Ecological Modelling 263 (2013) 103–108

Contents lists available at SciVerse ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Evaluating the predictive performance of stacked species distribution models applied to plant species selection in ecological restoration

Aitor Gastón*, Juan I. García-Viñas[1]

EGOGESFOR Research Group, Universidad Politécnica de Madrid, Escuela de Ingeniería Forestal y del Medio Natural, Ciudad Universitaria s/n, 28040 Madrid, Spain

ARTICLE INFO

Article history:
Received 10 October 2012
Received in revised form 21 March 2013
Accepted 29 April 2013
Available online 5 June 2013

Keywords:
Habitat models
Model validation
Revegetation

ABSTRACT

We propose an evaluation approach to validate stacked species distribution models applied to plant species selection in ecological restoration. The evaluation method does not assess the accuracy of individual species models, but focuses on the ability of the stacked models to discriminate between present and absent species in a vegetation relevé. We measured the discriminative ability using the area under the ROC curve (AUC) to avoid the drawbacks of converting occurrence probabilities into binary predictions. Using the proposed method, we compared competing sets of predictors and validated stacked species distribution models for plant species selection in ecological restoration projects in Spain. 120,938 vegetation relevés included in the Forest Map of Spain were used to train models for 188 species and an independent set of 100 vegetation relevés was used for validation. The best performing set of predictors included climate and soil related predictors derived from coarse resolution datasets. The model performance was acceptable on average (mean AUC: 0.88, sd: 0.07) and high (AUC > 0.9) in 42% of the relevés evaluated. We recommend the proposed evaluation approach to validate stacked species distribution models used to support species selection in ecological restoration projects.

1. Introduction

Ecological restoration is an intentional activity that initiates or accelerates the recovery of a degraded ecosystem with respect to its health, integrity and sustainability (SER, 2004). Restoring plant communities frequently involves planting or seeding (Clewell et al., 2005), therefore, a proper species selection is mandatory in such projects. Although many different criteria may be used in species selection, the selected species should be matched to local environmental conditions (Peman Garcia et al., 2008). Matching species to sites requires the estimation of habitat suitability for the considered species. Such estimations are part of the outputs of predictive species distribution models (SDMs), which relate field observations to environmental predictor variables (Guisan and Thuiller, 2005).

As with any predictive empirical model, SDMs should be validated through an evaluation of their predictive performance prior to use by practitioners (e.g. Fielding and Bell, 1997). Several authors have provided guidelines for model validation, although methodological decisions should be contingent on the user's intent (Araújo and Peterson, 2012). When models are used to predict areas for restoration, translocation, or reintroductions, performance evaluations should focus on the reduction of errors of omission (Araújo and Peterson, 2012). However, the goal of a restoration project is not always to predict areas for restoration of populations of given species, but to predict the most suitable species for a given site (e.g., restoring a degraded plant community). The number of suitable plant species can easily exceed the maximum number of species (n_i) that can be introduced for practical reasons, therefore, practitioners must identify the most suitable n_i species from among a limited number of commercially available species (N). This kind of decision requires one SDM for each one of the N species in order to predict the occurrence probability for each species in the site to be restored. The concern of the practitioner is the ability of the N stacked models to rank the N species according to habitat suitability at a particular site to be restored.

The evaluation of predictions from stacked SDMs against observed species assemblages has often been used in ecological research and environmental management. Some evaluations measure the likelihood of observing the assemblage given the predictions of the stacked SDMs using the log-likelihood (Oberdorff et al., 2001; Clarke et al., 2003). This approach allows us to directly compare predicted probabilities of occurrence with observed occurrences, although probabilities are more often converted to occurrences using a threshold prior to measuring predictive performance, which allows the comparison of predicted and observed species assemblages. Some authors use compositional

* Corresponding author. Tel.: +34 913367669; fax: +34 915446025.
E-mail addresses: aitor.gaston@upm.es (A. Gastón), juanignacio.garcia@upm.es (J.I. García-Viñas).
[1] Tel.: +34 913367558; fax: +34 915446025.

0304-3800/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecolmodel.2013.04.020

dissimilarity indices between observed and predicted assemblages, e.g., Bray–Curtis Index (Oberdorff et al., 2001; Clarke et al., 2003; Hallstan et al., 2012; Larsen et al., 2012), Simpson's index of dissimilarity (Baselga and Araújo, 2009), Jaccard's CC index (Tanaka and Koike, 2011) or Sorensen index (Pellissier et al., 2011). Other approaches focus on measures derived from the confusion matrix, e.g., false positive and false negative rates (Avery and Van Riper, 1990; Block, 1994; Feria and Peterson, 2002), true positive and true negative rates (Feria and Peterson, 2002; Kattwinkel et al., 2009), correct classification rate (Kattwinkel et al., 2009; Gabriels et al., 2007) or Cohen's Kappa (Kattwinkel et al., 2009; Gabriels et al., 2007).

