*Ann. Zool. Fennici* 48: 147–160 ISSN 0003-455X (print), ISSN 1797-2450 (online)
Helsinki 30 June 2011

# Modelling species' climatic distributions under habitat constraints: a case study with *Coturnix coturnix*

Francesc Sardà-Palomera[1,2] & David R. Vieites[3]

[1)] *Landscape Ecology group, Forest Sciences Center of Catalonia (CTFC), Crtra. vella de Sant Llorenç de Morunys km 2, ES-25280 Solsona, Spain (corresponding author's e-mail: francesc.sarda@ctfc.cat)*
[2)] *Department of Animal Biology, University of Barcelona, Avda. Diagonal 645, ES-08029 Barcelona, Spain*
[3)] *Department of Biodiversity and Evolutionary Biology, Museo Nacional de Ciencias Naturales (MNCN-CSIC), C/José Gutierrez Abascal n°2, ES-28006 Madrid, Spain*

*Received 22 July 2010, revised version received 14 Jan. 2011, accepted 6 Jan. 2011*

Sardà-Palomera, F. & Vieites, D. R. 2011: Modelling species' climatic distributions under habitat constraints: a case study with *Coturnix coturnix*. — *Ann. Zool. Fennici* 48: 147–160.

*Coturnix coturnix* is a climatically widespread farmland specialist bird with a well-known distribution in the Iberian Peninsula. Therefore, its current real distribution can be finely mapped. We used recent atlas data and high-resolution climate layers to model the climatic distribution of this species. Next, we compared the obtained predictive maps with the real distribution of the species. Precise information about underestimation and overestimation of model predictions at different sample sizes was obtained. An increase in sample size allowed the predicted distribution to approach the current real distribution, but unbalanced the climatic prediction. This is because the distribution of occurrences is mostly determined by habitat, which depends more on agricultural factors and less on climate. Modelling the climatic distribution of widespread species while ignoring habitat constraints could lead to a non-equilibrium scenario where obtained predictions may not be reliable.

## Introduction

Predicting actual and potential species' distributions has become a popular subject. A plethora of novel approaches and modelling techniques have recently been developed for this purpose, which are commonly known as species' distribution or bioclimatic envelope modelling (SDM) (Busby 1991, Carpenter *et al*. 1993, Phillips *et al*. 2004, Araújo & New 2007). The main assumption of such approaches is that the environmental conditions (e.g., climate) from the localities where species occur represent their ecological niches (Guisan & Zimmermann 2000) and can be used to develop predictive models. However, modelling the distribution of a species is not strictly equivalent to modelling its niche (Soberón & Nakamura 2009), and the selection of appropriate models and variables depends on which question needs to be answered.

Although some methods have been suggested as being "better" in terms of performance than others (Elith *et al*. 2006, Hernández *et al*. 2006), such comparisons between different methodologies are sometimes not appropriate. Different methods have different assumptions and the calculation of different objects may rely on different kinds of data (Segurado & Araújo 2004, Soberón & Nakamura 2009). Logically, the performance of these models can also vary depending on the quality and amount of data, being sensitive to the number of presence records (Wisz *et al.* 2008), the different types of absences (Brotons *et al.* 2004), the ratio of absences/presences (Jiménez-Valverde & Lobo 2006), the extent of the study area (VanDerWal *et al.* 2009) and the resolution of environmental layers (Araújo *et al.* 2005, Guisan *et al.* 2007).

Sample size is an important factor in the development of accurate species' distribution models. Most of the studies dealing with this issue have focused on how to minimize locality data points to properly predict the distribution area of a species (e.g., Hernandez *et al.* 2006, Pearson *et al.* 2007). This is an especially important issue in areas like the tropics, where species' distributions are poorly known, and the number of records per species is usually very low (Pearson *et al.* 2007, Vieites *et al.* 2008). Increasing the number of localities per species may improve the accuracy of the models (Wisz *et al.* 2008), although it also depends on the degree of environmental heterogeneity where species occur and the extent of their distribution ranges (Stockwell & Peterson 2002, Hernández *et al.* 2006). For example, species that occupy a homogeneous climatic space can be predicted with high accuracy by available modelling methods, even with a low number of presence records. In contrast, generalist species that occur across different environments are hard to model because of this heterogeneity, especially when available locality data are scarce (Stockwell & Peterson 2002, Segurado & Araújo 2004). Theoretically, the optimum number of localities to model species' distributions should be achieved when there are enough to cover the whole species distribution range. The existence of potential equilibrium in the accuracy and predictive power of models in parallel to the increase of the number of samples has not yet been fully explored.

Approaches that include randomly generated pseudo-absences are becoming very popular (Ferrier *et al.* 2002, Graham *et al.* 2004, Pearce & Boyce 2006). Presence-only data models are less sensitive than models based on presence-absence data (Brotons *et al.* 2004); however, predictions of those models are different when real absences are missing (*see* Soberón & Nakamura 2009). In parallel, the spatial resolution and scale of the models are relevant depending on which questions need to be addressed. For example, low resolution grid cells will be adequate to model species' distributions at a global scale, but at smaller scales, high-resolution environmental data are needed. Those models could also be affected by other independent parameters, such as species interactions, biogeographic history, evolutionary history and current habitat availability, which contribute to the actual species' distributions (Vieites *et al.* 2009).

Species' distribution modelling methods must be evaluated to asses their predictive errors. This is achieved by using measures of performance such as the area under the receiver operating characteristic curve (AUC). AUC measures the degree to which a species is restricted to a range of variation of the predictive variables. Hence, it gives information about the suitability of the presence of a species along the range of predictor conditions in the study area, but it does not allow comparisons between species (Lobo *et al.* 2008). A lower AUC is expected for a species with broad ecological and environmental tolerance, although it is not actually an indication of lower model performance. This method is threshold independent and is usually calculated using a percentage of the data used for modelling. In addition, this method is better when using independent data, although the assumption of data independency is often violated.

The optimal situation to test the performance of predictive models is to compare predicted distributions generated by the model with the real distribution of the species, although for most species, the real distribution is unknown. Here, we used high-resolution environmental data to develop a set of predictive models to test the impact of increasing sample sizes. We examined how sample size affected the performance and accuracy of models and their ability to predict the distribution of a widespread farmland bird species as well as the impact on the evaluation metrics. The species we selected is the common quail (*Coturnix coturnix)*, a climatically widespread farmland specialist with a very well known distribution range in the Iberian Peninsula.

