How to describe species richness patterns for bryophyte conservation?

Helena Hespanhol¹, Katia Cezón², Ángel M. Felicísimo³, Jesús Muñoz²,⁵ & Rubén G. Mateo⁴

¹CIBIO/MinBio, Centro de Investigação em Biodiversidade e Recursos Genéticos da Universidade do Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal
²Real Jardín Botánico (CSIC), Plaza de Murillo 2, 28014 Madrid, Spain
³Centro Universitario de Mérida, Universidad de Extremadura, 06800 Mérida, Spain
⁴Department of Ecology & Evolution, University of Lausanne, Biophore Building, 1015, Lausanne, Switzerland
⁵Universidad Tecnológica Indoaméricana, Bolívar 20-35, Ambato, Ecuador

Keywords
Biodiversity conservation, biological collections, bryophytes, richness models, species distribution models.

Abstract
A large amount of data for inconspicuous taxa is stored in natural history collections; however, this information is often neglected for biodiversity patterns studies. Here, we evaluate the performance of direct interpolation of museum collections data, equivalent to the traditional approach used in bryophyte conservation planning, and stacked species distribution models (S-SDMs) to produce reliable reconstructions of species richness patterns, given that differences between these methods have been insufficiently evaluated for inconspicuous taxa. Our objective was to contrast if species distribution models produce better inferences of diversity richness than simply selecting areas with the higher species numbers. As model species, we selected Iberian species of the genus Grimmia (Bryophyta), and we used four well-collected areas to compare and validate the following models: 1) four Maxent richness models, each generated without the data from one of the four areas, and a reference model created using all of the data and 2) four richness models obtained through direct spatial interpolation, each generated without the data from one area, and a reference model created with all of the data. The correlations between the partial and reference Maxent models were higher in all cases (0.45 to 0.99), whereas the correlations between the spatial interpolation models were negative and weak (−0.3 to −0.06). Our results demonstrate for the first time that S-SDMs offer a useful tool for identifying detailed richness patterns for inconspicuous taxa such as bryophytes and improving incomplete distributions by assessing the potential richness of under-surveyed areas, filling major gaps in the available data. In addition, the proposed strategy would enhance the value of the vast number of specimens housed in biological collections.

Introduction
Cryptogams, invertebrates, and other inconspicuous groups are of great importance in the functioning of ecosystems (Hafernik 1992; Vanderpoorten and Goffinet 2009). But, they have not received as much attention as charismatic taxonomic groups, such as flowering plants, mammals, or birds, for the establishment of conservation measures or the study of biodiversity patterns (Oliver and Beattie 1993; Hunter and Webb 2002). Due to the increasing pressure on the natural environment, many individual species and sites that are important for biodiversity conservation now face increasing threats (Brooks et al. 2002; Butchart et al. 2010). The estimation of species richness distribution is necessary to understand spatial patterns of biodiversity (Ricklefs 2004), to establish conservation strategies (Bombi et al. 2011), or to predict future patterns of biodiversity under global change (Algar et al. 2009).

However, information on species richness is often incomplete, especially for inconspicuous organisms due to
several reasons, including insufficient field surveys for most geographic areas, strong spatial biases in the survey data (Sérgio and Draper 2002; Wilson et al. 2005), or a lack of taxonomic knowledge about such “forgotten” groups (the “taxonomic impediment”) (Carvalho et al. 2007). Also, this is in part because the inclusion of invertebrates and nonflowering plants is perceived as being too time-consuming, costly, and difficult because of the shortage of specialists (Oliver and Beattie 1993). Finally, records of vertebrates or vascular plants are recurrently used as surrogates for estimates of total biodiversity. But, for example, bryophytes are characterized by high dispersal capacities (Muñoz et al. 2004) and strikingly different ecophysiological strategies as compared to flowering plants (Vanderpoorten and Goffinet 2009). As a result, the existence of a spatial congruence in species richness distribution patterns between flowering plants and bryophytes has been questioned (Shaw et al. 2005; Hedenäs 2007; Geffert et al. 2013). So, the use of charismatic taxonomic groups as indicators of total biodiversity should be revised and new techniques should be applied to know the richness patterns of inconspicuous groups. To overcome the lack of planned data sampling, natural history collections (NHCs) represent a useful solution. These sources of data have some advantages, including quantity, accessibility, taxonomic confidence, frequent updating, and wide temporal and geographic scales (Graham et al. 2004; García-Lan and Ezcurra 2011) and therefore are useful for conservation purposes (Loiselle et al. 2003; Gaubert et al. 2006; Sérgio et al. 2007b; Newbold 2010).

