Using species distribution models at local scale to guide the search of poorly known species: Review, methodological issues and future directions

Mauro Fois[a],[*], Alba Cuena-Lombraña[b], Giuseppe Fenu[a], Gianluigi Bacchetta[a],[c]

[a] Centro Conservazione Biodiversità, Dipartimento di Scienze della Vita e dell'Ambiente, Università degli Studi di Cagliari, Viale S. Ignazio da Laconi, 13, Cagliari 09123, Italy
[b] Dipartimento di Biologia Ambientale, 'Sapienza' Università di Roma, P.le A. Moro 5, 00185 Roma, Italy
[c] Hortus Botanicus Karalitanus (HBK), Università degli Studi di Cagliari, Viale Sant'Ignazio da Laconi, 9–11, Cagliari 09123, Italy

ABSTRACT

Among specific applications of species distribution models (SDMs), the use of SDMs probabilistic maps for guiding field surveys is increasingly applied. This approach is particularly used for poorly known and/or cryptic species in order to better assess their distribution. One of the most interesting aspects of these applications is that predictions could be clearly validated by real data, subsequently obtained in the field. Despite this important difference from other applications, to our knowledge, the efficiency of different algorithms, metrics for model evaluation and algorithm-specific settings have not yet been sufficiently investigated.

This research performs a literature survey to investigate which species, study area characteristics, variables, techniques and settings were used or suggested by previous authors. We then applied the most common approaches to guide field surveys for a set of 70 vascular plants in an endemic-rich area of Sardinia (Italy) of approx. 9000 ha, the flora of which was deeply investigated during the last two years. Our main aims were: (1) to use pre-model records for predicting the potential occurrence of plant species with different sample size, detectability and habitat preference, (2) to apply results for guiding searches for new populations of poorly known species, (3) to calculate the model performance according to independent real presence/absence data (testAUC) and (4) to compare different modelling data input and settings on the testAUC basis.

By emphasizing the importance of field verification, both the review and the worked example supported the reliability of SDMs for a wide range of species to understand where species could potentially be present and therefore to optimise resources for the search of new species localities. This study may help understand and summarise the most applied methodological approaches and to highlight future directions for this practical application. Without underrating the importance of most common recommendations, practitioners are encouraged to test the ability of this SDMs' application with their own data. Indeed, large gaps in species' types (e.g. insects) and in regions covered by these kind of studies (e.g. many African and Asian territories) were found. Furthermore, eventual biases due to lack of data, experience or staff, have in this experimental case less irreparable consequences than others, such as conservation assessments based on future projections, which cannot be otherwise adjusted by explicit data from ground validation.

1. Introduction

Species distribution models (SDMs) have become in recent years one of the most widely used tools in ecology (e.g. Guisan and Zimmermann, 2000; Guisan and Thuiller, 2005). The principle of SDMs is to relate species occurrences to environmental data in order to find ecological and evolutionary insights (Elith and Leathwick, 2009). This family of methods has a broad range of applications, including the study of niche evolution (e.g. Warren et al., 2008; Montemayor et al., 2017), the prediction of climate change impacts (e.g. Fois et al., 2016; Vessella et al., 2017), and the identification of areas for conservation (e.g. Kaky and Gilbert, 2016; Bosso et al., 2018). Another relevant application is the use of SDMs to guide efforts to locate new populations of poorly known species of conservation concern. Such model-based sampling of endangered and rare species, involving reiterative alternation of modelling and field sampling, is recognized by conservation biologists as very promising and cost-efficient (e.g. Le Lay et al., 2010; Volis, 2016). Indeed, information on the geographic distribution of en-

[*] Corresponding author.
Email address: mfois@unica.it (M. Fois).

dangered and rare species is usually scarce (Pulliam and Babbitt, 1997), and getting precise distribution maps for such species is difficult and often requires intensive surveys (Edwards et al., 2005; Guisan et al., 2006; Williams et al., 2009). Nonetheless, although poorly known species are increasingly in need of predictive distribution modelling, for both monitoring and conservation management purposes, they are at the same time difficult to model due to their limited number of available occurrences (Lomba et al., 2010; Bosso et al., 2016; Chen et al., 2017).

Especially in the last decade, there has been an increasing literature on methodological questions addressed in the enhancement of SDM applications. Examples are the relative performance of modelling methods (e.g. Araújo and New, 2006; Elith and Graham, 2009), the model performance evaluations (e.g. Lobo et al., 2008; Warren and Seifert, 2011), the application of different environmental predictors and resolutions (e.g. Peterson and Nakazawa, 2008; Bucklin et al., 2015), and differences among sample sizes (e.g. Hernandez et al., 2006; Proosdij et al., 2016), quality of the species occurrence data (e.g. Varela et al., 2014; Bloom et al., 2018), or species characteristics (e.g. McPherson and Jetz, 2007; Guillera-Arroita, 2017).

In case of poorly known species, the amount and quality of available occurrence data is implicitly low, and dealing with such limitations is more relevant than in other SDM applications. Maximum Entropy – MaxEnt (Phillips et al., 2006) is widely recognized as the most used technique, especially for small sample sizes and poorly known species distributions (Elith et al., 2006), since it only requires the information about presences, instead of both presences and absences and because, especially in the case of rare and elusive species, opportunistic data, a common example of presence-only data, often represent the largest set of available data (Virgili et al., 2017). Nonetheless, many other algorithms, such as - genetic algorithm for rule set production – GARP (Stockwell, 1999) and environmental niche factor analysis – ENFA (Hirzel et al., 2002), have been successfully used for modelling small sample sizes (e.g. Jiménez-Valverde et al., 2008a; Le Lay et al., 2010; Ortega-Huerta and Vega-Rivera, 2017). Alternatively, in the 'consensus approach' (Araújo and New, 2007) outputs of multiple algorithms are combined in order to reduce uncertainties in various model predictions. Differences among different settings within the same modelling technique have been also found to be relevant (e.g. Anderson and Gonzalez, 2011; Radosavljevic and Anderson, 2014; Halvorsen et al., 2016). Even if the omission of the entire regularization process has been advocated by Halvorsen (2013) and Halvorsen et al. (2015), regularization multiplier is a user-specified coefficient that is generally applied to set the degree of generality in the resulting MaxEnt models (Phillips et al., 2006, 2017); this parameter should be set in order to achieve a trade-off between possible overfitting (in case of low regularization multiplier values) and generalization (in case of high values) (Radosavljevic and Anderson, 2014; Moreno-Amat et al., 2015). The regularization multiplier selection is suggested to be based on an accurate analysis of each species-specific case (Anderson and Gonzalez, 2011) or adapted according to the specific purpose of the study (Moreno-Amat et al., 2015), and is often based on the Akaike information criterion (AIC), or on the Bayesian Information Criterion (BIC) (e.g. Warren and Seifert, 2011; Fourcade et al., 2018).

Another important aspect is how to measure the model performance. Generally, the initial sample is partitioned into test and training sets for model evaluation or, even in the case of the 'jackknife' approach suggested by Pearson et al. (2007) for small sample sizes, localities will still be spatially correlated, because the same surveyor(s) collected both data sets, and thus both sets of localities are constrained to come from the general areas the surveyor(s) visited (Veloz, 2009). The alternative is to train the model(s) with known locality data and test them with new, independent field surveys (Halvorsen, 2012; Searcy and Shaffer, 2014). Following this last suggestion, testing an SDM pre-

diction is theoretically comparatively easy; one only has to visit a set of previously unsampled localities and determine whether or not the target species is present. By repeating this procedure across multiple localities, one can make a statistical statement about the fit of a model to the true species distribution and compare alternative models. Nonetheless, although this kind of model validation sounds relatively straightforward, there are few published examples of ground truthing SDMs (Halvorsen, 2012; Searcy and Shaffer, 2014).

