Biogeography of the golden-striped salamander *Chioglossa lusitanica*: a field survey and spatial modelling approach

J. Teixeira, N. Ferrand and J. W. Arntzen

*Chioglossa lusitanica* is a streamside salamander endemic to the northwestern corner of the Iberian Peninsula. In an extensive field survey we found the species to be widely distributed across northwestern Portugal. The distribution data were used to identify macroenvironmental parameters related to the presence of the species in Portugal, and to produce a descriptive distribution model for *C. lusitanica*. We modelled the species distribution based on five environmental parameters (precipitation, relief, July temperature, number of frost months and hardness of water) selected by logistic regression analysis. The model yielded a high score of correct classification (95%). Excluding parameters unavailable for Spain the model included four parameters (precipitation, relief, July temperature and altitude) and had a 93% correct classification score. Extrapolation of the model to Spain yielded a correct classification of 92% and served as a validation of the modelling technique in an independent area. The models indicate the existence of several potentially suitable areas outside the known distribution of the species, providing clues about the species historical biogeography. The models, through the identification of highly suitable and stable protection regions, may contribute to informed conservation planning.

J. Teixeira (jteixeira@mail.icav.up.pt), N. Ferrand and J. W. Arntzen, Centro de Estudos de Ciência Animal, Campus Agrário de Vairão, PT-4485-661 Vairão, Portugal.
isolation in the early Pleistocene of populations separated by the Mondego river valley in the south of its present-day range (Alexandrino et al. 1997, 2000). Despite some detailed ecological studies (Goux 1957, Arntzen 1981, 1994a, b, 1995, Vences 1990) the distribution of the species was until recently (Sequeira et al. 1996) poorly documented. Accordingly, *C. lusitanica* has the conservation status of “insufficiently known” in Portugal (Anon. 1990), “rare” in Spain (Blanco and González 1992) and “indeterminate” in the IUCN red list (Anon. 1994). We here describe the biogeography of *C. lusitanica* in Portugal using the complementary approaches of field surveying and spatial modelling. The distribution data were used to identify macroenvironmental parameters related to the presence of the species in Portugal, and to produce a descriptive distribution model for *C. lusitanica*. The model was then applied to predict the species probability of occurrence across the Iberian Peninsula and to improve the understanding of its historical biogeography.

**Methods**

**Distribution**

The field survey was conducted in Portugal from 1994 to 1998 on a UTM 10 × 10 km grid basis. In each grid cell up to four potentially suitable habitats were surveyed for 45 min. Adults and other post-metamorphic salamanders were searched for under stones, litter and moss, while larvae were dipnetted from the backwater sections of brooks. A preliminary survey was restricted to the area circumscribed by 44 known records (Crespo and Oliveira 1989). The full survey encompassed all areas: 1) identified as “suitable” in preliminary analyses (probability of occurrence > 0.8 – including the mountains of Montesinho, Malcata, Aire e Candeeiros, Sintra and Monchique); 2) such that the documented range was delimited by a border of grid cells in which the species remained unrecorded. In total 374 localities were inspected over 281 grid cells.

**Environmental data**

Twenty-one ecologically meaningful variables were selected for analysis (Table 1). An altitude map was taken from the US Geological Survey (http://edcwww.cr.usgs.gov/doc/edchome/datasets/edcdata.html) and used to produce a relief map using a set of filter operations (Anon. 1997a). A vegetation index map (NDVI, Normalised Difference Vegetation Index) was obtained courtesy of the Royal Dutch Meteorological Institute (KNMI). Maps on the mean January and July temperature were digitised from the Portuguese Climate Atlas (Anon. 1974). Information on the remaining 16 parameters was available in digital form at DGA (Anon. 1995). Nonparametric correlation (rs) and hierarchical clustering were used to evaluate the level to which information appeared redundant. Thus, for parameters correlated at rs /p<0.8 a single variable was selected for analysis in Portugal.

