USES OF SPATIAL PREDICTIVE MODELS IN FORESTED AREAS TERRITORIAL PLANNING

Ángel M. Felicísimo
Área de Ingeniería Cartográfica, Geodesia y Fotogrametría, Escuela Politécnica,
Universidad de Extremadura, E10071 Cáceres, España

ABSTRACT

Territorial planning of areas of forestal vocation try to resolve a number of problems such as current forests conservation and degraded areas restoration. In this work we uphold that for a correct plannification it is necessary to know the forestry potential of the territory expressed cartographically by means of suitability models. Such models can be generated objetively through spatial analysis methods relating the current presence/absence of a forest type to a set of influent environmental variables. This work presents an example of the utility of such models as planning criteria for future actions. We present the suitability model for Quercus pyrenaica forests in the Extremadura Autonomous Community (Spain). The analysis of the generated models allows to establish criteria for conservation and restoration, and also to select areas where actuation should be prioritary.

1. INTRODUCTION

Forestry planning in a territory is an attempt to resolve a set of problems, including forest conservation and restoration. In a zone such as the Iberian Peninsula, where forests have been steadily eliminated over centuries, two of the major aims of any forestry zoning plan are to reduce forest fragmentation and to conserve its biodiversity.

To achieve these goals requires basic territorial information of good quality [15] that allow to know the territory as well as guide actions from a spatial point of view. However, disponibility of such geographical information is not enough, and must be accompanied by objective means of analysis that permit, among other aims, to spatially delimitate actuation areas as well as define which actions are prioritary or recommendable.

Suitability models are an alternative rarely employed in forestry planning, and its utility will be demonstrated in this work. A *suitability model* is a map in which each pixel has assignated a value reflecting suitability for a given use. Suitability models can be generated through diverse techniques, such as logistic regression or the non-parametric CART (*classification and regression trees*) or MARS (*multiple adaptative regression splines*). All techniques require of a vegetation map (dependent variable) and of a set of environmental variables (climate, topography, litology, etc.) potentially influential on vegetation distribution. The foundation of the method is to establish relationships between the environmental variables and the spatial distribution of the vegetation. Commonly, each vegetation type will respond to a different model as a consequence of their different environmental requeriments.

The area defined as “suitable” in a suitability model should reflect the potential area for the considered vegetation type. Usually, current distribution areas are less than the potential area, because the forest has been cleared from zones where it was present in the past. Finally, knowing the potential distribution area allows to get valuable data for restoration actions as well as to delimitate areas where such actions are prioritary .

In the following we present the suitability models generated for the *Quercus pyrenaica* (*rebollo*, *melojo*) forests in Extremadura (Spain), but the described methods are of general use, and independent of the vegetation type or the geographical study area.

2. STUDY AREA, MATERIAL AND METHODS

2.1. Study area

Extremadura is one of the 17 Autonomous Communities in Spain. With a surface of 41680 km^2 (Figure 1), it has a Mediterranean climate tempered by its relative proximity to the sea and the penetration of oceanic fronts from the W. Predominant tree species include *Quercus rotundifolia* (*encina*) (19600 km^2), *Quercus suber* (*alcornoque*)(3140 km^2) and *Quercus pyrenaica* (*rebollo* or *melojo*)(2160 km^2).

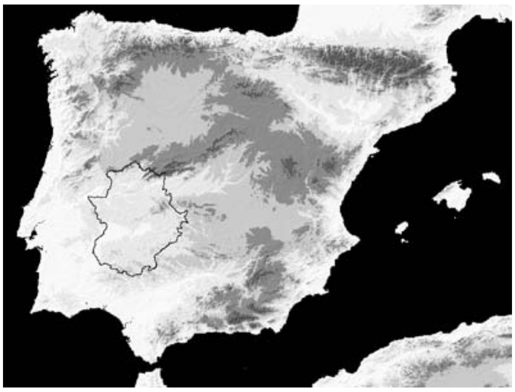

Figure 1. Geographic location of Extremadura in the Iberian Peninsula.

