Ann. Zool. Fennici 45: 200–210 ISSN 0003-455X (print), ISSN 1797-2450 (online)
Helsinki 26 June 2008 © Finnish Zoological and Botanical Publishing Board 2008

Challenging species distribution models: the case of *Maculinea nausithous* in the Iberian Peninsula

Alberto Jiménez-Valverde[1,*], José F. Gómez[1], Jorge M. Lobo[1], Andrés Baselga[1] & Joaquín Hortal[1,2]

[1)] *Departmento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales (CSIC). C/José Gutiérrez Abascal 2, ES-28006 Madrid, Spain (*corresponding author's e-mail: mcnaj651@mncn.csic.es)*
[2)] *NERC Centre for Population Biology, Division of Biology, Imperial College London. Silwood Park Campus, Ascot, Berkshire SL5 7PY, UK*

Received 7 May 2007, revised version received 22 Nov. 2007, accepted 26 Nov. 2007

Jiménez-Valverde, A., Gómez, J. F., Lobo, J. M., Baselga, A. & Hortal, J. 2008: Challenging species distribution models: the case of *Maculinea nausithous* in the Iberian Peninsula. — *Ann. Zool. Fennici* 45: 200–210.

*Maculinea nausithous*is the most endangered Iberian butterfly, being included in the Bern and Habitat directives and in the Spanish Red Book for invertebrates. We map its potential distribution in the Iberian Peninsula, identifying potentially suitable areas where to focus future surveys. Using presence/absence data and a set of environmental variables, the distribution of *M. nausithous* was modelled using Generalized Additive Models. Model results were filtered using land cover data to identify the anthropogenic grasslands inhabited by the species. The model obtained indicates that some Iberian areas that currently do not support *M. nausithous* populations could be environmentally suitable for the species. The reasons for these absences are discussed. Indeed, field surveys identified possible sources of uncertainty at finer scale, highlighting that deficiencies in land use GIS data might constitute an important source of error, able to explain both commission and omission errors (i.e., over and underpredictions) of the model.

Introduction

Species distribution modelling relates species records with a set of variables, building a mathematical function that can be interpolated or extrapolated to areas with absence of information on the focus species (Guisan & Zimmermann 2000). These techniques are being increasingly used in solving a variety of problems (*see* Peterson 2006 for a review). They can be used to define the environmental adaptations of species (e.g. Chefaoui *et al*. 2005, the 'Grinnellian niche' according to Soberón 2007), to locate new populations of rare and/or endangered taxa (e.g. Bourg *et al*. 2005, Peterson & Papeş 2006, Jiménez-Valverde & Lobo 2007a) or to identify areas with the environmental requirements of the focus species where to direct future survey efforts (Raxworthy *et al*. 2003). Other uses in conservation include highlighting areas for translocations and application for reserve system design (Peterson 2006).

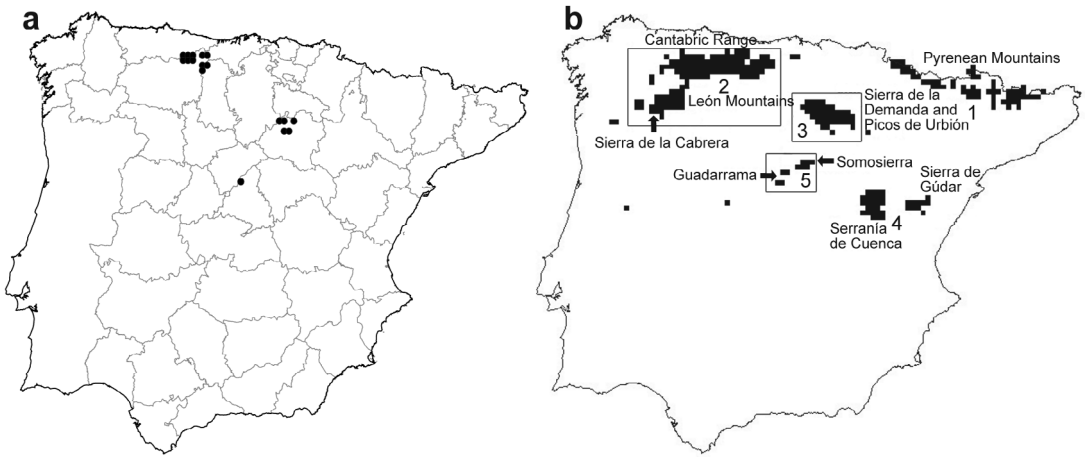

Fig. 1. (**a**) Observed and (**b**) potential distribution of *Maculinea nausithous* in the Iberian Peninsula. Black dots in **a** indicate recorded occurrences (data obtained from García-Barros *et al.* 2004), and black areas in **b** indicate locations where the probability of occurrence goes beyond the optimal threshold for *M. nausithous* (1 = Pyrenees; 2 = Cantabric Mountains; 3 = Sierra de la Demanda and Picos de Urbión; 4 = Serranía de Cuenca (left) and Sierra de Gúdar (right); 5 = Central System). Areas inside squares in **b** are those with known occurrences of the butterfly.

Maculinea is a Palaeartic butterfly genus comprising six species, four of them present in the Iberian Peninsula. *Maculinea* species are declining in most European countries as a result of habitat degradation (Wynhoff 1998, Munguira & Martín 1999). Their dependence on open habitats maintained by traditional land uses (grazing, mowing) puts their populations under extinction risk when changes in these uses occur (Thomas 1980, Munguira & Martín 1993, 1999). As a result, *Maculinea* species are listed as endangered in almost every European country having Red Lists (Munguira & Martín 1999).

