Modelling a species in expansion at local scale: is *Hyla meridionalis* colonising new areas in Salamanca, Spain?

Neftalí Sillero

*Centro de Investigação em Ciências Geo-Espaciais (CICGE), Universidade do Porto, Departamento de Matemática Aplicada, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal. E-mail: neftali.pablos@fc.up.pt*

**Abstract.** New records of the frog *Hyla meridionalis* have been reported in the province of Salamanca (Spain). However, it is not clear if those records correspond to an expansion process or to a lack of sampling. Thus, the species range was modelled in order to identify which is its potential distribution, which areas are available for expansion, and which variables drive the distribution. Ten MAXENT iterative models were produced at 1×1 km scale. The suitable areas in Salamanca occurred only in the south-east part of the province. This area corresponds mainly to the Alagón river basin, together with a small part of the Tormes river basin. The presence area was 1040 km², meanwhile the absence one corresponded to 11309 km². The models indicated a strong relationship with temperature and precipitation factors. Predictive models cannot confirm or reject the expansion process of the species. Then, *H. meridionalis* could be in equilibrium (or nearly) with the environment and may not spread further, or in expansion, but the process is almost complete. If the species is not in expansion, new records recently obtained in Salamanca could be the consequence of previous lack of sampling. Also, the expansion speed could be very low to be detected in human time-life.

**Keywords.** *Hyla meridionalis*, Spain, Salamanca, ecological modelling, GIS, areal expansion.

---

**INTRODUCTION**

New records of the frog *Hyla meridionalis* (Anura, Hylidae) have been reported recently in the Central mountain range of two Spanish provinces: Salamanca (Sillero and Carretero, 2007) and Madrid (Aceituno, 2001; Martínez-Solano and Aceituno, 2001). It is not clear if those records correspond to an expansion process or to a lack of sampling (Martínez-Solano and Fernández-González, 2003; Merchán et al., 2004; Sillero and Carretero, 2007). The species occurs in several countries of north-western Africa and south-western Europe (García-París, 1997): Morocco, Algeria, Tunisia, Portugal, Spain, south and south-eastern France and north-western Italy (Castanet and Guyetant, 1989; Bons
and Geniez, 1996; Schleich et al., 1996; Tejedo and Reques, 2002; Emanueli and Salvidio, 2006; Loureiro et al., 2008), as well as Madeira, the Canary Islands and Menorca (Balearic Islands), where it was introduced (Pleguezuelos et al., 2002; Loureiro et al., 2008). In the Iberian Peninsula, its range is divided in three isolated nuclei (García-París, 1997; Tejedo and Reques, 2002): one in the central and south-western parts of the Peninsula, one in the north-eastern extreme of Spain (Catalonia region), and one isolated population in the Basque Country (northern Spain).

Phylogenetic analysis concluded that the northern Iberia and southern France populations come from northern Morocco by human introduction, as well as the Canary Islands (Recuero et al., 2007); the southern Iberian populations come from southern Morocco by natural expansion. Recuero et al. (2007) described a biogeographical scenario based on two possible expansion movements. The species expanded naturally from southern to northern Morocco, and from there to the Iberian Peninsula. Subsequently, the population of northern Morocco became extinct. This area was recolonized with a differentiated population from the south of Morocco, and introduced by humans in the Mediterranean coast of France, expanding from there eastwards, southwards and northwards.

The first record of *H. meridionalis* in Salamanca (Spain) was observed by Bueno (1991), being the first location north to the Central mountain range in the Iberian Peninsula. Posteriorly, Merchán et al. (2004) and Sillero and Carretero (2007) defined the species’ distribution limits, and García (2007) found the most northern record. Aceituno (2001) and Martínez-Solano and Aceituno (2001) found new populations in the south-western part of the Madrid province. Martínez-Solano and Fernández-González (2003) and Merchán et al. (2004) suggested that the species is currently in expansion across the Central mountain range, considered as an effective geographic barrier. In Salamanca, the species remains apparently isolated from the southern populations (Sillero and Carretero, 2007). Merchán et al. (2004) proposed the global climatic warming as the most probable cause of the expansion, notwithstanding alternative explanations, namely the lack of intense sampling in Salamanca. However, Martínez-Solano and Fernández-González (2003) considered Madrid a well-sampled area (García-París et al., 1989).

