Acta Herpetologica 7(2): 221-237, 2012

The significance of using satellite imagery data only in Ecological Niche Modelling of Iberian herps

Neftalí Sillero[1], José C. Brito[2], Santiago Martín-Alfageme[3], Eduardo García-Meléndez[4], A.G. Toxopeus[5], Andrew Skidmore[5]

[1] *Centro de Investigação em Ciências Geo-Espaciais (CICGE), Universidade do Porto, Faculdade de Ciências, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal. Corresponding author.* E-mail: neftali.pablos@fc.up.pt

[2] *CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos da Universidade do Porto, Campus Agrário de Vairão, Rua Padre Armando Quintas, 4485-661 Vairão, Portugal*

[3] *Servicio Transfronterizo de Información Geografica, Universidad de Salamanca, C.M. San Bartolome, Pz. Fray Luis de León 1-8, 37008 Salamanca, Spain*

[4] *Área de Geodinámica Externa, Facultad de Ciencias Biológicas y Ambientales, Universidad de León, Campus de Vegazana s/n, 24071, León, Spain*

[5] *ITC, Faculty of Geo-Information Science and Earth Observation, University of Twente, Hengelosestraat 99, 7514 AE Enschede, The Netherlands*

Submitted on: 2011, 24th October; revised on: 2012, 20th March; accepted on: 2012, 11th April.

Abstract. The environmental data used to calculate ecological niche models (ENM) are obtained mainly from ground-based maps (e.g., climatic interpolated surfaces). These data are often not available for less developed areas, or may be at an inappropriate scale, and thus to obtain this information requires fieldwork. An alternative source of eco-geographical data comes from satellite imagery. Three sets of ENM were calculated exclusively with variables obtained (1) from optical and radar images only and (2) from climatic and altitude maps obtained by ground-based methods. These models were compared to evaluate whether satellite imagery can accurately generate ENM. These comparisons must be made in areas with well-known species distribution and with available satellite imagery and ground-based data. Thus, the study area was the south-western part of Salamanca (Spain), using amphibian and reptiles as species models. Models' discrimination capacity was measured with ROC plots. Models' covariation was measured with a Spatial Spearman correlation. Four modelling techniques were used (Bioclim, Mahalanobis distance, GARP and Maxent). The results of this comparison showed that there were no significant differences between models generated using remotely sensed imagery or ground-based data. However, the models built with satellite imagery data exhibited a larger diversity of values, probably related to the higher spatial resolution of the satellite imagery. Satellite imagery can produce accurate ENM, independently of the modelling technique or the dataset used. Therefore, biogeographical analysis of species distribution in remote areas can be accurately developed only with variables from satellite imagery.

Keywords. Remote Sensing, Geographical Information System, Ecological Niche Modelling (ENM), Reptiles, Amphibians, Spain.

ISSN 1827-9635 (print)
ISSN 1827-9643 (online)

INTRODUCTION

Predictive modelling of species distribution is an important tool for conservation biology and biogeography. For instance, it is used to calculate the potential distribution of species (Bombi et al., 2009; Brito et al., 2009), evaluate the effects of climatic warming on species distribution (Araújo et al., 2006), and the suitability of protected areas (García, 2006; Doko et al., 2011). Ecological niche models (hereafter referred to as ENM) are produced with environmental variables such as climatic, topographical and habitat data. Information about these environmental factors is obtained from a large variety of maps, produced by ground-based methods (climatic data from meteorological stations, altitude from geodesic methods, vegetation maps from analogical aerial photographies) and published in paper or in digital format. We defined ground-based methods as those making field observations, taking in situ measurements and performing land surveying (Kerle et al., 2004). These data are frequently available for developed areas of the world (Martínez Vega, 1996), but may not be available for less developed areas, where particular types of data may not exist. Currently, climate surfaces at 1 km^2 resolution have been produced for almost all the Earth by interpolation of data measured at meteorological stations (Hijmans et al., 2005). The variables obtained in that study were precipitation, as well as minimum and maximum temperature. However, data quality across climate surfaces was not consistent because the number of meteorological stations in some areas is scarce (e.g., in the Sahara and in Siberia, see Fig. 1 in Hijmans et al., 2005). Maximum data uncertainty occurred in mountains, in isolated islands and in poorly sampled areas (Hijmans et al., 2005). Furthermore, using climatic factors alone might not be sufficient for modelling species distributions. For instance, it has been shown that vegetation data clearly improves the explanatory power of bioclimatic models (Thuiller et al., 2004; Seoane et al., 2004). Therefore, in less developed areas, it may be preferable to carry out fieldwork to collect environmental data. This might not be feasible, however, when fieldwork costs are high, the study area is very large, or there is no time for data collection (Muldavin et al., 2001). These constraints hamper the development of ENM.

An alternative source of eco-geographical data comes from Remote Sensing platforms. Variables obtained from satellite imagery that are traditionally used in predictive modelling include vegetation indexes (such as the Normalized Difference Vegetation Index, NDVI), land cover (Kerr and Ostrovsky, 2003), and land surface temperature (Rogers et al., 2002). Other variables, such as ground humidity (Lavers et al., 1996), are used less frequently. Land-cover maps are obtained through a classification process, meanwhile vegetation indexes are obtained through mathematical calculations (Chuvieco, 2000). In comparison to data from ground-based maps, satellite imagery has several advantages, notably: 1) a worldwide coverage, allowing investigators to carry local, regional, national or continental studies; 2) several spatial resolutions, ranging from 1 km^2 in the Advanced Very High Resolution Radiometer (AVHRR) to 1 m^2 in the IKONOS; 3) several temporal resolutions, ranging from 16 days in the case of Landsat 5 TM to every day in the case of AVHRR; 4) a higher accuracy than ground-based methods, because satellite imagery data are not extrapolated as most climate data are; and 5) a lower cost in comparison with studies where it is necessary to obtain data from fieldwork. For example, many variables, such as altitude data from the Shuttle Radar Topography Mission (SRTM: www2.jpl.nasa.

gov/srtm/); NDVI data from the AVHRR sensor (http://edcsns17.cr.usgs.gov/EarthExplorer/); or Landsat imagery (www.glovis.usgs.org), are now publicly available at no cost (Sillero and Tarroso, 2010).

For these reasons, an important application for satellite imagery is the biogeographical analysis of species occurring in remote areas, where environmental data do not exist or are not publicly available, especially at high resolutions (Sader et al., 1991; Martínez Vega, 1996). Remote areas often correspond to biodiversity hotspots where rare and endemic species are frequently the target of conservation planning (Myers et al., 2000). Most of the 25 regions recognized as the most important hotspots of biodiversity in the world (see list in Myers et al., 2000) are located in less developed countries.