The simplest modeling process that can be applied to a collection of observations is based on interpolation of known locations and expert knowledge (Boitani et al. 2011). A number of studies have directly used raw species occurrence data or distribution maps derived from simple forms of spatial interpolation that estimate unknown data from neighbor values (i.e., direct interpolation of data) to evaluate patterns of biodiversity for conservation assessment (Ferrier 2002; Graham and Hijmans 2006; Hernandez-Stefanoni and Ponce-Hernandez 2006; Sérgio et al. 2012; Geffert et al. 2013). Historically, atlas works, which aim at mapping species distributions at large spatial scales, have been based on the collection of observations and their representation on a continuous spatial grid through interpolation techniques (Franklin and Miller 2009). This intuitive and simple approach that consists on aggregating the NHCs data and transferring the resulting information to a geographic space may be useful for coarse-scaled conservation assessments, and it appears as reasonable if data are spatially well distributed and the density of locations high (Ferrier 2002).

As the recording effort cannot be increased very much in most cases (Vanderpoorten et al. 2005), an alternative is to use the available NHCs to generate maps of potential richness (Guisan and Rahbek 2011) which can be found in the bibliography (Kruijer et al. 2010; Sérgio et al. 2011; Désamoré et al. 2012; Roux et al. 2012; Mateo et al. 2013b) in comparison with other organisms (e.g., vascular plants, birds, and mammals), and none of these studies has been applied yet to investigate species richness patterns.

The conservation of bryophytes is behind that of flowering plants (Schumacker and Martiny 1995; Vanderpoorten et al. 2005), although they are subject to many of the threats that flowering plants face. There are a number of programs which aim is to identify and protect a network of the best sites for biodiversity conservation (e.g., http://www.cbd.int; http://www.natura.org), based on biodiversity richness. One of them is the Important Plant Areas (IPA) program (http://www.plantlife.org.uk), focused on the identification of priority areas for wild plants, fungi, and their habitats around the world and to ensure their long-term survival. It offers guidelines to identify and protect regions with high diversity in habitats and species based on consistent criteria (Anderson 2002). Connected with the IPA project, some others programs have been developed, as the Important Bryophyte Areas or “IbrA” (Papp 2008). These programs offer the possibility to protect and properly manage the priority conservation sites, but one of the basic requirements is that the design of such networks must be based on sound knowledge on species’ distributions and typically uses species distribution maps based on raw occurrence data. Overall, the traditional

© 2015 The Authors. Ecology and Evolution published by John Wiley & Sons Ltd.
approach on bryophyte conservation planning is to collect all the available data, generally based on NHC, and after an implicit spatial interpolation process, to propose the richest areas as priority areas for conservation (Infante and Heras 2012; Sérgio et al. 2012).

