

A maximum entropy model for predicting wild boar distribution in Spain

Jaime Bosch^{1*}, Fernando Mardones², Andrés Pérez^{2,3}, Ana de la Torre¹
and María Jesús Muñoz¹

¹ Animal Health Research Center (CISA-INIA). Epidemiology and Environmental Health Department. Ctra. Algete-El Casar, s/n. 28130 Valdeolmos, Madrid, Spain. ² Center for Animal Disease Modeling and Surveillance (CADMS). Department of Medicine and Epidemiology. School of Veterinary Medicine. University of California Davis. One Shields Avenue, 1044 Haring Hall, Davis, CA 95616, USA.

³ National Scientific and Technical Research Council (CONICET). Rivadavia 1917 (C1033AAJ). Autonomous City of Buenos Aires, Argentine Republic

Abstract

Wild boar (*Sus scrofa*) populations in many areas of the Palearctic including the Iberian Peninsula have grown continuously over the last century. This increase has led to numerous different types of conflicts due to the damage these mammals can cause to agriculture, the problems they create in the conservation of natural areas, and the threat they pose to animal health. In the context of both wildlife management and the design of health programs for disease control, it is essential to know how wild boar are distributed on a large spatial scale. Given that the quantifying of the distribution of wild species using census techniques is virtually impossible in the case of large-scale studies, modeling techniques have thus to be used instead to estimate animals' distributions, densities, and abundances. In this study, the potential distribution of wild boar in Spain was predicted by integrating data of presence and environmental variables into a MaxEnt approach. We built and tested models using 100 bootstrapped replicates. For each replicate or simulation, presence data was divided into two subsets that were used for model fitting (60% of the data) and cross-validation (40% of the data). The final model was found to be accurate with an area under the receiver operating characteristic curve (AUC) value of 0.79. Six explanatory variables for predicting wild boar distribution were identified on the basis of the percentage of their contribution to the model. The model exhibited a high degree of predictive accuracy, which has been confirmed by its agreement with satellite images and field surveys.

Additional key words: *Sus scrofa*; environmental suitability; MaxEnt; spatial distribution; wildlife management; geographic information.

Introduction

Wild boar have become one of the most widely spread ungulates in the Iberian Peninsula (Vitorino &

Fonseca, 2004; Rosell & Herrero, 2007), in Europe (Massei & Genov, 2000; Keuling *et al.*, 2013), and in the world (Long, 2003; Oliver & Leus, 2008). Their spread has been linked to their biological traits that

* Corresponding author: jaime.bosch@inia.es
Received: 07-02-14. Accepted: 25-09-14.

This work has one supplementary table and three supplementary figures that do not appear in the printed article but that accompany the paper online.

Abbreviations used: AUC (area under the receiver operating characteristic curve); GAM (generalised additive models); GBIF (global biodiversity information facility); GIMMS (global inventory modeling and mapping studies); GLM (generalized linear model); GRASS (geographic resources analysis support system); MaxEnt (maximum entropy); NDVI (normalized difference vegetation index); ROC (receiver operating characteristic); SDM (species distribution model); SRTM (shuttle radar topography mission); VCF (vegetation continuous fields); VIF (variance inflated factor).

include their highly varied trophic spectrum (Herrero *et al.*, 2006), great adaptability to variable food resources and different ecological conditions (Abaigar, 1993; Herrero *et al.*, 2005), high reproductive rate (Taylor *et al.*, 1998; Rosell *et al.*, 2001), and, finally, an ability to adapt their spatio-temporal behavior to local conditions (Podgórski *et al.*, 2013). In previous decades, the remarkable increase in the number of wild boar was directly proportional to the progressive abandonment of rural areas, which provided wild boar with more areas in which to shelter – mostly scrub and wooded areas – and more trophic resources (Tellería & Sáez-Royuela, 1985; Sáez-Royuela & Tellería, 1986; Herrero *et al.*, 2005). In some countries such as Spain, the great scarcity of predators (Massei & Genov, 2000) that could naturally control wild boar populations also favored this species' expansion.

In many countries, wild boar are widely hunted and constitute an important economic resource. In Spain, it is estimated that 176,245 wild boar were killed in the 2006-2007 season (Bosch *et al.*, 2012). In some cases, the profitability of hunting has encouraged certain practices such as the use of artificial feeders and the legal or illegal relocation of individuals that have increased the number of boar (Wood & Barret, 1979; Spencer & Hamton, 2005). The negative effects of increases in wild boar populations include damage to crops (Herrero *et al.*, 2006; Schley *et al.*, 2008), traffic accidents (Rosell *et al.*, 2001; Peris *et al.*, 2005; Colino-Rabanal *et al.*, 2012), and the transmission of diseases since wild boar act as a reservoir for livestock, wildlife, and human diseases such as brucellosis, tuberculosis, salmonellosis, Aujeszky's disease, and classical and African swine fever. Some of these diseases can cause direct or indirect economic losses – mortality and poorer weight gain in livestock – and oblige the implementation of disease prevention, control, and eradication programs. A prerequisite for designing and implementing effective control programs is knowledge of the spatial distribution of the target species. Biodiversity models that consider species distribution, density, and abundance are of great importance for designing and implementing effective species management.

In countries such as Spain, extensive pig rearing is a very important economic activity. The resources offered by the vegetation – both food and shelter – are often shared by freeranging pigs and wild boar, which thus creates hotspot contact points and increases the risk of disease transmission. Moreover, stretches of

vegetation that cross national borders can act as corridors for wild boar and increase the risks of a transboundary spread of disease.

Climate is a key factor in explaining the species distribution in the world (Von Humboldt & Bonpland, 1807; De Candolle, 1855). Peninsular Spain (Canary and Balearic Islands, Ceuta and Melilla not included) is situated between latitudes 35° and 45° N and, due to its geographical position in the southern Palearctic, lies in a transition zone between contrasting climatic regions. This privileged location in the extreme southwest of Europe has meant that for millennia human influence has transformed the landscape and created a variety of unique semi-natural agroforestry systems. Spain is an area of highly heterogeneous topography, complex geomorphology, and remarkable geographical and lithological partitioning. It contains three biogeographic regions, Mediterranean, Atlantic and Alpine. The Mediterranean bioclimatic region is influenced by two floristic worlds, the Holarctic and the Palearctic, whose effects combine as they interact mutually (García *et al.*, 2002). Here, thermo-, meso-, and supra-Mediterranean levels predominate, while in the Atlantic bioclimatic region, thermo-, meso-, and orotemperate are the most common climatic levels and in Alpine region is criotemperate.

Climate in combination with other environmental factors are the main elements that determine vegetation patterns (Woodward, 1987; Ellenberg, 1988). Vegetation cover will influence the distribution of an animal species more than any other factor since it determines the land's ability to supply food and/or shelter for animals. Therefore, vegetation cover is a limiting factor for the spread of a species (Herrero *et al.*, 2006).

In the Iberian ecosystems, five climatic factors are responsible for modeling vegetation landscapes (Martí & del Moral, 2003): (1) the north-south variation between temperatures, (2) continentality, (3) the variation between the basic Mediterranean substrates and the western acidic Atlantic substrates, (4) altitude, and (5) anthropic influences. Overall, Spain is a mosaic of living, functioning agroforestry systems which possess a greater genetic diversity of flora and fauna (De Miguel, 2002) than more northerly regions (Papanastasis *et al.*, 2009; Pardini, 2009).

Biogeographical variation in wild boar density in western Eurasia has been evaluated by Melis *et al.* (2006) while Oliver & Leus (2008) have assessed this species' distribution in the Euroasiatic zone. In the case of the Iberian Peninsula, Bosch *et al.* (2012) have

recently created a habitat suitability map for wild boar based on the availability of vegetation resources, *i.e.*, food resources and shelter. These authors used different studies to assess the risk of introducing disease along vegetation corridors crossing national borders (De la Torre *et al.*, 2013). Furthermore, these studies also have been applied in Spanish epidemiological surveillance programs for certain diseases (Ministerio de Agricultura, Alimentación y Medio Ambiente, Spain, 2013). Finally, suitability maps are also very useful for identifying wildlife livestock interfaces (Hull *et al.*, 2014) and defining potential hotspots (De la Torre *et al.*, unpublished data).

