



Original research paper

Accounting for uncertainty in predictions of a marine species: Integrating population genetics to verify past distributions



Rosa M. Chefaoui*, Ester A. Serrão

Centro de Ciências do Mar (CCMAR), CIMAR Laboratório Associado, Universidade do Algarve, Campus de Gambelas, 8005-139 Faro, Portugal

ARTICLE INFO

Article history:

Received 12 September 2016
 Received in revised form 27 May 2017
 Accepted 7 June 2017
 Available online 14 June 2017

Keywords:

Ecological niche modelling
 Genetic diversity
 Last glacial maximum
 Ocean general circulation models
 Sampling bias
 Threshold

ABSTRACT

We develop a new perspective on the uncertainties affecting the predictions of coastal species distributions using patterns of genetic diversity to assess the congruence of hindcasted distribution models. We model the niche of the subtidal seagrass *Cymodocea nodosa*, for which previous phylogeographic findings are used to contrast hypotheses for the Last Glacial Maximum (LGM) in the Mediterranean and adjacent Atlantic coastal regions. We focus on amelioration of sampling bias, and explore the influence of other sources of uncertainty such as the number of variables, Ocean General Circulation Models (OGCMs), and thresholds used. To do that, we test geographical and environmental filtering of presences, and a species-specific weighted filter related to political boundaries for background data. Contrary to our initial hypothesis that reducing sampling bias by means of geographical, environmental or background filtering would enhance predictive power and reliability of the models, none of these approaches consistently improved performance. These counter-intuitive results might be explained by the higher relative occurrence area (ROA) inherent to linear coastal study areas in relation to terrestrial regions, which may cause worse predictions and, thus, higher variability among models. We found that the Ocean General Circulation Models (OGCMs), the threshold and, to a smaller extent, the number of variables used, conditioned greatly the variability of the predictions in both accuracy and geographic range. Despite these uncertainties, all models achieved the goal of identifying long-term persistence regions (glacial refugia) where the highest genetic diversity for *Cymodocea nodosa* is found nowadays. However, only the CCSM corroborated the hypothesis, raised in previous studies, of a vicariant process in shaping the species' genetic structure.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Although modelling the niche of species has proven to be an efficient approach to ecological, conservational and biogeographical questions during the last decades (see Guisan and Thuiller, 2005; Araújo and Peterson, 2012), a growing number of studies focused on the terrestrial realm highlights the importance of assessing the sources of uncertainty affecting species distribution models (SDMs) in order to obtain more reliable predictions. The potential sources of uncertainty may be related to a large list of factors ranging from the data included in the models to the algorithm used (see e.g. Rocchini et al., 2011; Beale and Lennon, 2012; Gould et al., 2014 for a review). Many of the studies on uncertainty in ecological niche modelling pay particular attention to sampling bias in occurrence data caused

by proximity to cities, rivers, roads or conservation reserves (e.g. Reddy and Dávalos, 2003; Kadmon et al., 2004) or historical bias (Hortal et al., 2008), as they may result in inaccurate estimations of niche. More recent attention has focused on the provision of methodologies to ameliorate geographic bias by filtering (or “thinning”) occurrences to avoid clumping and autocorrelation (Veloz, 2009; Beck et al., 2014; Aiello-Lammens et al., 2015), or to correct the derived bias in the environmental space which improved performance of the models in comparison to the use of geographic filters (Varela et al., 2014; de Oliveira et al., 2014). Finally, filters for background data have also been applied (Phillips et al., 2009; Kramer-Schadt et al., 2013), as the location of pseudo-absences or background data also affects predictions (Zaniewski et al., 2002; Chefaoui and Lobo, 2008; Merow et al., 2013).

Despite previous efforts to reduce uncertainty, much of the research up to now referred to biases shown by terrestrial species, while uncertainties related to the study of marine taxa have been poorly investigated. The relatively recent availability of satellite data on marine environmental variables has driven an increase in

* Corresponding author.

E-mail addresses: rosa.chefaoui@gmail.com (R.M. Chefaoui), eserrao@ualg.pt (E.A. Serrão).

the use of SDMs in this realm and, consequently, studies suited to its peculiarities are needed. One particularly relevant distinction is that most important marine species are coastal and therefore have a distribution that is more linear than bidimensional. Regarding sampling bias, coastal studies are not biased by the same geographic elements as terrestrial do (e.g. distance to roads). The proximity to cities or research institutions with diving centers may affect sampling effort, and at a large scale, the observed under-surveyed regions are coincident with countries with limited investment in research and political instability such as those located in coastal regions of North Africa (Chefaoui et al., 2016). In the Global Biodiversity Information Facility (GBIF, www.gbif.org/), it has been shown that more records of the common Eurasian butterfly are available from developed countries though the real occurrence of the species is higher in less developed ones (Beck et al., 2014). These differences among countries might be even more prevalent concerning subtidal marine habitats, where sampling marine campaigns are more expensive and require more infrastructure and technical expertise.