Many farmland bird populations have strongly decreased in the last 50 years due to agricultural intensification and habitat loss (Donald *et al*. 2001, Gregory *et al*. 2005). Predicting the distribution of these species and understanding the relation between climate and farmland habitat is essential for the effective management and conservation of these populations. However, in farmland landscapes, species presences can be conditioned not only by climate but also by landscape management, agricultural policies and cultural and economic aspects.

The common quail is a migratory bird that winters in the Sahel and breeds across Eurasia. Its distribution during the breeding season spans through different climates, from semi-arid in northern Africa and western Asia, to humid oceanic and continental in Europe and Russia (Guyomarc'h *et al*. 1998). Despite being a widespread species, the quail is a ground bird specialist of open habitats covered by seasonally dense ground cover, meadows and natural grasslands. However, in most of its breeding area, its natural habitat has been replaced by crops including winter cereal fields, which cover huge extensions in Eurasia, and other herbaceous crops (Guyomarc'h *et al*. 1998). Its current distribution in the Iberian Peninsula is very well known (Rodríguez-Teijeiro *et al*. 2003, Gonçalves 2008), and because of its habitat specialization, an accurate breeding distribution can be mapped. This peculiarity is not commonly found in many climatically generalist species and therefore makes the common quail a suitable candidate for exploring the accuracy of climate-based distribution models.

Currently, SDMs are the best tool to increase knowledge on the future and past distributions and the effects of climate change relative to other factors. To date, research on distribution modelling has mainly focused on model performance at relatively small sample sizes and the question of "how much data is enough" has not been investigated thoroughly. Moreover, insufficient attention has been paid to the challenges of modelling widespread generalist species. Here, we explore the ability of models to predict, from a climatic point of view, a precise distribution of a widespread farmland bird species and which variables contribute most to the different models based on different sample sizes.

## Material and methods

### Current quail distribution and locality data

Two recent fine-scale breeding atlases provided the distribution in 10 × 10 km UTM square grids for Spain (Rodríguez-Teijeiro *et al*. 2003) and Portugal (Gonçalves 2008). Those data were the basis for our models. Only squares where the species' reproduction was confirmed or probable were used, whereas squares were the species was not present, or its presence was possible but not confirmed, were not included. Suitable habitats for the species (Guyomarc'h *et al*. 1998, Guyomarc'h 2003) in the Iberian Peninsula were selected from the CORINE Land Cover 2000 dataset. The habitats selected were non-irrigated arable land, permanently irrigated arable land, pastures, annual crops associated with permanent crops, complex cultivation patterns, land principally occupied by agriculture with significant areas of natural vegetation, natural grassland, moors and heathland. We used this selection to create a habitat map for the presence of suitable breeding habitats for the species. We then intersected the known breeding distribution in 10 × 10 km UTM squares with the habitat map to obtain a constrained breeding distribution map for the species. This map is likely to represent the real distribution of the species in Iberia, hereafter called the "real distribution" at 200 m resolution (*see* Appendix 1).

### Modelling methods

High-resolution (200 m grid-cell size) climate digital layers for the Iberian Peninsula were obtained from the "Atlas Climático Digital de la Península Ibérica" (Ninyerola *et al*. 2005). This dataset consists of monthly and annual means of temperature and rainfall from 40 years for the Iberian Peninsula, based on the complete set of meteorological stations for Spain and a high number from Portugal. It improves other available datasets based on fewer stations for the area and coarser interpolations (e.g., Worldclim, Hijmans *et al*. 2005). This resolution is suitable for our purposes as it parallels the quail distribution and habitat datasets.

The common quail depends on dense herbaceous cover and cereal crops. The development of these crops is related to seasonal climate conditions (Austin *et al*. 1998) and shows some phenological variation from southern to northern Iberia (Rodríguez-Teijeiro *et al*. 2009). Thus, to model this species' distribution and identify the effect of climate on its habitat, seasonal climatic data are needed. We used the means of two consecutive months for temperature and precipitation. Annual means of temperature and rainfall were also included as predictive variables, resulting in a total of 14 variables: January–February, March–April, May–June, July–August, September–October and November–December means for temperature and rainfall, and annual means for temperature and precipitation.

Occurrence data points were randomly sampled from within the real distribution, but limited to only one point for each 10 × 10 km breeding UTM square to avoid duplications ($n$ = 3630, Appendix 1). We generated datasets with sample sizes of 5, 10, 25, 50, 75, 100, 250, 500, 1000, 1500, 2000, 2500 and 3000 randomly-selected points within the constrained breeding distribution. Those values range from a minimum of five localities to values close to the potential maximum coverage (3000, without duplicates) of the real species' distribution range in Iberia. This allowed us to test the influence of a whole range of sample sizes and investigate the potential stability of results towards a potential optimum. For each sample size, we randomized the modelling process by generating 25 different replicas for each sample size. Background samples (Ferrier *et al*. 2002) were generated by drawing a random sample of 10 000 sites within the Iberian Peninsula without any specification. Those were intended as a sample of the climatic space for the whole region, with the possibility that a background sample matched a presence record.

Models were developed using MaxEnt ver. 3.3, a maximum entropy method for predicting species' distributions (Phillips *et al*. 2006). This software has become very popular, mainly because it performs well with small sample sizes of presence-only data (Elith *et al*. 2006, Hernandez *et al*. 2006). We ran 325 SDMs using the 'logistic output format' that gives an estimate of the probability of presence (Phillips & Dudik 2008), and resulting models were projected to obtain presence probability maps.

## Evaluation of model performance

Four different methods were used to evaluate models. We used three different versions of the AUC index (Fielding and Bell 1997). When used to test the accuracy of distribution models with presence/absence data, the AUC can vary from 0.5 (for chance performance) to 1 (perfect fit). However, when the distribution of a certain species covers a fraction $a$ of the studied area, the maximum achievable AUC is less than 1 at exactly $1 - a/2$ (Phillips *et al*. 2006). In this study, we were able to use the real distribution presence/absence data to obtain the AUC (hereafter 'real AUC'), which was calculated using the R package 'presence–absence' version 1.1.3 in R ver. 2.8.1 (R Development Core Team 2005). Normally, when the real distribution of the species is not available, the most common method to calculate AUC is by randomly setting aside 30% of the occurrences as test points (Phillips *et al*. 2006) to obtain a test AUC (hereafter 'test AUC'). We also used the training AUC, calculated directly from the training data (hereafter 'training AUC'). Training and test AUC are given by MaxEnt.

