Ecological Modelling 225 (2012) 133–145

Contents lists available at SciVerse ScienceDirect

Ecological Modelling

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

Modelling commercial fish distributions: Prediction and assessment using different approaches

Miranda C. Jones[a,b,c,*], Stephen R. Dye[b], John K. Pinnegar[b], Rachel Warren[a,c], William W.L. Cheung[a,b,d]

[a] School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK
[b] Centre for Environment, Fisheries and Aquaculture Science (CEFAS), Lowestoft NR33 0HT, UK
[c] Tyndall Centre for Climate Change Research, Norwich, UK
[d] Fisheries Centre, University of British Columbia, Vancouver, BC, Canada

ARTICLE INFO

Article history:
Received 8 August 2011
Received in revised form 21 October 2011
Accepted 1 November 2011
Available online 14 December 2011

Keywords:
Species distribution modelling
North Sea
Range maps
Marine fishes
Model comparison

ABSTRACT

Species distribution models are important tools to explore the effects of future global change on biodiversity. Specifically, AquaMaps, Maxent and the Sea Around Us Project algorithm are three approaches that have been applied to predict distributions of marine fishes and invertebrates. They were designed to cope with issues of data quality and quantity common in species distribution modelling, and especially pertinent to the marine environment. However, the characteristics of model projections for marine species from these different approaches have rarely been compared. Such comparisons provide information about the robustness and uncertainty of the projections, and are thus important for spatial planning and developing management and conservation strategies. Here we apply the three commonly used species distribution modelling methods for commercial fish in the North Sea and North Atlantic, with the aim of drawing comparisons between the approaches. The effect of different assumptions within each approach on the predicted current relative habitat suitability was assessed. Predicted current distributions were tested following data partitioning and selection of pseudoabsences from within a specified distance of occurrence data. As indicated by the test statistics, each modelling method produced plausible predictions of relative habitat suitability for each species, with subsequent incorporation of expert knowledge generally improving predictions. However, because of the differences between modelling algorithms, methodologies and patterns of relative suitability, comparing models using test statistics and selecting a 'best' model are not recommended. We propose that a multi-model approach should be preferred and a suite of possible predictions considered if biases due to uncertainty in data and model formulation are to be minimised.

1. Introduction

Many pressures are currently affecting the marine environment and driving change in species composition and distribution. Fisheries are removing fishes at a rate considered to be unsustainable (Pauly et al., 2002), while essential habitat is being damaged or destroyed, for example through sand and gravel extraction, or chemically altered through release of endocrine-disrupting substances. Furthermore, concern over the impact of climate change on marine ecosystems is increasing (Root and Rosenzweigk, 2003), with longer term shifts in mean environmental conditions and climatic variability moving outside the bounds within which adaptations in marine communities have previously been associated (e.g. Beaugrand, 2004; King, 2005). The altered abundances and novel distributions resulting from these ocean-atmospheric changes (e.g. Beaugrand, 2009; Perry et al., 2005; Southward et al., 1995; Stebbing et al., 2002) may severely change the biological and environmental functioning of ecosystems or food webs, the goods and services derived from them, and conservation and resource management.

Species distribution modelling is widely used to study and predict the ecological effects of climate change (e.g. Hijmans and Graham, 2006; Beaumont and Hughes, 2002; Pearson and Dawson, 2003; Thuiller et al., 2008; Cheung et al., 2009). It uses statistically or theoretically derived response surfaces to relate observations of species occurrence or known tolerance limits to environmental predictor variables (Guisan and Zimmermann, 2000), thereby predicting a species' range as the manifestation of habitat characteristics that limit or support its existence at a particular location. It is thus grounded in ecological niche theory. The environmental conditions under which a species can survive and grow and

* Corresponding author at: School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK.
E-mail addresses: miranda.jones@uea.ac.uk (M.C. Jones), stephen.dye@cefas.co.uk (S.R. Dye), john.pinnegar@cefas.co.uk (J.K. Pinnegar), r.warren@uea.ac.uk (R. Warren), w.cheung@fisheries.ubc.ca (W.W.L. Cheung).

0304-3800/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2011.11.003

which therefore define the ecological properties of a species are described as the fundamental ecological niche (Hutchinson, 1957) or a species' potential distribution. The area within a fundamental niche into which a species is restricted due to the effects of competition and other biotic interactions is described as its realized niche (Austin et al., 1990; Guisan and Zimmermann, 2000), or distribution. To make use of the diversity of available data, a wide range of species distribution models (SDMs) have been proposed [see Guisan and Thuiller (2005) and Franklin (2009) for an overview], approaches varying widely in data requirements, mechanisms used and model performance (Guisan and Zimmermann, 2000; Elith et al., 2006; Austin, 2007; Wisz et al., 2008). The extent to which models are able to capture a species' realized or fundamental niche may thus vary depending on the modelling approach or data requirements.