Nevertheless, there is a general lack of knowledge concerning the accuracy of ENM calculated with satellite imagery in comparison with other data sources (Thuiller et al., 2004). It is important to evaluate the different environmental data sets available to modellers as well as to advice how to choose them from the available data, but to date this has generally not been well explored in the literature. It has been shown that climatic satellite imagery data can improve ENM accuracy when combined with ground-based maps (Suarez-Seoane et al., 2004). However, there are not consistent results about ENM reliability when using satellite imagery only. For instance, similar ENM were obtained with land-cover data (Venier et al., 2004; using bird species) and with climatic and NDVI data (Zimmermann et al., 2007; using tree species), but not with NDVI data only (Parra et al., 2004; using bird species also). Therefore, other remote sensing data, such as topographical data, should be evaluated and applied to other animal groups (e.g., amphibians, reptiles), where relationships between ground-based maps and satellite imagery variables might be different. Moreover, it is very important to study the suitability and accuracy of satellite imagery as a data source in ENM because in remote areas satellite imagery is frequently the only data source that is available.

The objective of this paper was to analyse how ecological niche models are affected by the type of environmental variables used. Can environmental variables obtained from different methods represent in the same way the ecological requirements of the species? Therefore, will ENMs using different sets of variables identify those relationships between the species and the environment? Specifically, we want to evaluate if ENM developed separately with satellite imagery and ground-based map data can produce similar results. To achieve this, such a comparison must be made in areas where the species distribution is well known and where both satellite imagery and ground-based data are available. The province of Salamanca, in Central Spain, was selected because robust chorological (Sillero et al., 2005) and environmental data (León Llamazares, 1991; Hijmans et al., 2005) are available. Amphibians and reptiles were selected as study species because they have a high correlation with environmental factors (Soares and Brito, 2007; Sillero et al., 2009). Also, to test the robustness of the statistical approaches and imagery used for predicting species models, the comparisons were performed with two ground-based maps and one satellite imagery dataset and four modelling methods. Specifically, the objective was to compare the predictive power and accuracy of four sets of models calculated using satellite imagery data and ground-based map data only, and next to evaluate if satellite imagery variables could be a useful data source to develop reliable ENM.

MATERIAL AND METHODS

Study area

The study area corresponded to the range of a scene of the Landsat 5 Thematic Mapper image, row 203, path 32, of June 23, 1999, on the Salamanca province (Fig. 1). This part of the province of Salamanca, located in Central Western Spain (Fig. 1) was selected because it is an area with good information about the distribution of species as well as the eco-geographical characteristics of the region. Salamanca is a 12500 km^2 plain (12884 raster cells) with an average altitude of 700 to 900 m. To the south, this plain is interrupted by the mountains of the Sistema Central, ranging from 900 m to 2425 m, and to the north-west by the Duero river canyon, ranging from 750 m to 112 m. The climate is Mediterranean, with low precipitation (average of 400 mm throughout the year), and summer drought, and with cool temperatures in winter. These general climate characteristics are modified perceptibly by the relief to the south and north-west. The precipitation in the mountain ranges increases with altitude and the average exceeds 700 mm per year.

Species

Amphibians and reptiles were selected because they are taxonomic groups that can be modelled more accurately than other animal groups, as they exhibit low mobility and high correlation

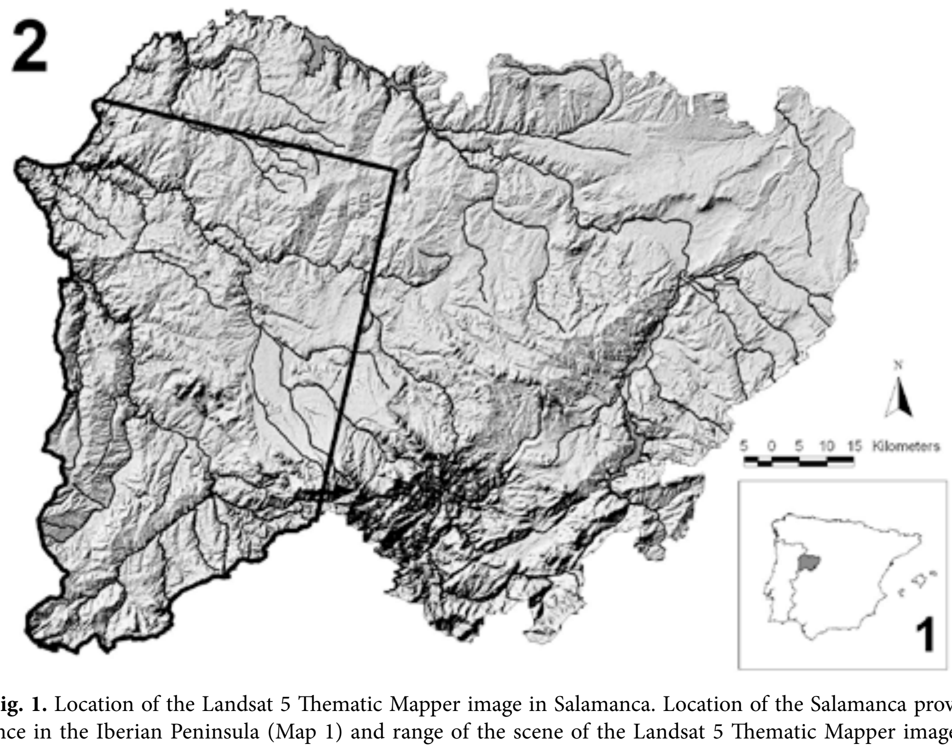

Fig. 1. Location of the Landsat 5 Thematic Mapper image in Salamanca. Location of the Salamanca province in the Iberian Peninsula (Map 1) and range of the scene of the Landsat 5 Thematic Mapper image, row 203, path 32, of June 23, 1999, on the Salamanca province (Map 2). The background is the hillshade model of the SRTM DEM with a spatial resolution of 100 m.

with environmental factors, such as temperature or precipitation (Soares and Brito, 2007; Sillero et al., 2009). Species distribution data were collected from the most recent herpetological atlas of Salamanca (Sillero et al, 2005), where observations were presented at the level of a 1 km x 1 km UTM grid. For modelling purposes, a subset of species was selected according to two criteria: 1) the distribution area was accurately delimited, following opinions from specialists published in herpetological atlases (Pleguezuelos et al., 2002; Sillero et al., 2005); and 2) the extension of their distributions, because the predictive and the explanatory power of ENM might depend on range size. Species with restricted ranges may be modelled with a higher accuracy than widespread species, as it is easier to define their ecological niche (Stockwell and Peterson, 2002; Seoane et al., 2005; Marmion et al., 2009). Using species of different biogeographical affinities (see Sillero et al., 2009), the effect of species types in ENMs' results is reduced. As in other statistical methods, in order to obtain reliable conclusion on method performances, it is necessary to include all the variation in the independent variables. Therefore, four widespread species and four restricted-range species were selected (Sillero et al, 2005). The widespread species were (Fig. 2): three amphibians, *Pleurodeles waltl* (n = 71 locations), *Salamandra salamandra* (n = 87), *Bufo calamita* (n = 129), and one reptile, *Natrix maura* (n = 83). The restricted species were (Fig. 2): one amphibian, *Rana iberica* (n = 41), and three reptiles, *Acanthodactylus erythrurus* (n = 37), *Lacerta schreiberi* (n = 35), *Podarcis carbonelli* (n = 46). In summary, four amphibians and four reptiles were used in the analysis.