Ecological modelling (Guisan and Zimmermann, 2000) is a useful tool for predicting and monitoring the expansion of species (e.g., Peterson and Vieglais, 2001). Pearson (2007) distinguished between potential models and habitat suitability models. Potential models are calculated with presence-absence data and predict the species’ realize niche (i.e., those areas that are truly occupied by the species). The model output is the probability of the species to be found in a particular place. However, habitat suitability models are calculated only with presence data and predict the species’ fundamental niche (i.e., the full range of abiotic conditions within which the species is viable). Then, the model output is the probability of a particular habitat to be suitable for the species (Brito and Crespo, 2002). In other words, in potential models is possible to include other factors (history, competition and so on) limiting the species’ distribution (this is obtained using absence data). Only with presence data, the predictive model is calculated without any limiting factor, being possible to assess the equilibrium state of the species with the environment (Santos et al., 2009).

The whole distribution of *H. meridionalis* has been recently modelled at global scale (10 × 10 km) in two environmental scenarios: in the present and in a near future (Sillero, unpubl. data). Model results suggested that the species can expand along the Mediterranean
Modelling *Hyla meridionalis* expansion

basin and Central Europe (including the British islands), with other areas in Asia, Australia and South Africa. The species' current potential distribution is very similar to the future one, using the same variables topographical and climatic variables. These global models considered the entire Salamanca province as a suitable area. This result depends probably on the model scale (Kaliontzopoulou et al., 2008) and it is expected that a more restricted suitable area may be predicted when modelling at a local scale. However, there are not records at 1 × 1 km UTM squares or GPS resolution for the whole distribution of the species. This situation hampers to calculate predictive models at local scale over larger areas (e.g., Iberian Peninsula). In fact, GPS records only exist for Salamanca (Sillero and Carretero, 2007) and partially for Portugal (Loureiro et al., 2008). Therefore, the *Hyla meridionalis* fundamental niche (sensu Pearson, 2007) was modelled in the province of Salamanca to verify if the species can expand or not, and which factors are driving the expansion. Specifically, the following questions were addressed: (1) which is the potential distribution of the species, (2) which areas are available for expansion, (3) which are the variables driving the distribution.

**MATERIAL AND METHODS**

*Study area*

The province of Salamanca is located in Central Western Spain (Fig. 1a). It is mainly a 12500 km² plain with an average altitude of 700 to 900 m. This plain is disrupted to the south by the mountains of the Central mountain range (900-2425 m), and to the north-west by the Duero river canyon (750-112 m). The climate is Mediterranean, with low precipitation (average of 400 mm throughout the year), and summer drought, and with cool temperatures in winter. These general climate characteristics are modified perceptibly by the relief to the south and north-west. Thus, three main ecosystems can be distinguished in the province: 1) the flat area, which has a strong continental climate. It is subdivided into two landscapes considering geological and economical factors, a) the cereal fields to the north-east located over the outer limit of the Duero sedimentary basin which provides the deepest soils in the province, and b) the Dehesa, principal landscape of Salamanca, consisting of large livestock exploitations spotted with more or less dense sets of Mediterranean oak trees (*Quercus ilex*) that grow in shallow and stony soils where granite rocks do not outcrop. 2) The Arribes canyon, which is the smallest area, with a very abrupt relief and a typical Mediterranean microclimate that provides mild temperatures in the winter. 3) The southern mountainous area, which has an average altitude of 900-950 m at its base and reaches 2425 m at the Calvitero peak, the summit of the province. The precipitation in the mountain ranges increases with altitude and the average exceeds 700 mm per year.

*Species and Environmental data*

Sillero and Carretero (2007) collected 51 GPS records of *H. meridionalis* in Salamanca, which were used as precise presence records for the models. A total of 22 independent ecogeographical variables (EGVs; global Spearman's correlation value lower than 0.7) were used to calculate the models (Table 1). All variables were included in the final models. Agro-climatic maps from León Llamazares (1991) were scanned into images and georeferenced using six points of known coordinates (two middle points and the four map corners) in the ArcInfo 9 software (ESRI, 2000a). Then,
they were automatically vectorized with the ArcScan extension (ESRI, 2000b) and the contour vector lines were interpolated to a raster file of 1 km² pixel size with a Trend Surface algorithm, using the Contour Grid Extension (Stuckens, 2003) of ArcView 3.2a software (ESRI, 1996). The climatic data cover a period from 1964 to 1990.

