

# The unpredictable characteristics of the localities where new Iberian species will be discovered

Ana Payo-Payo<sup>1</sup> and Jorge M. Lobo<sup>2\*</sup>

<sup>1</sup>Population Ecology Group, Department of Biodiversity and Conservation, Instituto Mediterráneo de Estudios Avanzados (IMEDEA, CSIC-UIB), Miquel Marqués 21, 07190, Esporles, Islas Baleares, Spain,

<sup>2</sup>Department of Biogeography and Global Change, Museo Nacional de Ciencias Naturales (CSIC), C/Gutiérrez Abascal 2, 28006, Madrid, Spain

## ABSTRACT

**Aim** To describe patterns in new species descriptions, examine the spatial distribution and the associated environmental conditions of the localities with newly discovered species, as well as to assess whether protected areas encompass the majority of the occurrences of undescribed species.

**Location** Ibero-Balear region.

**Methods** After an exhaustive literature review, the temporal and spatial patterns of the terrestrial animal species described in the Ibero-Balear region from 1994 to 2012 were mapped. Both linear and asymptotic functions were used to compare the rates of discovered species using Akaike weights. Generalized Linear Models were used to estimate the explanatory capacity of a large set of environmental, land use and distance variables on the presence–absence of newly discovered species and on the variation in the number of new discoveries.

**Results** The average annual description rate was 112 species; 99% of the species descriptions were invertebrates. Adjusted discovery curves suggest that we still have an incomplete animal inventory. Environmental variables possess a limited capacity to predict the characteristics of the localities where new species are discovered ( $\approx 12\%$ ).

**Main conclusions** Our results indicate that it is not possible to provide an advanced estimation of the environmental and land use characteristics of the Ibero-Balear regions in which new species will be discovered in the future. Current protected areas are not guaranteed to conserve these undiscovered species, and non-protected areas should also be considered important reservoirs of unknown biodiversity.

## Keywords

Description rates, Iberian Peninsula, Linnean shortfall, protected areas, undescribed species.

\*Correspondence: Jorge M. Lobo, Department of Biogeography and Global Change, Museo Nacional de Ciencias Naturales (CSIC), C/Gutiérrez Abascal, 2 28006, Madrid, Spain. E-mail: mcnj117@mncn.csic.es

## INTRODUCTION

Current habitat degradation is expected to cause severe biodiversity loss (Brooks *et al.*, 2002), leading to large-scale social and ecological consequences (Díaz *et al.*, 2006). It is widely accepted that this degradation is a global problem that requires a global response (Wood *et al.*, 2000). The existing protected area network provides an invaluable service in shielding habitat from destructive use and hence in reducing biodiversity loss (Rodrigues *et al.*, 2004). However, habitat degradation continues to expand along with human population growth. As a consequence, some now common

species will soon lose their habitat, while many others, which have not yet been described, will go extinct before man has ever known them (Wilson & Costello, 2005; Costello *et al.*, 2013). The scientific community is working to monitor these changes and to provide useful information to minimise their impact (Chapin *et al.*, 2000). These efforts include the compilation of available taxonomic and distributional information on species in global databases, such as the Global Biodiversity Information Facility (GBIF, [www.gbif.org](http://www.gbif.org)), and the estimation of the capacity of current protected areas to represent these species (Margules & Pressey, 2000). However, despite great improvements in biodiversity knowledge

(Chapman, 2009; Fontaine *et al.*, 2012), our information about biodiversity remains biased (Whittaker *et al.*, 2005). Many species are still not formally described (Linnean shortfall), the distribution of those already described is poorly understood (Wallacean shortfall), and in many cases, currently known distributions reflect opportunity circumstances rather than real patterns (Hortal & Lobo, 2006; Hortal *et al.*, 2008). Thus, we need to ascertain how far we are from complete inventories, how many species remain unknown, which taxa they are likely to belong to and where we should search if we want to find them. These issues are particularly worrisome in the case of invertebrates, whose inventories are in the early stages of development and for which failing to account for these difficulties might result in severe erosion of biodiversity.

The rates of species descriptions and the number of authors involved in description exercises have been used to estimate the proportion of species that remain to be discovered (Joppa *et al.*, 2010; Costello *et al.*, 2012, 2013; Fontaine *et al.*, 2012; Essl *et al.*, 2013), although the obtained extrapolations are not exempt from criticism (Löbl & Leschen, 2014). Beyond the evident gaps in biodiversity knowledge, there are also broad gaps in knowledge of the extent to which biodiversity is represented (Rodrigues *et al.*, 2004) and protected within existing reserve networks (Gaston *et al.*, 2008). Protected areas may be key for local, regional, and global strategies of biodiversity conservation, but are these areas also essential to safeguard as yet undiscovered species? In this study, we analyse data on the species description rates of the Ibero-Balear region as a case study to estimate the magnitude of undescribed terrestrial animal species present in this territory. Many studies have tried to predict the total number of species based on historical discovery rates. However, this study is the first attempt to describe the spatial and environmental predictors of the locations at which new species were discovered. Are these new species collected under some particular environmental or spatial conditions? Follow an environmentally predictable pattern these new discovered species? Using this information, we also evaluated the current reserve capacity for protecting unknown terrestrial biodiversity. Are newly discovered species mainly collected within or near natural protected areas? To what extent can extant reserves represent undiscovered species? Assuming a future scenario of continuous habitat alteration and fragmentation in which non-protected 'natural' areas may tend to progressively disappear, we assess the potential for protected areas to serve as biodiversity refuges for as yet undescribed species.

