

Distribución espacial de la riqueza de especies de plantas vasculares en México

Spatial distribution of species richness of vascular plants in Mexico

Gustavo Cruz-Cárdenas^{1,3}✉, José Luis Villaseñor¹, Lauro López-Mata² y Enrique Ortiz¹

¹Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México, Apartado postal 70-233, 04510 México, D. F., México.

²Colegio de Postgraduados, Campus Montecillo, Carr. México-Texcoco Km 36.5, 56230 Texcoco, Estado de México, México.

³CIIDIR, Instituto Politécnico Nacional-Michoacán, COFAA, Justo Sierra 28, 59510 Jiquilpan, Michoacán, México.

✉ gustavo.cruz.cardenas@gmail.com

Resumen. En diversos campos como biología y ecología, la información sobre la riqueza de especies y su distribución geográfica es fundamental para la toma de decisiones. Sin embargo, existen países que cuentan con información limitada a nivel nacional, como es el caso de México. Por lo tanto, consideramos importante generar un mapa de la distribución de la riqueza conocida y estimada de especies de plantas vasculares a nivel nacional. Para cumplir tal objetivo y mediante el uso de 2 métodos geoestadísticos (Kriging universal y Co-Kriging), se realizó la predicción espacial de riqueza de especies a partir de información contenida en celdas de $1^\circ \times 1^\circ$. Los resultados muestran que en México la riqueza varía desde 20 hasta 3 800 especies. Los estados con mayor riqueza conocida y estimada de especies son Chiapas, Guerrero y Oaxaca. Las 2 técnicas geoestadísticas empleadas demostraron ser una herramienta eficaz para calcular la predicción espacial de la riqueza de especies de plantas vasculares, debido a que el error medio y la media estandarizada del error de predicción fue cercano a 0.

Palabras clave: geoestadística, biotaxón, predictores ambientales, regresión-kriging.

Abstract. In many fields of biology, information on species richness and geographic distribution is essential for decision-making. However, Mexico as many other countries does not have this information at national level; therefore we consider important to generate information about the distribution of species richness both known and estimated at national level. In order to fulfill this objective and through the use of 2 geostatistical techniques (Kriging universal and Co-Kriging), we performed the spatial prediction of species richness from information contained in cells of $1^\circ \times 1^\circ$. Results showed the occurrence of areas in Mexico with richness varying from 20 to 3 800 species. The states with the highest number of species are Chiapas, Guerrero and Oaxaca. The 2 geostatistical techniques employed showed to be efficient tools to estimate spatial predictions of species richness.

Key words: geostatistics, biotaxon, environmental predictors, regression-kriging.

Introducción

La riqueza de especies, así como la ubicación de centros de diversificación y endemismo (hot-spots), son parámetros útiles en la toma de decisiones. Por ejemplo, si el objetivo es conservar áreas de excepcional valor biótico, determinar corredores para el libre movimiento de la biodiversidad o mitigar problemas ambientales o de salud, la información que proporciona la riqueza de especies siempre es atendida en primera instancia (Carroll y Pearson, 1998).

El conocer la riqueza de un sitio puede ayudar a evaluar la desaparición de especies promovida por el cambio de uso del suelo. Por ejemplo, en un estudio en Dinamarca hecho para diferentes tipos de vegetación se

ilustra cómo las actividades humanas han contribuido a la reducción de poblaciones de *Tilia cordata* Mill. y *Alnus glutinosa* (L.) Gaertn. y a la expansión de *Fagus sylvatica* L., y, consecuentemente, a la reducción de la biodiversidad (Bradshaw y Holmqvist, 1999).

En México algunos estudios han mostrado que los valores de riqueza de especies son de utilidad para determinar el estado de conservación de algunas áreas (Juárez-Jaimes et al., 2007; Domínguez-Domínguez et al., 2008; Aguirre y Duivenvoorden, 2010). Sin embargo, estos estudios han sido a nivel local o regional. A escala nacional se carece de información de calidad en cuanto a la taxonomía y distribución geográfica de las especies. Desde hace varios lustros, a partir de la revisión de literatura y de consultas en diversos herbarios, tanto nacionales como del extranjero, se ha generado información para todo México de los valores de riqueza de especies de plantas vasculares a escalas pequeñas, como, por ejemplo, celdas de $1^\circ \times$

1° de latitud y longitud (Villaseñor et al., 2005a, 2007). No obstante, esta información muestra autocorrelación espacial, es decir, la distribución espacial de la riqueza de especies presenta a nivel espacial un patrón de continuidad y de dependencia (Carroll y Pearson, 1998). Estos 2 factores son importantes de tomarse en cuenta cuando el objetivo es evaluar patrones de distribución.

Los métodos estadísticos clásicos ignoran este problema, pues dan por hecho la estacionalidad en el espacio y el tiempo, la independencia entre los datos y una distribución idéntica de los parámetros. Sin embargo, estos supuestos no siempre se cumplen (Rossi et al., 1992). Por lo tanto, partiendo de la condicional de que el valor de la riqueza de especies de una celda no es independiente del valor de la riqueza de especies de las celdas contiguas y en consecuencia tienen dependencia espacial, para su análisis se han propuesto técnicas geoestadísticas que toman en cuenta estas características de los datos espaciales (Wagner, 2003). Los métodos heurísticos, sin embargo, dan respuesta a este problema. Entre las soluciones propuestas se encuentran el factor de inflación de la varianza (vif), el cual no debe exceder de 10 y el valor más bajo del criterio de información de Akaike (Der y Everitt, 2002).

Las técnicas geoestadísticas se basan principalmente en la distancia geográfica a celdas vecinas más que en el tamaño de celda. Esto brinda la ventaja de que las celdas o las unidades geográficas a comparar no necesariamente

sean contiguas o se encuentren espaciadas a intervalos regulares. La estructura espacial de los datos se describe usualmente mediante un variograma experimental, el cual es básicamente una gráfica de la semivarianza entre pares de observaciones (por ejemplo, celdas) contra su distancia en un espacio geográfico. Un variograma se define mediante modelos teóricos permisibles (exponencial, esférico, logarítmico, etc.) y los parámetros sill (la diferencia del promedio al cuadrado de 2 observaciones independientes), range (la distancia máxima en la cual los pares de observaciones se pueden influenciar o están autocorrelacionados) y nugget (la varianza dentro de una unidad de muestreo) (Wagner, 2003). Una vez que la estructura de autocorrelación espacial se ha determinado con el variograma, es posible, por ejemplo, hacer una interpolación con el método Kriging para estimar matemáticamente la riqueza de especies y llenar las lagunas de información.

Otra aplicación de la geoestadística es utilizar taxa indicadores (sustitutos) para determinar la riqueza de especies en áreas que han sido pobemente inventariadas (Carroll y Pearson, 1998). Esto se puede hacer a través del análisis de un variograma cruzado (cross-variogram), el cual describe la dependencia espacial entre 2 variables medidas (Mulla y McBratney, 2002). Como en este caso son 2 variables (el taxón sustituto y los taxa desconocidos), se emplea un Co-Kriging para realizar la interpolación.