2.2. The dependent variable: current distribution of *Quercus pyrenaica* forests

The map with the current distribution of *Quercus pyrenaica* (Figure 2.) was generated from the "Mapa Forestal de España" (Dirección General de Conservación de la Naturaleza, Ministerio de Medio Ambiente).

2.3. The independent variables: digital terrain models

The environmental variables potentially influential were:

- elevation: the digital elevation model (DEM) was constructed using Delaunay's triangulation algorithm [19] from digitized hypsographic curves followed by a transform to a regular grid structure with a 100 m cell size.

- potential insolation: the models were constructed by simulation from the DEM, analyzing topographical shading [4] as a function of the sun's trajectory for standard date periods [8]. The result is an estimate of the amount of time that each point of the terrain receives direct solar radiation, with a 20-minute temporal resolution and 100-m spatial resolution.

- Mean maximum and minimum annual temperatures: interpolated from a total of 140 meteorological stations with the *thin plate splines* method ([10], [14]), with a 500-m spatial resolution.

- 3-month total precipitation: interpolated from a total of 276 meteorological stations with the *thin plate splines* method ([10], [14]), with a 500-m spatial resolution. Data for the two last sets come from the National Institute of Meteorology of Spain (*Instituto Nacional de Meteorología*).

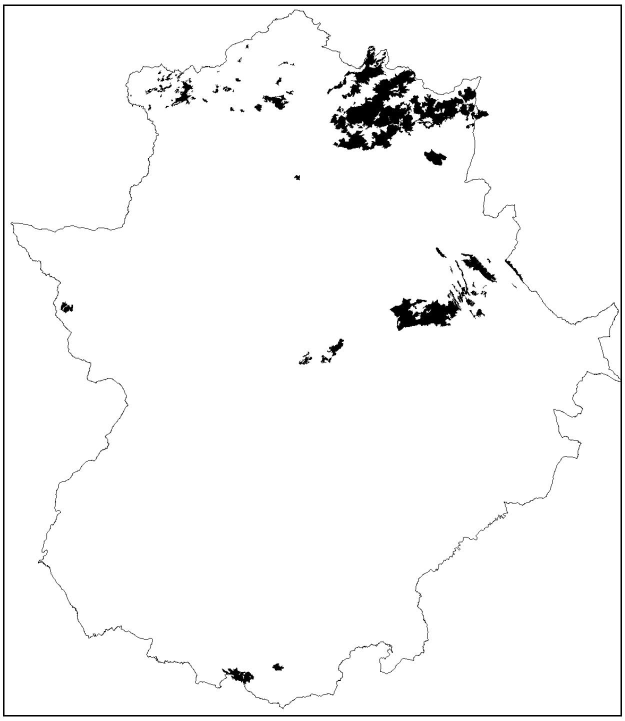

Figure 2. Current distribution of *Quercus pyrenaica* (*rebollo*) in Extremadura according to the "Mapa Forestal de España" (black, approximately 2160 km^2).

2.4. Methods: logistic mutiple regression (LMR)

Logistic multiple regression has been used as a forecasting method to generate probability models in a variety of fields, such as epidemiology [20], geological prospecting [1], silviculture or wildlife conservation [16][18]. LMR is adecuate because the dependent variable is dichotomous (presence/absence) and the model admits non-Gaussian independent variables. Finally, result values vary smoothly from 0 to 1 so that it is well-suited to generate a probability or suitability model [12].