In recent years, significant advances have been made in the statistical tools and techniques used to generate species distribution models (SDMs) (Guisan & Zimmermann, 2000; Guisan & Thuiller, 2005; Elith & Leathwick, 2009). SDMs predict species occurrence using mathematical models based on field data and environmental variables (Phillips *et al.*, 2006), which can indicate the suitability of habitats for developing populations of a particular species or community (Ferrier, 2002). Statistical methods employed for formulating SDMs include those that require presence/ absence data, as well as those such as the maximum entropy model (MaxEnt) that are based only on presence data (Phillips *et al.*, 2006; Phillips & Dudík, 2008). The MaxEnt (Phillips *et al.*, 2006) method has proven to be well suited to a wide range of presence-only datasets, most notably datasets with 11-13 environmental variables and >100 occurrences (Hernández *et al.*, 2006; Phillips & Dudík, 2008; Baldwin, 2009). This method applies the principle of maximum entropy to calculate the most likely geographical distribution for a species. It works in a similar – but not identical – way to generalized linear models (GLM) and general additive models (GAM) but with the difference that the equation is adjusted using an artificial intelligence method that assumes no predetermined pair-distribution data (Phillips *et al.*, 2006). MaxEnt employs a regularization function that prevents prediction caused by overfitting the data (Phillips *et al.*, 2006; Phillips, 2008). It estimates the probability of species occurrence by searching for the maximum entropy distribution (closest to uniform) that is subject to the constraint that the expected value of each environmental variable under this estimated distribution matches its empirical average (average values for the set occurrence data). This model expresses the value of habitat suitability for the species

as a function of environmental variables. A high value for the distribution function in a particular grid cell indicates that it has very favorable conditions for the presence of the species. Recent publications have demonstrated mathematically that MaxEnt is essentially equivalent to a non-homogeneous Poisson process and weighted logistic regression model with a background of properly weighted points (Fithian & Hastie, 2013). MaxEnt prevents overfitting better than the variable-selection methods such as generalized additive and generalized linear models that are commonly used for regression-based models (Phillips & Dudík, 2008). Unlike discriminative regression-based methods, MaxEnt is a generative approach that models species distribution directly. Previous studies have indicated that generative methods give better predictions than discriminative methods (Phillips & Dudík, 2008). In addition, some authors have argued that the MaxEnt model approach performs better than other presence-based algorithms (Elith *et al.*, 2006; Benito de Pando & Peñas de Giles, 2007; Elith & Leathwick, 2009; Mateo *et al.*, 2010) and usually guarantees accurate predictions of species' distribution (Elith *et al.*, 2006; Tsoar *et al.*, 2007). Besides MaxEnt employs a regularization function that prevents prediction from over-fitting the data (Phillips *et al.*, 2006; Phillips, 2008). Absence records are not as widely available in Spain as in many other regions and so the MaxEnt model represents a good approach for calculating the potential distribution of wild boar using the most important environmental variables that act as predictors of distribution and explain the occurrence of wild boar. As many authors have previously suggested, the first strategy for reducing the inconsistencies between different species-distribution models is to conduct thorough model comparison evaluations and adopt the most promising techniques for modeling (Elith *et al.*, 2006; Lawler *et al.*, 2006; Prasad *et al.*, 2006). The second strategy is to apply consensus methods (Laplace, 1820; Thuiller, 2004; Araújo & New, 2007; Marmion *et al.*, 2009).

The aims of the present study were to predict the potential distribution of wild boar in Spain and to identify the environmental variables that influence it by integrating animal presence and environmental data into a MaxEnt approach. This use of MaxEnt is intended as a starting point that will allow comparison with other models, as well as its future implementation in a consensus model that will increase the robustness of the prediction.

Materials and methods

Study area

In this study, the potential distribution of habitat suitability for wild boar was restricted to peninsular Spain (total area: 493,519.54 km²). Located in the southwestern Palearctic, Spain has a mean altitude of ~660 m a.s.l. (SD 1041.34) and a maximum height of 3,479 m. The Pyrenees act as a natural barrier that isolates Spain from the rest of northern Europe. The nature of the Iberian Peninsular – *e.g.*, its geographical position and topographical configuration – ensures that the typical environmental variables associated with wild boar distribution in Spain differ from those in many other areas of its world distribution.

Data source: Wild boar occurrence

The spatial distribution of *Sus scrofa* occurrence data (latitude and longitude) was obtained largely from the data portal of the Global Biodiversity Information Facility (GBIF: the world's largest online depository of records and provides access to specimen data from databases of biological surveys and collections from throughout the world. Retrieved information from the GBIF data portal gave a total 4,691 *S. scrofa* occurrence records in Spain in the period 1982-2013, with a resolution of ≤ 10 km, mainly consisting of field observations (~95% of the data). The main source of records in the GBIF was the Atlas and Red Data Book of the Terrestrial Mammals of Spain (3,669 out of 4,691 records) by Palomo *et al.* (2007) (National Biodiversity Inventory 2007, Ministry of Environment and Rural and Marine Affairs, Spain). This atlas provides information on the distribution of species in UTM 10 × 10 km grids corresponding to 15 years of work collating bibliographic data, data from collections in museums and scientific institutions, surveys and questionnaires conducted by technical staff in protected natural areas, and unpublished data from collaborators, partners and the authors' own personal observations and sampling. A further 119 presence records with coordinates were also used that were spatially and temporally distinct from those in the GBIF web. Data from Melis *et al.* (2006) and unpublished field data from Madrid and Andalusia (Spanish Ministerio de Agricultura, Alimentación y Medio Ambiente, 2012)

were also obtained. These presence records have GPS coordinates and were collected from animal trapping studies.

Typically, neighboring records are associated with similar values for environmental variables, which will potentially violate the assumption of independence (Heffner *et al.*, 1996). To mitigate pseudoreplication (Heffner *et al.*, 1996), a minimum distance between sampling sites that was greater than the minimum distance at which autocorrelation is generated was defined (Guisan & Zimmermann, 2000). To reduce this spatial autocorrelation, the distance between data pairs was widened and the density of points of presence was reduced to a minimum distance of 0.15 decimal degrees (~16 km) using the statistical software R (R Development Core Team, 2012). After applying the exclusion criteria and performing the selection process to reduce the density between occurrences, a total of 1,082 of the original 4,691 GBIF points became available for model building (Fig. 1).

Finally, we generated a random sample of 10,000 background points from the environmental data (Phillips & Dudík, 2008; Elith *et al.*, 2011), which are required by the MaxEnt method to mimic absences or pseudo-absences of the species.

Data source: environmental variables

After a review of the most important factors affecting the distribution of wild boar in Spain based on information available on a large scale, we selected 38 variables as potential predictors of wild boar distribution. Environmental variables were grouped into climatic predictors explaining the species macroecology, and into topographic, solar radiation, human influence, and vegetation predictors to determine the abundance and distribution of animals. The data sources are shown in Suppl. Table S1 [pdf online]. Briefly, we used 19 rasters from the WorldClim online database for the period 1950-2000 (Hijmans *et al.*, 2005) at a spatial resolution of 5 arc-minutes (~10 km). Topography layers included altitude (elevation), slope, and topographic diversity. Altitude (USGS, 2004) was obtained from the Global Land Cover Facility (<http://glcf.umd.edu/data/>) at a spatial resolution of 30 arc-seconds (~1 km); we then changed the spatial resolution to 5 arc-minutes (~10 km) and calculated the average of 1 × 1 km cells that occur inside each 10 × 10 km cell. Slope and topographic diversity were

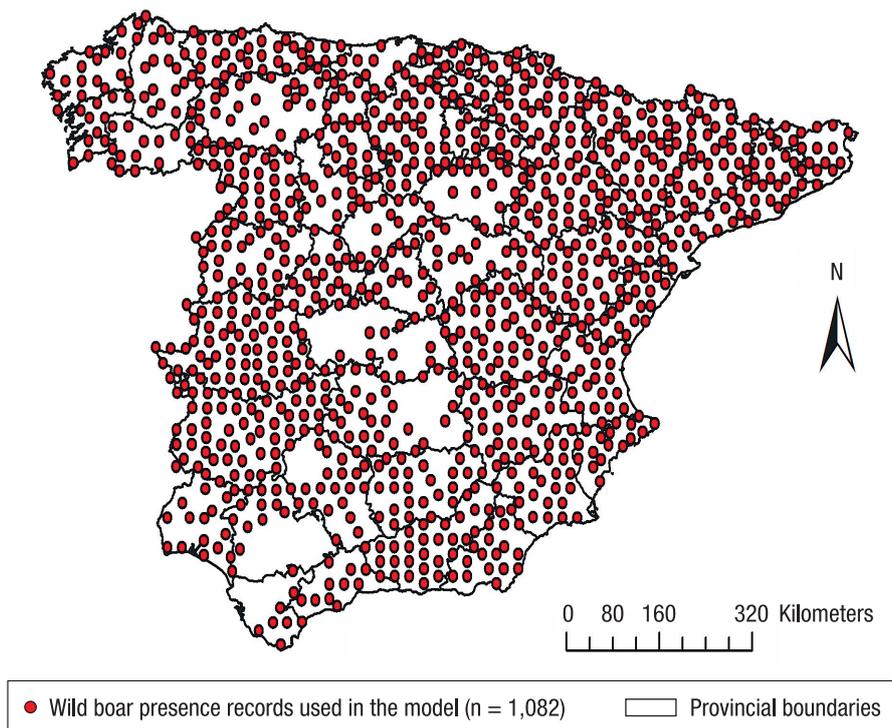


Figure 1. Distribution of points after the selection of the *Sus scrofa* presence records in Spain (occurrence data for wild boar in Spain). A total of 1,082 points were obtained, thereby reducing the density between points of presence.

