

ECOLOGICAL NICHE MODELLING WITH HERBARIUM DATA: A FRAMEWORK TO IMPROVE NATURA 2000 HABITAT MONITORING

AMICI V.^{1,*}, GERI F.², BONINI I.^{1,3}, ROCCHINI D.⁴

¹ *Department of Environmental Science "G.Sarfatti", University of Siena
Via P.A. Mattioli 4, 53100 Siena, Italy
(phone: +39 0577 232864; fax: +39 0577 232896)*

² *Laboratory of Ecology DICAM - Department of Civil, Environmental and Mechanical
Engineering, University of Trento,
Via Mesiano 77, 38123 Trento, Italy*

³ *Herbarium Universitatis Senensis
Via P.A. Mattioli, 4 - 53100 Siena
(phone: +39 0577 235409)*

⁴ *Department of Biodiversity and Molecular Ecology, Research and Innovation Centre,
Fondazione Edmund Mach,
Via E Mach 1, 38010 S. Michele all'Adige (TN), Italy.*

** Corresponding author
e-mail: valerio.amici@gmail.com;*

(Received 18th Jan 2012; accepted 22nd July 2014)

Abstract. The effectiveness of biodiversity conservation strategies depends on the knowledge about the distribution of habitats or single species. Despite this, efforts on biodiversity monitoring and conservation are currently hindered by a lack of information about the spatial distribution of species on large landscapes. Predictive species distribution models, can provide a powerful tool for solving this ecological problem. The vast majority of data available for modelling plants distribution are herbarium data, which lack reliable records of species absence. Although it has been found that herbarium records do not meet current standards for sampling in ecological studies, they remain often the only available source of sufficient magnitude with regard to relevant distribution data. Modifying existing statistical tools and developing new methods so that herbarium data, despite their shortcomings, can be used for modelling habitat suitability, is currently a growing field. The aim of this paper was to analyse the opportunities and bottlenecks for future application of distribution models in the mapping and monitoring of habitats of conservation interest in a complex Mediterranean area. Here we specifically concentrate on testing the Maximum entropy (Maxent) approach to estimate the distribution of a training habitat through the use of herbarium records and to explore a GIS-based integrated approach. The results obtained highlighted the important role that distribution models can have in individuating the areas where a targeted species or habitat type is most likely to be found, and in showing where to commit the limited available resources for inventories.

Keywords: *Ecological niche, GIS, Habitat distribution, Herbarium data, Maxent.*

Introduction

Identifying and mapping habitats of conservation interest around the globe represents a crucial step to individualize effective tools for biodiversity monitoring and conservation (Margules and Pressey, 2000; Groves et al., 2002). The effectiveness of conservation strategies crucially depends on the knowledge about the distribution of habitats or single species (Cianfrani et al., 2010). Alarming, efforts on biodiversity

monitoring and conservation are currently hindered by a lack of information about the spatial distribution of species on large landscapes (Wilson et al., 2004). In this context, predictive species distribution models (Guisan and Zimmermann, 2000) based on complex ecological databases, can provide a powerful tool for solving this ecological problem.

Species distribution models attempt to estimate a species' niche across a geographical area by relating presence records of the species to environmental/ecological predictors. These models estimate the probability that species occur in areas where it has not directly been observed given an array of measured environmental variables (Segurado and Araujo, 2004). These environmental variables can include any sort of biotic or abiotic features that may favor the presence of the species being modeled and that can be measured spatially. Distribution models (also called 'habitat models') now have an established place within conservation biology, where they support many crucial aspects as landscape planning, reserve placement, habitat monitoring, and identification of suitable restoration sites (Funk and Richardson, 2002; Rodriguez et al., 2007; Lobo, 2008).

Generally, distribution models use either presence-only data, or both presence and absence data (Guisan and Zimmermann, 2000; Elith et al., 2006) coming from a variety of sources as museums, herbaria, published inventories, transects and species lists (Stockwell and Peterson, 2002; Soberón and Peterson, 2004). While species data from planned surveys are ideal for modelling distributions, records for most species of the world are in a presence-only form (Cawsey et al., 2002). These are usually derived from ad-hoc observations that lack reliable records of species' absence (Dennis and Hardy, 1999; Graham et al., 2004; Lobo and Tognelli, 2011).

The vast majority of data available for modelling plants distribution are herbarium data, which are typically presence-only (Zaniewski et al., 2002; Wollan et al., 2008). However, it has been found that herbarium records provide limited accuracy in predicting distribution and do not meet current standards for sampling in ecological studies, as many populations may have been overlooked due to inadequate geographic coverage, or various collection biases (Rich and Woodruff, 1992; Heyligers, 1998; MacDougall et al., 1998; Hirzel and Guisan, 2002; Delisle et al., 2003; Ungricht et al., 2005). Nevertheless, herbarium data remain often the only available source of sufficient magnitude with regard to relevant and ample distribution data. Therefore, modifying existing statistical tools and developing new methods so that herbarium data, despite their shortcomings, can be used for modelling habitat suitability, is currently a growing field (e.g. Hirzel et al., 2002; Reutter et al., 2003; Engler et al., 2004; Hirzel et al., 2006).

A number of alternative modeling algorithms have been applied to classify the probability of species' presence as a function of a set of environmental variables and several studies have demonstrated that different modeling approaches have the potential to yield substantially different predictions (e.g. Loiselle et al., 2003; Thuiller, 2003; Brotons et al., 2004; Segurado and Araújo, 2004; Elith et al., 2006; Pearson et al., 2006). The Maximum entropy model (Maxent; Phillips et al., 2006; Elith et al., 2011) showed promising results in comparison with alternative methods (Elith et al., 2006; Pearson et al., 2006; Phillips et al. 2006; Peterson et al., 2007).