In this paper we first reviewed and analyzed the available literature about the use of SDMs to guide the search for unknown populations of poorly known and/or endangered species. Specifically, we focused on the following aspects: (1) where studies were focused and which was the study area extent, (2) what kind and how many variables were used, (3) which type of organisms and how many points were modelled, (4) which modelling techniques were used.

Therefore, we furtherly investigated this SDMs application through the case study of the Regional Park of Montarbu, an endemic-rich and environmentally diverse area in CE Sardinia (Italy) of about 9000 ha of extension. Specifically, an initial incomplete set of 346 presence records of 70 vascular plant species was used for modelling the potential distribution and for guiding field surveys. After two intensive field campaigns, several new localities were georeferenced and the independent dataset was used to measure model performances.

2. Literature review

2.1. The use of species distribution models (SDMs) to speed up the field search of new species localities

We conducted a literature review of modelling papers published from 2000 to December 2017. A research of the peer-reviewed articles was carried out in separate Google Scholar searches with the terms 'species distribution model' or 'environmental niche model' or 'habitat suitability model' combined with 'ground validation' or 'field validation' or 'field search' or 'ground search' or 'population discover' or 'guiding field surveys'. A number of 4386 papers were firstly screened by the title. The remaining 347 papers were reduced up to 89 after a further abstract revision. These last were fully revised and only 28 studies (Table 1) were retained because they were matching all the three following elements: (i) they were local experimental modelling studies, (ii) they were applying models to species of conservation concern, and (iii) they used models to guide searches for new populations.

2.2. Review's findings and discussion

Our literature search and selection procedure uncovered a diverse set of approaches in terms of number of occurrences used for modelling (from one to 1033) and extent of the study area (from 470 km^2 to 3×10^6 km^2; Table 1). Most of studies were carried out in Europe and America (Fig. 1a); this highlights the unbalanced efforts in this sense, which may increase especially in highly biodiverse areas, such as many African and Asian territories. In 16 papers, only one species was analyzed, confirming the specificity of such researches, which often require deep field efforts and knowledges that may not be affordable for many species at the same time. The most studied organisms were plants (13 studies), followed by birds (five studies), but SDMs were also used to search new population localities of most of the organism types, including butterflies, mammals, reptiles, amphibians and crustaceans (Table 1). Implicitly, this kind of applications is generally useful for cryptic species but, except for butterflies, other insects were only once considered (Rinnhofer et al., 2012). Among modelling techniques, the presence-only or presence-background methods were the most frequently applied with MaxEnt (15 studies) and GARP (six studies), while, among presence-absence methods, ENFA was the most fre-

Table 1

Summary, in chronological order, of the 28 studies applying species distribution models (SDM) to guide searches for new populations. In some cases all information was not available (NA). See the respective reference for each modelling technique description.

|	Organism(s)	N (min-max)	Surface (km^2)	Modelling technique(s)	Reference
1	Reptiles (9 spp.), Mammals (5), Birds (64), Plants (75)	9-NA	NA	GAM	Pearce et al. (2001)
2	Birds (89 spp.)	NA	112,000	GARP	Feria and Townsend Peterson (2002)
3	Plant (1 sp.)	187	1,215	Mahalanobis distance	Boetsch et al. (2003)
4	Reptiles (11 spp.)	4-108	587,041	GARP	Raxworthy et al. (2003)
5	Plant (1 sp.)	77	41,293	GLM-ENFA	Engler et al. (2004)
6	Plant (1 sp.)	132	2,272	CART	Bourg et al. (2005)
7	Lichens (5 spp.)	NA	3,000,000	BRT	Edwards et al. (2005)
8	Plant (1 sp.)	19	6,460	FloraMap	Jarvis et al. (2005)
9	Bat (1 sp.)	1033	19,000	GLM	Greaves et al. (2006)
10	Plant (1 sp.)	30	41,293	GAM	Guisan et al. (2006)
11	Butterfly (1 sp.)	17	NA	ENFA	Jiménez-Valverde et al. (2008b)
12	Plant (1 sp.)	1	19,622	Environmental distance analysis	de Siqueira et al. (2009)
13	Plants (6 spp.)	9-129	1200	MaxEnt, GLM, ANN, RF, ME	Williams et al. (2009)
14	Plants (3 spp.)	14-31	2,054	GAM, ENFA, ensemble	Le Lay et al. (2010)
15	Reptiles (49 spp.)	3-256	1,600	MaxEnt, Garp	Costa et al. (2010)
16	Bat (1 sp.)	17	90,100	ENFA, MaxEnt	Rebelo and Jones (2010)
17	Plant (1 sp.)	237	2,305	Maxent	Edvardsen et al. (2011)
18	Bird (1 sp.)	15	NA	MaxEnt	Botero-Delgadillo et al. (2012)
19	Insect (1 sp.)	7-21	470	MaxEnt	Rinnhofer et al. (2012)
20	Amphibian (1 sp.)	17	NA	MaxEnt	Groff et al. (2014)
21	Amphibian (1 sp.)	33	7123	MaxEnt	Peterman et al. (2013)
22	Plant (1 sp.)	4	NA	ENFA, MaxEnt	Särkinen et al. (2013)
23	Butterfly (1 sp.)	10	NA	MaxEnt	Verovnik et al. (2014)
24	Bird (1 sp.)	737	2,586	MaxEnt	Aizpurua et al. (2015)
25	Plant (1 sp.)	8	721	MaxEnt	Fois et al. (2015)
26	Plants (8 spp.)	5-295	110,000	MaxEnt	McCune (2016)
27	Birds (12 spp.)	26-151	NA	ENFA, GARP, MaxEnt	Ortega-Huerta and Vega-Rivera (2017)
28	Freshwater crustacean (1 sp.)	53-54	15,000	MaxEnt	Rhoden et al. (2017)
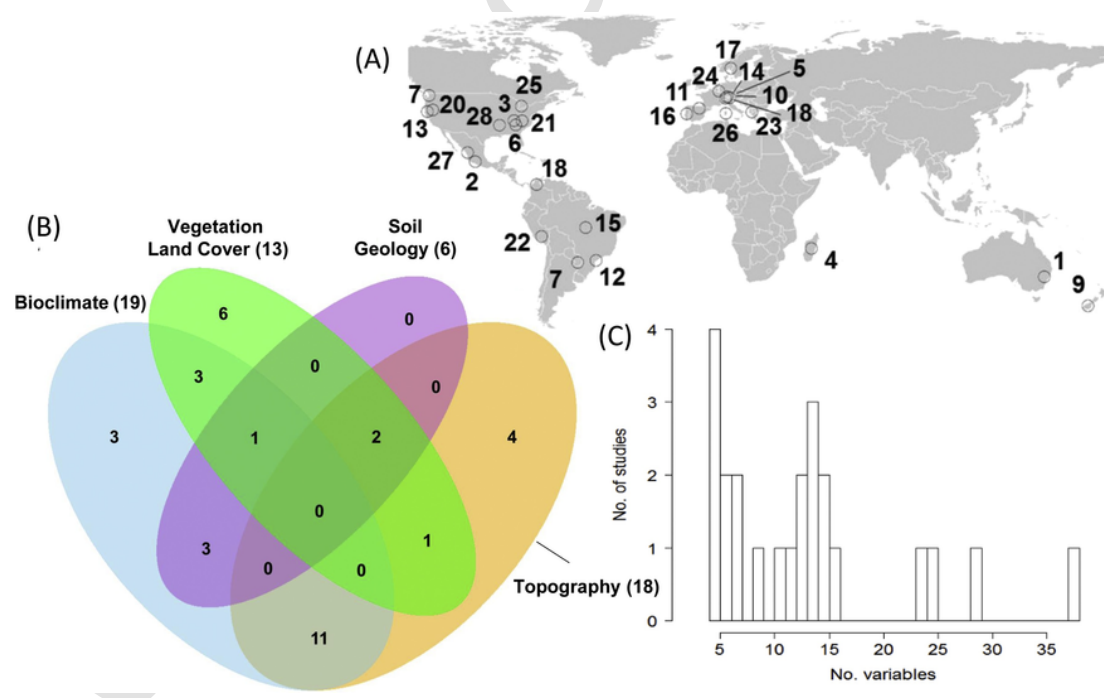
Fig. 1. Illustrative summary of (A) the distribution map of each one of the analysed study, (B) types of variables used illustrated by a Venn diagram and (C) number of variables used. Numbers in the map and all data are from the 28 articles cited in Table 1.

quently applied (five studies). In five cases, different model techniques were compared, while in only two cases model outputs were combined according to the ensemble approach (Table 1). In summary, main ef-

forts were generally destined for searching the species in field rather than for improving model complexity and, possibly, performances.