The introduction of a spatial component into the LMR to generate cartographic models is recent, and is usually integrated into geographic information systems. As an example, [7] use LMR in the ArcInfo GIS (ESRI Inc.) to generate a model of the distribution of a plant species, *Carex curvula* in the Swiss Alps. A similar study, applied to aquatic vegetation [21] used GRASS GIS (US Army Construction Engineering Research Laboratory). Finally, we have used ArcView (ESRI Inc.) to generate the logistic models for Cantabria (España) [3]. Procedure and the statistical foundations are detailed there. Here we just point out that the logistic model expresses suitability, P(*i*), respect to the values of *n* explanatory variables following the expression:

$$P(i) = \frac{1}{1 + e^{-(b(0) + b(1)\cdot x(1) + ... + b(n)\cdot x(n))}}$$

where P(*i*) represent the suitability value, *x*(1) ... *x*(n) the values of the environmental variables and *b*(1) ... *b*(n), coefficients. The results for each pixel vary between 0 (incompatibility) and 1 (ideal).

In the present work the regression coefficients were calculated by a stratified sampling over presence/absence areas of *Quercus pyrenaica* forest. The process is as follows:

1. We perform two random sampling over each forest/non-forest areas. One is used to generate the model (training sample) and the second to test its performance (test sample). According to the recommendations in [17], both samples have equal number o positive and negative cases.

2. Logistic regression was performed in a commercial statistical software package with the method *stepwise* by maximum likelihood.

3. Goodness-of-fit was calculated comparing the results of the training sample with the test sample for several cut values, measuring the AUC: Area Under the ROC (*Receiver Operating Characteristic*) curve.

Finally, we apply the logistic equation for the whole of the territory to generate the probability model. Results were compared with the map of present vegetation.

3. RESULTS

3.1. The logistic model

The model used 7 significative variables, which coefficients and basic statistics are shown in Table 1.

variable	b	e. e.	Wald	P	Description
dem	0.0055	0.0001	2425.3	<0.0001	elevation (digital elevation model)
ins12p	-0.1015	0.0114	79.2	<0.0001	potential insolation (solar declination +12º)
rain1	-0.0262	0.0006	1782.9	<0.0001	mean January-March rainfall
rain2	0.0146	0.0005	799.9	<0.0001	mean April-June rainfall
rain3	0.0565	0.0013	1886.6	<0.0001	mean July-September rainfall
rain4	0.0156	0.0003	2843.4	<0.0001	mean October-December rainfall
temp	-0.0581	0.0012	2418.9	<0.0001	annual mean minimum temperature
constant	10.2259	0.5780	313.0	<0.0001	

Table 1. Variables, coeficientes de regresión y significación para la regresión logística.
b = coeficiente de regresión, e. e. = error estándar, Wald = estadístico de Wald

Model predictive capacity (accuracy) can be evaluated as the percent of correctly classified cases, both for presences and absences. The common method is to consider the Area Under the ROC curve (AUC), calculated from sensibility and specificity values. AUC takes a maximum value of 1 for a model with a perfect fit. AUC in our study covers an area of 0.973, which represent a very good model accuracy.

3.2. *Quercus pyrenaica* suitability model

The last step to create the model is to generate the suitability map applying the logistic model to the whole of the territory. Our results are shown in the Figure 3. Values have been classified in 5 suitability classes to facilitate interpretation.