Finally, models were evaluated by comparing them to the quail's real distribution. To quantify underestimation (false negative) and overestimation (false positive) of model predictions, resultant predictive models were reclassified from continuous probabilities to raster binary maps of presence (reclassified as 1) and absence (reclassified as 0) using sensitivity-specificity sum maximization (Liu *et al*. 2005). Each binary map was added to the real distribution map, which was previously classified with value 2 for presence and value 0 for absence. Thus, we obtained one map for each model, with a range of pixels from 0 to 3. This resulting map was added to the map

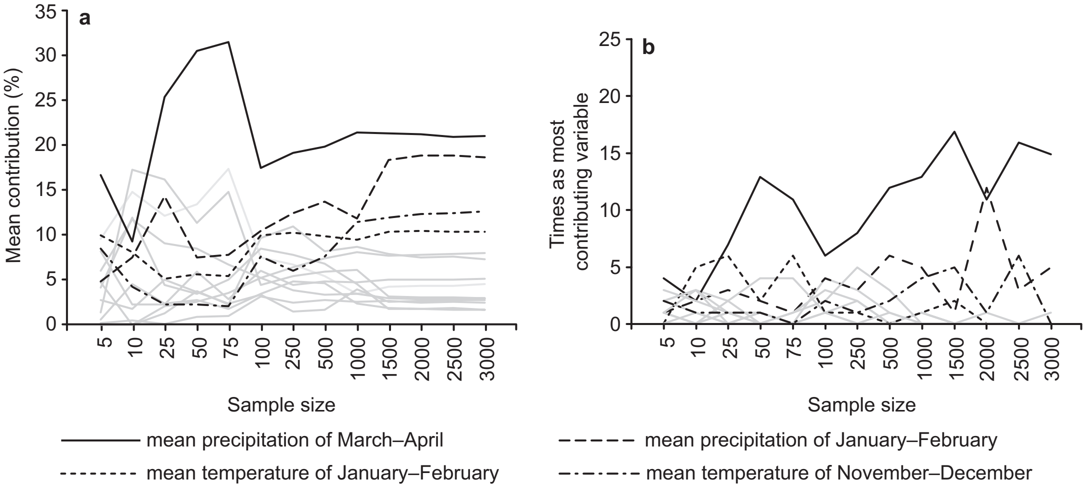


**Fig. 1.** Contribution of the four most important variables to the model with regard to sample size (black) and the rest of variables (grey): (**a**) mean conribution, and (**b**) times the variables contributed most to the models.

of the suitable habitat for the species which was previously reclassified, with value 3 for suitable habitat and value 0 for non suitable habitat. This last step allowed us to distinguish total overestimation from overestimation that occurred only within the areas of suitable habitat. From this analysis, we aim to quantify the effect of habitat constraints over the potential climatic distribution. Thus, for each model, we obtained a subsequent map which had pixel values from 0 to 6 (0 = nothing, 1 = total overestimation, 2 = 'value not possible', 3 = only suitable habitat, 4 = overestimation within suitable habitat, 5 = underestimation, 6 = overlap with the real distribution). The binary presence/absence maps were also added by sample size groups, and we obtained a final map for each sample size that ranged from 0 (any model predicted the species in a certain cell) to 25 (all models predicted the species in a certain cell).

## Results

The real distribution of the common quail in the Iberian Peninsula created from the intersection of atlas data and habitat suitability covers 30.54% of the territory, whereas the potential suitable habitat selected from CORINE land cover represents 42.49%. A total of 325 models were successfully run. Six models with five presence points and four models with ten presence points did not discriminate properly and were equivalent to random predictions. We obtained the contribution of each variable to the models (Fig. 1). The variable that contributed most was the accumulated precipitation in March–April. It was followed by the mean accumulated precipitation in January–February and the mean temperature in January-February. The precipitation of March–April was not the most important variable in models based on ten or less localities. This variable showed a maximum contribution in models based on 75 localities, a decrease towards models based on 100 localities, and then a relative stability in models based on higher sample sizes. From models based on 50 localities, the effect of this variable to the logistic prediction was negative in all the models.

A different contribution pattern with regard to sample size can be observed in other variables (Fig. 1a). For example, the contribution of precipitation of January–February increased with sample size, but was always below that of precipitation of March–April. The contributions of the variables 'mean temperature of January–February' and 'mean temperature of November–December' showed a similar tendency to increase, but always stayed below the contributions of the other two variables. When the replicates for each sample size are considered, the