Maculinea nausithous is the most endangered butterfly species in the Iberian Peninsula. Its status is vulnerable according to the Red Book of Invertebrates of Spain (Munguira *et al*. 2006), and is also included in the Bern and Habitat directives (two European directives aiming to preserve target species and habitats through national policies and the cooperation of member states; Munguira & Martín 1994, Munguira *et al*. 2006). *Maculinea nausithous* requires highly specialized habitats (Munguira *et al*. 2006). Its typical habitat is wet grasslands with dense and relatively tall vegetation (30–100 cm high), surrounded by forest, maintained by grazing or mowing, and with the presence of the host plant *Sanguisorba officinalis* (García-Barros *et al*. 1993, Munguira *et al*. 2006). This is the only known host plant for the species (Munguira *et al*. 1997, Munguira *et al*. 2006), which also grows in wet grasslands and river banks. Similar to other *Maculinea* species, *M. nausithous* is a parasite of *Myrmica* ants (Thomas *et al*. 1989, Elmes *et al*. 1998). Two host species have been reported for *M. nausithous*, *Myrmica rubra* and, less commonly, *Myrmica scabrinodis* (Munguira & Martín, 1999). The presence of *M. nausithous* has been confirmed only at 17 localities in the Iberian Peninsula, divided in three isolated core areas: the Cantabric Mountains, and Soria and Madrid provinces (Fig. 1a).

Here, we describe the potential distribution of *Maculinea nausithous* in the Iberian Peninsula with the aim of seeking potentially unknown populations. Thus, the two main goals of this study are: (1) to elaborate a potential distribution model for the species in the Iberian Peninsula highlighting areas that could potentially harbour this endangered and rare species, and (2) to identify those areas environmentally suitable for the butterfly but insufficiently sampled. To do this, we use environmental data on the factors known to affect the species; given the dependence of *M. nausithous* on its host plant and ant species, we take into account also their recorded and potential distributions. We use the results to stress some of the precautions and considerations that must be necessarily taken into account when

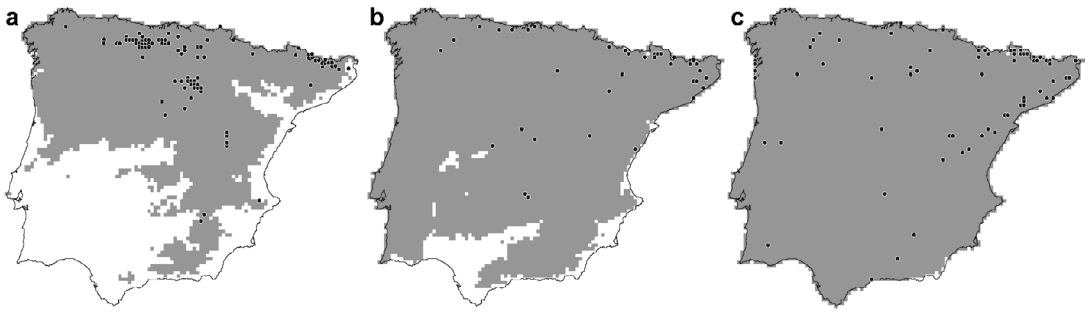

Fig. 2. Observed and potential distributions of the three *Maculinea nausithous* host species in the Iberian Peninsula: (**a**) *Sanguisorba officinalis*, (**b**) *Myrmica rubra* and (**c**) *Myrmica scabrinodis*. Black dots indicate observed occurrences of the three species, obtained from Anthos Project (www.programanthos.org) in the case of *S. officinalis*, and from www.hormigas.org in the case of the two *Myrmica* species. Areas with probability of occurrence higher than the selected optimal threshold score are shown in grey.

distribution models of specialist species are built for conservation purposes when reliable information about their absence is lacking.

Material and methods

Distribution data

Data on the presence of *Maculinea nausithous* in 100-km^2 UTM squares were extracted from García-Barros *et al*. (2004). In total, the species has been recorded only in 17 UTM squares (herein, presence points) in the Iberian Peninsula (Fig. 1a). The distribution of *Sanguisorba officinalis* (the host plant) (90 presence points) was obtained from the Anthos Project webpage (http://www.programanthos.org; Fig. 2a). In the same way, the distributions of *Myrmica rubra* (37 presence points) and *Myrmica scabrinodis* (68 presence points) were obtained from http://www.hormigas.org (Fig. 2b and c).

Environmental data

Four environmental variables were used to characterize the environmental conditions in each of the 100 km^2 Iberian Peninsula UTM squares: mean altitude, mean annual temperature, temperature range (summer maximum minus winter minimum) and total precipitation in spring. These variables were chosen due to their known influence on butterfly distributions: altitude is a widely used surrogate for environmental gradients (*see* Guisan & Hofer 2003), and temperature and precipitation are widely recognized to affect butterfly populations (Pollard & Yates 1993). We also selected spring precipitation as a specific predictor of the distribution of *M. nausithous*, since this variable is known to affect adult flight of this species (Nowicki *et al*. 2005a). Here, using a few variables with a recognized impact on the focus species is a preferable strategy as opposed to using a high number of variables and leaving the final selection of the retained variables to automatic procedures (e.g., stepwise selection), since models built from factors already known to affect the species are likely to be more robust than models built automatically by correlating numerous variables with the available data (Austin 2007).

Altitude was obtained from a Digital Elevation Model (Clark Labs 2000), and climate variables are courtesy of the Spanish Instituto Nacional de Meteorología and the Portuguese Instituto de Meteorologia. All variables were standardized to 0 mean and 1 standard deviation to eliminate the effect of measurement-scale differences. In addition, the distribution of anthropogenic grasslands was extracted from the CORINE Land Cover 2000 GIS layer (European Environment Agency 2005) (herein CLC) at 100-m resolution. This variable was used to filter potentially suitable areas from those defined by climate requirements, accounting for the importance of habitat characteristics in delimiting *M. nausithous* distribution.

