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, Diego Nieto-Lugilde c, Naia Morueta-Holme d, 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 Department of Botánica, Facultad de Ciencias, Universidad de Granada, 18071 Granada, Spain
d Section for Econinformatics and Biodiversity, Department of Bioscience, Aarhus University, Ny Munkegade 114, DK-8000 Aarhus C, Denmark
* Department of Integrative Biology, University of California—Berkeley, 3040 VLSB, Berkeley, CA 94720, USA

A R T I C L E   I N F O
Article history:
Received 6 February 2015
Received in revised form 30 May 2015
Accepted 31 May 2015

Keywords:
Pollen fossil
Corylus avellana
Alnus glutinosa
Model validation
Species distribution model
β-Multiplier

A B S T R A C T
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 (8000 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.
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. Pearson 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 (Merueta-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.
2.3. Presence data

Current occurrence data for the two species were obtained from the Global Biodiversity Information Facility database (GBIF; http://www.gbif.org/ accessed January 2012). We excluded occurrences with a reported coordinate precision ≥ 10 km as well as countries not represented in GBIF or with uneven sampling effort (Mateo et al., 2013). Thus, we only included occurrences from Andorra, Belgium, Finland, France, Germany, Netherlands, Norway, Spain, Sweden and United Kingdom (Fig. 1). From the extracted ∼50,000 presences per species, 10,000 records were randomly subsampled with 1 km minimum distance (Beck et al., 2014; Mateo et al., 2013) to reduce computing time. Non-metric multidimensional scaling (NMDS) analyses were performed with the distributions obtained from GBIF and the whole European distributions as in Atlas Florae Europaeae (Jalas and Suominen, 1972–1994) at 50 km for both species (Fig. D1). We discard the possibility of geographical bias in the calibration dataset leading to an environmental bias because there is an overlap in the environmental space between the GBIF data and the AFE data for both species as observed in (Figs. D1–D2).

We obtained paleorecords for the mid-Holocene in Spain (Fig. 1) from the European Pollen Database (2007-2015) (EPD; http://www.europeanpollendatabase.net), the Spanish research project Paleodiversitas (Carrión et al., 2013; Paleodiversitas, 2011, http://www.paleodiversitas.org/) and other references (Table A1 and references in Appendix A). The pollen percentages of each species were estimated relative to the total amount of terrestrial plant pollen grains (i.e. excluding wetland and aquatic species) for the 6 ± 0.25 kyr cal BP interval. Pollen data obtained from the EPD or directly provided by the authors were expressed as the average of pollen percentages. We estimated percentages for the remaining pollen records from pollen diagrams. To locate the studied time period when ages in the diagrams were not expressed as calendar years, we calibrated the radiocarbon dates using the program CALIB 6.0 (Stuiver et al., 2012) coupled with the INTCAL09 calibration curve (Reimer et al., 2009), and constructed an age-depth model using linear-interpolation between consecutive 14C datings. Only pollen records with chronologies built with two or more 14C datings were used.

We classified pollen percentages into presence/absence using a 0.5% and 1% threshold for Alnus and Corylus, respectively (Lisitsyna et al., 2011; Szczepanek et al., 2004). For Alnus, the 0.5% threshold provides the highest agreement between pollen and vegetation data (Lisitsyna et al., 2011). For Corylus, a 1% threshold was recommended by Lisitsyna et al. (2011).

A total of 55 pollen sites were studied (Table A1), with 27 presences and 28 absences for Alnus and 36 presences and 19 absences for Corylus in the Iberian Peninsula (Fig. 5). Since both Alnus and Corylus are wind pollinated taxa with high pollen production (Erdtman et al., 1969), underrepresentation of species in the pollen core is rare. Hence, absences in pollen sites can be considered reliable.

2.4. Climate data

Current climate data for Europe were downloaded from the Worldclim database (http://www.worldclim.org/; Hijmans et al.,
Table 1 Environmental variables (http://www.worldclim.org/ Hijmans et al., 2005) with indication of the sets of nested variables (4var, 10var, 13var, 15var and 19var) used to fit the models.

