis*	*Liquidambar styraciflua*	*Quercus martinezii*	*Telanthopora grandifolia*	*Viburnum acutifolium*	Frecuencia
Precipitación anual (bio12, mm)	1 (84)		1 (89)	1 (78)	1 (82)	4
Rango promedio de temperaturas diarias (bio2, °C)	1 (84)		1 (87)	1 (98)	1 (83)	4
Evapotranspiración anual (ETRA, mm)	1 (86)	1 (94)		1 (76)	1 (84)	4
Estacionalidad de precipitación (bio15, mm)	1 (87)		1 (83)		1 (83)	3
Precipitación del trimestre más frío (bio19, mm)	1 (83)		1 (88)		1 (80)	3
Isotermalidad (bio3, °C)	1 (84)			1 (88)	1 (81)	3
Temperatura mínima del mes más frío (bio6, °C)		1 (84)	1 (85)	1 (97)		3
Modelo digital de elevación (MDE, m)		1 (99)		1 (91)	1 (86)	3
Evapotranspiración de la estación lluviosa (ETRAH, mm)	1 (86)			1 (72)	1 (82)	3
K (cmol·Litro^{-1})		1 (99)	1 (97)	1 (94)		3
pH	1 (83)		1 (91)	1 (80)		3
Precipitación para la termporada de lluvia (PPH, mm)	1 (85)		1 (83)		1 (84)	3
Índice de rugosidad del terreno (TRI)		1 (71)	1 (81)		1 (79)	3
Precipitación del trimestre más húmedo (bio16, mm)			1 (83)		1 (84)	2
Temperatura máxima del mes más cálido (bio5, °C)				1 (98)	1 (81)	2
Carbono órganico (CO, kg·m^{-2})		1 (72)		1 (93)		2
Evapotranspiración del período seco (ETRAS, mm)				1 (76)	1 (83)	2
Na (cmol·Litro^{-1})	1 (86)	1 (81)				2
Aspecto	1 (82)					1
Temperatura media del trimestre más cálido (bio10, °C)		1 (77)				1
Precipitación del mes más seco (bio14, mm)	1 (84)					1
Precipitación del trimestre más seco (bio17, mm)			1 (84)			1
Estacionalidad de temperatura (bio4, °C)			1 (87)			1
Temperatura media del trimestre más seco (bio9, °C)		1 (85)				1
Ca (cmol·Litro^{-1})			1 (76)			1
Conductividad eléctrica (CE, dS·m^{-1})	1 (83)					1
calentamiento diurno anisotrópico (DAH)					1 (81)	1
NDVIMARCH2009		1 (77)				1
NDVIMAY2009	1 (89)					1
Materia órganica (OM, %)				1 (87)		1
Relación de absorción de sodio (SAR, %)		1 (74)				1
Temperatura para la estación seca (TS, °C)			1 (74)			
Total de covariables		14	13	13	13	14

ETRAH: Evapotranspiración de la temporada de lluvias (mayo a octubre); ETRAS: evapotranspiración de la estación seca (noviembre a abril). Covariables con el acrónimo NDVI son datos promedios mensuales del índice normalizado de diferencia de vegetación para el año 2009.

notes

TABLE 2. Relevant environmental covariables selected in the bootstrap analysis for each species. In brackets values of percentage of participation in the model are presented.

Covariable	*Catopheria chiapensis*	*Liquidambar styraciflua*	*Quercus martinezii*	*Telanthopora grandifolia*	*Viburnum acutifolium*	Frequency
Annual precipitation (bio12, mm)	1 (84)		1 (89)	1 (78)	1 (82)	4
Mean diurnal range (bio2, °C)	1 (84)		1 (87)	1 (98)	1 (83)	4
Annual evapotranspiration (ETRA, mm)	1 (86)	1 (94)		1 (76)	1 (84)	4
Precipitation seasonality (bio15, mm)	1 (87)		1 (83)		1 (83)	3
Precipitation of coldest quarter (bio19, mm)	1 (83)		1 (88)		1 (80)	3
Isothermality (bio3, °C)	1 (84)			1 (88)	1 (81)	3
Minimum temperature of coldest month (bio6, °C)		1 (84)	1 (85)	1 (97)		3
Digital elevation model (DEM, m)		1 (99)		1 (91)	1 (86)	3
Evapotranspiration of the wet season (ETRAH, mm)	1 (86)			1 (72)	1 (82)	3
K (cmol·liter^{-1})		1 (99)	1 (97)	1 (94)		3
pH	1 (83)		1 (91)	1 (80)		3
Precipitation for the wet season (PPH, mm)	1 (85)		1 (83)		1 (84)	3
Terrain roughness index (TRI)		1 (71)	1 (81)		1 (79)	3
Precipitation of wettest quarter (bio16, mm)			1 (83)		1 (84)	2
Maximum temperature of warmest month (bio5, °C)				1 (98)	1 (81)	2
Organic carbon (OC, kg·m^{-2})		1 (72)		1 (93)		2
Evapotranspiration of the dry season (ETRAS, mm)				1 (76)	1 (83)	2
Na (cmol·liter^{-1})	1 (86)	1 (81)				2
Aspect	1 (82)					1
Mean temperature of warmest quarter (bio10, °C)		1 (77)				1
Precipitation of driest month (bio14, mm)	1 (84)					1
Precipitation of driest quarter (bio17, mm)			1 (84)			1
Temperature seasonality (bio4, °C)			1 (87)			1
Mean temperature of driest quarter (bio9, °C)		1 (85)				1
Ca (cmol·liter^{-1})			1 (76)			1
Electrical conductivity (EC, dS·m^{-1})	1 (83)					1
Diurnal anisotropic heating (DAH)					1 (81)	1
NDVIMARCH2009		1 (77)				1
NDVIMAY2009	1 (89)					1
Organic material (OM, %)				1 (87)		1
Sodium absorption ratio (SAR, %)		1 (74)				1
Temperature for the dry season (TS, °C)		1 (74)				1
Total of covariables	14	13	13	13	14	

