



# Is current climatic equilibrium a guarantee for the transferability of distribution model predictions? A case study of the spotted hyena

Sara Varela<sup>1</sup>, Jesús Rodríguez<sup>2</sup> and Jorge M. Lobo<sup>1\*</sup>

<sup>1</sup>Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, Madrid and <sup>2</sup>Investigador Gestor de Colecciones/Curator, Área de Colecciones, CENIEH, Burgos, Spain

## ABSTRACT

**Aim** The transferability of species distribution models requires that species show climatic equilibrium throughout their entire distribution area. We test this assumption for the case of the spotted hyena, *Crocuta crocuta*, a large carnivore that has shifted its distribution over the last 100,000 years from a widespread Eurasian and African range to its current geographical distribution, restricted to the Sub-Saharan areas of the African continent.

**Location** Western Eurasia and Africa.

**Methods** The current realized distribution of *C. crocuta* was estimated using presences and reliable absences as well as climatic, land-cover and anthropic variables as predictors. The potential distribution was estimated using presences and a set of pseudo-absences selected from localities outside climatically suitable localities, with only climatic variables serving as predictors. The current potential distribution was transferred to the Last Interglacial period (126,000 yr BP) using the palaeoclimatic data yielded by the GENESIS 2 general circulation model, and validated with European fossil data. Generalized linear models were used on all occasions.

**Results** Climatic variables are able to predict the current distribution of the species with high accuracy. The geographical projection of this model indicates that the species is distributed over almost all of its potential suitable area, which allows us to suppose that the current distribution of this species is in climatic equilibrium. However, the time transference of model predictions for the western Eurasian region reveals almost no suitable conditions for hyenas, despite the widespread presence of *C. crocuta* fossil remains on this continent during the Last Interglacial period.

**Main conclusions** Our results indicate that, even when model results suggest a climatic equilibrium for a species distribution, the time transferability of such models does not necessarily provide realistic results. This occurs because the current geographical range does not allow estimations of all of the environmental requirements of a species. Therefore, any model trained with current data risks underestimating the potential suitable environmental and geographical range for species in a new area or time period.

## Keywords

Africa, climatic equilibrium, *Crocuta crocuta*, distribution models, historical distribution, model transferability, spotted hyena, western Eurasia.

\*Correspondence: Jorge M. Lobo, Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, c/ José Gutiérrez Abascal, 2, 28006 Madrid, Spain.  
E-mail: mcnj117@mncn.csic.es

## INTRODUCTION

Distribution is a dynamic characteristic of species that depends on a wide variety of environmental and contingent processes operating at different spatial and temporal scales (Maurer & Taper, 2002). Despite the difficulties involved in deriving causal processes from correlations, the available data on distribution are frequently related to a variety of explanatory variables in order to identify probable influential factors (Rahbek & Graves, 2001), generate distributional hypotheses for changing scenarios (Araújo *et al.*, 2005), or establish hypotheses about the current geographical range of a species when presence data are scarce (Guisan & Zimmermann, 2000). One of the most common sources of uncertainty in species distribution models (SDMs) derives from the lack of reliable absence information. Species may be absent from model-predicted favourable places owing to factors such as biotic interactions, dispersal limitation, extinctions, or historical processes (Pulliam, 1988, 2000; Ricklefs & Schluter, 1993; Hanski, 1998). On a geographical scale, most species are unable to colonize all potentially suitable areas. The inability to recolonize former distributional ranges after the Pleistocene glaciation events (Svenning & Skov, 2007) and range-size contractions as a consequence of human activities (Channell & Lomolino, 2000) are two well-established examples of the roles played by these restrictive forces in preventing the full geographical filling of environmentally suitable areas.

Fortunately, this drawback apparently does not invalidate one of the main abilities of SDMs: to construct potential distribution hypotheses capable of estimating the total territory that can be inhabited by a species according to a set of predictor variables (see Soberón & Peterson, 2005; Peterson, 2006; Soberón, 2007; Chefaoui & Lobo, 2008; Jiménez-Valverde *et al.*, 2008). This approach is useful for suggesting the likely future distribution of invaders (Peterson, 2003), for estimating the degree of species equilibrium with current environmental variables (Svenning & Skov, 2004), and for suggesting probable distribution shifts that might occur under a changing climatic scenario (Jiménez-Valverde & Lobo, 2007a). However, these potential distribution models have two main drawbacks. First, they constitute simulations that cannot be validated using real distributional data, since it is always possible to find a favourable place in which the species is absent (i.e. the absence in a favourable place cannot be used to test the accuracy of these potential model predictions). Second, as potential distribution models are parameterized with data from the species' native range, which is influenced by biotic or historical constraints, they may actually underestimate the potential range of a species. This happens because SDMs are built with the implicit assumption that the species occurs in all environmentally suitable localities; that is, in equilibrium with the environmental conditions (Brereton *et al.*, 1995; Dormann, 2007; Jiménez-Valverde *et al.*, 2008). These two drawbacks hinder both the evaluation and the transferability of potential distribution simulations to different areas and time periods, especially if predictions are based on

explanatory variables lacking a functional basis (Austin, 2002; Dunne *et al.*, 2004; Kearney & Porter, 2004; Kearney, 2006; Vanreusel *et al.*, 2007).