**Predictive models**

The secretive behaviour and activity patterns of *H. meridionalis* (García-Paris, 2004) complicate the accurate determination of absences, especially when the species may be in expansion. Therefore, a general-purpose machine learning method, Maximum Entropy (MAXENT 3.2.1 software version; http://www.cs.princeton.edu/~schapire/MAXENT), was used to identify the EGVs related to the fundamental species niche and to locate the areas of probable occurrence (Phillips et al., 2004, 2006; Phillips and Dudik, 2008). It uses presence-only occurrence data and deals with continuous and categorical EGVs. MAXENT models consistently outperform more established methods, either presence-only methods (Bioclim, Domain) or presence–absence methods (GAM, GLM), especially when samples sizes are low (Elith et al., 2006; Hernandez et al., 2006). MAXENT selects the model with the highest maximum entropy, i.e. with the most uniform distribution. The expected value of each EGV (or its transform and/or interactions) should match its empirical average, i.e. the average value for a set of sample points taken from the species distribution data (Phillips et al., 2004,
Modelling *Hyla meridionalis* expansion

**Table 1.** Description, units, abbreviated name and mean ± standard deviation percentage contribution (highest contributions in bold) of the ecogeographical variables used for calculating the 10 iterative MAXENT models. All variables were aggregated to a spatial resolution of 1 km². See Methods section for details. Ad. = Adimensional.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Units</th>
<th>Abbreviation</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>CORINE</td>
<td>Ad.</td>
<td>corine</td>
<td>3.68 ± 1.72</td>
</tr>
<tr>
<td>Variability of precipitation deficit lower than 50 mm in June</td>
<td>%</td>
<td>def50 jun</td>
<td>12.2 ± 2.5</td>
</tr>
<tr>
<td>Variability of precipitation deficit lower than 50 mm in October</td>
<td>%</td>
<td>def50 oct</td>
<td>0.07 ± 0.15</td>
</tr>
<tr>
<td>Duration of the hot period</td>
<td>no. days</td>
<td>hot per</td>
<td>19.34 ± 2.25</td>
</tr>
<tr>
<td>Duration of the cold period</td>
<td>no. days</td>
<td>cold per</td>
<td>0.36 ± 0.19</td>
</tr>
<tr>
<td>Duration of the dry period</td>
<td>no. days</td>
<td>dry per</td>
<td>0.11 ± 0.32</td>
</tr>
<tr>
<td>Annual real evapotranspiration</td>
<td>mm</td>
<td>etr</td>
<td>0.35 ± 0.14</td>
</tr>
<tr>
<td>Annual potential evapotranspiration</td>
<td>mm</td>
<td>evp</td>
<td>0.06 ± 0.13</td>
</tr>
<tr>
<td>Date of first frost</td>
<td>no. days</td>
<td>spring froz</td>
<td>0.15 ± 0.3</td>
</tr>
<tr>
<td>Date of last frost</td>
<td>no. days</td>
<td>last froz</td>
<td>0.45 ± 0.44</td>
</tr>
<tr>
<td>Annual mean temperature</td>
<td>ºC</td>
<td>annual isot</td>
<td>0.02 ± 0.03</td>
</tr>
<tr>
<td>Mean temperature in the hot period</td>
<td>ºC</td>
<td>hot isot</td>
<td>5.72 ± 3.44</td>
</tr>
<tr>
<td>Annual mean precipitation</td>
<td>mm</td>
<td>annual isoy</td>
<td>0.19 ± 0.2</td>
</tr>
<tr>
<td>Winter mean precipitation</td>
<td>mm</td>
<td>winter isoy</td>
<td>6.66 ± 4.47</td>
</tr>
<tr>
<td>Autumn mean precipitation</td>
<td>mm</td>
<td>autumn isoy</td>
<td>3.9 ± 2.4</td>
</tr>
<tr>
<td>Spring mean precipitation</td>
<td>mm</td>
<td>spring isoy</td>
<td>0.58 ± 0.38</td>
</tr>
<tr>
<td>Summer mean precipitation</td>
<td>mm</td>
<td>summer isoy</td>
<td>1.06 ± 1.47</td>
</tr>
<tr>
<td>no. days of August included in the hot period</td>
<td>no. days</td>
<td>hot august</td>
<td>0 ± 0</td>
</tr>
<tr>
<td>no. days of June included in the hot period</td>
<td>no. days</td>
<td>hot jul</td>
<td>0.84 ± 1.03</td>
</tr>
<tr>
<td>no. days of May included in the cold period</td>
<td>no. days</td>
<td>cold may</td>
<td>44.16 ± 4.19</td>
</tr>
<tr>
<td>no. days of October included in the cold period</td>
<td>no. days</td>
<td>cold oct</td>
<td>0.1 ± 0.1</td>
</tr>
<tr>
<td>Altitude a.s.l.</td>
<td>m</td>
<td>altitude</td>
<td>0.01 ± 0.02</td>
</tr>
</tbody>
</table>