In this article, we compare the traditional method used in bryophyte conservation planning to estimate richness patterns that aggregates raw distribution data and transfers the resulting information to a geographic space, and the widely method of generating SDMs that is being used in conservation planning and the design of reserve networks, by integrating the relationships between data on species distributions available in NHCs and meaningful environmental variables to build the habitat suitability of the species. We evaluate these two alternative approaches with different complexity to estimate species richness patterns (direct interpolation of data vs. S-SDMs) using museum collections data, given that differences between these methods have been insufficiently evaluated for inconspicuous taxa. Here, we use a spatial interpolation model as an equivalent approach to the traditional method commonly used in bryophyte conservation planning which consists on aggregating the NHC data and transferring the resulting information to a geographic space. Specifically, we investigate how these approaches can influence richness patterns, how much are affected by the data, and how this variation influences inferences drawn from these richness maps. Here, we also assess whether accurate predictions of potential richness can be achieved for bryophytes using S-SDMs, as little is known about the performance of S-SDMs in inconspicuous groups. In bryophytes, climatic filters and long-distance dispersal have indeed been traditionally assumed to shape species composition at the continental scale (Muñoz et al. 2004), so that macroclimatic factors proved excellent predictors of bryophyte-dominated ecosystems and bryophyte species distributions. Therefore, they are a good subject of study for SDMs (Sérgio et al. 2011; Mateo et al. 2013b). To perform this study, we used the genus *Grimmia* as models species because the taxonomy of this genus is well known, and thus, we can avoid the “taxonomic impediment” mentioned above. Moreover, the group has been relatively well collected in the Iberian Peninsula. This combination of factors makes the group ideal for testing our hypothesis: When species richness patterns are derived from direct interpolation of distributional data on inconspicuous taxa housed in NHCs, these data by themselves are not sufficiently informative to derive richness patterns. To test our hypothesis, we sought answer to several questions. Do S-SDMs generated from NHCs produce the same outcome, in terms of richness, as direct interpolation models of the information contained in the NHCs for inconspicuous taxa, namely bryophytes? Do S-SDMs generated from NHCs produce reliable results for inconspicuous taxa?

### Materials and Methods

#### Study area and species data

The genus *Grimmia* is an important component of the Iberian Peninsula bryoflora with regard to both the number of species and their distribution and ecological significance. The group was revised for several geographic areas by one of the authors (JM) who has studied most of the Iberian specimens deposited in herbaria worldwide. The genus includes 31 species from the Iberian Peninsula (Muñoz and Pando 2000; Casas et al. 2006; Muñoz et al. 2009). Thirteen species of the genus are known from fewer than five Iberian localities and have been excluded from this study (*G. anomalala, G. arenaria, G. atrata, G. capillata, G. crinitoleucophaea, G. elatior, G. elongata, G. horrida, G. incurva, G. longirostris, G. mollis, G. muehlenbeckii, and G. unicolor*). The taxonomic status of *G. dissimulata* and *G. meridionalis* is currently under study using molecular methods, and these two species are not discussed further in this study.

Information is available about the conservation status of the species included in this study at different scales and in different areas. *G. caespiticia* (Brid.) Jur. and *G. atrata* Hornsch. are considered rare in Europe (Schmacker and Martiny 1995). At the scale of the Iberian Peninsula, *G. arenaria* Hampe, *G. crinitoleucophaea* Cardot, *G. mollis* Bruch & Schimp., *G. muehlenbeckii* Schimp., and *G. unicolor* Hook. are considered vulnerable. *Grimmia capillata* De Not. is data deficient-new, *G. incurva* Schwägr. is near threatened, and the remaining species are included in the list of species of least concern and attention (Sérgio et al. 2007a).

We mapped the collections on a 0.01° (~0.72 km² in the study area) grid, the spatial resolution used for the study. At this spatial resolution, certain specimens collected in close proximity coincide in the same pixel and represent a single presence. The collection localities were resampled at a pixel size of 10 × 10 km to avoid the overrepresentation of very low numbers in the interpolation analyses (Fig. 1A). We employed two different data sets, the white dots in Figure 1A represent opportunistic sampling localities, whereas the black dots indicate systematically sampled localities, corresponding to the following data sets: (1) Creu Casas’ extensive and intensive collections across the Pyrenees and Francisco Lloret’s PhD dissertation; (2) the PhD thesis data of Helena Hespanhol (NW Portugal); (3) the PhD thesis data of Katia Cezón (Castilla-La Mancha); and (4) the PhD thesis data of Susana Rams (Sierra Nevada).
Interpolation model

We generated direct interpolation models to predict potential richness and to be compared with S-SDMs, using the inverse distance weighting (IDW) technique in ArcGIS 9.3 (ESRI, Redlands, California, USA). Only one or two species had been collected from most of the collection localities, irrespective of the number of specimens, which may reflect the fact that most of the collections were opportunistic and only detected the most common and conspicuous species (Fig. 2). For this reason and to avoid the overrepresentation of very low numbers in the interpolation analyses, we defined a fishnet of a 10 × 10 km cell size over the Iberian Peninsula and counted the number of different species occurring in each cell. This aggregated data set was interpolated using IDW with a radius of 12 and a power of 1.6 to reduce the influence of the closest points, thus producing a smoother final surface (Fig. 1B).