Table 1. Environmental parameters used to model the distribution of *C. lusitanica* in Portugal.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Code</th>
<th>Selected for analysis</th>
<th>Available for Spain</th>
<th>G-test</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acidity of the soil (pH)</td>
<td>ACID</td>
<td>YES</td>
<td>NO</td>
<td>482.5</td>
<td>Anon. (1995)</td>
</tr>
<tr>
<td>Altitude (m)</td>
<td>ALTI</td>
<td>YES</td>
<td>YES</td>
<td>128.0</td>
<td><a href="http://edcwww.cr.usgs.gov/">http://edcwww.cr.usgs.gov/</a></td>
</tr>
<tr>
<td>Chlorates content of subterranean water (Cl⁻ mg l⁻¹)</td>
<td>CHLO</td>
<td>NO</td>
<td>NO</td>
<td>471.1</td>
<td>Anon. (1995)</td>
</tr>
<tr>
<td>Water drainage (mm yr⁻¹)</td>
<td>DRAI</td>
<td>NO</td>
<td>NO</td>
<td>947.4</td>
<td>Anon. (1995)</td>
</tr>
<tr>
<td>Evapotranspiration (mm yr⁻¹)</td>
<td>EVAP</td>
<td>NO</td>
<td>YES</td>
<td>798.3</td>
<td>Anon. (1992), Anon. (1995)</td>
</tr>
<tr>
<td>Frost days (number)</td>
<td>FROD</td>
<td>YES</td>
<td>YES</td>
<td>8.8</td>
<td>Anon. (1992), Anon. (1995)</td>
</tr>
<tr>
<td>Frost months (number)</td>
<td>FROM</td>
<td>NO</td>
<td>YES</td>
<td>11.3</td>
<td>Anon. (1995)</td>
</tr>
<tr>
<td>Hardness of subterranean water (CaCO₃ mg l⁻¹)</td>
<td>HARD</td>
<td>YES</td>
<td>NO</td>
<td>443.2</td>
<td>Anon. (1995)</td>
</tr>
<tr>
<td>Humidity of the air (%)</td>
<td>HUMI</td>
<td>YES</td>
<td>YES</td>
<td>41.0</td>
<td>Anon. (1992), Anon. (1995)</td>
</tr>
<tr>
<td>Insolation (h)</td>
<td>INSO</td>
<td>YES</td>
<td>YES</td>
<td>548.9</td>
<td>Anon. (1992), Anon. (1995)</td>
</tr>
<tr>
<td>Lithology</td>
<td>LITH</td>
<td>YES</td>
<td>NO</td>
<td>217.4</td>
<td>Anon. (1995)</td>
</tr>
<tr>
<td>Vegetation index</td>
<td>NDVI</td>
<td>YES</td>
<td>YES</td>
<td>11.9</td>
<td>Royal Dutch Meterol. Inst. (KNMI)</td>
</tr>
<tr>
<td>Precipitation days (d yr⁻¹)</td>
<td>PRED</td>
<td>NO</td>
<td>YES</td>
<td>553.7</td>
<td>Anon. (1992), Anon. (1995)</td>
</tr>
<tr>
<td>Annual total precipitation (mm yr⁻¹)</td>
<td>PRET</td>
<td>YES</td>
<td>YES</td>
<td>891.3</td>
<td>Anon. (1992), Anon. (1995)</td>
</tr>
<tr>
<td>Solar radiation (kcal cm⁻²)</td>
<td>RADI</td>
<td>NO</td>
<td>YES</td>
<td>612.9</td>
<td>Anon. (1992), Anon. (1995)</td>
</tr>
<tr>
<td>Residual content of subterranean water (mg l⁻¹)</td>
<td>RESI</td>
<td>NO</td>
<td>NO</td>
<td>552.7</td>
<td>Anon. (1995)</td>
</tr>
<tr>
<td>Slope (%)</td>
<td>SLOP</td>
<td>YES</td>
<td>YES</td>
<td>438.1</td>
<td>produced from ALTI</td>
</tr>
<tr>
<td>Sulphates content of subterranean water (SO₄²⁻ mg l⁻¹)</td>
<td>SULP</td>
<td>NO</td>
<td>NO</td>
<td>448.48</td>
<td>Anon. (1995)</td>
</tr>
<tr>
<td>Annual mean temperature (°C)</td>
<td>TEMP</td>
<td>YES</td>
<td>YES</td>
<td>399.7</td>
<td>Anon. (1992), Anon. (1995)</td>
</tr>
<tr>
<td>July mean temperature (°C)</td>
<td>TJUL</td>
<td>YES</td>
<td>YES</td>
<td>574.6</td>
<td>Anon. (1974), Anon. (1992)</td>
</tr>
</tbody>
</table>
retained, selected on the basis of: 1) data availability for Spain, 2) promise in terms of known salamander life history, and 3) ease of use. Thirteen parameters were retained for spatial modelling (Table 1). Variables were introduced into a Geographical Information System (GIS; Idrisi 2.007 – Eastman 1997; Iwits 2.1 Anon. 1997a) as raster layers with 1 km spatial resolution. Values for UTM grid cells were mean (or modal, for the categorical variable LITH) pixel values. Cells covering <50% Portuguese terrestrial territory were excluded from the analysis.

