Ecological Modelling 266 (2013) 126–130

Contents lists available at ScienceDirect

Ecological Modelling

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

Short communication

Evaluating ensemble forecasts of plant species distributions under climate change

Shawn M. Crimmins*, Solomon Z. Dobrowski, Alison R. Mynsberge

Department of Forest Management, College of Forestry and Conservation, University of Montana, Missoula, MT 59802, USA

ARTICLE INFO

Article history:
Received 3 June 2013
Received in revised form 1 July 2013
Accepted 7 July 2013

Keywords:
Accuracy
California
Climate change
Ensemble forecast
Species distribution model
Transferability

ABSTRACT

Species distributions models (SDMs) are commonly used to assess potential species' range shifts or extinction risk under climate change. It has been suggested that the use of ensemble forecasts, where a variety of model algorithms are used to generate consensus predictions, are preferred to individual SDMs by avoiding bias or prediction error inherent in a single modeling approach. Whereas several studies have assessed the performance of ensemble predictions using cross-validation or data-partitioning approaches, few studies have assessed the predictive accuracy of ensemble forecasts under climate change by using temporally independent model validation data. We used five SDM approaches to develop consensus forecasts of distributions of 145 vascular plant species from California in the 1930s and tested their projections against current distributions, a span of approximately 75 years. When evaluated with a portion of the model training data, consensus forecasts were highly accurate with an average AUC value of 0.97. False positive and false negative error rates were also low, exhibiting similar performance to random forest models. However, when evaluated with temporally independent data, the accuracy of consensus forecasts was similar to that of generalized linear and generalized additive models, with an average AUC value of 0.83. Our results suggest that the high levels of predictive accuracy exhibited by consensus forecasts when using data partitioning approaches may not reflect their performance when predicting temporally independent data. We contend that consensus forecasts may not represent the best approach for predicting species distributions under future climatic change, as they may not provide superior predictive accuracy in novel temporal domains compared to traditional modeling approaches that more readily lend themselves to ecological interpretation of model structure.

1. Introduction

In recent years, the increasing availability of spatially explicit climate and species distribution data has led to the widespread use of species distribution models (SDMs) in ecological research. These models are often used to predict the impacts of climate change on biota by relating current species distributions to climate and then projecting future distributions under various climate change scenarios (Elith and Leathwick, 2009). There are now a variety of statistical techniques used to develop SDMs (Elith et al., 2006), although most studies continue to use a single modeling approach (Hanspach et al., 2010). Previous studies have advocated for the use of a single modeling algorithm, often suggesting that a single method may be superior (Lehman and Overton, 2002). However, recent studies comparing the accuracy of various modeling approaches have shown that prediction accuracy can vary substantially among modeling algorithms and have consistently failed to identify any single modeling approach as being superior to others (Segurado and Araújo, 2004; Elith et al., 2006; Pearson et al., 2006). Although these studies typically use data from a single time period to develop and test their models, there is also evidence to suggest that different modeling algorithms can lead to different predictions of species distributions over time (Araújo et al., 2005a; Dobrowski et al., 2011).

Because of this model based uncertainty in SDM accuracy, it has been suggested that the use of individual modeling algorithms should be abandoned in favor of ensemble forecasts (Araújo and New, 2007). The general premise behind ensembles is that uncertainty in individual model forecasts can be reduced by simultaneously considering the results from multiple models. This approach assumes that when averaging across multiple models that the true "signal" of interest will separate itself from the "noise" and bias associated with any individual models (Araújo and New, 2007). Recent studies have shown that consensus modeling methods may improve SDM predictions compared to individual modeling algorithms (Crossman and Bass, 2008; Marmion et al., 2009). As with the previous studies of SDM performance mentioned

* Corresponding author. Tel.: +1 636 751 3496.
E-mail addresses: shawn.crimmins@umontana.edu (S.M. Crimmins), solomon.dobrowski@cfc.umt.edu (S.Z. Dobrowski), alison.mynsberge@cfc.umt.edu (A.R. Mynsberge).

0304-3800/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecolmodel.2013.07.006

above, these works have used data within a single time frame to both develop and test models (Grenouillet et al., 2011). Conversely, recent studies using data from multiple time frames have shown inconsistent results with regards to ensemble forecast accuracy (Rapacciuolo et al., 2012; Smith et al., 2013). Because the factors affecting SDM accuracy within a single time frame may differ from those conferring accuracy when predicting into a novel time frame for individual modeling algorithms, evaluations of consensus forecasts across time are warranted.