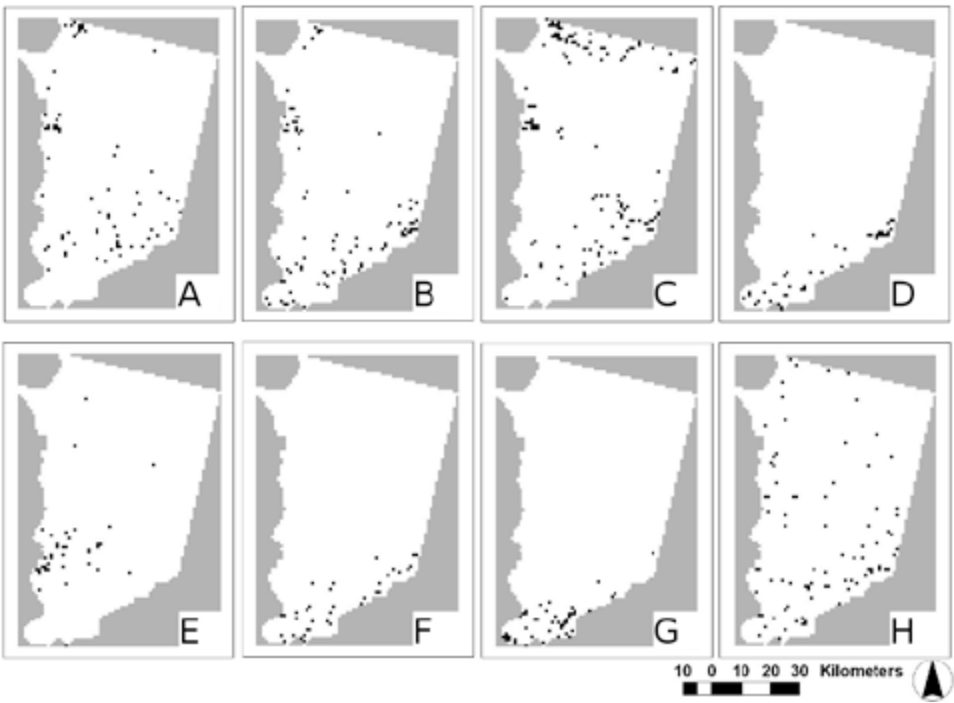

Fig. 2. Amphibian and reptile records for species distribution models in Salamanca province. Amphibians: A, *Pleurodeles waltl*; B, *Salamandra salamandra*; C, *Bufo calamita*; D, *Rana iberica*. Reptiles: E, *Acanthodactylus erythrurus*; F, *Lacerta schreiberi*; G, *Podarcis carbonelli*; H, *Natrix maura*. Each black square represents a record in a 1x1 km UTM square grid.

Environmental data sources

Three data sources were used to calculate ground-based map models: (1) a bioclimatic atlas of Salamanca (León Llamazares, 1991); (2) the WorldClim series (Hijmans et al., 2005; http://www.worldclim.org/); and (3) a Digital Elevation Model (DEM) built with the contour lines of 1:50000 topographic maps by the Junta de Castilla y León (Spain). Vegetation data were not included in ground-based variables as all maps available for the study area were produced by remote sensing techniques (in fact, all vegetation maps are currently performed using satellite imagery). These three data sources were combined in two datasets: GrB-L, with variables from León Llamazares (1991) and the DEM; and GrB-W, with variables from Worldclim series and the DEM. Only variables with a correlation lower than ± 0.8 were included in the ground-based map models: GrB-L dataset included 12 variables, and GrB-W dataset included 7 variables. In the Appendix 1 is described the methodology for obtaining and preparing the ground-based map variables (see Table A1).

Satellite imagery was selected from the Landsat 5 Thematic Mapper (TM) and the second version of the Shuttle Radar Topography Mission (SRTM) adapted to the official Spanish reference system (European Datum 1950 for Spain and Portugal; http://topografia.montes.upm.es/informacion/sig/mde/index.html). These two sensors have a high spatial resolution (30 m and 100 m, respectively) and a high spectral resolution (7 channels: three in the visible spectrum, four in the infrared; and 16 m of vertical precision, respectively) (Chuvieco, 2000; Rabus et al., 2003). Landsat 5 TM classifies vegetation more accurately than other sensors (e.g., Systeme pour l'Observation de la Terre, SPOT), because it has more channels specifically distributed to map vegetation types (Gao, 1999). A total of 14 variables with a correlation lower than ± 0.8 were included in the satellite imagery (SI) models (see Table A2). The methodology for obtaining and preparing the satellite imagery variables is described in the Appendix 1.

In order to analyse the similarity among the variables of the three datasets, a Spatial Principal Component Analysis (SPCA) was performed for each dataset. Then, the first component of each SPCA was used to calculate the correlation among the three datasets. Therefore, a Spatial Spearman correlation (SSC) was used to measure this covariation among datasets with ArcInfo software 9.2 (Band Collection Statistics tool). It was not possible to calculate the SSC directly with the datasets' variables due to the different number and types of variables.

Ecological niche models

The activity patterns of species complicate the accurate determination of reliable absence data in a given locality. Consequently, ground-based map and satellite imagery models were calculated using four modelling techniques that use presence-only occurrence data, thus calculating the ecological realized niche (*sensu* Sillero, 2011): Bioclim (Nix, 1986), Mahalanobis distance (Sokal and Rohlf, 1995), GARP (Stockwell and Noble, 1992), and Maxent (Phillips et al., 2004, 2006). Modelling techniques are explained in the Appendix 1. In other words, four sets of models were executed for each dataset (GrB-L, GrB-W and satellite imagery) and for each of the eight species, resulting in a total of 96 models.

Comparisons of model predictive ability

The models were compared in two ways: (1) all models were tested with receiver operated characteristics (ROC) plots. The area under the curve (AUC) of the ROC plot was taken as a measure of the discrimination capacity of the models (Liu et al., 2005; Elith et al., 2006; VanderWal et al., 2009). Comparisons were performed using an ANOVA test for repeated measures when variables or their logarithmic transformations were normal and homoscedastic; if not, the nonparametric Wilcoxon test was used (Sokal and Rohlf, 1995). (2) A Spatial Spearman correlation (SSC) was used to measure the covariation among models by pairs of datasets with ArcInfo software 9.2 (Band Collection Statistics tool). Models from several datasets can obtain high AUC values (thus models performed well), but they can be spatially dissimilar: species are then predicted in different places. For this reason, it is necessary to measure how spatially similar they are. The AUC values of both species groups (widespread and restricted) were compared using the Mann-Withney U test. Statistical analyses were performed with OpenStat software (http://www.statpages.org/miller/openstat/).

RESULTS

Similarity among datasets

The first pair of each SPCA explained 73.44, 74.46, and 47.12% of the variance of the GrB-L, GrB-W, and SI datasets, respectively. The most similar pair of datasets was GrB-W/SI (SSC = 0.58), followed by the pair GrB-L/SI (SSC = -0.37). Therefore, the less similar pair was GrB-W/GrB-L (SSC = -0.33).