2006; Phillips and Dudik, 2008). MAXENT selects uniformly pseudo-absence data randomly from the background squares. Hence, 10 iterative models were run automatically (using the MAXENT commands) with autofeatures (Araújo and New, 2007; Martínez-Freiría et al., 2008). Each EGV is weighted by a constant and the estimated probability distribution ranges from 0.0 to 1.0. The program starts with a uniform probability distribution and iteratively alters the gain one weight at a time to maximize the likelihood of the occurrence data set, converging to the optimum probable distribution. The gain is a measure of the likelihood of the samples or how much better the distribution fits the sample points than the uniform distribution does (Phillips et al., 2004, 2006; Phillips and Dudik, 2008). MAXENT was run randomly selecting 75% (38 records) of the presence records as training data and 25% as test data (12 records). MAXENT calculates the area under the curve (AUC) of the receiver operated characteristics (ROC) plots of the 10 iterative models (Liu et al., 2005; Pearson, 2007). The ROC curve describes the relationship between the proportion of observed
presences correctly predicted (sensitivity) and the proportion of observed absences incorrectly predicted (1 – specificity). Therefore, the ROC curve of a perfect model follows the left axis and top of the plot, and the curve of a random model follows the 1:1 line. The AUC expresses the proportion of the total area of the square defined by the axes of the ROC graph. It ranges from 0.5 for random models to 1.0 for perfect models. The importance of each EGV for explaining the species distribution was determined by: (1) Jack-knife analysis of the average AUC data; and (2) average percentage contribution of each EGV to the models. For this, MAXENT excludes a EGV in turn and a model was created with the remaining variables; then a model was created using each individual variable. Traditionally, a threshold is estimated to reclassify the occurrence map probability as presence/absence areas (Brito and Crespo, 2002), although this selection is rather subjective (Phillips et al., 2004, 2006; Phillips and Dudik, 2008). MAXENT offers three different types of thresholds for training data, such as: 'Maximum training sensitivity plus specificity logistic threshold' and 'Balance training omission, predicted area and threshold value logistic threshold'. Some authors considered that the maximum training sensitivity plus specificity threshold does not perform well (Jiménez-Valverde and Lobo, 2007). Therefore, the averages of the 10 raw models and the 'Balance training omission, predicted area and threshold value logistic threshold' were calculated (Araújo and New, 2007; Martínez-Freiría et al., 2008; Phillips and Dudik, 2008).

RESULTS

The suitable areas in Salamanca spread only through the south-east part, whereas the rest of the province appeared as unsuitable (Fig. 1). The AUC value was very high: 0.9912 ± 0.0008 (range: 0.9899-0.9925). The EGVs with the highest contribution to iterative models (Table 1) were ‘no. days of May included in the cold period’ (44.16%), ‘Duration of the hot period’ (19.34%) and ‘Variability of precipitation deficit lower than 50 mm in June’ (12.2%). Other variables with a lower contribution (Table 1) were ‘Winter mean precipitation’ (6.7%), ‘Mean temperature in the hot period’ (5.72%), ‘Autumn mean precipitation’ (3.9%) and ‘CORINE’ (3.68%). The remainder variables have a contribution lower than 1 (Table 1). The mean variable percentage contribution and AUC values of iterative models calculated without one variable were quite similar among variables, both with training and test data, i.e. variable absences did not affect strongly the models (Fig. 2). When the models were calculated only with one variable, the contribution varied strongly among EGVs, both with training and test data. ‘Summer mean precipitation’, ‘Mean temperature in the hot period’, ‘no. days of May included in the cold period’ and ‘Duration of the hot period’ were the most important variables (Fig. 2). Finally, the threshold for determining presence/absence areas was 0.0166 ± 0.0011. Therefore, the presence (suitable) area was 1040 km², meanwhile the absence (unsuitable) one corresponded to 11309 km².