## METHODS

### Data

The *locus typicus* of newly described animal species in the terrestrial environment of the Ibero-Balear region were gathered from the literature (see Appendix S1 in Supporting Information) and from a revision of the Zoological Record,

the world's oldest continuous database on animal biology provided by Thomson Reuters. Formerly well-known taxa that have been elevated to a species level were discarded. We assumed that the spatial and environmental conditions of the *locus typicus* reflect the average regional environmental conditions that each newly discovered species inhabits. Because the probability of finding a new species is proportional to its abundance, this assumption was justified by the general positive relationship between probability of occurrence, abundance and optimal niche conditions (Sagarin & Gaines, 2002). In total, information on 2123 newly described species identified in the Ibero-Balear region over a 19-year time span (1994–2012) was compiled in a database which can be available from the authors upon request. This relatively short period of time was selected for two main reasons: first, we can assume that no major changes in taxonomic effort occurred during those two decades. As information on the temporal variation in the number of taxonomists is lacking, selecting a wider temporal sequence of data would increase the probability that changes occurred in this unknown taxonomic workforce. Second, and most importantly, the probability of changes in the landscape is higher over increased time periods, thus undermining the search for relationships between new descriptions and land-use practices or the presence of protected areas. If a species was described a long time ago in a recently transformed urban area, our analyses could erroneously infer that urban areas harbour new species to be discovered. All the localities where specimens were found were georeferenced using the Geonames geographical database ([www.geonames.org](http://www.geonames.org)) and Google Maps (<https://maps.google.com>) when necessary.

### Description rates

Mean annual description rates ( $\pm$  95% confidence interval) were calculated for each of the nine highest-ranked taxonomic categories, which each account for more than 2% of the total species descriptions during the considered time period. These categories were as follows: the phyla Mollusca and Nemata, the subphylum Crustacea, the subclasses Arachnida and Collembola, and the insect orders Coleoptera, Lepidoptera, Diptera and Hymenoptera. Altogether, these taxonomic clades represented more than 86% of the newly described species considered herein and approximately 71% of all the terrestrial animal species recognised in the world (Scheffers *et al.*, 2012).

The number of different author combinations for works describing species within each taxon during the study period was also calculated. These values cannot be considered as a surrogate of taxonomic effort because they do not reflect the total number of taxonomists working in the region under consideration; this number was assumed to be constant across the short period of time examined. However, the correlation between the number of different author combinations and the number of described species was examined to determine if the description of new species is a participatory

process (positive correlation) or a practice dominated by a low number of taxonomists.

To estimate how far we are from relatively complete inventories in the Ibero-Balear region, we did not follow the effective procedure proposed by Wilson & Costello (2005), which is based on non-homogeneous renewal processes, because this method assumes a logistic-type relationship. A logistic curve is the most common shape obtained when the analysed data come from a large time period, beginning at the time of Linnaeus (Costello & Wilson, 2011). However, in our case, the initial phase of the logistic curve, in which growth is slow, should be lacking. Instead, the accumulated number of described species was regressed against year using a linear model and two sigmoidal models with an upper asymptote (the Morgan–Mercer–Flodin (MMF) and Weibull models; see Flather, 1996 and Tjørve, 2003; for a description of these models). For these calculations, the software CurveExpert 1.4 ([www.curveexpert.net](http://www.curveexpert.net)) was used, which enables a large number of regression models to be compared simultaneously (e.g., linear, exponential, power, sigmoidal, growth, etc.) in order to find the one with the greatest explanatory capacity. MMF and Weibull models were selected because they had a higher explanatory capacity when the total number of newly discovered species was considered. A logistic fitting procedure was also calculated in order to examine the general explanatory capacity of this function. Thus, we assessed whether the increase in the number of described species in the Ibero-Balear region is far from reaching an asymptotic tendency or follows a declining trend. In order to enable detection of even small asymptotic trends, the natural logarithm of the accumulated number of new species descriptions was used in the sigmoidal and logistic models. Model selection was based on the Akaike information criterion corrected for finite sample sizes (AICc; see Burnham & Anderson, 2002). The best model was determined by examining the normalised relative likelihood values given the data (Akaike weights;  $w_i$ ) following the recommendations of Burnham & Anderson (2002) to compare a model of the data against models based on transformed data.

### Environmental variables

Eighteen environmental variables (two topographic, nine climatic, three land uses and four geological) were calculated for each 100 km<sup>2</sup> UTM cell of the Ibero-Balear area. Environmental data from digital cartography at a resolution of 1 km<sup>2</sup> were averaged to generate the values for the 100 km<sup>2</sup> cells. This resolution (100 km<sup>2</sup>) was chosen because it is the standard scale used in Iberian invertebrate atlases (e.g., see Millán *et al.*, 2014) and is thus a compromise between the resolution of the biological data and that of the environmental information used to characterize the localities in which other, nearby populations might exist.

The mean altitude of each UTM cell, as well as the range of elevations, was extracted from a Digital Elevation Model (Clark-Labs, 2000). The climatic data for each cell were

obtained from WORLDCLIM (version 1.3, [www.worldclim.org](http://www.worldclim.org); see Hijmans *et al.*, 2005) using average monthly precipitation and temperatures values from 1950 to 2000. Aridity was calculated following the formula provided by Valencia-Barrera *et al.* (2002). Geological variables were the percentage of land with basic rocks, acidic rocks, basic deposits and acidic deposits in each UTM cell; these data were obtained by reclassifying the lithological map provided by the Instituto Geológico y Minero de España ([www.info.igme.es](http://www.info.igme.es)). These variables were used to examine if newly described species had been collected under specific environmental conditions.

Land use variables were extracted to assess if the cells harbouring newly described species were those experiencing lower human impact. To perform this extraction, we reclassified the 44 categories (label 3) established in the land use map provided by the Corine Land Cover 2006 project (see [www.eea.europa.eu](http://www.eea.europa.eu)), to obtain two main types of land uses: anthropogenic and natural (see Hernández-Manrique *et al.*, 2013 for more details). Next, we calculated the percentage of the area in the UTM cell covered by each of these two basic land use types. Finally, we estimated the percentage of the area in each UTM cell currently covered by protected areas to determine if the occurrence of cells with newly described species was associated with the proportion of protected area. Digital maps of protected areas were obtained from the Spanish Banco de Datos de la Naturaleza ([www.mma.es](http://www.mma.es)) and from the Portuguese Instituto de Conservação da Natureza (<http://portal.icn.pt>).

Three distance variables were also considered: the distance from the centre of each UTM cell to the border of any protected areas, the distance to areas of anthropogenic use, and the distance to urban areas catalogued as such for the Corine Land Cover project. These variables were included to estimate whether the proximity to protected areas or the distance to human altered conditions (proxies of naturalness and accessibility) are key factors explaining the likelihood of discovering new species.

