The application of distribution models and Geographical Information Systems for the study of biogeography and conservation of herpetofauna: *Chioglossa lusitanica* as a case study

JOSÉ TEIXEIRA¹² & NUNO FERRAND¹²

¹Centro de Estudos de Ciência Animal (CECA/UP), Campus Agrário de Vairão, 4485 - 661 Vairão, Portugal
²Departamento de Zoologia e Antropologia, Faculdade de Ciências, Universidade do Porto, Praça Gomes Teixeira, 4050 Porto, Portugal.

Summary: Environmental processes are complex, multidimensional and vary across different time scales. The study of this complexity is frequently based on the definition of rules or assumptions in the form of mathematical or logical expressions describing how these processes function. Developing and utilising such expressions is termed environmental modelling, a methodology that aims to better understand ecological processes and give scientific support to the management of natural resources.

The study of limiting factors of the distribution of species and the elaboration of descriptive models had a great impulse in the last decade, with the generalisation of the use of Geographical Information Systems (GIS). However, GIS-based modelling has wider potential uses for ecological studies and conservation planning and several other applications have been explored. In this paper the application of these techniques for the study of biogeography and conservation of herpetofauna are reviewed, using *Chioglossa lusitanica* as a case study.

Keywords: Biogeography, Conservation, GIS, Distribution models, *Chioglossa lusitanica*.

**INTRODUCTION**

Geographical Information Systems (GIS) is an advanced software technology conceived to store, transform, combine and represent information with a strong spatial component (GOODCHILD, 1993). One of the major uses of GIS in ecology is the elaboration and representation of environmental models.
(GLM) – namely multiple logistic regression and discriminant function analysis – (e.g. Nicholls, 1989; Pereira & Itami, 1991; Fielding & Haworth, 1995; Austin et al., 1996; Corsi et al., 1999), generalised additive models (GAM) (e.g. Yee & Mitchell, 1991; Bio et al., 1998), classification and regression trees (Walker, 1990; Smith et al., 1997), bioclimatic models generated by the program BIOCLIM (Sutherland et al., 1995; McGuigan et al., 1998), canonical correspondence analysis (Hill, 1991), models based on Bayes’ theorem (Aspinal & Veitch, 1993), artificial neural networks (Guégan et al., 1998) and overlap analysis (Brito et al., 1999a).

Modelling wildlife distributions with multivariate statistics and GIS is increasingly used to identify suitable habitat and to predict species’ potential distributions. Amphibians and reptiles are among the organisms to which these techniques have been applied, and several studies focus on predicting species occurrence and identifying a set of environmental variables that better describes/explains their distribution (e.g. Brito et al., 1996; Romero & Real, 1996; Teixeira et al., 2001). However, several other applications have been explored, such as the identification of conservation areas from occurrence (Brito et al., 1999b) or species diversity data (Lombard et al., 1993). Sutherland et al. (1995) produced a simulation of the potential area of expansion of the introduced cane toad (Bufo marinus) in Australia, and McGuigan et al. (1998) simulated the distribution of the tree frog Litoria personiana during the Last Maximum Glacial period.

Our case study, the Golden-striped salamander, Chioglossa lusitanica, is endemic to the Iberian Peninsula. It lives alongside uncontaminated streams in mostly mountainous areas of the provinces Galicia and Asturias in Spain and in north-western Portugal (Teixeira et al., 1999). The salamander’s current conservation status is ‘insufficiently known’ in Portugal (SNPRCN, 1990) and ‘rare’ in Spain (Blanco & Rodrigues, 1992).

This work reviews some applications of GIS-based techniques for biogeographical analysis and conservation assessment of the species.

**Methods**

We investigated 374 localities in 270 UTM 10x10km grid cells across and bordering the previously known salamander range in Portugal, from 1994 to 1998. Areas highlighted as ‘suitable’ in preliminary analyses (Teixeira et al., 1996) – i.e. with probability of occurrence higher than 0.8 in logistic regression based models – were also explored. The salamander was observed to be present in 202 out of 869 grid cells (Fig. 1, Teixeira et al., 2001).

Digital data on 22 ecologically meaningful environmental and geographic variables were chosen for modelling the salamander distribution (see Teixeira et al., 2001). Nonparametric correlation (r·) and hierarchical clustering methods were used to evaluate the level to which this information was redundant. Parameters correlated at r·(> 0.8 were selected using the following criteria: i) data availability for Spain, ii) ease of use, and iii) promise in terms of the known salamander life history. The 13 parameters (independent variables) retained for spatial modelling were: soil pH (ACID), altitude (ALTI), slope (SLOP), water hardness (HARD), number of frost days (FROD), number of frost months (FROM), air humidity (HUMI), insolation (INSO), lithology, (LITH), a vegetation index (NDVI), total precipitation (PRET), mean annual temperature (TEMP) and mean July’s temperature (TJUL). Techniques used for spatial modelling were: i) logistic regression (LR), ii) discriminant analysis (DA), iii) classification trees (CT) and iv) overlap analysis (OA) (for details see Teixeira, 1999; Teixeira et al., 2001).