Analysis and modelling

We choose to analyse the distribution of C. hispanica by logistic regression because this technique: 1) has been shown to be a powerful analytical tool, capable of analysing the effects of one or several independent variables, discrete or continuous, over a dichotomous variable (Hosmer and Lemeshow 1989), 2) relies on fewer statistical assumptions than its alternatives and 3) frequently produces robust results (Austin et al. 1996, Brito et al. 1999). Logistic regression takes the form: $\pi(x) = \frac{e^{\beta_0 + \beta_1x_1 + \beta_2x_2 + \cdots + \beta_px_p}}{1 + e^{\beta_0 + \beta_1x_1 + \beta_2x_2 + \cdots + \beta_px_p}}$, in which $\pi(x)$ represents the probability of occurrence of the target species and $g(x) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \cdots + \beta_px_p$, where $\beta_0$ is a constant, and $\beta_1, \beta_2, \ldots, \beta_p$ are the coefficients of the variables $x_1, x_2, \ldots, x_p$ included in the equation (Hosmer and Lemeshow 1989). Logistic regression analyses were performed with SPSS 8.0 for Windows (Anon. 1997b) by a forward stepwise addition with Bonferroni’s correction to the initial $\alpha = 0.05$ (Rice 1989). The variables were presented in the final equations by order of entrance in the models. Individual variables were tested for explanatory power with G-statistics in univariate logistic regression analysis.

To build distribution models from presence/absence data a preponderance of absence data may be desirable because of the larger environmental variation expected in areas where the target species is absent (Pereira and Itami 1991). This procedure, however, frequently produces biased results (Hosmer and Lemeshow 1989) and we equilibrated the impact of presences and absences through a weighting procedure (Anon. 1997b). The correct classification score (CCS) denotes the percentage of correctly described presences and absences at $\pi(x) = 0.5$. The behaviour of the models was further described by the omission error rate (number of presences where absence was predicted/total number of predictions (N)) and the commission error rate (number of absences where presence was predicted/N). Validation of the model derived for Portugal was achieved by an extrapolation for Spain, on the basis of nine environmental variables for which data were available (Table 1, Anon. 1992). Because the knowledge of the species distribution in Spain is not numerically well documented, we used the range of the salamander (Arntzen 1999) to evaluate model fit.

Results

Chioglosa hispanica was observed in 202 Portuguese UTM grid cells, including 198 grid cells out of 869 in the analysis (Fig. 1). The salamander was not found in any of the isolated areas identified as potentially suitable in the preliminary model (the mountains of Montesinho, Malcata, Aire e Candeiros, Sintra and Monchique).

The spatial model included five variables as follows: $g(x) = 5.499 + 0.009\text{PRET} - 0.687\text{HARD} - 0.799\text{FROM} + 0.082\text{SLOP} - 0.594\text{TJUL}$. CCS was 95% with no marked difference for presences and absences. The omission error rate was 0.012 and the commission error rate 0.168. The map derived from the model (Fig. 2) matches closely with the documented distribution of the salamander (Fig. 1). Areas classified as potentially suitable outside its documented range included the mountains of Montesinho, Malcata, Aire e Candeiros, Sintra and Monchique, in line with the results from the preliminary analysis. The model shows an east to west spatial constriction corresponding with the Mondego river basin.

The model produced with the nine environmental variables for which data were available for Spain was $g(x) = -0.017 + 0.009\text{PRET} + 0.131\text{SLOP} - 0.004\text{ALTI} - 0.521\text{TJUL}$. CCS of this model for Portugal was 93%. The extrapolation of the model to Spain yielded a CCS of 92% (98% of the species presence area and 91% of the species absent area) with omission and commission error rates of 0.002 and 1.091, respectively. The predicted range covers the known range for Spain entirely. The model also classifies large areas as suitable for the species, where its presence has not been documented, including a zone from the Cantabrian Mts up to and including the Pyrenees, as well as several isolated mountain ranges of the Central System (Figs 1 and 3).