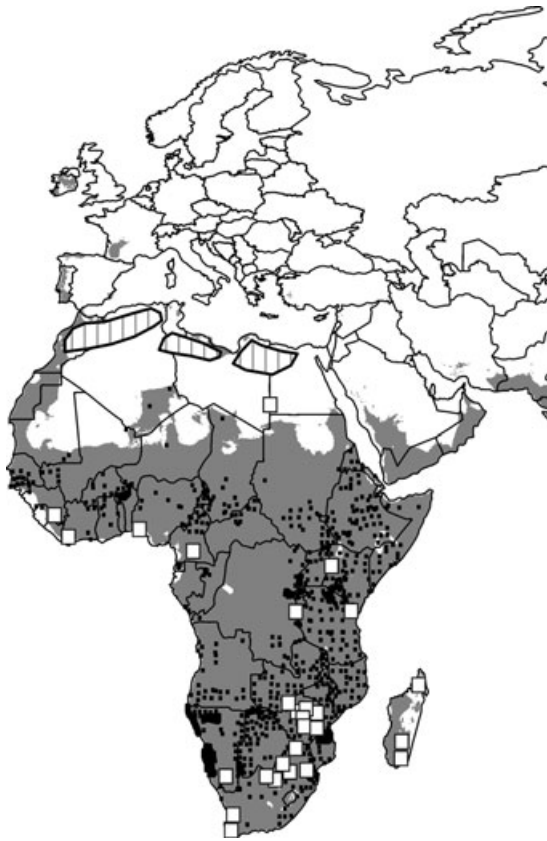
In this paper we examine the accuracy of SDMs when transferred. For that purpose, we build two different distribution models, one to estimate the realized distribution of the spotted hyena, *Crocuta crocuta* (Erxleben, 1777), and to test whether its current distribution is in equilibrium with climatic conditions, and another to identify potential environmentally suitable areas outside its current African distribution, specifically in the Western Palaearctic region, through geographical extrapolation of model results. The accuracy of these representations of potential distribution was evaluated by examining the past distribution of the spotted hyena in Europe during the Last (Eemian) Interglacial period (126,000 yr BP), the most recent period with similar climatic conditions to the current ones. The potential distribution model calibrated with current distribution and environmental data was transferred to the Last Interglacial period by using the climatic variables derived from the general circulation model (GCM) GENESIS 2 (Thompson & Pollard, 1997; Vavrus, 2001). By projecting a model fit with current environmental data onto past data, we assessed whether it is possible to predict the past distribution reliably when the current distribution apparently shows equilibrium with the environment, a basic assumption for the transfer of model predictions (Guisan & Thuiller, 2005). The specific objectives of this study were: (1) to discern the main variables accounting for the current known distribution of *C. crocuta*; (2) to perform a simulation of the climatic potential distribution of this species for the Afrotropical and Western Palaearctic regions, as well as one for the realized distribution in Africa (taking into account relatively reliable absence data); and (3) to test the accuracy of the potential model distribution predictions when they are transferred in time and space. The obtained results are used to discuss the necessary precautions needed in transferring model predictions and the frequent inability to assess the equilibrium state when the predicted distributions are based exclusively on current distribution data.

## MATERIALS AND METHODS

### Data origin

Presences and reliable absences for *C. crocuta* were selected from a number of sources. In total, 2814 cells of  $0.08^\circ \times 0.08^\circ$  were selected as presences and 13,727 cells of  $0.08^\circ \times 0.08^\circ$  as absences from a total of 766,452 cells (Western Palaearctic and Africa; see Fig. 1).

The fossil dataset was collected from bibliographic sources, using only those accurately and strictly dated for the Last Interglacial period (see Table S1 in Supporting Information). Nineteen current bioclimatic variables derived from the WorldClim interpolated map database (version 1.4, <http://www.worldclim.org>; Hijmans *et al.*, 2005) were used as climatic predictors. Furthermore, to consider the relevance of



**Figure 1** Presence and absence data locations for *Crocuta crocuta*. Presence data are plotted as black squares, and absences as white squares and as three striped polygons located in northern Africa. Pseudo-absences were selected randomly outside the multidimensional climatic envelope model (grey area).

non-climatic factors, 14 land-cover features, two measures of anthropic environmental impact, and one simple binomial variable (Barrier) were included to represent the likely barrier effect of the Sahara Desert. All variables (except the binary land-cover variables and the barrier) were standardized to means of zero and standard deviations of one to eliminate the effect of varying measurement scales. Palaeoclimatic data were downloaded from the GENESIS 2 GCM experiments (<http://ccr.aos.wisc.edu/>) at a resolution of  $4.5^\circ \times 7.5^\circ$  cells. Full information on the target species and the origins of biological and explanatory variables is provided in Appendix S1.

### Modelling methods and statistical analysis

Many different techniques are currently available to model species distributions (see Guisan & Zimmermann, 2000; Scott *et al.*, 2002), either considering only presence data (niche envelope models) or considering also absence data (group discrimination techniques). In general, group discrimination methods seem to offer more accurate predictions (Hirzel *et al.*, 2001; Brotons *et al.*, 2004; Segurado & Araújo, 2004), although the type of absence data used to calibrate the model strongly influences the type of distributional simulation obtained (see

Soberón & Peterson, 2005; Peterson, 2006; Chefaoui & Lobo, 2008). If the aim of a study is to generate distributional simulations to reflect the potential climatic distribution of a species (the set of environmentally suitable areas in which the species could occur according to a group of environmental variables), the absence information used must be located outside those environmentally suitable localities in which the species does not occur as a result of historical factors, biotic interactions, or dispersal limitation (Pulliam, 1988, 2000; Ricklefs & Schluter, 1993; Hanski, 1998). Including absences from *a priori* environmentally favourable localities will inevitably diminish the predicted range size (see Chefaoui & Lobo, 2008; Lobo *et al.*, 2008). Two types of models were generated with the available information for *C. crocuta*: (1) one for the Western Palaearctic and Afrotropical regions, using all available presence information ( $n = 2814$ ) and 10 times more pseudo-absences ( $n = 28,140$ ) selected outside the climatic conditions existing in the presence localities; and (2) another for the Afrotropical region, in which the same presences were used, but only previously well-determined absences were considered ( $n = 13,727$ ). The first approach allowed us to obtain a simulation of the potential climatic distribution of *C. crocuta* across the two biogeographical regions in which the species has been present since the beginning of the Pleistocene. The second approach offered a simulation of the realized distribution on the continent currently inhabited by this species (Africa). Pseudo-absences were selected by first tabulating each of the 19 climatic variables' maximum and minimum values for the observed distribution localities, in order to estimate the multidimensional envelope defined by the values of the locations where the species was observed (see Busby, 1986). The pseudo-absence cells were then randomly selected from the area falling outside this envelope; that is, within the generally inappropriate environmental conditions for the species according to the observed presence points (see Engler *et al.*, 2004 or Lobo *et al.*, 2006).