derived from the elevation variable and were generated at a spatial resolution of 10 km. Topographic diversity represents the topographic complexity of the terrain (the sum of the different number of slopes, elevations, and orientations in a 10 km radius around a given cell). Potential solar radiation variables were calculated using the elevation model (obtained from WorldClim in the Shuttle Radar Topography Mission (SRTM) elevation database), slope, orientation, and latitude and longitude maps. A series of equations that simulate the movement of the sun at certain dates and times – but taking into account the masking effect of the topography – were used. Solar radiation variables were generated with the Geographic Resources Analysis Support System (GRASS Development Team, 2011) software vers. 6.4.1 (<http://grass.osgeo.org>) and implemented in the module *r.sun* (Suri & Hofierka, 2004); the resolution was estimated at 5 arc-minutes (10 km). Anthropogenic or human influence was approximated using the human footprint raster (Sanderson *et al.*, 2002) obtained from the Socioeconomic Data and Applications Center (SEDAC, <http://sedac.ciesin.columbia.edu>) and by changing the spatial resolution (from 1 km to 10 km) in the same procedure as for the

altitude variable. Finally, the Normalized Difference Vegetation Index (NDVI) datasets (Tucker *et al.*, 2004) estimating the quantity and quality of vegetation development were obtained from the Global Inventory Monitoring and Modeling Studies (GIMMS), while the vegetation structure as a percentage of bare, herbaceous, or tree coverage were taken from the Vegetation Continuous Fields (MODIS-VCF). These data represent variations in the vegetation index during the 12 months of 2004. Despite covering only a short period of time, these data were selected because they offer good spatial resolution for this variable.

The 38 environmental predictors were evaluated to reduce collinearity by screening out a correlation tree or cluster dendrogram (Suppl. Fig. S1 [pdf online]). The evaluation process analyzed the correlation matrix of environmental variables according to distance (shortest distance = higher correlation), which identifies redundant variables using the raster package (Hijmans & van Etten, 2012) implemented in the R program. From the resulting correlation tree, variables were selected based on a cutoff or threshold minimum of 0.5. In each group of variables with node <0.5, only one variable was selected based on statistical and

biological criteria. Of the set of variables with the lowest correlations, the most representative ones for the wild boar were selected according to biological criteria and taking into account the particular environmental conditions in Spain. In addition to the above procedure, we calculated the variance inflation factor by sequentially removing variables with higher values (maximum value allowed = 5) since variables may exist that are a linear combination of other variables.

The working resolution for all environmental variables for habitat suitability mapping was 5 arc-minutes, ~10 km on the WGS 84 projection.

Model formulation and evaluation

We used a maximum entropy algorithm available in MaxEnt (Phillips *et al.*, 2006). Models based on 100 bootstrapped replicates were built and tested, *i.e.*, replicate sample sets were chosen by sampling with replacements, by selecting 'random seed', and cross-validating. For each replicate or simulation, 1,082 presence data were divided into two subsets that were used for model fitting (60% of the data) and cross-validation (40% of the data) (Fielding & Bell, 1997).

The accuracy of the final model was estimated by computing the area under the curve (AUC) of the receiver operating characteristic (ROC) curve, which is the preferred technique used to evaluate models based on presence-only data (Stockwell & Peters, 1999). Briefly, a ROC plot was built by plotting the sensitivity – the fraction of true positives out of the total number of positives (wild boar presences) – against the false positive fraction at various threshold settings (Manel *et al.*, 2001). Subsequently, the AUC was determined and used as a measure of the discriminating power of the fitted model (Pearce & Ferrier, 2000). The closer the AUC value is to 1, the greater its accuracy; values of 0.5 suggest that the model performs no better than random.

The model was fitted using an iterative process in which each iteration resulted in an increase in the regularized gain of the model due to the modification of a coefficient for a single feature. This gain was normalized to percentages in relation to the drop in the AUC values at the end of the reevaluation process. Variables were ranked based on the estimated percentage contribution, and values are shown as averages over replicate runs. The model's predictions are given in logistic format and can be interpreted as the predicted probability of *S. scrofa* presence in the region.

MaxEnt models were tested, selected, and evaluated using the default parameters in the MaxEnt software, vers. 3.3.3 (iterations (1,000) and by being stricter than recommended by authors of the algorithm (Phillips *et al.*, 2006). The resulting model was expressed on a map using the maximum value (point-wise) of the 100 replications. This map was drawn on ArcGIS 9.3 (ESRI®); map of the standard deviation of the 100 replications and a map of the 95% confidence level (lowerci) of the 100 replications are also included (Suppl. Fig. S2 [pdf online]).

Models based on presence-only data cannot be accurately validated by field data; however, it is possible to show the predictive ability of the model and whether or not the errors are acceptable (Lobo *et al.*, 2008). Thus, the 3,728 unused presence records ($n = 3,609$ from 4,691 GBIF and 119 from other field studies) were overlapped on the results of the model. The probabilities of presence values were classified according to their suitability for the wild boar as a means of comparing the model results with the pre-existing presence records for the species. Cells with probability values in the range 0-0.5 were classified as unsuitable, values in the range 0.5-0.6 as of low suitability, values in the range 0.6-0.7 of medium suitability, and values equal or higher than 0.7 as of high suitability.

Results

Retrieved data and correlation analysis

The result obtained after applying the exclusion criteria and performing the selection process to reduce the spatial autocorrelation between occurrences (decreased density of presences), a total of 1,082 points were available for model building.

The screening of the correlation tree and the variance-inflated factor (VIFs <5) restricted the number of variables included as predictors in the MaxEnt approach to 13 (out of 38). The selected variables were grouped as climatic ($n = 6$), potential solar radiation ($n = 1$), human footprint ($n = 1$), topography ($n = 2$), and vegetation ($n = 3$) (Table 1).

The final model proved to be accurate, with a mean AUC value of 0.79, a standard deviation of 0.007, and minimum and maximum values of 0.78 and 0.81, respectively. As AUC values above 0.75 are considered

Table 1. Environmental variables included in the model-building process (after screening based on the correlation tree) for modeling the distribution of *Sus scrofa* in Spain

Group	Variable name	Description
Climatic ¹	bio3	Isothermality (bio2/bio7) (* 100)
	bio5	Maximum temperature of warmest month
	bio6	Minimum temperature of coldest month
	bio8	Mean temperature of wettest quarter
	bio12	Annual precipitation
	bio15	Precipitation seasonality (coefficient of variation)
Potential solar radiation ²	sunh_ra	Sunshine hours, range of direct solar radiation (range of variable, difference between the maximum and minimum)
Human influence	hfp	Human footprint. Represents the human influence on the territory
Topography	tslope	Slope (terrain slope in degrees)
	tdiv	Topographic diversity represents the topographic complexity of the terrain (the sum of the different number of slopes, elevations, and orientations in a 10-km radius around a given cell)
Vegetation structure	covbare	Percentage of land area occupied by bare soil cover
	covtree	Percentage of land area occupied by tree cover
Vegetation index ³	ndvi_mn	Normalized difference vegetation index (annual minimum). NDVI estimates its quantity and quality, and vegetation development.

¹ The units of the temperature variables are expressed in tenths of degrees. ² Watts per square meter per hour (Wh / (m * m) / day).

³ Variations in the vegetation index during 2004.

informative (Phillips & Dudík, 2008) our uniform values indicates that all models provided good discrimination between true positive and false positive (Fielding & Bell, 1997; Pearce & Ferrier, 2000). Additional testing for each of the 100 replications using a binomial test of omission revealed statistical significance for the prediction ($p < 0.001$) (Phillips *et al.*, 2006), thereby supporting the reliability of the final model. The largest standard deviations (0.14-0.25) were located in very restricted areas such as the Pyrenees and Cantabrian mountains in northern Spain, and in southern Spain in the Alcornocales Natural Park and Sierra Nevada National Park (Suppl. Fig. S2 [pdf online]). In the other areas of Spain the data deviate little from the average values. In general, the results of the 95% confidence level (lowerci) coincide with montane and upland areas (Suppl. Fig. S2 [pdf online]).