Mainly due to its clear influence on the distribution pattern of many species, bioclimatic variables were the most used (Fig. 1b). The distribution of the number of variables among each study case was bimodal, with a peak of four variables used and another one around 13 variables (Fig. 1c). Such results could be explained by the common use of combining different variable categories. In most of cases, bioclimatic variables were related to topographic variables but other combinations of two (mostly) or three variable categories were also applied (Fig. 1b). The same pattern was observed in other reviews (e.g. Bradie and Leung, 2017; Fourcade et al., 2018) that identified bioclimatic variables and elevation as the most commonly employed predictors in SDM studies. Otherwise, the combination of different type of variables could increase the ability of describing the complexity of the processes that limit species' ranges (Pearson and Dawson, 2003). The other peak around the usage of four variables is mainly the result of efforts for pre-selection of variables to avoid strongly correlated variables entering the models. With this common practice to limit the potential for overfitting, candidate covariates are reduced in number based on their degree of collinearity, usually assessed by variance inflation factors (VIFs) or Pearson correlation test (Dormann et al., 2013).

3. The use of SDMs to guide the search of poorly known species: the example of the Regional Park of Montarbu

3.1. Study area and target species

The Regional Park of Montarbu (Fig. 2) is located in the central-eastern part of Sardinia, mainly within the territories of Seui municipality and part of the Gairo, Seulo and Ussassai municipalities. Its geo-graphical coordinates are 9°23′ 03″E and 39° 53′ 37″N. The study area has a surface of about 9000 ha and consists of a group of isolated limestone mountains which are commonly called "Tonneri" or "Tacchi" (Loi and Lai, 2001). It includes mountainous peaks extending above 1200 m, such as Margiani Pubusa (1324 m asl) and Perda 'e Liana (1293 m asl), but also other different environments, such as steep cliffs and gorges of more than 100 m, small wetlands and vast scrublands and woodlands. Also, there are restricted vegetation assemblages of biogeographic and conservation interests, such as the relict formations of some temperate trees like *Taxus baccata*, *Ilex aquifolium*, *Acer monspessulanum* and *Ostrya carpinifolia*. It has been recently recognized as a "micro-hotspot" of biodiversity, due to its exceptional richness in endemic vascular plants (Fois et al., 2018). Even if several researchers have studied this area in the past (e.g. Arrigoni, 1965; Loi and Lai, 2001; Loi et al., 2004), further endemic plant species were recently described or found, such as *Pinguicula sehuensis* and *Senecio morisii* (Bacchetta et al., 2014; Calvo and Aedo, 2015), until totalizing 99 endemic vascular plant species.

For this study, we selected 70 vascular plant species of conservation concern, of which we were able to collect a minimum of three point occurrences (see Appendix A). Geo-referenced occurrences were taken from available literature and from specimens stored in the herbaria of Cagliari (CA), Sassari (SS and SASSA), Florence (FI), Catania (CAT) and Turin (TO). Taxonomical treatments were checked in order to mainly exclude synonymies; furthermore, both records accompanied by a textual description of the locality or by coordinates, were expert-based revised according to the species ecology and distribution knowledge. Finally, within each cell in the grid of 250 × 250 m – which had a width double to the grain of the environmental rasters – all the points apart from one were randomly removed (Newbold, 2010). In this way, we
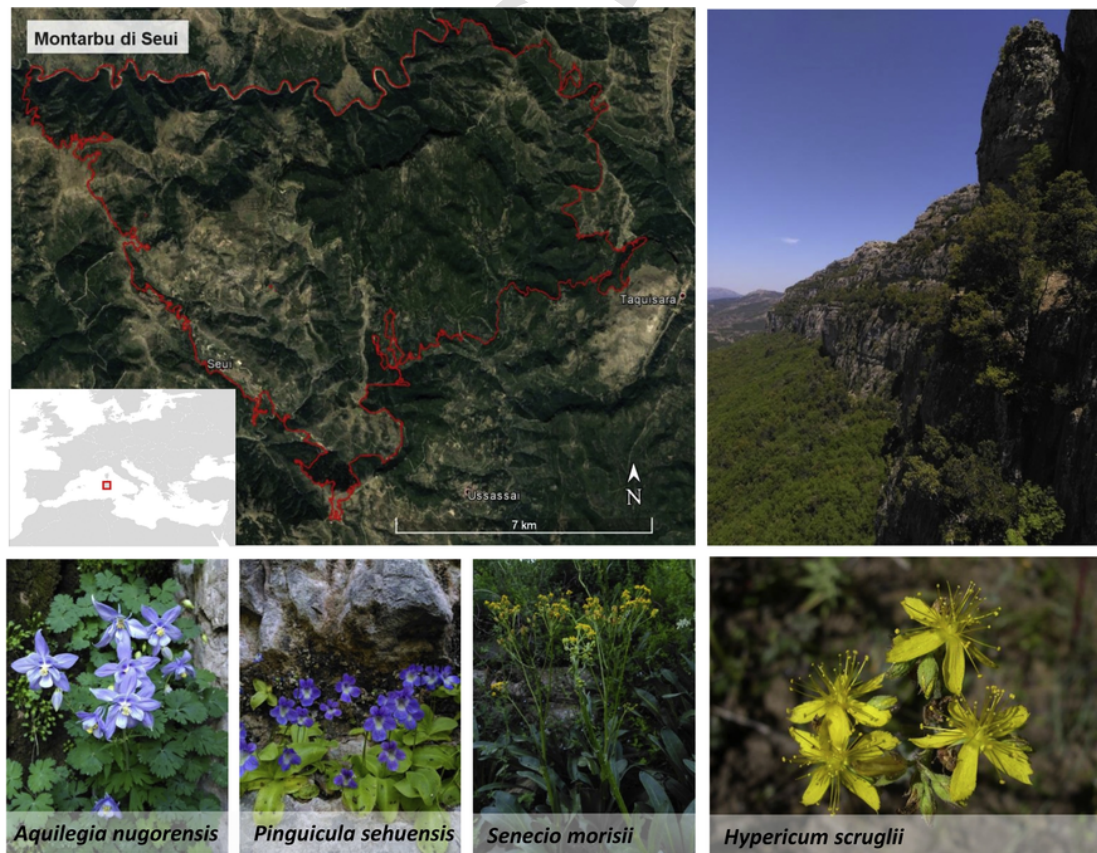

Fig. 2. The study area of Montarbu (central-eastern Sardinia, western Mediterranean Basin) with some typical vascular plant species and landscapes.

minimized spatial autocorrelation, given that closely situated localities are likely to represent similar environments (Mammola et al., 2018).