Stacked species distribution models (S-SDMs)

To generate the different SDMs described below, we used Maxent, one of the modeling technique most used in the literature, as it provides consistently reliable results in the work performed on the comparison of other modeling techniques (Elith et al. 2006). It is a machine-learning technique based on the principle of maximum entropy (Phillips et al. 2006; Phillips and Dudík 2008; Elith et al. 2011). This technique seeks a marginal suitability function for each variable that matches the empirical data, is maximally uninformativ e (close to the uniform distribution) elsewhere, and has a mean equal to that of the empirical data (Warren and Seifert 2011). Because the requirement of equal means can produce the undesirable result of overfitting the data used to train the model, Maxent has a regularization multiplier (β) that can tune the model to avoid such overfitting. In this study, we used Maxent 3.3.3e (MAXENT; Phillips et al. 2006, software available at https://www.cs.princeton.edu/~schapire/maxent/) with only the convergence threshold (0.00001) and the number of background points (10,000) set to their default values. To avoid overfitting, we increased the regularization multiplier (β = 2). This choice produced less open-ended response curves.

The nineteen Worldclim 1.4 bioclimatic variables at spatial resolution of about 1 × 1 km² (http://www.worldclim.org) were used as independent variables. These bioclimatic variables result from the global land area interpolation of the climate point data for the period 1950–2000 (Hijmans et al. 2005). They were intersected with the presence data and with 10,000 points selected at random and used as the background data set. To avoid multicollinearity, we performed a correlation analysis on the background data set and eliminated one of the variables in each pair that showed a Pearson correlation value >0.8. The final data sets included Isothermality (bio_03), Temperature Annual Range (bio_07), Mean Temperature of Warmest Quarter (bio_10), Mean Temperature of Coldest Quarter (bio_11), Precipitation Seasonality (bio_15, coefficient of variation), Precipitation of Wettest Quarter (bio_16), and Precipitation of Driest Quarter (bio_17). Because Grimmia species are sensitive to the chemical characteristics of the rocks on which they grow, we also included a soil acidity/alkalinity categorical vari-

![Figure 2](image-url). Percentage of collection localities (100 km² pixels) by Grimmia species count for (A) opportunistic sampling localities and (B) systematically sampled localities.
able that was derived from the European Soil Database v2.0 (Data S1).

The performance of the Maxent models was evaluated using the area under the curve (AUC) of the receiver operating characteristic (ROC) curve. Although the validity of this statistic as a technique for evaluating models has recently been challenged (Lobo et al. 2008; Peterson et al. 2008), particularly for presence-only data, its use in contexts similar to the present study has also been justified (Phillips et al. 2006; Anderson and Gonzalez 2011). For each species, the AUC was calculated through cross-validation based on 10 replicates, and the final model was the average of the replicates.

There are many ways to generate richness maps from original models (Ferrier and Guisan 2006). Following Calabrese et al. (2014), we summed the original values obtained for each individual taxon without reclassification for presence/absence. The resulting richness model can be considered as an estimate of the potential number of species that could be present in each particular pixel (Gelfand et al. 2005; Wilson et al. 2005).

Validation and comparison of the richness models

We used IDW to create four partial richness models for the entire Iberian Peninsula. Each partial model excluded one of the four systematically sampled data sets (Fig. 1B). We also used IDW to create a reference model using all of the available presence data. In addition, we generated four partial S-SDMs, each of which excluded the data for one of those four systematically sampled data sets, and we generated a reference model using all of the available presence data. For the partial models, the number of unique presences is indicated in Table 1; duplicate presences were removed in all cases. Partial models were generated with the aim to test the predictive power of the models (IDW and S-SDMs) in each of the four systematically sampled areas. Following Hernandez et al. (2006), the four partial models were compared with the reference model, considered to be the most representative of the true distribution of the species given the limitations of the modeling method, the species occurrence, and environmental data available. Lastly, for each of the four systematically sampled areas (i.e., setting the analysis window to only that area) and separately for IDW and S-SDMs, we calculated the Pearson correlation coefficient between the reference model and the partial model generated excluding the area being tested.