Area under the curve

Examples of ENM for two species are presented in Fig. 3. AUC values were higher than 0.5 in all the sets of models, i.e. results were not achieved by chance alone (Fig. 4). Only in three cases, AUC value was lower than 0.8 (Bioclim GrB-L and SI; and GARP GrB-L). In the GARP and Maxent models, *Natrix maura* was always the species with the lowest AUC value meanwhile *Rana iberica* and *Lacerta schreiberi* were the species with the highest values, in four and two datasets, respectively (Fig. 4). In the Bioclim and Mahalanobis models, this pattern is not evident. Considering each modelling method independently, there were highly significant differences in AUC values among the three datasets (Table 1 and Fig. 4), in the case of Bioclim (Wilcoxon test; $t = 2.366$; $P < 0.01$; $n = 24$; $df = 24$) and GARP (Wilcoxon test; $t = 1.96$; $P = 0.025$; $n = 24$; $df = 7$). AUC values from satellite imagery (SI) dataset (Bioclim, SI mean = 0.836 ± 0.080; GARP, SI mean = 0.869 ± 0.084) were higher than the other two datasets (Bioclim, GrB-L mean = 0.669 ± 0.173; GrB-W mean = 0.785 ± 0.072; GARP, GrB-L mean = 0.796 ± 0.114; GrB-W mean = 0.801 ± 0.057). The differences in AUC values were not significant when comparing the other two modelling methods: Mahalanobis distance (ANOVA for repeated measures: $n = 24$; $df = 20$; $F = 1.50$; $P = 0.262$); and Maxent (Wilcoxon test: $n = 24$; $df = 7$; $t = 1.153$; $P = 0.124$). There was not a clear pattern when comparing both groups of species (widespread vs. restricted); only in four cases (Bioclim GrB-L and GrB-W; Mahalanobis GrB-W; and Maxent SI), AUC values were higher in the restricted species (Table 2 and Fig. 4).

Spatial Spearman Correlations

Spatial Spearman Correlations values were in generally low; only few pairs were higher than 0.5, namely for the Bioclim, GARP and Maxent models (Fig. 5). Here again,

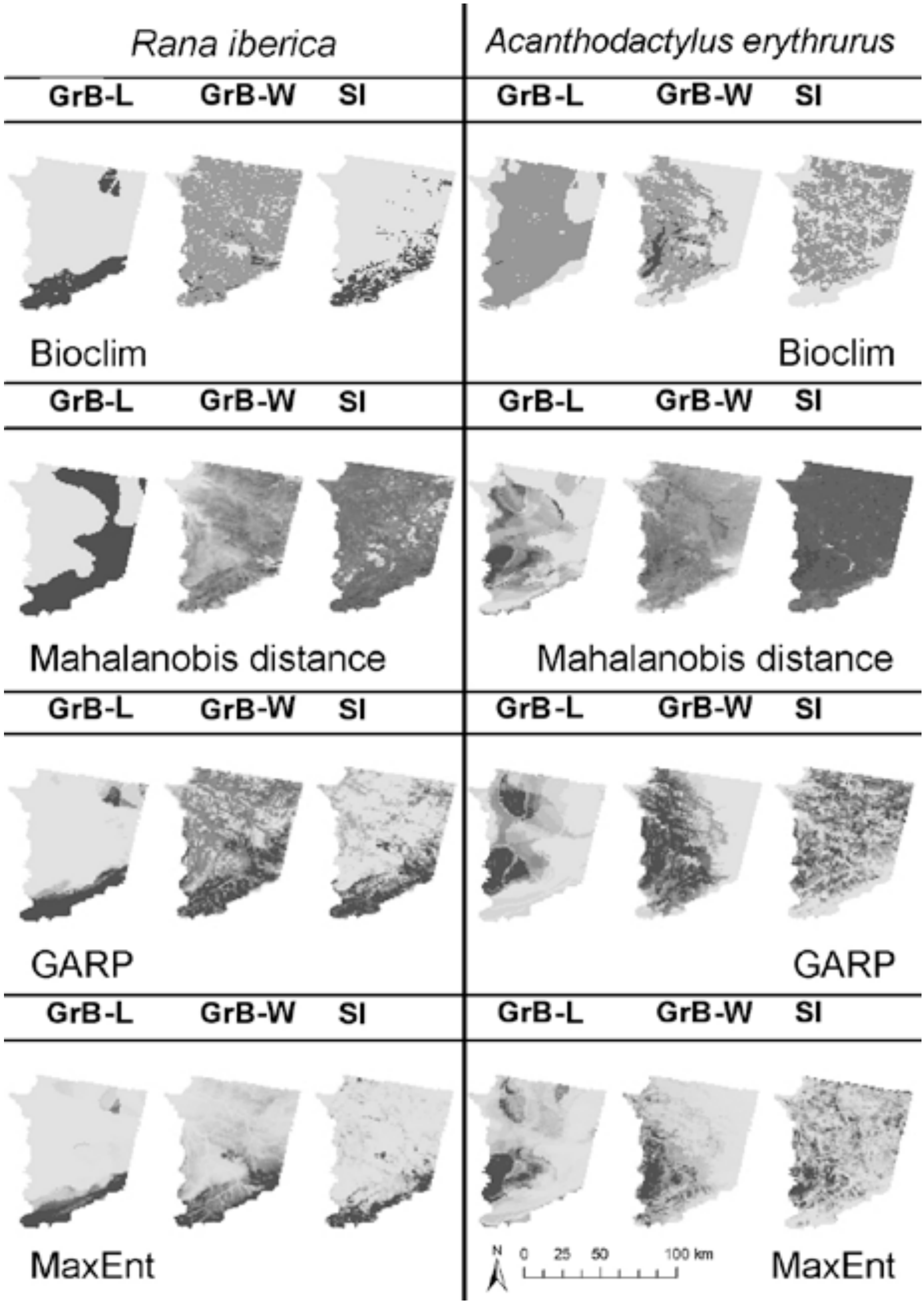

Fig. 3. Examples of species distribution models. The species are *Rana iberica* (first column of maps) and *Acanthodactylus erythrurus* (second column of maps). The first row corresponds to Bioclim models, the second to Mahalanobis models, the third to GARP models and the fourth to Maxent models. Inside each row, maps correspond to GrB-L, GrB-W and satellite imagery dataset, from left to right, respectively. Grey scale stands for habitat suitability values with darker tones corresponding to maximum values.

Table 1. Mean and Standard Deviation (SD) of AUC (Area Under the ROC Curve) values for the four modelling methods. Variables from two datasets of ground-based maps (GrB-L and GrB-W) and one of satellite imagery (SI) were used to calculate the species distribution models. See the text and Appendix 1 for details. *: There was no solution in the Mahalanobis model of *L. schreiberi* for the dataset GrB-L. Therefore, the AUC mean did not include this model.