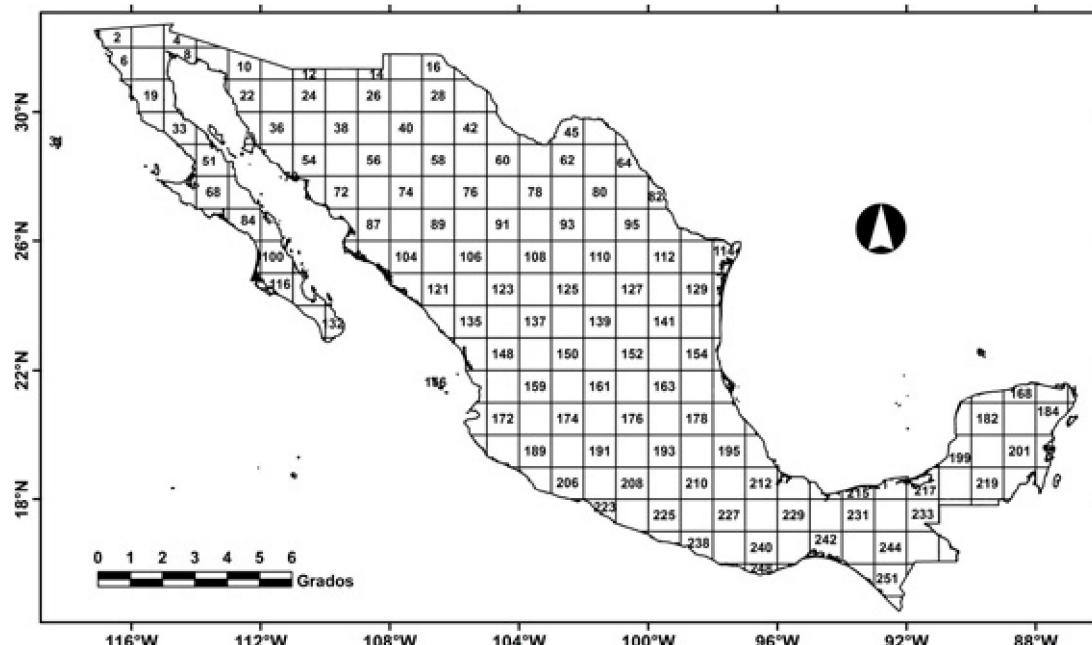


Figura 1. Mapa de México mostrando la división en celdas de 1° de latitud y longitud utilizada en este estudio (se indican el número de celdas).

El objetivo de este trabajo es generar un mapa de superficie de la riqueza estimada de especies de plantas vasculares para México, utilizando una malla de celdas de 1° de latitud y 1° de longitud. Para ello, se utilizaron técnicas geoestadísticas que permiten, como ya se indicó, analizar la distribución espacial de los valores de riqueza conocidos.

Materiales y métodos

Área de estudio. El área de estudio comprende el territorio de la República Mexicana, el cual abarca una superficie aproximada de 1 949 359 km². La superficie fue dividida en celdas de 1° de latitud y 1° de longitud (Fig. 1). En total se obtuvieron 253 celdas.

Datos de riqueza de especies. La información de riqueza de especies por celda se obtuvo a partir de la revisión intensa (aunque no exhaustiva) de la literatura florístico-taxonómica de México, de consultas a herbarios, tanto de México como del extranjero (detalles de gran parte de esta revisión en Villaseñor, 2003) y de algunos ejemplares citados en la base de datos en línea del Jardín Botánico de Missouri (<http://www.tropicos.org>). Se registraron 22 928 especies. Toda la información concerniente a la distribución geográfica de las especies se georreferenció, de tal manera que fuera posible asignarla a una celda en particular. De esta manera, fue posible calcular la riqueza de especies en cada celda del territorio nacional.

Mapas de riqueza estimada de especies. Kriging: a partir de los datos de riqueza conocida se estimó la riqueza total de especies con el método de interpolación Kriging y Co-Kriging (Apéndice 2). Para emplear la técnica de interpolación con el primer método, se debe ajustar el variograma experimental (ecuación 1; Goovaerts, 1999) de los valores de riqueza total de especies a un modelo permisible.

$$\hat{y}(h) = \frac{1}{2} n \sum_{i=1}^n [Z(x_i + h) - \tilde{Z}(x_i)]^2 \quad (1)$$

donde x_i y $x_i + h$ son localidades muestreadas separadas por una distancia h , y $Z(x_i)$ y $Z(x_i + h)$ son los valores de Z observados, en este caso, riqueza de especies para las localidades correspondientes.

Se evaluaron 7 modelos (esférico, exponencial, gaussiano, lineal, matern, bessel y pentaesférico) y se seleccionó aquel que tuviera el menor valor de error (ecuación 2; Cressie, 1985).

$$SSE = \sum_{i=1}^m w_i [\tilde{\gamma} - \gamma]^2 \quad (2)$$

donde m es número de lag (2 localidades separadas por una distancia determinada), $\tilde{\gamma}$ son los valores de semivarianza

para cada distancia, γ son los valores de semivarianza del modelo de predicción permisible y w_i son los factores de semivarianza obtenidos mediante la ecuación (ecuación 3; Cressie, 1985):

$$w_i = \frac{N}{\tilde{\gamma}^2} \quad (3)$$

donde N es el número de pares de puntos usados para calcular $\tilde{\gamma}$ de cada distancia. Despues de seleccionar el modelo permisible que mejor se ajusta a la semivarianza experimental de los valores de riqueza de especies observada por celda, se procedió a realizar la interpolación con Kriging. Se evaluaron 3 modelos Kriging (simple, ordinario y universal) con la técnica de “validación-cruzada 10-fold”. Se seleccionó el Kriging con menor error de precisión para realizar la interpolación con base en el error medio (ME; ecuación 4), el cual debe ser cercano a 0; de igual manera, la raíz del error cuadrático medio (RMSE; ecuación 5) debe ser menor que la varianza de la muestra y la media estandarizada del error de predicción (MSPE; ecuación 6) debe ser cercana a 0.

$$ME = \frac{1}{N} \sum_{i=1}^N [z(x_i) - \tilde{z}(x_i)] \quad (4)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N [z(x_i) - \tilde{z}(x_i)]^2} \quad (5)$$

$$MSPE = \frac{1}{N} \sum_{i=1}^N \frac{ME}{\sigma^2(x_i)} \quad (6)$$

donde $\tilde{z}(x_i)$ es el valor estimado de riqueza de especies, $z(x_i)$ es el valor de riqueza de especies conocido, N es el tamaño de la muestra y σ es la varianza de los valores medidos de riqueza de especies (Kravchenko y Bullock, 1999). Cabe mencionar que una prueba de χ^2 no se puede aplicar debido a que es sensible al estudio de residuales (Agresti, 2007).