Accordingly, the top six explanatory variables identified by their percentage contribution were: 1) sunshine hours 25.5% ('sunh_ra'); 2) precipitation seasonality 25.2% (coefficient of variation, 'bio15'); 3) isothermality 10.7% ('bio3'); 4) minimum tempera-

ture of coldest month 7.7% ('bio6'); 5) slope 6% ('tslope') and 6) annual precipitation 4.8% ('bio12'). The response curves (logistic output) produced by univariate models of the six most important predictor variables are given in Fig. 2. *Sus scrofa* habitat suitability increased with sunshine hours, annual precipitation and slope, but decreased when the coefficient of variation seasonality of the precipitation increased and had a varied response to isothermality. Both temperature seasonality and isothermality are a measure of variability in the temperature over the course of the year. Others variables (all with percentage contributions of less than 4.3) that increase the predicted probability in favorable situations include the percentage of land area occupied by tree cover ('covtree'), topographic diversity (topographic complexity of the terrain) ('tdiv'), and the maximum values of the normalized difference vegetation index (annual minimum) ('ndvi_mn'). However, the probability decreases with the percentage of land area occupied by bare soil cover ('covbare'), the minimum values of ('ndvi_mn') and, in general, to greater human influence.

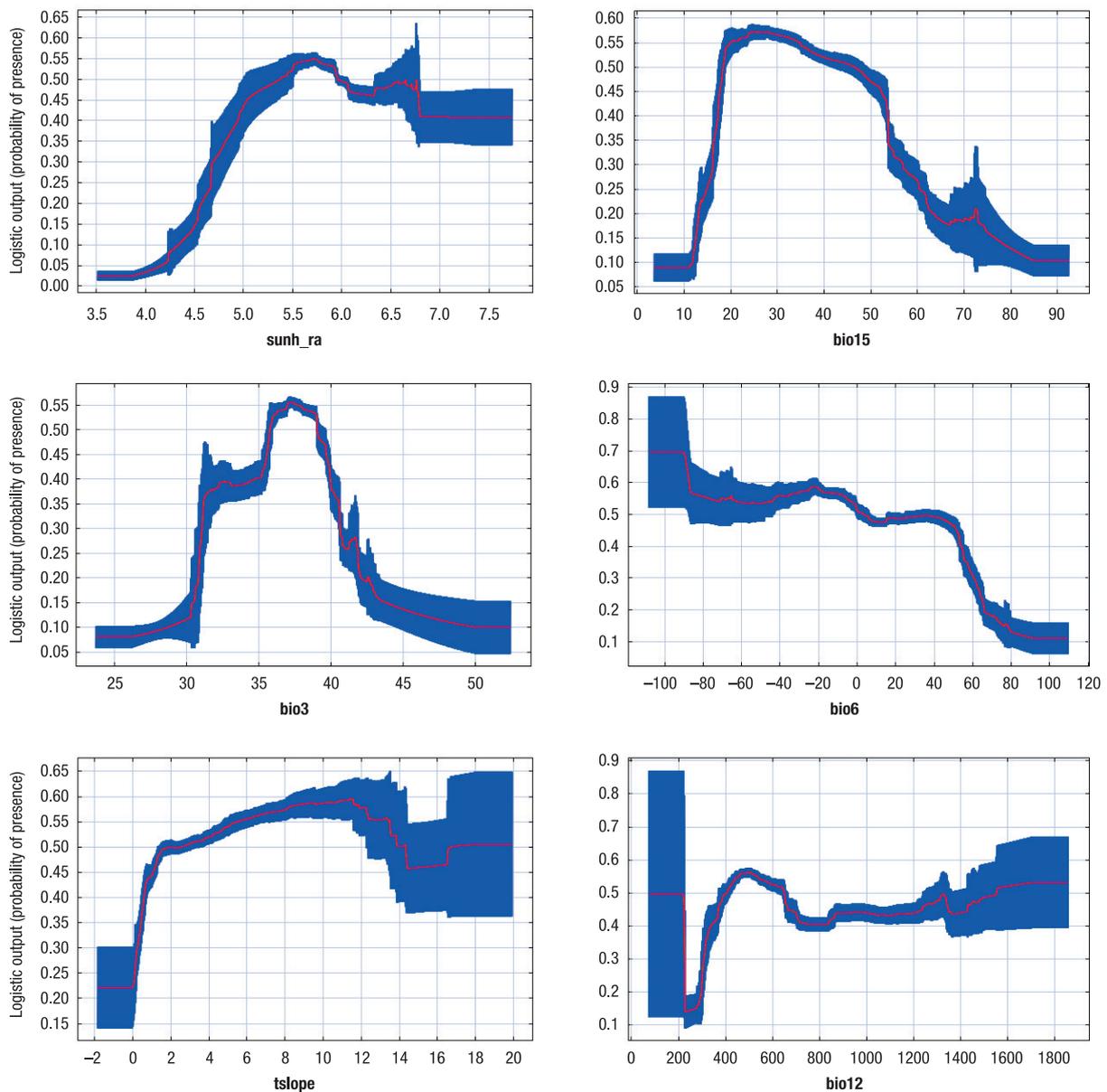


Figure 2. Response curves. The curves show the mean response of the 100 replicate MaxEnt runs (red) and the mean \pm one standard deviation (blue). Each curve represents a MaxEnt model created by using only the corresponding variable. These plots reflect the dependence of predicted suitability on the selected variable. On the Y-axis, logistic output (probability of presence) and on the X-axis (variable value). Abbreviations as follows: sunshine hours range (sunh_ra), precipitation seasonality (bio15), isothermality (bio3), minimum temperature of coldest month (bio6); slope (tslope) and annual precipitation (bio12).

Predicted habitat suitability distribution

Modelling the distribution of *S. scrofa* produced a broad potential distribution that runs across much of Spain (Fig. 3). In general, five areas stand out as the most suitable for the species: 1) the Cantabrian-Basque Mountains; 2) Pyrenees-Catalan Coastal Range; 3)

Iberian System (Valencia-Teruel); 4) Sierras de Cazorla, Sierra Morena and Mountains of Toledo; and 5) the Central System. The highest habitat suitability values (probability of presence per cell ≥ 0.7) coincide in general with upland areas, but also include certain lowland areas such as the Doñana National Park straddling the provinces of Huelva, Cadiz, and Seville,

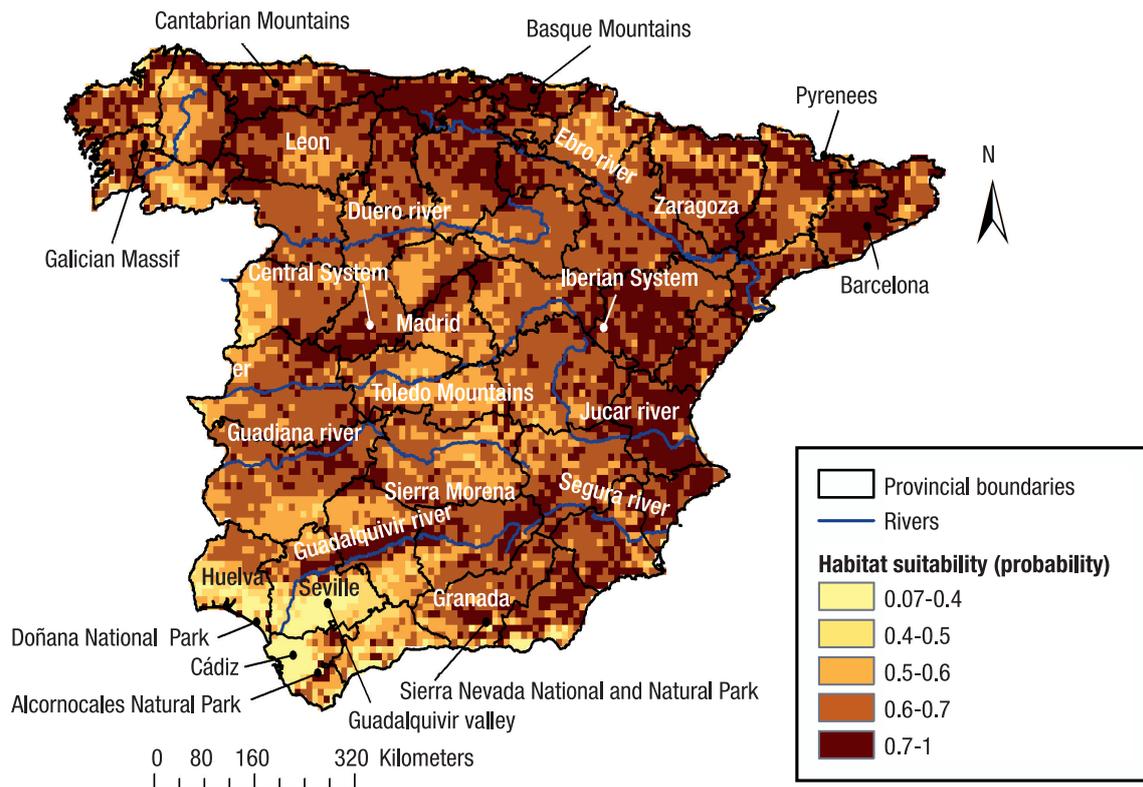


Figure 3. Potential geographic distribution of *Sus scrofa* in Spain (Model). Replicated MaxEnt model for *Sus scrofa* (using 60% and 40% for model fitting and for cross validation, respectively). Model with the maximum value (point-wise) of the 100 output grids.

