Short communication

Predicting soil micro-variables and the distribution of an endogeic earthworm species through a model based on large-scale variables

Daniel F. Marchán [a], [*], [1], Pablo Refoyo [a], [1], Marta Novo [b], Rosa Fernández [c], Dolores Trigo [a], Darío J. Díaz Cosín [a]

[a] Departamento de Zoología y Antropología Física, Facultad de Biología, Universidad Complutense de Madrid, C/José Antonio Nováis 2, 28040, Madrid, Spain
[b] Cardiff School of Biosciences, Cardiff University, BIOSI 1, Museum Avenue, Cardiff CF10 3AT, UK
[c] Museum of Comparative Zoology, Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA

ABSTRACT

Studies on spatial patterns of distributions of soil dwelling animals have usually relied on soil micro-variables or statistical analyses based on presence/absence data. Geographic Information Systems (GIS) allow easy access to large-scale variables to build species distribution models. In this study, we used MaxEnt to model the distribution of the endogeic earthworm Hormogaster elisae. Significant differences were found between the predicted suitability values of localities where the species was present and those where it was absent, validating the predictive model. Most of the large-scale training variables showed significant correlation with soil micro-variables known to influence the biology of the species, proving the ability of the model to predict (to an extent) soil variables from environmental ones. The methodology could be extended to other soil fauna.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Soil invertebrate fauna has traditionally been neglected by diversity and ecological research (Decaëns, 2010) due to the difficulty of sampling their populations (Smith et al. 2008). With the increasing recognition of their importance, more studies have started to explore the spatial patterns of their diversity and species distribution (Chefaoui et al., 2005; Guil et al. 2009; Crawford and Hoagland, 2010).

Concerning earthworms, both soil micro-variables and large-scale variables such as land use, lithology and climate have shown a significant effect on their distributions (Palm et al. 2013). However, most studies have relied exclusively on soil micro-variables at small scale (e.g. Joschko et al. 2006; Ouellet et al. 2008) despite requiring an intense effort to get regional coverages and being difficult to extrapolate. In contrast, large-scale variables are nowadays widely available through Geographic Information Systems (GIS), therefore facilitating the inference of species distribution models based on them (as in Shartell et al. 2013). It must be noted that the separation between both types of variables is not absolute, as some micro-variables show large-scale patterns and vice versa.

Other works used statistical analyses including absence data (Lindahl et al. 2009; Palm et al., 2013; Shartell et al. 2013), which are unreliable for soil fauna due to their difficult sampling. MaxEnt (Phillips et al. 2006) is an interesting statistical approach, as it is exclusively based on presence data and shows high predictive power when working with small datasets (Phillips and Dudík, 2008). It estimates the suitability of a location for the presence of a species as a function of the ecological similarity to its previously known habitat based on a set of environmental variables. MaxEnt has already been successfully implemented in soil dwelling arthropods (Crawford and Hoagland, 2010; Marek et al. 2012).

2. Model development

Hormogastrid earthworms (Oligochaeta, Hormogastridae) are especially difficult to sample due to their restricted distribution (Novo et al. 2012) and endogeic lifestyle. Obtaining distribution models that facilitate the selection of suitable sampling sites could help increasing knowledge of their currently poorly-known diversity and distribution. In this study, we chose the morphospecies Hormogaster elisae Álvarez, 1977 as a model system for the

---

* Corresponding author.
  E-mail address: danifermch@gmail.com (D.F. Marchán).
[1] These authors contributed equally.

inference of suitable habitats. It inhabits relatively acidic soils (Novo et al. 2012) with extreme conditions (low organic matter, cold winters and hot and dry summers) (Novo et al. 2009, 2010) and a high percentage of coarse sand (Hernandez et al. 2003, 2007). Novo et al. (2009, 2010) established its range in Central Iberia, finding 17 populations in the provinces of Segovia, Guadalajara and Madrid. These populations were the starting point for our MaxEnt distribution model (parameters in Suppl. Material 1). The large-scale variables potentially relevant for the biology of *H. elisae* were chosen as predictor variables, reflecting the influence of extreme temperature variation and water availability across the year (isothermality, precipitation of driest month/coldest quarter and mean diurnal temperature range), lithology, land use and anthropic activities (Suppl. Material 1).

