Estuarine, Coastal and Shelf Science 121-122 (2013) 8–19

Contents lists available at SciVerse ScienceDirect

Estuarine, Coastal and Shelf Science

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

Influence of model selection on the predicted distribution of the seagrass *Zostera marina*

Anna-Leena Downie [a,c,*], Mikael von Numers [b], Christoffer Boström [b]

[a] *Marine Research Centre, Finnish Environment Institute, PO Box 140, 00251 Helsinki, Finland*
[b] *Åbo Akademi University, Environmental and Marine Biology, Artillerigatan 6, FIN-20520 Turku, Finland*
[c] *Centre for Environment Fisheries and Aquaculture Science, Pakefield Road, Lowestoft NR33 0HT, UK*

ARTICLE INFO

Article history:
Received 29 June 2012
Accepted 30 December 2012
Available online 28 January 2013

Keywords:
Zostera marina
habitat suitability
GAM
Maxent
spatial predictions
ecological niche
Baltic Sea

ABSTRACT

There is an increasing need to model the distribution of species and habitats for effective conservation planning, but there is a paucity of models for the marine environment. We used presence (131) and absence (219) records of the marine angiosperm *Zostera marina* L. from the archipelago of SW Finland, northern Baltic Sea, to model its distribution in a 5400 km^2 area. We used depth, slope, turbidity, wave exposure and distance to sandy shores as environmental predictors, and compared a presence-absence method: generalised additive model (GAM), with a presence only method: maximum entropy (Maxent). Models were validated using semi-independent data sets. Both models performed well and described the niche of *Z. marina* fairly consistently, although there were differences in the way the models weighted the environmental variables, and consequently the spatial predictions differed somewhat. A notable outcome from the process was that with relatively equal model performance, the area actually predicted in geographical space can vary by twofold. The area predicted as suitable for *Z. marina* by the ensemble was almost half of that predicted by the GAM model by itself. The ensemble of model predictions increased the model predictive capability marginally and clearly shifted the model towards a more conservative prediction, increasing specificity, but at the same time sacrificing sensitivity. The environmental predictors selected into the final models described the potential distribution of *Z. marina* well and showed that in the northern Baltic the species occupies a narrow niche, typically thriving in shallow and moderately exposed to exposed locations near sandy shores. We conclude that a prediction based on a combination of model results provides a more realistic estimate of the core area suitable for *Z. marina* and should be the modelling approach implemented in conservation planning and management.

1. Introduction

Knowledge of where species and habitats occur and which factors limit or threaten their distribution is fundamental for both science and management. Currently there is a drive towards ecosystem-based management and spatial planning of the marine environment, see e.g. Gilliland and Laffoley (2008). If management and planning are to be successful, there is an overriding need to locate and delineate habitats and their communities across multiple scales (Cogan et al., 2009). In order to establish functional marine protected areas, information regarding the location of rare, sensitive or functionally important habitat-builders is central. Unfortunately, spatially explicit information of many habitat-forming species, such as corals, macroalgae and seagrasses, is still poor or lacking in many regions, thus preventing effective management and protection. These issues combined with often costly and relatively inefficient data collection methods in broad-scale field efforts have led to a rapid development, and an increasing application of remote sensing techniques, acoustic surveys, Geographic Information Systems (GIS), and various predictive modelling approaches in marine and coastal zone studies (Zacharias et al., 1999; Lathrop et al., 2001; Leathwick et al., 2008; Robinson et al., 2011; Brown et al., 2012).

Marine flowering plants, or seagrasses, globally comprise ca. 60 species distributed over 12 genera. Seagrasses form extensive meadows in many coastal areas, and provide several important ecological goods and services, including organic carbon production and export, sediment filtration, trapping and stabilization, nutrient cycling and provision of food and shelter for diverse faunal and floral communities (Hemminga and Duarte, 2000; Spalding et al., 2003). However, despite these important ecosystem functions,

* Corresponding author. Centre for Environment Fisheries and Aquaculture Science, Pakefield Road, Lowestoft NR33 0HT, UK.
E-mail address: anna.downie@cefas.co.uk (A.-L. Downie).

0272-7714/$ – see front matter
http://dx.doi.org/10.1016/j.ecss.2012.12.020

seagrass ecosystems still lack efficient protection, and multiple stressors, including global warming, increased nutrient loading, commercial fishing, physical disturbance and disease outbreaks, continue to decrease seagrass meadows throughout their distribution range (Hughes et al., 2009; Waycott et al., 2009; Boström et al., 2011).

Consequently, seagrasses have been identified as a habitat of conservation importance and effort is being expended to mapping seagrass meadows using multiple methodologies. In the management context, spatially explicit information on marine plant communities is important to identify areas where potentially large occurrences or diverse habitats coincide with local anthropogenic impacts, such as ferry routes, dredging, sand extraction and oil pollution (Lathrop et al., 2001; Zacharias and Gregr, 2005). Furthermore, knowledge of the spatial distribution of potential seagrass habitats is a prerequisite for successful transplantations or reintroductions (Van Katwijk et al., 2000; Short et al., 2002).

