

MODELIZACIÓN DEL HÁBITAT POTENCIAL DE FORMACIONES FORESTALES EN LA PROVINCIA DE HUELVA

M. ANAYA-ROMERO ⁽¹⁾, R. PINO ⁽²⁾, A. JORDÁN ⁽¹⁾, L. MARTÍNEZ-ZAVALA ⁽¹⁾, N. BELLINFANTE ⁽¹⁾

⁽¹⁾Departamento de Cristalografía, Mineralogía y Química Agrícola (Universidad de Sevilla). Autor para correspondencia, ajordan@us.es

⁽²⁾Departamento de Estadística e Investigación Operativa (Universidad de Sevilla)

Abstract. Three methods for modelling the potential distribution of forest formations in protected areas in southern Spain are discussed in this paper: logistic multiple regression, artificial neural net and decision tree. For every model, the potential habitat of every forest type has been designed using environmental information (soil, lithology, geomorphology and climate). The results were compared with the current area corresponding to every forest type, to determine the relative accuracy of every method. The multiple logistic regression seems to offer the best results.

Key words: Artificial neural net; decision tree; logistic regression; relationship vegetation - environment; Mediterranean forest.

Resumen. En este trabajo se discuten tres métodos de modelización de la distribución potencial de formaciones forestales utilizando datos medioambientales de áreas protegidas del sur de España: regresión logística múltiple, red neuronal artificial y árbol de decisión.

Para cada modelo, el hábitat potencial de cada tipo de vegetación forestal se ha diseñado utilizando información medioambiental (suelo, litología, geomorfología y clima). Los resultados obtenidos se compararon con el área actual de cada tipo de formación forestal, para determinar la precisión relativa de cada uno de los métodos. La regresión logística múltiple es el método que ha ofrecido mejores resultados.

Palabras clave: Red neuronal artificial; árbol de decisión; regresión logística; relaciones vegetación-medioambiente; bosque mediterráneo.

INTRODUCCIÓN

En general, la evaluación de los suelos está basada en modelos conceptuales de las relaciones entre las características del medio físico y su productividad. Por ello, a menudo, los modelos son demasiado simples para considerar relaciones complejas entre los distintos componentes del sistema.

Según Franklin (1995) y Guisan y Zimmermann (2000), las relaciones existentes entre

las diversas comunidades vegetales y el medio físico condicionan la distribución potencial de las especies. Tradicionalmente, el estudio de la vegetación potencial ha utilizado métodos intuitivos, basados en la experiencia previa y en el estudio de las series de vegetación. Un ejemplo de este tipo de análisis es el mapa de series de vegetación de España realizado por Rivas-Martínez (1987). Sin embargo, más recientemente, y aprovechando el desarrollo de nuevas herramientas informáticas, la modelización de

los procesos naturales y el análisis estadístico de la información espacial se ha hecho más eficiente.

La modelización de la vegetación y la distribución espacial de las especies se apoya frecuentemente en la utilización de la teledetección, los modelos digitales del terreno y los mapas de suelo (Neave y Norton, 1998; Vogelmann *et al.*, 1998; Münier *et al.*, 2001; Bellinfante *et al.*, 2003).

El principal objetivo de este trabajo es la comparación entre diferentes modelos de predicción del hábitat potencial (red neuronal artificial, árbol de decisión y regresión logística múltiple) para cuatro tipos de formaciones fo-

restales en el área de estudio: bosque de encina y alcornoque, pinar, eucaliptar y otras frondosas.

ÁREA DE ESTUDIO

El área de estudio incluye el Parque Natural de la Sierra de Aracena, así como parte del área natural del Andévalo occidental (Figura 1; aproximadamente 4770 km²). La elevación media en la sierra apenas supera los 1000 m, mientras que el Andévalo muestra un relieve muy suave. El principal tipo de uso de la Sierra de Aracena es el aprovechamiento forestal, y en el Andévalo predomina una agricultura de



FIGURA 1. Área de estudio.

baja producción. El clima es mediterráneo, con una gran diferencia estacional, caracterizado por inviernos fríos y lluviosos y por veranos cálidos secos.

El bosque de alcornoque (*Q. suber*) y la encina (*Q. rotundifolia*) constituyen la vegetación climática. También existen extensas áreas reforestadas, normalmente con eucaliptos y pinos, en el centro y sur del área de estudio. Ocasionalmente, se encuentran otras especies, ya sea cultivadas o formando parte del bosque de ribera.