Co-Kriging. Para emplear la técnica de Co-Kriging se utiliza una covariable o predictor que esté correlacionada con la variable de interés (en este caso, la riqueza total de especies). Como variable predictora se utilizaron los datos por celda de las especies de la familia Asteraceae. Esta familia es una buena indicadora a distintos niveles de la jerarquía taxonómica (Villaseñor et al. 2005b, 2007), además de que cumple con los supuestos mencionados por Pearson (1994) para un buen taxón indicador, es decir, su taxonomía es bien conocida y existen evidencias que sus patrones de distribución se correlacionan con otros taxa (Fig. 2).

Para realizar la interpolación con Co-Kriging, primero se debe ajustar un variograma-cruzado experimental, donde $\tilde{\gamma}_{ij}(h)$ describe la covarianza entre 2 especies, la

Figura 2. Relación entre el número total de especies conocidas por celda y el número de especies de Asteraceae (Número de celdas= 253).

i especie de interés y *j*, la especie conocida del taxón sustituto (ecuación 7; Goovaerts, 1999).

$$\begin{aligned} \tilde{\gamma}_{ij}(h) &= \frac{1}{2} n \\ \sum_{i=1}^n [Z(x_i + h) - Z(x_i)] \cdot [Z(x_j + h) - Z(x_j)] \end{aligned} \quad (7)$$

donde x_j y x_j+h son localidades muestreadas separadas por una distancia, h y $Z(x_j)$ y $Z(x_j+h)$ son los valores de Z medidos (riqueza de especies de Asteraceae) para las localidades correspondientes. Una vez ajustado el “cross-variograma”, se realizó la predicción especial con Co-Kriging (Waller y Gotway, 2004).

Clases en mapas de riqueza de especies. Para determinar el intervalo y número de clases de cada una de los mapas, se utilizó el método propuesto por Law et al. (2009), el cual consiste en tomar en cuenta la media y desviación estándar de los valores de la riqueza de especies del mapa para determinar las clases.

Software. Los análisis estadísticos y geoestadísticos se realizaron con el paquete estadístico R (Bivand et al., 2008; R Core Team, 2012). Para ello, se emplearon las librerías rgdal, spdep y gstat. La edición de los mapas se hizo en Quantum GIS 1.7.4 “Wroclaw”.

Resultados

La flora vascular de México registrada fue de 22 928 especies, número muy similar a la cifra de 22 185 especies reportada por Conabio (2008). La riqueza conocida de especies analizada en cada una de las 253 celdas reporta una media de 896 especies, con un valor mínimo de 19 y un máximo de 3 909 (Apéndice 1). Más de la mitad de las

celdas (67%) registra valores de riqueza menores a 1 000 especies, 21% de ellas tienen valores de riqueza entre 1 000 y 2 000 especies y el 12% restante registra más de 2 000 especies. Por otra parte, los valores de riqueza conocida de Asteraceae por celda registran una media de 157, con un valor mínimo de 12 y un máximo de 488. Un porcentaje de celdas similar al de la riqueza total (69%) registra menos de 200 especies, mientras que 24% registran entre 200 y 400 especies y el 7% restante de celdas reporta más de 400 especies de esta familia (Apéndice 1).

De los 7 modelos permisibles evaluados, el modelo pentaesférico fue el que mejor se ajustó a los datos de riqueza total de especies (Fig. 3a). El variograma muestra la existencia de autocorrelación espacial entre los datos de riqueza total de especies a una distancia no mayor de 2.46 grados (range) y más allá de esta distancia la autocorrelación disminuye. Además, presenta una estructura espacial fuerte porque la relación entre el valor de semivarianza del nugget y el sill es menor de 0 (Mulla y McBratney, 2002).

Para realizar la predicción espacial se utilizó Kriging universal, pues dicho método mostró el menor error en comparación con los otros 2 métodos (ordinario y simple), de acuerdo con la prueba de “validación-cruzada 10-fold” (Cuadro 1). Esta predicción espacial se muestra en la figura 3b.

El Kriging universal estima que los valores de riqueza total de especies oscilan desde 71 (la celda con menor riqueza) a 3 086 (la celda con mayor riqueza). Las celdas con menores valores de riqueza se ubican principalmente en sitios con matorral xerófilo o bosque estacionalmente seco. Las zonas con valores de riqueza intermedios se asocian con los bosques templados y los bosques tropicales húmedos, mientras que las celdas con mayor riqueza estimada se ubican en zonas con bosques templados y bosques húmedos de montaña.

El variograma cruzado de la riqueza conocida de especies con la riqueza de Asteraceae se ajustó mejor a un modelo exponencial (Fig. 4a). El mapa que resultó de

Cuadro 1. Estadísticas de la prueba de “validación-cruzada 10-fold” (errores) con las técnicas de Kriging y Co-Kriging

Método	ME	RMSE	MSPE
Co-Kriging	0.0044	0.3935	0.1550
Universal	-0.0162	0.8211	0.6743
Ordinario	-0.0131	0.8553	0.7320
Simple	0.4491	0.9830	0.9663

ME: error medio; RSME: raíz del error cuadrático medio; MSPE: media estandarizada del error de predicción.

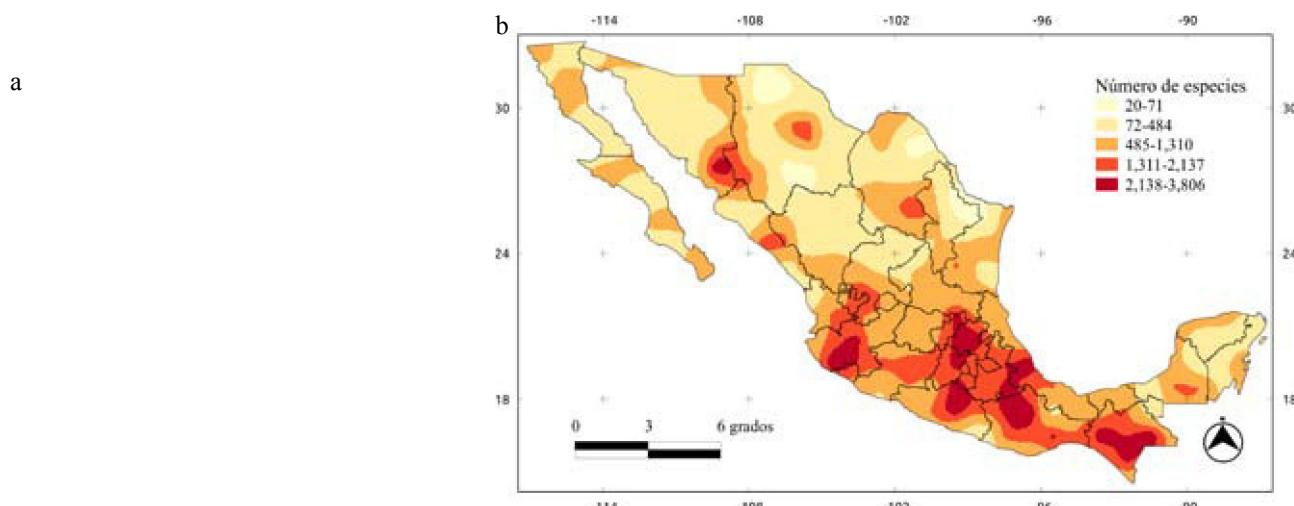


Figura 3. a, variograma que mejor se ajustó a la riqueza total de especies (nugget= 0, range= 2.46, sill= 0.94, modelo teórico= pentagonal); b, mapa de riqueza de especies generado con Kriging universal.