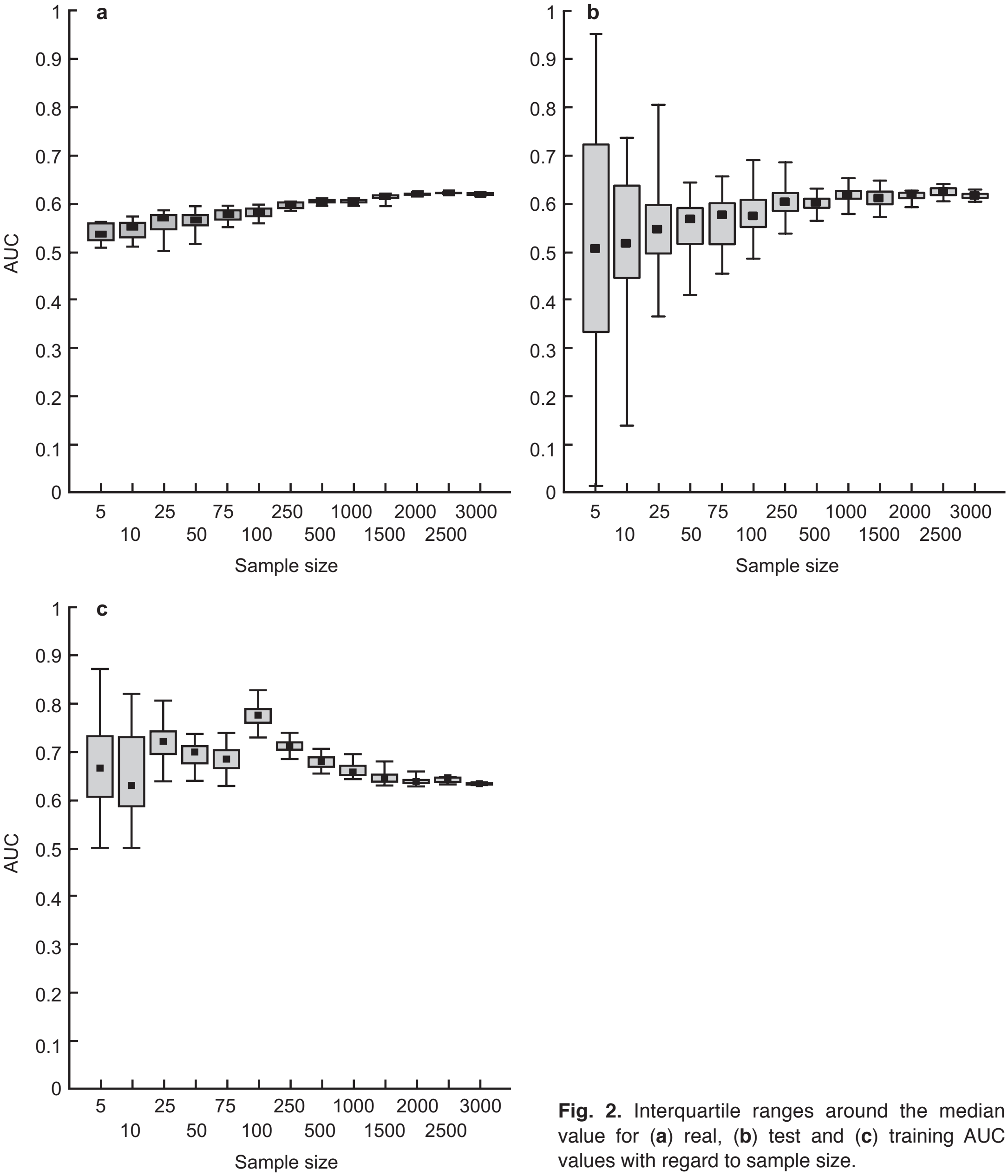


**Fig. 2.** Interquartile ranges around the median value for (**a**) real, (**b**) test and (**c**) training AUC values with regard to sample size.

contribution of the precipitation of March–April is the most important variable when sample size is increased (Fig. 1b)

Median real AUC values ranged from 0.55 to 0.65 (Fig. 2a). An increase in the median real AUC is observed in parallel to the increase in sample size, reaching a maximum when the sample size was equal to 3000 points. Median test AUC values showed a very similar pattern to the real AUC (Fig. 2b); however, test AUC showed much higher variation in the interquartile range, maximum and minimum values. Median training AUC values where higher than real and test AUCs, ranging between 0.6 and 0.8 (Fig. 2c). Like the test AUC, when sample sizes are low, the variation in training AUCs is high between models. However, the training AUC reached its maximum when the sample size was

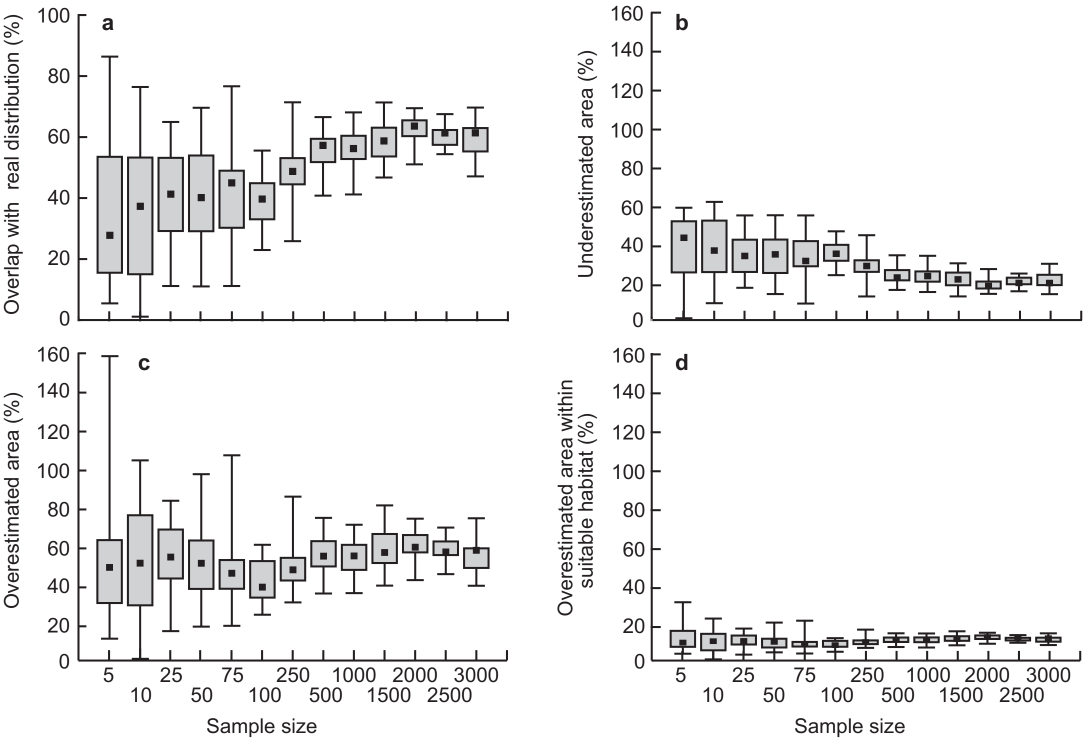


**Fig. 3.** Interquartile ranges around the median value of errors and overlap area when comparing model predictions with the real distribution of the species.

100 points and decreased towards higher values. In the three cases with sample sizes higher than 250, the variation between models was significantly reduced and the median training AUC stabilized around 0.65. The interquartile range of the real AUC was reduced when compared with those of the test and training AUCs. The real and test AUC never reached the maximum value obtained when using training data with 100 presences.

