



Impact of model complexity on cross-temporal transferability in Maxent species distribution models: An assessment using paleobotanical data



Elena Moreno-Amat^{a,*}, Rubén G. Mateo^{a,b}, Diego Nieto-Lugilde^{c,d},
Naia Morueta-Holme^{d,e}, Jens-Christian Svenning^d, Ignacio García-Amorena^a

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

^b Department of Ecology & Evolution, University of Lausanne, Biophore Building, 1015 Lausanne, Switzerland

^c Departamento de Botánica, Facultad de Ciencias, Universidad de Granada, 18071 Granada, Spain

^d Section for Ecoinformatics and Biodiversity, Department of Bioscience, Aarhus University, Ny Munkegade 114, DK-8000 Aarhus C, Denmark

^e Department of Integrative Biology, University of California–Berkeley, 3040 VLSB, Berkeley, CA 94720, USA

ARTICLE INFO

Article history:

Received 6 February 2015

Received in revised form 30 May 2015

Accepted 31 May 2015

Available online 22 June 2015

Keywords:

Pollen fossil

Corylus avellana

Alnus glutinosa

Model validation

Species distribution model

β -Multiplier

ABSTRACT

Maximum entropy modeling (Maxent) is a widely used algorithm for predicting species distributions across space and time. Properly assessing the uncertainty in such predictions is non-trivial and requires validation with independent datasets. Notably, model complexity (number of model parameters) remains a major concern in relation to overfitting and, hence, transferability of Maxent models. An emerging approach is to validate the cross-temporal transferability of model predictions using paleoecological data. In this study, we assess the effect of model complexity on the performance of Maxent projections across time using two European plant species (*Alnus glutinosa* (L.) Gaertn. and *Corylus avellana* L.) with an extensive late Quaternary fossil record in Spain as a study case. We fit 110 models with different levels of complexity under present time and tested model performance using AUC (area under the receiver operating characteristic curve) and AICc (corrected Akaike Information Criterion) through the standard procedure of randomly partitioning current occurrence data. We then compared these results to an independent validation by projecting the models to mid-Holocene (6000 years before present) climatic conditions in Spain to assess their ability to predict fossil pollen presence–absence and abundance. We find that calibrating Maxent models with default settings result in the generation of overly complex models. While model performance increased with model complexity when predicting current distributions, it was higher with intermediate complexity when predicting mid-Holocene distributions. Hence, models of intermediate complexity resulted in the best trade-off to predict species distributions across time. Reliable temporal model transferability is especially relevant for forecasting species distributions under future climate change. Consequently, species-specific model tuning should be used to find the best modeling settings to control for complexity, notably with paleoecological data to independently validate model projections. For cross-temporal projections of species distributions for which paleoecological data is not available, models of intermediate complexity should be selected.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Species distribution models (SDMs) are widely used algorithms for describing ecological patterns and estimating the ecological niche or the potential distribution of species across space and time (Elith and Leathwick, 2009). Mechanistic SDMs use functional traits

and physiological constraints to predict the response of an individual (or a population) to environmental conditions (Kearney et al., 2010; Morin and Thuiller, 2009), whilst correlative SDMs relate species known occurrences and environmental variables (Guisan and Zimmermann, 2000). While mechanistic models may provide a more reliable and realistic picture because they can include processes that limit species distribution (Kearney et al., 2010), correlative models are the most frequently used in climate change studies (Araújo and Peterson, 2012) and to inform decision-making in conservation (Araújo et al., 2011). For that reason, evaluating the uncertainty and predictive ability of correlative SDMs is crucial.

* Corresponding author. Tel.: +34 913367083.
E-mail address: emorenoamat@gmail.com (E. Moreno-Amat).

Model complexity is well known to affect uncertainty of models and their transferability across space and time (Warren et al., 2014). However, the complexity of models is usually disregarded, especially when projecting SDMs across time. In addition, evaluating models requires occurrences independent from the calibration datasets, particularly when projecting models into the future, where such data are lacking (Araújo et al., 2005). One approach is to project models back in time and validate them with paleoecological data (Svenning et al., 2011). For example, paleobotanical records, which provide information on vegetation composition of the past, can be used as independent datasets to validate the past predictions of SDMs calibrated on present-day plant distributions (e.g. Pearman et al., 2008; Record et al., 2013, among others). Model complexity, defined as the number of parameters included in a model, is crucial for inferring habitat quality and estimating the breadth of species' niches (Warren and Seifert, 2011). Besides, complexity can also negatively affect the predictive performance of a model as a result of overfitting (Warren and Seifert, 2011). Although recent studies have assessed different aspects of complexity on SDM performance (e.g. Merow et al., 2014; Muscarella et al., 2014; Syfert et al., 2013; Warren and Seifert, 2011, among others), only one study has explored the uncertainty associated with model complexity on SDM projections for future climate scenarios (Warren et al., 2014). However, the effects of complexity on the predictive ability of SDMs across time remain poorly known.

