

Aplicación de Modelos de Nicho Ecológico para estudios Epidemiológicos: *Triatoma dimidiata*, vector de la Enfermedad de Chagas en Ecuador

Yáñez-Cajo D.¹; Moreno M.¹; Chaguamate L.¹; Valencia N.² Rueda-Ayala V.³

¹Escuela Politécnica Nacional, Escuela de Formación de Tecnólogos (ESFOT), Quito, Ecuador

²Escuela Politécnica Nacional, Departamento de Ingeniería Civil y Ambiental

³Institute of Crop Science and Resource Conservation, University of Bonn

Resumen: Los modelamientos estadísticos y estocásticos son en la actualidad una herramienta importante para comprensión de fenómenos ecológicos. En este trabajo, se realizaron modelamientos de nicho de la especie *Triatoma dimidiata* vector de la enfermedad de Chagas con cuatro técnicas de modelamiento ecológico: Regresión logística, DOMAIN, Bioclim y Maxent; los cuales fueron evaluadas con el uso del test estadístico AUC. En virtud de este resultado, se determinó el modelo que permite comprender de una manera correcta la dinámica de esta especie y su distribución, la regresión logística es el modelo que presentó mayor exactitud, además se determinó que la relación de variables es importante para ejecutar cualquier tipo de modelamiento discutiendo la práctica de ejecutar modelamientos ecológicos simplemente aplicando todas las variables e insumos disponibles. El modelo de regresión logística fue utilizado para hacer una distribución de rangos con la finalidad de asignar valores categóricos y determinar la superficie con mayor idoneidad para la presencia de la especie, además de utilizar este como un ejemplo para la gestión de territorio, tomándolo como una herramienta de entomología médica. Se utilizó el sector geográfico con alta idoneidad como límite de un posible riesgo epidemiológico en el cual, se debería considerar trabajar en medidas de prevención para evitar una infección masiva de la enfermedad de Chagas.

Palabras clave: Modelo de distribución de especies, Chagas, regresión logística, variables ambientales, epidemiología.

Application of Ecological Niche Models for Epidemiologic studies: *Triatoma dimidiata*, vector of Chagas disease in Ecuador

Abstract: Nowadays, statistical and stochastic models and their application have become important tools for understanding natural phenomena. In this work, ecological niche modeling was performed for the specie *Triatoma dimidiata* vector of the Chagas disease. For this purpose, four ecological modeling techniques were used: Logistic regression, DOMAIN, Bioclim and Maxent. These models were evaluated using the statistical test AUC. Results showed that logistic regression determined the highest specificity (0.8). Categorical values were assigned through ranking, in order to locate the most suitable area for the presence of the studied species. Additionally, this model was applied for territory management, serving as a tool of medical entomology. The geographical area with higher suitability was established as a threshold at which an epidemiological risk occurs. This threshold should be considered for developing preventive measures to avoid a massive spread of Chagas.

Keywords: Species distribution model, Chagas, logistic regression, environmental variables, epidemiology

1. INTRODUCCIÓN

En Ecología, un modelo es una representación simplificada de un sistema o de una realidad compleja, que se elabora para facilitar su comprensión y el estudio de su comportamiento en situaciones difíciles de observar en la realidad. La biodiversidad es un sistema complejo, ya que los elementos que la conforman son organismos que tienen interacciones

ecológicas entre ellos difíciles de establecer con precisión (Molinier y Vignes, 1976) de ahí que la espacialización de la biodiversidad y sus aplicaciones representan un desafío en aras de plasmar la realidad de la forma más fiable. Los modelos de diversidad biológica proporcionan proyecciones simuladas de la distribución y abundancia de múltiples especies basados en un conjunto de condiciones ambientales. (Pereira et al., 2010).