We obtained detailed information about model underestimation and overestimation in relation to sample size (Fig. 3). In general, low sample-size models show an important variability in predicting the real distribution. This variability reduces towards higher sample sizes and stabilizes around 60% of the predicted overlap with the real distribution (Fig. 3a), reaching a maximum when the sample size was around 1000 presence points. Underestimation has an inverse pattern (Fig. 3b), decreasing inversely with sample size and stabilizing in models based on 1000 presence points. Overestimation shows huge variation when sample sizes are low (Fig. 3c), variation that is reduced significantly when sample sizes reach 1000 points or more. When considering overestimation only within the area with suitable habitat (Fig. 3d), the same pattern of overestimation is observed, but the error area decreases considerably.

We represented the sum of the replicates of the models for each sample size (Fig. 4), in which each grid cell can range from 0 (the species was not predicted by any model) to 25 (the species was predicted in 25 out of 25 models). There is a clear trend of spatial agreement between predictive models when sample size increases. Models with low sample sizes usually predict the species in different areas, and few or no pixels are predicted as presences in all replicates. With more than 250 records there is a significant reduction in the spatial variation of the presence of the species, while the differences between stacks of models with more than 1500 points are minimum. However, independently of sample size, an important underestimation in

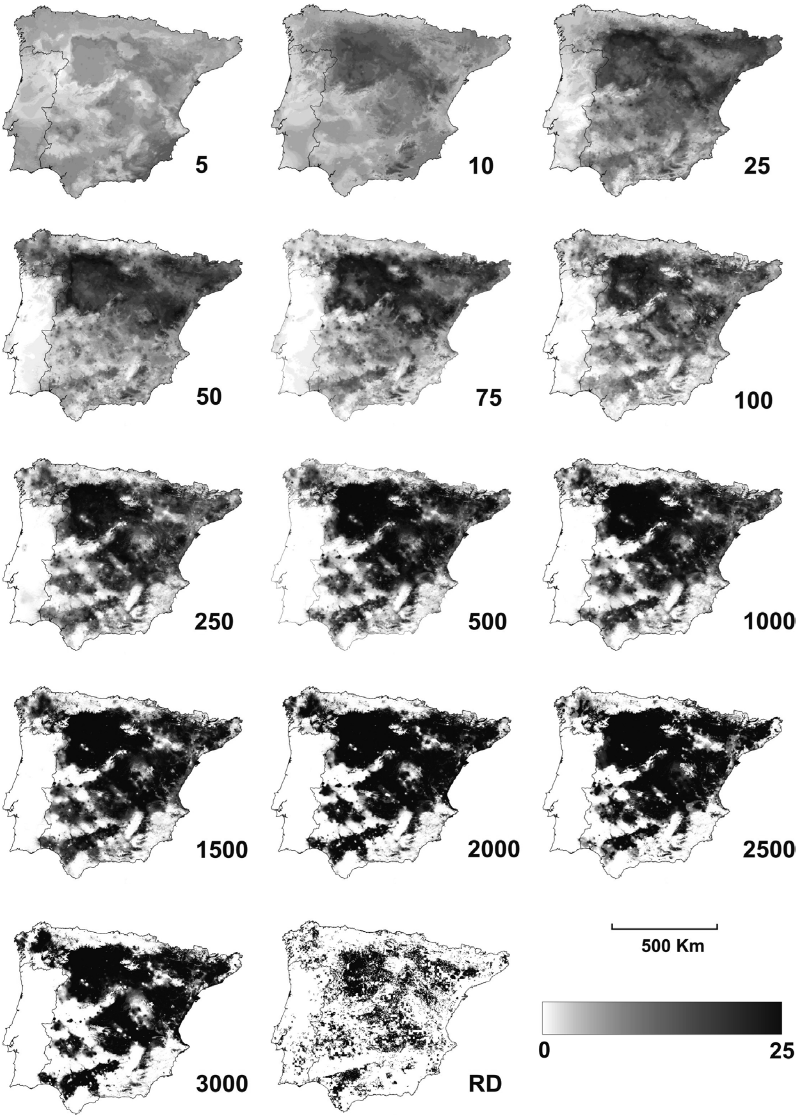


**Fig. 4.** Sum of presence binary maps predicted by the presence-only models for each sample size (5 to 3000) and the common quail's (*Coturnix coturnix*) real distribution (RD).

Portugal is observed, especially when sample sizes are higher than 250 presence points. This is more evident in southern Portugal, where the species is broadly present.

To further investigate this underestimation, we developed new models to account for the variation in the climatic space of Iberia. Mean March and April temperature and accumulated precipitation (the most contributing variable) on southern Portuguese occurrence localities (*n* = 144) showed a different climatic pattern (with higher mean temperatures but also precipitation above the mean) than the rest of the Mediterranean area of the Iberian Peninsula (Fig. 5). Also, this climatic area is represented in only 5% of the total sampled points in the Iberian Peninsula. Hence, it is possible that this underestimation error is caused by an uneven representation of this area in the whole Iberian climatic space, and models with low sample sizes are less affected by this. To overcome this potential bias in the models, we ran a model using presence points equally distributed in the Iberian Peninsula (*n* = 1992), and stratified them by the different fitoclimatic areas selected from the Spanish Bioclimatic Classification System (Rivas-Martínez

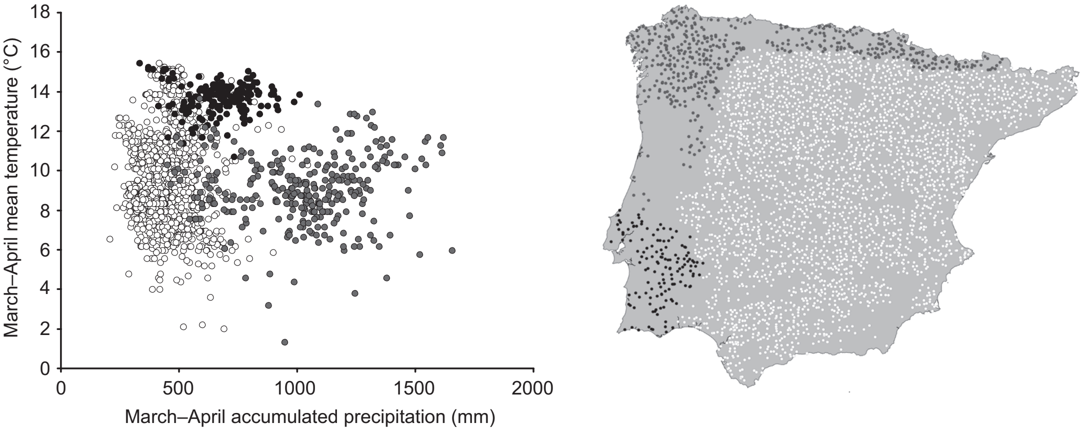


**Fig. 5.** Spatial distribution of common quail presences in the Iberian Peninsula related to March–April temperature and precipitation. Presences in main climatic areas are represented: Mediterranean climate (empty circles), Mediterranean climate with oceanic influence (black-filled circles) and oceanic climate (grey-filled circles).