Our main objectives were i) to validate the model as a sampling site selection tool by using new presence and absence data; and ii) to find correlations between the predictor variables and soil variables (obtained from presence points). Twenty five additional presence locations plus fifty one where *H. elisae* was absent (Suppl. Material 2) were selected for the statistical validation. The mean habitat suitability values obtained for presence/absence locations were compared by ANOVA analysis to test for significant differences. Soil factors were analyzed in presence locations (Suppl. Material 1, 2) and compared with large-scale variable/suitability values obtained for each location to study their correlations; ANOVA was performed with large-scale categorical variables as grouping variables.

The geographical representation of the predicted suitability values is shown in Fig. 1.

3. Model validation and correlations

The model obtained is potentially suitable for predicting the distribution of *H. elisae*. The average test AUC (a measure of predictive performance, see Suppl. Material 1) was 0.892, showing high predictive power. In addition, the validation based on the real distribution of the species confirmed its ability to discern suitable and unsuitable sampling points, as significant differences between the predicted suitability values of absence and presence locations were found (Suppl. Material 1). The mean value for locations where *H. elisae* was found was over 0.5, the threshold established by MaxEnt for random suitability.

Both Shartell et al. (2013) and Palm et al. (2013) needed to combine locally measured variables with their GIS-generated variables to reach AUC values 0.79/0.84−0.85. This highlights the usefulness of the present model, which achieved higher predictive power while relying only on large-scale variables.

Similar success was achieved by Crawford and Hoagland (2010) when modeling the distribution of the burying beetle *Nicrophorus americanus*, obtaining higher AUC values with MaxEnt compared to other approaches. In fact, the last four years have seen a proliferation of MaxEnt-based ecological niche modeling (ENM) works in invertebrates (e.g. Muhlfeld et al. 2011; Fuller et al. 2013; Staunton et al. 2014).

False negatives (frequent in endogeic earthworms due to their deep burrowing and patched distribution) could mean some absence points were actually suitable. However, this kind of error would only affect the validation of this model, while it would hinder distribution analysis based on presence/absence data. Some
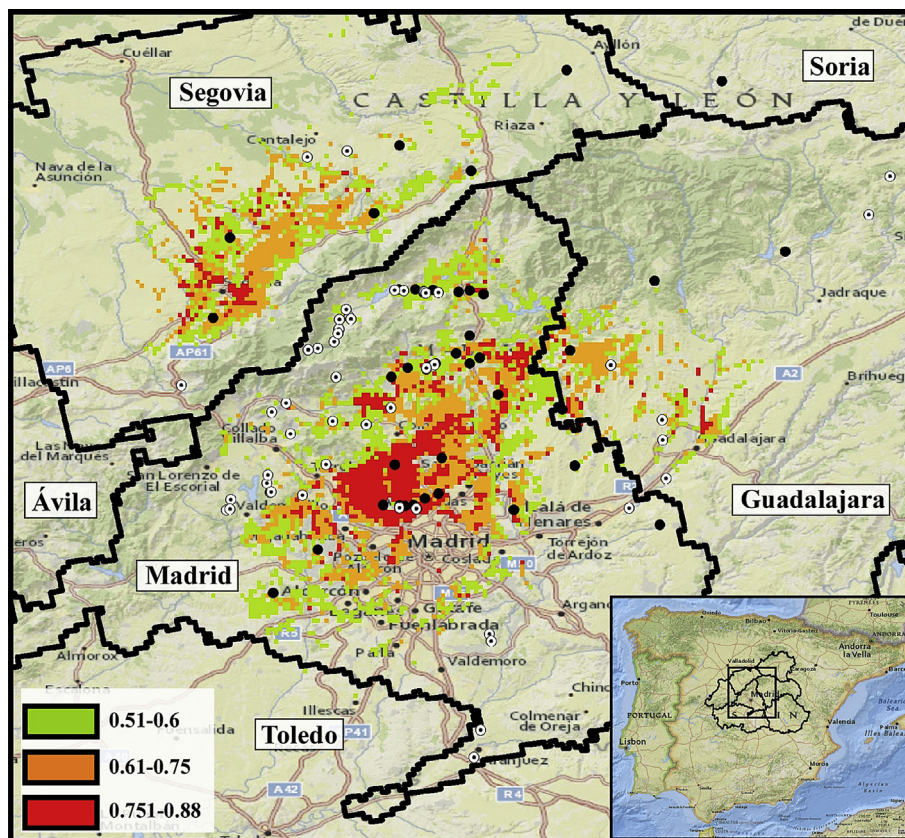

Fig. 1. Habitat suitability map including the distribution of presence (black dots) and absence (white dots) locations. Highest suitability values are shown in red, followed by orange and green. A map of the study area in the context of the Iberian Peninsula is shown for reference. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)