3.2. MaxEnt modelling and variable selection

Several techniques that have been used and projections can vary startlingly among these techniques (Naimi and Araújo, 2016). We chose to use MaxEnt 3.3.3k (Phillips et al., 2006), which is generally one of the most used methods for predicting species distribution (e.g. Merow et al., 2013) and, specifically, for guiding the search of unknown population localities (Table 1). In the setting panel, we selected the following options: random seed, remove duplicate presence records, write plot data, and 1000 maximum iterations. Models were repeated with four different regularization multiplier values, fixed at 0.5, 1, 2 and 5. A more recent version of MaxEnt (ver. 3.4.1; Phillips et al., 2017) is currently available; nonetheless, we preferred using the previous version since it is still the most tested one and because no comparisons with concrete examples, highlighting differences between the two versions, are nowadays available.

The use of most common bioclimatic variables, usually at the resolution of 1×1 km^2, has been recently criticized (e.g. Bradie and Leung, 2017; Fourcade et al., 2018). The large majority of these studies included all or some of the 19 bioclimatic variables popularized by the Worldclim project (Hijmans et al., 2005). Based on the assumption that climate must play an important role in driving species distributions, they have become a standard default predictor set in most modelling studies. Furthermore, it is recommended to include, at least, two different kinds of variables according to the ecology of each species and to test possible problems of multicollinearity. Nonetheless, high resolution and diverse variables are not always available and, despite their possible limitations, the globally retrievable datasets have been successfully tested in most of the reviewed cases. For these reasons, we used the most commonly employed predictors (bioclimatic, topographic and soil types) at the-at least in Europe-available resolution of 250×250 m. In particular, we used the freely available bioclimatic variables at the European scale (Metz et al., 2014), soil USDA classification at 250 m resolution (Hengl et al., 2017) and the Shuttle Radar Topographic Mission (SRTM) digital elevation data at the 90 m resolution (Jarvis et al., 2008). Aspect and slope were calculated on the basis of the digital elevation data using the terrain analysis tool in QGIS version 2.4.0 (QGIS Development Team, 2013). All 90 m rasters were resampled to the same pixel size (250×250 m) like the rest of variables by the neighbor method in R, using the package 'raster' (Hijmans, 2014). Multicollinearity problems were tested by computing variance inflation factors (VIFs) (Marquardt, 1970), which measure how strongly each predictor can be explained by the rest of the predictors and are based on the square of the multiple correlation coefficient (R^2) resulting from regressing the predictor variable against all other predictor variables (Naimi and Araújo, 2016). As a rule of thumb, a VIF of >10 signals that the model has a collinearity problem (Chatterjee and Hadi, 2006). We used a stepwise procedure, implemented through the 'sdm' package (Naimi and Araújo, 2016) in the R environment in order to remove all variables with VIFs >5, which was imposed as a precautionary threshold.

The variable importance was measured according to the permutation importance, which is calculated by randomly permuting training presence and pseudo-absence data in MaxEnt. In the absence of adequate variance partitioning methods for MaxEnt, permutation importance values were used as gross estimates of the variance explained by variables or groups of variables in SDM. It has to be noted, however, that permutation importance values are not strictly additive and do not account for shared variance between variables (Gallardo et al., 2015). The significance of differences between permutation importance of species with strict and wide habitat preference was evaluated by Wilcoxon's paired rank test.

3.3. Models' performance comparisons

Model performances were compared according to the 'threshold independent' test of the area under the curve (AUC), which is widely considered as unavoidable. The AUC has values that typically range from 0.5–1.0. Values close to 0.5 indicate a fit no better than that expected by random, while a value of 1.0 indicates a perfect fit (Baldwin, 2009). AUC values (trainAUC) of all replicate MaxEnt runs (as many as the number of occurrences we had), following the 'leave-one-out' approach, especially recommended for small sample sizes (Pearson et al., 2007), were firstly used. Few studies have had access to independent evaluation data, which are fundamental for proper model evaluation (Halvorsen, 2012). Following MaxEnt's results, two intensive field campaigns were conducted, in 2016 and 2017, when the entire study area was visited at least once. All the data obtained from the intensive field surveys were used for models' evaluations. Specifically, true records were included as test sample file in MaxEnt to calculate the AUC by the use of independent test data (testAUC). The minimum difference between trainAUC and testAUC (AUCdiff) (Warren and Seifert, 2011) was then calculated. As a role of thumb, low AUCdiff values are desirable because it means that the risk of overfitting models, due to a good performance with training data but poor with test data, is low (Warren and Seifert, 2011).

Models' comparisons were based on three main aspects: (i) number of input points, (ii) species traits and (iii) regularization multiplier settings. In order to test the discrimination ability among different species traits and distribution patterns, we also took into account the number of occurrences used for modelling each species. Therefore, we subdivided the species set into two categories: (i) strict habitat preference, for species living in less than two localized habitat types (e.g. cliffs, crevices, springs) or linear habitat types (e.g. riparian vegetation) and (ii) wide habitat preference, for species which were found in more than two localized or linear habitat types or are able to colonize habitats that predominate in the study area. We also tested the influence of species detectability by comparing model performances of the (i) highly detectable species (e.g. trees and scrubs) and (ii) low detectable species (e.g. ephemeral or cryptic species, species living in difficult-to-reach areas). The specific categorization of each species is reported in Appendix (Appendix A). A nonparametric two-group sample test (Wilcoxon signed-rank) was calculated to compare models' performances for species within the two categorizations according to their habitat preference and detectability. Significant differences among models performances with four different regularization multiplier values were tested with the Tukey honest significant difference method (TukeyHSD).

3.4. Results

A total of 280 model outputs (70 species $\times$ 4 regularization multiplier settings) were obtained from MaxEnt. Only six variables were retained, after testing for multicollinearity among predictors (Table 2). Occurrences used for modelling were limited (from 3 up to 19 points), for a total number of 346 records. After the two field campaigns, several new localities were georeferenced (1958 records) and used for the post-hoc ground validation through the testAUC.

In most of cases, the highest percent of contribution was recorded for bioclimatic variables, especially for Bio3 and Bio4 (Table 2). Significant differences (at $P < 0.05$) between species with strict and wide habitat preferences were only found for soil and slope.

Except for nine models with regularization value at 5 (Appendix B), all the rest of outputs scored, at least, trainAUCs higher than 0.7,

Table 2

Comparison among averages and standard deviation of permutation importance in MaxEnt models. All variables used at this stage were retained after VIF analysis, removing variables with VIFs of >5. Significant differences among the variable contribution of species with strict and wide habitat preference were tested by Wilcoxon's paired rank test.

Variable	Importance (general)	Importance (strict)	Importance (wide)	P
Mean Diurnal Range (Bio2)	3.34 (12.32)	0.9 (2.83)	6.01 (17.17)	n.s
Isothermality (Mean Diurnal Range/ Temperature Annual Range; Bio3)	19.62 (23.33)	19.18 (21.9)	20.09 (24.79)	n.s
Temperature Seasonality (Bio4)	8.23 (14.56)	5.80 (7.12)	10.9 (19.36)	n.s
Minimum Temperature of Coldest Month (Bio6)	56.97 (29.57)	57.69 (26.04)	56.19 (32.97)	n.s
Slope	7.64 (13.01)	11.19 (15.57)	3.78 (7.78)	0.04*
Soil	3.96 (12.50)	5.01 (14.33)	2.82 (9.98)	0.02*

which is a common threshold used to consider captured patterns far from random. Although significant correlation between trainAUC and testAUC was found, low testAUC resulted even in some cases of apparently high performances according to the trainAUC values (Fig. 3). Significant differences in model performances among regularization multipliers ($P < 0.01$) were found only in the case of trainAUC (Regularization = 5; Fig. 4).