la predicción espacial con Co-Kriging se presenta en la figura 4b. El mapa muestra una distribución de la riqueza total de especies estimada muy similar a la observada con el modelo Kriging universal. Las áreas con menor riqueza de especies se localizan al norte del país y las zonas con mayor riqueza de especies en algunos sitios de la Faja Volcánica Transmexicana y en los estados de Chiapas y Oaxaca.

Discusión

Aunque muy semejantes en sus patrones de riqueza, las diferencias en la distribución de la riqueza estimada de especies entre los 2 mapas generados (Figs. 3b, 4b) radican principalmente en la superficie que cada intervalo de clasificación predice, además del número mínimo y máximo de riqueza total de especies. El mapa generado con Kriging universal predice más áreas con mayor riqueza (más de 1 310 especies), mientras que el mapa obtenido con Co-Kriging predice más áreas con una riqueza total de especies menor a 1 310. Los intervalos de riqueza total de especies estimadas varían de 20 a 3 806 para el mapa generado con Kriging y de 61 a 2 421 para el mapa generado con Co-Kriging, con un promedio general de 765 especies.

En general, se puede afirmar que con las técnicas de predicción espacial utilizadas (Apéndice 2), los mapas generados muestran buena precisión, puesto que presentan errores cercanos a 0 (Cuadro 1), porque se puso atención a la autocorrelación espacial de los datos, ya que los valores de riqueza de especies de un sitio en particular están influenciados por los valores de riqueza de las

localidades vecinas (Jiguet et al., 2005). Sin embargo, de acuerdo con la prueba de “validación-cruzada 10-fold”, el mapa generado con Co-Kriging tiene menor error (Cuadro 1); estos resultados concuerdan con trabajos previos en Geoestadística, que recomiendan usar covariables si es que se cuenta con ellas, pues se obtienen mejores resultados en la predicción espacial (Mulla y McBratney, 2002; Waller y Gotway, 2004; Hengl, 2009).

Varias de las entidades que se registran en este trabajo con mayor riqueza de especies ya han sido discutidos en estudios previos. Por ejemplo, en Chiapas, González-Espinoza et al. (2004) evaluaron la riqueza de especies de árboles, utilizando cuadros de 5×5 minutos. No obstante, la diferencia en escalas, la ubicación de las áreas con mayor riqueza de árboles es bastante similar al patrón de distribución de riqueza de especies mostrado en la figura 4b para el estado. En el estado de Oaxaca la distribución de áreas con más de 1 300 especies sigue una orientación principalmente de norte a sur (Fig. 4b), que coinciden con la distribución de los bosques templados como lo mencionan Suárez-Mota y Villaseñor (2011); dicho bioma templado en este estado no sólo es uno de los que registran mayor riqueza total, sino también mayor número de endemismos en México.

Los resultados de la predicción espacial pueden mejorarse evaluando otras características de la estructura espacial de los puntos, así como empleando técnicas híbridas. Por ejemplo, en este trabajo no se evaluó la anisotropía de los datos, es decir, si la semivarianza tiene igual comportamiento a través del espacio. Una técnica híbrida que valdría la pena explorar es la de Regresión-Kriging (RK). Esta técnica utiliza una regresión con