ETRAH: Evapotranspiration of the wet season (May to October); ETRAS: Evapotranspiration of the dry season (November to April). Covariables with the acronym NDVI are monthly averages data of normalize difference vegetation index for the year 2009.

notas

(con la puntuación total más alta en la matriz resultante de la técnica de bootstrap) para cada especie fue: el índice normalizado de vegetación de mayo (NDVIMAY2009) para *C. chiapensis,* el modelo digital de elevación (MDE) para *L. styraciflua* y *V. acutifolium*, pH para *Q. martinezii,* y el rango promedio de temperaturas diarias (bio2) para *T. grandifolia*.

La técnica aquí propuesta para la selección de covariables evita el problema encontrado por Austin y Tu (2004b) con el remuestreo bootstrap. Estos autores identificaron 940 subgrupos de covariables como predictores independientes de 1000 muestras bootstrap y menos del 1% de esos subgrupos se repitieron cuatro veces. En este caso, se tomaron en cuenta otros criterios para seleccionar las covariables, y la prueba COR mostró una mayor consistencia. La técnica bootstrap es eficiente para la selección de las covariables, ya que permite una reducción *a priori* del número de covariables utilizados en el modelado de la distribución potencial de especies sin reducir la precisión. Por el contrario, se recomienda reducir el número de variables para evitar el sobreajuste del modelo, sobre todo cuando el número de registros es bajo (Stockwell & Peterson, 2002).

En cuanto a la validación del modelo con la prueba binomial, todos los modelos, excepto el modelo 1 para *V. acutifolium* fueron mejores que las expectativas al azar (Cuadro 3). Por otra parte, los análisis de la curva COR indicaron que no hubo diferencias significativas ($P < 0.05$) en la precisión existente entre los tres modelos para cada especie (Figura 2). La distribución de las especies de plantas de los tres modelos es mejor que la distribución al azar como se observa en las curvas de la Figura 2 que se proyectan hacia el noroeste o mayor a 0.95 ABC (Fawcett, 2006), concluyendo que las predicciones de los modelos generados contienen principalmente verdaderos positivos (sensibilidad) y un menor número de falsos negativos (1-especificidad). En el Cuadro 4 se muestran los intervalos de confianza bootstrap (BC_α) de los tres modelos para cada especie. El BC_α del área bajo la curva (ABC) indica que no hubo diferencias significativas ($P < 0.05$) entre los tres modelos; sin embargo, el modelo más parsimonioso fue consistentemente el modelo 3. En especies con menos de 50 registros, el valor BC_α del modelo 1 fue bastante grande, a diferencia del modelo 3, lo que sugiere que con pocos registros, la selección de variables con el método bootstrap resulta más adecuada.

En la Figura 3 se muestra la incertidumbre de los modelos de distribución de especies. Se observó diferencia significativa ($P < 0.05$) de la incertidumbre de los modelos generados con 56 variables y las generadas con las variables seleccionadas por el método bootstrap (Figuras 3a). Por el contrario, no se mostraron diferencias entre los modelos, ya que se usaron tanto las 19 covariables WorldClim y las covariables seleccionadas por la técnica propuesta. Sin embargo, los modelos de incertidumbre eran más pequeños utilizando las covariables seleccionadas. La gráfica de cajas muestra que hay una variación mayor en los valores de incertidumbre de los modelos de distribución de *C. chiapensis*. A partir de esto,