The aim of this paper is to disentangle the opportunities and drawbacks for future application of distribution models in the mapping and monitoring of habitats of conservation interest as these models could represent a straightforward but robust tool

to improve, among others, the monitoring of Natura2000 habitats. Here we specifically concentrated on testing the Maximum entropy (Maxent) approach to estimate the distribution of habitat characterized by an intrinsic an internal complexity through the use of herbarium records.

Material and Methods

Study area

The study area covers the entire territory of the Tuscany region, situated in central Italy between 9° and 12° east longitude and 42° and 44°30' north latitude. Tuscany is delimited by the Northern Apennines on the Northeast side and by the Tyrrhenian Sea on the West side (Fig. 1).



Figure 1. Study area (in red).

Forests vary from the evergreen Mediterranean forests dominated by *Quercus ilex*, along the coastlines, to the *Fagus sylvatica*, *Castanea sativa* and *Abies alba* forests of mountain sites. Corine land-cover data (Bossard et al., 2000; APAT, 2005) show a surface territory of about 19,720 km² of which 44% is covered by forests, while the agriculture area covers about 46%. The agriculture types that take up the larger surface area include intensive non-irrigated arable land alternated with traditional agro-ecosystems, while broad-leaved forests represent the major natural class. From an environmental point of view, the region is peculiar for its extremely heterogeneous morphological and climatic features. The topography varies from the plain areas near the coast line and around the principal river valleys to the hilly and mountainous zones towards the Apennine chain. From the climatic point of view, Tuscany is influenced by

its complex orographic structure and by the direction of the prevalent air flows: the central and southern parts of Tuscany fall within the temperate climatic zone with a dry season, the Mediterranean climate category, whereas the northern portion has some continental climate characteristics (Rapetti and Vittorini, 1995).

Target habitat and Herbarium data

The Natura2000 habitat “Arborescent matorral with *Juniperus* spp.” (5210, corresponding to 32.13 Corine Biotope code) was selected here to test an integrated Maximum Entropy approach in habitat distribution estimation. This habitat has been selected as target habitat in Maxent distribution as: *i*) it is a habitat of conservation interest in Tuscany, *ii*) has a narrow ecological niche, *iii*) is one of the better represented habitat, in terms of characterizing species, in the Herbarium Universitatis Senensis (University of Siena, Italy).

The target habitat consists of Mediterranean and sub-Mediterranean evergreen sclerophyllous bush and scrubs, organized around arborescent junipers, and it is found in the countries bordering the Mediterranean Sea (Calaciura and Spinelli, 2008). *Juniperus* spp. are evergreen shrubs or small trees with few vital needs that, thanks to their morpho-physiological characteristics, colonize harsh environments such as rocky coasts and dry, incoherent soils (Pignatti, 1982; Jordano, 1991). Juniper, with its deep and well-developed root system, is therefore an important species for soil retention and consolidation, preventing soil erosion caused by rain and wind (Mondino and Bernetti, 1998). It is important as an ecotone, since it is often a transition area between ecosystems (Calaciura and Spinelli, 2008).

In the study area, the characteristic physiognomic combination of the arborescent matorral is: *Juniperus oxycedrus*, *J. phoenicea*, *Pistacia lentiscus*, *Rhamnus alaternus*, *Phillyrea latifolia*, *Myrtus communis*, *Lonicera implexa*, *Smilax aspera*, *Rubia peregrina*, *Clematis flammula*, *Helichrysum stoechas*, *Brachypodium ramosum* (De Dominicis et al., 1988; Chiarucci et al., 1999; Viciani et al., 2005; Foggi et al., 2006a,b).

Field data on characterizing species were gathered by the digitization and georeferencing of the essiccata preserved in the Herbarium Universitatis Senensis, for a total of 254 records (*Table 1*).

Modelling procedure

We considered 13 environmental variables as potential predictors of the target habitat distribution (*Tab.2*). An ortho-Landsat ETM+ image (path 192, row 030, acquisition date June 20, 2000; spatial resolution 30 meters) was acquired. Bands 1 (blue, 0.45 – 0.515 μm), 2 (green, 0.525 – 0.605 μm), 3 (red, 0.63 – 0.69 μm), 4 (near infrared, 0.75 – 0.90 μm), 5 (middle infrared, 1.55 – 1.75 μm) and 7 (middle infrared, 2.09 – 2.35 μm) were considered; bands 6.1, 6.2 and 8 were not taken into account due to the different pixel size from that of the other bands (60 meters of ground resolution opposed to 30 meters of the other bands). In order to obtain topographic variables we relied on a 75 meters digital elevation model of Tuscany region (DEM) resampled by a nearest neighbour algorithm at a spatial resolution of 30 m, processed with an algorithm of terrain analysis with the production of derivate images like altitude, slope, aspect and solar radiation. The climatic variables (1990-2000) were obtained by spatial interpolation (inverse distance weighting method) of 130 climatic stations operated by

the Regional Agency for Innovation in Agriculture (ARSIA). All the image processing operations were performed using Grass GIS and QGIS software and resampled to 30 m pixels.

Table 1. Number of training records used in the Maxent models and obtained average training AUC and standard deviation values.

Maxent model	N° of records	AUC values	Standard deviation
<i>Brachypodium ramosum</i>	14	0.957	0.038
<i>Clematis flammula</i>	14	0.824	0.048
<i>Helichrysum stoechas</i>	17	0.980	0.026
Juniperus oxycedrus	34	0.971	0.012
Juniperus phoenicea	24	0.975	0.105
<i>Lonicera implexa</i>	23	0.916	0.016
<i>Myrtus communis</i>	18	0.911	0.015
<i>Phillyrea latifolia</i>	31	0.891	0.038
<i>Pistacia lentiscus</i>	17	0.903	0.030
<i>Rhamnus alaternus</i>	13	0.891	0.068
<i>Rubia peregrina</i>	20	0.852	0.036
<i>Smilax aspera</i>	29	0.913	0.024

A correlation analysis was performed, using the Spearman correlation coefficient (*cor* function within the R statistical software, R Development Core Team, 2013), in order to exclude pairs of related variables from the model. The analysis of the Spearman's rho statistic allowed the exclusion of pairs of highly correlated variables (Spearman's rho values ≥ 0.6), obtaining a set of 9 variables to be used as environmental layers in the Maxent modelling (Table 2).