When choosing and applying an SDM, it is therefore important to understand its performance, assumptions, characteristics and uncertainties, as well as how these might be affected by data availability and quality. Ideally, an SDM is developed from the relationship between direct or indirect environmental predictors and datasets of species presence and absence obtained by targeted surveys. Comprehensive data are, however, seldom available and instead frequently represent a restricted, patchy or biased view of species' distributions, leading to problems when data-driven modelling techniques are used to generate distribution predictions. Furthermore, it has been suggested that presence–absence data attribute superior performance, for example as measured by test statistics, to an SDM and thus a more reliable prediction (Brotons et al., 2004; Hirzel et al., 2001; Martinez-Meyer, 2005; Lobo et al., 2008). This would not be the case, however, if absence at a particular location is caused by factors not included in the model, such as dispersal limitations, biotic interactions or incorrect assessment (Pearson et al., 2007; Pearson and Dawson, 2003). Distributions predicted from recorded species' occurrence (presence) only may thus be more suitable for constructing models of potential habitat. Several studies show that SDM model accuracy decreases and variability in predictive accuracy increases with decreasing size of the species occurrence dataset (Wisz et al., 2008; Hernandez et al., 2006; Kadmon et al., 2003; Stockwell and Peterson, 2002). These issues of data paucity and quality are especially pertinent in the marine environment (Kaschner et al., 2006; MacLeod et al., 2005).

Model complexity is another important factor affecting the performance of SDMs. Complex models are suggested to be more effective (Elith et al., 2006; Tsoar et al., 2007; Wisz et al., 2008) and more accurate at finer resolutions (Kimmins et al., 2008). However, including more parameters or fitting complex response curves may result in a model that generalizes poorly (Drake et al., 2006), becoming less applicable to areas at a broader scale. Greater complexity also often reduces model transparency, which is important for the effective testing and reviewing of model outputs and soliciting additional information to improve model predictions. The complexity and transparency of a selected model may therefore depend not only on its perceived robustness but also on the specific application and the community by which it is being implemented.

Maxent, AquaMaps and the Sea Around Us Project model are three approaches commonly used to model distributions of marine fishes and invertebrates (Kaschner et al., 2008; Ready et al., 2010; Close et al., 2006; Bigg et al., 2008; Cheung et al., 2009). The Maxent software package (Phillips et al., 2006; Phillips and Dudík, 2008) was designed to overcome the problems of small sample sizes in presence-only datasets (Pearson et al., 2007). The AquaMaps procedure, based on a Relative Suitability Model (Kaschner et al., 2006), and the Sea Around Us Project model were also designed to overcome the lack of data and knowledge for many marine species. Generative modelling approaches, such as Maxent, may, however, be more vulnerable to biases from the skewed distribution of sampling effort present in many 'opportunistically' collected datasets, especially those with limited data-points. In these instances, discriminative methods (defined here as distribution models which restrict a species distribution, from a potential extent that encompasses the entire study area, based on a set of filters determined by known parameters or habitat preferences), such as that developed by the Sea Around Us Project (Close et al., 2006), might produce the more valid results. The incorporation of 'expert information' may also overcome this problem (Ready et al., 2010). Expert information may be defined as "habitat use information that is not directly available as raw data; published information about habitat use or preference that is based on quantitative investigations of species occurrence in relation to environmental knowledge" (Ready et al., 2010). It may be incorporated into a modelling procedure in various forms of knowledge such as species' behaviour, known depth range or geographic limits.

This study aims to assess the abilities of three statistical modelling approaches, representing a spectrum of theoretical frameworks and data-requirements, to predict current distributions of a range of marine species. Mentioned above, these are the correlative, presence-only modelling approaches Maxent (Phillips et al., 2004; http://www.cs.princeton.edu/~schapire/maxent) and AquaMaps (Kaschner et al., 2008; Ready et al., 2010; http://www.aquamaps.org), and the discriminative approach developed for the Sea Around Us Project (Close et al., 2006; http://www.seaaroundus.org). The comparison not only focuses on the perceived value of a modelling procedure as indicated by test statistics, but also considers the usability and practical application of the approaches and their results.

2. Methods

2.1. Model construction

2.1.1. Maxent

Maxent (Phillips et al., 2004) uses a generative approach (Phillips et al., 2006) to estimate the environmental co-variates conditioning species presence and bases the final prediction on the principle of maximum entropy. This specifies that the best approximation of an unknown distribution is the probability distribution with maximum entropy, subject to the constraints imposed by the sample of species presence observations (Phillips et al., 2006). Maxent has been shown to compete well with alternative approaches (Pearson et al., 2007; Phillips et al., 2006), perform better than classical presence-only methods (Elith et al., 2006) and perform well with small sample sizes (Pearson et al., 2007). Models were constructed using Maxent (version 3.3.3e) with default parameters for a random seed, regularization parameter (1, included to reduce over-fitting), maximum iterations (500), convergence threshold (0.00001) and maximum number of background points (10,000 points which have not been recorded as present). Selection of environmental features and their relative contribution to each iteration of the model was also carried out automatically.

2.1.2. AquaMaps

The AquaMaps approach to modelling species' distributions was based on a global distribution tool for marine mammals (Kaschner et al., 2006), and has now been applied to a large number of marine fishes (see FishBase, Froese and Pauly, 2011). AquaMaps uses simple, numerical descriptors of species relationships with environmental variables to predict distributions from publically available, global occurrence databases. This methodology does not allow complex, non-linear interactions to be fitted between predictors, but aims for transparency and understanding in the wider, non-modelling, community while also explicitly promoting incorporation of expert judgement.