### Statistical analyses

From total Iberian UTM cells ( $n = 6150$ ) the majority (81.5%;  $n = 5010$ ) do not have observations of newly discovered species. Generalized Linear Models (GLMs) were used to examine the relationship between environmental variables and either the presence–absence of newly discovered species or the number of new discoveries. Separate models were run because of the excess of unsampled areas within the study region and the possibility that the explanatory capacity of the selected environmental variables accounting for whether a new species was discovered in a UTM cell may differ from those accounting for variation in the number of newly discovered species (Heilbron, 1994). For the first case, we use a binomial error distribution linked to the set of predictor variables via a logit link function, while in the second case, a Poisson distribution was assumed and a logarithmic link function used. Our aim with these models was simply to

estimate the individual explanatory capacity of the selected variables in order to assess if the process of discovering new species follows an environmentally predictable pattern. We also built saturated models, which included all the considered variables, to estimate the maximum explanatory capacity. All calculations were performed using STATISTICA, version 12 (StatSoft, 2013). All explanatory variables were standardized to a mean of zero and standard deviation of one to eliminate the effect of measurement scale differences.

## RESULTS

### Species description rates

The mean annual discovery rate within the Ibero-Balear region was  $111.74 \pm 10.50$  species; invertebrates accounted for 99% of these new species descriptions. The description rates of each group were highly correlated with the number of species described worldwide (Spearman correlation coefficient,  $r_s = 0.73$ ;  $P < 0.03$ ) or within the Ibero-Balear region ( $r_s = 0.74$ ;  $P = 0.02$ ), such that more species were described from the most diverse groups. The number of newly described species in each group was also positively correlated with the number of author combinations ( $r_s = 0.92$ ;  $P < 0.001$ ), suggesting that the process of describing new species is not dominated by a small number of taxonomists.

Model selection (Table 1) revealed that a linear model was 3.8 times more likely than a Weibull model to represent the yearly increase in the cumulative number of newly discovered species ( $w_{iLIN}/w_{iWEI}$ ; 0.69/0.18) and 5.3 times more likely than the MMF model ( $w_{iLIN}/w_{iMMF}$ ; 0.69/0.13); a logistic relationship was highly unlikely. A linear increase in the cumulative number of newly discovered species was also the most probable relationship for Coleoptera, Arachnida, Diptera and Crustacea. In contrast, MMF models seemed to be the most likely for Hymenoptera, Collembola and Nemata, a Weibull model was most likely for Lepidoptera, and the logistic model was most probable in the case of Mollusca (Table 1, Fig. 1).

### Environmental correlates

A total of 1140 UTM Ibero-Balear grid cells (18.5% of total) contained localities in which a new species was described (Fig. 2). The mean number of described species in these UTM cells was  $1.84 \pm 0.14$ , although in three of the cells, more than 15 new species had been described. Two of the cells (with 99 and 17 described species, respectively) were located in the Los Monegros region, an unprotected steppe area with xerothermic conditions, which is noteworthy for having been subjected to a relatively intense survey effort (Ribera & Blasco-Zumeta, 1998; Blasco-Zumeta & Melic, 1999). The remaining UTM cell (with 25 described species) corresponds to a mid-mountain locality (El Ventorrillo) of the Sierra de Guadarrama, near Madrid City, which was also subjected to a relatively intense survey effort because it

is the field station of the National Museum of Natural Sciences.

Of the 21 considered variables, 17 were statistically significant when accounting for the presence–absence of newly described species in the UTM cells; 15 remained significant after applying a Bonferroni correction (Table 2). However, of these variables, only three explained more than 3% of the total deviance: altitude range (5.3%), distance to protected areas (4.9%) and protected area (3.2%). Considering curvilinear relationships by including quadratic terms of the considered variables did not greatly improve the explanatory capacity of these variables. The signs of the obtained parameters indicated that the cells where new species had been described had a higher topographic heterogeneity and a higher percentage of protected area and were usually located near reserves. When a saturated model including all the considered variables was carried out, only 12.1% of the total deviance could be explained.

Ten explanatory variables were statistically significant for explaining the frequency of species descriptions within the cells with positive counts, although only four variables remained significant after a Bonferroni correction (Table 2). However, only one variable accounted for more than 2% of the total deviance (protected area, 5.1%), such that a greater number of new species were described when the percentage of the protected area within the cell was higher. The explanatory capacity of the saturated model was also low for this case (12.8% of total deviance).

## DISCUSSION

A total of 112 terrestrial animal species were described per year in the Ibero-Balear region, which represent 17% of the species — including freshwater species— described yearly in Europe (Fontaine *et al.*, 2012) and approximately 1% of the total terrestrial species described yearly worldwide (IISE, 2012; Costello *et al.*, 2013). Our results show that the creation of the Ibero-Balear animal biodiversity inventory is a process that is relatively unbiased from the taxonomic point of view (more species are described in the *a priori* most diverse groups) and is participatory (with more author combinations describing species in the most diverse groups) and that taxonomic inventories are far from being considered completed. Because we examined species described over a short time period of almost two decades, in which the taxonomic workforce should not experience large variations, our results suggest that a yearly linear increase in the number of described species is the most plausible explanation. Of course, although we used the available information for all terrestrial animal taxa, the detected patterns inevitably reflect the description dynamics of the most diversified invertebrate groups. The description curves of some of these groups seem to be approaching a plateau. These asymptotic relationships can be the consequence of (1) a higher degree of difficulty for discovering new species assuming a similar sampling effort across the considered years, (2) an undetected decline

**Table 1** Number of newly described species (S) for each of the taxonomical groups accounting for more than 2% of total terrestrial animal species described in the Ibero-Balear region from 1994 to 2012; number of different author combinations describing species (Ac); annual rate (Ar) of species descriptions ( $\pm$  95% confidence interval); approximate number of animal species described worldwide (Ws) and in the Ibero-Balear region (IBs), according to Chapman (2009), Ramos *et al.* (2001) and Ortuño & Martínez Pérez (2011); explanatory capacity ( $R^2 \times 100$ ) of linear (LIN), logistic (LOG), Morgan–Mercer–Flodin (MMF) and Weibull (WEI) models on the cumulative number of newly discovered species; Akaike values corrected for finite sample sizes (AIC<sub>c</sub>); difference between the model with the lowest AIC ( $\Delta_i$ ); and Akaike weights ( $w_i$ ) for each one of the four considered models.