MÉTODOS

Vegetación actual

Los datos sobre cada tipo de formación forestal se extrajeron del mapa de usos y coberturas vegetales del suelo en Andalucía (Consejería de Medio Ambiente de la Junta de Andalucía, datos no publicados). A causa del elevado número de clases de la leyenda original, se seleccionaron cuatro grupos principales para este trabajo, de acuerdo con los grupos taxonómicos dominantes: *Quercus*, *Pinus*, *Eucalyptus* y otras frondosas (en adelante, frondosas).

Variables predictivas

Las variables se extrajeron de diferentes fuentes publicadas previamente o elaboradas por los autores. La información litológica se obtuvo a partir de los mapas geológicos publicados originalmente a escala 1:50.000. Los mapas se procesaron utilizando el software Arc/Info (ESRI, 1982-1997). Las variables litológicas seleccionadas fueron la naturaleza geológica (rocas volcánicas, ígneas intrusivas, metamórficas o sedimentarias), la acidez y la consolidación de la roca (roca dura o sedimento no cohesivo).

Los datos de suelo se obtuvieron de los mapas de unidades geomorfoedáficas del P. N. Sierra de Aracena (Recio y Núñez, 1997) y del Andévalo occidental (Martínez-Zavala, 2001).

Las variables edáficas seleccionadas fueron la acidez del suelo (pH), los elementos químicos asimilables (Fe, Mn, Cu, Zn, Mg, K y P), la materia orgánica, la capacidad de intercambio catiónico (CIC), la saturación del complejo de cambio (S), la proporción de elementos gruesos y la fracción de arcilla en la tierra fina.

Las variables geomorfológicas, fisiográficas y topográficas se extrajeron también de los mapas realizados por Recio y Núñez (1997) y Martínez-Zavala (2001). Estas variables fueron divididas en varias clases cualitativas. Como variables de tipo geomorfológico se consideraron los tipos de proceso erosivo (erosión laminar, en regueros y cárcavas, erosión por gota de lluvia o erosión vertical por cauces de agua), tipo de deslizamiento, sedimentación (por gravedad o por inundación), morfogénesis (fluvial, denudativa, endógena o kárstica) y clases fisiográficas, tal como aparecen definidas por De la Rosa y Moreira, (1987): lecho fluvial, llanura estable, superficie de erosión, meseta, loma, colina, cerro y relieve montañoso.

Las variables topográficas se obtuvieron a partir de un modelo digital del terreno (MDT) de resolución $20 \times 20 \text{ m}^2$. A partir de éste, se construyeron otros cuatro modelos derivados: elevación, pendiente, curvatura de la superficie y orientación de las laderas. La orientación se calculó como exposición N-S (seno del ángulo de orientación respecto al norte) y exposición E-W (coseno).

Los datos climáticos fueron procesados por Anaya-Romero (2003), usando una interpolación estadística de puntos de la red de estaciones meteorológicas. Las variables consideradas fueron la radiación solar, la precipitación media anual, la precipitación media en verano, la temperatura anual media, la temperatura media del mes más cálido y la temperatura media del mes más frío.

Preparación de los datos

La distribución actual de cada tipo de formación forestal, en formato digital, se utilizó para construir un modelo digital con una reso-

lución de 20 x 20 m². El modelo correspondiente a cada formación forestal posee una información binaria, de modo que el valor de cada celda es 1 (presencia) ó 0 (ausencia). Las variables predictivas también se procesaron de la misma manera. Cuando fue necesario, los valores decimales se corrigieron utilizando un factor de conversión apropiado (x 10, x 10², x 10³, etc.).

Debido al gran número de datos, se llevó a cabo un muestreo sistemático del área de estudio. Se seleccionaron 65.535 puntos uniformemente distribuidos (14 puntos por km²).

Para reducir la dimensionalidad de los datos edáficos, se realizó un análisis de componentes principales (ACP) sobre las 14 variables disponibles. La distribución espacial de los datos ausentes o erróneos se verificó visualmente. Su distribución era regular y sin tendencia, por lo que se eliminaron del análisis.