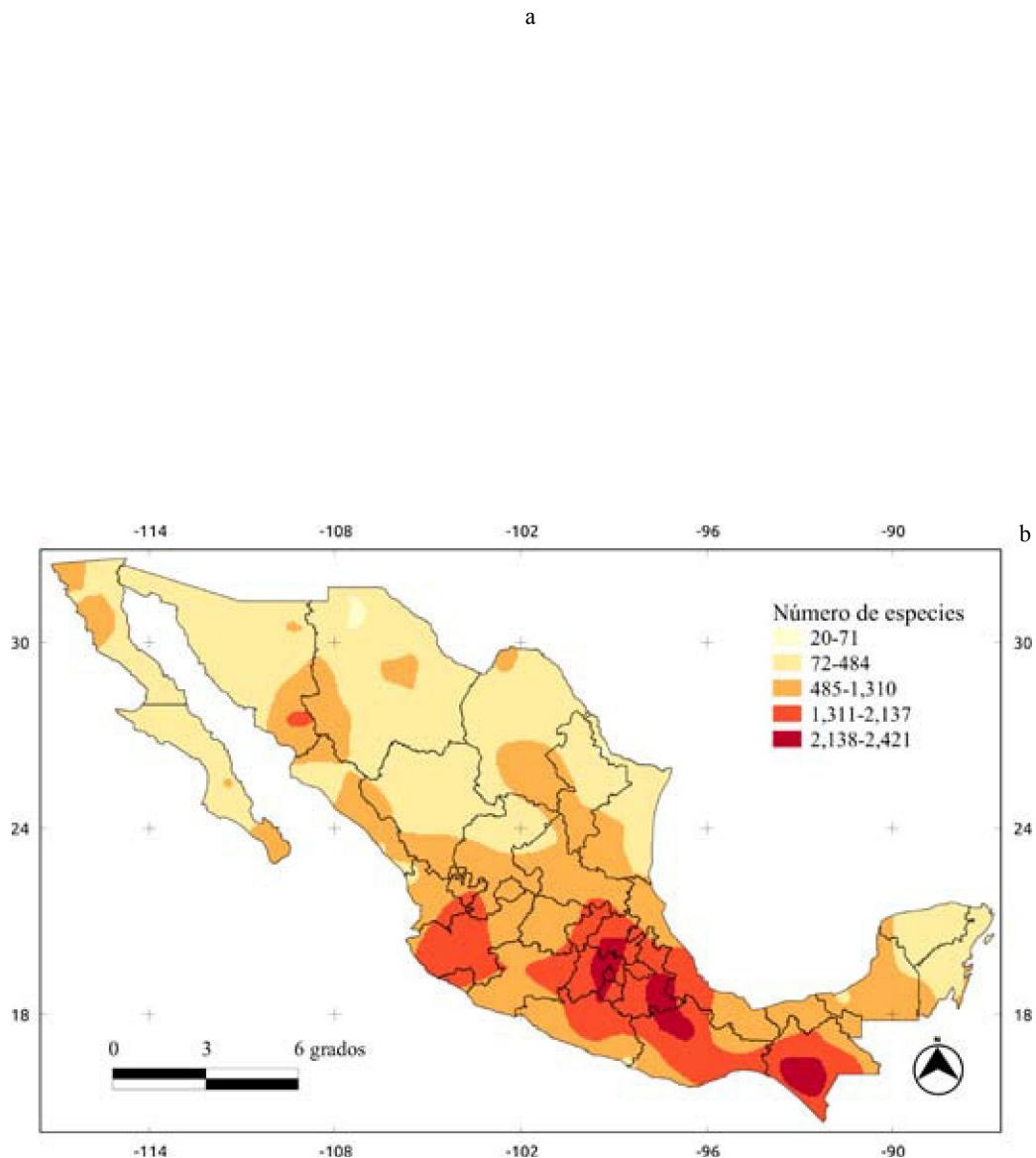


Figura 4. a, variograma-cruzado de la riqueza de especies (RE) conocidas y de la riqueza de Asteraceae (RA) por celda (Número de celdas= 253); b, mapa de riqueza de especies generado con Co-Kriging.

información auxiliar (variables ambientales o biotaxones) y después se usa un Kriging simple con media conocida (0) para interpolar los residuales del modelo de regresión (Hengl et al., 2007). Con la técnica de RK se han obtenido mejores resultados en la generación de mapas de propiedades de suelos (Hengl et al., 2004), tipos de vegetación (Miller et al., 2007) y distribución de especies (Allouche et al., 2008; Hengl et al., 2009).

Otro objetivo posterior importante sería realizar el análisis aumentando la escala de trabajo a minutos. Los mapas de riqueza de especies generados en este trabajo

se hicieron incorporando la información en celdas de 1 × 1 grados. Sin embargo, sería necesario generar mapas de riqueza de especies con información a celdas de mayor escala, para evaluar los patrones de distribución de la riqueza de especies y progresivamente incrementar la escala para su comparación. Esto podría hacerse con información auxiliar o covariables, ya sea con biotaxones o predictores ambientales que ya tengan información a escalas más grandes y emplear alguna de las técnicas de predicción espacial mencionadas. De esta manera se podrá, en el caso de México, evaluar en un futuro la hipótesis de

que la relación entre variable y covariables se mantiene a diferentes escalas de estudio (Pearson y Carroll, 1999).

Agradecimientos

El primer autor agradece la beca posdoctoral recibida a través del Programa de Becas Posdoctorales 2011 de su Dirección General de Asuntos del Personal Académico, UNAM. Así mismo, agradece al Departamento de Botánica del Instituto de Biología todas las facilidades otorgadas durante esta estancia posdoctoral.

Literatura citada

Aguirre, G. J. y F. J. Duivenvoorden. 2010. Can we expect to protect threatened species in protected areas? A case study of the genus *Pinus* in Mexico. *Revista Mexicana de Biodiversidad* 81:875-882.

Agresti, A. 2007. *An introduction to categorical data analysis*. Wiley-Interscience, New York. 360 p.

Allouche, O., O. Steinitz, D. Rotem, A. Rosenfeld y R. Kadmon. 2008. Incorporating distance constraints into species distribution models. *Journal of Applied Ecology* 45:599-609.

Bivand, R., E. Pebesma y V. Gómez-Rubio. 2008. *Applied spatial data analysis with R*. Springer, New York. 374 p.

Bradshaw, R. y B. Holmqvist. 1999. Danish forest development during the last 3000 years reconstructed from regional pollen data. *Ecography* 22:53-62.

Carroll, S. S. y L. D. Pearson. 1998. Spatial modeling of butterfly species richness using tiger beetles (Cicindelidae) as bioindicator taxon. *Ecological Applications* 8:531-543.

Conabio. 2008. Capital natural de México, vol. I: conocimiento actual de la biodiversidad. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, México. 621 p.

Cressie, N. 1985. Fitting variogram models by weighted least squares. *Journal of the International Association for Mathematical Geology* 17:563-586.

Der, G. y B. S. Everitt. 2002. *A handbook of statistical analyses using SAS*. Chapman and Hall/CRC, Boca Raton.

Domínguez-Domínguez, O., L. Zambrano, L. H. Escalera-Vázquez, R. Pérez-Rodríguez y G. Pérez-Ponce de León. 2008. Cambio en la distribución de goodeidos (Osteichthyes: Cyprinodontiformes: Goodeidae) en cuencas hidrológicas del centro de México. *Revista Mexicana de Biodiversidad* 79:501-512.

González-Espinosa, M., J. M. Rey-Benayas, N. Ramírez-Marcial, M. A. Huston y D. Golicher. 2004. Tree diversity in the northern Neotropics: regional patterns in highly diverse Chiapas, Mexico. *Ecography* 27:741-756.

Goovaerts, P. 1999. Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma* 89:1-45.

Hengl, T. 2009. *A practical guide to geostatistical mapping*. Segunda edición. University of Amsterdam. 291 p.

Hengl, T., G. M. B. Heuvelink y A. Stein. 2004. A generic framework for spatial prediction of soil variables based on regression-Kriging. *Geoderma* 122:75-93.

Hengl, T., G. Heuvelink y D. Rossiter. 2007. About regression-Kriging: from equations to case studies. *Computers and Geosciences* 33:1301-1315.

Hengl, T., H. Sierdsema, A. Radovi y A. Dilo. 2009. Spatial prediction of species' distributions from occurrence-only records: combining point pattern analysis, ENFA and regression-Kriging. *Ecological Modelling* 220:3499-3511.

Jiguet, F., R. Julliard, D. Couvet y A. Pettau. 2005. Modeling spatial trends in estimated species richness using breeding bird survey data: a valuable tool in biodiversity assessment. *Biodiversity and Conservation* 14:3305-3324.

Juárez-Jaimes, V., L. O. Alvarado-Cárdenas y J. L. Villaseñor. 2007. La familia Apocynaceae *sensu lato* en México: diversidad y distribución. *Revista Mexicana de Biodiversidad* 78:459-482.

Keitt, T., R. Bivand, E. Pebesma y B. Rowlingson. 2011. rgdal: bindings for the geospatial data abstraction library. R package version 0.7-5.

Kravchenko, A. N. y D. G. Bullock. 1999. A comparative study of interpolation methods for mapping soil properties. *Journal of Agronomy* 91:393-400.

Law, M. C., S. K. Balasundaram, M. H. Husni, O. H. Ahmed y M. H. Harun. 2009. Spatial variability of soil organic carbon in oil palm. *International Journal of Soil Science* 4:93-103.

Miller, J., J. Franklin y R. Aspinall. 2007. Incorporating spatial dependence in predictive vegetation models. *Ecological Modelling* 202:225-242.

Mulla, D. J. y B. A. McBratney. 2002. Soil spatial variability. In *Soil physics companion*, A. Warrick (ed.). CRC, Boca Raton. p. 343-373.