|                     | Total              | Coleoptera       | Hymenoptera      | Arachnida        | Diptera          | Lepidoptera     | Collembola      | Mollusca        | Nemata          | Crustacea       |
|---------------------|--------------------|------------------|------------------|------------------|------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| S                   | 2123               | 627              | 254              | 236              | 208              | 130             | 108             | 107             | 90              | 64              |
| Ac                  | 873                | 239              | 94               | 98               | 68               | 64              | 32              | 43              | 48              | 23              |
| Ar                  | 111.74 $\pm$ 10.50 | 33.00 $\pm$ 5.07 | 13.37 $\pm$ 2.96 | 12.42 $\pm$ 1.69 | 10.95 $\pm$ 3.41 | 6.84 $\pm$ 1.38 | 5.68 $\pm$ 2.49 | 5.63 $\pm$ 2.82 | 4.74 $\pm$ 1.51 | 3.37 $\pm$ 1.54 |
| Ws                  | 1,700,000          | 400,000          | 115,000          | 102,000          | 153,000          | 174,000         | 7500            | 85,000          | 25,000          | 47,000          |
| IBs                 | 58,000             | 10,000           | 7500             | 1600             | 6400             | 4200            | 600             | 850             | 1000            | 1600            |
| $R^2_{LIN}$         | 99.82              | 98.82            | 95.80            | 98.64            | 97.06            | 98.2            | 96.82           | 92.66           | 92.81           | 94.00           |
| AIC <sub>cLIN</sub> | <b>128.80</b>      | <b>121.62</b>    | 110.25           | <b>99.65</b>     | <b>91.77</b>     | 69.35           | 62.50           | 91.25           | 69.29           | <b>64.22</b>    |
| $\Delta_{LIN}$      | 0.00               | 0.00             | 49.68            | 0.00             | 0.00             | 16.00           | 5.98            | 16.54           | 14.70           | 0.00            |
| $w_{LIN}$           | 0.69               | 0.99             | 0.00             | 1.00             | 0.89             | 0.00            | 0.05            | 0.00            | 0.00            | 0.93            |
| $R^2_{LOG}$         | 99.11              | 97.86            | 99.55            | 98.21            | 97.03            | 98.66           | 98.14           | 97.19           | 95.54           | 89.43           |
| AIC <sub>cLOG</sub> | 166.62             | 131.51           | 85.00            | 159.18           | 109.72           | 74.40           | 70.19           | <b>74.72</b>    | 73.17           | 101.34          |
| $\Delta_{LOG}$      | 37.82              | 9.89             | 24.43            | 59.53            | 17.95            | 21.05           | 13.67           | 0.00            | 18.57           | 37.13           |
| $w_{LOG}$           | 0.00               | 0.01             | 0.00             | 0.00             | 0.00             | 0.00            | 0.00            | 0.99            | 0.00            | 0.00            |
| $R^2_{MMF}$         | 99.91              | 99.57            | 99.81            | 99.27            | 98.69            | 99.37           | 98.28           | 97.16           | 99.18           | 98.46           |
| AIC <sub>cMMF</sub> | 132.11             | 155.71           | <b>60.57</b>     | 160.95           | 97.47            | 55.11           | <b>56.52</b>    | 87.34           | <b>54.60</b>    | 80.95           |
| $\Delta_{MMF}$      | 3.30               | 34.09            | 0.00             | 61.31            | 5.70             | 1.76            | 0.00            | 12.62           | 0.00            | 16.74           |
| $w_{MMF}$           | 0.13               | 0.00             | 0.93             | 0.00             | 0.05             | 0.29            | 0.95            | 0.00            | 0.96            | 0.00            |
| $R^2_{WEI}$         | 99.91              | 99.57            | 99.78            | 99.27            | 98.69            | 99.40           | 98.41           | 97.03           | 99.14           | 98.47           |
| AIC <sub>cWEI</sub> | 131.44             | 165.12           | 65.77            | 160.97           | 97.22            | <b>53.35</b>    | 76.11           | 84.72           | 60.93           | 69.54           |
| $\Delta_{WEI}$      | 2.64               | 43.50            | 5.20             | 61.33            | 5.45             | 0.00            | 19.58           | 10.00           | 6.34            | 5.33            |
| $w_{WEI}$           | 0.18               | 0.00             | 0.07             | 0.00             | 0.06             | 0.71            | 0.00            | 0.01            | 0.04            | 0.07            |