### Results

**Species data**

Although Grimmia mosses grow on the tops of rocks, primarily in open areas, and are relatively conspicuous to nonspecialists, collections of Grimmia exist only for 557 of the ~8,100,100 km² cells covering the Iberian Peninsula. For opportunistic sampling localities (Fig. 2A), 86.3% of the collection localities have one or two species, 13.12% have three or four species, and the remaining 0.58% have five or six species. In the systematically sampled localities (Fig. 2B), 44.39% of the localities have one or two species of Grimmia, 42.44% have three or four species, and the remaining 13.18% have from five to eight species.

**Richness models**

As expected, the reference IDW interpolation model is biased toward those areas that have been systematically collected (Fig. 1B), even though we reduced the power parameter to increase the smoothness of the model. The final surface is highly heterogeneous, a result that may reflect the lack of appropriate surveys for most of the Iberian Peninsula.

The reference S-SDM for the genus is shown as the background in Figure 3. The more oceanic areas of the northwestern Iberian Peninsula and the mountain ranges in the interior exhibit conditions that support higher numbers of species. This result coincides with the expected richness pattern based on expert knowledge and the ecology of the resident species (Casas et al. 2006). The average test AUC over the 10 replicates is indicated in Table 2. The AUC values are generally high, except for those species whose main area of distribution is within one of the testing areas. The exclusion of those presences to train the model results in poor AUC values. For example, G. funalis is distributed primarily in the Pyrenees (Area 1), and, when this area is eliminated, only seven

### Table 1. Correlation between the reference and partial models.

<table>
<thead>
<tr>
<th>Area tested for correlation, data</th>
<th>Reference IDW vs. Partial IDW</th>
<th>Reference vs. Partial Maxent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excluded from the model generation</td>
<td>Presences</td>
<td></td>
</tr>
<tr>
<td>Area 1</td>
<td>492</td>
<td>−0.2099</td>
</tr>
<tr>
<td>Area 2</td>
<td>484</td>
<td>−0.0659</td>
</tr>
<tr>
<td>Area 3</td>
<td>426</td>
<td>−0.325</td>
</tr>
<tr>
<td>Area 4</td>
<td>523</td>
<td>−0.0558</td>
</tr>
</tbody>
</table>

*The reference models were generated using all of the presence data. For each partial model, the presences of the corresponding area were removed, and that window area was then used for the correlation calculation. The correlation between the reference IDW and Maxent models for the entire Iberian Peninsula was 0.1542. Areas as in Figure 1.*
presences remain to train the model; the sparseness of the remaining training data affects the accuracy of the model (Wisz et al. 2008; Mateo et al. 2010b). A more extreme example would be *G. tergestina*, with 14 of 16 known presences inside Area 3; training the model with the only two presences outside Area 3 explains the low AUC of the model in this area. Finally, the widespread *G. trichophylla* has moderate AUC values, as commonly found for generalist species (McPherson and Jetz 2007; Mateo et al. 2010a). Additionally, preliminary molecular results point that *G. dissimulata* and *G. meridionalis* represent taxa independent from *G. trichophylla*, and the merge of presences of taxa with different environmental requirements would explain the obtained AUC values.

Figure 3. Maxent richness model of *Grimmia* for the Iberian Peninsula. (A) Overlap between the richness model (background) and number of species recorded per 100 km² pixel. (B) and (C) Examples of areas in the Iberian Peninsula for which the richest areas according to the Maxent model match (solid lines) and do not match (dotted lines) the richest areas with more taxa recorded in the natural history collections.
The IDW approach produced the map with the lowest values of species richness (mean: 1.8, max: 8 species) while the S-SDMs approach provided the map with the highest values of species richness (mean: 2.2, max: 10 species).