and the Alcornocales Natural Park in Cadiz province. The unsuitable habitats in Spain (probability of presence per cell ≥ 0.5) were identified in two main areas: 1) provinces of Huelva, Seville, Malaga, and Cadiz; 2) southern Almeria, in southern Spain. The lowest suitability habitats in Spain, in general, the probabilities of presence between 0.5 and 0.6 correspond to many of the major river valleys and depressions such as those of the Miño, Tajo, Guadalquivir, Guadiana and Duero rivers, and low suitability areas were also identified in small areas in the north of Huesca and Navarra provinces in northern Spain. The results of the predictive ability measure of the model overlapping on the field data are as follows:

Using the presence records not employed in the model (from the decreased density of presences), observed presence coordinates were seen to coincide with the areas of high and medium probability of wild boar presence predicted by the model. Of the total points of presence not used in the model (3,609 GBIF records plus 119 from other studies), 97.2% of records matched suitability cells. 40.02% of records matched high

suitability cells, 47.49% medium suitability cells, and 10.41% low suitability cells. Only 2.09% of the presence data coincided with cells classified as unsuitable. The data are shown in Table 2 and the coordinates in Suppl. Fig. S3 [pdf online].

Discussion

In this study, the potential distribution of wild boar in Peninsular Spain was effectively predicted by a MaxEnt approach maximizing the use of information from open-source databases. Visualization of the predicted distribution reported in this study was similar to that of previous studies of wild boar distribution in the Euroasiatic zone, in particular in Spain (Spencer & Hampton, 2005; Oliver & Leus, 2008) and in the Iberian Peninsula (Palomo *et al.*, 2007; Araújo *et al.*, 2011a,b; Bosch *et al.*, 2012; Acevedo *et al.*, 2014). However, apart from using a different set of environmental variables, the main difference in the present study is that the fitted model also identifies the varia-

Table 2. The results of the predictive ability measure of the model

Probability model values	Suitability range	Percentage (%) ¹ n = 3,728	Average	SD	Min	Max
0-0.5 (Unsuitable)	0	2.5	0.41	0.089	0.16	0.49
0.5-0.6 (Low suitability)	1	7.73	0.57	0.025	0.5	0.59
0.6-0.7 (Medium suitability)	2	49.74	0.66	0.026	0.6	0.69
0.7-1 (High suitability)	3	40.02	0.75	0.041	0.7	0.94

¹ Percentage of cells classified as unsuitable, or of low, medium, or high suitability for wild boar presence. The 3,728 unused presence records (n = 3,609 from GBIF and 119 from other studies) were overlapped on the model results to test the predictive ability of the model. The meaning of the values of the average, standard deviation (SD), minimum (Min), and maximum (Max) correspond to the total number of presence records in each suitability range (0, 1, 2, and 3).

bles that have the greatest influence on species distribution – thereby providing information regarding the response profile of each variable – and combines their predictability to generate a higher resolution map.

When interpreting these data, it is important to consider that, although we have minimized the correlation between variables included in this model (see in material and methods section), it is possible that the variables reported to have a high percent contribution to the model are not actually the drivers of the distribution of *S. scrofa*. But rather, they are important in the model only because these variables are correlated with environmental variables that were not included in the model. Regardless of whether these parameters directly shape the distribution of *S. scrofa*, or are in fact only correlated with the true (unidentified) drivers of its distribution, these results could be used to identify suitable areas where *S. scrofa* may be found and provide a starting point for experimental work to elucidate the true environmental factors which are most important in driving the current distribution of wild boar in Spain.

In general, high suitability areas are characterized by mountainous terrain with forests, grassland, and sometimes wetlands. Such areas are particularly prevalent on the Central Plateau in the center of the Peninsula (the Central System and Mountains of Toledo), where there are grid cells with adequate environmental conditions (maximum likelihood) for the species occurrence (probability of presence ≥ 0.7). Some of the areas of highest suitability coincide with areas of known high density (western Pyrenees, Sierra Morena and Mountains of Toledo) from where wild boar populations are considered to have dispersed to the northwest and southeast of the Iberian Peninsula (Tellería & Sáez-Royuela, 1985). Nevertheless, the main areas identified as being unsuitable are located in the south

of Spain, in Guadalquivir valley and in the southeast. These areas are characterized by a strong anthropization of the environment. In addition, it is an area where there have been no reported presence of wild boar (Palomo *et al.*, 2007), due to its lack of potential resource for this species (Bosch *et al.*, 2012). According to our model, the specific factors that might be limiting the presence of wild boar in these areas are associated with human footprint, percentage of land areas occupied by bare soil cover and annual minimum of normalized difference vegetation index variables. In the northwest of Spain (closed to Galician Massif) where potential resource for this species had been previously reported (Bosch *et al.*, 2012), certain areas were identified as low suitability possibly due to the few collected presence data.

Of the variables with the greatest influence, precipitation seasonality (coefficient of variation) accounted for most variability in the prediction model. Due to the climatic characteristics of Spain, rainfall typically decreases in summer as in other Mediterranean bioclimatic zones with a high level of seasonal variability in precipitation. The precipitation seasonality variable expresses the variation in the level of rainfall over the different seasons in a given area, which in Spain is associated with latitude. As we move northwards towards the Atlantic bioclimatic region, the coefficient of variation of seasonal rainfall becomes more constant throughout the year, whereas in more southerly, Mediterranean influenced latitudes, the coefficient of variation of seasonal rainfall increases. Hence, at more northerly latitudes, the water regime is characterized by more constant rainfall, while in southern latitudes variation is greater and the rainfall regime is less constant over the year.

Temperature and precipitation have a significant influence on the distribution of terrestrial vertebrate

fauna since these two factors synthesize the flows of energy and water in the ecosystem and substantially limit the global distribution of biodiversity (Hawkins & Porter, 2003; Whittaker *et al.*, 2007). According to Austin (1985), the use of direct selection gradients and resources to calibrate ecological models improves the interpretation of the results. These criteria must be added to the spatial hierarchies that are subject to variables that control the distribution of vegetation (Huntley *et al.*, 1995; Neilson, 1995) since animals depend directly on the food and shelter that vegetation provides and their distribution is more affected by the structural characteristics of the vegetation than by other factors (except for human impact) (Markina-Lamonja, 1998). Bearing in mind the aforementioned points and the fact that Spain has a high level of species biodiversity (UNESCO, 1977; De Miguel, 1999; Ruiz de la Torre, 2002; Sainz *et al.*, 2010) and a great variety of habitats (due to topographic heterogeneity, climatic contrasts, complex geomorphology, and notable geographical and lithological partitioning), we can select either climatic variables, which determine distribution patterns at large scales (*e.g.* peninsular or European scale) using coarse resolutions (grid cells of 1-10 km²), or topographical and geological variables. Taking these aspects into consideration, through these gradients and ecological predictors, we were able to capture much of the study area's ecological and environmental variability and thus predict the potential distribution of wild boar in Spain with reasonable accuracy.

As stated in our results, the ecological predictor that best forecasts the presence of wild boar in Spain is precipitation associated with the energy flow in the ecosystem (precipitation, temperature, and sunshine hours), complemented by other factors such as slope and the diversity of terrain topography, vegetation structure and, in general, the low levels of human disturbance.

Due to the large variations in factors such as altitude, temperature, and climate that exist in Spain, the country possesses a high degree of habitat heterogeneity. Wild boar occupy a variety of habitats in Spain, from sea level to an altitude of around 2,400 m, with temperatures in the range -14.8-36.3°C and annual rainfall levels in the range 214-1,949 mm (Araújo *et al.*, 2011a). These ranges for wild boar habitats are supported by our results as shown by the response curves for precipitation and temperature: annual precipitation (bio12) in the range 400-1,800 mm, minimum temperature of coldest month (bio6) in the range -10-10°C

(Fig. 2), and maximum temperature of warmest month (bio5) in the range 10-40°C.