lower AIC<sub>c</sub> values in bold

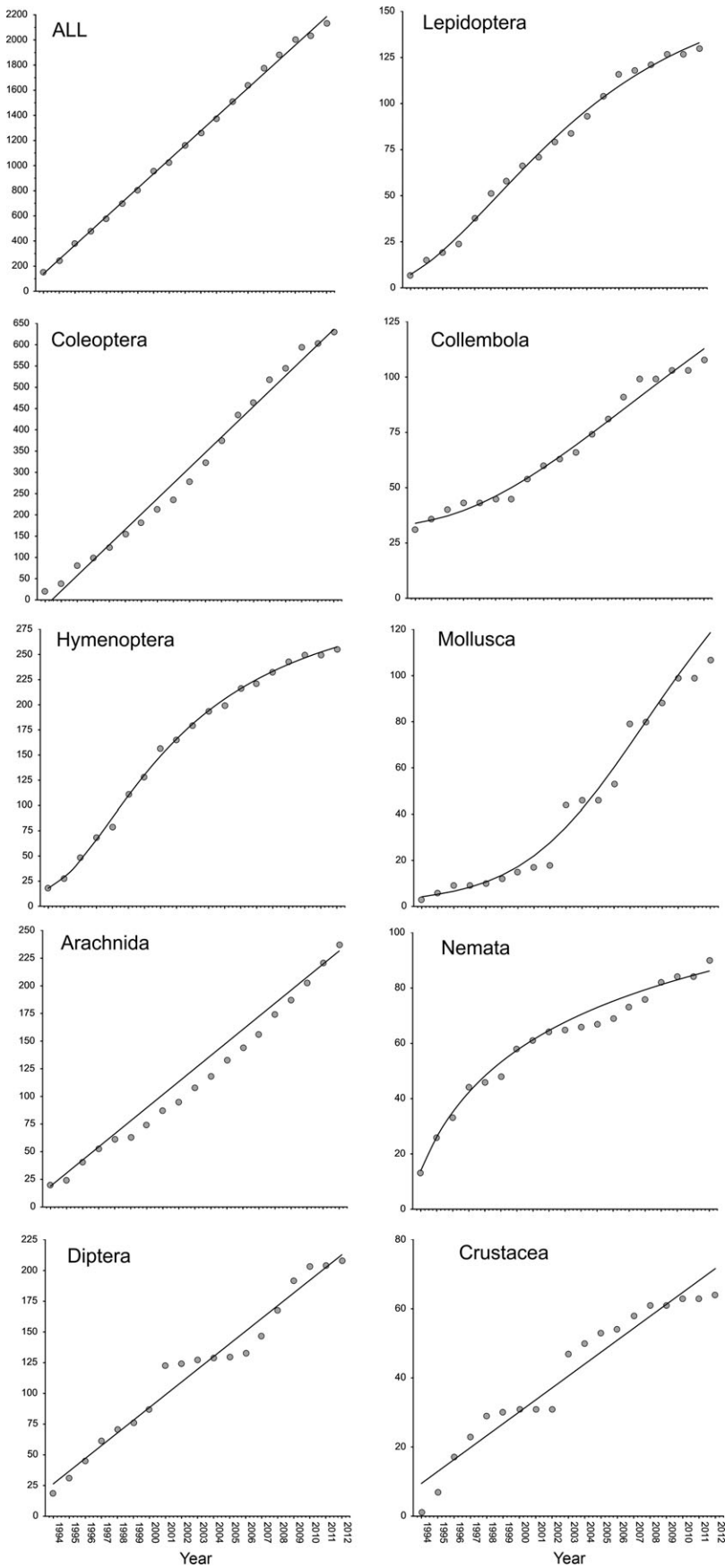
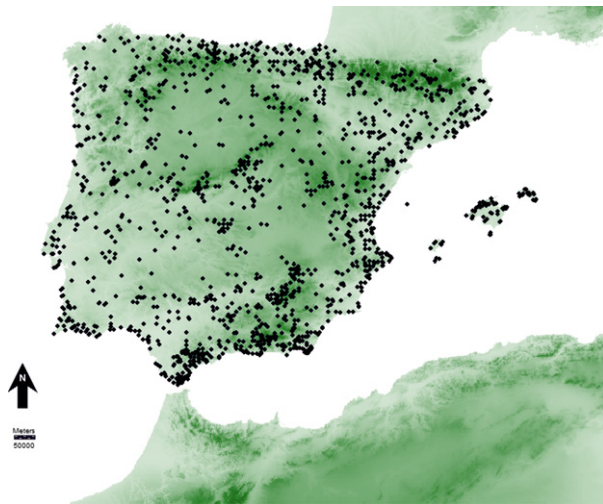


Figure 1 Increase in the accumulated number of newly described species in the Ibero-Balear region between 1994 and 2012 for all of the considered animal species and for the nine highest-ranked taxonomic categories that accounted for more than 2% of the total species descriptions during the considered time period.

in the supposed constant taxonomic workforce during recent years and/or (3) a simple consequence of random variation in the number of described species that cannot be ruled out



**Figure 2** Terrestrial 100 km<sup>2</sup> UTM cells within the Ibero-Balear region with at least one new animal species described between 1994 and 2012 (black dots). The background map shows altitude from lowland (light green) to mountain (dark green).

due to the short period of time considered. Thus, both the limited temporal scope of the data and the lack of a reliable measure to represent variation in taxonomic effort are drawbacks that limit the universal character of our predictions. Taking into account the available estimations on the Iberian species richness of animals (Lobo, 2015) as well as the perceived reduction in taxonomic resources (Lobo, 2000; Martín-Piera & Lobo, 2000; Martín Albaladejo, 2003), we are inclined to think that the observed asymptotic tendencies, when detected, underestimate the number of species remaining to be discovered.

In our opinion, the most interesting result of this study was the inability to identify relevant environmental and land use predictors able to explain the characteristics of the localities in which species have been found. Neither the presence/absence of newly discovered species nor the variation in their numbers within cells with positive values was associated with broad scale explanatory variables. Although there were some statistically significant environmental variables capable of discriminating those localities that were home to recently described species, the best predictive capacity that we calculated accounted for less than 13% of total variability. Thus, our results indicate that the localities in which most of future species descriptions will occur cannot be easily predicted or associated with a clear environmental pattern

**Table 2** Individual relationships between each of the explanatory variables and the presence–absence of newly discovered species (binomial models) or the number of new discoveries (Poisson models) using Generalized Linear Models. Wald statistics (Wald) were used to test the statistical significance of the effects, while Param are standardised parameter estimates. The change in deviance of each explanatory variable from a full model, in which any one predictor is entered in the model, was calculated to estimate the percentage of explained deviance (%).