danilo.yanez@epn.edu.ec

En ecología, las técnicas estadísticas y los sistemas de información geográfica (SIG) se conjugaron para conseguir interpretaciones que permitan tomar decisiones en la gestión de la biodiversidad, convirtiéndose en una herramienta para el diagnóstico y predicción de determinados incidentes. Así, los SIG se han utilizado para el estudio de los vectores de la enfermedad de Chagas, por ejemplo, para estratificar el riesgo de infección con el parásito (Ramsey et al., 2010; López-Cárdenas et al., 2005). De igual manera, los SIG se han usado para asociar variables de especies huésped-vector (Peterson et al., 2012) y para identificar las fuentes y los patrones de infestación después del control químico (Porcasi et al., 2006). Asimismo, se han utilizado mapas computarizados de distribución del vector, modelos de nicho ecológico y modelos de riesgo de transmisión de la enfermedad con el propósito de ayudar en el control de los vectores de enfermedad y realizar esfuerzos de prevención (Beard et al., 2003; Carpenter et al., 1993; Peterson et al., 2002). Si analizamos este contexto, los modelamientos de distribución de especies constituyen un elemento útil para estipular las áreas donde existe mayor probabilidad de encontrar una especie determinada, dadas las aptitudes ambientales del sector geográfico en análisis.

La tendencia creciente de la migración humana amenaza con expandir la enfermedad de Chagas de forma exponencial desde las zonas rurales a las zonas urbanas y, de regiones endémicas a no endémicas (Franco-Paredes et al., 2007); por ello, el estudio de la distribución de los insectos vectores se convierte en una tarea importante en este escenario epidemiológico.

Este estudio se realizó con el fin de analizar los modelamientos de distribución de especies referente a la relación de variables y el algoritmo utilizado por cada técnica de modelado, además de someterlos a evaluación y aplicar su utilidad para comprender la dinámica de *Triatoma dimidiata* vector más importante sobre la transmisión de Chagas en Ecuador (Abad-Franch y Aguilar, 2003).

2. METODOLOGÍA

2.1. Variables

Existen numerosas colecciones científicas en herbarios y museos que actualmente facilitan la información de presencia de especies, esto ayuda a los investigadores al acceso de ésta por medio de sistemas distribuidos (Guralnick et al., 2007) (eg. GBIF, VERTNET). Éstas son una fuente de información que debe ser depurada para obtener datos más precisos, además de estas fuentes de datos se utiliza los registros propios de muestreos o datos de instituciones que se encargan de levantar información de la biodiversidad. Se debe considerar que a mayor número de ocurrencias en diferentes sitios del espacio geográfico de análisis, el modelo presentará mayor precisión. Se estudió la especie *Triatoma dimidiata* de la familia Triatominae, vector del microorganismo *Trypanosoma cruzi*, por su relevancia en la transmisión de la enfermedad de Chagas en el Ecuador. Las ocurrencias de esta especie se adquirieron de muestreos de campo, datos de sistemas distribuidos y registros de instituciones de investigación de la biodiversidad, se realizó un depuramiento de esta información lo que nos permitió obtener un total de 69 registros.

Con estos registros se generó la base de datos con distintos atributos ambientales (coordenadas geográficas, ubicación, localidades, altitud, etc), esta información biológica fue convertida en una capa temática de puntos. Se utilizaron todos los puntos depurados de la presencia de la especie con el fin de no perder información valiosa que permite acercarnos mejor a una realidad del nicho de la especie (Papes y Gaubert, 2001).

Las variables geofísicas utilizadas fueron las generadas por el proyecto WorldClim (<http://www.worldclim.org/download>), se utilizaron 20 variables, las cuales corresponden a: temperatura, precipitación, elevación del terreno y, las derivadas de estas; las características espaciales de las mismas son: formato ráster con una resolución espacial de 30 segundos de arco = 0.93 km.

2.2. Métodos de modelado ecológico

Utilizamos la estrategia de modelizar la distribución de una sola especie (*Triatoma dimidiata*), por considerarlo más realista que los modelos a nivel de comunidad, ya que trabajos donde se agrupan varios modelos para determinar la distribución de varias especies de un mismo género presentan un sesgo al considerar una mezcla de los datos de presencias de especies con la interacción de las variables ambientales. (Guisan y Zimmermann, 2000).