1987) and the Atlas do Ambiente do Portugal (Amaral Franco 2007). In this model, the variables were similarly weighted, and March–April was not the most contributing variable (Table 1). The probability of occurrence of the species was higher in Portugal than the previous models (Appendix 2), yet the overall prediction of the species was not better than other models of lower sample size (real AUC = 0.59, training AUC = 0.63).

## Discussion

We modelled the distribution of a widespread species that occurs in many climates and is a habitat specialist. Modelling such species from a climatic point of view is a challenge, as the performance of species' distribution models has been shown to be negatively correlated with the niche width of the species, with generalist species being more poorly modelled (Stockwell & Peterson 2002). Quails can breed at all latitudes in the Iberian Peninsula, and at elevations lower than 2100 m a.s.l. with an alpine climate (Guyomarc'h 2003). In spite of being a habitat specialist (breeding in open habitats with dense herbaceous cover), this habitat can have different characteristics, with different plant species and varieties depending on the climatic properties of each agricultural area (Instituto Geográfico Nacional 1991). However, the most suitable habitats for the species are dry cereal cultures, which represent 19% of the surface of the Peninsula and 41% of the suitable habitat (CORINE 2000). Dry cereal cultures are mainly located in the Mediterranean climatic area of the Peninsula, encompassing the central plateau and southern and eastern Spain (Instituto Geográfico Nacional 1991).

### Modelling

The current distribution of the species covers less than one third of the Iberian Peninsula,

**Table 1.** Contribution of main variables affecting the climatic distribution of the common quail *Coturnix coturnix* from a model with stratified data.

| Variable | Contribution (%) | Overall effect |
|---|---|---|
| **Precipitation** | | |
| July–August | 15.7 | + |
| September–October | 15.7 | + |
| **Temperature** | | |
| July–August | 12.8 | + |
| **Precipitation** | | |
| March–April | 12.3 | – |
| May–June | 8.5 | – |
| November–December | 7.8 | – |

and the species is not present in all available suitable habitats. Some of the models run with sample sizes of five and ten did not predict better than random. However, others seemed to discriminate well between species' presence and absence. This can be interpreted as a chance result, depending on where presence points fell. In some cases, due to the heterogeneity of the species' climate requirements, it was not possible for MaxEnt to build a model, as the data were too random. The question of how few data points are needed to properly model species' distribution has been investigated before (Hernandez *et al*. 2006, Pearson *et al*. 2007), with different degrees of success. We were interested in addressing this issue, as well as the issue of the potential maximum or stability in model parameters and predictions towards this hypothetical optimum. In general, species distribution models from few locality data points, with the exception of species occurring in small areas or which are climatically narrow (Stockwell & Peterson 2002), perform poorly. When considering the whole set of predictive models and their randomizations, an increase in sample size also increases the predictive power of the models for the quail as well as decreases the variance between models per sample size category. However, a gradual increase towards the maximum number of localities is not observed in terms of overlap with the real distribution and predictive errors. Instead, there is a degree of stability when the model is based on more than 1000 locality data points.

We must take into account that the effective sample size is not the same as the actual sample size, and realize that this can affect the comparison of model performances. As the sample size increases, the proportion of overlap in the samples increases, and therefore the models are bound by the input data to be increasingly similar. This explains the important variability in predicting the real distribution and why it is reduced towards larger sample sizes. The models stabilize at 1000 presence points, which is approximately 30% of all the possible points in this study.

The most significant variable was the mean precipitation of March–April and its contribution increased with sample size. Areas with less precipitation during this period showed higher presence of the species. The quail is a migratory bird that depends on winter cereals (mainly wheat and barley) that represent most of the herbaceous cover in the Iberian Peninsula. Cereals do not need an important water input (Austin *et al*. 1998), and cover huge extensions in the Mediterranean area, which are characterized by lower precipitation (*see* Fig. 5), and represent a major part of the peninsula's surface. Thus, although the species is widespread in all of Eurasia, even in places with higher precipitation rates than the Iberian Peninsula (Guyomarc'h *et al*. 1998, Guyomarc'h 2003), some areas of the Iberian Peninsula are climatically miss-represented. A large increase in sample size in these areas with a dominant climate unbalances the models and cause significant errors. However, although the Iberian Peninsula is a good scenario as it presents a high variety of climatic conditions and habitats that the species can experience in its distribution range, only including data from part of the breeding range of the species could have potential effects. The distribution in Asia and Africa indicates that the species can tolerate a broader range of climatic conditions, some of them probably not found within the Iberian range. It is possible that if data from the broader distribution had been included, then climate variables might have had greater importance and the models might improve even for the Iberian proportion of the range. Unfortunately, such 'high quality' data for the whole distributional range of this species, especially in Africa and Asia, do not exist.

## Model evaluation

Some models based on small sample sizes predicted a very high percentage of the real distribution of the species, but also had important overestimation. Overestimation was reduced considerably when sample size increased. However, overestimation was important even with large sample sizes. In presence-only models overestimation is actually a relative error (Zaniewski *et al*. 2002). In our case, overestimation is probably showing the potential climatic niche of the species without habitat constraints. These results show that it is risky to generate current distribution maps based on climate factors only, when the species is a habitat specialist. In our case, overestimation represents around 50% of the real distribution and increasing the sample size did not reduce the median error but reduced the interquartile range.

Although models had considerable overestimation, correcting by the suitable habitat reduced this ‘error’ significantly. Our results suggest that models based on few samples may give the wrong impression of predicting the distribution of a species well. In fact they are not doing that, and this cannot be confirmed if the species’ real distribution is not known, which is true in most cases.