Métodos de Clasificación

Regresión logística. Dada una respuesta binaria 0-1 y p variables predictoras cuantitativas x_1, \dots, x_p , (algunas de las cuales pueden ser variables ficticias 0-1 empleadas para codificar variables cualitativas), bajo el modelo de regresión logística (RL) la probabilidad de que se observe la respuesta objetivo (codificada como 1) es:

$$\pi(x_1, \dots, x_p) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}$$

La función *glm* de R (Venables and Ripley, 1999) intenta calcular el estimador de máxima verosimilitud de los $p+1$ parámetros por un método iterado de mínimos cuadrados (IWLS). Existen diversos procedimientos inferenciales para contrastar la significación estadística del modelo completo y la significación individual de cada variable. Además, este modelo es fácilmente interpretable y se dispone de una gran variedad de criterios para identificar observaciones atípicas y observaciones influencia. Asimismo existen varios algoritmos

propuestos para realizar un proceso de selección secuencial de variables. En nuestro contexto, la aplicación de un modelo RL sobre un modelo digital del terreno (MDT), proporciona una medida de adecuación para cada uno de los usos forestales estudiados. El resultado es un nuevo MDT en el que cada para celda de la parrilla se obtiene una estimación de la probabilidad de cada uso. Si se desea obtener un modelo automático se necesita determinar un valor umbral, de manera que si la probabilidad estimada es superior a dicho umbral, el punto es clasificado en la clase codificada como 1.

Árboles de Decisión. Un Árbol de Decisión (AD) se construye mediante un particionamiento binario recursivo a partir de la variable respuesta y construyendo las divisiones mediante las variables predictoras. En cada etapa se elige la división que maximiza la reducción de la impureza (una medida de diversidad para la variable respuesta calculada sobre un conjunto específico de nodos). La división continúa hasta que los nodos terminales contienen un número de observaciones demasiado pequeño como para continuar el proceso de división. Para clasificar un punto se identifica primero el nodo terminal que le corresponde y se determina la clase más frecuente entre los casos que corresponden a dicho nodo terminal. En nuestro estudio, para cada posible tipo forestal los dos posibles valores de la variable respuesta son Sí (presencia de dicho tipo) y No (ausencia). Hemos utilizado el paquete *rpart* de R (Venables and Ripley, 1999), que implementa la metodología CART (Classification and Regression Trees) que fue propuesta por Breiman *et al.* (1984). Con esta implementación, el usuario debe ajustar un parámetro fundamental, el número de nodos terminales, también llamado tamaño del árbol.

Perceptrón Multinivel. Una Red Neuronal Artificial (RNA) proporciona una gran variedad de modelos matemáticos no lineales,

útiles para afrontar problemas estadísticos muy diversos. Existen varios resultados teóricos que proporcionan un gran respaldo a una arquitectura particular, como el Perceptrón Multinivel (PM), como puede verse en Bishop (1995). De acuerdo con este resultado, hemos considerado un perceptrón de tres capas (una de entrada, una oculta y una capa de salida) con la función de activación logística $g(w) = \frac{e^w}{e^w + 1}$ en la capa oculta y la función de activación identidad en la capa de salida. Sean H el tamaño de la capa oculta, $\{v_{ih}, i=0,1,2,\dots,p, h=1,2,\dots,H\}$ los coeficientes sinápticos para las conexiones entre la capa de entrada, de tamaño p , y la capa oculta $\{w_{hj}, h=0,1,2,\dots,H, j=1,2,\dots,q\}$ los coeficientes sinápticos para las conexiones entre la capa oculta y la capa de salida, de tamaño q . Las salidas de la red son

La salida de cada nodo es reescalada me-

$$o_j = w_{0j} + \sum_{h=1}^H w_{hj} g(v_{0h} + \sum_{i=1}^p v_{ih} x_i), \quad j = 1, 2, \dots, q$$

dante los coeficientes sinápticos y se propaga como entrada a los nodos de la siguiente capa de la red (Gardner y Dorling, 1998).

En nuestro problema de clasificación la respuesta fue codificada mediante un vector $z=(z_1, z_2)$ formado por dos variables 0-1, una para cada clase, por lo que el número de salidas fue $q=2$. Dada una entrada a la red, su clasificación es la clase correspondiente al máximo de las dos predicciones de las variables auxiliares. En general, para q clases, la siguiente expresión proporciona una estimación de la probabilidad de la clase $j, j=1,2,\dots,q$:

La función *net* de R (Venables and Ri-

$$\hat{P}[j] = \frac{e^{o_j}}{\sum_{r=1}^q e^{o_r}}$$

pley, 1999) ajusta modelos PM mediante el procedimiento BFGS, un algoritmo cuasi-Newton publicado en 1970 por Broyden, Fletcher,

Goldfarb y Shanno, que intenta minimizar un criterio de mínimos cuadrados que permite incluir un término de penalización λ diseñado para intentar evitar problemas de sobreajuste, que surgen cuando la red se especializa en exceso en los datos disponibles, perdiendo la capacidad de generalización ante nuevos datos. Si W denota el vector de todos los M coeficientes sinápticos de la red, el método BFGS se aplica al siguiente problema no lineal de mínimos cuadrados:

La principal desventaja de los modelos PM es el hecho de que no se tiene asegurada la

$$\text{Min}_W \sum_{i=1}^n \|z_i - \hat{z}_i\|^2 + \lambda \left(\sum_{i=1}^M W_i^2 \right)$$

convergencia a soluciones globales mediante ningún algoritmo de entrenamiento.