Table 2. Environmental variables used for Ecological niche modelling of the occurrence of *Arborescent matorral with Juniperus spp. habitat*.

Spectral variables (Ortho-Landsat ETM+ image: path 192, row 030, acquisition date June 20, 2000; spatial resolution 30 meters)	band 1 (blue, 0.45 – 0.515 μm) band 4 (near infrared, 0.75 – 9.90 μm) band 7 (middle infrared, 2.09 – 2.35 μm)
Geomorphologic variables (DTM; spatial resolution 30 m)	Altitude (ALT) Slope (SLO) Solar radiation (SR)
Climatic variables (Spatial interpolation of 130 climatic stations operated by the Regional Agency for Innovation in Agriculture, ARSIA; spatial resolution 30m)	Mean winter temperature (MWR) Mean summer temperature (MST) Mean Annual Rainfall (MAR)

We applied here the Maximum Entropy modelling method (Maxent; Phillips et al., 2006; Phillips and Dudik 2008; Elith et al., 2011) in order to estimate the potential habitat distribution for the species characterizing the selected target habitat. Maxent is a

maximum entropy based machine learning program (<http://www.cs.princeton.edu/~schapire/maxent/>) that uses the principle of maximum entropy on presence-only data in order to approximate the species' niche and potential geographic distribution (Elith et al., 2006; Ortega-Huerta and Peterson, 2008; Hernandez et al., 2006; Pearson et al. 2007; Papes and Gaubert, 2007; Wisz et al., 2008; Kumar and Stohlgren, 2009; Warren et al., 2011). In estimating the probability distribution of a species across an area, Maxent formalizes the principle that the estimated distribution must agree with everything that is known (or inferred from the environmental conditions where the species has been observed) but should avoid making any assumptions that are not supported by the data (Shannon, 1948; Phillips et al., 2006). The approach attempts to estimate the maximum entropy (the distribution that is most spread-out, or closest to uniform) subject to constraints imposed by the information available regarding the observed distribution of the species and environmental conditions across the study area (Phillips et al., 2004; Baldwin, 2009).

The model evaluates the suitability of each pixel as a function of environmental variables and estimates the most uniform distribution given the constraint that the expected value of each environmental predictor variable matches its empirical average (average values for the set of presence-only occurrence data) (Hernandez et al., 2008; Ficetola et al., 2010). Output from Maxent models is a raster map that represents the probability of presence of a species and that is made up of pixels whose values varies from 0 to 1, where 0 is the lowest and 1 the highest probability.

In order to assess the predictive performance of the models we followed the most commonly used approach that involves the use of the Receiver Operating Characteristic curves (ROC; Hanley and McNeil, 1982; Zweig and Campbell 1993). The area under curve (AUC) value indicates the model accuracy (Phillips et al., 2006; Jimenez-Valverde, 2011). For random prediction, AUC is 0.5. The main advantage of ROC analysis is that the AUC provides a single measure of model performance, independent of any particular choice of threshold.

In this work we performed a bootstrap replicated run to do multiple runs (100) for each characterizing species (a total of 12 distribution models); through this method the training data is selected by sampling with replacement from the presence points, with the number of samples equaling the total number of presence points (Phillips et al., 2006).

Then, the 12 Maxent raster outputs have been processed through overlay operations using the *raster map calculator* module of GRASS GIS software (*Fig.2*), in order to obtain a distribution map of the target Natura2000 habitat.

Accuracy assessment

In order to assess the accuracy of the distribution map of the target habitat, we generated 100 random test points, 50 in the areas characterized by an high probability to found the habitat and e 50 in the remaining areas. The threshold value of probability used to classify area with high or low suitability was equal to 0.5 (average result of multiple runs). Once located with a high precision GPS, the test points have Been verified in the field on the basis of the presence of the characteristic physiognomic combination in a neighborhood of 30 meters.

In order to assess the reliability of the model, the Kappa Index of Agreement (KIA) (Cohen, 1960) was computed using the following algorithm:

$$K = (p_o - p_e) / (1 - p_e)$$

where p_o is the probability or proportion expected by chance and p_e is the probability or proportion observed.

The obtained KIA index allowed the assessment of the spatial agreement between the test points and the distribution map of the target habitat. The evaluation of the degree of agreement has been performed according to the interpretation scale proposed by Landis and Kock (1977), where values equal to 0 means null agreement and values equal to 1 means maximum agreement.

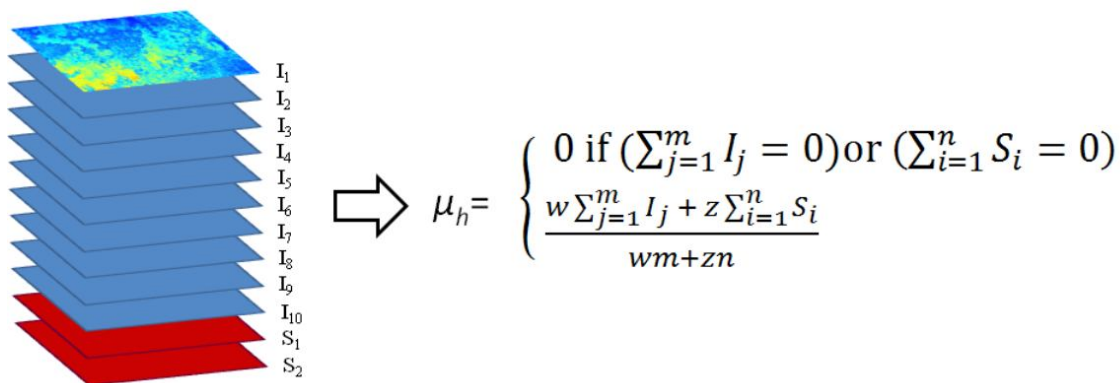


Figure 2. Overlay algorithm used to obtain the habitat probability of occurrence (μ_h). I =key species (*Juniperus*); m = number of key species; S =other characterizing species; n = number of characterizing species; w = value of weight for key species; z = value of weight for characterizing species.