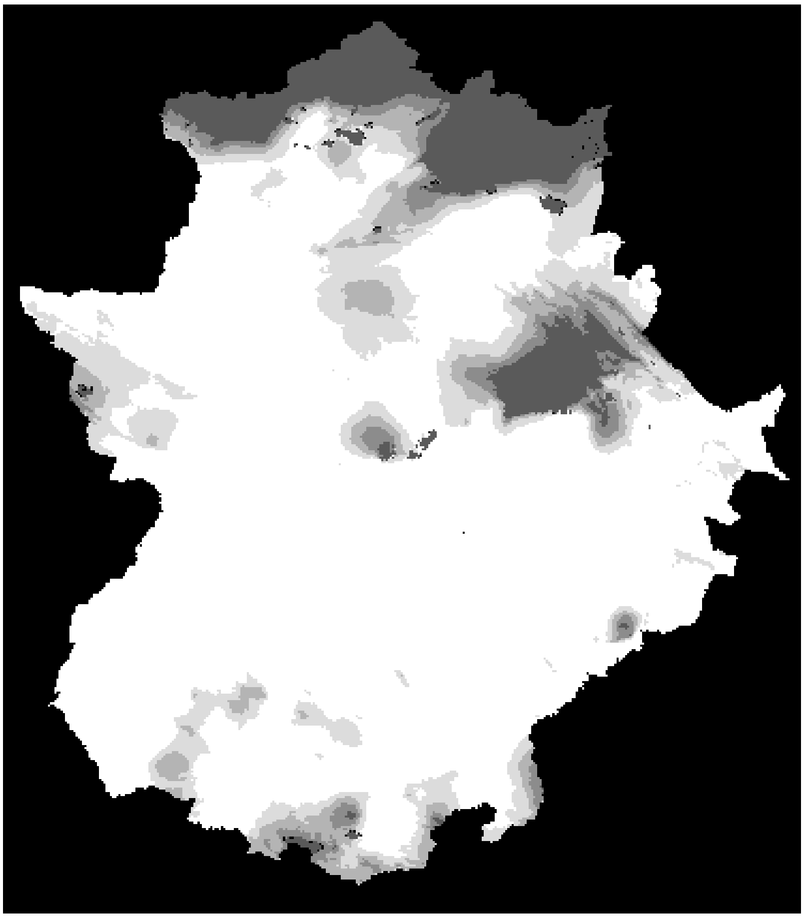

Figure 3. Application of the logistic model to the study area. Values leyend - white: 0.00-0.10 (incompatible); light grey: 0.11-0.25; medium grey: 0.26-0.50; dark grey: 0.51-0.75; very dark grey: 0.76-1.00. Area outside the study area in black.

3.3. Potencial area disponibility

Superposing the probability model over the Forest Map of Spain (*Mapa Forestal de España*) we check which is the current vegetation present for each suitability value of *Q. pyrenaica*. Such analysis is similar to that in [9], who estimated diversity loss resulting from invasive plants expansion. This comparison between both maps, suitability and present vegetation, allows to get results needed to understand some aspects of vegetation dynamics and facilitate procedure proposals. Results on woodlands, forests and plantations are shown as an example.

3.3.1. Other forests

Some 1198 km^2 of the potential area of *Q. pyrenaica* are at present covered with two *Quercus* species typical of other vegetation series. Figure 4 shown frequency histograms for both on the *Q. pyrenaica* distribution area. *Quercus pyrenaica* suitability values have been grouped in 9 classes. From the histograms it is clear that both *Quercus suber* (*alcornoque*) and *Q.*

rotundifolia (*carrasca*) grow under different environmental conditions, as their presence on the areas of highest suitability for *Q. pyrenaica* is small.