A significant relationship was found between the number of input points and the testAUC performance (Fig. 5a), revealing that model performance decreases with the number of points. Such trend was however not confirmed by the AUCDiff (Fig. 5c). No significant differences were found between species with high or low detectability and strict or wide distribution in terms of model performances (Fig. 5b,c)

3.5. Discussion

Model reliability, measured in terms of trainAUC values, were, with some exception, generally high; nonetheless testAUCs measured with
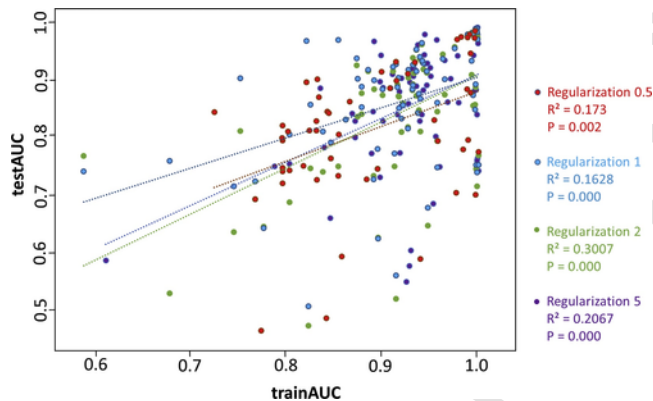

Fig. 3. Plot of the relationship between trainAUC and testAUC for models set with four regularization multiplier values (0.5, 1, 2, 5). The R^2 and statistical significances of the two simple regressions are reported.

independent data, were low even in the case of high performances of the trainAUC, confirming the cautionary notes of several authors on the use of AUC, especially when based on presence-only data (e.g. Hernandez et al., 2006; Lobo et al., 2008; Merow et al., 2013; Proosdij et al., 2016). In the case of trainAUCs, high values indicate that the model can distinguish between presences and potentially unsampled locations (background), which is not necessarily a relevant distinction. In addition, exceptionally high values of trainAUC for models based on small sample sizes of narrow-ranged species are generally considered as overestimations since the chance that a random presence has a higher probability of occurrence than a random absence for such species is high (Proosdij et al., 2016). For these reasons, trainAUCs may be used to identify SDMs that perform better than random expectations, but not for comparing different models generated through MaxEnt, particularly for small sample sizes (Warren and Seifert, 2011; Morales et al., 2017). Accordingly, our results confirms caution when using SDMs, since also when trainAUC values were particularly high, realized distribution was in some cases far from the predicted one (Fig. 3). Even if part of such drawbacks could be overcome by practices widely discussed (e.g. Halvorsen, 2012; Jiménez-Valverde et al., 2008a; Warren and Seifert, 2011; Morales et al., 2017), practitioners may always consider that predictions have always a degree of uncertainty and only very accurate field data may reduce at minimum SDM uncertainties. Such uncertainties could be considerably diminished through measuring model performances by real independent records (which is our case). Our testAUCs generally confirmed the usefulness of modelling techniques to guide field surveys of poorly known species, even in the case of a very low number of input points used or for species with a restricted distribution range. The partially decreasing performance with the number of input points is likely to be not in line the general role (e.g. Proosdij et al., 2016). Indeed it is generally considered that model performance decrease with sample size, even though there are many example of successful model applications with small samples (Pearson et al., 2007; McCune, 2016; Rus et al., 2017). We tried to explain this result by two assumptions: first, that species with initial small sample size were effectively present in further previously unknown localities and, second,
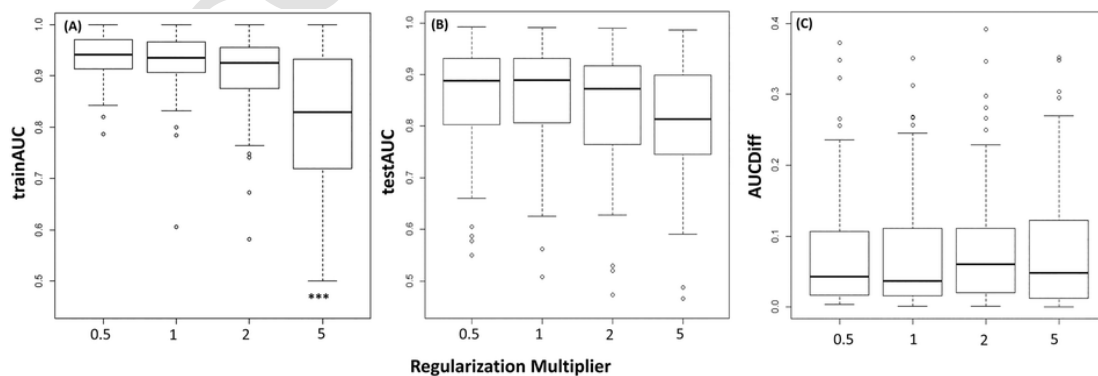

Fig. 4. Boxplot showing (A) trainAUC, (B) testAUC and (C) AUCDiff values among the four regularization multiplier settings Significant differences ($P < 0.001$) were found only for the trainAUCs of models with regularization multiplier at 5.
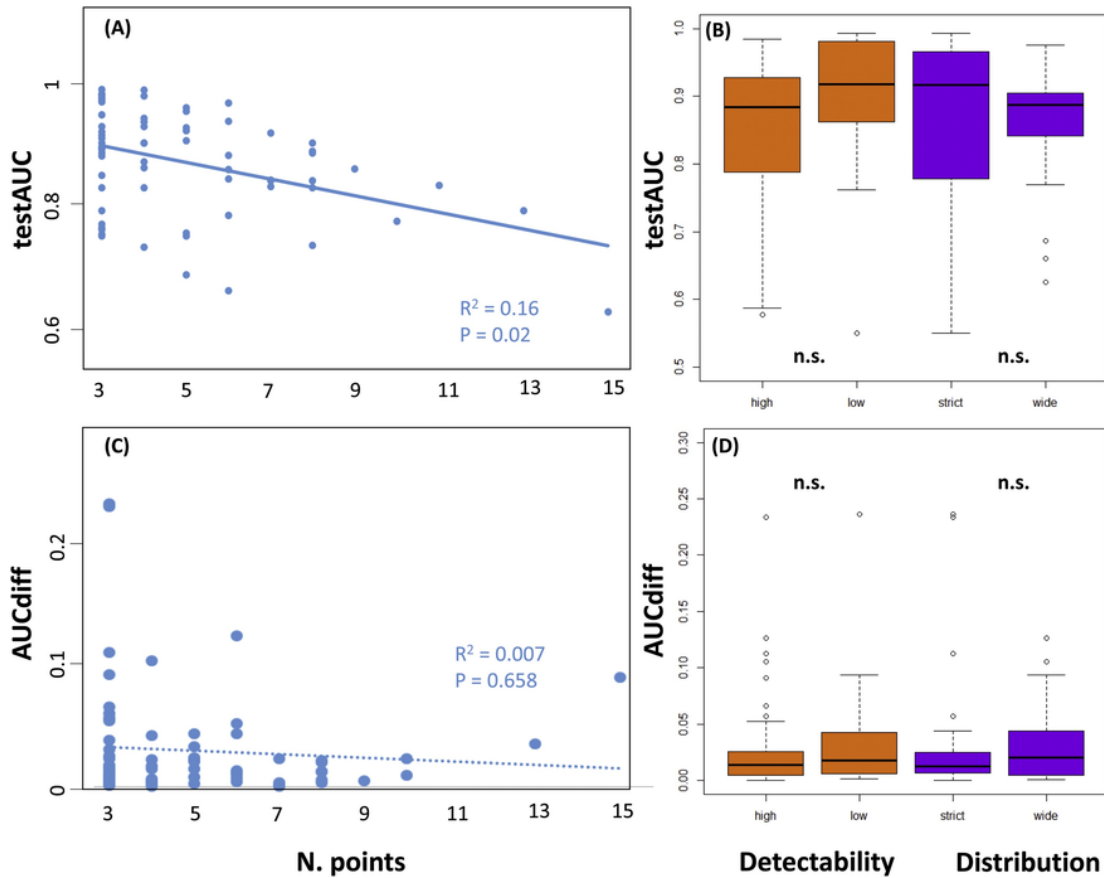
Fig. 5. Plot of the relationship between (A) testAUC and (C) AUCDiff for models with increasing number of input points. The R² and statistical significances of the two simple regressions are reported. On the right side, boxplots are showing (B) testAUC and (D) AUCDiff values between species of high and low detectability and between species with strict and wide habitat preferences; no significant differences (P > 0.01) between the two pairs were found.

that species with several known localities were likely to be potentially present in a great part of the study area, but no-ecological factors, such as human disturbance, were effectively limiting the real distribution.