Presences and absences (or pseudo-absences) were logistically regressed against continuous and dummy explanatory variables using generalized linear models (GLMs; Crawley, 1993), an extension of linear models that allows for nonlinearity in the data and also a range of independent non-normal variable distributions. Recent studies recommend the use of simple GLM techniques that do not overfit when good-quality species data are used (Meynard & Quinn, 2007) and when model results are transferred to different regions and environments (Randin *et al.*, 2006). We first related the dependent variables separately to each predictor variable in order to identify the most relevant predictors. To consider possible curvilinear relationships, we selected the linear, quadratic or cubic function of each variable that accounts for the most important change in deviance with significant terms (Austin, 1980). Next, we built a complete model using all of the formerly detected statistically significant variables, selecting between competing models by a forward-backward stepwise selection procedure. The purpose of this complete model is not to estimate the parameters of each explanatory variable

correctly, but to obtain a reliable model able to forecast the distribution of the species throughout the entire territory. The STATISTICA 6.0 software package was used for these computations (StatSoft, Inc., 2003). The stability of the final model results was examined by a cross-validation procedure in which the original data were partitioned 10 times in training (75% of total data) and testing (25% of data) sets, and model accuracy measures were calculated for these testing data. Continuous probabilities derived from GLM results were converted into a presence-absence variable using the recommended threshold that minimizes the difference between sensitivity and specificity (Jiménez-Valverde & Lobo, 2007b). Sensitivity (the ratio of correctly predicted presences to the total number of presences, or true-positive rate), specificity (the true-negative rate), and the receiver operating characteristic curve (ROC; Zweig & Campbell, 1993) were used to estimate model accuracy. The area under the ROC curve (AUC) provides a single performance measure across all possible ranges of thresholds (Fielding & Bell, 1997). This score can be used to estimate the accuracy of distribution-model predictions, provided that they are carried out on the same species and extent (Lobo *et al.*, 2008). A similar procedure was used to derive the potential distribution of the species using current climatic layers from the GENESIS 2 GCM in order to transfer the predictions to the conditions characterizing the Last Interglacial period (the GENESIS 126,000 yr BP climatic scenario).

## RESULTS

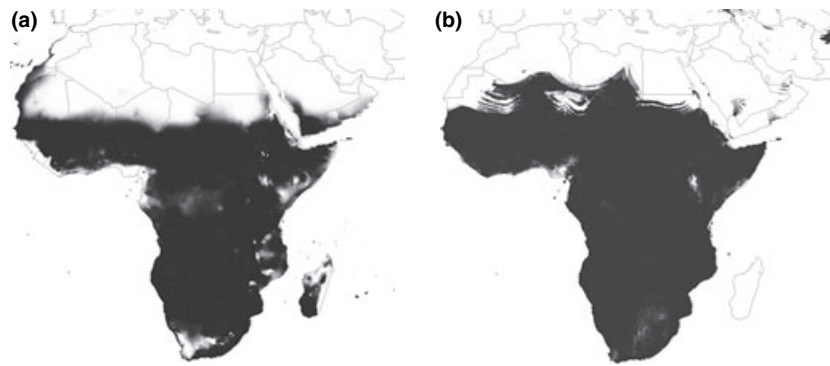
### Current potential distribution

The complete model included eight explanatory variables accounting for 96.5% of the total deviance (see coefficients in Table S2). Accuracy measures for this model are remarkably high (AUC = 0.999; sensitivity and specificity = 0.994), whereas cross-validation values (mean  $\pm$  95% CI) are  $0.9990 \pm 0.0004$  for the AUC;  $0.9936 \pm 0.0007$  for sensitivity; and  $0.9936 \pm 0.0005$  for specificity. Examination of the individual environmental variables indicated (Table 1) that those related to annual temperature variability (isothermality and temperature seasonality) explained the highest percentage of total deviance (around 80% and 70%, respectively). The probability of *C. crocuta* presence increased when isothermality values (a measure of the annual temperature range experienced on a daily basis) were higher as well as when temperature seasonality was low. In general, temperature variables seemed to be more relevant to explaining the potential species distribution than precipitation variables. In Africa, the species had high probabilities of occurrence in almost all Sub-Saharan territories, including in some high-precipitation and xerothermic localities (Fig. 2a).

Using only the two most important detected climatic variables (isothermality and temperature seasonality), a selection of pseudo-absences outside the climatic conditions

**Table 1** Deviance and change in deviance from a full model in which the number of parameters is equal to the number of observations for *Crocota crocuta* distribution in Africa and western Eurasia for each of the explanatory climatic variables used (see Hijmans *et al.*, 2005). The model was computed by selecting pseudo-absences outside the environmental range defined by these variables (potential distribution). The linear, quadratic, or cubic functions of each variable accounting for a significant change in deviance ( $P \leq 0.05$ ) were selected.

	Function	Sign	Deviance	Change in deviance	Percentage change in deviance
Full model			18,859.41		
Isothermality	Cubic	+++	3817.35	15,042.06	79.76
Temperature seasonality	Quadratic	--	5577.58	13,281.84	70.43
Mean temperature of wettest quarter	Cubic	++-	10,835.28	8024.13	42.55
Annual temperature range	Quadratic	--	11,173.01	7686.41	40.76
Mean temperature of coldest quarter	Cubic	+++	11,667.75	7191.66	38.13
Mean temperature of driest quarter	Cubic	+++	12,001.04	6858.38	36.37
Minimum temperature of coldest month	Quadratic	+-	12,102.89	6756.53	35.83
Annual mean temperature	Cubic	+++	13,375.04	5484.37	29.08
Precipitation seasonality	Cubic	++-	14,212.21	4647.20	24.64
Mean temperature of warmest quarter	Cubic	++-	15,242.14	3617.27	19.18
Maximum temperature of warmest month	Cubic	++-	15,906.70	2952.71	15.66
Precipitation of wettest month	Quadratic	+-	16,968.52	1890.89	10.03
Precipitation of wettest quarter	Quadratic	+-	17,130.96	1728.45	9.16
Precipitation of driest month	Cubic	-+-	17,184.50	1674.92	8.88
Precipitation of warmest quarter	Cubic	++-	17,438.27	1421.14	7.54
Precipitation of driest quarter	Cubic	-+-	17,480.80	1378.61	7.31
Mean diurnal range	Cubic	+++	17,687.00	1172.41	6.22
Annual precipitation	Cubic	++-	18,060.17	799.25	4.24
Precipitation of coldest quarter	Quadratic	-+	18,262.37	597.05	3.17



**Figure 2** Geographic predicted representations of the (a) potential and (b) realized distributions of *Crocuta crocuta* (see Tables S2 and S3). The probability maps are in a white (0) to black (1) scale. The striped pattern in the north is a consequence of the interpolation procedure used for the seasonal precipitation variable. IUCN estimations of the distribution of the spotted hyena identify Central African evergreen rain forest as an absence area. However, as our dataset includes presence points within evergreen rain forests in Equatorial Guinea, the model does not identify the areas with this type of habitat as potentially unsuitable.