Para generar el modelamiento ecológico analizamos el comportamiento de los datos, en este estudio se llevaron a cabo los siguientes procedimientos: estadísticos descriptivos, test normalidad, valores atípicos y test de independencia de variables, con la ayuda del programa estadístico SPSS v 18 (SPSS Inc., 2006).

Usando estos valores se construyeron los modelos potenciales para la distribución de la especie; Para este propósito empleamos cuatro técnicas.

Regresión logística binaria Ecuación (1) (Ortega-Calvo y Cayuela-Domínguez, 2002): esta técnica fue muy adecuada para el modelado ecológico:

$$f(z) = \frac{1}{1+e} \quad (1)$$

Su resultado está comprendido entre 0 (incompatible) y 1 (idóneo) (Hosmer y Lemeshow, 2000). La expresión matemática del modelo logístico binario en la Ecuación (2) es la siguiente:

$$f(z) = \frac{1}{1+e^{\alpha+\sum \beta_i x_i}} \quad (2)$$

El objetivo es hallar los coeficientes ($\beta_0 = \alpha, \beta_1, \dots, \beta$) que mejor se ajusten a la expresión funcional, para calcular las probabilidades de presencia se utilizó SPSS v.18 (SPSS Inc., 2006).

Domain: esta técnica utiliza la distancia de Gower como índice de similaridad ambiental entre cada píxel y el sitio más próximo en el que está presente la especie (Carpenter et al.,

1993). Para aplicar este modelo fue suficiente definir la presencia de la especie como variable.

Bioclim (Nix, 1986): esta técnica nos permitió definir una regla para identificar áreas con clima similar a los lugares de la presencia de la especie. Para este algoritmo utilizamos las variables ambientales descritas anteriormente para determinar la distribución potencial de la especie, pero se cabe destacar que este algoritmo pueden usar variables no climáticas como: tipo de suelo y cobertura vegetal natural. La distribución potencial se identifica interpolando las variables climáticas, lo que da como resultado gráfico un mapa binario; Se utilizó el software Diva-GIS v. 5.2 (Hijmans et al., 2005).

Maxent (Phillips et al., 2006), es un programa de modelado de distribución de especies para datos de registro solo de presencia de especies (Elith et al., 2011), basado en el principio de máxima entropía. Las ventajas de MaxEnt sobre otro software son las siguientes: 1. Solo requiere datos de presencia junto con la información ambiental para el área de estudio. 2. Puede utilizar datos tanto continuos como categóricos y puede incorporar iteraciones entre diferentes variables. 3. La distribución de probabilidad de MaxEnt tiene una definición matemática concisa y por tanto susceptible de análisis. 4. El sobre-aprendizaje puede ser evitado con el uso de L1-regulaciones. 5. Los resultados son continuos, esto permite obtener sutiles distribuciones a diferencia de las diferentes áreas modeladas en modelos binarios. 6. También se puede utilizar datos de presencia/ausencia de especies, esto modificando a un modelo condicional. 7. Es un enfoque generativo, en lugar de discriminativo, esto puede ser una ventaja cuando la cantidad de datos de entrenamiento es limitada.

$$P(y = 1|z) = \frac{f_1(z)P(y=1)}{f(z)} \quad (3)$$

Que se centra en comparación de densidades de probabilidad en el espacio de covarianza, es decir, se trata de reducir al mínimo la distancia de $f_1(z)$ a $f(z)$ Ecuación (3). Esta distancia se denomina entropía relativa. Así, el objetivo sería minimizar la ganancia de información (entropía relativa) Ecuación (4) sujeta a restricciones de límite de error,

$$\max_{\alpha, \beta} \frac{1}{m} \sum_{i=1}^m \ln(f(z_i) e^{g(z_i)}) - \sum_{j=1}^n \lambda_j \quad (4)$$

Sujeta a Ecuación (5):

$$\int f(z) e_{g(z)} \partial z = 1 \quad (5)$$

2.3. Evaluación de modelos

Para comparar la capacidad predictiva de los modelos y de cada uno de los métodos aplicados, se usó el test estadístico AUC (Area Under the Curve), a partir de la curva ROC (Receiver Operating Characteristics) (Hanley y McNeil, 1982). Esta prueba produce un valor dentro de los rangos de 0.5 y 1. Un valor de 0,5 equivale a una clasificación al azar,

mientras que un valor de 1 indica un ajuste perfecto (Fielding y Bell, 1997). AUC nos permitió evaluar cómo influyó las ocurrencias de *Triatoma dimidiata* en el entrenamiento del modelo.