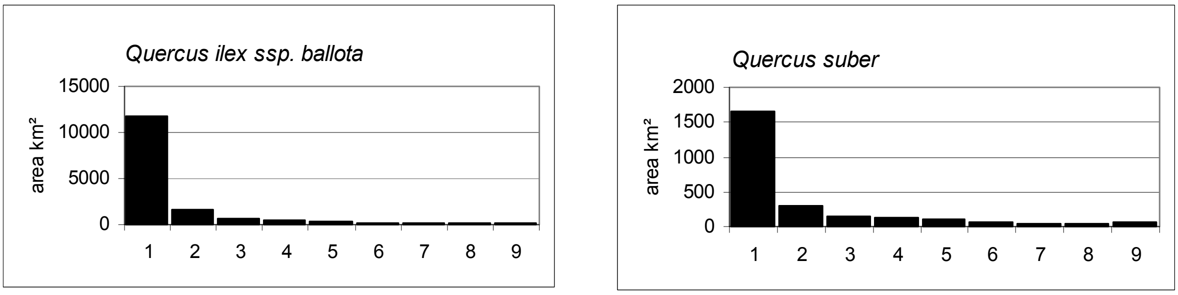

Figure 4. Left: Presence of *Q. ilex* ssp. *ballota* (= *Q. rotundifolia*) on *Q. pyrenaica* suitability classes. Both forests are mutually excluded, as *Q. rotundifolia* almost dissapear from areas with high suitability for *Q. pyrenaica*. Right: Presence of *Q. suber* on *Q. pyrenaica* suitability classes. As in the previous case, there exist exclusion between both forest types, althoughh not so sharp (see text). Suitability classes: 1 (0.00-0.11); 2 (0.12-0.22); 3 (0.23-0.33); 4 (0.34-0.44); 5 (0.45-0.55); 6 (0.56-0.66); 7 (0.67-0.77); 8 (0.78-0.88) and 9 (0.89-1.00).

In forest planning this zones should be excluded from the potential area of *Q. pyrenaica*, as they are covered by other vegetation type that must be conserved. It is interesting to note that the mutual exclusion with *Q. suber* is not so strict as with *Q. rotundifolia*. In fact, *Q. pyrenaica* and *Q. suber* form mixed stands of considerable extension. Although considered at times as a xerophyte, *Quercus suber* has in fact humidity requeriments that exclude it from the more arid zones, where *Q. rotundifolia* is the dominant species.

3.3.2. Plantations

In the study area there are plantations of *Castanea sativa* and several species of *Pinus* and *Eucaliptus*. Frequency histograms are shown in Figure 5.

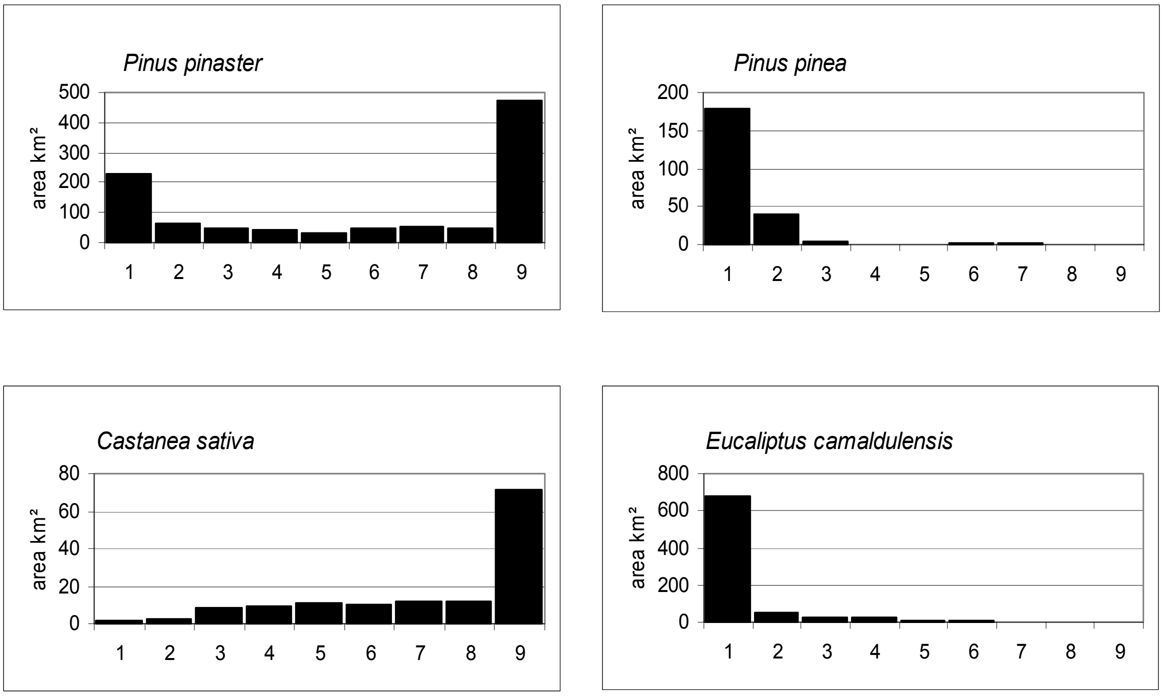

Figure 5. Presence of the most common plantation speices in the suitability classes of *Q. pyrenaica*. Suitability classes as in Figure 4.