**Table 2** Deviance and change in deviance from a full model in which the number of parameters is equal to the number of observations for *Crocuta crocuta* distribution in Africa for each explanatory climatic and non-climatic variable used (see Hijmans *et al.*, 2005). The model was computed using reliable absences (realized distribution). The linear, quadratic, or cubic functions of each variable accounting for a significant change in the deviance ( $P \leq 0.05$ ) were selected. Land-cover variables came from the Global Land Cover Characterization of the US Geological Survey (version 2; see <http://edcns17.cr.usgs.gov/glcc/>), and the distance to populated areas (urban distance) and distance to wilderness areas (wilderness distance) were created from the Gridded Population of the World project (version 3; see <http://sedac.ciesin.org/gpw>) and the World Wilderness Areas dataset of the United Nations Environment Programme (see <http://www.grid.unep.ch>).

	Function	Sign	Deviance	Change in deviance	Percentage change in deviance
Full model			15,088.06		
Climatic variables					
Isothermality	Cubic	+++	1440.18	13,647.88	90.45
Temperature seasonality	Cubic	+++	2560.68	12,527.38	83.03
Annual temperature range	Cubic	+++	5442.38	9645.68	63.93
Mean temperature of coldest quarter	Cubic	+++	5878.14	9209.92	61.04
Mean temperature of wettest quarter	Cubic	+++	6520.79	8567.27	56.78
Precipitation of warmest quarter	Cubic	+++	6867.08	8220.98	54.47
Minimum temperature of coldest month	Cubic	+++	6988.50	8099.56	53.68
Precipitation of wettest month	Quadratic	++	9532	9327.41	49.46
Precipitation of wettest quarter	Quadratic	++	9637.34	5450.72	36.13
Mean temperature of driest quarter	Quadratic	--	9787.3	9072.11	48.10
Maximum temperature of warmest month	Cubic	+++	9857	5231.45	34.67
Precipitation seasonality	Linear	+	10,811.63	4276.43	28.34
Annual precipitation	Cubic	+++	10,940.4	4147.64	27.49
Mean diurnal range	Cubic	+++	12,405.08	2682.98	17.78
Mean temperature of warmest quarter	Quadratic	++	12,650.82	2437.24	16.15
Precipitation of coldest quarter	Cubic	+++	13,224	1863.99	12.35
Annual mean temperature	Quadratic	++	13,462.4	5397.01	28.62
Precipitation of driest month	Quadratic	++	14,026.33	1061.73	7.04
Precipitation of driest quarter	Cubic	+++	14,080.4	1007.67	6.68
Non-climatic variables					
Barrier		-	1141.50	13,946.56	92.43
Urban distance	Quadratic	++	10,420.19	4667.87	30.94
Land use – Bare ground		+	12,239.09	2848.97	18.88
Land use – Woodland		-	13,468.71	1619.35	10.73
Wilderness distance	Quadratic	++	13,506.22	1581.84	10.48
Land use – Wooded grassland		-	13,973.03	1115.03	7.39
Land use – Evergreen broadleaf forest		-	14,688.24	399.82	2.65
Land use – Grassland		-	14,826.23	261.83	1.73
Land use – Closed shrubland		-	14,944.76	143.30	0.95
Land use – Cropland		-	15,058.18	29.88	0.20
Land use – Deciduous broadleaf forest		-	15,060.45	27.61	0.18

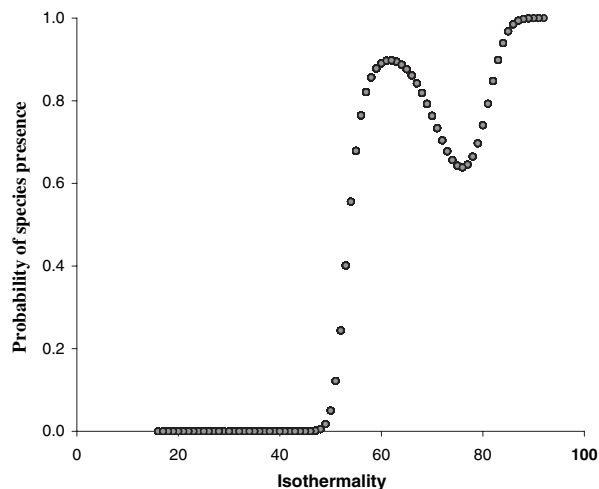


existing in the presence localities generated a similar potential distribution (not shown).