Modelling process

Given that we had reliable data only for the presence of the studied species, we used a simple modelling technique to gather data on absence before modelling their potential distributions. In creating pseudo-absence (i.e. data from sites where the species is highly likely to be absent) from outside the environmental domain of presence, restricting predictions were allowed where needed (Engler *et al*. 2004, Lobo *et al*. 2006) to avoid the use of unreliable absence information in the modelling process (Jiménez-Valverde *et al*. 2007). As proposed by Engler *et al*. (2004), we used Ecological Niche Factor Analysis (ENFA; Hirzel *et al*. 2002) to create these pseudo-absences for *M. nausithous*, *S. officinalis* and the two *Myrmica* species. ENFA develops habitat suitability maps using a specific ordination technique to explore the relationships between the presence of a species and environmental gradients, and assigns a degree of suitability to each point on a map (typically from 0 to 100) (*see* Hirzel *et al*. 2002 for a complete description of the process). We performed ENFA analysis using Biomapper 3.1 software (Hirzel *et al*. 2003), applying the geometric mean distance algorithm to compute habitat suitability maps (*see* Hirzel *et al*. 2002, Hirzel & Arlettaz 2003). All squares with habitat suitability values of zero were considered as pseudo-absences. *Maculinea nausithous* is able to disperse over distances of around 5 km (Munguira & Martín 1999, Nowicki *et al*. 2005b), so some individuals could be found in relatively unsuitable areas located near the core populations. Therefore, a buffer zone of one pixel (10 km) was established around confirmed presences, and pixels with zero habitat suitability occurring inside these buffer zones were not considered as pseudo-absences. If available, as many as ten times more pseudo-absence than presence points (n = 170 for *M. nausithous*; n = 900 for *S. officinalis*; n = 148 for *Myrmica rubra*; n = 77 for *Myrmica scabrinodis*) were randomly selected to be used in the following modelling process, since lower presence/absence ratio scores could bias the model results obtained (King & Zeng 2001).

The presence/absence of the butterfly and its host species were modelled using Generalized Additive Models (GAMs; Hastie & Tibshirani 1990), with a logit link function. GAMs are semi-parametric extensions of Generalized Linear Models, and are known to perform reasonably better than other widely-used modelling techniques (e.g. Ferrier & Watson, 1997). Given that the selection of the appropriate degrees of freedom is not a straightforward task, GAMs with penalized regression splines were used to build predictive models (Wood & Augustin 2002). As a first step, smoothed terms with 4 initial degrees of freedom were regressed independently against the response variable in order to determine significant predictors. Afterwards, significant terms were introduced in the model and selected using a manual backward procedure following the criteria proposed by Wood and Augustin (2002) (*see* this reference for a complete explanation of GAMs with penalized regression splines). Models were fitted in R (R Development Core Team 2004) using the mgcv package (Wood 2004).

Accuracy in model predictions was assessed using a jackknife procedure, a technique which yields relatively unbiased estimates of model performance (Olden *et al*. 2002). Data-demanding evaluation techniques like splitting data into independent training and evaluation sets were discarded due to low sample sizes of our biological datasets. Sample size has a great impact on model accuracy (Stockwell & Peterson 2002), so we preferred to use all the available information for training, employing an alternative measure of overfitting such as jackknife validation (Vaughan & Ormerod 2005). Here, an observation is excluded and the model is parameterized again using the remaining $n - 1$ observations to obtain a predicted probability score for the excluded observation. This procedure is repeated n times (one per observation), and the receiver operating characteristics (ROC) technique is applied to these new jackknife probabilities, using the area under the curve (AUC) as a measure of discrimination power when models are validated with data not used in the training process (Fielding & Bell 1997). AUC varies from 0.5 (discrimination ability no better than random) to 1 (perfect discriminatory ability). In addition, sensitivity and specificity (presences correctly predicted as presences and absences correctly predicted as absences, respectively) were calculated from these new jackknife probabilities; as these two