It has been suggested that using data from two time periods represents the only way to directly assess the performance of SDMs for predicting climate change impacts (Nogues-Bravo, 2009). However, such data are extremely rare and, as such, there have been comparatively few instances of such data being used in evaluations of consensus forecasts. In one of the first studies to use such datasets, (Araújo et al., 2005b) found that consensus forecasts yielded more accurate predictions of coarse-resolution changes in bird species range size than individual (non-consensus) models, although the spatial resolution of their study limited the applicability of their results. However, more recent studies have presented conflicting evidence regarding how ensemble forecasts will compare to individual algorithms when projecting distributions across time, with some evidence suggesting ensembles provide superior prediction accuracy (Rapacciuolo et al., 2012) while other research suggests they do not (Smith et al., 2013). The objectives of this study were to (1) test whether consensus forecasts yield more accurate predictions of plant species distributions under climate change than individual models, and (2) determine if measures of consensus forecast accuracy based on data from a single time frame correlate to measures of accuracy in a novel time frame.

2. Materials and methods

2.1. Study area

Our study area comprised the dominant mountain ranges of California, USA, an area of approximately 220,000 km^2. This area was ideal for assessing climate change impacts on species distributions as it covered major biophysical gradients and has experienced substantial and diverse climate change during the 20th century which has led to major shifts in plant species distributions (Crimmins et al., 2011).

2.2. Vegetation data

We used plant species presence and absence data from two time periods to develop and test ensemble forecasts. We used historical species presence and absence data from 13,746 plots collected between 1928 and 1940 as part of the Vegetation Type Map Project (VTM) (Wieslander, 1935a,b). Georeferenced plot locations were estimated to be accurate within 200 m (Kelly et al., 2005). We compiled a collection of 33,596 contemporary vegetation plots from a variety of sources including the US Forest Service, National Park Service, California Fish and Game, US Geological Survey, California Native Plant Society, and academic institutions. All plots were georeferenced with similar accuracy to our historical data. From both datasets we extracted presence and absence locations for 145 plant species with sufficient representation in both time periods ($\geq$30 occurrences).

2.3. Climate data

We developed a parsimonious suite of four climatic predictor variables that we hypothesized would exhibit direct influence on species distributions and displayed low levels of correlation ($r < 0.6$). We used gridded 800-m resolution climate data from two time periods approximately representing 30-year time frames prior to vegetation data collection (1906–1935, 1976–2005). We used two climatic variables from the parameter-elevation regression on Independent Slopes Model (PRISM) (Daly et al., 2008) dataset, maximum temperature and minimum temperature. We also used two hydrologic variables that have been shown both theoretically and empirically to affect vascular plant distributions: mean annual actual evapotranspiration, and mean annual climatic water deficit (Stephenson, 1990, 1998). The hydrologic variables were developed using a modified climatic water balance model (Lutz et al., 2010) that accounts for snowmelt and soil moisture storage (Dobrowski et al., 2013).

2.4. Species distribution models

We used five widely applied statistical methods to model plant species distributions. These included two regression techniques (generalized linear models, GLM; generalized additive models, GAM), two machine learning techniques (generalized boosting machines, GBM; random forest, RF), and one classification technique (classification trees; CT). These techniques have been widely used in species distribution modeling applications, represent a broad range of analytical approaches, and were used in a recent study that concluded consensus forecasts yielded very accurate predictions of species distributions for climate change impact studies (Marmion et al., 2009). We built and calibrated models using a randomly selected 75% of the historical data. We tested the accuracy of predicted species distributions using the withheld 25% of the original data (internal evaluation) and using 100% of the contemporary data (independent evaluation). We computed model accuracy using the area under the curve (AUC) statistic (Fielding and Bell, 1997). Because AUC is often considered inadequate for assessing species distribution model performance (Lobo et al., 2008) we also converted probability of occurrence values into binary presence–absence predictions using the sensitivity–specificity equality approach to select prediction thresholds for each model (Cantor et al., 1999). Using these binary predictions we calculated false positive (FP) and false negative (FN) fractions. Although our FP and FN fractions may be biased due to mismatch in spatial resolution between our vegetation and climate data, we assumed such bias was minimal.

2.5. Consensus forecasts

We use the term "consensus forecast" to mean a single prediction that represents a measure of central tendency across a suite of individual models (Araújo et al., 2005b). In our case we chose to use the average predicted probability of occurrence across the five models as our consensus forecast. This is analogous to the Mean(all) approach used by Marmion et al. (2009), which they suggested was one of the best methods for developing consensus predictions. These predictions were evaluated using AUC and threshold dependent FP and FN fractions as described above. Prediction accuracy metrics between individual models and consensus forecasts were evaluated using paired *t*-tests to control for species effects. Variances of accuracy metrics between modeling approaches were compared using *F*-tests.

3. Results

Spatial patterns in predicted probability of occurrence varied among model algorithms and between time periods (see Fig. 1). When using internal evaluations, consensus forecasts had high mean AUC values across our 145 study species ($\bar{x} = 0.97$) that exceeded AUC values for all individual model algorithms except for