### Current realized distribution

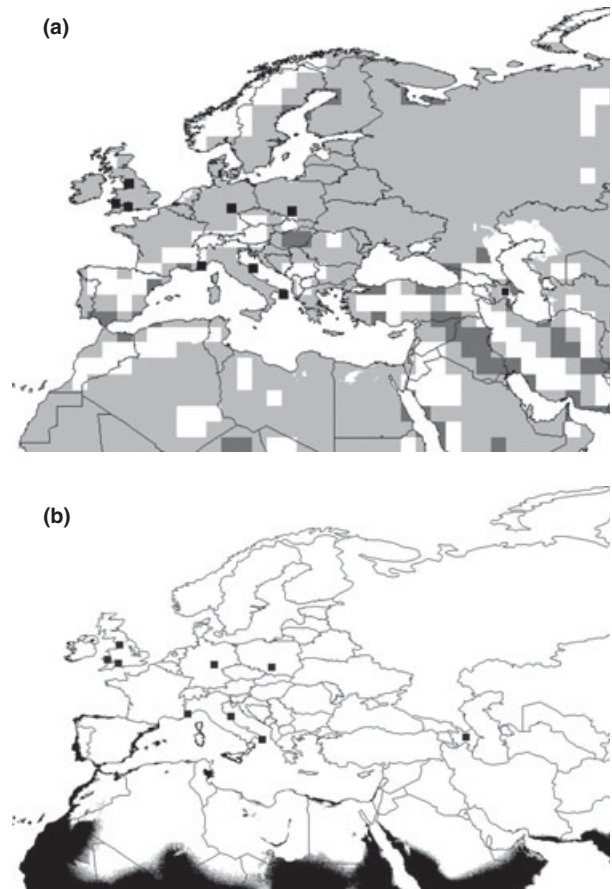
The climatic variables related to temperature variability were also able to explain a high percentage of the total deviance (see Table 2). The Sahara Desert barrier variable accounted for 92% of the total variability; anthropic variables (distance to populated places and wilderness areas) accounted for 37%; and land cover accounted for 31%. Adding all these significant non-climatic variables to the complete climatic model increased the explained variability by only 1%. Land-cover and anthropic variables did not seem to be particularly relevant for explaining the general pattern of presence and absence on a global scale, although it will be necessary to study the relationship between these variables and the species distribution at a local scale. The complete model generated with all of the former significant variables included nine explanatory variables (see coefficients in Table S3), accounting for 98% of total deviance. Accuracy measures were also high (AUC = 0.979; sensitivity = 0.95; and specificity = 1), and cross-validation values (mean  $\pm$  95% CI) were  $0.9992 \pm 0.0007$  for the AUC;  $0.9963 \pm 0.0010$  for sensitivity; and  $0.9964 \pm 0.0008$  for specificity.

Apart from the Sahara Desert barrier, the five most important variables were the same as those in the potential climatic model. Again, isothermality and temperature seasonality were the two most explicative variables, accounting for



**Figure 3** Relationship between isothermality (the ratio between mean diurnal range and temperature annual range) and the predicted probability of presence for the realized distribution. *Crocota crocuta* has a low probability of occurrence when temperature scores exhibit strong variation throughout the year (deserts and mediterranean-type climate areas); the species has a high probability of occurrence in environments with more stable temperature conditions. The shift in the slope (when isothermality values are around 80) approximately coincides with the evergreen rain forest data points.

90% and 83% of total variability, respectively. *Crocota crocuta* was predicted to be absent when isothermality scores are  $< 50$  (Fig. 3), representing places with a high variation in temperature throughout the year. Thus, mediterranean-type climate and desert areas were not covered by the model (Fig. 2b). Predicted values also declined when isothermality values were around 80, as observed in the Congo and Central African tropical rain forests. In West Africa, coastal areas with high values of precipitation and minimum temperature, from Guinea to Gabon, were predicted as unsuitable for *C. crocuta*, as were the rain forest areas of Gabon and Congo. The mediterranean-type climate areas of South Africa and the



**Figure 4** (a) Differences between the Last Interglacial (126,000 yr BP) average annual temperature and the current values for Europe according to three categories: white, current colder temperatures ( $\leq 2^\circ\text{C}$ ); dark grey, current warmer temperatures ( $\geq 2^\circ\text{C}$ ); grey, similar temperatures ( $\pm 2^\circ\text{C}$ ). Climate layers extracted from the GENESIS 2 general circulation model (Thompson & Pollard, 1995, 1997; P. Batra, unpublished data). The correlation between the two temperature values for all  $2^\circ$  cells is statistically significant ( $r = 0.97$ ;  $P < 0.0001$ ) with a regression slope of 1.006. Fossil sites are represented as black squares. (b) The predicted suitable area in Europe (in black) according to the potential prediction model generated using current African data. Note that favourable locations are restricted to the coast of the Iberian Peninsula and that fossil sites appear outside the predicted potential distribution.

areas with low values of isothermality that surround Lake Malawi and the coast of Mozambique, as well as the desert coast of Djibouti and Eritrea, were other areas where the model predicted the species to be absent.

### Model transferability

The geographic representation of the aforementioned potential distribution showed that the Western Palaearctic region currently does not seem to exhibit the climatic conditions necessary for the presence of *C. crocuta* (Fig. 4). In spite of the similarity between current and Last Interglacial temperatures, the fossil presence is outside supposedly climatically favourable European areas, as derived from current distribution locations (Fig. 4b).

Using current GENESIS 2 GCM climatic variables, we were able to explain 67% of the total deviance. The potential distribution thus generated (Fig. 5a) was transferred to the Last Interglacial without success, since fossil records were located in cells with zero or very low suitability values (mean  $\pm$  SD;  $0.032 \pm 0.064$ ; Fig. 5b).

## DISCUSSION

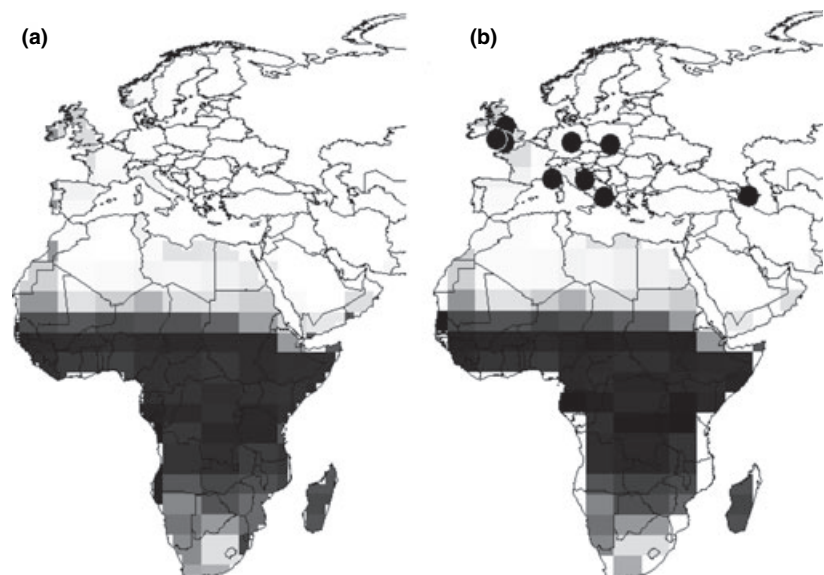
### Ecological implications of model results

