



Species distribution models predict range expansion better than chance but not better than a simple dispersal model

Marta Rodríguez-Rey^a, Alberto Jiménez-Valverde^b, Pelayo Acevedo^{b,c,d,*}

^a Area of Zoology, Department of Environmental Sciences, University of Castilla-La Mancha, E-45071 Toledo, Spain

^b Biogeography, Diversity, and Conservation Research Team, Faculty of Sciences, University of Málaga, E-29071 Málaga, Spain

^c Instituto de Investigación en Recursos Cinegéticos (CSIC-UCLM-JCCM), E-13071 Ciudad Real, Spain

^d CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBio Laboratório Associado, Universidade do Porto, 4485-661 Vairão, Portugal

ARTICLE INFO

Article history:

Received 18 October 2012

Received in revised form 22 January 2013

Accepted 26 January 2013

Keywords:

Autocorrelation

Deer species

Geographic distance

Predictive performance

Range expansion

Species distribution models

ABSTRACT

The evaluation of species distribution models (SDMs) is a crucial step; usually, a random subsample of data is used to test prediction capacity. This procedure, called cross-validation, has been recently shown to overestimate SDMs performance due to spatial autocorrelation. In the case of expanding species, there exists the possibility to test the predictions with non-random geographically structured data, i.e., a new data set which corresponds to the last occupied localities. The aim of this study was to evaluate the capacity of SDMs to predict the range expansion pattern of six free-living deer species in Great Britain and to assess whether SDMs perform better than a simple dispersal model – a null model that assumes no environmental control in the expansion process. Distribution data for the species prior to 1972 were used to train the SDMs (ENFA, MAXENT, logistic regression and an ensemble model) in order to obtain suitability maps. Additionally, the geographical distance to the localities occupied in 1972 was considered a proxy of the probability that a certain locality has to be occupied during an expansion process considering only dispersal (GD model). Subsequently, we analysed whether the species increased their ranges between 1972 and 2006 according to the estimated suitability patterns and whether or not SDMs predictions outperformed GD predictions. SDMs showed a high discrimination capacity in the training data, with the ensemble models performing the best and ENFA models the worst. SDMs predictions also worked better than chance in classifying new occupied localities, although differences among techniques disappeared and the predictions showed no difference with respect to GD. Spatial autocorrelation of both the environmental predictors and the expansion process may explain these results which illustrate that GD is a much more parsimonious model than any of the SDMs and may thus be preferable both for prediction and explanation. Overestimation of SDMs performance and usefulness may be a common fact.

Crown Copyright © 2013 Published by Elsevier B.V. All rights reserved.

1. Introduction

The use of species distribution modelling (SDM) has grown exponentially in the last two decades and has shown its potential in the fields of biodiversity conservation and ecosystem management (Franklin, 2009; Peterson et al., 2011). It is usually implicitly made the assumption that species distributions are in equilibrium with the environment (Franklin, 2009). However, in nature, equilibrium is the exception rather than the rule (Gaston, 2009). Paradoxically,

it is under such circumstances when species distribution models (SDMs) are particularly needed. For instance, species in the first stages of an invasion process is a clear example of such a scenario (Peterson, 2003). Explicitly recognizing non-equilibrium as a probable working scenario is important because it has methodological implications in SDM (Jiménez-Valverde et al., 2011).

At present, one of the greatest challenges in SDM is model evaluation (Vaughan and Ormerod, 2005; Lobo et al., 2008; Hijmans, 2012; Jiménez-Valverde, 2012; Jiménez-Valverde et al., 2013; Smith, 2013). Discrimination capacity is the property that obtains most of the attention, and it is generally accepted that model testing should be performed on data that have not been used in the training step; otherwise, model performance would be overestimated. To get this independent testing data, modellers usually perform the so-called cross-validation, i.e., they randomly divide the data into a training set and a validation set (Fielding and Bell, 1997). However,

* Corresponding author at: CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBio Laboratório Associado, Universidade do Porto, 4485-661 Vairão, Portugal. Tel.: +351 252 660411; fax: +351 252 661780.

E-mail addresses: pelayo.acevedo@gmail.com, pavevedo@irec.csic.es (P. Acevedo).

Hijmans (2012) compared the predictive performance (AUC values) of two SDM techniques with that of a purely distance-based method (the null model) and showed that because training and testing presences are closer to each other than training presences and testing absences, cross-validation still overestimated SDMs discrimination capacity.

When working with species that are in the process of expanding their geographic ranges, model evaluation can be performed using a new data set from the most recent time period (Araújo et al., 2005), which will correspond to the observed area of expansion instead of a random subsample of the data (Jiménez-Valverde et al., 2011). Usually, this new testing data set will be strongly spatially structured in some way. To avoid overestimating the distribution of the species, it has been suggested that spatial and environmental predictors should be considered together when modelling species in disequilibrium (De Marco et al., 2008; Sullivan et al., 2012). Including spatial variables in SDMs is desirable when the aim is to predict the most vulnerable localities that would be occupied in the short term during an expansion course (De Marco et al., 2008). Explicitly including the spatial variables in the modelling framework helps to maintain range cohesion (De Marco et al., 2008), and may be a way to account for spatially structured non-environmental factors, such as dispersal behaviour (Sullivan et al., 2012). These factors may significantly affect species distribution, especially in the initial phases of a range expansion process (Muñoz and Real, 2006).