Surprisingly, the highest training AUCs were reached when sample sizes were 100, and therefore, those could be considered the best models according to this index. In contrast, real and test AUCs showed a different pattern when we increased sample size. Compared with the real distribution, models with 100 presence points show lower overestimation, higher underestimation and lower overlap with the real distribution, showing a more conservative distribution model. This coincides with the sample size that has the most homogeneous variable contribution, as the climate space is represented more evenly. However, when the sample size increases to over 250 presence points, distribution models seem to be biased towards a Mediterranean climate with lower precipitation where southern Portugal is under-represented.

An increase in sample size represents an improvement of the accuracy of the model to represent the current distribution of the species until certain limits. The real AUC, test AUC and direct comparisons of quail distribution models with the real distribution are indices that give a measure of how models fit the actual current distribution. However, training AUC is an index that weighs up when predictor variables are better represented. The current distribution therefore is not shaped only by climatic factors, but other factors such as geography, agricultural history, human needs or economy. These are important factors that define the agricultural landscape and influence the habitat and the species’ distribution.

Our results suggest that large sample sizes perform better, yet models that are unable to fully represent the real distribution of the species show a high degree of overestimation. However, if we consider the training AUC results, it seems that there is an optimum sample size where models perform better around 100 points. This is probably because in our case, this sample size provides a more adequate weighting of the predictor variables. This suggests that stratified sampling of localities by climate for modelling may be a good approach when only climate data are available to model such species. Many species are not habitat specialists, or their habitat preferences are completely unknown, making it more challenging to model their distribution. Hence, in such cases, and whenever possible, habitat variables should be included *a priori* or *a posteriori*. In our case, the *a posteriori* correction of the models with habitat significantly decreased overestimation to less than 20% of the species distribution. Although climate models can be corrected by habitat, in our case some areas of occurrence of the species were never predicted by the models.

## Conclusions

Species distribution models based on climate variables are broadly used to predict species distribution according to different climatic scenarios. These models are built from presences and/or absences in which the distribution is assumed to be in equilibrium with climate and habitat. However, the distribution of farmland species strongly depends on the presence of adequate farmland habitat, which at the same time depends on non-natural processes like agricultural needs and human management. Thus, these processes unbalance species occurrence within its climatic range, biasing the models. Studies that include only climate data as predictor variables may incur mismatches with real species’ distributions by not considering habitat variables.

Previous knowledge about the biology and requirements of the study species is crucial for optimal modelling. The omission of some predictor variables can generate significant errors.

These errors are even more relevant when trying to predict past or future distributions in a changing world where human activities (i.e., agriculture) have direct implications on species' distributions.

## Acknowledgements

We sincerely thank all volunteers who collected the bird data for both atlases. We also thank José Domingo Rodríguez-Teijeiro, Manel Puigcerver and two anonymous reviewers who contributed to improve the manuscript with their helpful comments and suggestions. This work was supported by a predoctoral travel grant for research in foreign institutions (AGAUR 2007 BE-1 00214) to FSP and a Spanish Ministry of Science and Innovation grant (CGL2009-10198) to DRV.

## References

Amaral Franco, J. 2007. *Atlas Digital do Ambiente*. — Agência Portuguesa do Ambiente, Lisboa.

Araújo, M. B., Thuiller, W., Williams, P. H. & Reginster, I. 2005: Downscaling European species atlas distributions to a finer resolution: implications for conservation planning. — *Global Ecology and Biogeography* 14: 17–30.

Araújo, M. B. & New, M. 2007: Ensemble forecasting of species distributions.— *Trends in Ecology & Evolution* 22: 42–47.

Austin, R. B., Cantero-Martínez, C., Arrúe, J. L., Playán, E. & Cano-Marcellán, P. 1998: Yield-rainfall relationships in cereal cropping systems in the Ebro river valley of Spain. — *European Journal of Agronomy* 8: 239–248.

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. 1991: BIOCLIM — a bioclimate analysis and prediction system. — *Plant protection quarterly* 6: 8–9.

Carpenter, G., Gillison, A. N. & Winter, J. 1993: DOMAIN: a flexible modeling procedure for mapping potential distributions of plants, animals. — *Biodiversity and Conservation* 2: 667–680.

Donald, P. F., Green, R. E. & Heath, M. F. 2001: Agricultural intensification and the collapse of Europe's farmland bird populations. — *Proceedings of the Royal Society London B* 268: 25–29.

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

Ferrier, S., Watson, G., Pearce, J. & Drielsma, M. 2002: Extended statistical approaches to modelling spatial pattern in biodiversity: the north-east New South Wales experience. I. Species-level modelling. — *Biodiversity and Conservation* 11: 2275–2307.

Fielding, A. H. & Bell, J. F. 1997: A review of methods for the assessment of prediction errors in conservation presence/absence models. — *Envirommental Conservation* 24: 38–49.

Gonçalves, D. 2008: *Coturnix coturnix*. — In: *Atlas das aves nidificantes em Portugal (1999–2005)*. Instituto da Conservação da Naturaleza e da Biodiversidade, Sociedade Portuguesa para o Estudo das Aves, Parque Natural da Madeira e Secretaria Regional do Ambiente do Mar. Assirio & Alvim, Lisboa.

Graham, C. H., Ferrier, S., Huettmann, F., Moritz, C. & Townsend Peterson, A. 2004: New developments in museum-based informatics and applications in biodiversity analysis. — *Trends in Ecology and Evolution* 19: 497–503.

Gregory, R. D., van Strien, A., Vorisek, P., Gmelig Meyling, A. W., Noble, D. G., Foppen, R. P. B., Gibbons, D. W. 2005: Developing indicators for European birds. — *Philosophical Transactions of the Royal Society B* 360: 269–288

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

Guisan, A., Graham, C. H., Elith, J., Huettmann, F. & NCEAS Predicting Species Distribution Working Group 2007: Sensitivity of predictive species distribution models to change in grain size. — *Diversity and Distributions* 13: 332–340.

Guyomarc'h, J. C., Combreau, O., Puigcerver, M., Fontoura, P. A. & Aebischer, N. J. 1998: Quail. — In: Olgivie, E. (ed.), *Birds of the Western Palaearctic*, update 2: 27–46. Oxford University Press, Oxford.

Guyomarc'h, J. C. 2003: Elements for a common quail (*Coturnix c. coturnix*) management plan. — *Game and Wildlife Science* 20: 1–92.