The spatial representation of the models was produced in the GIS software Idrisi for Windows v. 2.0 (Eastman, 1997). Because we only found nine environmental variables for Spain...
(IGN, 1992) from the thirteen used for Portugal, we produced additional models with those parameters to allow their extrapolation of the models for that country. This procedure aimed to test the performance of the modelling techniques with an independent data set and to help understand the biogeography of the species in Spain, where the knowledge of the distribution is not numerically well documented.

The potential salamander distribution under changing climatic conditions (2-3 °C temperature increase over 50-80 years (HadCM2 models) (IPCC, 2001)) was forecasted using LR and DA models (the easiest to manipulate in GIS environments among those used) (Teixeira & Arntzen, in press).

We also combined an ecological (LR) model of the distribution of *C. lusitanica* with synthetic genetic diversity maps constructed from multivariate analysis of allozyme data with the aim of identifying conservation units (Alexandrino, 2000; Alexandrino et al., submitted). Genetic data consisted of allele frequencies at six polymorphic allozyme loci (PGM1, PEPB, PEPC, PEPD, ADH and PGD) scored from 17 populations distributed across the whole range of *C. lusitanica* (Alexandrino et al., 2000). A Principal Components Analysis (PCA) was performed with the most common allele (frequency >0.05) at five nearly diallelic loci and four out of five alleles at the highly polymorphic PGM1 locus. The factor scores for the first two principal components (PC) resulting from PCA were used to construct geographical maps with the kriging interpolation procedure, using the Surfer 6.0 geostatistical software (Golden Software, 1996). The two maps (PC1 and PC2) were

---

**Figure 1.** A) Range of *C. lusitanica* on the Iberian Peninsula (adapted from Arntzen, 1999), B) distribution of *C. lusitanica* in Portugal. Solid dots correspond to occurrence, open dots to absence, and grey dots to literature citings covering cells in which the species was not observed (Teixeira et al., 2001).
overlayed and weighted for the variance explained by each of the PC factors and then combined with the LR potential distribution map of *C. lusitanica* in the GIS software Idrisi (EASTMAN, 1997). This procedure aims to enhance the areas with high genetic diversity that coincide with the areas of high habitat suitability.

### RESULTS AND DISCUSSION

**Distribution modelling**

The fit of the four descriptive models for Portugal was good, with correct classification scores ranging from 88.2% to 95.7% (Fig. 2). Although the CT-based model had slightly higher accuracy for Portugal and Spain (the validation area), LR and DA-based models showed better potential for manipulation with GIS and produced descriptive maps with higher resolution than CT- and OA-based models.

In total 11 variables were selected in the different modelling techniques (Fig. 2). From these, only precipitation was selected in all models, appearing to be the major explicative variable of *C. lusitanica* distribution, and few (e.g. slope, mean July’s temperature and a suite of soil-type characteristics) were consistently selected. Although some caution should be used with the ecological interpretation of the results because any environmental model is by definition a simplification of complex phenomena, the set of selected variables corroborates the empirically described dependence of the species to humid mountain areas of the north-western Iberian Peninsula (GOUX, 1957; ARNTZEN, 1981).

The maps derived from the models closely match the documented distribution of the salamander in Portugal (Figs. 1 and 2). The models show an east to west spatial constriction in the habitat suitability corresponding to the Mondego river basin. This result is concordant with the genetic structure of the species, which shows high differentiation between populations located on both sides of the river valley (ALEXANDRINO et al. 2000). Areas consistently classified as potentially suitable outside its documented range included the mountains of Montesinho, Malcata, Aire e Candeeiros, Sintra and Monchique. These areas provide important clues about the historical biogeography of the salamander. The existence of isolated populations of other amphibians and reptiles with similar habitat requirements and overall distribution pattern in some of these areas (e.g. *Rana iberica* and *Lacerta schreiberi*) suggests that the golden-striped salamander range did not extend so far south in the past.

The extrapolation of the models to Spain yielded a lower CCS (90.0% to 93.8%, Tab. 1). Although the predicted range covers the entire known distribution of the species in Spain, it also includes large areas where its presence has not been documented, including a zone from the Cantabrian Mts. to the Pyrenees, as well as several isolated mountain ranges of the Central System (Figs 1 and 3). The absence of the salamander is likely due to the lack of suitable habitat in these areas.
mander in north-eastern Spain from Asturias to the Pyrenees, in areas classified as suitable by the present models, suggests that either we did not include some important limiting variable for the species in Spain (e.g. the unavailability of soil type related variables could lead to overestimated predictions), or that the species is still expanding eastwards from the last glacial refugia located south of Douro river to yet unoccupied suitable habitats (Teixeira et al., 2001). Vences (1997) describes that the salamander range limit in Asturias coincides with a transition to calcareous soils, which seems to support the first hypothesis.