|                                 | Binomial            |         |          |      | Poisson            |         |          |      |
|---------------------------------|---------------------|---------|----------|------|--------------------|---------|----------|------|
|                                 | Wald                | Param   | Deviance | %    | Wald               | Param   | Deviance | %    |
| Altitude                        | 8.56*               | 0.0941  | 5888.6   | 0.14 | 6.27*              | 0.0444  | 1654.7   | 0.37 |
| Altitude range                  | <b>310.31***</b>    | 1.7030  | 5582.6   | 5.33 | <b>31.00***</b>    | 0.0908  | 1631.1   | 1.79 |
| Aridity                         | <b>11.12**</b>      | −0.1136 | 5885.6   | 0.19 | <b>12.89**</b>     | 0.0614  | 1648.5   | 0.74 |
| Winter precipitation            | <b>21.15***</b>     | 0.1434  | 5876.5   | 0.35 | 0.65 <sup>NS</sup> |         |          |      |
| Autumn precipitation            | <b>72.29***</b>     | 0.2557  | 5828.0   | 1.17 | 2.62 <sup>NS</sup> |         |          |      |
| Spring precipitation            | <b>48.55***</b>     | 0.2104  | 5850.7   | 0.79 | 2.22 <sup>NS</sup> |         |          |      |
| Summer precipitation            | <b>25.66***</b>     | 0.1541  | 5872.5   | 0.42 | 8.81*              | −0.0557 | 1651.8   | 0.55 |
| Winter mean temperature         | 7.06*               | 0.0877  | 5889.9   | 0.12 | 0.13 <sup>NS</sup> |         |          |      |
| Autumn mean temperature         | 0.001 <sup>NS</sup> |         |          |      | 0.08 <sup>NS</sup> |         |          |      |
| Spring mean temperature         | 2.31 <sup>NS</sup>  |         |          |      | 0.07 <sup>NS</sup> |         |          |      |
| Summer mean temperature         | <b>14.60**</b>      | −0.1238 | 5882.5   | 0.25 | 4.24 <sup>NS</sup> |         |          |      |
| Anthropogenic land use          | <b>92.70***</b>     | −0.3311 | 5800.5   | 1.64 | 1.88 <sup>NS</sup> |         |          |      |
| Natural land use                | <b>53.14***</b>     | 0.2428  | 5843.1   | 0.91 | 4.17 <sup>NS</sup> | −0.0473 | 1656.7   | 0.25 |
| Protected area                  | <b>197.92***</b>    | 0.3794  | 5711.2   | 3.15 | <b>93.19***</b>    | 0.1309  | 1576.2   | 5.10 |
| Acidic rocks                    | <b>18.01***</b>     | −0.1451 | 5878.5   | 0.31 | 1.75 <sup>NS</sup> |         |          |      |
| Basic rocks                     | <b>55.17***</b>     | 0.2167  | 5845.4   | 0.87 | 0.00 <sup>NS</sup> |         |          |      |
| Acidic sediments                | 0.00 <sup>NS</sup>  |         |          |      | 1.63 <sup>NS</sup> |         |          |      |
| Basic sediments                 | <b>25.14***</b>     | −0.1723 | 5871.0   | 0.44 | 7.29*              | 0.0625  | 1653.7   | 0.43 |
| Distance to protected areas     | <b>226.53***</b>    | −0.6765 | 5610.7   | 4.86 | 7.17*              | −0.0719 | 1653.4   | 0.45 |
| Distance to urban areas         | <b>19.14***</b>     | −0.1479 | 5877.5   | 0.33 | 6.83*              | −0.0546 | 1654.0   | 0.42 |
| Distance to anthropogenic areas | 0.26 <sup>NS</sup>  |         |          |      | <b>12.67**</b>     | −0.1029 | 1647.3   | 0.82 |

Values that were significant at a Bonferroni-corrected significance level (0.05/21 = 0.002) are shown in bold.

\**P* ≤ 0.01, \*\**P* ≤ 0.001, \*\*\**P* ≤ 0.0001. NS = not significant statistically

(e.g., more species are not discovered in mountain than in valley areas). Both ecological and collector reasons could be added if an environmental determinism were to appear on the occurrence of newly discovered species; the higher species richness in some environments could favour the collection of new species, and/or influencing the detectability, accessibility or collector interest. Discriminating between these causes would require knowledge about the temporal variation in survey effort. In any case, this lack of environmental signal is a consequence of the highly dispersed character of these localities across the Iberian territory. An uneven geographical or environmental distribution of the taxonomic and survey effort could be affecting our results, provided that this effort is higher in areas harbouring a low number of unnamed species. However, the large number of new species found under contrasting environmental conditions, when those localities were subjected to a relatively intense taxonomic effort (as in the cases of Los Monegros and El Ventorrillo), suggests that environmental conditions play a minor role in the discovery of new species and that it is enough to carry out a relatively intense taxonomic and survey effort to identify new species. The three most outstanding localities, which harbour almost 7% of the total species descriptions, seem to have reached this prominence by contingent and unpredictable circumstances. Could we obtain similar description rates in other localities if they were subjected to the same intensity of taxonomic work? We believe so.

The unpredictable and serendipitous characteristics of the localities in which new species could be discovered in the future are linked to the low relevance of land uses and protected areas for explaining the location of newly discovered species. Forty-three percent of the cells in which a new species had been discovered had some protected area, while 19% of cells without data on newly discovered species had some protected area. However, despite the fact that the amount of area protected and the distance to protected areas were among the best predictors, these variables only accounted for 3–5% of total variability. Although the limitations of the used land cover data (Díaz-Pacheco & Gutiérrez, 2014) and the probable role played by micro-environmental conditions in explaining the occurrence of invertebrates are uncertainty sources, our results show that the descriptions of new species did not seem to be exclusive to the cells with natural land uses; 34% of the cells with new descriptions had at least 50% of their area affected by anthropogenic activity (48% in the remaining cells). The special authorization required by amateur taxonomists to collect specimens in protected areas may also influence these results. However, similar legislation requirements are in place in Spain in natural, unprotected environments. In our opinion, the occurrence of newly described species outside protected areas and within human-altered landscapes would be the consequence of the incapacity of reserves to fully represent the terrestrial biodiversity in a Mediterranean region where biodiversity is deeply influenced by a long history of intense human settlement (Blondel & Aronson, 1995). Previous studies show that

current Iberian protected areas are clearly insufficient, not only for ensuring the representation of endangered invertebrate species (Hernández-Manrique *et al.*, 2012), terrestrial vertebrates and plants (Araújo *et al.*, 2007; López-López *et al.*, 2011), but also for guaranteeing adequate future climatic conditions for the long-term survival (Sánchez-Fernández *et al.*, 2013) or the maintenance of the phylogenetic diversity (Abellán *et al.*, 2013) of some groups. If reserves have not been selected to represent biodiversity, especially in the case of invertebrates, it is not surprising that a large proportion of newly discovered species are collected outside protected areas. This reasoning does not deny the conservation value of protected areas, but emphasizes the biodiversity value of the non-protected territory.