The unavailability of predictors is another source of uncertainty affecting niche projections to the past for marine species in comparison with terrestrial ones. Uncertainties affecting the General Circulation Models (GCMs) pertaining to the Coupled Model Inter-comparison Project (CMIP) have been reported both for oceanic variables (Wang et al., 2013) and for atmospheric ones (Svenning et al., 2008; Braconnot et al., 2012; Varela et al., 2015). However, although those uncertainties are common to both marine and terrestrial realms and despite the improved reconstructions of climate during the LGM (Last Glacial Maximum; 23 000–18 000 years BP), there is a scarcity of ocean paleoclimatic variables corresponding to the Ocean GCMs (OGCMs) for the LGM in comparison with the models of atmospheric variables used in terrestrial studies. This limits the possibilities to use a wider set of oceanic predictors for the LGM to hindcast marine niches (see e.g. Chefaoui et al., 2017). There is also an observable incompleteness of data for all seas and oceans, as important gaps are found in some models (e.g. Black Sea), and some OGCMs are available just at a coarse resolution ($\sim 2^\circ$).

To provide a new perspective for marine coastal species, we explore these uncertainties in SDMs for the subtidal seagrass *Cymodocea nodosa* (Ucria) Ascherson. Although *C. nodosa* is the most common seagrass in the eastern Mediterranean, no exact mapping has been carried out and scanty reports of its distribution exist from some countries (e.g., Egypt, Lebanon...) in comparison with the western basin (Green and Short, 2003). We will estimate if sampling bias linked to geographical boundaries is affecting data input of *C. nodosa* to calibrate niche models and, consequently, derived past projections. If so, we will try to ameliorate sampling bias under the expectation that the accuracy of niche models for coastal species may be improved similarly as terrestrial ones by the use of filters for species data. Thus, in addition to testing geographical and environmental filters for presences, we will seek to ameliorate a possible political bias by creating a weighted probability filter for background data in accordance to political boundaries. This “weight” filter will combine an estimation of sampling effort for *C. nodosa* in relation to the available distribution records of similar marine plants from GBIF, with the developmental level of each country.

We will also test if the lack of oceanic predictors pertaining to climate simulations for the LGM might be hindering the reliability of SDM-based hindcasting marine studies. Under the assumption that differences found using current climatic predictors may prevail through niche hindcasting, we give an account of the uncertainty caused by the reduced availability of oceanic LGM variables examining performance differences among current climate models obtained using nested groups of predictors varying in number. Finally, we will also take into consideration the effect on our predic-

tions of the threshold used to validate and transform probabilities into binary outputs, which has been identified as a yet unexplored source of uncertainty modelling species range shifts (Nenzén and Araújo, 2011).

To assess these uncertainties on our current and LGM models for *C. nodosa*, we will take advantage of previous ecological (Chefaoui et al., 2015) and genetic (Alberto et al., 2008; Masucci et al., 2012) findings allowing us to assess congruence between the hypotheses supported by different approaches. Species distribution shifts throughout the Quaternary leave genetic signatures on populations (Hewitt, 2004). High congruence between the distribution of present intraspecific genetic diversity and regions of long-term population persistence since the LGM have been found for several marine species (Assis et al., 2014; Neiva et al., 2014; Assis et al., 2016; Chefaoui et al., 2017). Thus, finding a congruence between genetic information and SDMs for the LGM could help us identify possible uncertainties in the SDMs. We will contrast our hindcast results with the predictions made based on the present genetic structure of the species along its distribution, namely environmental barriers limiting gene flow inferred by Alberto et al. (2008) and Masucci et al. (2012) for the species across the Mediterranean and Atlantic. Alberto et al. (2008) identified four regions with a strong genetic structure, and two potential imprints of vicariance located in the range areas with presumably higher sea surface temperature (SST), namely the low-latitude Atlantic (AL) and the Eastern Mediterranean (EM) regions (Fig. 1). Masucci et al. (2012) inferred the gene flow directionality among *C. nodosa* populations. We will examine if these hypothesized glacial refugia (persistence regions with high genetic diversity) are supported by the LGM projections obtained by all models to evaluate the most accurate prediction on the basis of independent genetic data.

Our goals are to examine for the first time how sampling bias, and other sources of uncertainty such as the availability of variables, the Ocean General Circulation Model (OGCM), and the threshold used, affect the predictions of a coastal marine species, in this case a subtidal seagrass, comparing predictions derived from our models with previous predictions derived from data on genetic diversity throughout the species range. This paper investigates those uncertainties for current and Last Glacial Maximum (LGM) predictions, using the observed geographical distribution of extant genetic diversity for assessment of congruence of the predictions.

2. Methods

2.1. Effect of the number of variables on performance of models under current conditions

We used 299 records of *Cymodocea nodosa* compiled from the literature, the Global Biodiversity Information Facility (GBIF, www.gbif.org/), and Algaebase (Guiry and Guiry, 2014), covering the entire species range (Mediterranean Sea, North-East Atlantic coasts and the Black Sea) at any depth along its known distribution (0–35 m depth). A total of 210 presence cells remained after georeferencing data to a $0.083^\circ \times 0.083^\circ$ (~ 9.2 km) grid resolution (Fig. 1). To assess the influence of a reduction in the number of variables on performance, we compared different sets of nested predictors under current conditions, using a specific SDM for this purpose (see Fig. 2). The complete set of 18 variables was comprised of the environmental variables and landscape metrics previously found to be the best for modelling the niche of *C. nodosa* in Chefaoui et al. (2016). Environmental predictors were: sea surface temperature (SST) of winter and summer, diffuse attenuation coefficient (Kd), wave height, nitrate, phosphate, pH, photosynthetically available radiation (PAR), and salinity (Appendix A in the Supplementary material, Table A.1). Landscape metrics were: mean edge contrast index dis-

Download English Version:

<https://daneshyari.com/en/article/5742134>

Download Persian Version:

<https://daneshyari.com/article/5742134>

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