Para editar o tratar los modelos y utilizarlos como una herramienta para la gestión del territorio, se necesitó introducir los archivos ráster generados en los software Diva-GIS (Hijmans, 2005) y Maxent (Phillips et al., 2004) a un software de procesamiento geográfico, para este trabajo Arcgis 9.3 (ESRI, 2008), estos archivos están denominados con la extensión “.asc”.

Además se realizó una síntesis cartográfica con el archivo de geoinformación temática de “poblados”, el cual dispone de información sobre los asentamientos humanos, esto permitió determinar los poblados que estarían dentro del área de influencia epidemiológica.

3. RESULTADOS Y DISCUSIÓN

Al aplicar las técnicas de modelado ecológico descritas, obtuvimos como resultado que el modelo de mayor evaluación fue el de de regresión logística (0.8) (Fig 2); seguido por el modelo de Máxima entropía (0.7); Domain (0.5) y Bioclim (0.2). Estos resultados permiten determinar que el modelo que expresa una mayor probabilidad de ubicar a la especie según su ecología es regresión logística, siendo el más idóneo para trabajar en propuestas de prevención de la enfermedad causada por *Triatoma dimidiata*.

3.1 Idoneidad

Al ser regresión logística el modelo más efectivo para predecir el nicho de la especie, se editó este para determinar el área idónea para la distribución de la misma. Los rangos numéricos de este archivo ráster fueron utilizados para el uso de los quintiles, esto nos permitió obtener cinco rangos categóricos: 1:0-16 Muy baja; 2:17-35 Baja; 3:36-51 Media; 4:52-67 Alta; 5:68-100 Muy alta Figura 3.

Dentro de la superficie de muy alta idoneidad se determinaron los centros poblados más amenazados por un posible riesgo epidemiológico, estos poblados se encuentran en su mayoría en la región Amazónica del Ecuador en las provincias de Sucumbios, Napo y Pastaza; pero en su mayor riesgo en la parte oriental de la sub-cordillera occidental de la cordillera de los Andes.

4. CONCLUSIÓN

La especie *Triatoma dimidiata* se distribuye en zonas tropicales de elevadas temperaturas (World Health Organization, 2010), especialmente en zonas de la región amazónica, pero cabe destacar que existe un sector de presencia en la región noroccidente de la provincia de Esmeraldas. Por esta razón, los modelos utilizados se analizaron tomando en consideración este patrón ecológico. Los cuatro modelos determinaron superficies de idoneidad dentro de la región amazónica Figura 1 pero no todos dentro de la Provincia de Esmeraldas.

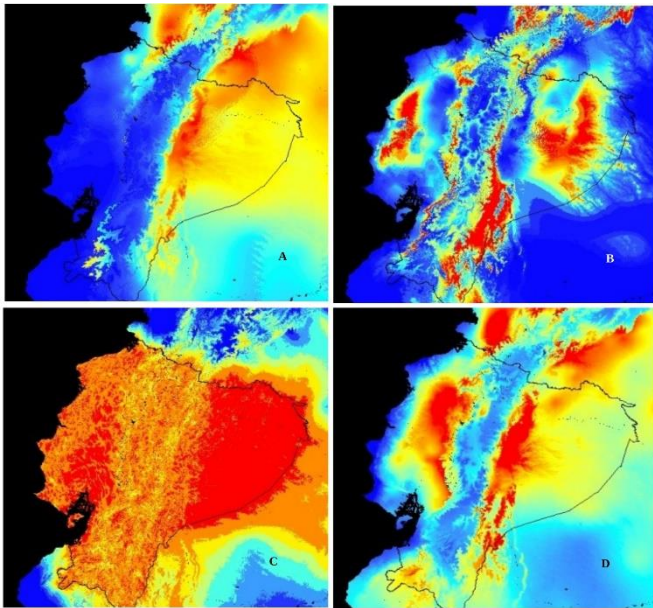


Figura 1. Modelos de distribución de *Triatoma dimidiata* obtenidos mediante el uso de cuatro algoritmos A) Regresión logística; B) Domain; C) Bioclimático; D) Maxent.