According to the presence probability expressed by the previously described response curves, there is a direct relationship between wild boar presence and environmental factors. Bioclimatic levels depend on direct and indirect resource gradients including flow variables such as energy and water temperature, sunshine hours, and rainfall, which also vary depending on altitude, longitude, and orientation (abiotic interactions). These sets of gradients either limit or encourage environmental conditions in which different types of vegetation adapted to these biotopes can thrive. Hence, wild boar presence is also linked directly to the type of vegetation (land cover) since it provides the habitat in which boars develop and survive. It is important to note that apart from rainfall, water input may also come in the form of dew in areas of high environmental humidity or from the alteration of the ecosystem balance associated with modern agriculture through water supply using new water infrastructures. Huge steppe areas have been becoming irrigated, with high productivity in crops; thus artificially providing food and shelter for animals and finally causing authentic population explosions of these suids in Europe during the last decades (Sáez-Royuela & Tellería, 1986). This species, along with others, found here an opportunity to colonize an environment, *a priori*, not suitable for them. One of the most predominant species in these "wet deserts" is corn (*Zea mays*), forming large areas of this crop. These cornfields act as an "artificial forest" with abundant food, shelter or refuge, tranquility and water, *i.e.*, an ideal artificial and temporary habitat for wild boar.

However, this variability in the data is not taken into account in this study because the large-scale distribution of this crop in Spain is not available to date. More accurate estimates of the geographical distribution of the species would require more sophisticated methodological approaches, which may explicitly include the mechanisms responsible for local population dynamics (Keith *et al.*, 2008; Anderson *et al.*, 2009), that is, dispersal mechanisms and biotic interactions (Araújo & Luoto, 2007; Hirzel & Le Lay, 2008), limiting dispersion factors (*i.e.* natural or artificial geographical barriers), and the role of absences data (Lobo *et al.*, 2010).

It is important to take note of the limitations of the AUC statistic when true instances of absence are not available to validate the model error as previously

described (Lobo *et al.*, 2008; Peterson *et al.*, 2008; Jiménez-Valverde, 2012). Depending on the species and the territory, the factors causing these absences vary. Unfortunately, in the case of the wild boar, absence data are not available and they are difficult to estimate accurately, largely owing to the high ecological plasticity of this species and the human factors that affect it (*e.g.* introduction for hunting). Therefore, future research should be geared to identifying these absences and developing real wild boar (realized) distributions in order to improve our predictive ability and to validate the model error (Jiménez-Valverde, 2012).

The inclusion of biotic interactions or absence data, for instance, in these models gives more realistic distributions (Araújo & Luoto, 2007; Heikkinen *et al.*, 2007; Baselga & Araújo, 2009). Unfortunately, the use of these factors is still under study and it was not possible to include them in our model.

Other models capable of estimating the response of species to climate change or other changes in the environment are still at an experimental stage (Brook *et al.*, 2009) and require parameters that are not available for most species. However, a number of approaches for analyzing the influence of climate change and in current conditions have been developed for the study of wild boar in the Iberian Peninsula (Araújo *et al.*, 2011a,b).

Finally, the predictive ability of the model was assessed using field data as described above (Lobo *et al.*, 2008). The results showed many habitat suitability areas in the sampling points of other authors and presences that were not used in the model but which coincide with actual *Sus scrofa* field occurrence localities, which confirms the model accuracy at these points (Suppl. Fig. S3 [pdf online]). The results show that the predictive ability of the model is high in the areas where we compare the species' presence and that the error is quite acceptable since only 2.09% of the 3,728 presence records distributed throughout Spain coincided with cells classified as unsuitable. However, it is worth noting that these latter records were located very close to cells with good habitat suitability for the species (Suppl. Fig. S3 [pdf online]).

Habitat models provide information about the environmental requirements of species, facilitate the application of this information, and fill the gap between science and management by focusing on conservation biology (Elith *et al.*, 2006; Phillips *et al.*, 2006; Peterson *et al.*, 2011). The model generated will help identify areas where hunting is of concern, that are

close to urban or rural centers where wild boars are more likely to cause traffic accidents, and those that are near croplands. This in turn facilitates the detection of true hotspot contact areas between wild boar and livestock and dispersion corridors for this species between countries, particularly those located in the altitudinal range of 500-2,500 m a.s.l. The temperature and precipitation characteristics in this altitudinal zone favor the presence of wild boar and are reflected in the variety of altitudinal environments that arise in the transition from Atlantic to Mediterranean bioclimatic areas.

Unfortunately, data of presence from Portugal are not currently available. It would have been very interesting to have interpreted the results with data from this country because its Atlantic climate probably influences wild boar populations in a different way. Bearing in mind that to obtain a distribution model for a species such as wild boar with a worldwide range, the selection and interpretation of environmental and climatic variables should be done very carefully as very significant regional peculiarities exist and these variables may not fully explain the probability of presence. Other variables related to biotic interaction and absence data could be added to the analysis to help determine the best explanation for the presence of the species. However, these data are not currently available.

Acevedo *et al.* (2014) have recently published a study focused on Spain determining the abundance of wild boar that is based on hunting yields and environmental predictors (above all climatic predictors and predictors related to the most important land cover for wild boar). Previously, Bosch *et al.* (2012) undertook a complete review and used a standardized European land-cover program to develop a habitat suitability map for the Iberian Peninsula which, moreover, included a unified habitat and a density map per grid cell. Suitable potential habitats where the wild boar might thrive were determined on the basis of selected land uses and assigned specific weights related to the land's ability to supply food and/or shelter to the wild boar.

Both of these studies used hunting data but both seem to oversimplify the true situation given that they did not employ various important – but currently unavailable – biological variables such as biotic interactions. Much effort – including the present study, in which a probability of presence score is calculated that gives the habitat suitability index per grid cell – has been made to typify wild boar habitat in Spain since the first potential habitat model for wild boar using

presence data and environmental variables was presented in 2012 (Bosch *et al.*, unpublished data). We anticipate that these studies can be used to compare strategies, results, and methodologies to obtain an evermore exact map of wild boar distribution, abundance, and density in Spain.

One of the inherent challenges in the present study was to develop a methodology based on presence data rather than hunting data since many authors have criticized the use of the latter as a source of data in scientific or technical work given that, among other reasons, hunting statistics are often incomplete, disperse, and rarely homogeneous over time. Likewise, the complexity of hunting practices is great since there are many different methods of hunting; hunting effectiveness varies and there is great heterogeneity in hunting grounds and management practices (Martínez-Jaúregui *et al.*, 2011; Sarasa & Sarasa, 2013). In general, the results of the present study do not differ greatly from those obtained using other methods.

This fact implies that the models that use hunting data to calculate densities (Bosch *et al.*, 2012) or abundances (Acevedo *et al.*, 2014) may be as valid as those that are based on presence data since they are very similar when compared on a spatial level. A large number of methods and techniques exist and all require distinct types of data and generate results with differing predictive abilities. Nevertheless, the tendency of the SDM (species distribution models) is to use consensus methods to combine predictions (Laplace, 1820; Thuiller, 2004; Araújo & New, 2007) in order to decrease the predictive uncertainty of single models (Araújo *et al.*, 2005). Only through the efforts such as those of the present study and the other abovementioned studies will it be possible to develop a fully accepted method that will improve the prediction of wild boar distribution in Spain.

Our model generates highly accurate predictions, as confirmed by satellite images and field surveys and could be used in studies concerning the distribution, management, and conservation of wild boar and wildlife research in general.

Acknowledgements

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 311931.