DISCUSSION

Suitable areas for *H. meridionalis* were predicted in the south-western part of Salamanca, a Mediterranean high plateau between the two highest local Central System mountains (Hastiala, 1783 m and Canchal Negro, 2430 m). This area corresponds mainly to the Alagón river
Modelling *Hyla meridionalis* expansion

Therefore, the species is currently occupying almost all the suitable area in Salamanca: it can only spread further to the north and west. However, these two zones have a low suitability index. Thus, *H. meridionalis* may be in equilibrium in Salamanca, or nearly, with the environment. However, predictive models for the global distribution of *H. meridionalis* confirm that the species can expand along the Mediterranean basin and Central Europe (including the British islands), with other areas in Asia, Australia and South Africa (Sillero, unpubl. data). Most part of the Iberian Peninsula is suitable for *H. meridionalis*, including the province of Salamanca. These differences in the results were probably due to the different scale used in both studies. The global study was calculated with a grid of 10 km², meanwhile the present study used a more precise grid (1 km²). Kaliontzopoulou et al. (2008) proved that variables with higher spatial resolutions produce better predictive models. Therefore, if the global model uses a higher spatial resolution, might include a smaller potential area.

Another hypothesis is that the species is not in expansion, but has rather been in equilibrium for a long time. Hence, new records in Salamanca were observed recently as a consequence of previous lack of sampling (García, 2007; Sillero and Carretero, 2007). Martínez-Solano and Fernández-González (2003) rejected this hypothesis for Madrid. Nevertheless, the expansion may not be fast enough in order to be noticed in human time-life. It would be necessary to model past environmental conditions in order to identify original suitable areas. If the expansion is not recent and the current suitable areas were occupied many years ago, the current distribution should be very similar to the past one. If the expansion is recent, both suitable areas (past and present) should be different. In that hypothetical case, climatic conditions changed and allowed the species expansion. Recuero et al. (2007) cannot determine the date of the expansion process. The models built indicated a strong relationship with temperature and precipitation factors. Hence, climatic change might affect the distribution and biogeography of the species, as suggested by Merchán et al. (2004). Araújo et al. (2006) forecast a general expansion to the north due to global warning. However, future forecasts for the whole global species distribution (calculated with topographical and

---

**Fig. 2.** Jack-knife results of the contribution and AUC values of each variable from the 10 iterative MAX-ENT models for the training and testing records. TRWO: Training data without one variable; TRWI: Training data with only one variable; TEWO: Test data without one variable; TEWI: Testing data with only one variable; AUCWO: AUC without one variable; AUCWI: AUC with only one variable. See more details in the Materials and Methods section. See abbreviations on Table 1.
climatic variables) were very similar to current potential models (Sillero, unpubl. data). The temperature increment could not be enough for producing a expansion movement. In fact, models built for the Salamanca province confirmed that absence of ecologically favourable habitats prevented further expansion. Recuero et al. (2007) also suggested the presence of *H. arborea* as a main cause for no further expansion (Brito and Crespo, 2002). Nevertheless, *H. meridionalis* is syntopic with *H. arborea* in a large part of its range, and no competition problems have been described until now (García-Paris, 2004).

In summary, *H. meridionalis*, in Salamanca, is in equilibrium (or near) with the environment and only can spread in two small areas located north and west of the Salamanca populations. Thus, if the species is currently expanding, the process is almost complete. However, predictive models cannot confirm or reject the expansion process of the species. Further studies are needed to monitor the *H. meridionalis* populations in Salamanca and Madrid, in order to verify the hypothetical expansion process; and for study the influence and competition of *H. arborea* in the distribution of *H. meridionalis*.

ACKNOWLEDGEMENTS

Thanks to A. Kaliontzopoulou and F. Martínez-Freiría for improving an earlier version of the manuscript. I am supported by a post-doctoral grant (SFRH/BPD/26666/2006) from FCT. This work is dedicated to the memory of Daniel Omolo.

REFERENCES


