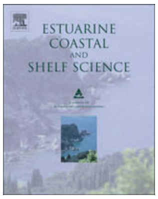
Estuarine, Coastal and Shelf Science 83 (2009) 90–96

Contents lists available at ScienceDirect

Estuarine, Coastal and Shelf Science

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



Transferability of predictive fish distribution models in two coastal systems

Göran Sundblad [a,b,*], Meri Härmä [c,d], Antti Lappalainen [c], Lauri Urho [c], Ulf Bergström [a]

[a] *Swedish Board of Fisheries, Institute of Coastal Research, Box 109, SE-74071 Öregrund, Sweden*
[b] *Department of Limnology, Evolutionary Biology Centre (EBC), Uppsala University, Box 573, SE-751 23 Uppsala, Sweden*
[c] *Finnish Game and Fisheries Research Institute, Box 2, FI-00791 Helsinki, Finland*
[d] *Department of Biological and Environmental Sciences, Faculty of Biosciences, University of Helsinki, Box 56, FI-00014 Helsinki, Finland*

ARTICLE INFO

Article history:
Received 31 October 2008
Accepted 19 March 2009
Available online 5 April 2009

Keywords:
prediction
modelling
fish larvae
habitat selection
GAM
Baltic Sea

ABSTRACT

Species distribution modelling has emerged as a tool both for exploring niche theory and for producing distribution maps for management. To understand and predict potential effects of large scale habitat change there is a need for proper model validation and applicability also in unstudied areas. However, knowledge about factors influencing the transferability of distribution models, i.e. the accuracy of the models when applying them in a new geographical area, is limited. We have successfully modelled the larval distribution of two fish species, northern pike (*Esox lucius* L.) and roach (*Rutilus rutilus* L.), on a regional scale in the Baltic Sea using a few and easily measured environmental variables. When models were transferred from the training area to the testing area the models showed reasonable to very good discrimination (ROC 0.75 and 0.93) based on external validation using independent data separated also in time (1–2 years). The predicted larval distribution also overlapped with the distribution of young-of-the-year fish later in the season. Performance when reversing the transfer, by constructing the models in the testing area and predicting back to the original training area, was less successful. This discrepancy was species-specific and could be explained by differences in the species presence ranges along the predictor variables in the testing area compared to the training area. Our results illustrate how transferability success can be influenced by area-specific differences in the range of the predictor variables and show the necessity of validating model predictions properly.

1. Introduction

Maps of potential species distributions are usually made by combining statistical models of species-environment relationships with maps of the environmental variables in a geographic information system. These distribution maps are valuable tools in management and conservation efforts and contribute to studies of habitat and niche theory (Araújo et al., 2004; Pearman et al., 2008). The rapid growth of species distribution models during the last decade has seen the development of several quantitative methods for mapping habitat distribution and quality (Guisan and Thuiller, 2005). Although based on different assumptions, approaches such as generalised linear and additive models, classification and regression trees and multivariate adaptive regression splines are all frequently used to describe the often complex, non-linear species-environment relationships (Guisan and Zimmerman, 2000). In addition, many other techniques such as quantile regression, boosted regression trees and structural equation modelling are continuously being developed (see Austin, 2007; De'ath, 2007; and Vaz et al., 2008 for applications). Considering the multitude of existing techniques it is perhaps not surprising that the specific statistical method applied can be considered of less importance than a suitable combination of ecological knowledge and general statistical skill in the modeller (Austin, 2007). Despite of (or possibly due to) the recent increase in species distribution models in peer reviewed literature, aspects such as the importance of validation (Olden et al., 2002) and geographical transferability of the models, i.e. building a model in one area and exporting it to another area (Randin et al., 2006), are still under development.

Maps of potential distribution will be of little use, or may even do more damage than good, e.g. by spatial misallocation of conservation efforts, if the predictive capability is not communicated to end-users. Validation can be made in a variety of ways and even though the most suitable approach is to use independent data (Fielding and Bell, 1997) this is seldom the case (Olden et al., 2002), often due to time or cost constraints. The simplest form of validation of a models predictive performance is the resubstitution approach, i.e. using the same data in model construction and validation. However, the resubstitution approach has been shown to

---

* Corresponding author. Swedish Board of Fisheries, Institute of Coastal Research, Box 109, SE-74071 Öregrund, Sweden.
E-mail address: goran.sundblad@fiskeriverket.se (G. Sundblad).

0272-7714/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ecss.2009.03.025

produce biased and usually overoptimistic estimates (Olden et al., 2002 and references therein). Other validation approaches are different forms of cross validation. An advantage of cross validation is that all data can be used in model building, although its predictive performance can still be overestimated compared to independent data (Sandman et al., 2008). The standard method to assess the accuracy of species distribution models is to look at the area under curve value of a receiver operating characteristic (ROC) plot (Fielding and Bell, 1997). The advantages of ROC plots are considered to be their independence of a particular threshold and that they provide a single measure of model and prediction accuracy (e.g. Pearce and Ferrier, 2000), although recent criticism suggests that additional performance measures should be provided (Lobo et al., 2008).