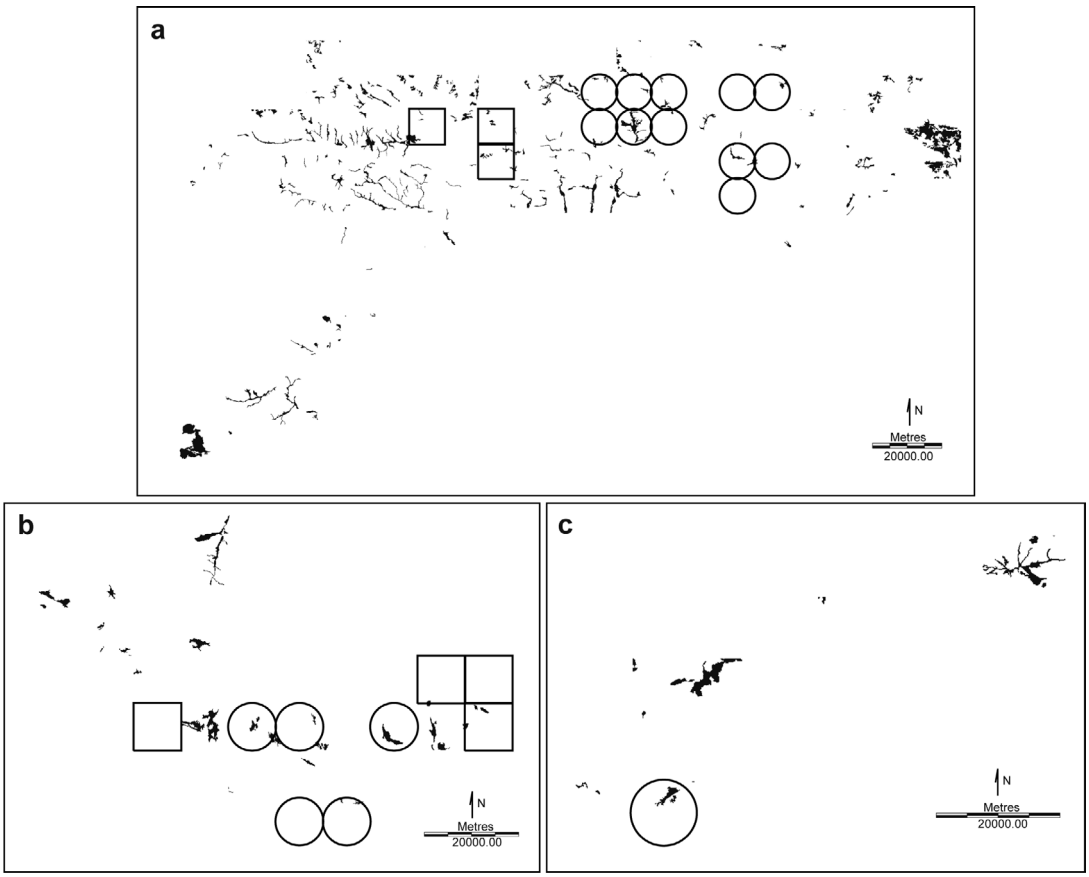

Fig. 3. Grasslands (black areas) potentially suitable for *Maculinea nausithous* in (**a**) the Cantabrian Mountain range (area 2), (**b**) the Sierra de la Demanda and Picos de Urbion (area 3) and (**c**) the Central System Mountain range (area 5) after deleting the area without anthropogenic pastures (circles = records of *M. nausithous*; squares = sampling locations).

accuracy measures depend on the value above which probabilities are considered as presences, we applied the threshold which minimizes the difference between sensitivity and specificity (Jiménez-Valverde & Lobo 2007b). All these computations were run in R (R Development Core Team 2004) using the ROCR package (Sing *et al*. 2005) and own scripts.

Field surveys

Several locations where the model predicts the potential presence of the species but where the species has not yet been recorded were visited to evaluate model reliability (Fig. 3). A pilot field study carried out by three trained surveyors in the areas where the presence of the butterfly was already known showed that the detectability of the species is high: if *M. nausithous* is present, flying individuals are seen in no more than 5 minutes once the suitable habitat (i.e., wet areas with *S. officinalis* stems in open grasslands) is located. Therefore, all grassland patches within each UTM 100 km^2 identified as potentially suitable by the model were sampled during 15 min by the three abovementioned surveyors. If the butterfly was not detected after this time, its absence from the grassland patch was assumed, and confirmed for the UTM once all suitable land patches were surveyed. Surveys were carried out on 17–25 July 2005 to coincide with the species flight peak (*see* Munguira & Martín 1993), and between 10.00–13.00 and 17.00–19.00, in order to avoid the hottest hours when flight activity is reduced. Sampling locations were selected

according to accessibility because of the extent of the study and resource limitations.

Results

The probability of presence of *M. nausithous* is highly related with the four variables considered (Table 1). Mean altitude and mean annual temperature were the two factors that accounted for the highest percentage of explained variation, and were the only variables retained in the final model; they accounted for 75.3% of the deviance. Model accuracy was also high, according to its jackknife AUC score (0.98), and to its high sensitivity and specificity scores, 0.94 and 0.95 respectively.

The extrapolation of this model to the entire Iberian territory identifies five areas that could be suitable for the butterfly (Fig. 1b): the Pyrenees Mountain range (area 1), León Mountains and Sierra de la Cabrera at the Cantabric Mountain Range (area 2), Sierra de la Demanda and Picos de Urbión in the northern part of the Iberian Mountain Range (area 3), Serranía de Cuenca and Sierra de Gúdar in the southern part of the Iberian Mountain Range (area 4), and Sierra de Guadarrama and Somosierra in the Iberian Central System Mountain Range (area 5).

The distributions of the three hosts of *M. nausithous* (either known or potential), are much wider than the area potentially suitable for the target butterfly species (Fig. 2). The model for *S. officinalis* retained only mean annual temperature (2.24 e.d.f.), which explained 97.2% of the variance (AUC = 0.99, specificity = 0.98, sensitivity = 0.98). The potential distribution of this plant (Fig. 2a) extends through most of the northern half of the Iberian Peninsula, the Southern Central Plateau, and the Baetic Mountain Range, being much wider than the potential distribution of *M. nausithous*. The model for *Myrmica rubra* retained mean annual temperature (1.832 e.d.f.) and total precipitation in spring (1.000 e.d.f.), explaining 99% of the variance (AUC = 0.99, specificity = 0.99, sensitivity = 1.00) and predicting suitable environmental conditions practically all over the Iberian Peninsula, except in the southern east and in the Guadalquivir river basin (Fig. 2b). The model for *Myrmica scabrinodis* retained only mean annual temperature (1.816 e.d.f.), which explained 99% of the variance (AUC = 0.99, specificity = 0.99, sensitivity = 1.00) and predicted suitable environmental conditions all over the Iberian territory (Fig. 2c).