Our results showed that regularization multiplier settings could affect model performances only in terms of trainAUCs, while no substantial differences were found at the time of measuring the performance with independent real data. Even when more complex models (i.e with lower regularization values) show higher performances in terms trainAUCs, such apparently higher ability was not confirmed by the post-model validation through the data obtained in field. According to many authors (Anderson and Gonzalez, 2011; Radosavljevic and Anderson, 2014; Moreno-Amat et al., 2015; Halvorsen et al., 2016), best regularization values can sensibly vary depending on many factors, such as sample size, study area extent and so, even this choice should be carefully pondered on a species-by-species basis.

In line with most of SDMs users (see Fig. 1 and most of references reported in Table 1), our example confirmed that bioclimatic variables are useful inputs for modelling species. Indeed, in most of the cases, the highest percent of contribution was recorded for such variables (Table 2). Significant differences (at P < 0.05) between species with strict and wide habitat preferences were only found for soil and slope (Table 2), suggesting that species with strict habitat preferences are more linked to these variables than other species. Differently, bioclimatic variables are generally determining the distribution of any kind of species. Nonetheless, it is important to highlight the high standard deviations for all variables. This suggests, once again, that general rules are tricky, while variable selection should be species-specifically carried out. Accordingly, starting from an as much as possible diverse set of variables (including bioclimatic but preferably also combined with

other kind of available information categories) could allow to find a representative set of environmental variables for each species.

4. General conclusions and future directions

Despite it is commonly considered one of main useful SDMs applications, our review highlighted that the use of SDMs to guide the search of unknown species populations was rarely experimented. This is particular evident in under-investigated countries, even where their biodiversity heritage and the need of increasing data are widely recognized (e.g. Kier et al., 2005; Grenyer et al., 2006). We may then consider that field surveys are in some cases not feasible due to lack of resources or because some areas are less safe or less easy to reach than others. Secondly, even if the Web is supporting the emergence of new forms of scientific interaction and engagement, specific training opportunities on modelling issues are uncommon in many countries. These are the first reasons of why we would like to encourage even early stage practitioners to experiment this application with data from their local study sites. Furthermore, among the variability in terms of organism and modelling characteristics, the experimental use of SDMs in this field was positively rated by the majority of authors, even if a certain degree of false presences was always depicted. This degree of uncertainty is an implicit characteristic of all predictions, and model predictions should be always considered as a suggestion/proposal of distribution mechanisms and always followed by expert-based interpretations.

Our results adhere to the calls of several studies that already include recommendations how to use and set up MaxEnt or other SDM techniques (e.g., Elith et al., 2006; Warren and Seifert, 2011; Merow et al., 2013; Radosavljevic and Anderson, 2014; Morales et al., 2017). In

particular, we confirm the caution expressed by other authors in considering consequences of considering the AUC alone as a performance measure. Due to the complexity of all the raised questions, which were mainly never discussed before for this specific application, it was unfeasible to deeply investigate all issues in one research paper. For this reason, only one shaft of light was provided by the experimental study of Regional Park of Montarbu and further investigations are required to improve the applicability of this approach.

This said, practitioners are once again encouraged to test the ability of SDMs in helping their next species searches in field, with their own data, even if limited. Indeed, we demonstrated that even with small and possibly biased initial samples, reliable results could be obtained. Without underrating the importance of relevant recommendations, we would like to highlight that eventual biases due to lack of data, experience or staff, have in this experimental application less irreparable consequences than others, such as conservation assessments based on future projections, which cannot be adjusted by explicit data from ground validation.

In conclusion, our take-home keyword is 'local': local are such kind of experimental SDMs applications and local should be the practitioners, since the knowledge of the territory and the willingness needed to hold up the field trip efforts are, in this particular case, as important as modelling skills. Furthermore, local experts may facilitate the often complex and artificial data filtering of species localities retrieved from the Web or other sources by their better knowledge on local species or by implementing small-scale pilot surveys to obtain accurate locality information and refine sampling protocols (Peterman et al., 2013). On the other side, the global expert community may strengthen cooperation for the interchange of knowledge and funding for increasing the quality and quantity of globally available environmental data, and for strengthening research capacity by training local researchers.

Acknowledgements

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ecolmodel.2018.07.018.

References

Aizpurua, O., Cantú-Salazar, L., San Martin, G., Biver, G., Brotons, L., Titeux, N., 2015. Reconciling expert judgement and habitat suitability models as tools for guiding sampling of threatened species. J. Appl. Ecol. 52, 1608–1616.

Anderson, R.P., Gonzalez, I.J., 2011. Species-specific tuning increases robustness to sampling bias in models of species distributions: an implementation with Maxent. Ecol. Model. 222, 2796–2811.

Araújo, M.B., New, M., 2007. Ensemble forecasting of species distributions. Trends Ecol. Evol. 22, 42–47.

Arrigoni, P.V., 1965. Ricerche geobotaniche su Linaria muelleri Moris e notizie su Hypericum aegyptiacum L., nuovo reperto per la flora sarda. Webbia 20, 307–330.

Bacchetta, G., Cannas, M., Peruzzi, L., 2014. A new diploid butterwort species (Pinguicula, Lentibulariaceae) from Sardinia. Phytotaxa 186, 279–286.

Baldwin, R.A., 2009. Use of maximum entropy modeling in wildlife research. Entropy 11, 854–866.

Bloom, T.D., Flower, A., DeChaine, E.G., 2018. Why georeferencing matters: introducing a practical protocol to prepare species occurrence records for spatial analysis. Ecol. Evol. 8, 765–777.

Boetsch, J.R., van Manen, F.T., Clark, J.D., 2003. Predicting rare plant occurrence in Great Smoky Mountains National Park, USA. Nat. Areas J. 23, 229–237.

Bosso, L., Mucedda, M., Fichera, G., Kiefer, A., Russo, D., 2016. A gap analysis for threatened bat populations on Sardinia. Hystrix 27, 212–214.

Bosso, L., Smeraldo, S., Rapuzzi, P., Sama, G., Garonna, A.P., Russo, D., 2018. Nature protection areas of Europe are insufficient to preserve the threatened beetle Rosalia

alpina (Coleoptera: Cerambycidae): evidence from species distribution models and conservation gap analysis. Ecol. Entomol. 43, 192–203.

Botero-Delgadillo, E., Páez, C.A., Bayly, N., 2012. Biogeography and conservation of Andean and Trans-Andean populations of Pyrrhura parakeets in Colombia: modelling geographic distributions to identify independent conservation units. Bird Conserv. Int. 22, 445–461.

Bourg, N.A., McShea, W.J., Gill, D.E., 2005. Putting a CART before the search: successful habitat prediction for a rare forest herb. Ecology 86, 2793–2804.