The main objective of this study was to compare the predictive performance of four classic SDM techniques and a method that accounted for dispersal alone as a null model that assumes no environmental control in the expansion process. Using data about the recent range expansion of six species of deer in Britain, we attempted to answer the following two questions: (1) do SDMs provide significantly better-than-chance predictions of the species' range expansions? and (2) do SDMs perform better than the dispersal model?

2. Materials and methods

2.1. The species

There are six species of deer living wild in Britain: two are native (red deer *Cervus elaphus* and roe deer *Capreolus capreolus*); one is naturalized (introduced by the Normans around 1000 years ago; fallow deer *Dama dama*); and three are non-native species introduced between 50 and 150 years ago (sika deer *Cervus nippon*, Reeves' Muntjac *Muntiacus reevesi* and Chinese water deer *Hydropotes inermis*). It has been estimated that the six species have expanded their ranges in Britain between 1972 and 2002 (Ward, 2005), a tendency that is still occurring (Ward et al., 2008), and seems to be happening throughout Europe (Apollonio et al., 2010).

Species distribution data refer to a 10 km × 10 km grid superimposed on a map of Britain comprising 2800 grid squares. For modelling purposes, the study area was restricted to 2283 grid squares to avoid potential bias in modelling arising from including those smaller than 14 ha (coastline). Data on deer species distribution (Fig. 1) were obtained from Ward (2005) and supplemented with data from Acevedo et al. (2010). The idea was to replicate a common modelling exercise in which data for a species in disequilibrium are modelled and the geographic projection is interpreted as a map of potential ways for future colonization. To do so, the data from 1972 were used to train the models and to obtain suitability maps for the six species in Great Britain. Subsequently, we analysed whether the species increased their ranges between 1972 and 2006 according to the estimated suitability surfaces, i.e., whether they expanded their ranges occupying preferably those localities with higher suitability values (as estimated using the data from 1972).

Table 1
Variables used to model the distribution of the six wild ungulates.

Factors	Variables
Climate ^a	BIO1: annual mean temperature (°C × 10) BIO2: mean diurnal range (°C × 10) BIO3: isothermality (BIO2/BIO7) × 100 (°C × 10) BIO4: temperature seasonality (standard deviation × 100) BIO5: max temperature of warmest month (°C × 10) BIO6: min temperature of coldest month (°C × 10) BIO7: temperature annual range (BIO5–BIO6) (°C × 10) BIO8: mean temperature of wettest quarter (°C × 10) BIO9: mean temperature of driest quarter (°C × 10) BIO10: mean temperature of warmest quarter (°C × 10) BIO11: mean temperature of coldest quarter (°C × 10) BIO12: annual precipitation (mm) BIO13: precipitation of wettest month (mm) BIO14: precipitation of driest month (mm) BIO15: precipitation seasonality (coefficient of variation) BIO16: precipitation of wettest quarter (mm) BIO17: precipitation of driest quarter (mm) BIO18: precipitation of warmest quarter (mm) BIO19: precipitation of coldest quarter (mm)
Topography ^b	Range of altitude (m) Mean altitude (m above sea level) Max altitude (m above sea level) Mean slope (°) Max slope (°)
Spatial	Longitude (m) Latitude (m)

^a Bioclimatic variables were available at ~1 km² pixel width from the Worldclim project database (details in Hijmans et al., 2005).

^b Topographic variables were extracted from the European Digital Elevation Model carried out by the Shuttle Radar Topography Mission (European Environment Agency, 2000) with a spatial resolution of 100 m.

2.2. Predictors and modelling techniques

Twenty-four environmental predictors, grouped into two main factors (climate and topography), were chosen on the basis of their potential predictive power (Table 1). Although land use variables are usually taken into account when modelling the distribution of wild ungulates (e.g. Acevedo et al., 2010, 2011), they were not considered in this study because, to the best of our knowledge, land use information was unavailable for the training period. Two models were run with each SDM technique (see below); one using only the environmental variables as predictors and another one which also included latitude and longitude (spatial factor; Table 1) to account for spatially structured non-environmental factors (De Marco et al., 2008; Sullivan et al., 2012).

Four distinct SDM techniques selected to represent different levels of model complexity and data requirements (see Jiménez-Valverde et al., 2011) were used to model the occurrence of the species (see details in Appendix A): Ecological Niche Factor Analysis (ENFA; Hirzel et al., 2002), Maxent (Phillips et al., 2006; Phillips and Dudík, 2008), logistic regression (GLM; Hosmer and Lemeshow, 2000) and an ensemble model (EM; Thuiller et al., 2009) of four techniques (generalized linear models, multivariate adaptive regression splines, generalized boosted models and random forests).

2.3. The dispersal model

For each species, the geographical distance from each non-occupied locality in 1972 to the nearest occupied locality in the same period was calculated. This distance was considered a simple proxy of the probability that a certain location has to be occupied in an expansion process, i.e., it is more probable that a species will disperse to those localities that are closer to the species range limit. Geographic distances were standardized between 0 and 1 and their

Download English Version:

<https://daneshyari.com/en/article/6297227>

Download Persian Version:

<https://daneshyari.com/article/6297227>

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