References

- Abaigar T, 1993. Régimen alimentario del jabalí (*Sus scrofa*, L. 1758) en el sureste ibérico. Doñana Acta Vert 20(1): 35-48.
- Acevedo P, Quiros-Fernandez F, Casal J, Vicente J, 2014. Spatial distribution of wild boar population abundance: Basic information for spatial epidemiology and wildlife management. Ecol Indic 36: 594-600.
- Anderson BJ, Akçakaya HR, Araújo MB, Fordham DA, Martínez-Meyer E, Thuiller W, Brook BW, 2009. Dynamics of range margins for metapopulations under climate change. Proc R Soc B 276: 1415-1420.
- Araújo MB, Luoto M, 2007. The importance of biotic interactions for modelling species distributions under climate change. Glob Chang Biol 16: 743-753.
- Araújo MB, New M, 2007. Ensemble forecasting of species distributions. Trends Ecol Evol 22: 42-47.
- Araújo MB, Guilhaumon F, Neto DR, Pozo I, Calmaestra R, 2011a. Biodiversidade e alterações climáticas/biodiversidad y alteraciones climáticas. Ministerio do Ambiente e Ordenamiento do Território & Ministerio de Medio Ambiente y Medio Rural y Marino. Lisboa /Madrid. 656 pp. Available in http://www.ibiochange.mncn.csic.es/iberiachange/wordpress/wp-content/uploads/2008/07/Libro-1-junio-2012.pdf?goback=.gde_2716697_member_121228272. [15 September 2012].
- Araújo MB, Guilhaumon F, Neto DR, Pozo I, Calmaestra R, 2011b. Impactos, vulnerabilidad y adaptación al cambio climático de la biodiversidad española. 2. Fauna de vertebrados. Dirección General de Medio Natural y Política Forestal. Ministerio de Medio Ambiente, y Medio Rural y Marino. Madrid, 640 pp. Available in http://www.magrama.gob.es/es/biodiversidad/temas/inventarios-nacionales/inventarioespecies-terrestres/efectos_cambio_climatico.aspx. [02 July 2012].
- Araújo MB, Whittaker RJ, Ladle RJ, Erhard M, 2005. Reducing uncertainty in projections of extinction risk from climate change. Global Ecol Biogeogr 14: 529-538.
- Austin MP, 1985. Continuum concept, ordination methods, and niche theory. Annu Rev Ecol Syst 16: 39-61.
- Baldwin RA, 2009. Use of maximum entropy modeling in wildlife research. Entropy 11(4): 854-866.
- Baselga A, Araújo MB, 2009. Individualistic vs. community modelling of species distributions under climate change. Ecography 32: 55-65.
- Benito de Pando B, Peñas de Giles J, 2007. Aplicación de modelos de distribución de especies a la conservación de la biodiversidad en el sureste de la Península Ibérica. GeoFocus (Artículos) 7: 100-119.
- Bosch J, Peris S, Fonseca C, Martínez M, De la Torre A, Iglesias I, Muñoz MJ, 2012. Distribution, abundance and density of the wild boar, *Sus scrofa* L., on the Iberian Peninsula, based on the CORINE program and hunting statistics. Folia Zool 61(2): 138-151.
- Brook BW, Akçakaya HR, Keith DA, Mace GM, Pearson RG, Araújo MB, 2009. Integrating bioclimate with population models to improve forecasts of species extinctions under climate change. Biol Lett 5: 723-725.

- Colino-Rabanal VJ, Bosch J, Muñoz MJ, Peris SJ, 2012. Influence of new irrigation croplands on wild boar *Sus scrofa* road kills in NW Spain. *Anim Biodiv & Conserv* 35(2): 97-102.
- De Candolle AI, 1855. *Géographie botanique raisonnée*. Masson, Paris.
- De la Torre A, Bosch J, Iglesias I, Muñoz MJ, Mur L, Martínez-López B, Martínez M, Sánchez-Vizcaíno JM, 2013. Assessing the risk of African swine fever introduction into the European Union by wild boar. *Transbound Emerg Dis* doi:10.1111/tbed.12129.
- De Miguel JM, 1999. Nature and configuration of the agricultural-forestry-pasture landscape in the conservation of biological diversity in Spain. *Rev Chil Hist Nat* 72: 547-557.
- De Miguel JM, 2002. Ecología, diversidad y desarrollo sostenible en sistemas agroforestales tradicionales en España. *Cuad Soc Esp Cien For* 14: 23-32.
- DiMiceli CM, Carroll ML, Sohlberg A, Huang C, Hansen MC, Townshend JRG, 2011. Annual global automated MODIS vegetation continuous fields (MOD44B) at 250 m spatial resolution for data years beginning day 65, 2000-2010, Collection 5 Percent Tree Cover, University of Maryland, College Park, MD, USA.
- Ellenberg H, 1988. *Vegetation ecology of Central Europe*, 4th ed. Cambridge University Press, Cambridge, UK.
- Elith J, Leathwick JR, 2009. Species distribution models: ecological explanation and prediction across space and time. *Annu Rev Ecol Evol Syst* 40: 677-697.
- Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick J, Lehmann A, *et al.*, 2006. Novel methods improve prediction of species distributions from occurrence data. *Ecography* 29: 129-151.
- Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ, 2011. A statistical explanation of MaxEnt for ecologists. *Diversity Distrib* 17: 43-57.
- Ferrier S, 2002. Mapping spatial pattern in biodiversity for regional conservation planning: Where to from here? *Syst Biol* 51: 331-363.
- Fielding AH, Bell JF, 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ Conserv* 24(1): 38-49.
- Fithian W, Hastie T, 2013. Finite-sample equivalence of several statistical models for presence-only data. *Ann Appl Stat* 7(4): 1917-1939.
- García M, Maldonado J, Morla C, Sainz H, 2002. Fito-geografía histórica de la península Ibérica. In: *La diversidad biológica de España* (Pineda FD, de Miguel JM, Casado JM, Montalvo J, eds.), Pearson Educación, Prentice Hall, Madrid. pp: 45-63.
- Guisan A, Zimmermann NE, 2000. Predictive habitat distribution models in ecology. *Ecol Model* 135(2-3): 147-186.
- Guisan A, Thuiller W, 2005. Predicting species distribution: offering more than simple habitat models. *Ecol Lett* 8(9): 993-1009.
- Hansen MC, DeFries RS, Townshend JRG, Carroll M, Dimiceli C, Sohlberg RA, 2003. Global percent tree cover at a spatial resolution of 500 meters: First results of the MODIS vegetation continuous fields algorithm. *Earth Interact* 7: 1-15.
- Hawkins BA, Porter EE, 2003. Relative influences of current and historical factors on mammal and bird diversity patterns in deglaciated North America. *Glob Ecol Biogeogr* 12: 475-481.
- Heffner RA, Butler MJ, Reilly CK, 1996. Pseudoreplication revisited. *Ecology* 77: 2558-2562.
- Heikkinen R, Luoto M, Virkkala R, Pearson RG, Körber JH, 2007. Biotic interactions improve prediction of boreal bird distributions at macro-scales. *Glob Ecol Biogeogr* 16: 754-763.
- Hernández PA, Graham CH, Master LL, Albert DL, 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29: 773-785.
- Herrero J, Irizar I, Laskurain NA, García-Serrano A, García-González R, 2005. Fruits and roots: wild boar foods during the cold season in the southwestern Pyrenees. *Ital J Zool* 72(1): 49-52.
- Herrero J, García-Serrano A, Couto S, Ortuño V, García-González R, 2006. Diet of wild boar *Sus scrofa* L. and crop damage in an intensive agroecosystem. *Eur J Wildl Res* 52: 245-250.
- Hirzel AH, Le Lay G, 2008. Habitat suitability modelling and niche theory. *J Appl Ecol* 45: 1272-1381.
- Hijmans RJ, van Etten J, 2012. Raster: Geographic analysis and modeling with raster data. R package version 2.0-12. Available in <http://CRAN.R-project.org/package=raster> [10 December 2012].
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A, 2005. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25(15): 1965-1978.
- Hull V, Zhang J, Zhou S, Huang J, Viña A, Liu W, Tuanmu MN, Li R, Liu D, Xu W, *et al.* 2014. Impact of livestock on giant pandas and their habitat. *J Nat Conserv* 22(3): 256-264.
- Huntley B, Berry PM, Cramer W, McDonald AP, 1995. Modelling present and potential future ranges of some European higher plants using climate response surfaces. *J Biogeog* 22(6): 967-1001.
- Jiménez-Valverde A, 2012. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecol Biogeogr* 21: 498-507.
- Keith DA, Akçakaya HR, Thuiller W, Midgley GF, Pearson RG, Phillips SJ, Regan H, Araújo M, Rebelo T, 2008. Predicting extinction risks under climate change: Coupling stochastic population models with dynamic bioclimatic habitat models. *Biol Lett* 4: 560-563.
- Keuling O, Baubet E, Duscher A, Ebert C, Fischer C, Monaco A, Podgórski T, Prevot C, Ronnenberg K, Sodeikat G, 2013. Mortality rates of wild boar *Sus scrofa* L. in central Europe. *Eur J Wildl Res* 59(6): 805-814.
- Laplace PS, 1820. *Théorie analytique des probabilités*. Courcier. Paris.
- Lawler JJ, White D Neilson, RP, Blaustein AR, 2006. Predicting climate-induced range shifts: model differences and model reliability. *Glob Chang Biol* 12: 1-17.