Bradie, J., Leung, B., 2017. A quantitative synthesis of the importance of variables used in MaxEnt species distribution models. J. Biogeogr. 44, 1344–1361.

Bucklin, D.N., Basile, M., Benscoter, A.M., Brandt, L.A., Mazzotti, F.J., Romañach, S.S., Speroterra, C., Watling, J.I., 2015. Comparing species distribution models constructed with different subsets of environmental predictors. Divers. Distrib. 21, 23–35.

Calvo, J., Aedo, C., 2015. A taxonomic revision of the Eurasian/Northwestern African Senecio doria Group (Compositae). Syst. Bot. 40, 900–913.

Chatterjee, S., Hadi, A.S., 2006. Simple linear regression. Regression Analysis by Example, fourth edition John Wiley & Sons, Inc., Hoboken.

Chen, Y., Zhang, J., Jiang, J., Nielsen, S.E., He, F., 2017. Assessing the effectiveness of China's protected areas to conserve current and future amphibian diversity. Divers. Distrib. 23, 146–157.

Costa, G.C., Nogueira, C., Machado, R.B., Colli, G.R., 2010. Sampling bias and the use of ecological niche modeling in conservation planning: a field evaluation in a biodiversity hotspot. Biodivers. Conserv. 19, 883–899.

De Siqueira, M.F., Durigan, G., de Marco Júnior, P., Peterson, A.T., 2009. Something from nothing: using landscape similarity and ecological niche modeling to find rare plant species. J. Nat. Conserv. 17, 25–32.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., García Marquéz, J.R., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36, 27–46.

Edvardsen, A., Bakkestuen, V., Halvorsen, R., 2011. A fine-grained spatial prediction model for the red-listed vascular plant Scorzonera humilis. Nord. J. Bot. 29, 495–504.

Edwards, T.C., Cutler, D.R., Zimmermann, N.E., Geiser, L., Alegria, J., 2005. Model-based stratifications for enhancing the detection of rare ecological events. Ecology 86, 1081–1090.

Elith, J., Graham, C.H., 2009. Do they? How do they? Why do they differ? On finding reasons for differing performances of species distribution models. Ecography 32, 66–77.

Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40, 677–697.

Elith, J., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijimans, R., Huettmann, F., Leatwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., et al., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129–151.

Engler, R., Guisan, A., Rechsteiner, L., 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. J. Appl. Ecol. 41, 263–274.

Feria, A., Townsend Peterson, A., 2002. Prediction of bird community composition based on point-occurrence data and inferential algorithms: a valuable tool in biodiversity assessments. Divers. Distrib. 8, 49–56.

Fois, M., Fenu, G., Cuena-Lombraña, C., Cogoni, D., Bacchetta, G., 2015. A practical method to speed up the discovery of unknown populations using Species Distribution Models. J. Nat. Conserv. 24, 42–48.

Fois, M., Cuena-Lombraña, A., Fenu, G., Cogoni, D., Bacchetta, G., 2016. The reliability of conservation status assessments at regional level: past, present and future perspectives on Gentiana lutea L. ssp. lutea in Sardinia. J. Nat. Conserv. 33, 1–9.

Fois, M., Fenu, G., Bacchetta, G., 2018. Identifying and assessing the efficiency of different networks of a fine-scale hierarchy of biodiversity hotspots. Plant Ecol. Divers. 11, 121–132.

Fourcade, Y., Besnard, A.G., Secondi, J., 2018. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. Glob. Ecol. Biogeogr. 27, 245–256.

Gallardo, B., Zieritz, A., Aldridge, D.C., 2015. The importance of the human footprint in shaping the global distribution of terrestrial, freshwater and marine invaders. PLoS One 10, e0125801.

Greaves, G.J., Mathieu, R., Seddon, P.J., 2006. Predictive modelling and ground validation of the spatial distribution of the New Zealand long-tailed bat (Chalinolobus tuberculatus). Biol. Conserv. 132, 211–221.

Grenyer, R., Orme, C.D.L., Jackson, S.F., Thomas, G.H., Davies, R., Davies, T.J., Jones, K.E., Olson, V.A., Ridgely, R.S., Rasmussen, P.C., et al., 2006. Global distribution and conservation of rare and threatened vertebrates. Nature 444, 93–96.

Groff, L.A., Marks, S.B., Hayes, M.P., 2014. Using ecological niche models to direct rare amphibian surveys: a case study using the Oregon Spotted Frog (Rana pretiosa). Herpetol. Conserv. Biol. 9, 354–368.

Guillera-Arroita, G., 2017. Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. Ecography 40, 281–295.

Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecol. Lett. 8, 993–1009.

Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N.G., Lehmann, A., Zimmermann, N.E., 2006. Using niche-based models to improve the sampling of rare species. Conserv. Biol. 20, 501–511.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135, 147–186.

Halvorsen, R., 2012. A gradient analytic perspective on distribution modelling. Sommerfeltia 35, 1–165.

Halvorsen, R., 2013. A strict maximum likelihood explanation of MaxEnt, and some implications for distribution modelling. Sommerfeltia 36, 1–132.

Halvorsen, R., Mazzoni, S., Bryn, A., Bakkestuen, V., 2015. Opportunities for improved distribution modelling practice via a strict maximum likelihood interpretation of MaxEnt. Ecography 38, 172–183.

Halvorsen, R., Mazzoni, S., Dirksen, J.W., Næsset, E., Gobakken, T., Ohlson, M., 2016. How important are choice of model selection method and spatial autocorrelation of presence data for distribution modelling by MaxEnt. Ecol. Model. 328, 108–118.

Hengl, T., de Jesus, J.M., Heuvelink, G.B., Gonzalez, M.R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., et al., 2017. SoilGrids250m: global gridded soil information based on machine learning. PLoS One 12, e0169748.

Hernandez, P.A., Graham, C.H., Master, L.L., Albert, D.L., 2006. The effect of sample size and species characteristics on performance of different species distribution modelling methods. Ecography 29, 773–785.

Hijmans, R.J., 2014. Raster: Geographic Data Analysis and Modeling. R package ver 2.2–12.

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. Int. J. Climatol. 25, 1965–1978.

Hirzel, A.H., Hausser, J., Chessel, D., Perrin, N., 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data?. Ecology 83, 2027–2036.

Jarvis, A., Williams, K., Williams, D., Guarino, L., Caballero, P.J., Mottram, G., 2005. Use of GIS for optimizing a collecting mission for a rare wild pepper (Capsicum flexuosum Sendtn.) in Paraguay. Genet. Resour. Crop Evol. 52, 671–682.

Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E., 2008. Hole-Filled SRTM for the Globe Version 4. Available from the CGIAR-CSI SRTM 90m Database http://srtm.csi.cgiar.org.

Jiménez-Valverde, A., Lobo, J.M., Hortal, J., 2008. Not as good as they seem: the importance of concepts in species distribution modelling. Divers. Distrib. 14, 885–890.

Jiménez-Valverde, A., Gómez, J.F., Lobo, J.M., Baselga, A., Hortal, J., 2008. Challenging species distribution models: the case of Maculinea nausithous in the Iberian Peninsula. Ann. Zool. Fenn. 45, 200–210.

Kaky, E., Gilbert, F., 2016. Using species distribution models to assess the importance of Egypt's protected areas for the conservation of medicinal plants. J. Arid Environ. 135, 140–146.

Kier, G., Mutke, J., Dinerstein, E., Ricketts, T.H., Küper, W., Kreft, H., Barthlott, W., 2005. Global patterns of plant diversity and floristic knowledge. J. Biogeogr. 32, 1107–1116.

Le Lay, G., Engler, R., Franc, E., Guisan, A., 2010. Prospective sampling based on model ensembles improves the detection of rare species. Ecography 33, 1015–1027.