Los modelos ecológicos deben generarse a partir de las variables ambientales apropiadas (Ottaviani, 2004). Es probable que un modelo de distribución pueda presentar algún tipo de sesgo a causa de no elegir las variables significativas para el tema de estudio o por el desconocimiento de la biología de la especie.

La técnica de regresión logística demostró ser la que más se apega a la realidad para predecir la distribución de la especie. Esto fue debido a que se la utilizó para estudiar el efecto de múltiples variables explicatorias sobre una variable respuesta categórica, y, de más de dos categorías en escala ordinal. En este último caso cuando la variable independiente puede tomar más de dos posibles valores podemos codificarlas usando variables indicadoras “variable indicatriz”.

El uso de esta regresión la planteamos como alternativa al no poder aplicar un modelo de “regresión lineal” por no cumplirse el supuesto de continuidad y de distribución normal de la variable dependiente, la clave de la popularidad para la regresión logística se basa en la función logística.

La asignación categórica permite determinar las superficies más idóneas para la distribución de la especie en el modelo que se lo aplique, siempre y cuando esta sea continua. En el caso de la regresión logística, el rango de 68 a 100 representado con color rojo indica que existe una muy alta idoneidad para la presencia de la especie. En función de la perspectiva del investigador, los objetivos del proyecto o de la biología de la especie, estos rangos pueden variar por tanto, hay que asignarlos con el criterio más adecuado posible.

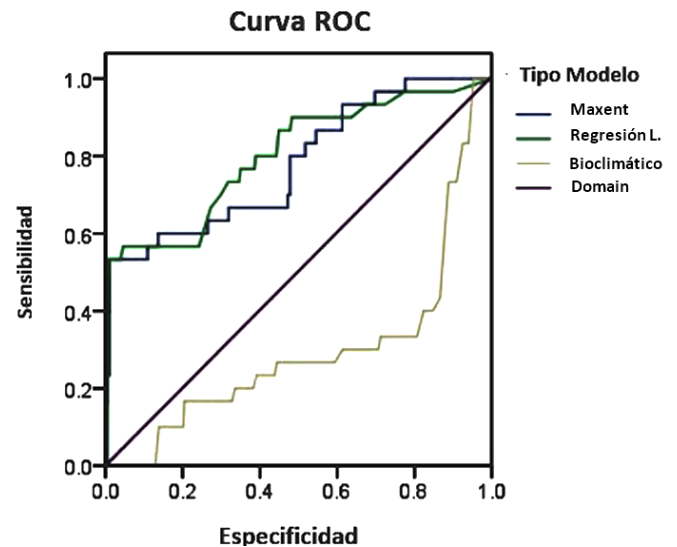


Figura 2. Curva ROC, expresando los valores de la evaluación de cada una de las técnicas de modelamiento ecológico.