A recent meta-analysis (Coetzee *et al.*, 2014) determined that protected areas worldwide generate significant biodiversity benefits compared to surrounding areas, although their impact in mitigating land use impacts have most likely been exaggerated (Joppa & Pfaff, 2010). Without a doubt, natural reserves should be a cornerstone in any conservation strategy, but their effectiveness for conserving biodiversity and providing ecosystem services is highly dependent on their management, location, number and connectivity as well as on the sustainable management of the surrounding unprotected areas (Mora & Sale, 2011). Here, we show that current protected areas may not be enough to guarantee the conservation of those species that might be discovered in the future, and that adequate conservation management should be required in unprotected areas to diminish our uncertainty about real biodiversity patterns.

## ACKNOWLEDGEMENTS

We would like to thank Luis Castelo and Francisco Yagüe for their help. APP was partially supported by a pre-doctoral-FPU Fellowship (ref. FPU2012-000869) from the Spanish Ministry of Education. Partial financial support provided by the project CGL2011-25544.

## REFERENCES

- Abellán, P., Sánchez-Fernández, D., Picazo, F., Millán, A., Lobo, J.M. & Ribera, I. (2013) Preserving the evolutionary history of freshwater biota in Iberian National Parks. *Biological Conservation*, **162**, 116–126.
- Araújo, M.B., Lobo, J.M. & Moreno, J.C. (2007) The effectiveness of Iberian protected areas in conserving terrestrial biodiversity. *Conservation Biology*, **21**, 1423–1432.
- Blasco-Zumeta, J. & Melic, A. (1999) Introducción a Los Monegros. *Boletín de la Sociedad Entomológica Aragonesa*, **24**, 5–8.
- Blondel, J. & Aronson, J. (1995) Biodiversity and ecosystem function in the Mediterranean basin: human and nonhuman determinants. *Mediterranean-type ecosystems: the function of biodiversity* (ed. by G.W. Davis and D.M. Richardson), pp. 43–119. Springer Verlag, Berlin.



- Brooks, T.M., Mittermeier, R.A., Mittermeier, C.G., Da Fonseca, G.A.B., Rylands, A.B., Konstant, W.R., Flick, P., Pilgrim, J., Oldfield, S., Magin, G. & Hilton-Taylor, C. (2002) Habitat loss and extinction in the hotspots of biodiversity. *Conservation Biology*, **16**, 909–923.
- Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multimodel inference – a practical information-theoretic approach*. Springer Verlag, New York, USA.
- Chapin, F.S. III, Zavaleta, E.S., Eviner, V.T., Naylor, R.L., Vitousek, P.M., Reynolds, H.L., Hooper, D.U., Lavorel, S., Sala, O.E., Hobbie, S.E., Mack, M.C. & Diaz, S. (2000) Consequences of changing biodiversity. *Nature*, **405**, 234–242.
- Chapman, A.D. (2009) *Numbers of living species in Australia and the world*. Australian Biodiversity Information Services, Toowoomba, Australia. Available at: <http://www.environment.gov.au/node/13875> (accessed January 2015).
- Clark-Labs (2000) Global change data archive. 1 km global elevation model CD-ROM.
- Coetzee, B.W.T., Gaston, K.J. & Chown, S.L. (2014) Local scale comparisons of biodiversity as a test for global protected area ecological performance: a meta-analysis. *PLoS ONE*, **9**, e105824.
- Costello, M.J. & Wilson, S.P. (2011) Predicting the number of known and unknown species in European seas using rates of description. *Global Ecology and Biogeography*, **20**, 319–330.
- Costello, M.J., Wilson, S. & Houlding, B. (2012) Predicting total global species richness using rates of species description and estimates of taxonomic effort. *Systematic Biology*, **61**, 871–883.
- Costello, M.J., May, R.M. & Stork, N.E. (2013) Can we name Earth's species before they go extinct? *Science*, **339**, 413–416.
- Díaz, S., Fargione, J., Chapin, F.S. III & Tilman, D. (2006) Biodiversity loss threatens human well-being. *PLoS Biology*, **4**, e277.
- Díaz-Pacheco, J. & Gutiérrez, J. (2014) Exploring the limitations of CORINE Land Cover for monitoring urban land-use dynamics in metropolitan areas. *Journal of Land Use Science*, **9**, 243–259.
- Essl, F., Rabitsch, W., Dullinger, S., Moser, D. & Milasowszky, N. (2013) How well do we know species richness in a well-known continent? Temporal patterns of endemic and widespread species descriptions in the European fauna. *Global Ecology and Biogeography*, **22**, 29–39.
- Flather, C. (1996) Fitting species–accumulation functions and assessing regional land use impacts on avian diversity. *Journal of Biogeography*, **23**, 155–168.
- Fontaine, B., van Achteberg, K., Alonso-Zarazaga, M.A. et al. (2012) New species in the Old World: Europe as a frontier in biodiversity exploration, a test bed for 21st century taxonomy. *PLoS ONE*, **7**, e36881.
- Gaston, K.J., Jackson, S.F., Cantú-Salazar, L. & Cruz-Piñón, G. (2008) The ecological performance of protected areas. *Annual Review of Ecology, Evolution, and Systematics*, **39**, 93–113.
- Heilbron, D.C. (1994) Zero-altered and other regression models for count data with added zeros. *Biometrical Journal*, **36**, 531–547.
- Hernández-Manrique, O.L., Numa, C., Verdú, J.R., Galante, E. & Lobo, J.M. (2012) Current protected sites do not allow the representation of endangered invertebrates: the Spanish case. *Insect Conservation and Diversity*, **5**, 414–421.
- Hernández-Manrique, O.L., Sanchez-Fernandez, D., Numa, C., Galante, E., Verdú, J.R. & Lobo, J.M. (2013) Extinction trends of threatened invertebrates in peninsular Spain. *Journal of Insect Conservation*, **17**, 235–244.
- 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.
- Hortal, J. & Lobo, J.M. (2006) A synecological framework for systematic conservation planning. *Biodiversity Informatics*, **3**, 18–45.
- Hortal, J., Jiménez-Valverde, A., Gómez, J.F., Lobo, J.M. & Baselga, A. (2008) Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos*, **117**, 847–858.
- IISE (2012) *State of observed species*. Arizona state University. Available at: <http://www.esf.edu/species/SOS.htm> (accessed January 2015).
- Joppa, L.N. & Pfaff, A. (2010) Global protected area impacts. *Proceedings of the Royal Society B*, **278**, 1633–1638.
- Joppa, L.N., Roberts, D.L. & Pimm, S.L. (2010) How many species of flowering plants are there? *Proceedings of the Royal Society B*, **278**, 554–559.
- Löbl, I. & Leschen, R.A.B. (2014) Misinterpreting global species numbers: examples from Coleoptera. *Systematic Entomology*, **39**, 2–6.
- Lobo, J.M. (2000) La riqueza entomológica ibérica y el estado actual de los recursos taxonómicos: apuntes para un diagnóstico. *Boletín de la Sociedad Entomológica Aragonesa*, **27**, 173–176.
- Lobo, J.M. 2015. *Biodiversidad entomológica Ibérica*. Revista IDE@-SEA, 3, 1-8. Available at: <http://www.sea-entomologia.org/ide@/>(accessed April 2015).
- López-López, P., Maiorano, L., Faluccci, A., Barba, E. & Baitani, L. (2011) Hotspots of species richness, threat and endemism for terrestrial vertebrates in SW Europe. *Acta Oecologica*, **37**, 399–412.
- Margules, C.R. & Pressey, R.L. (2000) Systematic conservation planning. *Nature*, **405**, 243–253.
- Martín Albaladejo, C. (2003) *Tendencias de la taxonomía entomológica española*. Doctoral Thesis, Universidad Complutense de Madrid. Available at: <http://eprints.ucm.es/tesis/19972000/X/3/X3059701.pdf> (accessed April 2015).
- Martín-Piera, F. & Lobo, J.M. (2000) Diagnóstico sobre el conocimiento sistemático y biogeográfico de tres órdenes de insectos hiperdiversos en España: Coleoptera, Hymenoptera y Lepidoptera. *Hacia un Proyecto Cytad para el Inventario y Estimación de la Diversidad Entomológica en*