From the histograms it is clear that both *Pinus pinaster* and *Castanea sativa* were planted on areas of high suitability for *Q. pyrenaica*. On the other hand, *Pinus pinea* and *Eucaliptus camaldulensis* shown different preferences and grow in areas with low suitability for *Q. pyrenaica*. Another species not shown, *Pinus sylvestris*, with only 11.5 km^2 coverage in the area, appears exclusively in the class of highest suitability.

Finally, it must be noted the presence on classes of high suitability of shrublands dominated by *Cytisus* spp. (*escobonales*) and *Erica* spp. (*brezales*). This formations must be considered as substitution stages where the primary forest has been cleared; if soil degradation is moderate, *Cytisus* formations dominate, whilst *Erica* formations commonly develop on more degraded soils.

4. APLICATIONS IN FOREST PLANNING

Suitability maps, along with ocupation by other species statistics let forest planning and also define priority areas of application. For example, in the study area can be defined three main zones:

Type 1 Zones, where conservation of present vegetation in prioritary. This zones are those currently occupied by *Q. pyrenaica* forests and *Castanea sativa* woodlands, an exotic species but well integrated in local vegetation.

Type 2 Zones, where actions should tend to reduce forest fragmentation through recovery of areas occupied by secondary formations (in the sense of ecological succession). They comprise *Cytisus* spp. shrublands, with moderate soil degradation, and *Erica* spp. heaths, where soil degradation is more severe. Action zones will be those combining presence of those substitution secondary formations with the highest suitability values for *Q. pyrenaica*. Priority actions will take place in the zones of suitability classes 8 and 9 of largest extension adjacent to *Type 1 Zones*, which facilitate seed arrival and, consequently, rapid regeneration.

Type 2 Zones, where the exotic species, mainly *Pinus* and *Eucalyptus* plantations, will be progressively replaced. Priority actions will be taken, as in the previous case, on zones with high suitability values and proximity to *Type 1 Zones*.

Figure 6 shows the "Map of Use Orientation" of the enumerated actions in one of the two zones of highest suitability values in the study area: North Zone in the Cáceres Province, and Las Villuercas region (Figure). In the latter, besides relatively large areas of well-preserved *Q. pyrenaica* forest, there exist the possibility of reducing forest fragmentation by recuperation of surrounding shrublands separating forest patches, and also enlarge the final extension operating also on the *Pinus* plantations.

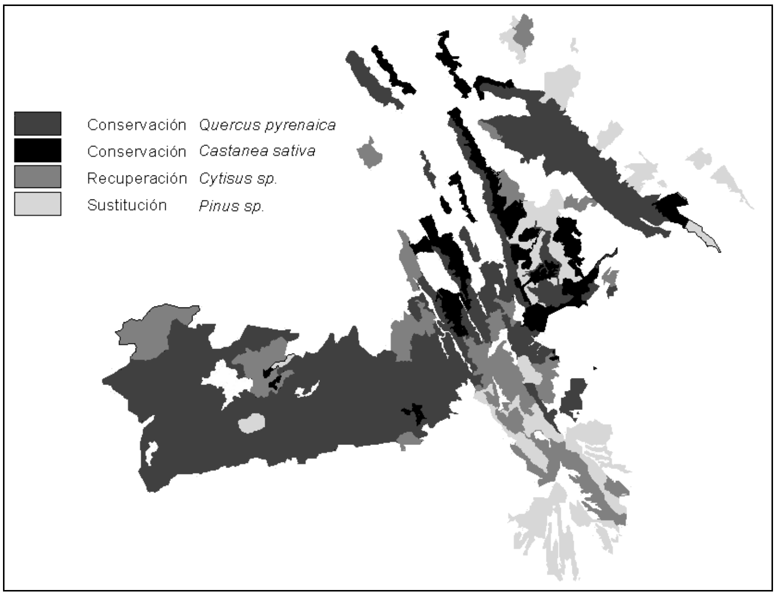

Figure 6. Map of Use Orientation based on the criteria explained in the text: Las Villuercas zone (E Extremadura).