Selección de Parámetros

Ninguno de los tres métodos puede ser considerado automático, por lo que se requiere un proceso cuidadoso para seleccionar los parámetros. En nuestra metodología de trabajo se ha realizado para cada parámetro una búsqueda en un espacio o parrilla de posibles valores. Para que la selección fuera más segura, se ha seguido el siguiente proceso de validación cruzada:

1. Dividir el conjunto de datos en validación (el primer 10% de casos) y entrenamiento (el restante 90%).
2. Ajustar el modelo sobre el conjunto de entrenamiento.
3. Calcular la proporción de errores sobre el conjunto de validación.
4. Repetir los pasos 1, 2, y 3 para los pares de conjuntos de validación segundo a décimo.
5. Calcular la media de las 10 proporcio-

nes de errores de validación.

Para cada modelo de clasificación, el conjunto de parámetros seleccionado se determinó con el objetivo de minimizar la medida de error calculada en el paso 5.

RESULTADOS

Reducción de variables edáficas

Para reducir el número de datos edáficos, tras el ACP se seleccionaron los cuatro componentes con mayor autovalor (AV), que explican una varianza acumulada del 65 % (tabla 1). Las variables que mostraron una mayor correlación con estos ejes se incluyeron en el análisis posterior (tabla 2).

El componente 1 (AV=3.7, VE=29%) se relacionó de manera significativa con el contenido en micronutrientes (como Cu y Fe, $r=0.95$ y 0.83 , respectivamente). El componente 2

(AV=2.0, VE=16%) se relacionó con el contenido en gravas ($r=0.84$) y la saturación del complejo de cambio (0.71). El componente 3 (AV=1.6, VE=12%) se relacionó con el pH y el contenido en potasio, magnesio y manganeso. Finalmente, el componente 4 (AV=1.17, VE=9%) se relacionó con el contenido en arcilla, la CIC y el contenido en materia orgánica.

Selección del modelo de predicción

La predicción obtenida con cada método predictivo se comparó con la distribución actual utilizando la matriz de confusión (Kleinbaum, 1994). Para ello, se calculó un índice de error (IE), que consiste en la suma de los errores de predicción dividida entre el número total de puntos analizados tanto para datos de entrenamiento como de validación (tabla 3).

TABLA 1. Autovalor (AV) de los componentes principales y porcentaje de varianza explicada (VE).

Eje	AV	VE (%)	AV-acumulado	VE-acumulada
1	3.70	28.46	3.70	28.46
2	2.03	15.58	5.73	44.03
3	1.55	11.92	7.27	55.95
4	1.17	8.97	8.44	64.92

TABLA 2. Correlación entre las variables edáficas incluidas en el análisis de componentes principales y los factores determinados. Sólo se muestran las correlaciones significativas y superiores a 0.6.

Variable	Factor 1	Factor 2	Factor 3	Factor 4
CIC	-	-	-	0.6709
Cobre	0.9488	-	-	-
Hierro	0.8351	-	-	-
Grava	-	0.8358	-	-
Arcilla	-	-	-	0.6000
Potasio	-	-	0.7476	-
Materia orgánica	-	-	-	0.6712
Magnesio	-	-	0.6083	-
Manganeso	-	-	0.6236	-
Fósforo	-	-	-	-
pH	-	-	0.6972	-
Sat. del complejo de cambio	-	0.7097	-	-
Zinc	-	-	-	-

TABLA 3. Cálculo del índice de error para los datos de entrenamiento (E) y de validación (V). RLM: regresión logística múltiple; RNA: red neuronal artificial; AD: árbol de decisión.