Method	Dataset	Mean	SD
Bioclim	GrB-L	0.669	0.173
	GrB-W	0.785	0.072
	SI	0.836	0.080
Mahalanobis	GrB-L*	0.986	0.010
	GrB-W	0.995	0.008
	SI	0.985	0.021
GARP	GrB-L	0.796	0.114
	GrB-W	0.801	0.057
	SI	0.860	0.084
Maxent	GrB-L	0.864	0.073
	GrB-W	0.830	0.074
	SI	0.874	0.058

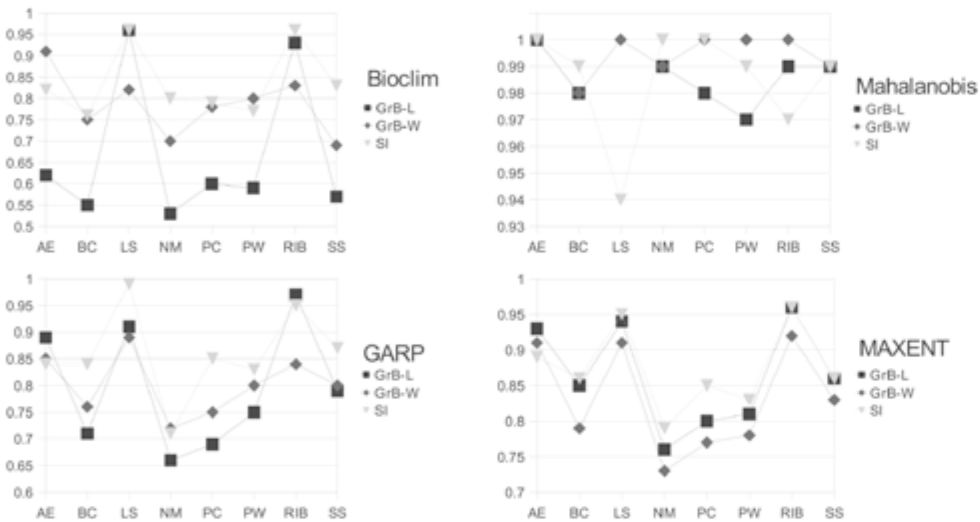

Fig. 4. AUC values of species distribution models. Models were calculated with two datasets of ground-based maps (GrB-L and GrB-W) and one of satellite imagery (SI). AE: *Acanthodactylus erythrurus*; BC, *Bufo calamita*; LS: *Lacerta schreiberi*; NM, *Natrix maura*; PC, *Podarcis carbonelli*; PW: *Pleurodeles waltl*; RIB, *Rana iberica*; SS: *Salamandra salamandra*. There was no solution in the Mahalanobis model of *L. schreiberi* for the dataset GrB-L.

Table 2. Results from Mann-Whitney U test (the first line corresponded to Z statistics and the second line to the P value). AUC (Area Under the ROC Curve) and Spatial Spearman Correlations (SSC) values were compared between two species groups (widespread and restricted) for the four modelling methods. The widespread species were: *Pleurodeles waltl*, *Salamandra salamandra*, *Bufo calamita*, *Natrix maura*; the restricted species were: *Rana iberica*, *Acanthodactylus erythrurus*, *Lacerta schreiberi*, *Podarcis carbonelli*. Variables from two datasets of ground-based maps (GrB-L and GrB-W) and one of satellite imagery (SI) were used to calculate the species distribution models. See the text and Appendix 1 for details. There was no solution in the Mahalanobis model of *L. schreiberi* for the dataset GrB-L. Therefore, AUC and SSC value was not included this model.

Method	AUC			SSC		
	GrB-L	GrB-W	SI	GrB-L/SI	GrB-L/GrB-W	GrB-W/SI
Bioclim	2.31 0.01 (**)	2.02 0.02 (**)	1.44 0.07 (NS)	1.15 0.12 (NS)	0.72 0.24 (NS)	0.29 0.39 (NS)
Mahalanobis	0.35 0.36 (NS)	1.73 0.04 (*)	0.29 0.39 (NS)	1.3 0.1 (NS)	2.12 0.02 (NS)	1.24 0.11 (NS)
GARP	1.44 0.07 (NS)	1.44 0.07 (NS)	1.59 0.06 (NS)	1.44 0.07 (NS)	1.15 0.12 (NS)	1.01 0.16 (NS)
Maxent	1.44 0.07 (NS)	1.44 0.07 (NS)	1.73 0.04 (*)	0.87 0.19 (NS)	1.15 0.12 (NS)	0.58 0.28 (NS)

**: Highly significant; *: Significant; NS: Non significant.

Table 3: Mean and Standard Deviation (SD) of Spatial Spearman Correlations (SSC) for the four modelling methods. Variables from two datasets of ground-based maps (GrB-L and GrB-W) and one of satellite imagery (SI) were used to calculate the species distribution models. See the text and Appendix 1 for details. *: There was no solution in the Mahalanobis model of *L. schreiberi* for the dataset GrB-L. Therefore, SSC were not calculated for the pair combinations with this model.

Method	Dataset	Mean	SD
Bioclim	GrB-L/SI	0.29	0.17
	GrB-W/GrB-L	0.32	0.19
	SI/GrB-W	0.20	0.10
Mahalanobis	GrB-L/SI*	0.08	0.10
	GrB-W/GrB-L*	0.34	0.10
	SI/GrB-W	0.43	0.10
GARP	GrB-L/SI	0.33	0.24
	GrB-W/GrB-L	0.23	0.29
	SI/GrB-W	0.29	0.12
Maxent	GrB-L/SI	0.49	0.20
	GrB-W/GrB-L	0.60	0.11
	SI/GrB-W	0.47	0.16

**: Highly significant; *: Significant; NS: Non significant.

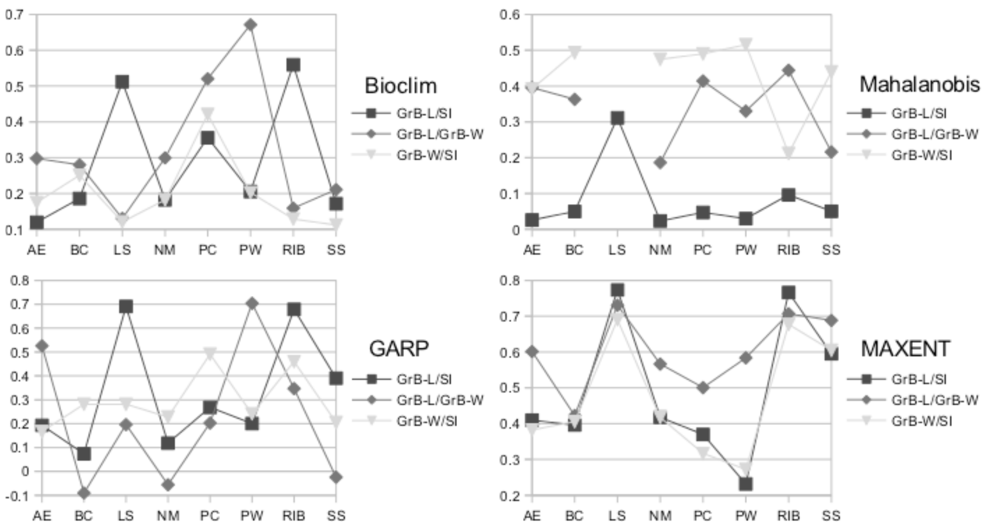

Fig. 5. Spatial Spearman Correlation (SSC) values of species distribution models. SSC were calculated comparing pairs of different dataset models. AE: *Acanthodactylus erythrurus*; BC, *Bufo calamita*; LS: *Lacerta schreiberi*; NM, *Natrix maura*; PC, *Podarcis carbonelli*; PW: *Pleurodeles waltl*; RIB, *Rana iberica*; SS: *Salamandra salamandra*. There was no solution in the Mahalanobis model of *L. schreiberi* for the dataset GrB-L.