5. DISCUSSION

Traditionally, environmental factors have been considered as the main determinant in vegetation distribution [13]. First attemps to clarify the relationships between environment and vegetation were analytical, exploring association between species and one variable [5]. Later on, computers allow generalization of multivariate analyses (factorial analysis, classification, etc.). Suitability models generated by logistic regression synthesize the response of the species to a set of environmental variables affecting its spatial distribution. They are models based on real data (*data driven*), and in statistical techniques, which makes them objective methods. In our experience those methods give good results in montane areas because limiting factors are mainly physical: elevation, potential insolation, slope, etc. These factors are generally known (elevation), or they can be derived with enough accuracy (potential insolation).

Quercus pyrenaica is a widely distributed species in the Iberian Peninsula, but always associated to the so call “submediterranean sector”: mountain areas with larger rainfall than surrounding plains, which allows the growing of species which do not tolerate summer dryness [2]. The generated suitability model reflects rather exactly the current spatial distribution of the species, and predicts a potential area distribution of reasonable extension and characteristics from a biological point of view. Another argument improving model aparent reliability is the analysis of the current vegetation in the potential area. In all cases they are sustitution formations in the sense of community ecological succession. Shrublands and heaths were generated by forest clearing and subsequent soil loss. Pine plantations were most likely done on those previously degraded areas, and current management includes selective cut for wood production [11].

To transfer the model into actions is an issue needing complementary information, specially on soil property and its current economical management. To deal with this aspect is beyond the aim of this paper, and the proposed actions should be considered only as examples of direct application of the information generated by the model. It must be pointed, though, that the generated maps and statistics reflect objective data of obvious utility. Combination of model maps with territory current uses and landscape management data could successfully complete an objective decission system extremely useful in territorial planning.

ACKNOWLEDGEMENTS

This paper is part of the Project "*Desarrollo de modelos de idoneidad para la vegetación en Extremadura mediante Sistemas de Información Geográfica. Aplicación a la conservación y regeneración de recursos, a la gestión y conservación de la diversidad biológica y a la recuperación de zonas degradadas*" co-funded by the Junta de Extremadura (Consejería de Educación, Ciencia y Tecnología - II Plan Regional de Investigación, Desarrollo Tecnológico e Innovación de Extremadura) and FEDER (Fondo Europeo de Desarrollo Regional).

JUNTA DE EXTREMADURA

6. BIBLIOGRAPHY

[1] Agterberg, F.P. Estimating the probability of occurrence of mineral deposits from multiple map patterns. En *Use of Microcomputers in Geology*, D.F. Merriam and H. Kürzl (Eds.), pp. 73-92. Plenum Press, New York., 1992.

[2] Blanco Castro, E., Casado González, M.A., Costa Tenorio, M., Escribano Bombín, R., García Antón, M., Génova Fuster, M., Gómez Manzaneque, A., Gómez Manzaneque, F., Moreno Saiz, J.C., Moria Juaristi, C., Regato Pajares, P. and Sainz Ollero, H. *Los bosques ibéricos. Una interpretación geobotánica*, Planeta, Barcelona, 1997.

[3] Felicísimo, A.M., Francés, E., Fernández, J.M., González-Díez, A. and Varas, J. Modeling the potential distribution of forests with a GIS. *Photogrammetric Engineering and Remote Sensing*, 68, 5, 455-461, 2002.

[4] Fernández Cepedal, G. and Felicísimo, A.M. Método de cálculo de la radiación solar incidente en áreas con apantallamiento topográfico. *Revista de Biología de la Universidad de Oviedo*, 5, 109-119, 1987.

[5] Gauthier, B., Godron, M., Hiernaux, P. and Lepart, J. Un type complementaire de profil écologique: le profil écologique 'indicé'. *Canadian Journal of Botany*, 33, 2859-2865, 1977.