**Simulation of potential effects of climate change**

The simulation of the impact of global warming for Portugal predicted a substantial reduction of the salamander’s range between 15 to 25%, for 2050, and 19 to 35%, for 2080 (Fig. 4, Teixeira & Arntzen, in press). Al-

---

**Figure 2.-** Probability of occurrence of *C. lusitanica* in Portugal as given by: A) logistic regression analysis (Teixeira et al., 2001), B) Discriminant analysis, C) Classification Tree, and D) Overlap analysis.
though a 3 °C change in mean annual temperature corresponds to a shift in isotherms of 300-400 km in latitude (Hughes, 2000), it was not predicted a range shift, but an overall range reduction and a fragmentation in the areas corresponding to the Mondego and Douro rivers. The pattern of range reduction is coincident with the scenario of past allopatry associated with these two river valleys, inferred from genetic studies (ALEXANDRINO et al., 2000).

The forecasted increase in temperature for the end of this century exceeds that of any period over the last 120,000 years and it remains unclear how species and ecosystems will respond to these unparalleled changes (GRAHAM & GRIMM, 1990). The direct effects of temperature on amphibian physiology and survival are well known (e.g. GOUX, 1957; WHITFORD, 1973; Feder, 1978; ROME et al., 1992; TRIPEPI et al., 1998). However, temperature also indirectly affects many environmental variables important for amphibians, such as soil moisture and air humidity, vegetation composition, permanency of brooks and concentration of

Figure 3.- Probability of occurrence of *C. lusitanica* in Spain as given by: A) logistic regression analysis (Teixeira et al., 2001), B) Discriminant analysis, C) Classification Tree, and D) Overlap analysis.
Figure 4.- Logistic Regression (LR) and Discriminant analysis (DA) predictions of potential impact of climate change on *C. lusitanica* distribution in Portugal. A) Present day model, B) year 2050 scenario (increase of 2 °C), C) year 2080 scenario (increase of 3 °C) (Teixeira & Arntzen, *in press*).
dissolved oxygen in the water (e.g. THORSON SVIHLA 1943, JÆGER 1971, SUGALSKI CLAUSEN 1997). Recently it was also shown that temperature can produce synergetic effects with other factors, resulting in, for example, increased exposure of embryos to ultraviolet (UV-B) radiation and raising vulnerability to infections (KIESECKER et al., 2001). The highly complex influences of temperature on species complicate our ability to simulate the effects of climate change. Nonetheless, modelling is a valuable approach for gaining insight into the direction and magnitude of the repercussions of environmental change. Our models may help to highlight this potential threat and aid the development of conservation measures to minimise the impact of climate change, such as the identification and protection of stronghold areas and the development of habitat corridors that allow individuals to migrate to favourable areas in the future (TEIXEIRA & ARNTZEN, in press).

Representation of Genetic Structure

Multivariate analysis of allozyme genetic variation explained 70% of the total variation along the first axis, showing a south-north genetic differentiation (Fig. 5A). The second axis explained 25% of the total variation and shows a diffusion gradient centred in Serra de Montemuro in northern Portugal (Fig. 5B). The combination of both maps with the LR model for the Iberia Peninsula depicts the region south of Mondego river and the region between the Mondego and Douro rivers, as important areas for conservation of genetic diversity (Fig. 6) (ALEXANDRINO et al., submitted). The importance of genetic variation as a component of biodiversity is widely recognised (HUMPHRIES et al., 1995) and its assessment is central to identify and prioritise areas for monitoring, management and protection, but neither the concepts nor the methods for calculating the conservation value of spatial diversity have been resolved (MORITZ & FAITH, 1998). We anticipate that this approach can be successfully applied to a wide variety of organisms as a practical tool to aid the identification of biodiversity hotspots within single species, specially in refugial areas where intraspecific diversification has occurred (ALEXANDRINO et al., submitted).

Figure 5.- Principal Component Analysis of 9 independent allelic frequencies at the loci PGM1, PEPB, PEPC, PEPD, PGD and ADH in 17 populations of C. lusitanica: A) Synthetic map for the first principal component, representing 70% of the total variation in allele frequency; B) Synthetic map for the second principal component, representing 25% of the total variation (Alexandrino et al., submitted).
CONCLUSIONS

The ectothermal physiology and high ecological constraints of the amphibians and reptiles make them good organisms for modelling. The application of ecological models and Geographic Information Systems is here shown to be of great help for the study of biogeography and conservation of herpetofauna, namely through the identification of a set of environmental variables that describes organismal distribution, prediction of species occurrence and habitat suitability, simulation of the impact of environmental change and representation of genetic structure. These applications do not aim to represent the whole spectrum of potential uses of the techniques, but simply to exemplify several of them. In the near future it is expected that GIS, spatial statistics and computer capacity will continuously improve and newer applications will arise. Nonetheless, the great challenge of these techniques will be centred on gaining the needed confidence to apply them at a large scale to conservation planning and ecosystem management.

Acknowledgments

We thank J.W. Arntzen for assistance in developing the models and Steve Weiss and Patrick Osborne for comments on an earlier draft of the manuscript. The work was supported by the Instituto da Conservação da Natureza, the ‘LIFE’ program of the European Community, and by the Fundação para a Ciência e Tecnologia (PRAXIS XXI /BM/ 12745/97 MSc and SFRH/BD/3376/2000 PhD grants to JT).
REFERENCES