After excluding the areas without anthropogenic pastures, no sites with suitable habitat for the butterfly were found in the southern part of the Iberian Mountain Range (Serranía de Cuenca and Sierra de Gúdar, area 4). The Pyrenees Mountain Range (area 1), on the contrary, harbours favourable combinations of environmental conditions and habitats, but the species has not been recorded in that area (in spite of been erroneously included in the Habitat Directive of the Catalonian regional government, C. Stefanescu pers. com.). The butterfly has been recorded in the three remaining large areas (Fig. 3). The additional sampling carried out in areas 2 and 3 (*see* locations in Fig. 3) failed to detect any previously unconfirmed *M. nausithous* population.

Discussion

Potential distribution of *Maculinea nausithous*

The large-scale geographic range of *M. nausithous* does not seem to be limited by the

Table 1. Relationship between the probability of presence of *Maculinea nausithous* and the four environmental variables used as predictors (e.d.f, estimated degrees of freedom).

Variable	Explained deviance (%)	e.d.f	χ^2	$p <$
Altitude	66.8	1.947	11.58	0.01
Precipitation at spring	22.8	1.000	24.94	0.01
Temperature range	25.2	2.274	12.70	0.01
Mean temperature	68.8	1.916	10.57	0.05

presence of its host plant or host ant species, which present a wider potential (and actual) distribution in the Iberian Peninsula. This is not surprising, as guest insects usually have narrower distribution ranges than their hosts (Gaston 2003). The model for *M. nausithous* highlights potentially suitable areas in most mountain areas of the northern half of the Iberian Peninsula, as expected by the altitudinal preferences reported in the literature for this species (Munguira & Martín 1993).

The extrapolation of the potential distribution of *M. nausithous* identifies some potentially suitable areas where the species has not been recorded (and is likely to be absent) at the Pyrenees and Serranía de Cuenca-Sierra de Gúdar (areas 1 and 4, respectively). This implies that other factors apart from environment are shaping its current distribution. Absence of *M. nausithous* from area 4 is easily explained by the absence of a suitable habitat (anthropogenic grasslands). Its absence from the Pyrenees, however, should be attributed to non-environmental factors such as extinction, biotic interactions, or simply the species has not been detected due to an insufficient sampling effort.

Although temperature constraints may be excluding *M. nausithous* from the mountain ranges of southern and western Iberia, something else could also be responsible of this distributional pattern. Most Iberian butterfly species are of Eurosiberian origin and colonized the Iberian Peninsula through the Pyrenees (Dennis *et al.* 1995a, 1995b), spreading most times throughout the interconnected Iberian mountain ranges. Due to this, the peninsular effect has been outlined as an explanation for the diversity gradient shown by Iberian butterflies (Martín & Gurrea 1990, 2003, but *see* Hortal *et al.* 2004). Licaenids are especially sensible to this effect because of their relatively low dispersal capacity when compared with the rest of butterfly families (Martín & Gurrea, 1990). Thus, *M. nausithous* could be excluded from the southern and western Iberian mountain ranges simply because its populations have not been able to reach them. The outcome of this historical process will show a spatial pattern highly correlated with some environmental gradients (especially at the west, *see* Hortal *et al.* 2004); this fact hinders the task of discriminating the relative importance of historical and environmental hypotheses with correlative techniques, such as the one used in this study. Therefore, although our results indicate that environmental limitations are the most plausible explanation for the absence of the species in Portugal in the west or the Baetic mountain range in the southeast, we cannot discard the alternative hypothesis of limited post-glacial dispersal as a cause for this pattern, especially taking into account that the species is also absent from some areas that we identify as potentially suitable (*see* above).

Areas to survey in the future

We overlaid the 50×50 km Iberian UTM squares identified as well-sampled (> 90% butterfly species recorded) by Romo and García-Barros (2005) on our model extrapolations, to identify areas environmentally suitable for *M. nausithous* but insufficiently sampled (Fig. 4). The absence of sampling effort is apparent in the western part of the Cantabrian Mountain range, Sierra de la Cabrera and Montes de León, as well as in Somosierra and Sierra del Cadí and surroundings in the Pyrenees Mountain Range. Given that additional intensive surveys could reveal many new sites for *M. nausithous* (*see* Settele 2005), we suggest that intensive survey campaigns should be carried out in these areas in the future.

Cautions when modelling specialists´ distributions

Negative results in our field surveys may be due to non-detection. However, field surveys with negative results in area 3 highlight one important source of uncertainty in predictive models: errors in environmental layers. These survey locations coincide with anthropogenic grasslands, but clearly not the kind of grasslands that *M. nausithous* requires; they were unmanaged and much drier than the grasslands where presence locations have been confirmed. Moreover, several records in areas 2 and 3 are located in UTM squares with no grasslands according to the available CORINE Land Cover data, as

