



# Incorporating plant fossil data into species distribution models is not straightforward: Pitfalls and possible solutions



Elena Moreno-Amat <sup>a,\*</sup>, Juan Manuel Rubiales <sup>a,b</sup>, César Morales-Molino <sup>a,c</sup>,  
Ignacio García-Amorena <sup>a</sup>

<sup>a</sup> Departamento de Sistemas y Recursos Naturales, Universidad Politécnica de Madrid, Ciudad Universitaria s/n, 28040 Madrid, Spain

<sup>b</sup> Departamento de Biología Vegetal I, Facultad de Ciencias Biológicas, Universidad Complutense de Madrid, C/ José Antonio Novais, 12, 28040 Madrid, Spain

<sup>c</sup> Institute of Plant Sciences and Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland

## ARTICLE INFO

### Article history:

Received 15 December 2016

Received in revised form

21 June 2017

Accepted 22 June 2017

### Keywords:

Global  
Paleoclimate modelling  
Quaternary  
Data analysis  
Fossil pollen  
Hindcasting  
Model calibration  
Model validation  
Palaeo-database  
Palaeo-species distribution models

## ABSTRACT

The increasing development of species distribution models (SDMs) using palaeodata has created new prospects to address questions of evolution, ecology and biogeography from wider perspectives. Palaeobotanical data provide information on the past distribution of taxa at a given time and place and its incorporation on modelling has contributed to advancing the SDM field. This has allowed, for example, to calibrate models under past climate conditions or to validate projected models calibrated on current species distributions. However, these data also bear certain shortcomings when used in SDMs that may hinder the resulting ecological outcomes and eventually lead to misleading conclusions. Palaeodata may not be equivalent to present data, but instead frequently exhibit limitations and biases regarding species representation, taxonomy and chronological control, and their inclusion in SDMs should be carefully assessed. The limitations of palaeobotanical data applied to SDM studies are infrequently discussed and often neglected in the modelling literature; thus, we argue for the more careful selection and control of these data. We encourage authors to use palaeobotanical data in their SDM studies and for doing so, we propose some recommendations to improve the robustness, reliability and significance of palaeo-SDM analyses.

© 2017 Elsevier Ltd. All rights reserved.

## 1. Introduction

The development of global databases has increased the availability of biodiversity data at the planetary scale and triggered rapid progress in the use of biodiversity informatics to resolve a wide spectrum of macroecological, biogeographical and evolutionary questions (Elith and Leathwick, 2009; Peterson et al., 2010). Following this trend, the rapid and steady increase in accessible fossil data is broadening our knowledge on the functioning of ecosystems over long timescales (Willis et al., 2010). The palaeorecord provides a unique and large body of data documenting the past occurrence of species and communities at decadal to millennial timescales that is crucial to extend in time our understanding of the response of biodiversity to environmental changes.

Within the emerging discipline of palaeoecoinformatics, i.e. the

coupled use of palaeodata and computing tools (Brewer et al., 2012), correlative species distribution models (SDMs) (i.e. algorithms that relate species occurrences to environmental and geographical predictors; Guisan and Zimmermann, 2000) have attracted most attention (Maguire et al., 2015). Specifically, SDMs have been extensively used to study the impacts of climate change (e.g. Thuiller et al., 2005) and invasive species on biodiversity (e.g. Ficetola et al., 2007), to guide nature conservation (e.g. Anderson and Martínez-Meyer, 2004) and decision making (e.g. Schwartz, 2012), among others. SDMs have also been applied to locate potential migratory routes in the past (Waltari and Guralnick, 2009), study past range dynamics and taxa responses to climate change (Nogués-Bravo et al., 2008), integrate the spatial distribution of species and genetic diversity (Espíndola et al., 2012), explore species extinctions (Lorenzen et al., 2011), test species niche conservatism over time (Stigall, 2012), examine species diversification or speciation mechanisms (Peterson and Nyári, 2008), correlate species richness and climatic change (Sandel et al., 2011) and identify habitable areas for humans during the Last Glacial Maximum

\* Corresponding author.

E-mail address: [emorenoamat@gmail.com](mailto:emorenoamat@gmail.com) (E. Moreno-Amat).

(Banks et al., 2008). The predictive accuracy of SDMs is generally defined as the ability of the model to correctly classify the presence and absence of species. To assess their performance, model predictions can be validated using a subset of the original dataset not overlapping with the training subset, using an independent dataset from the one used in model fitting or using data from different regions (Araújo et al., 2005). However, the reliability of SDM predictions at millennial scales cannot be demonstrated unless fossil records are used to verify the model predictions (Martínez-Meyer et al., 2004; Rodríguez-Sánchez et al., 2010) and palaeoecological data have recently been applied to assess the robustness of SDM predictions of past time periods (e.g. Roberts and Hamann, 2012a, b).

A number of reviews have discussed the achievements, limitations and prospects related to the use of SDMs to address phylogeographical, biogeographical, and macroecological research questions over long timescales. First, Nogués-Bravo (2009) noted the theoretical assumptions underlying niche modelling that may influence the resulting inferences and provided suggestions to improve hindcasting. Then, both Varela et al. (2011) and Svenning et al. (2011) focused on methodological and conceptual issues (data, predictors and modelling algorithms) of palaeo-SDMs coupled with palaeobiological data. Later, Collevatti et al. (2013) explored the uncertainties derived from atmosphere-ocean global circulation models, species and their interactions when predicting palaeodistributions, proposing a framework coupling palaeoclimatic distributions, the fossil record and phylogeographical studies to partially overcome them. More recently, Franklin et al. (2015) underscored the potential usefulness of hindcasting the past distribution of edible plants and game animals to reconstruct resource palaeolandscapes that in turn improve the understanding of human prehistory. Finally, Maguire et al. (2015) highlighted the importance of palaeodata to test the predictive ability of SDMs and their assumptions with an emphasis on the potential of community-level models.