<table>
<thead>
<tr>
<th>Description</th>
<th>Variable set</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIO 01 Annual mean temperature</td>
<td>X</td>
</tr>
<tr>
<td>BIO 02 Mean diurnal range (Mean of monthly (max temp − min temp))</td>
<td>X</td>
</tr>
<tr>
<td>BIO 03 Isothermality (BIO2/BIO7 × 100)</td>
<td>X</td>
</tr>
<tr>
<td>BIO 04 Temperature seasonality (standard deviation × 100)</td>
<td>X</td>
</tr>
<tr>
<td>BIO 05 Max temperature of warmest month</td>
<td>X</td>
</tr>
<tr>
<td>BIO 06 Min T° coldest month</td>
<td>X</td>
</tr>
<tr>
<td>BIO 07 Temperature annual range (BIO5–BIO6)</td>
<td>X</td>
</tr>
<tr>
<td>BIO 08 Mean temperature of wettest month</td>
<td>X</td>
</tr>
<tr>
<td>BIO 09 Mean temperature of driest quarter</td>
<td>X</td>
</tr>
<tr>
<td>BIO 10 Mean T° warmest quarter</td>
<td>X</td>
</tr>
<tr>
<td>BIO 11 Mean temperature of coldest quarter</td>
<td>X</td>
</tr>
<tr>
<td>BIO 12 Annual precipitation</td>
<td>X</td>
</tr>
<tr>
<td>BIO 13 Precipitation of wettest month</td>
<td>X</td>
</tr>
<tr>
<td>BIO 14 Precipitation of driest month</td>
<td>X</td>
</tr>
<tr>
<td>BIO 15 Precipitation seasonality (Coefficient of variation)</td>
<td>X</td>
</tr>
<tr>
<td>BIO 16 Precipitation of wettest quarter</td>
<td>X</td>
</tr>
<tr>
<td>BIO 17 Precipitation of driest quarter</td>
<td>X</td>
</tr>
<tr>
<td>BIO 18 Precipitation of warmest quarter</td>
<td>X</td>
</tr>
<tr>
<td>BIO 19 Precipitation of coldest quarter</td>
<td>X</td>
</tr>
</tbody>
</table>

2005) at 30 arcs (~1 km²) resolution. This database provides 19 bioclimatic variables derived from monthly temperature and precipitation data for the 1950–2000 period. These variables represent annual trends, seasonality and extremes in climate that act as limiting environmental factors for many organisms (Hijmans et al., 2005; Kumar and Stohlgren, 2009).

In order to reproduce mid-Holocene climatic conditions in the Iberian Peninsula and to determine the effect of the General Circulation Model, we used two Ocean Atmospheric General Circulation Models (OA-GCMs): the CCSM (http://www.cccsm.ucar.edu; Kiehl and Gent, 2004) and the MIROC3.2 (http://www.ccsr.u-tokyo.ac.jp/ehml/etopindex.shtml). Monthly values of temperature and precipitation for the two OA-GCMs were compiled from the Paleoclimate Modeling Intercomparison Project Phase 2 (PMIP2) database. The variables were extracted at the original resolution (2.5° cell sizes) covering the entire Iberian Peninsula. Following the standard change-factor approach to downscale the data, we computed anomalies between mid-Holocene and current climatic conditions from the OA-GCMs output, and downscaled the anomalies to a spatial resolution of 30 arcs using regularized splines. To avoid potential estimations of negative values of precipitation, the anomalies for this variable were calculated as relative differences (Hijmans and Graham, 2006; Lima-Ribeiro et al., 2013). Finally, mid-Holocene climates were calculated adding the interpolated differences to the current climate (i.e. pre-industrial values) from the Worldclim database. This approach accounts for the effect of topography, ensuring consistency of the climatic patterns across time (Hijmans and Graham, 2006). Finally, the 19 bioclimatic variables were calculated from the downscaled data of past monthly temperatures and precipitations.

2.5. Model calibration and tuning of complexity

We ran the models with the Maxent modeling algorithm version 3.3.3k (Phillips et al., 2006). Present distribution models for both species were developed with the occurrence data and 10,000 random background points, representing the distribution of environmental conditions in the study region (Phillips and Dudík, 2008). To avoid a geographical bias in our models, we randomly chose the background points from the same European countries for which presences were used. Because we were interested in studying the variation in performance across models, we built all the models with the same occurrence and background data to avoid any potential source of variation due to differences in data input. We randomly selected 70% of the data (both presences and background) to fit the models and held the remaining 30% for testing purposes running one replicate per model. To project the models under mid-Holocene conditions we used clamping, the default setting in Maxent in case of extrapolation, i.e. making the response constant outside of the range of the training data (Elith et al., 2010). Finally, we used the logistic output for the final geographical predictions (Phillips and Dudík, 2008).