Pearson, D. 1994. Selecting indicator taxa for quantitative assessment of biodiversity. *Philosophical Transactions of the Royal Society of London B* 345:75-79.

Pearson, D. y S. Carroll. 1999. The influence of spatial data on cross-taxon congruence patterns and prediction accuracy of species richness. *Journal of Biogeography* 26:1079-1090.

Pebesma, E. 2004. Multivariable geostatistics in S: the gstat package. *Computers and Geosciences* 30:683-691.

Suárez-Mota, M. E. y J. L. Villaseñor. 2011. Las compuestas endémicas de Oaxaca, México: diversidad y distribución. *Boletín de la Sociedad Botánica de México* 88:55-66.

R Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <http://www.R-project.org/>; última consulta: 08.IV.2013.

Rossi, R. E., D. J. Mula, A. G. Journel y E. H. Franz. 1992. Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecological Monographs* 2:277-314.

Villaseñor, J. L. 2003. Diversidad y distribución de las magnoliophyta de México. *Interciencia* 28:160-167.

Villaseñor, J. L., P. Maeda, J. J. Colín-López y E. Ortiz. 2005a. Estimación de la riqueza de especies de Asteraceae mediante extrapolación a partir de datos de presencia-ausencia. *Boletín de la Sociedad Botánica de México* 76:5-18.

Villaseñor, J. L., G. Ibarra-Manríquez, J. A. Meave y E. Ortiz. 2005b. Higher taxa as surrogates of plant biodiversity in a megadiverse country. *Conservation Biology* 19:232-238.

Villaseñor, J. L., P. Maeda, J. A. Rosell y E. Ortiz. 2007. Plant families as predictors of plant biodiversity in Mexico. *Diversity and Distributions* 13:871-876.

Wagner, H. H. 2003. Spatial covariance in plant communities: integrating ordination, geostatistics, and variance testing. *Ecology* 84:1045-1057.

Waller, L. y C. Gotway. 2004. Applied spatial statistics for public health data. Wiley-Interscience, Hoboken, New Jersey. 494 p.

Apéndice 1. Se indica para cada celda (primera cifra; Fig. 1) los valores de riqueza de especies, tanto conocida (segunda cifra) como estimada con Kriging universal (tercera cifra) y con Co-Kriging (cuarta cifra).

1,38,190,64;	2,1181,218,971;	3,177,338,192;	4,373,336,358;	137,606,518,523;	138,260,634,322;	139,460,555,448;
5,357,296,319;	6,335,248,442;	7,568,249,510;	8,378,264,334;	140,536,647,262;	141,1486,506,1560;	142,378,458,319;
9,461,219,477;	10,457,277,535;	11,611,216,520;	143,215,369,115;	144,63,122,21;	145,114,98,207;	
12,115,347,144;	13,1444,246,1383;	14,936,248,636;	147,265,443,284;	148,610,739,1154;	149,2001,741,1341;	
15,47,219,52;	16,83,110,166;	18,97,241,82;	9,776,187,687;	150,1156,738,885;	151,464,765,840;	152,975,797,1165;
20,115,236,120;	21,32,231,17;	22,123,199,143;	23,99,231,123;	153,589,854,913;	154,607,602,314;	155,242,432,345;
24,144,282,228;	25,954,317,760;	26,957,236,762;	27,19,266,30;	156,334,254,115;	157,523,534,651;	158,1626,864,1383;
28,60,187,85;	29,75,125,40;	30,9,231,16;	32,125,211,176;	159,1369,1159,1519;	160,1202,992,1049;	161,961,937,867;
33,234,170,318;	34,103,184,120;	35,314,172,313;	36,106,231,66;	37,166,280,170;	38,143,436,142;	39,544,375,586;
40,470,263,302;	41,901,204,530;	42,1250,259,1162;	40,576,557,330;	41,632,354,46;	42,164,454,253,463;	
43,142,276,155;	44,953,181,660;	45,1153,228,965;	46,656,362,298;	47,785,345,407;	48,166,486,292;	
46,453,147,287;	47,16,155,23;	49,161,125,153;	50,230,145,181;	49,171,1235,674;	50,546,1233,523;	51,1058,1389,1272;
51,132,244,168;	52,266,196,358;	53,323,239,277;	51,2534,1512,2192;	52,2573,1300,2951;	53,783,903,774;	
54,372,234,423;	55,133,352,100;	56,1004,420,1000;	54,180,128,590,240;	55,181,730,202,469;	56,182,361,336,425;	
57,1444,482,1151;	58,99,477,211;	59,390,251,666;	57,183,348,398,541;	58,184,315,442,366;	59,185,527,347,285;	
60,1260,201,811;	61,119,202,106;	62,149,306,183;	60,186,2,179,14;	61,187,1089,337,477;	62,188,3364,518,3429;	
63,959,212,580;	64,49,216,102;	65,37,144,49;	63,189,2070,771,1861;	64,190,1658,864,1334;	65,191,2025,1083,1924;	
67,480,159,423;	68,571,209,482;	69,543,251,491;	67,192,1831,1397,2048;	68,193,2522,1786,2386;	69,194,1705,1757,1956;	
71,223,381,276;	72,2383,507,3363;	73,1794,585,1316;	71,195,2225,1228,2076;	72,196,2983,542,2696;	73,197,15,821,19;	
74,838,344,835;	75,39,295,50;	76,52,199,65;	74,198,35,383,41;	75,199,643,302,633;	76,200,210,384,420;	
78,151,261,205;	79,236,281,306;	80,96,259,207;	78,201,419,390,360;	79,202,774,342,542;	80,203,180,24,47;	
82,43,125,58;	83,96,229,97;	84,258,224,267;	82,204,13,618,15;	83,205,815,282,1145;	84,206,519,465,1222;	
86,592,395,321;	87,1163,397,1207;	88,742,326,484;	86,207,1283,729,927;	87,208,995,1404,1180;	88,209,2852,1488,2269;	
89,95,224,154;	90,69,199,115;	91,135,181,100;	89,210,1488,1824,1451;	91,211,2627,1609,2741;	92,212,1518,1527,1238;	
93,912,357,568;	94,1098,344,653;	95,642,187,286;	93,213,1406,601,1232;	94,214,2302,475,1751;	95,215,98,679,198;	
96,22,147,41;	97,20,183,22;	98,654,109,421;	96,216,757,368,305;	97,217,284,392,375;	98,218,1479,358,753;	
100,735,189,631;	101,100,253,115;	102,154,226,103;	99,219,1286,283,552;	100,220,381,396,351;	101,221,192,426,151;	
103,106,219,57;	104,404,331,763;	105,533,248,374;	102,222,8,491,2;	103,223,911,301,857;	104,224,1384,598,1508;	
106,210,203,363;	107,84,223,72;	108,181,291,233;	105,225,2906,787,2991;	106,226,1737,1177,2067;	107,227,2515,1390,2325;	
109,713,327,475;	110,1417,371,1156;	111,1137,294,1012;	108,228,3909,1153,3024;	109,229,345,1253,761;	110,230,1108,1076,500;	
112,72,219,96;	113,25,262,41;	114,823,169,601;	111,231,1033,850,810;	112,232,1466,702,1634;	113,233,285,594,508;	
115,109,165,101;	116,291,177,241;	117,260,195,394;	114,234,39,436,85,235,57,434,203;	115,236,56,505,48,237,613,348,289;	116,238,138,770,326;	
118,62,169,52;	119,13,193,27;	120,1487,135,975;	117,238,1533,1019,1435;	118,240,1520,1347,2467;	119,241,2427,1109,1402;	
121,1720,216,986;	122,121,341,94;	123,249,289,451;	118,242,1403,1006,2023;	119,243,3100,896,2888;	120,244,2113,844,2476;	
124,132,345,246;	125,197,493,230;	126,364,505,380;	119,245,3416,591,1314;	121,246,1117,359,557;	122,247,61,809,51;	
127,459,507,804;	128,459,355,981;	129,759,349,532;	120,248,1608,634,1598;	123,249,1274,945,1127;	124,250,582,1085,342;	
130,364,289,314;	131,287,135,373;	132,992,94,811;	125,251,2509,759,2796;	126,252,477,843,653;	127,253,198,1011,224;	