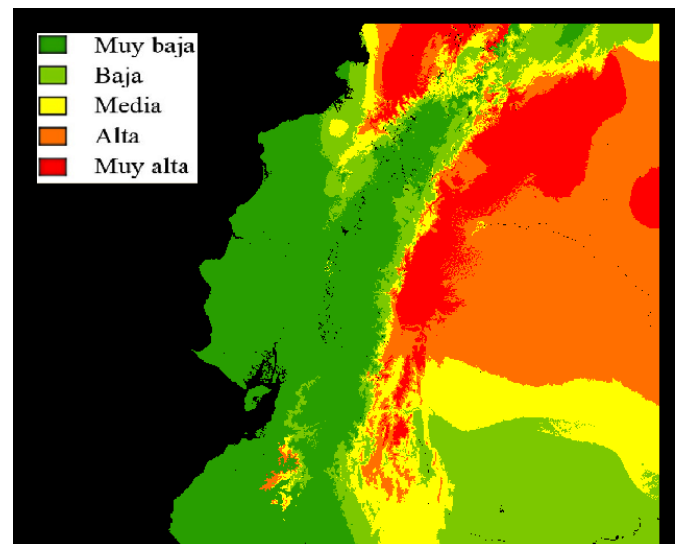


Figura 3. Rangos categóricos a partir de quintiles determinando la zona de idoneidad de la distribución de la especie; aplicado al modelo de regresión logística (rojo zona de muy alta idoneidad).

4.1 Aplicación

Para aplicar este modelo como herramienta en la gestión de territorio, se utilizó los resultados categóricos de muy alta idoneidad de la distribución, esto permite determinar las zonas de mayor vulnerabilidad a enfermedades; para facilidad de uso de estos resultados es necesario transformar el insumo ráster a vector, y, de la misma manera trabajar en el sistema de proyección geográfica más útil para el investigador, como ejemplo de lo mencionado, se puede proyectar de datos geográficos a datos UTM, lo que permite calcular la superficie de influencia.

Con estas herramientas los planificadores del área de salud pueden ejecutar medidas de prevención y evitar contagios en poblaciones amenazadas por la enfermedad de Chagas. En este contexto, los modelos ecológicos se han convertido en una

herramienta muy útil al momento de realizar planificación sobre estudios biológicos, ya que optimizan el tiempo y la disposición de recursos. Como fin de este estudio, se evidenció la utilidad del modelo para ser aplicado en la planificación del territorio en temas de salubridad ambiental y en gestión de la salud pública.

AGRADECIMIENTO

A Luis Escobar Ph.D, profesor de la Universidad de Minnesota, por su aporte en la realización de este trabajo; a la Escuela de formación de tecnólogos (ESFOT) a la Ing. Mariana Veintimilla M.Sc., a la Ing. Mónica Vinuesa M.Sc. y, al Ing. Carlos Romo. M.Sc. Al Departamento de Ingeniería Civil y Ambiental (DICA) de la Escuela Politécnica Nacional (EPN).