We analyzed three potential sources of complexity: number of variables, number and type of features included in the model, and β-multiplicators. To do so, we calibrated a total of 110 models for each species. First, we selected five subsets of nested predictor variables (with 4, 10, 13, 15 and 19; Table 1). The 4 variables set was based on ecological criteria (Pearman et al., 2008; Prentice et al., 1992) and avoiding correlated variables (Pearson $r > 0.75$), with increasing correlation allowed in the subsequent set of variables. Second, we compared two different sets of feature classes: the default “auto-features” option (which allows all possible features), and allowing only linear, quadratic, and product (LQP) features. We used LQP all together because species responses to ecological gradients are frequently nonlinear. Quadratic responses are suitable for unimodal curves, as expected for fundamental niches (Austin, 2007), and product, i.e. interaction between variables, is not rare. Auto-features include for LQP plus threshold and hinge features, which are useful when there is a physiological tolerance limit (Merow et al., 2013). Finally, we tested 11 different β-multiplicators (from 0 to 5 in steps of 0.5).

The most complex model resulted from the combination of the highest number of variables (19 variables), the lowest β-multiplicator ($β = 0$) and enabled auto-features, whereas the simplest model arose from using the smallest number of variables (four variables), the highest β-multiplicator ($β = 5$), and allowing only LQP features in the model.

2.6. Model evaluation

In order to evaluate the performance of models under current climate conditions, we used two methods. First, we calculated the AUC (Area Under the receiver operating characteristic Curve) using the testing dataset (30% of the presences and background; Araújo et al., 2005). AUC is a threshold independent statistic that informs about the ability of a model to discriminate between presences and
absences from the study area (Fielding and Bell, 1997; Lobo et al., 2008). AUC has become the standard to assess the predictive accuracy of SDMs, allowing comparisons among different models for a single species and a single study region (Lobo et al., 2008). Second, we calculated the Akaike Information Criterion corrected for small sample sizes (AICc; Johnson and Omland, 2004) since AUC does not account for goodness-of-fit of the model, model complexity, or the increase of false absences due to the use of background data (Lobo et al., 2008; Warren and Seifert, 2011). AICc is a balanced statistic between the goodness of fit and the number of parameters of the model (Johnson and Omland, 2004), which allows selecting models with optimal complexity (Warren and Seifert, 2011). Additionally, we assessed the differences in suitability of the best current models selected according to AUC and AICc projected to geographical space.

To evaluate projections to mid-Holocene climate conditions in continental Spain, we used the fossil pollen data as independent datasets. Here, we calculated AUC, which tests the ability of the models' mid-Holocene projections to discriminate between real mid-Holocene presence and absence of the species derived from thresholded pollen data (e.g. Pearman et al., 2008). Because pollen from wind pollinated taxa can represent presences at coarser spatial scales than our analysis (Nieto-Lugilde et al., 2015), we also calculated the correlation (Spearman coefficient) between pollen percentages and the habitat suitability values predicted by the model. Pollen percentages should be less sensible to scale than presences/absences derived from pollen thresholds. This correlation also served as a sensitivity test to assess whether the percentage and threshold analyses lead to the same results, since most SDM studies using pollen data only use thresholded values (Pearman et al., 2008; Record et al., 2013) and there is no consensus in the percentage that defines the presence for each species (Lisitsyna et al., 2011). AICc was not calculated in the past because it was obtained only for the calibrated models which were the current ones.

All the analysis were conducted with the R software (packages: vegan, dismo and raster; http://cran.r-project.org/) except for the AICc analysis, which was run with the software ENMtools (Warren et al., 2010). Past climate variables were transformed with GRASS GIS software (GRASS Development Team, 2012).