Apéndice 2. “Script” para realizar un ejemplo de cómo calcular la riqueza de especies con los métodos de Kriging y Co-Kriging en R. Para mayor detalle se pueden consultar las librerías gstat (Pebesma, 2004) y rgdal (Keitt et al., 2011). Los comentarios en el Script están marcados con el símbolo de #. En “C” se debe de crear una carpeta que se llame “Ejemplo k_ck” y copiar el archivo “coor_sr.csv”, para después poder ejecutar el Script.

```

library(gstat) #Cargar librería
#####Cargar archivo#####
#Se escribe la ruta y el nombre del archivo "coor_sr.csv"
rte=read.csv("C:/Ejemplo_k_ck/coor_sr.csv")
#Se muestra la estructura del objeto "rte". Es una hoja de datos
#que tiene 5 columnas y 48 filas.
str(rte)
# ##### A n á l i s i s
# Visual#####
#Cálculo de correlación entre riqueza de especies y Asteraceae,
# la cual es de 0.89
cor(rte$tsp,rte$asp)
#Graficar riqueza total de especies vs riqueza de Asteraceae
plot(rte$asp,rte$tsp,xlab="Número de Asteraceae",ylab="Número Total de Especies",
pch=20)
abline(lm(rte$tsp~rte$asp))
text(50,3000,"r=0.89")
#Evaluación de la distribución de las valores de riqueza de especies
#Se realiza una análisis visual de la distribución de la riqueza total de especies y
#Asteraceae. El histograma muestra que no existe distribución normal para
#los valores de riqueza total de especies
par(mfrow=c(2,1))
hist(rte$tsp,main="",xlab="Riqueza total de especies",ylab="Frecuencia")
hist(rte$asp,main="",xlab="Riqueza de Asteraceae",ylab="Frecuencia")
#Se aplica logaritmo a los valores de riqueza para ajustar a una distribución normal.
#En análisis posteriores los valores de riqueza total de especies serán transformados con #logaritmo. Existen otras funciones para realizar la transformación de las variables como #raíz cuadrada y cuarta, log (1 + x), etc.
#Se grafican los valores de riqueza de especies transformados
hist(log(rte$tsp),main="",xlab="Riqueza total de especies",ylab="Frecuencia (log transformación)")
#####Selección de variograma y Kriging#####
#La hoja de datos "rte" se transforma a "SpatialPointsDataFrame",
#formato requerido por la librería "gstat" para aplicarle sus funciones
rte.coor=rte
coordinates(rte.coor)=~x+y
str(rte.coor)
###Selección y ajuste de semivariograma###
#Primero se evalúa la estructura espacial de los datos con la ayuda de variograma
#que grafica la semivarianza y la distancia entre pares de puntos
Variograma.Var=variogram(log(rte$tsp)~x+y,rte.coor)
#Se grafica el semivariograma para obtener los parámetros de inicio para
#hacer su ajuste (sill, range y nugget).
plot(Variograma.Var)
# En el punto 4 se observa la asintota del variograma (sill).
#El valor de 4 se sustituye en los parámetros de sill y range para obtener los parámetros
# de inicio para el ajuste de modelos permisibles
sill=Variograma.Var[4,3]
range=Variograma.Var[4,2]
nugget=min(Variograma.Var[3,])
#Se generan 7 modelos permisibles con los parámetros de inicio
Esférico=vgm(sill,"Sph",range,nugget,variogramModel=Variograma.Var)
Exponencial=vgm(sill,"Exp",range,nugget,variogramModel=Variograma.Var)
Gaussiano=vgm(sill,"Gau",range,nugget,variogramModel=Variograma.Var)
Lineal=vgm(sill,"Lin",range,nugget,variogramModel=Variograma.Var)
Matern=vgm(sill,"Mat",range,nugget,variogramModel=Variograma.Var)
Bessel=vgm(sill,"Bes",range,nugget,variogramModel=Variograma.Var)
Pentaesf=vgm(sill,"Pen",range,nugget,variogramModel=Variograma.Var)
#Se ajustan los modelos permisibles
fit.Esférico=fit.variogram(Variograma.Var,Esférico)
fit.Exponencial=fit.variogram(Variograma.Var,Exponencial)
fit.Gaussiano=fit.variogram(Variograma.Var,Gaussiano)
fit.Lineal=fit.variogram(Variograma.Var,Lineal)
fit.Matern=fit.variogram(Variograma.Var,Matern)
fit.Bessel=fit.variogram(Variograma.Var,Bessel)
fit.Pentaesf=fit.variogram(Variograma.Var,Pentaesf)
#Se evalúan los modelos y se selecciona el que tenga menor error, en este caso, fue el "Lineal"
attr(fit.Esférico,"SSErr")
attr(fit.Esférico,"SSErr")
attr(fit.Exponencial,"SSErr")
attr(fit.Gaussiano,"SSErr")
attr(fit.Lineal,"SSErr")
attr(fit.Matern,"SSErr")
attr(fit.Bessel,"SSErr")
attr(fit.Pentaesf,"SSErr")
#Se observan los parámetros con los que se ajustó el variograma
#sill parcial (psill)=0.574, nugget=0.078 y range=1.719
fit.Lineal