[6] Godron, M. *Les principaux types de profils écologiques*. CNRS-CEPE, 8 pp. Montpellier, 1965.

[7] Guisan, A., Theurillat, J.-P. and Kienast, F. Predicting the potential distribution of plant species in an alpine environment. *Journal of Vegetation Science*, 9, 65-74, 1998.

[8] Heywood, H. Standard date periods with declination limits. *Nature*, 204, 678, 1964.

[9] Higgins, Steven I., Richardson, David M., Cowling, Richard M. and Trinder-Smith, Terry H. Predicting the landscape-scale distribution of alien plants and their threat to plant diversity. *Conservation Biology*, 13, 2, 303-313, 1999.

[10] Hutchinson, M.F. The application of thin plate smoothing splines to continent-wide data assimilation. En *Data Assimilation Systems*, J.D. Jasper (Ed.), pp. 104-113. BMRC Research Report No. 27, Bureau of Meteorology, Melbourne, 1991.

[11] ICONA. *Segundo Inventario Forestal Nacional 1986-1995. Extremadura, Cáceres*, Ministerio de Agricultura, Pesca y Alimentación, Madrid, 1994.

[12] Jongman, R.H.G., Ter Braak, C.J.F. and van Tongeren, O.F.R. (Eds.) *Data analysis in community and landscape ecology*, Cambridge University Press, Cambridge, 1995.

[13] Leduc, A., Drapeau, P., Bergeron, Y. and Legendre, P. Study of spatial components of forest cover using partial Mantel tests and path analysis. *Journal of Vegetation Science*, 3, 69-78, 1992.

[14] Lennon, Jack J. and Turner, John R.G. Predicting the spatial distribution of climate temperatura in Great Britain. *Journal of Animal Ecology*, 64, 370-392, 1995.

[15] Lund, H. G. and Iremonger, S. Omissions, commissions, and decisions: the need for integrated resource assessments. *Forest Ecology and Management*, 128, 3-10, 2000.

[16] Mladenoff, David J., Sickley, Theodore A. and Wydeven, Adrain P. Predicting gray wolf landscape recolonization: logistic regression modelos vs. new field data. *Ecological Applications*, 9, 1, 37-44, 1999.

[17] Narumalani, S., Jensen, J.R., Althausen, J.D., Burkhalter, S. and Mackey, H.E. Aquatic macrophyte modeling using GIS and logistic multiple regression. *Photogrammetric Engineering and Remote Sensing*, 63, 1, 41-49, 1997.

[18] Pereira, J.M.C. and Itami, R.M. GIS-based habitat modeling using logistic multiple regression: a study of the Mt. Graham Red Squirrel. *Photogrammetric Engineering and Remote Sensing*, 57, 11, 1475-1486, 1991.

[19] Peucker, T.K., Fowler, R.J.. Little, J.J. and Mark, D.M. The triangulated irregular network. Proceedings of the ASP Digital Terrain Models (DTM) Symposium, pp. 516-540. American Society of Photogrammetry, Falls Church, Virginia, 1978.

[20] Thomson, M.C., Elnaiem, D.A., Ashford, R.W. and Connor, S.J. Towards a kala azar risk map for Sudan: mapping the potential distribution of Phlebotomus orientalis using digital data of environmental variables. *Tropical Medicine and International Health*, 4, 2, 105-113, 1999.

[21] van de Rijt, C.W.C.J., Hazelhoff, L. and Blom, C.W.P.M. Vegetation zonation in a former tidal area: a vegetation-type response model based on DCA and logistic regression using GIS. *Journal of Vegetation Science*, 7, 505-518, 1996.

Wilson, W.L., Day, K.R. and Hart, E.A. Predicting the extent of damage to conifer seedlings by the pine weevil (Hylobius abietis L.): a preliminary risk model by multiple logistic regression. *New Forests*, 12, 203-222, 1996.