Hernandez, P. A., Graham, C. H., Master, L. L. & Albert, D. L. 2006: The effect of sample size and species characteristics on performance of different species distribution modeling methods. — *Ecography* 29: 773–785.

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.

Instituto Geográfico Nacional 1991: *Atlas Nacional de España*. — Ministerio de Obras Públicas y Transportes, Dirección General del Instituto Geográfico Nacional, Madrid.

Jiménez-Valverde, A. & Lobo, J. M. 2006: The ghost of unbalanced species distribution data in geographical model predictions. — *Diversity and Distributions* 12: 521–524.

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

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.

Ninyerola, M., Pons, X. & Roure, J. M. 2005: *Atlas climático digital de la Península Ibérica: metodología y aplicaciones en bioclimatología y geobotánica*. — CREAF, Bellaterra.

Pearce, J. & Boyce, M. 2006: Modelling distribution and abundance with presence-only data. — *Journal of Applied Ecology* 43: 405–412.

Pearson, R. G., Raxworthy, C. J., Nakamura, M. & Peterson, A. T. 2007: Predicting species' distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. — *Journal of Biogeography* 34: 102–117.

Phillips, S. J., Dudik, M. & Schapire, R. E. 2004: A maximum entropy approach to species distribution modeling. — In: Brodley, C. E. (ed.), *Proceedings of the 21st International Conference on Machine Learning*: 655–662. ACM Press, New York.

Phillips, S. J., Anderson, R. P. & Schapire, R. E. 2006: Maximum entropy modelling of species geographic distributions. — *Ecological Modelling* 190: 231–259.

Phillips, S. J. & Dudik, M. 2008: Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. — *Ecography* 31: 161–175.

R Development Core Team 2005: *R: A language and environment for statistical computing*. — R Foundation for Statistical Computing, Vienna, Austria, available at http://www.R-project.org.

Rivas-Martínez, S. 1987: *Memoria del mapa de las Series de Vegetación de España*. — Instituto para la conservación de la Naturaleza, Madrid.

Rodríguez-Teijeiro, J. D., Puigcerver, M. & Gallego, S. 2003: *Codorniz común*. — In: Martí, R. & del Moral, J. C. (eds.), *Atlas de las Aves Reproductoras de España*: 218–219. Dirección General de Conservación de la Naturaleza-Sociedad Española de Ornitología, Madrid.

Rodríguez-Teijeiro, J. D., Sardà-Palomera, F., Nadal, J., Ferrer, X., Ponz, C. & Puigcerver, M. 2009: The effects of mowing and agricultural landscape management on population movements of the common quail. — *Journal of Biogeography* 36: 1891–1898.

Segurado, P. & Araújo, M. B. 2004: An evaluation of methods for modelling species distributions. — *Journal of Biogeography* 13: 1555–1568.

Soberón, J. & Nakamura, M. 2009: Niches and distributional areas: Concepts, methods, and assumptions. — *Proceedings of the National Academy of Science* 106: 19644–19650

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

VanDerWal, J., Shoo, L. P., Graham, C. & Williams, S. E. 2009: Selecting pseudo-absence data for presence-only distribution modelling: How far should you stray from what you know? — *Ecological Modelling* 220: 589–594.

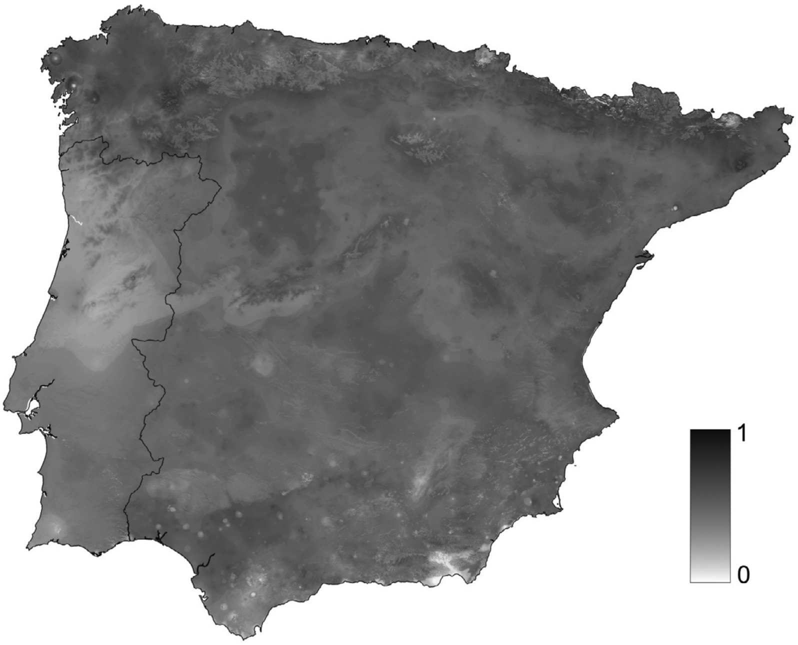
Vieites, D. R., Nieto-Román, S. & Vences, M. 2008: Towards understanding the spatial pattern of amphibian diversity in Madagascar. — In: Andreone, F. (ed.), *A conservation strategy for the amphibians of Madagascar*: 307–410. Monografie XLV, Museo Regionale di Scienze Naturali, Torino.

Vieites, D. R., Nieto-Román, S. & Wake, D. B. 2009: Reconstruction of the climate envelopes of salamanders and their evolution through time. — *Proceedings of the National Academy of Sciences* 106: 19715–19722.

Wisz, M. S., Hijmans, R. J., Peterson, A. T., Graham, C. H., Guisan, A. & NCEAS Predicting Species Distribution Working Group 2008: Effects of sample size on the performance of species distribution models. — *Diversity and Distributions* 14: 763–773.

Zaniewski, A. E., Lehmann, A. & Overton, J. M. C. 2002: Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. — *Ecological Modelling* 157: 261–280.

**Appendix 1.** Real distribution (in grey) of the common quail *Coturnix coturnix* in the Iberian Peninsula, created from the recent atlas from Spain (Rodríguez-Teijeiro *et al.* 2003) and Portugal (Gonçalves 2008), and CORINE land cover 2000. Random presence points ($n = 3639$) are shown.



**Appendix 2.** Optimal climate map for the presence of the common quail *Coturnix coturnix* in the Iberian Peninsula.