Rana iberica and *Lacerta schreiberi* were the species with the highest values when using the GARP and Maxent models; however, these two species had the lowest values using the Bioclim model. In the Mahalanobis models, *Natrix maura* had the lowest values in two comparisons. There was no solution in the Mahalanobis model of *L. schreiberi* for the dataset GrB-L; therefore, this model was not included in the subsequent analysis.

There were significant differences among dataset pairs of comparisons using the Mahalanobis (Wilcoxon test: n = 24; df = 7; t = 2.366; P = 0.009) and Maxent models (Wilcoxon test: n = 24; df = 7; t = 1.820; P = 0.034; Table 3 and Figure 5). SSC were higher when comparing the pair of datasets GrB-L and GrB-W than the other two pairs, in Mahalanobis models, and when comparing the pair of datasets GrB-W and SI, in Maxent models. For the other two sets of models, there were no significant differences among pairs of comparisons: Bioclim (Wilcoxon test: n = 24; df = 7; t = 0.700; P = 0.242; Table 3); GARP (Wilcoxon test: n = 24; df = 7; t = 0.700; P = 0.242; Table 3). Also, there was no significant difference among SSC values of widespread and restricted species groups (Table 2 and Fig. 5).

DISCUSSION

The high accuracies obtained in almost all models based on satellite imagery and ground-based datasets indicated that, in general, all models performed reliably and are robust. The predictive power of the four sets of modelling methods was similar, with the

AUC values and Spatial Spearman correlations (SSC) not being significantly different in almost all cases. When there were significant differences, (Bioclim and GARP models), it was the satellite imagery dataset which obtained higher AUC values. Probably, this result is not due only to the dataset itself, but also may be produced by the modelling technique employed: Bioclim and GARP have a lower accuracy compared with other models (Elith et al., 2006); and the dataset GrB-L is probably the less accurate as it results from data interpolation. Therefore, both methods together with the dataset may produce worst results. The lower accuracy of GrB-L is confirmed by the SSC of SPCA results: SI dataset is more similar to the other two datasets, than those datasets (GrB-L and GrB-W) between them. For the Mahalanobis and Maxent models, the SSC was higher when comparing the pair of datasets GrB-L/GrB-W and GrB-W/SI. As confirmed by the results of the SPCA SSC, dataset SI is a good surrogate of GrB-W. However, the high SSC between the models of the datasets GrB-L and GrB-W is not concordant: the value should be lower as they are the most dissimilar datasets.

However, the P-value for Maxent SSC comparisons was very low. In some models (Bioclim, Mahalanobis, and Maxent; see Table 2), species with a restricted range had high AUC values, while widely dispersed species had lower AUC values, suggesting that species with a restricted range could be more accurately modelled than widely dispersed species. On the other hand, in presence only models, the maximum achievable AUC is lower in widespread species (Phillips et al., 2006). This might further affect the AUC differences between widespread and restricted range species. Nevertheless, SSC values were similar for both groups of species. Although this result had been confirmed clearly in earlier studies (e.g., Stockwell and Peterson, 2002; Seoane et al., 2005; VanDerWal et al., 2009), our results do not allow to assess the same conclusion. Restricted species are modelled with a larger accuracy because their ecological niches have a smaller geographical extension, and therefore they are more easily identified (see VanDerWal et al., 2009). Hence, higher SSC values among restricted species should be expected in comparison with widespread species.

In other studies, ENM were improved when climate data from satellite imagery were combined with habitat variables (Suarez-Seoane et al., 2004). Also, the incorporation of land cover data improved the explanatory power of ENM calculated only with ground-based variables, although it did not improve the predictive power (Thuiller et al., 2004). These results confirmed that satellite imagery data improve model results when added to ground-based data. However, ENM calculated separately or in combinations with regional climate surfaces and from land cover data also had similar accuracy (Venier et al., 2004). Model performance was also similar when NDVI variables were combined with climatic and topographical datasets using ground-based data. Zimmermann et al. (2007) found that ground-based variables outperform the satellite imagery predictors, although models calculated with only satellite imagery data provided high model accuracies, due to the correlation between climate and satellite imagery variables. Our findings support results by Vernier et al. (2004) and Zimmermann et al. (2007), although these studies used only two datasets and one modelling technique. In another study (Parra et al., 2004), models performed relatively well with two ground-based datasets (one with only climatic variables and other with only topographical data) and poorly with NDVI variables (used as unique data source, for deriving several variables). Probably, these partial datasets were not able to define completely the species' ecological niche. Therefore, results could depend on the technique or dataset implemented. In fact, Bioclim was considered as the main factor for general low model performance (Parra et al., 2004). As here, the aim of Parra et al. (2004) was not to test among modelling methods but to test among environmental datasets.

Our results showed that satellite imagery data can produce ENM of acceptable accuracy, independently of the modelling technique or the dataset used. This is very important when the study area is in a remote region, or when spatial resolution of the ground-based layers is coarse, as satellite imagery may be the only feasible source of data (Martínez Vega, 1996; Sillero and Tarroso, 2010). In fact, remote sensing data have been applied for many years in epidemiological studies (Hay et al., 1996), particularly in tropical areas, where it is very difficult to measure environmental variables that require continuous monitoring (Rogers et al., 2002). For example, accurate data on altitude (lower than 1 km^2) is produced currently by radar images (e.g., SRTM and ASTER GDEM; see Sillero and Tarroso, 2010). Moreover, for some countries (e.g., Mauritania) digital topographical maps are not available. A source for altitude data is radar images. Climatic data come from interpolation of weather stations, but in areas without such stations like the Sahara desert, the interpolated results can be very inaccurate (see Hijmans et al., 2005). Using again the example of Mauritania, all vegetation maps have been produced by satellite imagery data. However, works by Parra et al. (2004), Vernier et al. (2004) and Zimmermann et al. (2007) did not compare the satellite imagery models with models calculated exclusively with ground-based data, to measure which type of variables produced more accurate ENM. According to the present study, biogeographical analysis of species distribution in remote areas can be accurately calculated using variables from satellite imagery only. Small differences may be found when modelling other species, namely in groups with low environmental dependency (e.g., birds, mammals). Everything will depend on the capacity of remote sensing data in explaining the habitat species' requirements.