- Lobo JM, Jiménez-Valverde A, Real R, 2008. AUC: a misleading measure of the performance of predictive distribution models. *Glob Ecol Biogeogr* 17(2): 145-151.
- Lobo JM, Jiménez-Valverde A, Hortal J, 2010. The uncertain nature of absences and their importance in species distribution modelling. *Ecography* 33: 103-114.
- Long JL, 2003. *Introduced mammals of the world*. CSIRO Publishers, Collingwood, Australia.
- Manel S, Williams HC, Ormerod SJ, 2001. Evaluating presence-absence models in ecology: the need to account prevalence. *J Appl Ecol* 38: 921-931.
- Markina-Lamonja FA, 1998. Estudio de las poblaciones de corzo (*Capreolus capreolus* L.) y jabalí (*Sus scrofa* L.) y análisis de su explotación cinegética en el territorio histórico de Álava. Doctoral thesis. University of León, Spain.
- Marmion M, Parviainen M, Luoto, M, Heikkinen RK, Thuiller W, 2009. Evaluation of consensus methods in predictive species distribution modelling. *Divers Distrib* 15: 59-69.
- Martí R, del Moral JC (Eds.), 2003. Atlas de las aves reproductoras de España. Dirección General de Conservación de la Naturaleza-Sociedad Española de Ornitología. Madrid.
- Martínez-Jauregui M, Arenas C, Herruzo AC, 2011. Understanding long-term hunting statistics: the case of Spain (1972-2007). *Forest Syst* 20(1): 139-150.
- Massei G, Genov P, 2004. The environmental impact of wild boar. *Galemys* 16: 135-145.
- Mateo RG, Croat TB, Felicísimo AM, Muñoz J, 2010. Profile or group discriminative techniques? Generating reliable species distribution models using pseudo-absences and target-group absences from natural history collections. *Divers Distrib* 16: 84-94.
- Melis C, Szafrńska PA, Jędrzejewska B, Bartoń K, 2006. Biogeographical variation in the population density of wild boar (*Sus scrofa*) in western Eurasia. *J Biogeogr* 33: 803-811.
- Neilson RP, 1995. A model for predicting continental-scale vegetation distribution and water balance. *Ecol Appl* 5 (2): 362-385.
- Oliver W, Leus K, 2008. *Sus scrofa*. In: IUCN 2012. IUCN Red List of Threatened Species. Vers. 2012.2. Available in www.iucnredlist.org [11 February 2012].
- Palomo LJ, Gisbert J, Blanco JC, 2007. Atlas y libro rojo de los mamíferos terrestres de España. Dirección General para la Biodiversidad-SECEM-SECEMU, Madrid, 588 pp.
- Papanastasis VP, Mantzanas K, Dini-Papanastasi O, Ispikoudis I, 2009. Traditional agroforestry systems and their evolution in Greece. In: *Agroforestry in Europe: current status and future prospects* (Rigueiro-Rodríguez A, McAdam J, Mosquera-Losada MR, eds). Springer Science + Business Media B.V., Dordrecht, pp: 89-109.
- Pardini A, 2009. Agroforestry systems in Italy: traditions towards modern management. In: *Agroforestry in Europe: current status and future prospects* (Rigueiro-Rodríguez A, McAdam J, Mosquera-Losada MR, eds). Springer Science + Business Media BV, Dordrecht, pp: 255-267.
- Pearce J, Ferrier S, 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecol Model* 133: 225-245.
- Peris S, Baquedano R, Sánchez A, Pescador M, 2005. Mortalidad del jabalí (*Sus scrofa*) en carreteras de la provincia de Salamanca (NO de España). ¿Influencia de su comportamiento social? *Galemys* 17 (1-2): 13-23.
- Peterson AT, Papes M, Soberón J, 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecol Model* 213: 63-72.
- Peterson AT, Soberón J, Pearson RG, Anderson RP, Martínez-Meyer E, Nakamura M, Araújo MB, 2011. *Ecological niches and geographic distributions* (MPB-49). Princeton University Press.
- Phillips S, 2008. Response to transferability and model evaluation in ecological niche modelling. *Ecography* 31: 272-278.
- Phillips SJ, Dudík M, 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31: 161-175.
- Phillips SJ, Anderson RP, Schapire RE, 2006. Maximum entropy modeling of species geographic distributions. *Ecol Model* 190: 231-259.
- Pinzon J, Brown ME, Tucker CJ, 2005. Satellite time series correction of orbital drift artifacts using empirical mode decomposition. In: *Hilbert-Huang Transform: introduction and applications* (Huang NE & Shen SSP, eds). World Sci Publ Co. Pte. Ltd, Singapore. pp: 167-186.
- Podgórski T, Baś G, Jędrzejewska B, Sönnichsen L, Śnieżko S, Jędrzejewski W, Okarma H, 2013. Spatiotemporal behavioral plasticity of wild boar (*Sus scrofa*) under contrasting conditions of human pressure: primeval forest and metropolitan area. *J Mammal* 94: 109-119.
- Prasad AM, Iverson LR, Liaw A, 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9: 181-199.
- Rosell C, Herrero J, 2007. *Sus scrofa* Linnaeus, 1758. In: *Atlas y libro rojo de los mamíferos terrestres de España* (Palomo LJ, Gisbert J, Blanco JC, eds). Dirección General para la Biodiversidad-SECEM-SECEMU, Madrid (Spain), pp: 348-351.
- Rosell C, Fernández-Llario P, Herrero J, 2001. El jabalí (*Sus scrofa* Linnaeus, 1758). *Galemys* 13: 1-25.
- Ruiz de la Torre J, 2002. Vegetación forestal española. In: *La diversidad biológica de España* (Pineda FD, de Miguel JM, Casado MA, Montalvo J, coords.). Prentice Hall, Madrid (Spain). pp: 65-79.
- Sáez-Royuela C, Tellería JL, 1986. The increased population of the wild boar (*Sus scrofa* L.) in Europe. *Mammal Review* 16: 97-101.
- Sainz H, Sánchez de Dios R, García-Cervigón A, 2010. La cartografía sintética de los paisajes vegetales españoles: una asignatura pendiente en geobotánica. *Ecología* 23: 249-272.
- Sanderson E, Jaiteh M, Levy M, Redford K, Wannebo A, Woolmer G, 2002. The human footprint and the last of the wild. *Bioscience* 52(10): 891-904.

- Sarasa M, Sarasa JA, 2013. Intensive monitoring suggests population oscillations and migration in wild boar *Sus scrofa* in the Pyrenees. *Anim Biodivers Conserv* 36(1): 79-88.
- Schley L, Dufrêne M, Krier A, Frantz AC, 2008. Patterns of crop damage by wild boar (*Sus scrofa*) in Luxembourg over a 10-year period. *Eur J Wildl Res* 54: 589-599.
- Spencer PBS, Hampton JO, 2005. Illegal translocation and genetic structure of feral pigs in Western Australia. *J Wildl Manag* 69: 377-384.
- Stockwell D, Peters D, 1999. The GARP modelling system: Problems and solutions to automated spatial prediction. *Int J Geogr Inf Sci*. 13: 143-158.
- Suri M, Hofierka J, 2004. A new GIS-based solar radiation model and its application to photovoltaic assessments. *Trans GIS* 8:175-190.
- Taylor R, Hellgren E, Gabor T, Ilse L, 1998. Reproduction of feral pigs in Southern Texas. *J Mammal* 79 (4): 1325-1331.
- Tellería JL, Sáez-Royuela C, 1985. L'evolution demographique du sanglier (*Sus scrofa*) en Espagne. *Mammalia* 49(2): 195-202.
- Thuiller W, 2004. Patterns and uncertainties of species' range shifts under climate change. *Global Change Biol* 10(12): 2020-2027.
- Tsoar A, Allouche O, Steinitz O, Rotem D, Kadmon R, 2007. A comparative evaluation of presence-only methods for modeling and mapping studies, NA94apr15b.n11-V1g, 2.0, Global Land Cover Facility, University of Maryland, College Park, Maryland 04/15/1994.
- Tucker CJ, Pinzon JE, Brown ME, Slayback D, Pak EW, Mahoney R, Vermote EF, El Saleous N, 2005. An extended AVHRR 8-km NDVI data set compatible with MODIS and SPOT vegetation NDVI data. *Int J Remote Sens* 26(20): 4485-5598.
- UNESCO, 1977. Mediterranean forest and maquis: ecology, conservation and management. MaB technical notes 2, France, 79 pp.
- USGS, 2004. Shuttle radar topography mission, 1 arc second scene SRTM_u03_n008e004, Unfilled Unfinished 2.0, Global Land Cover Facility, University of Maryland, College Park, Maryland, February 2000.
- Vitorino FJ, Fonseca JM, 2004. Wild boar in Portugal. *Galemys* 16 (NE): 243-251.
- Von Humboldt A, Bonpland A, 1807. *Essai sur la géographie des plantes*. Facsimile reprint, Sherborn Fund no 1. Society for the Bibliography of Natural History, London.
- Whittaker RJ, Nogués-Bravo D, Araújo MB, 2007. Geographical gradients of species richness: a test of the water-energy conjecture of Hawkins *et al.* (2003) using European data for five taxa. *Glob Chang Biol* 16: 76-89.
- Wood GW, Barrett RH, 1979. Status of wild pigs in the United States. *Wildl Soc Bull* 7: 237-246.
- Woodward FI, 1987. *Climate and plant distribution*. Cambridge University Press, Cambridge, UK. 174 pp.