Lobo, J.M., Jiménez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. Glob. Ecol. Biogeogr. 17, 145–151.

Loi, M.C., Lai, A., 2001. The flora of Mount Tonneri and Mount Arquerì: Mesozoic calcareous outcrops of Central-Eastern Sardinia. Fl. Medit. 11, 385–418.

Loi, M.C., Marras, G., Maxia, A., 2004. The flora of Monte Perda 'e Liana (CE-Sardinia). Fl. Medit. 14, 153–172.

Lomba, A., Pellissier, L., Randin, C., Vicente, J., Moreira, F., Honrado, J., Guisan, A., 2010. Overcoming the rare species modelling paradox: a novel hierarchical framework applied to an Iberian endemic plant. Biol. Conserv. 143, 2647–2657.

Mammola, S., Goodacre, S.L., Isaia, M., 2018. Climate change may drive cave spiders to extinction. Ecography 41, 233–243.

Marquardt, D.W., 1970. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. Technometrics 12, 591–612.

McCune, J.L., 2016. Species distribution models predict rare species occurrences despite significant effects of landscape context. J. Appl. Ecol. 53, 1871–1879.

McPherson, J., Jetz, W., 2007. Effects of species' ecology on the accuracy of distribution models. Ecography 30, 135–151.

Merow, C., Smith, M.J., Silander, J.A., 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography 36, 1058–1069.

Metz, M., Rocchini, D., Neteler, M., 2014. Surface temperatures at the continental scale: tracking changes with remote sensing at unprecedented detail. Remote Sens. 6, 3822–3840.

Montemayor, S.I., Melo, M.C., Scattolini, M.C., Pocco, M.E., del Río, M.G., Dellapé, G., 2017. The fate of endemic insects of the Andean region under the effect of global warming. PLoS One 12, e0186655.

Morales, N.S., Fernández, I.C., Baca-González, V., 2017. MaxEnt's parameter configuration and small samples: are we paying attention to recommendations? A systematic review. PeerJ 5, e3093.

Moreno-Amat, E., Mateo, R.G., Nieto-Lugilde, D., Morueta-Holme, N., Svenning, J.C., García-Amorena, I., 2015. Impact of model complexity on cross-temporal transferability in Maxent species distribution models: an assessment using paleobotanical data. Ecol. Model. 312, 308–317.

Naimi, B., Araújo, M.B., 2016. Sdm: a reproducible and extensible R platform for species distribution modelling. Ecography 39, 368–375.

Newbold, T., 2010. Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. Prog. Phys. Geogr. 34, 3–22.

Ortega-Huerta, M.A., Vega-Rivera, J.H., 2017. Validating distribution models for twelve endemic bird species of tropical dry forest in western Mexico. Ecol. Evol. 7, 7672–7686.

Pearce, J., Ferrier, S., Scotts, D., 2001. An evaluation of the predictive performance of distributional models for flora and fauna in north-east New South Wales. J. Environ. Manage. 62, 171–184.

Pearson, R.G., Dawson, T.P., 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful?. Glob. Ecol. Biogeogr. 12, 361–371.

Pearson, R.G., Raxworthy, C.J., Nakamura, M., Townsend Peterson, A., 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. J. Biogeogr. 34, 102–117.

Peterman, W.E., Crawford, J.A., Kuhns, A.R., 2013. Using species distribution and occupancy modeling to guide survey efforts and assess species status. J. Nat. Conserv. 21, 114–121.

Peterson, A.T., Nakazawa, Y., 2008. Environmental data sets matter in ecological niche modelling: an example with Solenopsis invicta and Solenopsis richteri. Glob. Ecol. Biogeogr. 17, 135–144.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190, 231–259.

Phillips, S.J., Anderson, R.P., Dudík, M., Schapire, R.E., Blair, M.E., 2017. Opening the black box: an open-source release of Maxent. Ecography 40, 887–893.

Proosdij, A.S., Sosef, M.S., Wieringa, J.J., Raes, N., 2016. Minimum required number of specimen records to develop accurate species distribution models. Ecography 39, 542–552.

Pulliam, H.R., Babbitt, B., 1997. Science and the protection of endangered species. Science 275, 499–500.

Radosavljevic, A., Anderson, R.P., 2014. Making better Maxent models of species distributions: complexity, overfitting and evaluation. J. Biogeogr. 41, 629–643.

Raxworthy, C.J., Martinez-Meyer, E., Horning, N., Nussbaum, R.A., Schneider, G.E., Ortega-Huerta, M.A., Peterson, A.T., 2003. Predicting distributions of known and unknown reptile species in Madagascar. Nature 426, 837–841.

Rebelo, H., Jones, G., 2010. Ground validation of presence-only modelling with rare species: a case study on barbastelles Barbastella barbastellus (Chiroptera: Vespertilionidae). J. Appl. Ecol. 47, 410–420.

Rhoden, C.M., Peterman, W.E., Taylor, C.A., 2017. Maxent-directed field surveys identify new populations of narrowly endemic habitat specialists. PeerJ 5, e3632.

Rinnhofer, L.J., Roura-Pascual, N., Arthofer, W., Dejaco, T., Thaler-Knoflach, B., Wachter, G.A., Erhard, C., Steiner, F.M., Schlick-Steiner, B.C., 2012. Iterative species distribution modelling and ground validation in endemism research: an Alpine jumping bristletail example. Biodivers. Conserv. 21, 2845–2863.

Rus et al., 2017 Rus J.D., Ramírez-Rodríguez R., Amich F., Melendo-Luque M., Habitat distribution modelling, under the present climatic scenario, of the threatened endemic Iberian Delphinium fissum subsp. sordidum (Ranunculaceae) and implications for its conservation, Plant Biosyst. 152 (2017) 891–900.

Särkinen, T., Gonzáles, P., Knapp, S., 2013. Distribution models and species discovery: the story of a new Solanum species from the Peruvian Andes. PhytoKeys 31, 1–20.

Searcy, C.A., Shaffer, H.B., 2014. Field validation supports novel niche modeling strategies in a cryptic endangered amphibian. Ecography 37, 983–992.

Stockwell, D., 1999. The GARP modelling system: problems and solutions to automated spatial prediction. Int. J. Geogr. Inf. Sci. 13, 143–158.

Varela, S., Anderson, R.P., García-Valdés, R., Fernández-González, F., 2014. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. Ecography 37, 1084–1091.

Veloz, S.D., 2009. Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. J. Biogeogr. 36, 2290–2299.

Verovnik, R., Popović, M., Šašić, M., Cuvelier, S., Maes, D., 2014. Wanted! Dead or alive: the tale of the Brown's Grayling (Pseudochazara amymone). J. Insect Conserv. 18, 675–682.

Vessella, F., López-Tirado, J., Simeone, M.C., Schirone, B., Hidalgo, P.J., 2017. A tree species range in the face of climate change: cork oak as a study case for the Mediterranean biome. Eur. J. For. Res. 136, 555–569.

Virgili, A., Racine, M., Authier, M., Monestiez, P., Ridoux, V., 2017. Comparison of habitat models for scarcely detected species. Ecol. Model. 346, 88–98.

Volis, S., 2016. Species-targeted plant conservation: time for conceptual integration. Isr. J. Plant Sci. 63, 232–249.

Warren, D.L., Seifert, S.N., 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. Ecol. Appl. 21, 335–342.

Warren, D.L., Glor, R.E., Turelli, M., 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. Evolution 62, 2868–2883.

Williams, J.N., Seo, C., Thorne, J., Nelson, J.K., Erwin, S., O'Brien, J.M., Schwartz, M.W., 2009. Using species distribution models to predict new occurrences for rare plants. Divers. Distrib. 15, 565–576.