Among the climatic variables, those related to temperature stability during the year had the highest predictive capacity in both models. Spotted hyenas seem to avoid areas with regular precipitation throughout the year (evergreen forests) as well as those with high values of temperature seasonality (mediterranean-type climate and desert zones). It is likely that temperature stability is not linked to any physiological constraint, but to the species' response to certain habitat structures (Mills & Hofer, 1998). The barrier variable (the Sahara Desert) also had a high explanatory capacity to

predict the realized distribution, suggesting that this large desert could be conditioning the current northern dispersal of *C. crocuta*. Interestingly, genetic studies (Rohland *et al.*, 2005) indicate that the last connection between the European and African populations occurred during the Last Interglacial, a period of intensified monsoons and increased precipitation. As a consequence of this increase in humidity, 97% of North Africa was covered by vegetation, mainly grasses (van Andel & Tzedakis, 1996; Montoya, 2007). The Sahara Desert also experienced other much less arid periods during the late Pleistocene and Holocene (Issar, 2003; Drake & Bristow, 2006). In particular, during the period from 10,300 to 9500 yr BP, when the European population went extinct, the Sahara was significantly more humid than it is at present (Lézine, 2007). Studies on the viability of the spotted hyena metapopulations highlight the possibility that isolation could be a key factor for the extinction of this species (Mills & Hofer, 1998). If this is the case, further investigation will be necessary to determine why the African spotted hyena populations failed to spread northwards and rescue the European populations from extinction.

### Climatic equilibrium and transferability

Both potential and realized distribution models of *C. crocuta* were highly accurate, describing around 96% of the total variability and offering extremely high cross-validation results. According to our model, climate is the main conditioning factor capable of explaining the current distribution of this species. The predicted area for the potential distribution in Africa is only 8% higher than the realized area. Hence, our results suggest that the current *C. crocuta* distribution is highly conditioned by climatic variables and that the species is currently present throughout the major portion of its climatically suitable area. Thus, the assumption of equilibrium with the climatic conditions is apparently not violated, potentially



**Figure 5** (a) Predicted suitable area for *Crocuta crocuta* using current data on annual temperature range and annual precipitation derived from the GENESIS 2 general circulation model climatic data ( $4.5 \times 7.5$  resolution), and (b) projected potential distribution using the Last Interglacial climate scenario (GENESIS 2 GCM). Western Eurasia is not predicted as a suitable area for the species, and the probability of occurrence differs only slightly from zero in southern Europe. *Crocuta crocuta* fossil locations during the Last Interglacial (black circles) are located outside the predicted suitable areas for the species. Probabilities oscillate between zero (white) and one (black).

allowing this model to be transferred in space and time. Transferability of the model results to western Eurasia predicts almost no suitable area for the species, and the potential model built using GENESIS 2 climatic layers showed similar results, again excluding western Eurasia as a favourable area for the species during the Last Interglacial period. However, fossil data available for a climatically similar Pleistocene period clearly indicate that this species inhabited regions with climatic conditions very different from those experienced currently. An overwhelming amount of fossil evidence indicates that *C. crocuta* inhabited the European Pleistocene under both glacial and interglacial conditions (Werdelin & Solounias, 1991; Testu, 2006), coexisting both with cold-adapted species, such as *Coelodonta antiquitatis* (woolly rhinoceros) and *Mammuthus primigenius* (woolly mammoth) (Altuna & Mariezkurrena, 2000), and with warm-adapted species, such as *Dama dama* (fallow deer) and *Sus scrofa* (wild boar) (Bologna *et al.*, 1994). The failure of the models to predict the palaeodistribution of the species suggests that current distribution data do not represent the entire climatic range where spotted hyenas are able to live. Similar results could be obtained for many other large mammals currently extinct in Europe but present in Africa (Barnosky *et al.*, 2004).

When seeking to minimize the difference between the estimated population parameters and the sample statistics, it is universally agreed that the use of statistical methods to draw inferences requires that the sample from which relationships among variables will be inferred be representative of the population described (Zar, 1999). In the case of species distribution models (SDMs), such samples frequently comprise the available information on the presence of the species (i.e. species occurrences), which is a subset of the total number of inhabited localities. The relationships that SDMs aim to infer constitute the 'niche' or set of environmental conditions that would allow any species' population to attain a positive rate of demographic increase in the absence of immigration (Soberón, 2007). The subsequent geographical representation of such an inferred niche could provide reliable hypotheses on the potential distribution of species, provided that the inferred relationships represent all the combinations of environmental conditions where the species can persist. Hence, we sought to determine whether it is possible to attain reliable knowledge of the niche using only distributional information. Darwin (1859, p. 346) himself wrote in the *Origin of Species* that '...neither the similarity nor the dissimilarity of the inhabitants of various regions can be accounted for by climatal and other physical conditions'. That is, the environmental conditions of the localities with data on the presence of the species will not provide information on the whole spectrum of suitable environmental conditions. The distribution of a species is (almost always) the consequence of the interplay of a large number of current and past processes, many of them exerting contingent effects on species populations (Ricklefs, 2004). Most species distributions are not in equilibrium with current environmental conditions (Svenning & Skov, 2004), owing to the impact of past environmental

conditions, dispersal limitations and the particular history of regions and organisms. Such a lack of equilibrium hinders the precise estimation of potential distributions/niches from current distribution data (Jiménez-Valverde *et al.*, 2008). Recently, the lack of success in the transferability of the results of SDMs has been presented as evidence of niche shifts (Broennimann *et al.*, 2007; Fitzpatrick *et al.*, 2007). Such an argument neglects both the role played by the phenotypic plasticity of species (Miner *et al.*, 2005) and the fact that a species can be predicted to be absent from areas with suitable environmental conditions simply because these conditions do not exist within its native distribution area. On these premises, the transference of the set of suitable environmental conditions inferred from the sample used to another time or region is subject to an unknown degree of uncertainty, even if SDMs are developed using an adequate resolution and the presence data used cover the species' entire geographical distribution. SDMs developed from current distributional information may underestimate the set of environmental conditions where a species can occur (Dormann, 2007). Extreme caution is thus mandatory if the aim is transferring the results of merely correlative models over both space and time. The apparent equilibrium with climate that might be estimated from current distributional information does not guarantee the transferability of model predictions. Rather, to estimate the potential distributions of species it is advisable to use distributional data coming from different regions and time periods (Mau-Crimmins *et al.*, 2006; Nogués-Bravo *et al.*, 2008) and/or functional and physiological data (Kearney *et al.*, 2008).