Results and discussion

The approach proposed in this study yielded an habitat distribution map, highlighting how the application of a such integrated method could represent an accurate and cost-effective tool to obtain distribution maps for large and ecologically complex areas (Amici, 2011). Moreover, the modelling approach (Maxent) and the accurate geo-referencing of the herbarium specimens, allowed to use a fair amount of data that are often required to estimate species or habitat distribution for large areas. This aspect is essential as Herbarium records are often perceived as untrustworthy since they have been obtained without any planned sampling scheme, and their geographical locations are often imprecise (Elith et al., 2006; Hortal et al., 2008). Compared with contemporary data, most Herbarium data were collected using outdated methodologies, with associated uncertainties and inaccuracies and might not provide strong inference on species distribution modeling (Golding 2004; Hernandez and Navarro, 2007).

In this context, choosing the most suitable modelling approach becomes crucial: numerous recent studies have successfully applied niche-based models, using presence-only data (e.g. data from herbarium collections; Loiselle et al., 2008) to map habitat in space (Phillips et al., 2006; Pearson et al., 2007). By isolating the niche relationships in occurrence data, presence-only habitat models supply knowledge of species' environmental and spatial distributions with less dependence on observational factors (Elith et al., 2006). In particular Maxent has been found as one of the most effective methods for species distribution modelling, as it showed high quality performance with both small and large sample sizes (Elith et al., 2006; Wisz et al., 2008) and for its capacity to predict novel presence localities for poorly known species (Pearson et al., 2007). The distribution maps provided through Maxent have a probabilistic interpretation, giving a smooth gradation from most to least suitable conditions.

In this work Maxent models output resulted in twelve maps showing the mean probability of occurrence for each characterizing species.

The Maxent internal test of model performance showed AUC values ranging from 0.98 (*Helichrysum stoechas*) to 0.82 (*Clematis flammula*), which indicate that predictive quality of the models can be considered good (Table 1). How close the AUC is to its potential maximum, can ultimately only be assessed if it is known how specialized is the environmental niche that the species occupies, because a wider niche corresponds generally with a lower AUC value (Phillips et al., 2006; Elith et al., 2011). In fact, lower values of AUC are related to *Clematis flammula*, that could be considered a sinantropic species, and to species as *Rubia peregrina*, *Phyllirea latifolia* and *Rhamnus alaternus*, that represent species with a wider ecological niche than the other characterizing species. Those species could be found in different types of Mediterranean habitats, from the thermophilous mixed oak forests to mixed forest dominated by *Quercus ilex* and Mediterranean maquis (Ahrends et al., 2011). Species with wider distributions often encompass a greater range of environmental conditions (Gaston, 2003) and thus may be harder to model, especially when the number of training points is limited (e.g. see Thuiller et al., 2004).

The highest AUC values are related to *Juniperus oxycedrus*, *J. phoenicea* and *Helichrysum stoechas* (that represent a species with a narrow ecological niche). In the study area *Helichrysum stoechas* is mostly found in scrubland and coastal maquis along the Tyrrhenian coast (Arrigoni et al., 1985; De Dominicis, 1993). The two *Juniperus* species, besides being the main species that characterize the target habitat, are found on rocky slopes, deforested areas and coastal sands, mostly on calcareous substrates often with southern exposures (Angiolini et al., 2002; Landi et al., 2007).

The analysis of the contribution to the models of the selected predictors may reinforce the understanding of the ecological properties beyond the characterizing species and, consequently, of the target habitat. In this regard, analyzing the results on the contribution of each environmental variable to the models (Table 3), it is not surprising that the variables that have the greatest contribution to the models are, in most cases, the climatic one and in particular the rainfall and the solar radiation. In fact, the habitat type used as a test in this study presents an ecological niche characterized by xeric, warm, sunny and harsh situations and can be considered semi-arid, occupying unsheltered, steep biotopes (Blasi et al., 2005; Calaciura and Spinelli, 2008).

Concerning the spectral variables (Landsat ETM+ bands), is interesting to note that the highest contribution could be found for the models of species identified as dominant of the top layer of a scrubland (e.g. *Juniperus oxycedrus*, *Juniperus phoenicea*,

Phyllirea latifolia) or species that are frequently associated with the Mediterranean maquis (e.g. *Myrtus communis*, *Pistacia lentiscus*). Instead, the contribution of spectral reflectance with regard to herbaceous species, has generally the lowest values, since these species, in most cases, could not be identified by remote sensor due to the complexity and heterogeneity of habitats in which they are present.

Table 3. Relative contributions of the variables to the MAXENT models (%). For details about variables acronyms, see section Table 2. **Br** – *Brachypodium ramosum*, **Cf** - *Clematis flammula*, **Hs** - *Helichrysum stoechas*, **Jo** - *Juniperus oxycedrus*, **Jp** – *Juniperus phoenicea*, **Li** - *Lonicera implexa*, **Mc** – *Myrtus communis*, **Pla** - *Phillyrea latifolia*, **Ple** - *Pistacia lentiscus*, **Ra** - *Rhamnus alaternus*, **Rp** – *Rubia peregrina*, **Sa** – *Smilax aspera*.

