



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

Mauro Fois<sup>a,\*</sup>, Alba Cuena-Lombrana<sup>b,c</sup>, Giuseppe Fenu<sup>a</sup>, Gianluigi Bacchetta<sup>a,c</sup>

<sup>a</sup> 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

<sup>b</sup> Dipartimento di Biologia Ambientale, 'Sapienza' Università di Roma, P.le A. Moro 5, 00185 Roma, Italy

<sup>c</sup> Hortus Botanicus Karalitanus (HBK), Università degli Studi di Cagliari, Viale Sant'Ignazio da Laconi, 9–11, Cagliari 09123, Italy

## ARTICLE INFO

### Keywords:

Ground validation  
MaxEnt  
Mediterranean flora  
Independent presence-absence data  
Plant distribution patterns  
Regularization multiplier

## 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 a 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 biological groups (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 endangered and

\* Corresponding author.

E-mail address: [mfois@unica.it](mailto:mfois@unica.it) (M. Fois).

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, 2007; 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; Prosdij 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-Aroita, 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 prediction 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 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<sup>2</sup> to 3 × 10<sup>6</sup> km<sup>2</sup>; Table 1). Most of the 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 frequently 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 efforts were generally destined for searching the species in field rather

Download English Version:

<https://daneshyari.com/en/article/8845959>

Download Persian Version:

<https://daneshyari.com/article/8845959>

[Daneshyari.com](https://daneshyari.com)