## ACKNOWLEDGEMENTS

We thank Pablo Sastre Olmos and Joaquín Hortal for their help in the management of geographical data and their valuable comments. S.V. was supported by a doctoral fellowship from the Fundación Atapuerca. J.R. and S.V. were supported by the Spanish MEC project (CGL2006-13532-C03/BTE).

## REFERENCES

- Altuna, J. & Mariezkurrena, K. (2000) Macromamíferos del yacimiento de Labeko Koba (Arrasate, País Vasco). *Munibe*, **52**, 107–181.
- van Andel, T.H. & Tzedakis, P.C. (1996) Palaeolithic landscapes of Europe and environs, 150,000–25,000 years ago: an overview. *Quaternary Science Reviews*, **15**, 481–500.
- Araújo, M.B., Pearson, R.G., Thuiller, W. & Erhard, M. (2005) Validation of species–climate impact models under climate change. *Global Change Biology*, **11**, 1504–1513.
- Austin, M.P. (1980) Searching for a model for use in vegetation analysis. *Vegetatio*, **42**, 11–21.
- Austin, M.P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101–118.



- Barnosky, A.D., Koch, P.L., Feranec, R.S., Wing, S.L. & Shabel, A.B. (2004) Assessing the causes of late Pleistocene extinctions on the continents. *Science*, **306**, 70–75.
- Bologna, P., Di Stefano, G., Manzi, G., Petronio, C. & Sardella, R. (1994) Late Pleistocene mammals from the melpignano (LE) “Ventarole”: preliminary analysis and correlations. *Bolletino della Societa Paleontologica Italiana*, **32**, 265–274.
- Brereton, R., Bennett, S. & Mansergh, I. (1995) Enhanced greenhouse climate change and its potential effect on selected fauna of south-eastern Australia: A trend analysis. *Biological Conservation*, **72**, 339–354.
- Broennimann, O., Treier, U.A., Müller-Schärer, H., Thuiller, W., Peterson, A.T. & Guisan, A. (2007) Evidence of climatic niche shift during biological invasion. *Ecology Letters*, **10**, 701–709.
- Brotons, L., Thuiller, W., Araújo, M.B. & Hirzel, A.H. (2004) Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, **27**, 437–448.
- Busby, J.R. (1986) A biogeographical analysis of *Notophagus cunninghamii* (Hook.) in south-eastern Australia. *Australian Journal of Ecology*, **11**, 1–7.
- Channell, R. & Lomolino, M.V. (2000) Trajectories to extinction: spatial dynamics of the contraction of geographical ranges. *Journal of Biogeography*, **27**, 169–179.
- Chefaoui, R.M. & Lobo, J.M. (2008) Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling*, **210**, 478–486.
- Crawley, M.J. (1993) *GLIM for ecologists*. Blackwell Scientific Publications, Oxford.
- Darwin, C.R. (1859) *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. John Murray, London.
- Dormann, C.F. (2007) Promising the future? Global change projections of species distributions. *Basic and Applied Ecology*, **8**, 387–397.
- Drake, N.A. & Bristow, C. (2006) Shorelines in the Sahara: geomorphological evidence for an enhanced monsoon from paleolake Megachad. *The Holocene*, **16**, 901–911.
- Dunne, J.A., Saleskamar, S.R., Fisher, M.L. & Harte, J. (2004) Integrating experimental and gradient methods in ecological climate change research. *Ecology*, **85**, 904–916.
- Engler, R., Guisan, A. & Rechsteiner, L. (2004) An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology*, **41**, 263–274.
- Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.
- Fitzpatrick, M.C., Weltzin, J.F., Sanders, N.J. & Dunn, R.R. (2007) The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? *Global Ecology and Biogeography*, **16**, 24–33.
- Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.
- Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.
- Hanski, I. (1998) Metapopulation dynamics. *Nature*, **396**, 41–49.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.
- Hirzel, A.H., Helfer, V. & Métral, F. (2001) Assessing habitat-suitability models with a virtual species. *Ecological Modelling*, **145**, 111–121.
- Issar, A.S. (2003) *Climate changes during the Holocene and their impact on hydrological systems*. Cambridge University Press, Cambridge.
- Jiménez-Valverde, A. & Lobo, J.M. (2007a) Potential distribution of the endangered spider *Macrothele calpeiana* (Araneae, Hexathelidae) and the impact of climate warming. *Acta Zoologica Sinica*, **53**, 865–876.
- Jiménez-Valverde, A. & Lobo, J.M. (2007b) Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologica*, **31**, 361–369.
- Jiménez-Valverde, A., Lobo, J.M. & Hortal, J. (2008) Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distributions*, **14**, 885–890.
- Kearney, M. (2006) Habitat, environment and niche: what are we modelling? *Oikos*, **115**, 186–191.
- Kearney, M. & Porter, W.P. (2004) Calculating and mapping the fundamental niche: physiology, climate and the distribution of nocturnal lizards across Australia. *Ecology*, **85**, 3119–3131.
- Kearney, M., Phillips, B.L., Tracy, C.R., Christian, K.A., Betts, G. & Porter, W.P. (2008) Modelling species distributions without using species distributions: the cane toad in Australia under current and future climates. *Ecography*, **31**, 423–434.
- Lézine, A.M. (2007) Pollen records. *Postglacial Africa. Encyclopedia of Quaternary Science* (ed. by S.A. Elias), pp. 2681–2699. Elsevier, Amsterdam.
- Lobo, J.M., Verdú, J.R. & Numa, C. (2006) Environmental and geographical factors affecting the Iberian distribution of flightless *Jekelius* species (Coleoptera: Geotrupidae). *Diversity and Distributions*, **12**, 179–188.
- Lobo, J.M., Jiménez-Valverde, A. & Real, R. (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, **17**, 145–151.
- Mau-Crimmins, T.M., Schussman, H.R. & Geiger, E.L. (2006) Can the invaded range of a species be predicted sufficiently using only native-range data? Lehmann lovegrass (*Eragrostis lehmanniana*) in the southwestern United States. *Ecological Modelling*, **193**, 736–746.
- Maurer, B.A. & Taper, M.L. (2002) Connecting geographical distributions with population processes. *Ecology Letters*, **5**, 223–231.