Remote sensing is a very effective tool to map seagrass meadows in intertidal areas and in subtidal areas with good visibility (e.g. Pu et al., 2012), but is of reduced utility in turbid conditions. Studies using acoustic methodologies, including sidescan sonar and multibeam echosounders have provided promising new tools for mapping seagrass meadows at a local scale (e.g. Paul et al., 2011; Van Rein et al., 2011; Micallef et al., 2012). Because of the high cost and other limitations of direct mapping methodologies, distribution modelling has been put forward as an alternative approach to mapping seagrasses for management purposes (Kelly et al., 2001; Bekkby et al., 2008; Grech and Coles, 2010; Valle et al., 2011).

The species distribution modelling (SDM) and habitat suitability (HS) modelling tools, with a long history of use in the terrestrial environment (Guisan and Zimmermann, 2000; Elith and Leathwick, 2009), are currently gaining popularity in marine areas (Robinson et al., 2011). An advantage of the modelling approach is that it has a dual role. Predictive models of seagrass distribution may provide useful tools for scientific research as well as management and conservation efforts of valuable seagrass habitats. Such maps may serve as a baseline for predicting shifts or declines in species range in response to e.g. climate change scenarios, contributing to studies in many areas of applied ecology. For seagrasses, prospective changes include e.g. changed growth and survival rates as a result of increasing seawater temperatures (Reusch et al., 2005), spatial (both vertical and horizontal) changes in distribution due to increased turbidity (Krause-Jensen et al., 2001), or range shifts in response to reduced salinity levels due to increased precipitation caused by climate change (Dippner et al., 2008). Habitat suitability maps may also provide useful tools in restoration efforts (Bos et al., 2005), and can guide invasion ecology by identifying areas potentially suitable to alien species (Hirzel et al., 2004). The value of predictive maps to study ecology and potential scenarios of change depends on their accuracy and ability to produce consistent results.

With the proliferation of SDMs, there are increasing studies using a wide range of methods. The large list of available methods from which to choose is continuously growing. Here we concentrate on two currently very commonly used methods. One of the most consistently used methods, with good performance in comparison studies (e.g. Elith et al., 2006), is generalized additive modelling (GAM, Hastie and Tibshirani, 1990), which has been applied to predicting species distributions of, among others, marine fish (e.g. Stoner, 2001; Florin et al., 2009; Sundblad et al., 2009), coral reefs (e.g. Garza-Pèrez et al., 2004) and benthic macrophytes (e.g. Bekkby et al., 2008; Sandman et al., 2008; Bekkby et al., 2009; Nyström Sandman et al., 2012). The drawback of the traditional regression methods, such as GAM, is that they require reliable information on both presence and absence of the modelled response. In many studies, however, obtaining reliable absence data is problematic and modelling using presence-only information has recently gained popularity in marine studies (e.g. Bryan and Metaxas, 2007; Ready et al., 2010; Howell et al., 2011; Jones et al., 2012). The problem of potential false absences is especially apparent in many marine data sets with incomplete sampling, such as data derived from underwater video or other semi-quantitative methods. At the same time, the inherent methodological error introduced by each modelling method has prompted the use of forecast ensembles to quantify and control for uncertainty in predictions. In ensemble modelling several model specifications and methods are used to predict distribution and the resulting predictions are combined using a variety of approaches (Araujo and New, 2007).

Each modelling method relates the potential for a species presence or absence to the environmental conditions where the training data specifies the species as present or absent. The different ways in which the methods fit this relationship inevitably leads to some differences in the predicted distributions. The fact that most models deliver similar accuracy validation scores does not directly imply they produce a similar map. This study aimed to investigate the consequences that selecting between two popular SDM methods: (1) generalized additive modelling (GAM) and (2) maximum entropy modelling (Maxent), would have on the predicted spatial distribution of the seagrass *Zostera marina* L. *Z. marina* was chosen for the comparison, due to its important habitat building and modifying role, and consequent management importance, and the extensive information available on its distribution in the study area. We expected that a strong relationship between the species distribution and its environment would lead to a consistent description of the niche of *Z. marina*, but methodological differences would lead to some differences in the predicted distributions. We compare the distribution predictions of the two methods, to quantify the difference in the proportion of the environment predicted suitable, to investigate how each method describes the habitat of *Z. marina* and finally to assess the effect of using the two methods as an ensemble on the predicted distribution of *Z. marina*.

2. Materials and methods

2.1. Study area

The Baltic Sea is the largest brackish water basin in the world. This enclosed sea is characterized by steep physico-chemical gradients, limited water exchange, low biological diversity and high levels of nutrient pollution affecting the entire biota (Elmgren, 2001). The study area (total area ca. 7300 km^2, of which water covers ca. 5400 km^2, Fig. 1) is situated in the northern Baltic Sea (59′ 40″ N, 21′ 00″ E to 60′ 35″ N, 22′ 25″ E), and its delineation is based on the available observations of *Zostera marina* occurrences in SW Finland. It includes a large part of the Archipelago Sea (salinity range 5–6; Suominen et al., 2010), located between the mainland of Finland and the Åland Islands (Fig. 1). The area is shallow, with a mean depth of 23 m, and topographically extraordinarily complex. Although dominated by crystalline bedrock, a characteristic feature of this area is the presence of extensive ice-marginal sand moraines (Rainio, 1995). The bedrock and glacial deposits form a highly convoluted archipelago with over 30 000 scattered islands and small skerries, accompanied by equally complex underwater topography. The sand deposits on the glacial formations provide a patchily distributed substrate for rooted macrophytes, including the eelgrass *Z. marina* (Boström et al., 2006a).

Occurrence of *Zostera marina* has not been observed at depths exceeding 10 m in the area. We thus limited our study to include