Método	Encinar/ alcornocal		Pinar		Fronzosas		Eucaliptos		Promedio	
	E	V	E	V	E	V	E	V	E	V
RLM	0.272	0.255	0.056	0.061	0.011	0.013	0.173	0.176	0.128	0.119
RNA	0.252	0.556	0.050	0.449	0.007	0.292	0.144	0.766	0.113	0.516
AD	0.136	0.247	0.039	0.076	0.005	0.009	0.089	0.173	0.067	0.126

DISCUSIÓN

La interpretación de los resultados debe ser diferente para cada método utilizado. Por supuesto, cada método posee ventajas e inconvenientes. El modelo RNA es útil para la predicción, especialmente cuando no se puede construir un modelo teórico, o cuando se trabaja con sistemas no lineales (Gardner y Dorling, 1998). Quizás, uno de los principales problemas para los científicos sea decidir cuál es la arquitectura de la red, ya que no hay reglas predeterminadas (Gardner y Dorling, 1998). Las RNAs y los ADs son capaces de clasificar datos en más de dos categorías. Por otro lado, los resultados que muestra la RLM deben ser interpretados como probabilidad de presencia. La RLM es el método usado más comúnmente en el análisis de datos ecológicos, ya que es muy fácil de aplicar y posee una gran capacidad de explicación de los resultados (Gevrey *et al.*, 2003; Robertson *et al.*, 2003). Una gran diferencia entre los modelos de RLM y RNA/AD es la incapacidad de estudiar relaciones no lineales que tiene el primero.

En general, la RLM y el AD ofrecen un IE similar y menor que el obtenido por la RNA en el caso de los datos de validación. La respuesta no es tan clara para los datos de entrenamiento. Sin embargo, cuando se compara el promedio de IE para los tres métodos ocurre que de nuevo RLM y AD son más precisos que la RNA.

La RLM posee el mínimo IE cuando se utilizan los datos de validación. Por lo tanto, puede concluirse que, en este caso, la RLM ofrece la mayor fiabilidad cuando se usan nuevos datos en el modelo.

No obstante, esta conclusión no puede ser tomada como una regla general. De hecho, existen algunas limitaciones. En primer lugar, los resultados pueden depender de la calidad de los datos de entrada. El primer requerimiento para la mejora de la predicción del modelo es una elevada precisión y resolución de los

mapas biofísicos de entrada (Guisan y Zimmermann, 2000), especialmente cuando se utilizan variables como la litología, las unidades de suelo o la cobertura vegetal (Fischer, 1994). La competencia ecológica entre diferentes especies no se ha tenido en cuenta durante este trabajo, y probablemente posee una gran relevancia en la distribución real de las comunidades vegetales. Factores históricos, como la actividad antrópica sobre el sistema original tampoco se han tenido en cuenta. Esto es, quizás, un problema sin solución, excepto para sistemas naturales no intervenidos. De este modo, y de acuerdo con otros autores (Guisan y Zimmermann, 2000; Cairns, 2001) la variabilidad en la capacidad y exactitud obtenida por los tres métodos utilizados en este trabajo sugiere que el mejor modelo predictivo depende en gran manera del tipo y calidad de los datos empleados.

CONCLUSIONES

Se han evaluado tres modelos distintos para elegir aquél que mejor se adapta a los datos de entrada y se han comparado mediante el cálculo del IE de cada uno de ellos.

- El uso de cartografía temática básica publicada o elaborada por el equipo de trabajo permite la elaboración de modelos de evaluación de aptitud del suelo.

- El modelo predictivo más preciso, teniendo en cuenta el tipo y calidad de los datos utilizados en el proceso de evaluación, ha sido la regresión logística múltiple.

- La distribución actual de las distintas formaciones no tiene por que adecuarse a los modelos de distribución potencial, ya que la actuación de factores de tipo antrópico ha podido afectar de manera artificial a las condiciones ecológicas o a las especies.

- Todo modelo de predicción está basado sólo en procesos conocidos. Por lo tanto, no es posible modelizar la distribución de formaciones no contempladas en el modelo utili-

zado.

AGRADECIMIENTOS

Esta investigación forma parte de los resultados del proyecto de investigación Propuesta metodológica para la realización de un modelo de distribución potencial de usos forestales basado en parámetros edáficos, geomorfológicos, climáticos y topográficos (OG-096/01), financiado mediante un convenio entre el Departamento de Cristalografía, Mineralogía y Química Agrícola (Universidad de Sevilla) y la Consejería de Medio Ambiente de la Junta de Andalucía.

Los autores desean agradecer a A. Cerdá (Universidad de Valencia) sus aportaciones para la mejora del trabajo.