As satellite imagery (e.g., NDVI) observe the actual spatial variation (Han et al., 2004), satellite imagery can describe the environment more accurately than ground-based data sources. Satellite imagery is produced with higher spatial and spectral resolutions than conventional field methods used to produce ground-based data (Kerr and Ostrovsky, 2003; Turner et al., 2003). This supposes a larger diversity of values in the variables obtained from the satellite imagery than those derived from the ground-based data. This is proved by the low proportion of explained variance presented by the first component of the PCA analysis of the SI dataset. Whereas a climatic map obtained by interpolation shows only few values inside a particular area, satellite imagery may capture subtle differences in either the soil or the vegetation characteristics. For instance, an artificial grass field is not distinguished from a natural one by conventional aerial photography, but the infra-red image is able to detect the chlorophylical activity of the natural grass field (Chuvieco, 2000). When using satellite imagery, two adjacent grass fields, one artificial and the other natural, therefore would be identified as two different structures. In an ENM, the habitat suitability index of the two fields probably would be the same if satellite imagery was not used. This is also applicable to adjacent micro-habitats. Furthermore, climate variables alone cannot distinguish possible reflectance differences between early and late plant species. Therefore, in some situations, areas with different habitat suitability indexes can be detected more accurately with satellite imagery, resulting in more accurate ENM (Suarez-Seoane et al., 2004). An example of this could be an ENM of deciduous broadleaved trees, as multi-temporal imagery can distinct phenology compared to evergreen needle-leaf trees (Zimmermann et al., 2007).

Satellite imagery has also disadvantages or short comings. The temporal record of satellite imagery is recent and short (Chuvieco, 2000). Thus, our capacity to perform historical studies with satellite imagery is lower than with interpolated-climate data (Hijmans et al., 2005). Moreover, one of the most important programmes of acquiring satellite imagery (Landsat) is not completely functional any more (both Landsat 5 and 7 are not working properly). Other disadvantage of satellite imagery is the relatively low number of environmental variables that sensors can provide (Sillero and Tarroso, 2010). Manual photo-interpretation can obtain more exact products than automatic image classification algorithms (although there are examples proving the opposite, such as the first edition of CORINE project for Spain; pers. obs.).

Despite these disadvantages, an important advantage of satellite imagery data over ground-based data is its capacity to produce temporal series of information from the same area (Hay et al., 1998; França and Setzer, 1998; Azzali and Menenti, 2000; Lucas et al., 2000a; Lucas et al., 2000b; Sobrino, 2000; Green and Hay, 2002). Therefore, it is possible to study the same ecological processes over time, such as habitat monitoring (Kobayashi and Dye, 2005), land cover mapping (Stow et al., 2006) and climate change detection (Roerink et al., 2003). Future work should include more complete temporal series of climatological data, land cover and vegetation productivity.

ACKNOWLEDGEMENTS

We would like to Prof. J. Martins for checking the English. NS has a post-doctoral grant (SFRH/BPD/26666/2006) from Fundação para a Ciência e Tecnologia (FCT, Portugal) and JCB has a contract (Programme Ciência 2007) from FCT.

SUPPLEMENTARY MATERIAL

Supplementary material associated with this article can be found at: < http://www.unipv.it/webshi/appendix > Manuscript number 9891: Appendix 1.

REFERENCES

Araujo, M.B., Thuiller, W., Pearson, R.G. (2006): Climate warming and the decline of amphibians and reptiles in Europe. J. Biogeogr. **33**: 1712-1728.

Azzali, S., Menenti, M. (2000): Mapping vegetation-soil-climate complexes in southern Africa using temporal Fourier analysis of NOAA-AVHRR NDVI data. Int. J. Remote Sensing **21**: 973-996.

Bombi, P., Luiselli, L., Capula, M., Salvi, D. (2009): Predicting elusiveness: potential distribution model of the Southern smooth snake, *Coronella girondica*, in Italy. Acta Herpetol. **4:** 7-13.

Brito, J.C., Acosta, A.L., Álvares, F., Cuzin, F. (2009): Biogeography and conservation of taxa from remote regions: An application of ecological-niche based models and GIS to North-African canids. Biol. Conserv. **142**: 3020-3029.

Chuvieco, E. (2000): Fundamentos de Teledetección Espacial, 3ª ed. revisada. Ed. Rialp, Madrid.

Doko, T, Fukui, H., Kooiman, A., Toxopeus, A.G., Ichinose, T., Chen, W., Skidmore, A.K. (2011): Identifying habitat patches and potential ecological corridors for remnant Asiatic black bear (*Ursus thibetanus japonicus*) populations in Japan. Ecol. Model. **222**: 748-761.

Elith, J., Graham, C., Anderson, R., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R., Huettmann, F., Leathwick, J., Lehmann, A., Li, J., Lohmann, L., Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R., Soberón, J., Williams, S., Wisz, M., Zimmermann, N.E. (2006): Novel methods improve prediction of species' distributions from occurrence data. Ecography **29**: 129-151.

França, H., Setzer, A.W. (1998): AVHRR temporal analysis of a savanna site in Brazil. Int. J. Remote Sens. **19**: 3127-3140.

Gao, J. (1999): A comparative study on spatial and spectral resolutions of satellite data in mapping mangrove forest. Int. J. Remote Sens. **20**: 2823-2833.

Garcia, A. (2006): Using ecological niche modelling to identify diversity hotspots for the herpetofauna of Pacific lowlands and adjacent interior valleys of Mexico. Biol. Conserv. **130**: 25-46.

Green, R.M., Hay, S.I. (2002): The potential of Pathfinder AVHRR data for providing surrogate climatic variables across Africa and Europe for epidemiological applications. Remote Sens. Environ. **79**: 166-175.

Han, K.S., Champeaux, J.L., Roujean, J.L. (2004): A land cover classification product over France at 1 km resolution using SPOT4/VEGETATION data. Remote Sens. Environ. **92**: 52-66.

Hay, S.I., Snow, RW., Rogers, D.J. (1998): Predicting malaria seasons in Kenya using multitemporal meteorological satellite sensor data. Trans. R. Soc. Trop. Med. Hyg. **92**: 12-20.

Hay, S.I., Tucker, C., Rogers, D.J., Packer, M. (1996): Remotely sensed surrogates of meteorological data for the study of the distribution and abundance of arthropod vector of disease. Ann. Trop. Med. Parasitol. **90**: 1-19.

Hijmans, R.J., Cameron, E., Parra, J.L., Jones, P.G., Jarvis, A. (2005): Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. **25**: 1965-1978.

Kerle, N. Janssen, L.L.F., Huurneman, G.C. (Eds) (2004): Principles of remote sensing : an introductory textbook. ITC, Enschede.

Kerr, J.T., Ostrovsky, M. (2003): From space to species: ecological applications for remote sensing. Trends Ecol. Evol. **18**: 299-305.

Kobayashi, H., Dye, D.G. (2005): Atmospheric conditions for monitoring the long-term vegetation dynamics in the Amazon using normalized difference vegetation index. Remote Sens. Environ. **97**: 519-525.

Lavers, C.P., Haynes-Young, R.H., Avery, M.I. (1996): The habitat associations of dunlin (*Calidris alpina*) in the Flow Country of nourthern Scotland and an improved model for predicting habitat quality. J. Appl. Ecol. **33**: 279-290.