In this paper, we focus on the use of palaeobotanical data in SDMs because this type of data is well suited to exploring the limitations inherent in the use of fossil data in SDMs and analysing the differences with modern datasets. In addition to the temporal and spatial uncertainties (Brewer et al., 2012), palaeobotanical data share certain features with other palaeobiological records whose limitations have been previously discussed by Varela et al. (2011). Nevertheless, additional factors that may hinder the validity and reliability of SDM interpretations must be considered, including the processing of palaeobotanical information from different sources, which likely constitutes the main obstacle to successfully using these data in SDMs.

Here, we discuss the constraints on the use of palaeobotanical data in SDM studies by focusing on problems observed in published papers. We highlight the assumptions and limitations related to the use of palaeobotanical data in SDM studies and we assess the use of this type of data in the SDM field up to the present. Finally, we provide suggestions to improve the use of palaeodata for modelling purposes.

## 2. Palaeobotanical constraints for SDMs

Palaeobotanical data included in SDM analyses are generally used as a direct proxy for past species occurrences, which involves a number of underlying assumptions that do not always hold true, such as the lack of correspondence between fossil assemblages and the species spectra and abundance of the original living assemblages (Behrensmeyer et al., 2000). The size and resistance to decay of the remains and the suitability of sedimentary deposits for preserving organic material (i.e. temperature, pH, and lack of

oxygen) constrains the fossilization process. Hence, the fossil record is biased towards certain taxa, time periods and sites that hosted favourable conditions for fossil preservation (Behrensmeyer et al., 2000; Varela et al., 2011).

Additionally, the taxonomic resolution reached in the analysis (usually lower in pollen than in macrofossil assemblages) and taphonomic processes (including production, transport, and deposition) limit the correspondence between fossil assemblages and past vegetation (Goring et al., 2013), thus limiting the usefulness of fossil data. Moreover, the chronology of the fossil record must be accurate and precise for the data to be fully valuable and to allow for comparisons with other independent data, and these conditions also apply when fossil records are correlated with modelled past climates. These climatic models are simulated for particular time periods, and SDMs consequently require palaeobotanical data within the same time period. A suitable chronology consists of a sufficient number of absolute dates, robust age estimates for control points, and appropriate age-depth models when using fossil data in SDMs.

SDM users should be aware of the limitations of the different palaeobotanical proxies with respect to the data's representation, taxonomy, chronology, data density, data source and spatial distribution to improve the ecological interpretations of the analysis.

### 2.1. Species representation

Fossil pollen is the most important source of data in articles that use palaeobotanical information in SDMs (see Table 1). Pollen is continuously deposited in sedimentary environments, and constant taphonomic regimes allow for sequential reconstructions of vegetation over time (Fægri and Iversen, 1950). Terrestrial pollen records may represent vegetation over a variable area (100–10<sup>4</sup> km<sup>2</sup>) depending on site type, size and topographical location; therefore, collected information should be considered at local to regional scales (Conedera et al., 2006; Hellman et al., 2008). Moreover, the representation of different taxa in pollen records may be biased by differences in pollen production, dispersion, deposition, preservation and manipulation (Birks and Birks, 2000; Goring et al., 2013). In general terms, entomophilous and ambophilous taxa are underrepresented (e.g. *Acer* L. and *Aesculus* L.) compared with anemophilous taxa (e.g. *Pinus* L., *Betula* L., *Fagus* L.) (Prentice, 1988) because of their relatively low pollen production. Additionally, non-anemophilous taxa generally produce pollen grains that present reduced dispersal ability compared with anemophilous taxa as a result of the size, weight and morphology of the pollen grain (Erdtman, 1969). For instance, the differences in the representation of *Tilia* L. and *Quercus* L. in lake sediments are explained by the poor dispersal ability of the relatively heavy and large pollen grains of *Tilia* as well by its lower pollen production (Prentice, 1988). Important dissimilarities in dispersal capacity are observed even within anemophilous taxa, such as between *Abies alba* Mill. and other co-occurring tree species like such as *Fagus sylvatica* L. and *Pinus sylvestris* L. (Poska and Pidek, 2010), and within taxa among different localities (Pidek et al., 2013).

Differences in pollen morphology affect pollen transportation distances. Bisaccate pollen grains, such as *Pinus*, can be transported long distances by the wind (Erdtman, 1969), although certain exceptions occur such as *Abies* Mill., whose large pollen size limits long-distance dispersal and therefore its finding in the fossil record indicates proximity to the pollen source (Erdtman, 1969). Finally, there may be additional biases in pollen analysis associated with laboratory treatments, particularly for large-sized pollen types (e.g. *Abies*; 150–160 μm; Beug, 2004) that may not be retained during the sieving process if the mesh size is too small such as the mesh recommended in several reference protocols for pollen sample

Download English Version:

<https://daneshyari.com/en/article/5786564>

Download Persian Version:

<https://daneshyari.com/article/5786564>

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