The correlation between the reference IDW and reference S-SDMs models was low ($r = 0.1542$), indicating that the models differ in the information they provide (Table 1). The partial S-SDMs (i.e., generated without information from a given area) are good predictors of the number of species in that area, as shown by the medium to high correlation values (0.4504 to 0.9954). The lowest correlation is obtained from the tests performed with Area 3 because this area includes 19.16% of the presences. Although the correlation is still high, eliminating this information from the model affects the accuracy of the reference model. In contrast, removing the information from the IDW interpolation has a dramatic effect on the models, as indicated by the low and negative correlation values ($-0.0558$ to $-0.3250$).

**Discussion**

**Potential richness models**

Our results demonstrate for the first time that the combination of individual species models (i.e., S-SDMs) offers a useful tool for identifying detailed richness patterns for inconspicuous taxa such as bryophytes. The richness patterns generated by S-SDMs coincide with the expected richness pattern based on expert knowledge and therefore are a powerful tool for basic biodiversity applications (e.g., biogeography and conservation).

Our results confirm that S-SDMs are less affected by collection biases than are spatial interpolation models and that the information about the richness provided by the S-SDMs is improved relative to the raw information contained in the NHCs. In this study, we found a correlation between the reference (i.e., created using all of the available presence data) IDW and S-SDM models of 0.1542, confirming that the richest areas according to the SDMs differ from the areas with more taxa recorded in the NHCs. In the case of the S-SDMs, we found high correlations between the partial models generated eliminating important pieces of information and the model generated using all of the available information. This result shows that the SDMs investigated were robust and could extract latent information from the data. In contrast, the spatial interpolation models performed poorly and were unable to represent the actual richness (Table 1), although, in some cases, with a good spatial coverage, the implementation of spatial interpolation techniques may increase the predictive accuracy of habitat models, especially for species with loose association with environmental variables (Brotons et al. 2007).

However, SDMs are not free of uncertainty when applied to conservation programs (Rondinini et al. 2006; Carvalho et al. 2010; Underwood et al. 2010; Aranda and Lobo 2011; Mateo et al. 2013a). For example, they may not include all environmental, ecological, and histo-