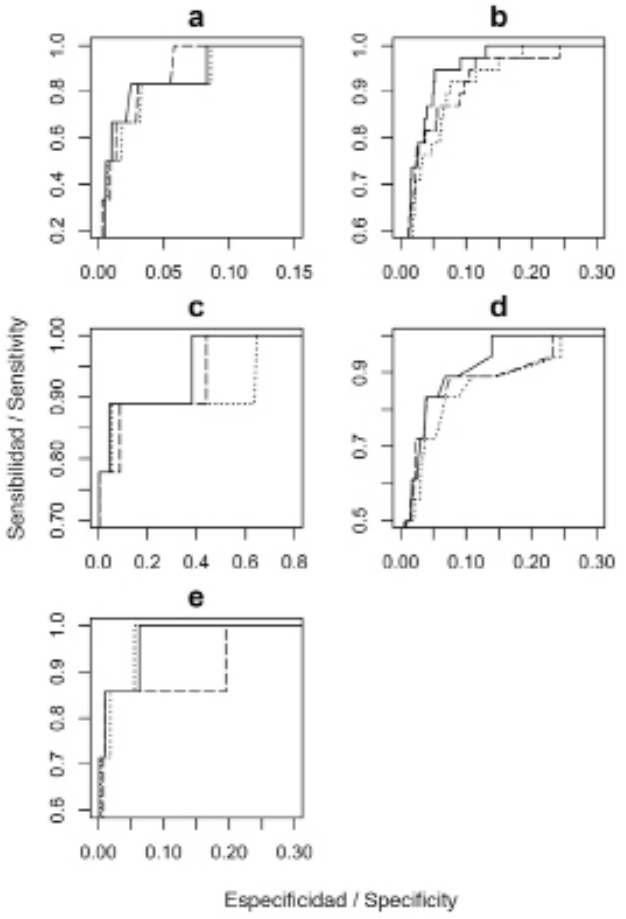

FIGURA 2. Curva COR de las especies evaluadas: a) *Catopheria chiapensis,* b) *Liquidambar styraciflua,* c) *Quercus martinezii,* d) *Telanthophora grandifolia,* e) *Viburnum acutifolium*. Modelo 1 (56 covariables; línea continua), el modelo 2 (19 covariables; línea de puntos) y modelo 3 (covariables seleccionadas utilizando el método propuesto; línea discontinua).

FIGURE 2. Receiver operating characteristic (ROC) curves for the species evaluated: a) *Catopheria chiapensis,* b) *Liquidambar styraciflua,* c) *Quercus martinezii,* d) *Telanthophora grandifolia,* e) *Viburnum acutifolium*. Model 1 (56 covariables; solid line), model 2 (19 covariables; dotted line) and model 3 (covariables selected using the method proposed; longdash line).

HMF the abundant rainfall as well as a low rate of loss for evapotranspiration. In addition, Vogelman (1973) suggested that the moderate temperatures and the high atmospheric humidity are the main environmental determinants of the HMF vegetation. Moreover, the most important covariable (i. e., with the highest point total in the matrix resulting from the bootstrap method) for each species was: the normalized vegetation index of May (NDVIMAY2009) for *C. chiapensis,* the digital elevation model (DEM) for *L. styraciflua* and *V. acutifolium*, pH for *Q. martinezii,* and the average range of daily temperatures (bio2) for *T. grandifolia*.

notes

CUADRO 3. Importancia de los modelos con número distinto de covariables mediante la prueba binominal.

Especies	Modelo 1 (56 covariables)	Modelo 2 (19 covariables WORLDCLIM)	Modelo 3 (covariables selecionadas para bootstrap)
Catopheria chiapensis	0.0160	0.0160	0.0160
Liquidambar styraciflua	< 0.0001	< 0.0001	< 0.0001
Quercus martinezii	0.0190	0.0190	0.0190
Telanthopora grandifolia	< 0.0001	<0.0001	<0.0001
Viburnum acutifolium	0.0620	0.0070	0.0070

TABLE 3. Significance of models with distinct number of covariables using the binomial test.

Species	Model 1 (56 covariables)	Model 2 (19 covariables WORLDCLIM)	Model 3 (covariables selected for bootstrap)
Catopheria chiapensis	0.0160	0.0160	0.0160
Liquidambar styraciflua	< 0.0001	< 0.0001	< 0.0001
Quercus martinezii	0.0190	0.0190	0.0190
Telanthopora grandifolia	< 0.0001	<0.0001	<0.0001
Viburnum acutifolium	0.0620	0.0070	0.0070

CUADRO 4. Intervalos de confianza (BC_α) de 10,000 repeticiones de valores del área bajo la curva (ABC). Entre paréntesis se muestra el valor absoluto de la diferencia entre los valores mínimos y máximos del intervalo de confianza.

