Estuarine, Coastal and Shelf Science 180 (2016) 11–20

Contents lists available at ScienceDirect

Estuarine, Coastal and Shelf Science

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

Predicting fish species distribution in estuaries: Influence of species' ecology in model accuracy

Susana França*, Henrique N. Cabral

MARE – Marine and Environmental Sciences Centre, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749 – 016, Lisboa, Portugal

ARTICLE INFO

Article history:
Received 18 September 2015
Received in revised form
16 June 2016
Accepted 19 June 2016
Available online 21 June 2016

Keywords:
Species distribution models
Fish species
Model accuracy
AUC
Sensitivity
Specificity
Estuaries

ABSTRACT

Current threats to biodiversity, combined with limited data availability, have made for species distribution models (SDMs) to be increasingly used due to their ability to predict species' potential distribution, by relating species occurrence with environmental estimates. Often used in ecology, conservation biology and environmental management, SDMs have been informing conservation strategies, and thus it is becoming crucial to understand how trustworthy their predictions are. Uncertainty in model predictions is expected, but knowing the origin of prediction errors may help reducing it. Indeed, uncertainty may be related not only with data quality and the modelling algorithm used, but also with species ecological characteristics. To investigate whether the performance of SDM's may vary with species' ecological characteristics, distribution models for 21 fish species occurring in estuaries from the Portuguese coast were examined. These models were built at two distinct spatial resolutions and seven environmental explanatory variables were used as predictors. SDMs' accuracy was assessed with the area under the curve (AUC) of receiver operating characteristics (ROC) plots, sensitivity and specificity. Relationships between each measure of accuracy and species ecological characteristics were then examined. SDMs of the fish species presented small differences between the considered scales, and predictors as latitude, temperature and salinity were often selected at both scales. Measures of model accuracy presented differences between species and scales, but generally higher accuracy was obtained at smaller spatial scales. Among the ecological traits tested, species feeding mode and estuarine use functional groups were the most influential on the performance of distribution models. Habitat tolerance (number of habitat types frequented), species abundance, body size and spawning period also showed some effect. This analyses will contribute to distinguish, based on species ecological characteristics, between species whose occurrence is reliably or less reliably predicted by distribution models. Fish species assigned to the benthivores feeding group, and using estuaries opportunistically (either as feeding or nursery grounds) seem to be more accurately predicted by SDM's.

1. Introduction

Current threats to biodiversity combined with the difficulty in obtaining detailed species inventories, have made for species distribution models (SDMs) to be increasingly used in the fields of ecology and conservation biology (Guisan and Thuiller, 2005; Guisan et al., 2013). SDMs aim to predict species environmental requirements and potential distribution based on correlations of species occurrence with environmental attributes, which may have a direct or indirect effect on their establishment and survival (Guisan and Zimmermann, 2000; Elith and Graham, 2009; Franklin, 2010).

SDMs play a crucial role in supporting spatial conservation decision making, as researchers are constantly asked for recommendations about conserving biodiversity based on limited species-distribution data (Ortega-Huerta and Peterson, 2004; Guisan et al., 2013). One fundamental issue in their development is the assessment of predictive accuracy: as SDMs influence policy management and consequently conservation strategies, it is important to know how trustworthy and reliable their predictions are (Fielding and Bell, 1997; Guisan and Thuiller, 2005). Model reliability and prediction errors depend on factors such as, the methodological aspects (comprising limitations imposed by the algorithm used and data-gathering process) (Fielding and Bell,

* Corresponding author.
E-mail address: sofranca@fc.ul.pt (S. França).

http://dx.doi.org/10.1016/j.ecss.2016.06.010
0272-7714/

1997; Segurado and Araújo, 2004) and the nature of data availability (including the processes arising directly from the species' ecology) (McPherson and Jetz, 2007).

Species ecological characteristics (e.g. range size and ecological specialization) have been shown to influence model performance (Segurado and Araújo, 2004; Brotons et al., 2004; McPherson et al., 2004; Elith et al., 2006), either by influencing the quality of the available data (Boone and Krohn, 1999) or by making the relationship between species' occurrence and environmental factors more complex and difficult to be assessed (Brotons et al., 2004; McPherson and Jetz, 2007). Previous studies on the influence of species ecological characteristics on model accuracy yielded different results, with no effect reported by some authors (Elith and Burgman, 2002; Huntley et al., 2004) and significant impacts caused by ecological characteristics found (Stockwell and Peterson, 2002; Segurado and Araújo, 2004).

A central and recurrent problem in SDM building is identifying the appropriate scale for modelling (Wiens, 2002). Conceptually, there is no single natural scale at which ecological patterns should be studied. Rather, the most appropriate scale should be chosen based on the study goals, the system, and available data (Elith and Leathwick, 2009). Commonly, the scale is imposed through data choice or model structure, as the chosen predictors often vary, having effects on species at markedly different spatial scales. One way to assess the importance and effects of scale on SDMs performance is to change the scale at which the models are built and to measure the qualitative and quantitative effects on their performance (Guisan et al., 2007). Studies relating the effect of species' ecological characteristics versus scale on models accuracy and performance are scarce, and these topics have been mostly addressed for terrestrial species (Elith and Leathwick, 2009; Guisan et al., 2013).