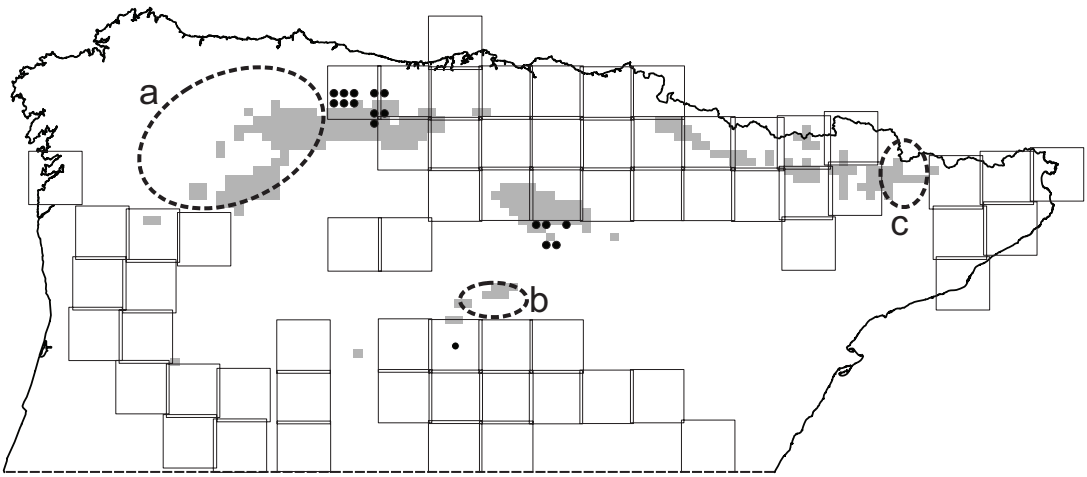

Fig. 4. Potential distribution of *Maculinea nausithous* and 50 × 50 km UTM squares identified as well-sampled by Romo and García-Barros (2005). Black dots are the 17 known records for the butterfly. Circles indicate the areas suggested for future surveys (a = western part of the Cantabrian Mountain range, Sierra de la Cabrera and Montes de León; b = Somosierra; c = Sierra del Cadí and surroundings).

shown in Fig. 3. CLC data are known to suffer from classification errors (European Environment Agency 2002, Erhard *et al*. 2004), i.e, the allocation of an incorrect category to a pixel, which may negatively affect a fine-resolution work. Apart from errors in the classification of satellite data, there are problems associated with the difficulty of separating overlapping classes of the CLC legend (Felicísimo & Gago 2002). Thus, although CLC data are readily available and their use for *Maculinea* habitat studies has previously been suggested (Munguira & Martín 1999), they should be used with extreme caution when taking fine-scale management decisions. Their inclusion is, nevertheless, necessary when modelling habitat-specific species at a coarse resolution, given that ignoring them would considerably overestimate potential distributions.

Grasslands present in sampling sites from area 2 were, a priori, appropriate for *Maculinea*, based on our field experience. Suitable grasslands, usually with high densities of the host plant, were found but no butterfly populations were detected. Although the absence of records is not the equivalent of the true absence of the butterfly (given that the species could be present but not yet detected by surveyors), the conspicuousness of this species makes these field absences quite reliable. Thus, absences assumed to be reliable could be attributed to (i) extinction of current populations, (ii) areas never reached by the butterfly, (iii) metapopulation dynamics (i.e., sink habitats), and (iv) a missing fine-scale key factor, such as absence or local spatial structure and size of *Myrmica* populations (Elmes *et al*. 1998, Anton *et al*. 2005), grass height, wind protection, incident solar radiation, cover of shrubs, etc., which make the grasslands unsuitable for the butterfly.

Concluding remarks

Predictive models of species distribution based on climate and topographic variables can be used to delimit the potential distribution range on broad scales, and highlight areas that can potentially sustain undiscovered populations of rare species. The methodology followed in this study can be applied to other rare species from different taxonomic groups. The climate that describes the environmental requirements of a species is first constrained by historical and unique events that cause the absence of the species from environmentally suitable places (Soberón 2007). Then, the resultant potential distribution is shaped a posteriori by historical and dispersal limitations, biotic interactions and habitat selection. Thus, although broad-scale distribution models are a necessary first step in any large-scale study of the

geographic range of an endangered species, their output should never be used as a direct estimate of its actual distribution. Rather, their most direct use could be the detection of suitable locations, which might become the focus of additional survey effort to find new populations of the species and update, in this way, its distribution map (*see* also Raxworthy *et al*. 2003).

Nevertheless, this study has pointed out several drawbacks that must be considered when working on a finer scale with a species presenting highly specific habitat and biological requirements. In the case of *M. nausithous*, an additional problem arises from the absence of information on a key factor such as the fine-grain data on the distribution of its hosts (*S. officinalis* and the two *Myrmica* species), which can only be overcome with additional field work designed to construct highly-detailed maps (*see*, for example, Holloway *et al*. 2003). Indeed, land-use information obtained from widely available sources, such as the CLC project, must be used with caution. If fine-resolution results are needed, other sources of information should be employed, such as aerial photographs or detailed field work. Errors or lack of fine-resolution information such as the here highlighted can be serious over large areas, aggregating along the territory and, so, directly affecting broad-scale potential distribution models.

This work illustrates a real example of field application of a distribution predictive model, something that is rarely done. The main problem is found when the model is applied to the area of interest but the focus species is not found. These may be the cause of historical factors that deserve further attention and, then, the hypotheses generated by the models are interesting tools for exploring biogeography questions (*see* Peterson 2006). However, many times the absence of the species is caused by the lack of a key required factor. Distinguishing between the two origins of overestimation is a challenging task not always easily approachable. This lack of knowledge about habitat requirements may be the rule rather than the exception for most of the species, especially for hyperdiverse groups such as invertebrates, the most ignored taxa but contributing most to biodiversity. Overestimation in conservation is dangerous as it might imply the location of conservation efforts in places which do not harbour the focus species. As a final conclusion we stress that predictive models must be used with caution in conservation, and they are a complement of rather than a replacement for expert knowledge.

Acknowledgements

Juan Ignacio de Arce Crespo and Silvia Jiménez Mendoza provided helpful field information. The Junta de Castilla y León and the Diputación General de Aragón provided the licences to survey *M. nausithous*. This paper was supported by a Fundación BBVA project (Diseño de una red de reservas para la protección de la Biodiversidad en América del sur austral utilizando modelos predictivos de distribución con taxones hiperdiversos) and a MEC Project (CGL2004-04309). AJ-V was supported by a Museo Nacional de Ciencias Naturales/C.S.I.C./Comunidad de Madrid PhD grant. AB position at MNCN was supported by MEC (Ministerio de Educación y Ciencia) Juan de la Cierva Program. JH was supported by a Portuguese FCT (Fundação para a Ciência e Tecnologia) grant (BPD/20809/2004) and the U.K. Natural Environment Research Council.