3. Results

3.1. Evaluation of current species distributions

The AUC on the test dataset varied from 0.765 to 0.889 (mean ± s.d. = 0.844 ± 0.032) for Alnus (Fig. 2A–C) and from 0.748 to 0.879 (0.823 ± 0.034) for Corylus (Fig. 2D–F). For both species, AUC increased with model complexity. The highest values were thus observed in the most complex models—i.e. models resulting from combining 19 variables, β = 0 and autofeatures. Conversely, the lowest AUC values were obtained with the simplest model—i.e. when combining four variables, β = 5, and allowing only LQP features (Fig. 2C and F). When the number of parameters exceeded 100–125 the AUC values reached a plateau. Differences in AUC between the five variable sets were small, except for the models run with 4 variables.

Contrary to AUC, AICc values did not converge on the most complex model (Fig. 3). For both species, the best model according to AICc was the one run with autofeatures and highest number of variables (19), but differed in the optimal β-multiplier. The best model for Alnus was obtained when the β-multiplier equaled 2.5, resulting in 81 parameters, while the default β-multiplier (β = 1) produced 133 parameters (Fig. 3A and B). The best model for Corylus was at β-multiplier equal 1.5, resulting in an intermediate number of 107
parameters, whereas the default $\beta$-multiplier ($\beta = 1$) produced 136 parameters (Fig. 3D and E).

Both AUC and AIcc values were constant across $\beta$-multipliers when only LQP features were allowed in the models, being only affected by the number of variables.

The projected suitable areas for the best models selected according to AUC and AIcc showed similar patterns (Figs. G1 and G2). However, the models with lowest AIcc tended to predict higher suitability values in each pixel than models with the highest AUC (Figs. G1, C and F).

### 3.2. Evaluation of mid-Holocene projections

The results for both species were similar under both OA-GCMs (Fig. 4 and Fig. E1 for MIROC and Figs. F1–F2 for CCSM), with no significant differences in mean AUC values (Student’s $t$-test, $n = 220$, $t = –1.335$, $p$-value = 0.183) nor in the correlation values between pollen percentages and the suitability index of the models ($n = 220$, $t = –0.303$, $p$-value = 0.762). Hence, we report the results for MIROC in the rest of the manuscript, while CCSM results are reported in Appendices F1–F2 for the sake of simplicity.

AUC values for mid-Holocene projections evaluated with pollen data decreased slightly relative to the tests with current occurrences (Fig. 4). AUC varied from 0.663 to 0.808 (0.741 ± 0.035) for *Alnus*, and from 0.762 to 0.872 (0.823 ± 0.021) for *Corylus*. The highest AUC for *Alnus* was obtained with 19 variables, autofeatures, and $\beta$-multiplier equal to 1.5, whereas for *Corylus*, it was obtained with four variables, autofeatures, and $\beta$-multiplier equal 1.5 (Fig. 5). AUC generally increased with increasing $\beta$-multipliers (Fig. 4), with the exception of models with 4 variables, which showed constant values for all $\beta$-multipliers. In the case of *Alnus*, all the models (except those with four variables) reached a maximum at $\beta = 1.5–2.5$ and then slightly decreased, whereas for *Corylus*, all the models (except those with four variables) stabilized at $\beta = 3$.

The effect of model complexity tested against pollen percentages was similar to that described for AUC, except for the models of *Corylus* run with four variables (Figs. E1 and F2). *Alnus* models reached a maximum at $\beta = 1.5–2.5$ and then decreased, whereas *Corylus* models stabilized at $\beta = 3$. Similar to the evaluation with current data, the past AUC and the correlation values were the same along all the $\beta$-multipliers used when only LQP features were allowed in the models (Fig. 4B and D and Fig. E1, B and D).

### 4. Discussion

The complexity of SDMs affects both model performance and the area predicted as suitable (Warren and Seifert, 2011; Merow et al., 2014). Even when complex models show high performance predicting current distributions, they may produce poor predictions at new sites or different time periods (Warren et al., 2014). The poor transferability of complex correlative models could be a consequence of not capturing the species niche correctly by fitting the models too closely to the training data (overfitting) and preventing them from generalizing well (Warren et al., 2014). Assessing the effects of complexity in the transferability of correlative SDMs predictions is still fundamental given their broad implementation.