Variables	Br	Cf	Hs	Jo	Jp	Li	Mc	Pla	Ple	Ra	Rp	Sa
Band1	2.2	3.2	5.9	7.4	5.9	2.8	4.4	7.3	12.9	6.1	6.1	6
Band4	2.1	3.1	0.5	8.6	3.8	3.6	0	3.1	1.2	1.8	1.8	2.1
Band7	5.4	6.1	2.8	9.5	8.7	6.1	19.7	18.4	21.4	22	37.1	17.7
ALT	8.1	6.5	10.1	6.2	8.7	7.6	36.2	15.9	17.3	9.5	11.6	12.6
SLO	8.5	9.8	6.9	10.1	3.8	4.1	12.2	10.1	7.7	21.7	3.4	3.4
SR	8.9	14.6	16.7	5.7	8.2	18.5	12.3	10.9	19.1	6.7	5.6	12.5
MWT	7.7	8.1	12.4	12.1	24	11.7	2.4	5.9	10.6	8.8	10.8	13.2
MST	9.8	8.8	15.8	13.9	4.2	16.7	5.4	0.9	5.1	5.8	3.8	8.8
MAR	47.3	39.8	28.9	26.5	32.7	28.9	7.4	27.5	4.7	17.6	19.8	23.7

The performed overlay of the characterizing species models resulted in a raster map representing the probability of presence of the habitat “Arborescent matorral with *Juniperus* spp.” (Fig.3). The map shows the probability of occurrence values ranging from 0 to 0.58 as the result of the overlay between the individual distribution maps of the characterizing species, each representing the point-wise mean of 100 output grids. The habitat distribution map describes currently suitable areas for the target habitat, although that are delineated as suitable may in fact be unoccupied due to factors like human disturbance.

Nevertheless, as demonstrated by the accuracy assessment in a GIS environment, the results seem to be in agreement with the distribution of this habitat in Tuscany. Based on the field data, the KIA results showed a value equal to 0.82 corresponding to a very good strength of agreement according to Landis and Koch (1977) scale. Nonetheless, unless the performed independent validation data, the expert opinion, as part of the validation process, will probably be useful for improving model accuracy (see Araùjo and Guisan, 2006 for a discussion of model evaluation).

Distribution predictions also may be improved by giving greater attention to sampling design, parameterization of the model, model selection and evaluation (Araùjo and Guisan, 2006).

In summary, despite the fact that herbaria and museums contain a vast storehouse of valuable information on biodiversity that offers great potential to advance conservation planning and decision-making (Graham et al., 2004; Soberón and Peterson, 2004), many natural history collections most species are represented by few specimens. Nevertheless, in recent years is increasing the efforts of ecologists to make available georeferenced

and taxonomically verified records in electronic format for large-scale modelling (Loiselle et al., 2008). For this reason, even if in geographical areas where expert knowledge regarding species' distributions is limited because of relatively sparse geographical sampling, species distribution models might also be used by experts to improve their understanding of the geographical and environmental distributions of species.

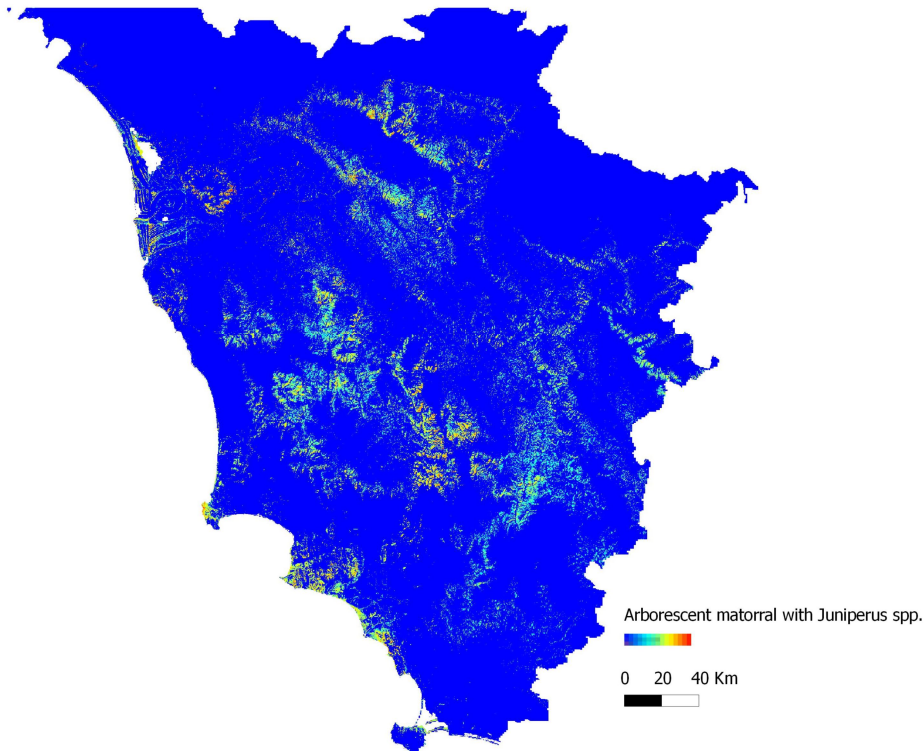


Figure 3. Estimated distribution map of the Natura2000 habitat “Arborescent matorral with *Juniperus* spp.” (5210) for Tuscany.

Conclusions

The goal of this study was to test a Maximum entropy approach to predict habitat distributions in a complex Mediterranean area, given that many applications of habitat distribution modelling rely on existing natural history collections deposited at herbaria or museums.

The modelling approach followed by this study, taking into account the uncertainty proper of the ecosystems (Salski and Sperlbaum, 1991; Foody, 1999), could represent an invaluable tool for conservation management since it can aid in the process of focusing conservation action onto the right geographical locations and in ranking areas in terms of conservation values.