References

Anton, C., Musche, M., Hula, V. & Settele, J. 2005: Which factors determine the population density of the predatory butterfly *Maculinea nausithous*? — In: Settele, J., Kühn, E. & Thomas, J. A. (eds.), *Studies on the ecology and conservation of butterflies in Europe, vol. 2. Species ecology along a European gradient: Maculinea butterflies as a model*: 57–59. Pensoft, Sofia/Moscow.

Austin, M. 2007: Species distribution models and ecological theory: a critical assessment and some possible new approaches. — *Ecological Modelling* 200: 1–19.

Bourg, N. A., McShea, W. J. & Gill, D. E. 2005: Putting a CART before the search: successful habitat prediction for a rare forest herb. — *Ecology* 86: 2793–2804.

Chefaoui, R. M., Hortal, J. & Lobo, J. M. 2005: Potential distribution modelling, niche characterization and conservation status assessment using GIS tools: a case study of Iberian Copris species. — *Biological Conservation* 122: 327–338.

Clark Labs 2000: *Gobal change data archive, vol. 3. 1 km global elevation model*. — Clark Labs, Worcester, MA.

Dennis, R. L. H., Shreeve, T. G. & Williams, W. R. 1995a: Affinity gradients among European butterflies: evidence and historical component to species distributions. — *Entomologis's Gazette* 46: 141–153.

Dennis, R. L. H., Shreeve, T. G. & Williams, W. R. 1995b: Taxonomic differentiation in species richness gradients among European butterflies (Papilionoidea, Hesperioidea): contribution of macroevolutionary dynamics. —

Ecography 18: 27–40.

Elmes, G. W., Thomas, J. A., Wardlaw, J. C., Hochberg, M. E., Clarke, R. T. & Simcox, D. J. 1998: The ecology of *Myrmica* ants in relation to the conservation of *Maculinea* butterflies. — *Journal of Insect Conservation* 2: 67–78.

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.

Erhard, M., Hennings, V. & Werner, B. 2004: Using CORINE land-cover and statistical data for the assessment of soil erosion risks in Germany. — In: Mohaupt-Jahr, B., Keil, M. & Kiefl, R. (eds.), *CORINE land cover 2000 in Germany and Europe and its use for environmental applications*: 124–130. Federal Environmental Agency, Berlin.

European Environment Agency 2002: *I&CLC2000 Technical Reference Document*. — Joint research Centre, European Commission.

European Environment Agency 2005: *CORINE land cover 2000 vector data for Spain*.

Felicísimo, A. M. & Gago, L. M. S. 2002: Thematic and spatial accuracy: a comparison of the CORINE land cover with the forestry map of Spain. — In: Ruiz, M., Gould, M. & Ramon, J. (eds.), *5th AGILE Conference on Geographic Information Science*: 109–118. Palma, Balearic Islands, Spain.

Ferrier, S. & Watson, G. 1997: *An evaluation of the effectiveness of environmental surrogates and modelling techniques in predicting the distribution of biological diversity*. — Environment Australia, Canberra, available on the web at http://www.deh.gov.au/biodiversity/publications/technical/surrogates.

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.

García-Barros, E., Munguira, M. L., Martín, J. & Viejo, J. L. 1993: *Maculinea nausithous* (Bergsträsser, 1779) en Madrid. — *SHILAP Revista de lepidopterología* 21: 255–257.

García-Barros, E., Munguira, M. L., Martín Cano, J., Romo Benito, H., García-Pereira, P. & Maravalhas, E. S. 2004: *Atlas de las mariposas diurnas de la Península Ibérica e islas Baleares (Lepidoptera: Papilionoidea & Hesperioidea)*. — Monografías SEA vol. 11, Zaragoza.

Gaston, K. J. 2003: *The structure and dynamics of geographic ranges*. — Oxford University Press, Oxford.

Guisan, A. & Zimmermann, N. E. 2000: Predictive habitat distribution models in ecology. — *Ecological Modelling* 135: 147–186.

Guisan, A. & Hofer, U. 2003: Predicting reptile distributions at the mesoscale: relation to climate and topography. — *Journal of Biogeography* 30: 1233–1243.

Hastie, T. J. & Tibshirani, R. J. 1990: *Generalized additive models*. — Chapman & Hall, London.

Hirzel, A. H., Hausser, J., Chessel, D. & Perrin, N. 2002: Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? — *Ecology* 83: 2027–2036.

Hirzel, A. H. & Arlettaz, R. 2003: Modeling habitat suitability for complex species distributions by environmental-distance geometric mean. — *Environmental Management* 32: 614–623.

Hirzel, A. H., Hausser, J. & Perrin, N. 2003: *Biomapper 3.1*. — Laboratory for Conservation Biology, University of Lausanne, Lausanne, Switzerland, available on the web at http://www.unil.ch/biomapper/.

Holloway, G. J., Griffiths, G. H. & Richardson, P. 2003: Conservation strategy maps: a tool to facilitate biodiversity action planning illustrated using the heath fritillary butterfly. — *Journal of Applied Ecology* 40: 413–421.

Hortal, J., Garcia-Pereira, P. & García-Barros, E. 2004: Butterfly species richness in mainland Portugal: Predictive models of geographic distribution patterns. — *Ecography* 27: 68–82.

Jiménez-Valverde, A. & Lobo, J. M. 2007a: Potential distribution of the endangered spider *Macrothele calpeiana* (Walckenaer, 1805) (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., Ortuño, V. M. & Lobo, J. M. 2007: Exploring the distribution of *Sterocorax* Ortuño, 1990 (Coleoptera, Carabidae) species in the Iberian Peninsula. — *Journal of Biogeography* 34: 1426–1438.