REFERENCIAS

- Anaya-Romero, M. (2003): Modelo de distribución potencial de usos forestales basado en parámetros edáficos, geomorfológicos, climáticos y topográficos. Tesis Doctoral. Universidad de Sevilla. Sevilla, España.
- Auer, P., Holte, R.C., y Maass, W. (1996): Theory and applications of agnostic PAC-learning with small decision trees. Pp. 21-29. En: *Machine Learning: Proceedings of the Twelfth International Conference*. Prieditis, A., y Russell, S. (eds.), *Morgan Kaufmann Publishers Inc.* San Francisco, EEUU.
- Bellinfante, N., Martínez-Zavala, L., y Panque, G. (2000): Cartografía de unidades geomorfoedáficas de la comarca del Anévalo (NW de Huelva). Relaciones suelo/geomorfología/uso. *Edafología* 7-3:287-299.
- Bellinfante, N., Anaya-Romero, M., Jordán, A., y Martínez-Zavala, L. (2003): Propuesta metodológica para la realización de un modelo de distribución potencial de usos forestales basado en parámetros edáficos, geomorfológicos, climáticos y topográficos. Consejería de Medio Ambiente (Junta de Andalucía) – Universidad de Sevilla. Sevilla, España.
- Bishop, C.M. (1998): *Neural Networks and Machine Learning*. NATO ASI Series. Series F: Computer and System Sciences. Vol. 68. *Springer-Verlag*. New York, EEUU.
- Breimann, K., Friedman, J.H., Olshen, R.A., y Stone, C.J. (1984): *Classification and Regression Trees*. *Chapman and Hall*. New York, EEUU.
- Cairns, D.M. (2001): A comparison of methods for predicting vegetation type. *Plant Ecology* 156:3-18.
- De la Rosa, D., y Moreira, J.M. (1987): Evaluación Ecológica de Recursos Naturales de Andalucía. Agencia de Medio Ambiente (Junta de Andalucía). Sevilla, España.
- ESRI (1982-1997): ArcInfo v.7.1.2. ESRI. Redlands, EEUU.
- Fischer, H.S. (1994): Simulation of the spatial distribution of plant communities based on maps of site factors: investigated in the MaB test site Davos. *Veröff. Geobot. Inst. Eidgenöss. Tech. Hochsch. Stift. Rübel Zür.* 122:1-136.
- Franklin, J. (1995): Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Prog. Phys. Geogr.* 19, 474-186.
- Gardner, M.W., y Dorling, S.R. (1998): *Atmospheric Environment* 32(14/15), 2627-2636.
- Gevrey, M., Dimopoulos, I., y Lek, S. (2003): Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling* 160:249-264.
- Guisan, A., y Zimmermann, E. (2000): Predictive habitat distribution models in ecology. *Ecological Modelling* 135:147-186.

- Kleinbaum, D.G. (1994): Logistic regression. A self-learning test. *Springer Verlag*. New York, EEUU.
- Münier, B., Nygaard, B., Ejrnaes, R., y Bruñi, H.G. (2001): A biotope landscape model for prediction of semi-natural vegetation in Denmark. *Ecological Modelling* 139:221-233.
- Neave, H.M., y Norton, T.W. (1998): Biological inventory for conservation evaluation. IV. Composition, distribution and spatial prediction of vegetation assemblages in southern Australia. *Forest Ecology & Management* 196:259-281.
- Martínez-Zavala, L. (2001). Análisis territorial de la comarca del Andévalo Occidental: una aproximación desde el medio físico. Tesis doctoral. Universidad de Sevilla. Sevilla.
- Recio, J.M., y Núñez, M.A. (1997): Cartografía de unidades geomorfoedáficas del Parque Natural Sierra de Aracena y su entorno. Consejería de Medio Ambiente (Junta de Andalucía) – Universidad de Córdoba. Sevilla, España.
- Rivas-Martínez, S. (1987): Memoria del mapa de series de vegetación de España. ICONA, Serie Técnica. *Servicio de Publicaciones del Ministerio de Agricultura Pesca y Alimentación*. Madrid, España.
- Robertson, M.P., Peter, C.I., Villet, M.H., y Ripley, B.S. (2003) : Comparing models for predicting species' potential distributions: a case study using correlative and mechanistic predictive modelling techniques. *Ecological Modelling* 164:153-167.
- Venables, W.N., y Ripley, B.D. (1999): Modern Applied Statistics with S-PLUS. *Springer*. New York, EEUU.
- Vogelmann, J.E., Sohl, T., y Howard, S.M.. (1998): Regional characterization of land cover using multiple sources of data. *Photogrammetric Engineering and Remote Sensing* 64(1):45-57.