Estuaries are important coastal ecosystems which are among some of the most biologically productive areas on Earth (Costanza et al., 1997). As transitional systems, estuaries establish links with marine and freshwater ecosystems (Beck et al., 2001), which will lead to persistent environmental fluctuations. This will limit the occurrence of several fish species, as considerable physiological demands are required. Nevertheless, these ecosystems play important nursery roles for several fish species, presenting high prey availability, refuge from predators and good conditions for a rapid growth (Beck et al., 2001). Furthermore, the dynamic nature of estuaries (which depend greatly on the season) and their nursery role highlight the differential use of these ecosystems by many fish species. This brings difficulties in studying these species distribution and building their SDMs is thus, challenging. More accurate predictions of fish species distributions in estuarine systems may be obtained by taking their ecological characteristics into consideration. In this study, fish species ecological characteristics such as functional guilds covering estuarine use and feeding mode, habitat tolerance, spawn period, species abundance, vulnerability and body size were taken into account, as they might influence species distribution within estuaries and thus influence the accuracy at which fish species distributions are predicted.

In order to determine which of these ecological characteristics explain mostly the accuracy of model predictions, the relationship between SDMs accuracy and those characteristics were examined. To reach this goal, firstly SDMs were built for 21 fish species occurring in estuaries from the Portuguese coast and their accuracy values were then related with species ecological characteristics. This will contribute to distinguish, based on their ecological characteristics, between species whose occurrence will be more or less reliability predicted by distribution models, which may be crucial for future assessments of conservation planning. As the spatial scale at which predictions are made and models are built are known to influence models' reliability, two different spatial resolutions were used to build the models in order to determine at which scale predictions were more accurate.

2. Material and methods

2.1. Overview

Generalized linear models were used to relate the presence and absence of 21 fish species in nine estuaries from the Portuguese coast with environmental features and variables measured locally. The accuracy of these distribution models was measured on data withheld from model training, using three metrics: the area under the curve (AUC) of receiver operating characteristics (ROC) plots, sensitivity and specificity. Generalized linear models were then used to examine the relationships between model accuracy and species ecological characteristics.

2.2. Species distribution data

Nine estuarine systems along the Portuguese coast were considered in the present study: Minho, Douro, Ria Aveiro, Mondego, Tejo, Sado, Mira, Ria Formosa and Guadiana (Fig. 1). These systems differ substantially in terms of their geomorphologic and hydrologic characteristics: Tejo and Sado are large systems with areas of 320 km^2 and 180 km^2, respectively, while Mira is the smallest with 5 km^2. Mean river flow values are considerably higher in the Tejo estuary (300 $m^3\ s^{-1}$) and this system also presents the largest estuary mouth width (5.3 km). Ria Aveiro and Ria Formosa are shallow coastal lagoon systems with large intertidal areas. Shallow areas are a common feature in all the estuarine systems, with mean depths varying between 1 and 6 m (Table 1).

Sampling was performed in order to cover simultaneously the whole estuarine gradient and the different estuarine habitats. The considered habitats were: tidal freshwater (in tidal estuarine areas, this is the zone upstream of saline influence), salt marsh (intertidal, sediment based, macrophyte-dominated, saline-influenced habitats), intertidal soft substratum (areas of unvegetated intertidal habitats, lying between the highest and lowest tides, and composed predominantly of sediments from fine silt to coarse sands), subtidal soft substratum (permanently subtidal unvegetated habitats, composed predominantly of sediments ranging from fine silts to coarse sands) and seagrass beds (vegetated habitats, based on soft substrata, dominated by halophytic macrophytes adapted to complete and continuous submergence in water of low to high salinity). Present habitat classification was adapted from Pihl et al. (2002).

Fish assemblages were sampled in May and July 2006 and seasonally in 2009 (January, April, July and October). Sampling was conducted during the night, using a 2 m beam trawl with a tickler chain and 5 mm mesh size in the cod end. Hauls were towed at a constant speed (0.8 m s^{-1}) and lasted for 10 min whenever there was sufficient habitat area available (never lasted less than 8 min). At the beginning of each haul, salinity, dissolved oxygen in the water (%), water temperature (ºC) and conductivity (mS/cm) were measured with a multi-parameter probe (WTW ProfiLine Oxi 197) and depth (m) was also registered. Three replicates of sediment were collected using a Van Veen grab (0.05 m^2) for the determination of mean mud content in the sediment (percentage of dry sediment not retained in a 0.063 mm calibrated sieve). The number of tows differed for each estuary according with the system dimension and the presence of different habitats, which differ in number, for the sampled systems. A total of 770 tows were performed.