#### 4.1. Effects of complexity in current and past species distributions

As expected, we found that complex models performed well under current conditions. The best models according to the AUC for predicting current distributions were the most complex, reflecting that AUC does not consider model complexity nor goodness-of-fit (Lobo et al., 2008; Warren and Seifert, 2011). In contrast, using the AIcc and thus selecting the model by balancing
Fig. 4. AUC values for Maxent models projected to the mid-Holocene under MIROC General Circulation Model for Alnus (A and B) and Corylus (C and D) as a function of β multiplier (A and C) and nested sets of variables (B and D). The boxplots show AUC variability within each variable sets (B and C). When only linear, product and quadratic features are included in the model (autofeatures option deselected) AUC values remained constant, and are therefore plotted as red crosses in B and D. Empty circles indicate outliers. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 5. Projections to the mid-Holocene in Spain of the best Maxent models according to the correlation between fossil pollen percentage and the suitability predicted by the models under CCSM General Circulation Model: presences (blue crosses) and absences (pink squares) for Alnus (A) and Corylus (B). The best model was the one with autofeatures, 10 variables and β multiplier = 2.5 for Alnus (A) and the one with autofeatures, 13 variables and β multiplier = 2.5 for Corylus (B). The predicted suitability of the species in the area is displayed in a color gradient: red indicates high suitability and blue indicates low suitability. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
overparameterization and model fit, resulted in the selection of models with less parameters (Fig. 3), which had slightly lower AUC (Fig. 2). Furthermore, the AICc-selected models showed better cross-time transferability than AUC-selected models, with the former resulting in higher AUCs and correlation values when hindcasted. Thus, while the more complex models performed well under current conditions, they performed poorly when transferred in time to mid-Holocene climatic conditions (Fig. 4 and Figs. E1, F1 and F2).

The effects of model complexity were also apparent as geographic differences in the predicted current distribution of each species (Figs. G1–G2). For instance, the less complex AICc-selected models predicted a wider potential distribution for both species than the overfitted AUC-selected models. However, these differences varied across regions (Figs. G1–G2). For example, AUC-selected model predictions were narrower in the northern parts of the distribution for both species, although this trend was not as clear in the Iberian Peninsula, especially in the binary predictions (Fig. G2).

The two independent paleobotanical validations used (past AUC using pollen thresholds and correlation using pollen percentages) overall led to the same trends. Selecting a threshold to estimate local species presence from pollen percentage is complex due to the lack of agreement in the percentage that could be considered as presence for a species in a pollen core (Lískiysna et al., 2011). However, since the percentage and threshold analyses produced similar results, the threshold selection did not introduce bias.

Our results are thus in line with Warren et al. (2014), who found that model complexity strongly impacted both the modeled niche breadth and the projected suitability under future climate scenarios. Complex models calibrated on current conditions may thus overestimate the reduction of environmental suitability under climate change. As correlative SDMs are increasingly used as a tool to identify future suitable areas for hosting species (Araújo et al., 2011) such erroneous projections can have strong consequences in conservation plans and management strategies.

4.2. How do Maxent settings affect model complexity?

Our results show that the factors most affecting Maxent model complexity are primarily the features classes included in the model, second the number of environmental predictors, and finally the β-multiplier selected (Fig. 3, and Table B1).

In terms of model performance, the feature types had a much larger effect than the number of variables or the β-multiplier (Figs. 2 and 4 and Figs. F1–F2). This is because forcing the model to use only LQP features heavily restricted the number of parameters allowed in the model (Table B1). In current projections, the AUC-selected and AICc-selected models agreed in the use of autofeatures as well as AUC in hindcasted models. Thus, in both present and past times, autofeatures were the option with greatest predictive performance. In contrast, Syfert et al. (2013) found that the choice of feature types had relatively minor effects on model performance when compared to the correction for sampling bias, and suggested that the autofeatures may be capturing local idiosyncratic effects rather than broad physiological responses of the species. They recommended fitting simple features like linear and quadratic and use them when they present similar performance values than with the autofeatures option (Syfert et al., 2013). Furthermore, Shcheglovitova and Anderson (2013) stated that it is possible to achieve models that outperform those obtained with default settings by coupling complex features with higher regularization to ensure that the model has a low number of parameters. However, in our study the number of variables had a greater effect than the β-multiplier. Too few variables could lead to under-parameterization and thus a poor representation of the full ecological range of the species (Synes and Osborne, 2011; Syfert et al., 2013) as observed in our results of both species in the 4 variable model set.