As also evidenced by the results of this paper, there are several advantages of using species distribution modeling to support conservation planning in complex landscapes, especially when a great amount of field data is not available. Maps based only on field occurrence data do not provide information on the likelihood of occurrence in areas that have not been surveyed, and also when surveyed, ‘false absence’ may occur

(MacKenzie, 2005). Moreover range maps from field and similar data are often too coarse or clustered to be useful for on-the-ground conservation action or research (Dormann et al., 2007); then accurate predictive distribution maps should be used with the aim of making field inventories more efficient and effective (Willis et al., 2003).

Our results confirm the important role that distribution models can have in highlighting the areas where a targeted species or habitat type is most likely to be found, and showing where to commit the limited available resources for inventories (Loiselle et al., 2008).

REFERENCES

- [1] Ahrends, A., Rahbek, C., Bulling, M.T., Burgess, N.D., Platts, P.J., Lovett, J.C., Wilkins Kindemba, V., Owen, N., Sallu, A.N., Marshall, A.R., Mhoro, B.E., Fanning, E., Marchant, R. (2011): Conservation and the botanist effect. - *Biological Conservation* 144: 131-140.
- [2] Amici, V. (2011): Dealing with vagueness in complex forest landscapes: A soft classification approach through a niche-based distribution model. - *Ecological informatics* 6: 371-383.
- [3] Angiolini, C., Riccucci, C., De Dominicis, V. (2002): La Flora vascolare della Riserva Naturale Lago di Burano (Grosseto, Toscana meridionale). - *Webbia* 57: 115-152.
- [4] ANPA (2001): La biodiversità nella regione biogeografica Mediterranea. – ANPA, Dipartimento Stato dell’Ambiente, Controlli e Sistemi Informativi, Roma.
- [5] APAT (2005): La realizzazione in Italia del progetto europeo Corine Land Cover 2000. - *Rapporti APAT* 36: 1-86.
- [6] Araújo, M.B., Guisan, A. (2006): Five (or so) challenges for species distribution modelling. - *Journal of Biogeography* 33: 1677-1688.
- [7] Arrigoni, P.V., Nardi, E., Raffaelli, M. (1985): La vegetazione del Parco naturale della Maremma (Toscana). - *Arti Grafiche Giorgi & Gambi, Firenze*.
- [8] Baldwin, R.A. (2009): Use of maximum entropy modeling in wildlife research. - *Entropy* 11: 854-866.
- [9] Blasi, C., Boitani, L., La Posta, S., Manes, F., Marchetti, M. (2005): Stato della Biodiversità in Italia, Contributo alla strategia nazionale per la biodiversità. - Ministero dell’Ambiente e della Tutela del Territorio. Direzione per la protezione della natura, Roma.
- [10] Bossard, M., Feranec, J., Otahel, J. (2000): Technical report no.40. CORINE land cover technical guide-Addendum 2000. - European Environment Agency, Copenhagen.
- [11] Brotons, L., Thuiller, W., Araujo, M.B., Hirzel, A.H. (2004): Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. - *Ecography* 27: 437-448.
- [12] Calaciura, B., Spinelli, O. (2008): Management of Natura 2000 habitats. 5210 Arboreal scrub with *Juniperus* spp. - European Commission, Brussels.
- [13] Cawsey, E.M., Austin, M.P., Baker, B.L. (2002): Regional vegetation mapping in Australia: a case study in the practical use of statistical modeling. - *Biodiversity and Conservation* 11: 2239-2274.
- [14] Chiarucci, A., Foggi, B., Selvi, F. (1999): The *Juniperus oxycedrus* ssp. *oxycedrus* scrub communities of Tuscan serpentine soils. - *Atti della Società Toscana di Scienze Naturali - Memorie serie B* 105: 51-57.
- [15] Cianfrani, C., Le Lay, G., Hirzel, A.H., Loy A. (2010): Do habitat suitability models reliably predict the recovery areas of threatened species? - *Journal of Applied Ecology* 47: 421-430.

- [16] Cohen J. (1960): A coefficient of agreement for nominal scale. - *Educational and Psychological Measurement* 20: 37-46.
- [17] De Dominicis, V., Casini, S., Mariotti, M., Boscagli, A. (1988): La vegetazione di Punta Ala (Prov. di Grosseto). - *Webbia* 42: 101-143.
- [18] De Dominicis, V. (1993): La vegetazione. - In: Giusti, F. (ed.) *La Storia Naturale della Toscana Meridionale*, Pizzi Editore, Cinisello Balsamo.
- [19] Delisle, F., Lavoie, C., Jean, M., Lachance, D. (2003): Reconstructing the spread of invasive plants: taking into account biases associated with herbarium specimens. - *Journal of Biogeography* 30: 1033-1042.
- [20] Dennis, R.L.H., Hardy, P.B. (1999): Targeting squares for survey: predicting species richness and incidence of species for a butterfly atlas. - *Global Ecology and Biogeography* 8: 443-454.
- [21] Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kühn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M., Wilson R. (2007): Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. - *Ecography* 30: 609-628.
- [22] Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.McC., Peterson, A.T., Phillips, S.J., Richardson, K.S., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., Zimmermann, N.E. (2006): Novel methods improve prediction of species' distributions from occurrence data. - *Ecography* 29: 129-151.
- [23] Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J. (2011): A statistical explanation of MaxEnt for ecologists. - *Diversity and Distributions* 17: 43-57.
- [24] Engler, R., Guisan, A., Rechsteiner, L. (2004): An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. - *Journal of Applied Ecology* 41: 263-274.
- [25] Ficetola, G.F., Maiorano, L., Falcucci, A., Dendoncker, N., Boitani, L., Padoa-Schioppa, E., Miaud, C., Thuiller, W. (2010): Knowing the past to predict the future: land-use change and the distribution of invasive Bullfrogs. - *Global Change Biology* 16: 528-537.
- [26] Foggi, B., Cartei, L., Pignotti, L., Signorini, M.A., Viciani, D., Dell'olmo, L., Menicagli, E. (2006a): Il paesaggio vegetale dell'Isola d'Elba (Arcipelago Toscano). *Studio di fitosociologia e cartografico*. - *Fitosociologia* 43: 3-95.
- [27] Foggi, B., Chegia, B., Viciani, D. (2006b): Contributo alla conoscenza della vegetazione del Promontorio di Piombino (Livorno - Toscana). - *Parlatorea* 8: 121-139.
- [28] Foody, G.M. (1999): The continuum of classification fuzziness in thematic mapping. - *Photogrammetric Engineering and Remote Sensing* 65: 443-451.
- [29] Funk, V.A., Richardson, K.S. (2002): *Systematic Data in Biodiversity Studies: Use It or Lose It*. - *Systematic Biology* 51: 303-316.
- [30] Gaston, K.J. (2003): *The structure and dynamics of geographic ranges*. - Oxford University Press, Oxford.
- [31] Golding, J.S. (2004): The use of specimen information influences the outcomes of Red List assessments: the case of southern African plant specimens. - *Biodiversity and Conservation* 13: 773-780.
- [32] Graham, C.H., Ferrier, S., Huettmann, F., Moritz, M., Peterson, A.T. (2004): New developments in museum based informatics and applications in biodiversity analysis. - *Trends in Ecology and Evolution* 19: 497-503.
- [33] Groves, C.R., Jensen, D.B., Valutis, L.L., Redford, K.H., Shaffer, M.L., Scott, J.M., Baumgartner, J.V., Higgins, J.V., Beck, M.W., Anderson, M.G. (2002): *Planning for Biodiversity Conservation: Putting Conservation Science into Practice*. - *BioScience* 52: 499-512.