We are not aware of any existing performance evaluation of stacked SDMs applied to plant species selection in ecological restoration. As mentioned above, the concern of the practitioner is the ability of the stacked models to discriminate between present and absent species at a particular site. This issue may be solved using a discrimination measure that compares predictions from stacked SDMs against observed species assemblages at the sites evaluated. The measures previously used to evaluate stacked SDMs require the adoption of a threshold to convert probabilities into predicted occurrences. One problem with the threshold dependent measures is their failure to use all of the information provided by the model (Fielding and Bell, 1997). In the case of species selection for ecological restoration, converting probabilities into predicted occurrences equates species with different suitability values for the restoration site (e.g. 0.2 and 0.9 are equal for a threshold of 0.15). One way of avoiding this problem is to use threshold-independent measures, such as the area under the receiver operating characteristic curve (AUC). The AUC measures the probability that a random selection from the positive group will have a score greater than a random selection from the negative class (Fielding and Bell, 1997). In this case, the AUC would be computed for each site and would estimate the probability that the stacked SDMs offer a higher suitability estimate for a present species than for an absent species (both of them picked at random). Here we propose the use of site AUC to evaluate the predictive performance of stacked SDMs applied to plant species selection in ecological restoration. First we fitted SDMs for 188 plant species native to Spain (see Section 2.1) using national scale datasets (see Sections 2.1 and 2.2) and penalized logistic regression (see Section 2.3). We then applied the proposed evaluation approach to validate stacked SDMs for plant species selection in ecological restoration projects in Spain (see Section 2.4) using an independent set of vegetation relevés (see Section 2.1). We focused on two potential applications for the proposed evaluation approach: the estimation of the overall performance of the stacked models and the comparison of the performance of different sets of predictors (see Fig. 1).

The proposed evaluation approach allows the modeller to build SDMs better suited for plant species selection in ecological restoration. More accurate models are expected to decrease the likelihood of selecting unsuitable species and, consequently, improve the success rates of the restoration projects.

2. Materials and methods

2.1. Species occurrence data

We fitted SDMs for vascular plant species native to continental Spain using data from 120,938 vegetation relevés carried out between 1986 and 1997 as part of the field survey of the Forest Map of Spain (Ruiz de la Torre, 1990). The vegetation relevés in the Forest Map of Spain include all the woody species as well as some large and dominant grasses. The relevés include non-forest species (e.g. shrubs like *Atriplex halimus* or grasses like *Lygeum spartum*) as the map covers not only forests, but also open woodlands, shrublands and grasslands. Species with less than 15 occurrences were excluded, resulting in 188 species to be modelled (77 trees, 104 shrubs and 7 large grasses). The number of occurrences ranged from 15 to 42,720 with 90% of the species between 28 and 8735 occurrences (see Fig. 1).

An independent set of vegetation relevés by Prof. Ruiz de la Torre (Gastón et al., 2011) was used for model evaluation. Only relevés from locations with evidently low degrees of human perturbation were used (tree cover more than 75%, more than 3 tree species and no evidence of human perturbation). 100 relevés (see Fig. 2) met the requirements with an average of 8.3 tree species and a mean tree cover of 87%. The presence or absence of the 188 modelled species was extracted for the 100 relevés that met the requirements.

2.2. Environmental variables

Climatic data grids were derived by applying a multiple regression model based on meteorological station data (Sánchez Palomares et al., 1999) to the STRM 3″ (≈90 m) elevation dataset (Farr et al., 2007). A set of 17 climatic variables commonly used in tree species autoecology in Spain (e.g., Gandullo and Sánchez Palomares, 1994) were initially considered as candidate predictors: mean seasonal rainfalls (4), mean annual rainfall, mean seasonal temperatures (4), mean annual temperature, mean maximum temperature of the warmest month, mean minimum temperature of the coldest month, dry season length, dry season intensity, mean annual potential evapotranspiration, mean annual water surplus, and mean annual water deficit.

A variable clustering approach was used as a variable reduction strategy (Harrell, 2001). Hierarchical clustering was conducted on a similarity matrix (squared Spearman correlation coefficients) using the complete linkage clustering method. Once the variable groups had been defined, the first principal component of each group was taken as representative of that group. The data reduction procedure resulted in five groups of variables related to various environmental conditions (Fig. 3): (1) mean thermal conditions (*SpT*, *AuT*, *T*, *PET*), (2) summer thermal conditions (*SuT*, *Tw*), (3) winter thermal conditions (*WiT*, *Tc*), (4) water availability during the dry season (*DSL*, *DSI*, *SuR*, *WD*), and (5) mean water availability (*WiR*, *AuR*, *WS*, *SpR*, *R*).

The European Soil Database (ESDB, Van Liedekerke et al., 2006) was used as a soil data source. The ESDB comprises a coarse-scale soil map (1 km resolution grid) and an associated database with the values of several soil-related variables for each cell of the map. Three soil-related predictors were extracted from the ESDB: calcareous nature of the parent material (a binary value), presence of gypsum (a binary value), and major FAO soil group (14 classes).

Three subsets of increasing complexity predictors were tested: (1) climate only (15 parameters), (2) climate and lithology (17 parameters) and (3) climate, lithology and major soil group (30 parameters).

2.3. Modelling strategy

We used penalized logistic regression (Harrell, 2001) to fit the SDMs. The penalized regression outperformed an alternative regularization technique called *lasso* (Tibshirani, 1994) with small sample sizes in a comparison of regularization methods applied to species distribution models (Reineking and Schröder, 2006) and performed at least as well as Maxent in a wide range of sample sizes (Gastón and García-Viñas, 2011).