Especies	Modelo 1 (56 covariables)		Modelo 2 (19 covariables WORLDCLIM)	Modelo 3 (covariables seleccionadas para bootstrap)
*Catopheria chiapensis**	0.925-0.994	[0.069]	0.940-0.991 [0.051]	0.962-0.993 [0.031]
Liquidambar styraciflua	0.971-0.989	[0.018]	0.953-0.983 [0.030]	0.951-0.984 [0.033]
*Quercus martinezii**	0.819-0.995	[0.176]	0.642-0.996 [0.354]	0.695-0.989 [0.294]
Telanthopora grandifolia	0.940-0.986	[0.046]	0.895-0.978 [0.083]	0.902-0.983 [0.081]
*Viburnum acutifolium**	0.886-0.998	[0.112]	0.963-0.997 [0.034]	0.961-0.999 [0.038]

* La especie tuvo menos de 50 registros.

TABLE 4. Confidence intervals (BC_α) from 10,000 repetitions of values of the area under the curve (AUC). The absolute value of the difference between the maximum and minimum values of the confidence interval is indicated in brackets.

Species	Model 1 (56 covariables)		Model 2 (19 covariables WORLDCLIM)	Model 3 (covariables selected for bootstrap)
*Catopheria chiapensis**	0.925-0.994	[0.069]	0.940-0.991 [0.051]	0.962-0.993 [0.031]
Liquidambar styraciflua	0.971-0.989	[0.018]	0.953-0.983 [0.030]	0.951-0.984 [0.033]
*Quercus martinezii**	0.819-0.995	[0.176]	0.642-0.996 [0.354]	0.695-0.989 [0.294]
Telanthopora grandifolia	0.940-0.986	[0.046]	0.895-0.978 [0.083]	0.902-0.983 [0.081]
*Viburnum acutifolium**	0.886-0.998	[0.112]	0.963-0.997 [0.034]	0.961-0.999 [0.038]

* Species had less than 50 records.

notas

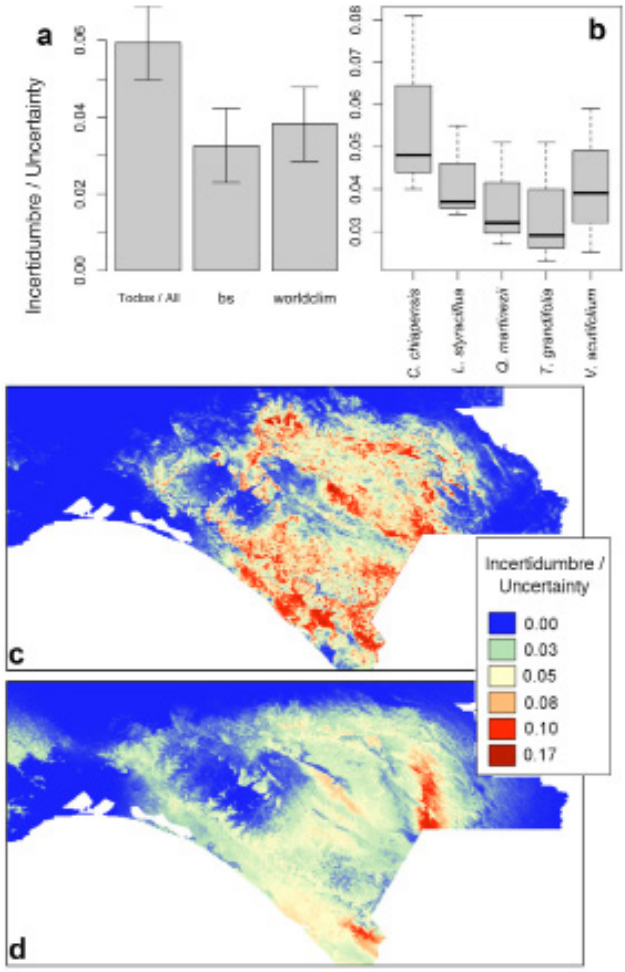

FIGURA 3. Incertidumbre en los modelos de distribución de especies. a) barra de error de incertidumbre de modelos establecidos de covariables: Todos = 56 covariables, bs = covariables seleccionadas utilizando el método propuesto y worldclim = 19 covariables; b) gráfica de cajas de incertidumbre del modelo de cada especie: *Catopheria chiapensis, Liquidambar styraciflua, Quercus martinezii, Telanthophora grandifolia, Viburnum acutifolium;* c) Incertidumbre de distribución del modelo de *C. chiapensis*, 56 covariables como datos de entrada; d) Incertidumbre de distribución del modelo de *C. chiapensis*, covariables seleccionadas utilizando el método propuesto como datos de entrada.