- [34] Guisan, A., Zimmermann, N.E. (2000): Predictive habitat distribution models in ecology. - *Ecological Modelling* 135: 147-186.
- [35] Hanley, J.A., McNeil, B.J. (1982): The meaning and use of the area under a receiver operating characteristic (ROC) curve. - *Radiology* 143: 29-36.
- [36] Hernandez, P.A., Graham, C.H., Master, L.L., Albert, D.L. (2006): The effect of sample size and species characteristics on performance of different species distribution modeling methods. - *Ecography* 29: 773-785.
- [37] Hernández, H.M., Navarro, M. (2007): A new method to estimate areas of occupancy using herbarium data. - *Biodiversity and Conservation* 16: 2457-2470.
- [38] Hernandez, P.H., Franke, I., Herzog, S.K., Pacheco, V., Paniagua, L., Quintana, H.L., Soto, A., Swenson, J. J., Tovar, C., Valqui, T.H., Vargas, J., Young, B.E. (2008): Predicting species distributions in poorly-studied landscapes. - *Biodiversity and Conservation* 17: 1353-1366.
- [39] Hirzel, A.H., Guisan, A. (2002): Which is the optimal sampling strategy for habitat suitability modelling? - *Ecological Modelling* 157: 331-341.
- [40] Hirzel, A.H., Hausser, J., Chessel, D., Perrin, N. (2002): Ecological-niche factor analysis: how to compute habitat- suitability maps without absence data? - *Ecology* 83: 2027-2036.
- [41] Hirzel, A.H., Le Lay, G., Helfer, V., Randina, C., Guisan, A. (2006): Evaluating the ability of habitat suitability models to predict species presences. - *Ecological modelling* 199: 142-152.
- [42] Hortal, J., Jiménez-Valverde, A., Gómez, J.F., Lobo, JM, Baselga, A. (2008): Historical bias in biodiversity inventories affects the observed realized niche of the species. - *Oikos* 117: 847-858.
- [43] Jiménez-Valverde, A. (2011): Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. - *Global Ecology and Biogeography*, doi: 10.1111/j.1466-8238.2011.00683.x
- [44] Jordano, P. (1991): Gender variation and expression of monoecy in *Juniperus phoenicea* (L.) (Cupressaceae). - *Botanical Gazette* 152: 476-785.
- [45] Kumar, S., Stohlgren, T.J. (2009): Maxent modeling for predicting suitable habitat for threatened and endangered tree *Canacomyrica monticola* in New Caledonia. - *Journal of Ecology and The Natural Environment* 1: 94-98.
- [46] Landi, M., Zoccola, A., Crudele, G., Del Prete, C. (2007): Indagine sulla popolazione e caratterizzazione fitosociologica della vegetazione a *Juniperus phoenicea* L. subsp. *turbinata* (Guss.) Nyman dell 'Isola di Montecristo (Arcipelago Toscano). - *Atti Società Toscana di Scienze Naturali-Memorie Serie B* 114: 115-123.
- [47] Landis, J.R., Koch, G.G. (1977): An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. - *Biometrics* 33: 363-374.
- [48] Lobo, J.M., Tognelli, M.F. (2011): Exploring the effects of quantity and location of pseudo-absences and sampling biases on the performance of distribution models with limited point occurrence data. - *Journal for Nature Conservation* 19: 1-7.
- [49] Loiselle, B. A., Howell, C.A., Graham, C.H., Goerck, J.M., Brooks, T., Smith, K.G., Williams, P.H. (2003): Avoiding pitfalls of using species distribution models in conservation planning. - *Conservation Biology* 17: 1591-1600.
- [50] Loiselle, B.A., Jørgensen, P.M., Consiglio, T., Jiménez, I., Blake, J.G., Lohmann, L.G., Montiel, O.M. (2008): Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? - *Journal of Biogeography* 35: 105-116.
- [51] MacDougall, A.S., Loo, J.A., Clayden, S.R., Goltz, J.G. Hinds, H.R. (1998): Defining conservation priorities for plant taxa in southeastern New Brunswick, Canada using herbarium records. - *Biological Conservation* 86: 325-338.