```

```

#Se grafica el variograma ajustado que fue el lineal
plot(Variograma.Var, pl=F,model=fit.Lineal,
,col="blue",pch=20,main="Variograma ajustado con modelo
Linear",xlab="Distancia",ylab="Semivarianza")
###Análisis de de validacion cruzada para calcular la precisión
de3 Kriging###
###Aviso checar que no haya valores con iguales
coordenadas###
#La validación cruzada se realiza empleando 10-fold para medir
los errores de ajuste de3 #kriging, universal, ordinario #y
simple
#Universal
U.cross=krige.cv(log(tsp)~x+y,rte.coor,fit.Lineal,nfold=10)
#Ordinario
O.cross=krige.cv(log(tsp)~1,rte.coor,fit.Lineal,nfold=10)
#Simple
S.cross=krige.cv(log(tsp)~1,rte.coor,fit.Lineal,nfold=10,beta=5)
#Transformar a hoja de datos
Ures=as.data.frame(U.cross)$residual
Ores=as.data.frame(O.cross)$residual
Sres=as.data.frame(S.cross)$residual
#Mean Error(ME)
U.ME=mean(Ures)
O.ME=mean(Ores)
S.ME=mean(Sres)
ME=c(U.ME,O.ME,S.ME)
#Root Mean Squared Error (RMSE)
U.RMSE=sqrt(mean(Ures^2))
O.RMSE=sqrt(mean(Ores^2))
S.RMSE=sqrt(mean(Sres^2))
RMSE=c(U.RMSE,O.RMSE,S.RMSE)
#Mean Standardized Prediction Error (MSPE)
U.MSPE=mean(Ures^2)
O.MSPE=mean(Ores^2)
S.MSPE=mean(Sres^2)
MSPE=c(U.MSPE,O.MSPE,S.MSPE)
datos=c(ME,RMSE,MSPE)
Evaluación=matrix(datos,nrow=3,ncol=3,byrow=TRUE,
dimnames = list(c("ME", "RMSE","MSPE"),
c("Universal", "Ordinario", "Simple")))
#Se imprime una tabla donde se muestran los errores obtenidos
por
#la validación cruzada
Evaluación#Kriging Universal el de menor error
#####Selección de co-variograma y
Co-Kriging#####
#Se modela la co-regionalización de los datos empleando la
función gstat. Donde los modelos se#ajustan simultáneamente
en forma directa y con variograma-cruzado
#Se crea el objeto gstat para especificar los variogramas
experimentales
#Se llena el primer marco del objeto gstat con los valores de
riqueza total
g2=gstat(id = "rte", fórmula = log(tsp)~1, data = rte.coor)
#Posteriormente, se adicionan los valores de riqueza de
Asteraceae, el objeto gstat,
#ahora tiene 2 marcos
g2=gstat(g2,id = "Asteraceae", formula = asp~1,data = rte.
coor)
#Se adicionan parámetros de inicio para ajustar los modelos
g2=gstat(g2,id      = "rte",      model=vgm(psill=0.5744,
"Lin",range=1.72, nugget=0.078))
g2=gstat(g2,id      = "Asteraceae", model=vgm(psill=0.5744,
"Lin",range=1.72, nugget=0.078))
v.cross2 <- variogram(g2)
g2=gstat(g2,id      = "rte",      model=vgm(psill=0.5744,
"Lin",range=1.72, nugget=0.078),fill.all=T)
#Se realiza el ajuste de los modelos
(g2 <- fit.lmc(v.cross2, g2))
#Se grafica variogramas ajustados directo y cruzado
plot(variogram(g2), model=g2$model,col="blue",pch=20)
###Validación cruzada de CoKriging###
c.v=gstat.cv(g2,nfold=10)
#Se obtienen los valores de error del ajuste de los modelos
MEco=mean(c.v$residual)
RMSEco=sqrt(mean(c.v$residual^2))
MSPEco=mean(c.v$residual^2)
MEco
RMSEco
MSPEco
#####Predicción
espacial#####
#Delimitar área de estudio con base en las coordenadas mínimas
y máximas
#de los puntos de muestreo
xmin=min(rte$x)
xmax=max(rte$x)
ymin=min(rte$y)
ymax=max(rte$y)
#Resolución de 1km aprox
reso=0.008333
#Hacer Grid donde se almacenan datos interpolados
grid.xy <- expand.grid(x = seq(xmin,xmax,by=reso),y=
seq(ymax,ymin,by=-reso))
coordinates(grid.xy) <- ~x + y
gridded(grid.xy)=T
###Interpolación Kriging###
#Se realiza la predicción espacial empleando el modelo teórico
lineal y Kriging universal
Predicción=krige(log1p(tsp)~1,rte.coor,grid.xy ,fit.Lineal)
#Se elimina el logaritmo a valores de riqueza total, si es que se
aplica, y se guarda el resultado en #nueva columna la "tsp"
Predicción$tsp=exp(Predicción$var1.pred)
#Se elimina el logaritmo a valores de varianza de la riqueza
total, si es que se aplica y se guarda #los valores en nueva
columna la "vtsp"
Predicción$vtsp=exp(Predicción$var1.var)
###Interpolación Co-Kriging###
Ck=predict.gstat(g2,grid.xy)
# Se elimina el logaritmo a valores de riqueza total, si es que
se aplica, y se guarda el resultado en #en nueva columna
la "ck"

```

```
Ck$ck=exp(Ck$rte.pred)
# Se elimina el logaritmo a valores de varianza de la riqueza
  total, si es que se aplica y se guarda #los valores en nueva
  columna la "ckv"
Ck$ckv=exp(Ck$rte.var)
###Presentación de mapas###
#Se grafican los mapas de riqueza total y la varianza de los
  valores de riqueza total
krig=spplot(Predicción[“tsp”],main=list(“a”),cex=2),scales
  = list(draw = T),ylab=list(“Coordenada Y”,cex=1.3),
col.regions = bpy.colors(100)
co.krig=spplot(Ck[“ck”],main=list(“b”),cex=2),scales=list(draw
  = T),xlab=list(“Coordenada X”,cex=1.3),
col.regions = bpy.colors(100))
print(krig, position=c(0,0,0.5,1), more = TRUE)#a)Riqueza de
  especies con Kriging
```

```
print(co.krig, position=c(0.5,0,1,1), more = FALSE)#b)Riqueza
  de especies con Co-Kriging
#####
#####Exportación de
  cobertura#####
library(rgdal) #cargar librería
###Se exporta la capa de riqueza de especie generada por Kriging
  en formato Geotiff###
rts=Predicción[“tsp”]#riqueza total de especies
writeGDAL(rts,”C:/Ejemplo
  k_ck/rtsKU.
  tif”,drivername=”GTiff”,options=NULL)
### Se exporta la capa de riqueza de especie generada por Co-
  Kriging en formato Geotiff###
rts_ck=Ck[“ck”]#riqueza total de especies Co-Kriging
writeGDAL(rts_ck,”C:/Ejemplo
  k_ck/rtsCk.
  tif”,drivername=”GTiff”,options=NULL)
```

Nota: Se recomienda consultar las siguientes fuentes de información: <http://spatial-analyst.net/>
Fortin, M. J. y M. Dala. 2008. Spatial analysis, a guide for ecologists. Cambridge University Press. 392 p.