### Table 2. Maxent test AUC obtained by 10-fold cross-validation.

<table>
<thead>
<tr>
<th>Species</th>
<th>Unique presences</th>
<th>Area 1</th>
<th>Area 2</th>
<th>Area 3</th>
<th>Area 4</th>
<th>Reference model</th>
</tr>
</thead>
<tbody>
<tr>
<td>G. alpestris</td>
<td>23</td>
<td>0.9927</td>
<td>0.9934</td>
<td>0.9939</td>
<td>0.9929</td>
<td>0.994</td>
</tr>
<tr>
<td>G. anodon</td>
<td>8</td>
<td>0.8983</td>
<td>0.9033</td>
<td>0.9033</td>
<td>0.8576</td>
<td>0.9033</td>
</tr>
<tr>
<td>G. caespitica</td>
<td>20</td>
<td>0.9818</td>
<td>0.9858</td>
<td>0.9855</td>
<td>0.985</td>
<td>0.9859</td>
</tr>
<tr>
<td>G. crinita</td>
<td>25</td>
<td>0.9655</td>
<td>0.965</td>
<td>0.9713</td>
<td>0.9703</td>
<td>0.9705</td>
</tr>
<tr>
<td>G. decipiens</td>
<td>203</td>
<td>0.8585</td>
<td>0.8106</td>
<td>0.8704</td>
<td>0.8592</td>
<td>0.8596</td>
</tr>
<tr>
<td>G. funalis</td>
<td>19</td>
<td>0.7555</td>
<td>0.9522</td>
<td>0.9539</td>
<td>0.9559</td>
<td>0.9528</td>
</tr>
<tr>
<td>G. hartmanii</td>
<td>41</td>
<td>0.9426</td>
<td>0.8927</td>
<td>0.9419</td>
<td>0.945</td>
<td>0.9492</td>
</tr>
<tr>
<td>G. laevigata</td>
<td>130</td>
<td>0.8617</td>
<td>0.7767</td>
<td>0.8751</td>
<td>0.8591</td>
<td>0.8526</td>
</tr>
<tr>
<td>G. lisae</td>
<td>74</td>
<td>0.8213</td>
<td>0.8396</td>
<td>0.7792</td>
<td>0.8238</td>
<td>0.8165</td>
</tr>
<tr>
<td>G. montana</td>
<td>178</td>
<td>0.9287</td>
<td>0.8829</td>
<td>0.9364</td>
<td>0.9257</td>
<td>0.9312</td>
</tr>
<tr>
<td>G. orbicularis</td>
<td>149</td>
<td>0.8695</td>
<td>0.8769</td>
<td>0.8924</td>
<td>0.8611</td>
<td>0.8627</td>
</tr>
<tr>
<td>G. ovais</td>
<td>28</td>
<td>0.7952</td>
<td>0.8442</td>
<td>0.8425</td>
<td>0.831</td>
<td>0.8421</td>
</tr>
<tr>
<td>G. pulvinata</td>
<td>229</td>
<td>0.8176</td>
<td>0.81</td>
<td>0.8552</td>
<td>0.8103</td>
<td>0.806</td>
</tr>
<tr>
<td>G. ramondii</td>
<td>12</td>
<td>0.963</td>
<td>0.9774</td>
<td>0.9615</td>
<td>0.954</td>
<td>0.9477</td>
</tr>
<tr>
<td>G. reflexidens</td>
<td>9</td>
<td>0.9778</td>
<td>0.9883</td>
<td>0.9883</td>
<td>0.9621</td>
<td>0.9883</td>
</tr>
<tr>
<td>G. tergestina</td>
<td>16</td>
<td>0.8135</td>
<td>0.8858</td>
<td>0.5898</td>
<td>0.8589</td>
<td>0.8375</td>
</tr>
<tr>
<td>G. torquata</td>
<td>9</td>
<td>0.9007</td>
<td>0.9104</td>
<td>0.9617</td>
<td>0.8726</td>
<td>0.9007</td>
</tr>
<tr>
<td>G. trichophylla</td>
<td>180</td>
<td>0.7447</td>
<td>0.7135</td>
<td>0.752</td>
<td>0.7386</td>
<td>0.7397</td>
</tr>
</tbody>
</table>

1Areas as in Figure 1.
torical factors that affect species distributions (Guisan and Zimmermann 2000). The main caveat of using S-SDMs to generate species richness patterns is that it tends to overestimate actual species richness (Trotta-Moreu and Lobo 2010; Dubuis et al. 2011; Mateo et al. 2012). A suggested solution to this problem has been to use macroecological models of species richness to constrain S-SDMs (see Dubuis et al. 2011; Guisan and Rahbek 2011). Other alternative, “hybrid” models that include basic mechanisms, such as dispersal and demography into SDMs (Boulangeat et al. 2012; Dullinger et al. 2012), have provided more realism and better predictive performance than traditional models (Thuiller et al. 2013). On the other hand, little is known about the performance of S-SDMs in inconspicuous groups. These results suggest that this distributional proposal should be considered as a preliminary step and that a higher number of database records are needed for the insufficiently surveyed cells.

Considerations for conservation strategies

Two main types of distribution data are frequently used in conservation planning to obtain distribution maps: observed and predicted data (Rondinini et al. 2006). Additionally, previous studies have discussed the advantages and drawbacks of different approaches that are used to generate species richness patterns as the basis for subsequent conservation measures (Freitag and Jaarsveld 1998; Bombi et al. 2011). Currently, one of the criteria employed in spatial conservation networks is to focus logistic and economic efforts in richness areas (Prendergast et al. 1999; Margules and Pressey 2000; Myers et al. 2000). The most intuitive and simple approach for reserve design is aggregating the NHC data and transferring the resulting information to a geographic space (equivalent to the spatial interpolation model used here). However, the results of this study show that such networks may be strongly biased toward those areas with higher numbers of collections and would not necessarily represent the richest areas. However, it is rarely the case that the areas of interest are sufficiently surveyed, and the existing surveys are certainly not adequate for this purpose in the case of small and inconspicuous organisms and for most geographic areas, including the megadiverse tropics (Cayuela et al. 2009). But, see the limitations described in the first part of this discussion.