Finally, the β-multiplier also had some influence on model performance and cross-temporal predictions. We found that the β-multiplier selected as the best compromise between performance and overprediction was higher than the default (β=1), in accordance with former studies (Anderson and González, 2011; Muscarella et al., 2014; Radosavljevic and Anderson, 2014; Shcheglovitova and Anderson, 2013). AICc-selected models had higher β-multiplier and thus less parameters than the default β-multiplier models, which were not the models with highest AUC as also observed by Muscarella et al. (2014). Decreasing AUC of the models with higher regularization multiplier has also been demonstrated in other studies (e.g. Shcheglovitova and Anderson, 2013).

4.3. The importance of species-specific tuning in Maxent models

The two species studied obtained different optimal parameters for best cross-temporal transferability among them (same features, but different number of variables and beta-multipliers) but higher (regularization parameter) than default Maxent settings in both species. Therefore, our findings jointly with other studies (e.g. Cao et al., 2013; Halvorsen et al., 2015; Muscarella et al., 2014; Warren et al., 2014), highlight the importance of testing a range of parameter settings to see which perform best in each study case when using Maxent to predict the distribution of a species in different time periods.

5. Conclusions and further recommendations

Model complexity is a major issue that affects model performance and transferability and should be taken into account when projecting models across time. In general, Maxent users select default settings for complexity which, as we demonstrated in this study, can lead to the generation of over-complex models. In our study, complex models performed more poorly over time, indicating that standard methods for evaluating models that do not consider complexity (Fielding and Bell, 1997; Lawson et al., 2014) cannot be recommended. We highly encourage conducting species-specific tuning when modeling distributions with Maxent to determine the best modeling options, as suggested by other authors (e.g. Halvorsen et al., 2015; Muscarella et al., 2014; Shcheglovitova and Anderson, 2013). When paleoclimatological data are available, we recommend their use to make validations of SDMs projected in time. The use of this data allows us to understand the effects of model complexity through time and constitutes a reliable and independent dataset to test hindcasted models. Unfortunately, however, such data are scarce. When no paleoclimatological data are available and the aim of the study includes cross-temporal projections of species distributions, our results suggest that simpler models, with an intermediate complexity (intermediate number of variables and intermediate β-multiplier coupled with autofeatures option selected) will produce the best cross-temporal transferability.

Data accessibility

Species current distribution can be found in GBIF database:
http://www.gbif.org/species
Species mid-Holocene distribution can be obtained from: European Pollen Database (http://www.europeanpollendatabase.net/fpd-epd) and references uploaded as online supporting data.
Climate data: http://www.worldclim.org/download.
Acknowledgments

EMA was supported by a Technical University of Madrid (Spain) grant (UPM-RR01/2011) and a visiting grant from Graduate School Science and Technology (Aarhus University, Denmark). This research was funded by the project CGL2011-27229 (DINECOFOR) from the Spanish Ministry of Economy and Competitiveness, and CLIMIFORAD from the Inter-American Development Bank (RG-T1837). RGM was funded by a Marie Curie Intra-European Fellowship within the 7th European Community Framework Programme (ACONITE, PIEF-GA-2013-622620). DNL was supported by Universidad de Granada (Spain) throughout a Perfeccionamiento de doctoros Fellowship. NMH acknowledges support from the Aarhus University Research Foundation. JCS was supported by the European Research Council (ERC-2012-SIG-310886-HISTFUNC). The authors are especially grateful to César Morales del Molino, Lourdes López Merino, Maria Fernanda Sánchez Goñi, José Antonio López Sáez, Francesc Burjachs Casas, Miriam Dorado Valiño, Tony Stevenson and Penélope González Sampériz for providing the raw pollen data. We also thank the data providers of the Global Biodiversity Information Facility and European Pollen Database for making biodiversity data freely available. We are indebted to Salvia García Álvarez for compiling and harmonizing fossil data in the Iberian Peninsula. The authors declare that they have no conflict of interest.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ecolmodel.2015.05.035

References