- [52] MacKenzie, D.I. (2005): What are the issues with presence-absence data for wildlife managers? - *Journal of Wildlife Management* 69: 849-860.
- [53] Margules, C.R., Pressey, R.L. (2000): Systematic conservation planning. - *Nature* 405: 243-253.
- [54] Mondino, G.P., Bernetti, G. (1998): I tipi forestali. Boschi e macchie della Toscana. - Regione Toscana - Servizio Foreste e Patrimonio Agro-forestale regionale, Firenze.
- [55] Ortega-Huerta, M.A., Peterson, A.T. (2008): Modeling ecological niches and predicting geographic distributions: a test of six presence-only methods. - *Revista Mexicana De Biodiversidad* 79: 205-216.
- [56] Papes, M., Gaubert, P. (2007): Modelling ecological niches from low numbers of occurrences: assessment of the conservation status of poorly known viverrids (Mammalia, Carnivora) across two continents. - *Diversity and Distribution* 13: 890-902.
- [57] Pearson, R.G., Thuiller, T., Araújo, M.B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T.P., Lees, D.C. (2006): Model-based uncertainty in species range prediction. - *Journal of Biogeography* 33: 1704-1711.
- [58] Pearson, R.G., Raxworthy, C.J., Nakamura, M., Peterson, A.T. (2007): Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. - *Journal of Biogeography* 34: 102-117.
- [59] Peterson, A.T., Papes, M., Eaton, M. (2007): Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. - *Ecography* 30: 550-560.
- [60] Phillips, S.J., Dudík, M., Schapire, R.E. (2004): A maximum entropy approach to species distribution modeling. - *Proceedings of the 21th International Conference on Machine Learning*, ACM Press, New York.
- [61] Phillips, S.J., Anderson, R.P., Schapire, R.E. (2006): Maximum entropy modeling of species geographic distributions. - *Ecological Modelling* 190: 231-259.
- [62] Phillips, S.J., Dudík, M. (2008): Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. - *Ecography* 31: 161-175.
- [63] Pignatti, S. (1982): *Flora d'Italia*. - Edagricole, Bologna.
- [64] R Development Core Team (2013): *R: A language and environment for statistical computing*. - R Foundation for Statistical Computing, Vienna.
- [65] Rapetti, F., Vittorini, S. (1995): *Carta climatica della Toscana*. - Pacini Editore, Pisa.
- [66] Reutter, B.A., Helfer, V., Hirzel, A.H., Vogel, P. (2003): Modelling habitat-suitability on the base of museum collections: an example with three sympatric *Apodemus* species from the Alps. - *Journal of Biogeography* 30: 581-590.
- [67] Rich, T.C.G., Woodruff, E.R. (1992): Recording bias in botanical surveys. - *Watsonia* 19: 73-95.
- [68] Rodríguez, J.P., Brotons, L., Bustamante, J., Seoane, J. (2007): The application of predictive modelling of species distribution to biodiversity conservation. - *Diversity and Distributions* 13: 243-251.
- [69] Salski, A., Sperlbaum, C. (1991): Fuzzy logic approach to modelling in ecosystem research. - *Lecture Notes in Computer Science* 20: 520-527.
- [70] Segurado, P., Araújo, M.B. (2004): An evaluation of methods for modelling species distributions. - *Journal of Biogeography* 31: 1555-1568.
- [71] Shannon, C.E. (1948): A Mathematical Theory of Communication. - *The Bell System Technical Journal* 27: 379-423.
- [72] Soberón, J., Peterson, A.T. (2004): Biodiversity informatics: managing and applying primary biodiversity data. - *Philosophical Transaction of the Royal Society B* 359: 689-698.
- [73] Stockwell, D.R.B., Peterson, A.T. (2002): Effects of sample size on accuracy of species distribution Models. - *Ecological Modelling* 148: 1-13.

- [74] Thuiller, W. (2003): BIOMOD – optimizing predictions of species distributions and projecting potential future shifts under global change. - *Global Change Biology* 9: 1353-1362.
- [75] Thuiller, W., Brotons, L., Araújo, M.B. Lavorel, S. (2004): Effects of restricting environmental range of data to project current and future species distributions. - *Ecography* 27: 165-172.
- [76] Ungricht, S., Rasplus, J., Kjellberg, F. (2005): Extinction threat evaluation of endemic fig trees of New Caledonia: Priority assessment for taxonomy and conservation with herbarium collections. - *Biodiversity and Conservation* 14: 205-232.
- [77] Viciani D., Gabellini A., Gonnelli V., De Dominicis V. (2005): La vegetazione della Riserva Naturale Monti Rognosi (Arezzo, Toscana) ed i suoi aspetti di interesse botanico-conservazionistico. - *Atti Società Toscana di Scienze Naturali-Memorie Serie B* 111: 27-42.
- [78] Warren, D.L., Seifert, S.N. (2011): Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. - *Ecological Applications* 21: 335-342.
- [79] Willis, F., Moat, J., Paton, A. (2003): Defining a role for herbarium data in Red List assessments: a case study of *Plectranthus* from eastern and southern tropical Africa. - *Biodiversity and Conservation* 12: 1537-1552.
- [80] Wilson, R.J., Thomas, C.D., Fox, R., Roy, D.B., Kunin, W.E. (2004): Spatial patterns in species distributions reveal biodiversity change. - *Nature* 432: 393-396.
- [81] Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., NCEAS Predicting Species Distributions Working Group (2008): Effects of sample size on the performance of species distribution models. - *Diversity and Distribution* 14: 763-773.
- [82] Wollan, A.K., Bakkestuen, V., Kausrud, H., Gulden, G., Halvorsen, R. (2008): Modelling and predicting fungal distribution patterns using herbarium data. - *Journal of Biogeography* 35: 2298-2310.
- [83] Zaniwski, A.E., Lehmann, A., Overton, J.M. (2002): Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. - *Ecological Modelling* 157: 261-280.
- [84] Zweig, M.H., Campbell, G. (1993): Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. - *Clinical Chemistry* 39: 561-77.