Discussion

The field survey markedly increased the knowledge of the distribution of C. hispanica in Portugal. The main results are that: 1) the species’ range extends in the south to near the Tejo river, and 2) the distribution is continuous at the selected spatial scale. The species was not found in five grid cells listed by Malkmus (1995), three of which correspond to old and imprecise records (e.g., “Coimbra” – Bocage 1864) and thus they may have been allocated to an incorrect UTM grid cell. Following Arntzen (1981) we consider the old record
for Elvas (Boscá 1880) erroneous because local habitat characteristics are clearly unfavourable. We failed to find the species in Sintra Mts where it was introduced ca 60 yr ago (Seabra 1943) and last seen in 1993 (G.-D. Guex in Arntzen 1999).

The distribution of _C. lusitanica_ is well described by a limited number of environmental parameters, although it should be kept in mind that the good fit of a model does not necessarily imply correct inference of causation (James and McCulloch 1990). Areas determined as suitable for _C. lusitanica_ are hilly or mountainous with a mild summer and winter climate, high precipitation and soft water. This set of parameters to some extent confirms the empirically described dependence of the species to humid mountain areas (Goux 1957, Arntzen 1981).

Teixeira (1999), using different modelling techniques (overlap analysis, discriminant analysis and classification trees), obtained similar results: precipitation appeared in all the models; slope, mean July temperature and a soil-related characteristic (hardness of water, acidity of soil or lithology) were consistently included, and altitude and the number of frost months were occasionally incorporated. As measured by their performance in univariate analysis the variables contribute to explain individually the distribution of the species in order of PRET, TJUL, INSO, ACID, HARD and SLOP.

In view of these results, precipitation appears to be the major explicative variable for the occurrence of golden-striped salamander. This is not surprising because precipitation determines most of the ecological requirements of the species: humidity of the air and soil moisture, vegetation composition and permanency and water flow of the brooks. The high relief of the terrain is frequently referred to as a typical characteristic of _C. lusitanica_ habitats (Vences 1997, Teixeira et al. 1999), being often associated with the presence of clean and fast-running brooks with little human disturbance. Temperature also seems to play an important role in the distribution of the species since it is absent from the regions with mean July’s temperature above 22.5°C. In fact, Goux (1957) reports critical thermal values for _C. lusitanica_ in captivity as 28°C for adults and 25°C for larvae. In contrast, _Salamandra salamandra_ has critical temperature values of 35°C for adults and juveniles (Degani 1994), which appears to be reflected in a much wider distribution including the southern Iberian Peninsula. Temperature, directly or indirectly, also affects other important environmental variables such as soil moisture and air humidity, permanency of brooks and concentration of dissolved oxygen in the water. Water
hardness serves as an indicator of soil chemistry, with high values in calcareous regions. The influence of this parameter on the distribution of the species may be related with the lower capacity of these substrates to retain water. Indeed, Jaeger (1971) suggested that soil moisture affected the distribution of some American salamander species.

An intriguing situation occurs in Spain where the model suggests a zone of suitable habitats from Galicia to the eastern Pyrenees. This discrepancy with the known distribution may in part be caused by the absence in the Spanish model of the parameters HARD and FROM. The eastern limit of the *C. lusitanica* range appears to coincide with a transition to calcareous soils (i.e. the species is absent in regions with hard water, Vences 1997). Alternatively, either our model does not locally apply because the distribution of the species is limited by different variables in different geographical regions (Brown and Lomolino 1998), or northern Spain provides suitable habitat that is not yet occupied by the species.

The logistic regression models proved helpful in the study of the biogeography of the golden-striped salamander, in particular to describe its present-day distribution, to formulate biogeographical hypotheses to explain its absence in apparently suitable areas and to plan fieldwork surveys. The model may also contribute to informed conservation planning, through the identification of highly suitable and stable protection regions, and by simulating the effects of environmental change.
Acknowledgements

We thank Pedro Beja, José Carlos Brito, James Harris and Steve Weiss for comments and João Alexandrino, Fernando Sequeira and others (too many to mention them all) for assistance in the field. The work was supported by the Instituto da Conservação da Natureza, the “LIFE” program of the European Union, and by the Fundação para a Ciência e Tecnologia (PRAXIS XXI/BM/12745/97 and SFRH/BD/3376/2000 grants to JT, and PRAXIS XXI/BCC/11959/97 grant to JWA).

References


Anon. 1997b. SPSS for Windows. Release 8.0. – SPSS.