Among the different SDM algorithms, maximum entropy (Maxent; Phillips et al., 2006) is extensively used for projecting current species distributions to different time periods (Elith and Leathwick, 2009). Maxent is a very flexible modeling algorithm widely used because of its better predictive performance relative to other modeling algorithms (Elith et al., 2006) even with low sample sizes (Pearson et al., 2007), its applicability to presence-only data (Phillips et al., 2006), and its user-friendly interface as well as availability through the dismo package (Hijmans and Elith, 2013). Maxent has been described as a modeling method able to fit overly complex response curves (Elith and Leathwick, 2009), particularly when using default parameters (Merow et al., 2013). Recently Maxent has been proved to be mathematically equivalent to a Poisson regression, a particular type of generalized linear models (GLM; Renner and Warton, 2013). The default settings of Maxent have been tested over a wide range of species and environmental conditions (Phillips and Dudík, 2008), but not in relation to cross-temporal transferability.

The complexity of Maxent models can be adjusted mainly through the choice of (1) the number of environmental predictors, (2) the feature classes allowed, and (3) the regularization (β -multiplier) selected in the model. The initial selection of the number of environmental predictors that best describe the species' ecological niche has a great influence on model performance (Harris et al., 2013; Synes and Osborne, 2011). Generally, this selection is based on previous knowledge of the ecology of the species and/or statistical assessments (Morueta-Holme et al., 2010). The Maxent algorithm consists of an iterative process that automatically selects the features that best explain the species' distribution (Merow et al., 2013). Features are basis functions and other transformations of the environmental predictors (i.e. linear, quadratic, product – i.e. interaction between variables – threshold and/or hinge; Phillips and Dudík, 2008). The features considered can be manually set by the user (Phillips and Dudík, 2008), or automatically by the algorithm when using the “autofeatures” option. Finally, given a fixed number of predictors and features allowed in the model, Maxent controls for model complexity through the regularization parameter (a set of parameters called beta-multipliers). Maxent forces the predicted values of the variables such as mean and variance to match the values of the presence locations. Thus, the regularization parameter prevents Maxent from over-fitting assuring that the

predicted values do not fit too exactly the empirical constraints of the predictor (Merow et al., 2013). Higher values of regularization parameter penalize the inclusion of parameters, thus creating less complex models than the default (regularization parameter = 1). Also the β -multiplier limits the number of features included in the model based on the number of presences (with more data allowing for an increasing number of features; Merow et al., 2013). Even though Maxent-users can control model complexity by modifying default setting options (i.e. model specifications; Phillips et al., 2006), in most studies using Maxent, model complexity is completely ignored (Muscarella et al., 2014; Warren et al., 2014; Yackulic et al., 2013).

Here, we study the effect of model complexity on the ability of Maxent models to predict species distributions across time. Specifically, we fit models with current presence data for *Alnus glutinosa* (L.) Gaertn. and *Corylus avellana* L. and different levels of complexity, and compare their performance when predicting both present and mid-Holocene distributions. Furthermore, we analyze whether the Maxent default settings are optimal to project these species under both current and past climate conditions. These two European species are ideal for evaluating model projections across time due to their wide current distribution across Europe and their strong representation in the fossil record.

2. Materials and methods

2.1. Study area

Correlative SDMs assume the equilibrium of species with climate and that the training data are illustrative of environment to which the models are predicted (Elith et al., 2010). In order to account for the full ecological range of the species (Hijmans and Elith, 2013), we fit models using available current occurrences from Europe. We used the paleorecord from continental Spain to validate the model projections to past climatic conditions. Consequently, we avoided truncating the niche estimation and reduced the likelihood of extrapolating, i.e. projecting outside the climatic conditions present in the training data (Elith et al., 2010). Following Elith et al. (2010) we checked for extrapolation, with a multivariate environmental similarity surface analysis (MESS), by comparing the current climatic conditions contained in the calibration dataset and mid-Holocene climatic conditions (Fig. C1).

2.2. Study species

A. glutinosa (black alder) and *C. avellana* (hazel) are widely distributed in the Atlantic and Centro-European Region, and have their southern geographical limits in the Mediterranean Region (Fig. 1). The distribution of *A. glutinosa* ranges from Ireland to western Siberia and from mid-Scandinavia to Northern Africa (Morocco and Algeria; Kajba and Gračan, 2003). Meanwhile, *C. avellana* populations are distributed from Ireland to the Ural Mountains in Russia and from Scandinavia to southern Spain (Castroviejo et al., 1986–2012).

Within Spain, the two species are mainly found in the northern mountainous region with some scattered populations in the southwestern part of the country (Castroviejo et al., 1986–2012). According to the pollen record, and considering the lack of mid-Holocene pollen records in Western and Southwestern continental Spain, the current distribution of both species in Spain is similar to that of the mid-Holocene (Fig. 1). Additionally, identification of their pollen is very reliable, as they are the only species within their respective genera present during the Holocene in the Iberian Peninsula.

Download English Version:

<https://daneshyari.com/en/article/6296563>

Download Persian Version:

<https://daneshyari.com/article/6296563>

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