King, G. & Zeng, L. 2001: Logistic regression in rare events data. — *Political Analysis* 9: 137–163

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.

Martín, J. & Gurrea, P. 1990: The peninsular effect in Iberian butterflies (Lepidoptera: Papilionoidea and Hesperioidea). — *Journal of Biogeography* 17: 85–96.

Martín, J. & Gurrea, P. 2003: La distribución de las zigenas (Lepidoptera, Zygaenidae) ibéricas: una consecuencia del efecto península. — *Graellsia* 59: 273–285.

Munguira, M. L. & Martín, J. 1993: The conservation of endangered lycaenid butterflies in Spain. — *Biological Conservation* 66: 17–22.

Munguira, M. L. & Martín, J. 1994: La conservación de las *Maculinea* españolas. — *Butlletí Societat Catalana Lepidopterología* 73: 20–28.

Munguira, M. L. & Martín, J. 1999: *Action plan for Maculinea butterflies in Europe*. — Nature and environment, No. 97, Council of Europe Publishing.

Munguira, M. L., García-Barros, E. & Martín, J. 1997: Plantas nutricias de los licénidos y satirinos españoles (Lepidoptera: Lycaenidae y Nymphalidae).— *Boletín de la Asociación española de Entomología* 21: 29–53.

Munguira, M. L., Martín, J. & García-Barros, E. 2006: *Maculinea nausithous* (Bergsträsser, 1779). — In: Galante, E. & Verdú, J. R. (eds.), *Libro Rojo de los Invertebrados de España*: 227–228. Dirección General para la Biodiversidad, Ministerio de Medio Ambiente, Madrid.

Nowicki, P., Witek, M., Skórka, P., Settele, J. & Woyciechowski, M. 2005a: Population ecology of the endan-

gered butterflies *Maculinea teleius* and *M. nausithous* and its implications for conservation. — *Population Ecology* 47: 193–202.

Nowicki, P., Settele, J., Thomas, J. A. & Woyciechowski, M. 2005b: A review of population structure of *Maculinea* butterflies. — In: Settele, J., Kühn, E. & Thomas, J. A. (eds.), *Studies on the ecology and conservation of butterflies in Europe, vol. 2. Species ecology along a European gradient: Maculinea butterflies as a model*: 144–149. Pensoft, Sofia/Moscow.

Olden, J. D., Jackson, D. A. & Peres-Neto, P. 2002: Predictive models of fish species distributions: a note on proper validation and chance predictions. — *Transactions of the American Fisheries Society* 131: 329–336.

Peterson, A. T. 2006: Uses and requirements of ecological niche models and related distributional models. — *Biodiversity Informatics* 3: 59–72.

Peterson, A. T. & Papeş, M. 2006: Potential geographic distribution of the Bugun liocichla *Liocichla bugunorum*, a poorly-known species from north-eastern India. — *Indian Birds* 2: 146–149.

Pollard, E. & Yates, T. J. 1993: *Monitoring butterflies for ecology and conservation*. — Chapman & Hall, London.

R Development Core Team 2004: *R: A language and environment for statistical computing*. — R Foundation for Statistical Computing, Vienna, Austria, available on the web at http://www.R-project.org.

Raxworthy, C. J., Martínez-Meyer, E., Horning, N., Nussbaum, R. A., Schneider, G. E., Ortega-Huerta, M. A. & Peterson, A. T. 2003: Predicting distributions of known and unknown reptile species in Madagascar. — *Nature* 426: 837–841.

Romo, H. & García-Barros, E. 2005: Distribución e intensidad de los estudios faunísticos sobre mariposas diurnas en la Península Ibérica e Islas Baleares (Lepidoptera, Papilionoidea y Hesperioidea). — *Graellsia* 81: 37–50.

Settele, J. 2005: How endangered is *Maculinea nausithous*? — In: Settele, J., Kühn, E. & Thomas, J. A. (eds.), *Studies on the ecology and conservation of butterflies in Europe, vol. 2. Species ecology along a European gradient: Maculinea butterflies as a model*: 230. Pensoft, Sofia/Moscow.

Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. 2005: *ROCR: Visualizing the performance of scoring classifiers. R package version 1.0-1*. — R Foundation for statistical computing, Vienna, Austria, available on the web at http://rocr.bioinf.mpi-sb.mpg.de.

Soberón, J. 2007: Grinnellian and Eltonian niches and geographic distribution of species. — *Ecology Letters* 10: 1115–1123.

Stockwell, D. R. B. & Peterson, A. T. 2002: Effects of sample size on accuracy of species distribution models. — *Ecological Modelling* 148: 1–13.

Thomas, J. A. 1980: Why did the large blue become extinct in Britain? — *Oryx* 15: 243–247.

Thomas, J. A., Elmes, G. W., Wardlaw, J. C. & Wovejechowski, M. 1989: Host specificity among *Maculinea* butterflies in *Myrmica* ant nests. — *Oecologia* 79: 452–457.

Vaughan, I. P. & Ormerod, J. 2005: The continuing challenges of testing species distribution models. — *Journal of Applied Ecology* 42: 720–730.

Wood, S. N. 2004: *mgcv: GAMs with GCV smoothness estimation and GAMMs by REML/PQL. R package version 1.1-8*. — R Foundation for Statistical Computing, Vienna, Austria, available on the web at http://www.R-project.org.

Wood, S. N. & Augustin, N. H. 2002: GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. — *Ecological Modelling* 157: 157–177.

Wynhoff, I. 1998: The recent distribution of the European *Maculinea* species. — *Journal of Insect Conservation* 2: 15–27.