- Meynard, C.N. & Quinn, J.F. (2007) Predicting species distributions: a critical comparison of the most common statistical models using artificial species. *Journal of Biogeography*, **34**, 1455–1469.
- Mills, G. & Hofer, H. (1998) *Status survey and conservation action plan. Hyaenas*. IUCN/SSC Hyaena specialist group. International Union for Conservation of Nature, Gland, Switzerland and Cambridge.
- Miner, B.G., Sultan, S.E., Morgan, S.G., Padilla, D.K. & Relyea, R.A. (2005) Ecological consequences of phenotypic plasticity. *Trends in Ecology and Evolution*, **20**, 685–692.
- Montoya, M. (2007) The Last Interglacial. *Encyclopedia of Quaternary Science* (ed. by S.A. Elias), pp. 1916–1927. Elsevier, Amsterdam.
- Nogués-Bravo, D., Rodríguez, J., Hortal, J., Batra, P. & Araújo, M.B. (2008) Climate change, humans, and the extinction of the Woolly Mammoth. *PLoS Biology*, **6**, 685–692.
- Peterson, A.T. (2003) Predicting the geography of species' invasions via ecological niche modeling. *The Quarterly Review of Biology*, **78**, 419–433.
- Peterson, A.T. (2006) Uses and requirements of ecological niche models and related distributional models. *Biodiversity Informatics*, **3**, 59–72.
- Pulliam, H.R. (1988) Sources, sinks and population regulation. *The American Naturalist*, **132**, 652–661.
- Pulliam, H.R. (2000) On the relationship between niche and distribution. *Ecology Letters*, **3**, 349–361.
- Rahbek, C. & Graves, G.R. (2001) Multiscale assessment of patterns of avian species richness. *Proceedings of the National Academy of Sciences USA*, **98**, 4534–4539.
- Randin, C.F., Dirnbock, T., Dullinger, S., Zimmermann, N.E., Zappa, M. & Guisan, A. (2006) Are niche-based species distribution models transferable in space? *Journal of Biogeography*, **33**, 1689–1703.
- Ricklefs, R.E. (2004) A comprehensive framework for global patterns in biodiversity. *Ecology Letters*, **7**, 1–15.
- Ricklefs, R.E. & Schluter, D. (eds) (1993) *Species diversity in ecological communities. Historical and geographical perspectives*. University of Chicago Press, Chicago, IL.
- Rohland, N., Pollack, J.L., Nagel, D., Beauval, C., Airvaux, J., Pääbo, S. & Hofreiter, M. (2005) The population history of extant and extinct *hyaenas*. *Molecular Biology and Evolution*, **22**, 2435–2443.
- Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A. & Samson, F.B. (2002) *Predicting species occurrences. Issues of accuracy and scale*. Island Press, Washington, DC.
- Segurado, P. & Araújo, M.B. (2004) An evaluation of methods for modelling species distributions. *Journal of Biogeography*, **31**, 1555–1568.
- Soberón, J. (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, **10**, 1115–1123.
- Soberón, J. & Peterson, A.T. (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, **2**, 1–10.
- StatSoft, Inc. (2003) *STATISTICA (data analysis software system and user's manual)*, version 6.1. StatSoft Inc., Tulsa, OK.
- Svenning, J.C. & Skov, F. (2004) Limited filling of the potential range in European tree species. *Ecology Letters*, **7**, 565–573.
- Svenning, J.C. & Skov, F. (2007) Could the tree diversity pattern in Europe be generated by postglacial dispersal limitation? *Ecology Letters*, **10**, 453–460.
- Testu, A. (2006) *Etude paléontologique et biostratigraphique des Felidae et Hyaenidae pléistocènes de l'Europe méditerranéenne*. Centre Européen de Recherches Préhistoriques de Teutavel, Perpignan.
- Thompson, S.L. & Pollard, D. (1995) A global climate model (GENESIS) with a Land-Surface transfer scheme (LXS) Part I: present climate simulation. *Journal of Climate*, **8**, 732–761.
- Thompson, S.L. & Pollard, D. (1997) Greenland and Antarctic mass balances for present and doubled atmospheric CO<sub>2</sub> from the GENESIS Version-2 global model. *Journal of Climate*, **10**, 871–900.
- Vanreusel, W., Maes, D. & van Dyck, H. (2007) Transferability of species distribution models: a functional habitat approach for two regionally threatened butterflies. *Conservation Biology*, **21**, 201–212.
- Vavrus, S. (2001) A comparison of feedback processes in the Arctic during past and future warm climates. *Preprints, 6th Conference on Polar Meteorology and Oceanography*, pp. 5–8. American Meteorological Society, San Diego, CA.
- Werdelin, L. & Solounias, N. (1991) *The Hyaenidae: taxonomy, systematics and evolution*. Fossils and Strata, Universitetsforlaget, Oslo.
- Zar, J.H. (1999) *Biostatistical analysis*. Prentice Hall, Upper Saddle River, NJ.
- Zweig, M.H. & Campbell, G. (1993) Receiver-operating characteristics (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, **39**, 561–577.

## SUPPORTING INFORMATION

The following supporting information is available for this article online:

**Appendix S1** Basic information on the target species, origin of biological data, explanatory variables and palaeoclimatic data.

**Table S1** List of the European Last Interglacial localities with fossil data of *Crocota crocuta*.

**Table S2** Parameter estimates from the final potential-distribution generalized linear model.

**Table S3** Parameter estimates from the final realized-distribution generalized linear model.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

**BIOSKETCHES**

**Sara Varela** is a palaeontologist and PhD student at the Museo Nacional de Ciencias Naturales (Spain) interested in the past processes able to explain the current distributions of mammal species and especially in estimating the effectiveness of model simulations carried out with modern and past biological data.

**Jesús Rodríguez** is a palaeoecologist with research interests in macroecology and the factors influencing the assemblage of ecological communities. His research is focused on the evolution of Pleistocene European mammalian communities in relation to environmental and historical factors.

**Jorge M. Lobo** is interested in the patterns and processes of species distributions from a macroecological perspective and in the management of biodiversity and conservation biogeography. He is a specialist in the biogeography and ecology of dung beetles (Scarabaeoidea).

---

Editor: Jens-Christian Svenning