- Iberoamérica: PrIBES-2000* (ed. by F. Martín-Piera, J.J. Morrone and A. Melic), pp. 287–308. m3 m: Monografías Tercer Milenio, vol. 1, SEA, Zaragoza.
- Millán, A., Sánchez-Fernández, D., Abellán, P., Picazo, F., Carbonell, J.A., Lobo, J.M. & Ribera, I. (2014) *Atlas de los coleópteros acuáticos de España peninsular*, 820 pp. Ministerio de Agricultura, Alimentación y Medio Ambiente, Madrid.
- Mora, C. & Sale, P.F. (2011) Ongoing global biodiversity loss and the need to move beyond protected areas: a review of the technical and practical shortcomings of protected areas on land and sea. *Marine Ecology Progress Series*, **434**, 251–266.
- Ortuño, V.M. & Martínez Pérez, F.D. (2011) Diversidad de artrópodos en España. *Memorias de la Real Sociedad Española de Historia Natural*, **9**, 235–284.
- Ramos, M.A., Lobo, J.M. & Esteban, M. (2001) Ten years inventorying the Iberian fauna: results and perspectives. *Biodiversity and Conservation*, **10**, 19–28.
- Ribera, I. & Blasco-Zumeta, J. (1998) Biogeographical links between steppe insects in the Monegros region (Aragón, NE Spain), the eastern Mediterranean, and central Asia. *Journal of Biogeography*, **25**, 969–986.
- Rodrigues, A.S.L., Andelman, S.J., Bakarr, M.I. *et al.* (2004) Effectiveness of the global protected area network in representing species diversity. *Nature*, **428**, 640–643.
- Sagarin, R. & Gaines, S. (2002) The “abundant centre” distribution: to what extent is it a biogeographical rule? *Ecology Letters*, **5**, 137–147.
- Sánchez-Fernández, D., Abellán, P., Picazo, F., Millán, A., Ribera, I. & Lobo, J.M. (2013) Do protected areas represent species’ optimal climatic conditions? A test using Iberian water beetles. *Diversity and Distributions*, **19**, 1407–1417.
- Scheffers, B.R., Joppa, L.N., Pimm, S.L. & Laurance, W.F. (2012) What we know and don’t know about Earth’s missing biodiversity. *Trends in Ecology and Evolution*, **27**, 501–510.
- StatSoft, Inc. (2013) *STATISTICA (data analysis software system), version 12*. Available at: [www.statsoft.com](http://www.statsoft.com).
- Tjørve, E. (2003) Shapes and functions of species–area curves: a review of possible models. *Journal of Biogeography*, **30**, 827–835.
- Valencia-Barrera, R., Comtois, P. & Fernandez-Gonzalez, D. (2002) Bioclimatic indices as a tool in pollen forecasting. *International Journal of Biometeorology*, **46**, 171–175.
- Whittaker, R.J., Araújo, M.B., Paul, J., Ladle, R.J., Watson, J.E.M. & Willis, K.J. (2005) Conservation biogeography: assessment and prospect. *Diversity and Distributions*, **11**, 3–23.
- Wilson, S.P. & Costello, M.J. (2005) Predicting future discoveries of European marine species by using a non-homogeneous renewal process. *Journal of the Royal Statistical Society, Series C*, **54**, 897–918.
- Wood, A., Stedman-Edwards, P. & Mang, J. (2000) *The root causes of biodiversity loss*. Wold Wildlife Fund and Earth scan Publications, London.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

**Appendix S1.** Database literature review.

## BIOSKETCHES

**Ana Payo Payo** is a PhD student at University of Barcelona and she is currently working in the Ecology Population Group (GEP) at Spain. Her research focuses on biodiversity conservation, ecological modelling and demography.

**Jorge M. Lobo** is a researcher at the National Museum of Natural History (MNCN-CSIC, Madrid, Spain). His research focuses on biogeography, macroecology and insect conservation, working in several general aspects related with the detection of biodiversity patterns and the estimation of the main factors implicated in them.

Author contributions: Both authors contributed equally to the article. J.M.L. conceived the idea; A.P.P. collected the data; A.P.P. and J.M.L. analysed the data; and A.P.P. and J.M.L. led the writing.

---

Editor: Enrico Di Minin