León Llamazares, A. (1991): Caracterización agrícola de la provincia de Salamanca, 2ª Ed. Ministerio de Agricultura, Pesca y Alimentación, Madrid.

Liu, C., Berry, P.M., Dawson, T.P., Pearson, R.G. (2005): Selecting thresholds of occurrence in the prediction of species distributions. Ecography **28**: 385-393.

Lucas, R.M., Honzák, M., Curran, P.J., Foody, G.M., Milnes, R., Brown, T., Amaral, S. (2000): Mapping the regional extent of tropical forest regeneration stages in the Brazilian Legal Amazon using NOAA AVHRR data. Int. J. Remote Sens. **21**: 2855-2881.

Lucas, R.M., Honzák, M., Curran, P.J., Foody, G.M., Nguele, D.T. (2000): Characterizing tropical forest regeneration in Cameroon using NOAA AVHRR data. Int. J. Remote Sens. **21**: 2831-2854.

Marmion, M., Luoto, M., Heikkinen, R.K., Thuiller, W. (2009): The performance of state-of-the-art modelling techniques depends on geographical distribution of species. Ecol. Model. **220**: 3512-3520.

Martínez Vega, J.J. (1996): Una revisión sobre las imágenes espaciales como fuentes cartográficas. Revista de Teledetección **6**: 1-13.

Muldavin, E.H., Neville, P., Harper, G. (2001): Indices of Grassland Biodiversity in the Chihuahuan Desert Ecoregion Derived from Remote Sensing. Conserv. Biol. **15**: 844-855.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Kent, J. (2000): Biodiversity hotspots for conservation priorities. Nature **403**: 853-858.

Nix, H.A. (1986): A biogeographic analysis of Australian Elapid Snakes. In: Atlas of Elapid Snakes of Australia, Australian Flora and Fauna Series Number 7, pp. 4-15. R. Longmore Ed, Australian Government Publishing Service, Canberra.

Parra, J.L., Graham, C.C., Freile, J.F. (2004): Evaluating alternative data sets for ecological niche models of birds in the Andes. Ecography **27**: 350-360.

Phillips, S.J., Anderson, R.P., Schapire, R.E. (2006): Maximum entropy modeling of species geographic distributions. Ecol. Model. **190**: 231-259.

Phillips, S.J., Dudík, M., Schapire, R.E. (2004): A maximum entropy approach to species distribution modeling. Proc. 21 Int. Conf. Machine Learning: 655-662.

Pleguezuelos, J.M., Márquez, R., Lizana, M. (2002): Atlas de distribución y Libro Rojo de los Anfibios y Reptiles de España. Dirección General de Conservación de la Naturaleza-Asociación Herpetológica Española, 2ª impresión, Madrid.

Rabus, B., Eineder, M., Roth, A., Bamler, R. (2003): The Shuttle Radar Topography Mission-a new class of digital elevation models acquired by spaceborne radar. ISPRS J. Photogramm. Remote Sens. **57**: 241-262.

Roerink, G.J., Menenti, M., Soepboer, W., Su, Z. (2003): Assessment of climate impact on vegetation dynamics by using remote sensing. Phys. Chem. Earth, Parts A/B/C **28**: 103-109.

Rogers, D.J., Randolph, S.E., Snow, R.W., Hay, S.I. (2002): Satellite imagery in the study and forecast of malaria. Nature **415**: 710-715.

Sader, S.A., Powell, G.V.N., Rappole, J.H. (1991): Migratory bird habitat monitoring through Remote Sensing. Int. J. Remote Sens. **12**: 363-372.

Seoane, J., Bustamante, J., Diaz-Delgado, R. (2004): Are existing vegetation maps adequate to predict bird distributions? Ecol. Model. **175**: 137-149.

Seoane, J., Carrascal, L.M., Alonso, C.L., Palomino, D. (2005): Species-specific traits associated to prediction errors in bird habitat suitability modelling. Ecol. Model. **185**: 299-308.

Sillero, N. (2011): What does ecological modelling model? A proposed classification of ecological niche models based on their underlying methods. Ecol. Model. **222**: 1343-1346.

Sillero, N., Brito, J.C., Toxopeus, B., Skidmore, A.K. (2009): Biogeographical patterns derived from remote sensing variables: the amphibians and reptiles of the Iberian Peninsula. Ampibia-Reptilia **30**: 185-206.

Sillero, N., Celaya, L., Martín-Alfageme, S. (2005): Using GIS to Make An Atlas: A Proposal to Collect, Store, Map and Analyse Chorological Data for Herpetofauna. Rev. Esp. Herpetol. **19**: 87-101.

Sillero, N., Tarroso, P. (2010): Free GIS for herpetologists: free data sources on Internet and comparison analysis of proprietary and free/open source software. Acta Herpetol. **5**: 63-85.

Soares, C., Brito, J.C. (2007): Environmental correlates for species richness and biogeographic relationships among amphibians and reptiles in a climate transition area. Biodivers. Conserv. **16**: 1087-1102.

Sobrino, J.A. (2000): Teledetección. Valencia: Servicio de Publicaciones, Universidad de Valencia.

Sokal, R.R., Rohlf, F.J. (1995): Biometry, 3rd Ed. Freeman, New York.

Stockwell, D.R.B., Noble, I.R. (1992): Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. Math. Comp. Simul. **33**: 385-390.

Stockwell, D.R.B., Peterson, A.T. (2002): Effects of sample size on accuracy of species distribution models. Ecol. Model. **148**: 1-13.

Stow, D.A., Hope, A., McGuire, D., Verbyla, D., Gamon, J., Huemmrich, F., Houston, S., Racine, C., Sturm, M., Tape, K., Hinzman, L., Yoshikawa, K., Tweedie, C., Noyle, B., Silapaswan, C., Douglas, D., Griffith, B., Jia, G., Epstein, H., Walker, D., Daeschner, S., Petersen, A., Zhou, L., Myneni, R. (2004): Remote sensing of vegetation and land-cover change in Arctic Tundra Ecosystems. Remote Sens. Environ. **89**: 281-308.

Suarez-Seoane, S., Osborne, P.E., Rosema, A. (2004): Can climate data from METEOSAT improve wildlife distribution models? Ecography **27**: 629-636.

Thuiller, W., Araujo, M.B., Lavorel, S. (2004): Do we need land-cover data to model species distributions in Europe? J. Biogeogr. **31**: 353-361.

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., Steininger, M. (2003): Remote sensing for biodiversity science and conservation. Trends Ecol. Evol. **18**: 306-314.

VanDerWal, J., Shoo, L.P., Graham, C., Williams, S.E. (2009): Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? Ecol. Model. **220**: 589-594.

Venier, L.A., Pearce, J., McKee, J.E., McKenney, D.W., Niemi, G.J. (2004): Climate and satellite-derived land cover for predicting breeding bird distribution in the Great Lakes Basin. J. Biogeogr. **31**: 315-331.

Zimmermann, N.E., Edwards, T.C., Moisen, G.G., Frescino, T.S., Blackard, J.A. (2007): Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. J. Appl. Ecol. **44**: 1057-1067.