REFERENCIAS

- Abad-Franch, F., Aguilar, M. (2003). "Control de la enfermedad de chagas en el Ecuador". Ministerio de salud Pública del Ecuador. Quito. Ecuador.
- Beard, C.B., Pye, F.J., Steurer, R., Rodríguez, R., Campman, A., Peterson, J.T., Ramsey, R.A., Wirtz, L.E., Robinson, E. (2003). "Chagas Disease in a domestic transmission cycle in Southern Texas, USA". *Emerg. Infect. Dis.*, 9: 103-105.
- Carpenter, S., Caraco, R., Correll, R., Howarth, R., Sharpley, R., Smith, H. (1993). Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological Applications* 8: 559-568.
- Costa, J., Peterson, A.T., Beard, C.B. (2002). Ecologic niche modeling and differentiation of populations of *Triatoma brasiliensis*. *Am J Trop. Med Hyg.* 67: 516-520.
- ESRI. (2008). *ArcGIS a Complete Integrated System. Environmental Systems Research Institute.* Inc. Redlands, California.
- Fielding, A., Bell, H. (1997). Review of methods for the assessment of prediction errors in conservation presence/ absence models. *Environmental Conservation*, 24:38-49.
- Franco-Paredes, C., Von, A., Hidron, A.J., Rodríguez-Morales, A., Tellez, I., Barragán, M., Jones, D.C., Náquira, G., Mendez J. (2007). *Chagas disease: an impediment in achieving the Millennium Development Goals in Latin America.* BMC Int. Health Hum. Rights.
- Guisan, A.E., Zimmermann, N.E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147-186.
- Ortega-Calvo, M., Cayuela-Domínguez, A. (2002). Regresión logística no condicionada y tamaño de muestra: una revisión bibliográfica. *Rev. Esp. Salud Pública.* 76: 85-93.
- Guralnick, P., Hill, A., Lane, M. (2007). Towards a collaborative, global infrastructure for biodiversity assessment. *Ecology Letters*, 10: 663-672.
- Hanley, H., McNeil, N. (1982). Calculation of the Standard Error of the Area under the Curve (AUC) and of the difference between two AUCs. *Radiology*, 143(1): 29-36.
- Hijmans, R., Luigi Guarino, J., Jarvis, A., Brien, R., Mathur, P., Bussink, C., Cruz, M., Barrantes, I., Rojas, E. (2005). *DIVA-GIS Version 5.2 Manual.* LizardTech, Inc., and the University of California.
- Hosmer, D.W., Lemeshow, S. (2000). *Applied Logistic Regression.* (2nd ed). John Wiley and Sons.
- Elith, J., Steven, J., Phillips, T.H., Miroslav, D., Yung, C., Colin-Yates, J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions.* 17: 43-57
- López-Cárdenas, J., González-Bravo, F.E., Salazar-Schettino, P.M., Gallaga-Solórzano, J.C., Ramírez-Barba, E., Martínez-Méndez, J., Sánchez-Cordero, V., Peterson, T., Ramsey, J.M. (2005). Fine-scale predictions of distribution of Chagas disease vectors in the state of Guanajuato, Mexico. *J. Med. Entomol.* 42: 1068-1081.
- Molinier, R., Vignes, R. (1976). *Introducción a la ecobiocenología: los biotopos, las biocenosis, los ecosistemas.* Colección Tropos. Barcelona. Vicens-Vives.
- Nix, H.A. (1986). BIOCLIM. A Bioclimatic Analysis and Prediction System. Research report, *CSIRO Division of Water and Land Resources*, 1983-1985.
- Ottaviani, E., Malagoli, D., Franchini, A. (2004). Invertebrate humoral factors: cytokines as mediators of cell survival. *Prog. Mol. Subcell. Biol.* 34:1-25.
- Papes, M., Gaubert, C. (2001). Modelling ecological niches from low numbers of occurrences: assessment of the conservation status of poorly known viverrids (Mammalia, Carnivora) across two continents. *Diversity and distributions*, 12(2): 125-136.
- Pereira, H.M., Leadley, P.W., Proenca, V., Alkemade, R., Scharlemann, P., Fernandez-Manjarres, J.F. (2010). Scenarios for global biodiversity in the 21st century. *Science.* 330: 1496-1501.
- Peterson, A.T., Sanchez-Cordero, V., Beard, C.B. (2002). Ramsey. Ecologic niche modeling and potential reservoirs for Chagas disease, Mexico. *Emerg Infect Dis.* 8:662-667.
- Phillips, S.J., Anderson, R.P., Schapire, R.E. (2006). Maximum entropy modeling of species, geographic distribution?. *Ecological Modelling*, 190: 231-259.
- Phillips, S.J., Dudík, M., Schapire, R.E. (2004). *A maximum entropy approach to species distribution modeling. Proceedings of the Twenty-First International Conference on Machine Learning.* 655-662.
- Porcasi, X., Catalá, S.S., Hrellac, H., Scavuzzo, M.C., Gorla, D.E. (2006). Infestation of rural houses by *Triatoma infestans* (Hemiptera: Reduviidae) in southern area of Gran Chaco in Argentina. *J. Med. Entomol.* 43: 1060-1067.
- Ramsey, J.M., Ordóñez, R., Cruz-Celis, A., Alvear, A.L., Chávez, V., López, R. (2000). Distribution of domestic Triatominae and stratification of Chagas disease transmission in Oaxaca, Mexico. *Med. Vet. Entomol.* 14: 1-12.
- SPSS Inc. Released. 2006. *PASW Statistics for Windows, Version 18.0.* Chicago: SPSS Inc.
- World Health Organization. (2010). *Chagas Disease: American Trypanosomiasis.* Fact sheet.