Reserve selection is sensitive to the type of distribution data used for the selection (Freitag and Jaarsveld 1998; Wilson et al. 2005; Rondinini et al. 2006). For areas in which detailed surveys have been performed, the richest areas according to the S-SDM match those with more taxa in the NHCs (Fig. 3B and C, solid lines), whereas such agreement does not occur in the areas for which only opportunistic samples are available (Fig. 3B and C, dashed lines). In general, the richest areas identified by spatial interpolation models coincide with the best-surveyed areas, indicating that such models represent a restricted approach that excludes poorly sampled areas. In contrast, the S-SDMs show a richness pattern that is independent of the collection effort and is, therefore, more appropriate for biodiversity patterns studies.

There is a consensus that SDMs are an appropriate tool for management and conservation programs and for the identification of suitable areas for threatened or otherwise rare species (Parviainen et al. 2008, 2009; Williams et al. 2009). Our results show that the S-SDMs derived from natural history collections improve the incomplete information inherent in the scattered nature of sampling distributions, particularly in the case of opportunistic sampling, by assessing the potential richness of clearly under-surveyed sites. Therefore, these models contribute additional information that is not obvious from the limited presence data.

We consider that S-SDMs should be used to produce potential maps of species richness when information is limited and different types of distribution data are available (opportunistically sampled vs. systematically sampled localities). Consequently, S-SDMs could be highly useful for reserve design at the national scale. This conclusion is in agreement with several studies that stressed that predictive models can effectively address the problem of insufficient field survey and museum data (Loiselle et al. 2003; Maes et al. 2005; Rodriguez et al. 2007; Sérgio et al. 2007b; Braunisch and Suchant 2010; Costa et al. 2010; Mateo et al. 2013a) and offer benefits for conservation prioritization (Elith and Leathwick 2009). Using models to predict distributions is also likely to become increasingly important as climate change and other dynamic processes are incorporated into conservation planning efforts (Rondinini et al. 2006; Underwood et al. 2010). Although the S-SDMS seems to outperform the IDW approach, both approaches combined may provide improved information for directing future efforts to conserve these inconspicuous taxa and targeting areas for monitoring and management. Sánchez-Fernández et al. (2011) demonstrated that species distribution models, combined with a survey effort map, might be used to select the location of future surveys by prioritizing those species-rich areas with a low level of sampling effort. Here, the IDW output could be seen as a proposal of a survey effort map which overlaid with the S-SDM output could allow us to locate those areas where more sampling effort is necessary. Another possibility is the combination of S-SDMs with macroecological models (Guisan and Rahbek 2011).
Conclusions

From our study, we can conclude that (1) using a modeling approach based on the combination of individual species models (stacking species distribution models) allowed the identification of detailed richness patterns for inconspicuous taxa such as bryophytes. (2) Stacking species distribution models are less affected by collection biases or types of distribution data than are spatial interpolation models (traditional approach) and therefore could be highly useful for conservation purposes. (3) Spatial interpolation models may provide a complementary view to the modeling approach.

For the Iberian Peninsula, further studies should aim at refining this modeling approach with more taxonomic groups within bryophytes, and also with additional geographic information that would inform about new areas for conservation for inconspicuous groups such as bryophytes.

Acknowledgments

We thank the researchers who made this work possible through the collection and identification of the specimens. The first author is funded by the Fundação para a Ciência e Tecnologia (FCT) under a Postdoctoral fellowship (SFRH/BPD/64665/2009) co-funded by the Programa Operacional Ciência e Inovação – 2010 and Fundo Social Europeu. RGM is supported by a Marie Curie Intra-European Fellowship within the 7th European Community Framework Programme (ACONITE, Nº 622620).

Conflict of Interest

None declared.

References


Modelling Bryophyte Species Richness Patterns

H. Hespanhol et al.


Supporting Information

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

Data S1. Simplified grid specifying the acidic/basic nature of each pixel used.