Assessing the generality of a species distribution model is a necessity before inferences in unknown areas or situations can be made (Fielding and Haworth, 1995). The ability to transfer a model in space and time is dependent on where in the process chain the predictor variables act (Austin, 2002). Predictor variables with more direct mechanistic effects should have a higher transferability potential than indirect predictor variables, since the relationship between a response variable and an indirect predictor variable are expected to change with region and (or) time, while the relationship between a response variable and a direct predictor variable are expected to change less. As a consequence, an increased distance in space or time when reallocating a model should therefore negatively influence its potential transferability, especially when using indirect predictor variables. Preferably the range of the predictor variables should be the same, or wider, in the training region than in the target region, but potential effects of the quality and type of predictor variables are still poorly known (Randin et al., 2006). Previous attempts of transferring descriptions of fish habitat suitability are only available from river systems and have given mixed results. Freeman et al. (1997) showed that models could be successfully transferred between regulated and non-regulated streams, and suggested that transferability success was higher for fish with more specific habitat requirements. In contrast, Glozier et al. (1997) demonstrated that the habitat preferences of a fish can be stream-specific and that the transferability potential between river systems may, thus, be low. Ultimately it is the realised niche that is described in distribution models, and potential differences in biotic pressure (e.g. prey availability, competition, predation or historic dispersal) between areas can also affect model transferability, even when the abiotic conditions are within the same range in both areas (Fielding and Haworth, 1995).

This study focuses on habitat characteristics and distribution of the larval stage of two common fish species in the brackish coastal waters of the Northern Baltic Sea, northern pike (*Esox lucius* L.) and roach (*Rutilus rutilus* L.). Both marine and freshwater species in the Baltic Sea exist on the edge of their tolerance limits, which together with the strong environmental gradients should provide suitable conditions for successful species distribution modelling with abiotic factors (Florin et al., 2009). Additionally, the juvenile and larval stages of fish often require more specific environmental conditions than subsequent stages and should therefore show a stronger preference for specific habitats. For example, both low temperatures and increased salinities have been shown to have direct negative effects on pike and roach ontogenic development (Klinkhardt and Winkler, 1989; Engström-Öst et al., 2005; Jacobsen et al., 2007). Water visibility, and thus Secchi depth, is related to productivity, with higher production in more turbid and eutrophic areas. Besides being an indirect variable indicating the abundance of food for larval pike and roach, Secchi depth may also directly affect the prey capture rate of fish (Ljunggren and Sandström, 2007), as well as predator avoidance and habitat choice (Lehtiniemi et al., 2005). Wave exposure may also act as a proxy variable for more direct variables such as zooplankton abundance and temperature, thus potentially influencing the distribution of larval fish. Sheltered areas often provide a more stable and suitable environment for larval fish than exposed areas, with higher temperatures in spring and a higher zooplankton production (e.g. Sandström and Karås, 2002).

Pike is a top predator in coastal areas of the Baltic Sea and is highly valued as a target species for both recreational and commercial fishery. Roach is one of the most abundant species in the inner coastal areas, and being a benthivore as well as an important prey species for predatory fish it has a key role in the coastal system. Both pike and roach spawn in early spring and prefer sheltered, vegetated shallow shores as spawning and larval areas (Karås and Hudd, 1993; Casselman and Lewis, 1996) and reed belts (*Phragmites australis* (Cav.) Steud.) have been reported to be important reproduction habitats at least in the Finnish parts of the Northern Baltic Sea (Urho et al., 1990; Lappalainen et al., 2008). These habitats are subject to intense physical exploitation by for example dredging, boating and construction of jetties, which can have detrimental effects on the recruitment habitats of pike, roach and other coastal species by decreasing vegetation cover and altering the abiotic environment (Eriksson et al., 2004; Sandström et al., 2005; Snickars et al., 2009). The importance of these recruitment habitats in combination with a high exploitation pressure stresses the need for mapping potential recruitment habitats to aid conservation management. For this, species distribution modelling provides a cost-effective tool for identification of areas of key interest.

The aims of this study were: (1) to produce species-specific distribution models for pike and roach larvae by relating larval distribution to regional environmental variables; and (2) to export the models and generate spatial habitat predictions in a new coastal area. We show that accurate model predictions may be obtained by transferring models to new areas. Our results emphasise the importance of considering predictor variable ranges when transferring models, as well as the need for using external independent data for validation of predicted distributions.

2. Methods

2.1. Study areas and field sampling

The Baltic Sea is one of the largest brackish water basins in the world. The two study areas were situated in the archipelago of southwest Finland (Fig. 1), which consists of numerous islands and displays strong environmental gradients on small spatial scales. There is also a general larger gradient from the off-shore areas, with less saline, warmer and more eutrophic waters closer to the mainland. Sampling of newly-hatched pike and roach larvae, concurrently with abiotic variable measurements (temperature, salinity and Secchi depth), were conducted yearly between the beginning of May and the end of June over a period of four years (Fig. 1). Secchi depth was measured using a white Secchi disk (Ø 25 cm) and salinity was measured with a portable conductivity meter (Thermo Fisher Scientific Inc., Waltham, MA, USA) using the Practical Salinity Scale.

In the first 2 years (May 4 to June 21 2004 and May 17 to June 17 2005) larvae of both species were sampled in area 1 (training area), and in area 2 (testing area) the following year (May 16 to June 21 2006). Sampling of roach larvae was carried out at 112 sampling sites in the training area and at 94 sites in the testing area, and sampling of pike larvae at 91 sites in the training area and at 81 sites in the testing area. Sampling sites were randomly distributed from the inner to the outer archipelago in both areas (Fig. 1) and