FIGURE 3. Uncertainty in species distribution models. a) error bar of the uncertainty of the models set of covariables: all = 56 covariables, bs = Covariables selected using the method proposed, y worldclim = 19 covariables; b) boxplot of model uncertainty by species: *Catopheria chiapensis, Liquidambar styraciflua, Quercus martinezii, Telanthophora grandifolia, Viburnum acutifolium;* c) model distribution uncertainty from *C. chiapensis*, 56 covariables as input data; d) model distribution uncertainty from *C. chiapensis*, covariables selected using the method proposed as input data.

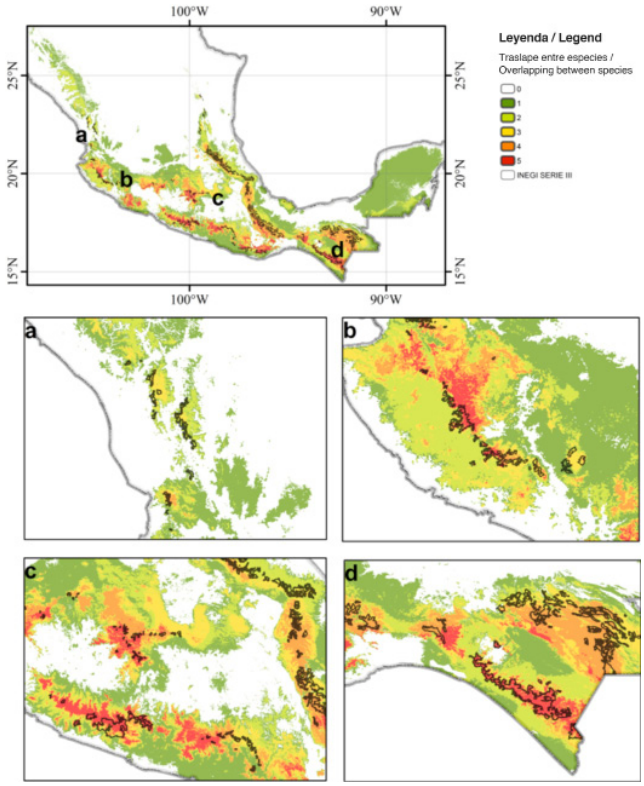

FIGURA 4. Mapa que muestra el traslape de la distribución potencial de *Catopheria chiapensis, Liquidambar styraciflua, Quercus martinezii, Telanthophora grandifolia, Viburnum acutifolium* y las áreas de BHM de acuerdo con el INEGI (2005).

FIGURE 4. Map showing the overlay of the potential distribution of *Catopheria chiapensis, Liquidambar styraciflua, Quercus martinezii, Telanthophora grandifolia, Viburnum acutifolium* and the areas of HMF according to INEGI (2005).

The technique proposed here for the selection of covariables avoids the problem found by Austin and Tu (2004b) with the bootstrap resampling. These authors identified 940 subgroups of covariables as independent predictors from 1,000 bootstrap samples and less than 1 % of those subgroups were repeated four times. In our case, other criteria were taken into account to select covariables, and the ROC test showed higher consistency. The bootstrap technique for the selection of covariables is efficient because it allows an *a priori* reduction of the number of covariables used in the modeling of the potential distribution of species without reducing precision. On the contrary, it is recommended to reduce the number of variables to avoid overfitting the model, especially when the number of records is low (Stockwell & Peterson, 2002).

In terms of model validation with the binomial test, all models except model 1 for *V. acutifolium* were better than random expectations (Table 3). Moreover, the ROC curve analyses indicated that no significant differences ($P <$

notes

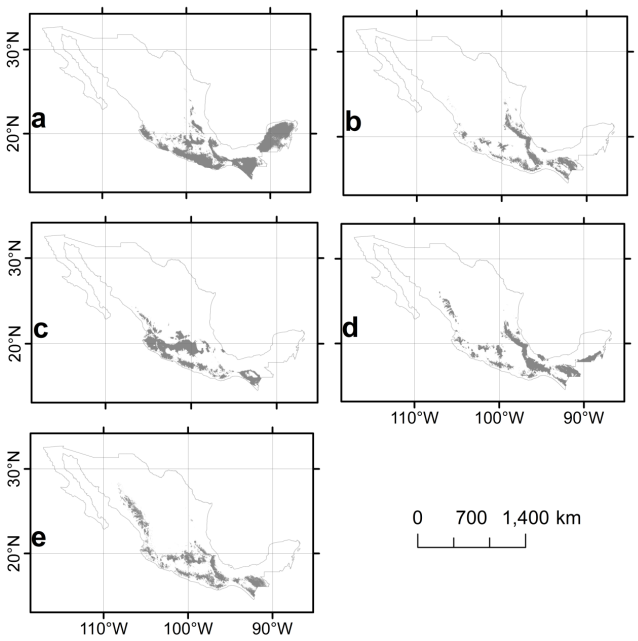

FIGURA 5. Distribución potencial de las especies utilizadas en el análisis: a) *Catopheria chiapensis,* b) *Liquidambar styraciflua,* c) *Quercus martinezii,* d) *Telanthophora grandifolia,* e) *Viburnum acutifolium*.

FIGURE 5. Potential distribution of species used in the analysis: a) *Catopheria chiapensis,* b) *Liquidambar styraciflua,* c) *Quercus martinezii,* d) *Telanthophora grandifolia,* e) *Viburnum acutifolium*.

se utilizaron los modelos de distribución de *C. chiapensis* para ilustrar la variación geográfica de incertidumbre (Figura 3c y 3d). En la Figura 3c (con 56 covariables) se muestra una incertidumbre mayor en casi toda el área prevista en comparación con la Figura 3d (con las covariables seleccionadas por bootstrap). Esto sugiere que una selección *a priori* de las covariables permite reducir la incertidumbre desde el punto de vista espacial.

Las covariables que explican mejor la distribución potencial del conjunto de especies evaluadas están estrechamente relacionadas con el medio ambiente en el que se distribuye el BHM en México (Instituto Nacional de Estadística y Geografía [INEGI], 2005; Figura 4). Sin embargo, algunas de las especies analizadas no son exclusivas del BHM, por lo tanto, los modelos se ampliaron a los sitios con otros biomas (Figura 5). Por ejemplo, *C. chiapensis* también se encuentra en ambientes tropicales y templados del sureste de México, tal y como preside su modelo de distribución potencial (Figura 5a). Lo mismo sucede para *Q. martinezii*, cuya distribución potencial incluye áreas húmedas y templadas de montaña en el centro y sureste de México (Figura 5d).

Las áreas de co-ocurrencia de las cinco especies coinciden con las delimitaciones anteriores del BHM en México (Figura 4) (Challenger & Caballero, 1998; Luna-Vega et al., 2006; Ramírez-Marcial et al., 2001; Rzedowski, 1996; 0.05) in precision existed among the three models for each species (Figure 2). The distribution of species of plants from the three models is better than random distribution as the curves in Figure 2 are projected to the northwest o greater than 0.95 AUC (Fawcett, 2006), concluding that the predictions of the generated models contain mainly true positives (sensitivity) and fewer false negatives (1-specificity). The bootstrap confidence intervals (BC_α) of the three models for each species are presented in Table 4. The BC_α of the area under the curve (AUC) indicate that no significant differences ($P < 0.05$) between the three models existed; however, the most parsimonious model was consistently model 3. In species with less than 50 records the BC_α value of model 1 were quite large, as opposed to model 3, suggesting that with few records the selection of variables with the bootstrap method resulted more adequate.

Figure 3 shows the uncertainty of the species distribution models. There was a significant difference ($P < 0.05$) of the uncertainty of the models generated with 56 variables and that generated with the variables selected by bootstrap (Figures 3a). In contrast, there were not differences among models as they used both the 19 worldclim covariables and the covariables selected by the proposed technique. However, the models uncertainty was smaller using the selected covariables. The boxplot shows that there is a higher variation in the values of uncertainty of the distribution models of *C. chiapensis*. Based on that, the *C. chiapensis* distribution models were used to illustrate the uncertainty geographical variation (Figure 3c and 3d). Figure 3c (with 56 covariables) shows a higher uncertainty in most of the predicted area as compared with Figure 3d (with the covariables selected by bootstrap). That suggests that *a priori* selection of covariables allows to reduce the uncertainty from the spatial point of view.

The covariables that best explained the potential distribution of the set of species evaluated were closely related to the environment in which the HMF is distributed in Mexico (Instituto Nacional de Estadística y Geografía [INEGI], 2005; Figure 4). Nevertheless, some of the species analyzed are not exclusive to the HMF, thus models extended to sites with other biomes (Figure 5). For example, *C. chiapensis* also inhabits tropical and temperate environments of southeastern Mexico, just as its model of potential distribution predicts (Figure 5a). The same is true for *Q. martinezii*, whose potential distribution includes areas of humid and temperate mountain environments of central and southeastern Mexico (Figure 5d).

The areas of co-occurrence of the five species concur with previous delimitations of the HMF in Mexico (Figure 4) (Challenger & Caballero, 1998; Luna-Vega et al., 2006; Ramírez-Marcial et al., 2001; Rzedowski, 1996; Villaseñor, 2010), supporting our results. This biome is mainly found along chains of mountains, in the Sierra Madre Occidental, from Sonora to Jalisco, in the Sierra Madre Oriental, from

Villaseñor, 2010), respaldando nuestros resultados. Este bioma se encuentra principalmente a lo largo de las cadenas montañosas, en la Sierra Madre Occidental, desde Sonora hasta Jalisco, en la Sierra Madre Oriental, desde el sur de Tamaulipas al centro de Veracruz, en la Sierra Madre del Sur, en Guerrero y Oaxaca, en el Sierra Norte de Oaxaca, a lo largo de la Faja Volcánica Transmexicana, y en la Sierra Madre de Chiapas.

CONCLUSIONES

Los modelos de distribución de especies analizados fueron mejorados con la inclusión de covariables de suelo, topografía y datos de sensores remotos. Por lo tanto, se recomienda el uso de covariables auxiliares y bioclimáticas, para el modelado de la distribución de especies. El método de bootstrap es útil para la selección *a priori* de covariables utilizadas en el modelado. Sin embargo, se requiere la presencia y la ausencia de datos para que este método pueda ser aplicado. Una alternativa para la definición de la ausencia se basa en el conocimiento de la historia de vida de las especies objetivo; por ejemplo, al conocer los requisitos climáticos o el rango de altitud donde se encuentra una especie, es posible definir los sitios que pueden funcionar como ausencias.

AGRADECIMIENTOS

El primer autor agradece a la Dirección General de Asuntos del Personal Académico de la Universidad Nacional Autónoma de México (UNAM), por dos años de beca postdoctoral (2010-2012) para llevar a cabo esta investigación bajo la supervisión del Dr. José Luis Villaseñor del Instituto de Biología.

LITERATURA CITADA

Austin, P. C., & Tu, J. V. (2004a). Bootstrap methods for developing predictive models. *The American Statistician*, *58*(2), 131–137. Obtenido de http://www.jstor.org/discover/10.2307/27643521?uid=3738664&uid=2129&uid=2&uid=70&uid=4&sid=21102531764551

Austin, P. C., & Tu J. V. (2004b). Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. *Journal of Clinical Epidemiology, 57*(11), 1138–1146. doi: 10.1016/j.jclinepi.2004.04.003

Bruijzeel, L. A., Waterloo, M. J., Proctor, J., Kuiters, A. T., & Kotterink, B. (1993). Hydrological observations in montane rain forest on Gunung Silam, Sabah, Malasya, with special reference to the 'Massenerhebung' effect. *Journal of Ecology*, *81*(1), 141–167. http://www.jstor.org/discover/10.2307/2261231?uid=3738664&uid=2&uid=4&sid=21103215050991

Challenger, A., & Caballero, J. (1998). *Utilización y conservación de los ecosistemas terrestres de México: Pasado, presente y futuro*. México: Comisión Nacional para el Conocimiento y Uso de la Biodiversidad.

the south of Tamaulipas to Central Veracruz, in the Sierra Madre del Sur, in Guerrero and Oaxaca, in the Sierra Norte of Oaxaca, along the Trans-volcanic Belt, and in the Sierra Madre of Chiapas.

CONCLUSIONS

The species distribution models analyzed were improved with the inclusion of covariables of soil, topography and remote sensing data. Therefore, we recommend using ancillary covariables plus bioclimatic, for species distribution modeling. The bootstrap method was useful for selecting covariables *a priori* modeling. However, it requires the presence and absence data for this method may be applied. Absence data are usually not available HMF one alternative for defining absences is based on the knowledge of the life history of target species; for example, by knowing the climate requirements or the altitude range where a species is found, it is possible to define sites that can function as absences.

ACKNOWLEDGMENTS

First author thanks to the Dirección General de Asuntos del Personal Académico of Universidad Nacional Autonoma de Mexico (UNAM), for a two years postdoctoral fellowship support (2010-2012) to conduct this research under the supervision from José Luis Villaseñor at the Instituto de Biología.

End of English Version

Cimmery, V. (2010). *SAGA User Guide, updated for SAGA version 2.0.5*. USA: Geosystem Analysis. Obtenido de http://www.saga-gis.org/en/index.html

Cruz-Cárdenas, G., López-Mata, L., Ortiz-Solorio, C. A., Villaseñor, J. L., Ortiz, E., Silva, J. T., & Estrada-Godoy, F. (2014). Interpolation of Mexican soil properties at a scale of 1: 1,000,000. *Geoderma*, *213*, 29–35. doi: 10.1016/j.geoderma.2013.07.014

Der, G., & Everitt, B. S. (2002). *Handbook of statistical analyses using SAS*. USA: CRC Press.

D'heygere, T., Goethals, P. L., & De Pauw, N. (2003). Use of genetic algorithms to select input variables in decision tree models for the prediction of benthic macroinvertebrates. *Ecological Modelling, 160*(3), 291–300. doi: 10.1016/S0304-3800(02)00260-0

D'heygere, T., Goethals, P. L., & De Pauw, N. (2006). Genetic algorithms for optimization of predictive ecosystems models based on decision trees and neural networks. *Ecological Modelling, 195*(1-2), 20–29. doi: 10.1016/j.ecolmodel.2005.11.005

DiCiccio, T., & Efron, B. (1996). Bootstrap confidence intervals. *Statatical Science, 11*, 189–228. Obtenido de http://www.jstor.org/discover/10.2307/2246110?uid=3738664&uid=2129&uid=2&uid=70&uid=4&sid=21102531814331

Dimitris, R. (2009). *Bootstrap stepAIC. R package version 1.2-0*. Vienna, Austria: R Foundation for Statistical Computing. Obtenido de http://cran.r-project.org/web/packages/bootStepAIC/bootStepAIC.pdf

Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 677–697. doi: 10.1146/annurev.ecolsys.110308.120159

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*(1), 43–57. doi: 10.1111/j.1472-4642.2010.00725.x

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861–874. doi: 10.1016/j.patrec.2005.10.010

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978. doi: 10.1002/joc.1276

Instituto Nacional de Estadística y Geografía (INEGI). (2005). Mapa de uso de suelo y vegetación 1:250000. Obtenido de http://www.inegi.org.mx/geo/contenidos/recnat/usosuelo/

Kumar, S., & Stohlgren, T. J. (2009). Maxent modeling for predicting suitable habitat for threatened and endangered tree *Canacomyrica monti*cola in New Caledonia. *Journal of Ecology and natural Environment*, *1(*4), 94–98. Obtenido de http://www.academicjournals.org/jene/PDF/Pdf2009/July/Kumar%20and%20Stohlgren.pdf

Luna-Vega, I., Alcantara-Ayala, O., Ruíz-Pérez, C. A., & Contreras-Medina, R. (2006). Composition and structure of humid montane oak forests at different sites in central and eastern Mexico. In Kapelle, M. (Ed.), *Ecology and conservation of neotropical montane oak forests* (pp. 101–112). New York, USA: Springer-Verlag.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling, 190*(3), 231–259. doi: 10.1016/j.ecolmodel.2005.03.026

Phillips, S. J., & Dudik, M. (2008). Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography, 31*(2), 161–175. doi: 10.1111/j.0906-7590.2008.5203.x

R Development Core Team (2010). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Obtenido de http://www.R-project.org/

Ramıírez-Marcial, N., González-Espinosa, M., & Williams-Linera, G. (2001). Anthropogenic disturbance and tree diversity in montane rain forests in Chiapas, Mexico. *Forest Ecology and Management, 154*(1), 311–326. doi: 10.1016/S0378-1127(00)00639-3

Riley, S. J., DeGloria, S. D., & Elliot, R. (1999). A terrain ruggedness that quantifies topographic heterogeneity. *Intermountain Journal of Sciences, 5*(1-4), 23–27. Obtenido de http://download.osgeo.org/qgis/doc/reference-docs/Terrain_Ruggedness_Index.pdf

Rzedowski, J. (1996). Análisis preliminar de la flora vascular de los bosques mesófilos de montaña de México. *Acta Botánica de México, 35*, 25–44. Obtenido de http://www.redalyc.org/articulo.oa?id=57403504

Sappington, J., Longshore, K. M., & Thompson, D. B. (2007). Quantifying landscape ruggedness for animal habitat analysis: A case study using bighorn sheep in the Mojave Desert. *Journal of Wildlife Management, 71(5)*, 1419–1426. doi: 10.2193/2005-723

Stockwell, D. R., & Peterson, A. T. (2002). Effects of sample size on accuracy of species distribution models. *Ecological Modelling 148*(1), 1–13. doi: 10.1016/S0304-3800(01)00388-X

Turc, L. (1954). Le bilan d'eau des sols: Relations entre les precipitation, l'évaporation et l'écoulement. *Annales Agronomiques, 5*, 491–596.

United State Geological Survey (USGS). (2010). Global 30 Arc-Second Elevation (GTOPO30). Obtenido de https://lta.cr.usgs.gov/GTOPO30

United State Geological Survey (USGS). (2010). The USGS global visualization viewer. Obtenido de http://glovis.usgs.gov/

Vázquez-García, J. A. (1995). Cloud forests archipelagos: Preservation of fragmented montane ecosystems in tropical America. In L. S. Hamilton, J. O. Juvik, & and F. N. Scatena (Eds.), *Tropical montane cloud forests* (pp. 315–332). London: Springer.

Venables, W. N., & Ripley, B.D. (2010). *stepAIC: MASS. R package version 7.3-9*. Vienna, Austria: R Foundation for Statistical Computing. Obtenido de http://cran.stat.ucla.edu/

Villaseñor, J. L. (2010). *El bosque húmedo de montaña en México y sus plantas vasculares: Catálogo florístico-taxonómico.* México: CONABIO-UNAM.

Vogelman, H. M. (1973). Fog precipitation in the cloud forest of Eastern Mexico. *BioScience*, *23(2),* 96–100. Obtenido de http://www.jstor.org/stable/1296569