Ecological Modelling 309–310 (2015) 72–81

Contents lists available at ScienceDirect

Ecological Modelling

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

Can multilayer perceptron ensembles model the ecological niche of freshwater fish species?

R. Muñoz-Mas[a,*], F. Martínez-Capel[a], J.D. Alcaraz-Hernández[a], A.M. Mouton[b]

[a] Institut d'Investigació per a la Gestió Integrada de Zones Costaneres (IGIC), Universitat Politècnica de València, C/Paranimf 1, 46730 Grau de Gandia, València, Spain
[b] Research Institute for Nature and Forest (INBO), Kliniekstraat 25, B-1070 Brussels, Belgium

ARTICLE INFO

Article history:
Received 23 December 2014
Received in revised form 24 April 2015
Accepted 25 April 2015
Available online 15 May 2015

Keywords:
Artificial neural networks
Barbus haasi
Data mining
Species distribution modelling
Uncertainty analysis

ABSTRACT

The potential of Multilayer Perceptron (MLP) Ensembles to explore the ecology of freshwater fish species was tested by applying the technique to redfin barbel (*Barbus haasi* Mertens, 1925), an endemic and montane species that inhabits the North-East quadrant of the Iberian Peninsula. Two different MLP Ensembles were developed. The physical habitat model considered only abiotic variables, whereas the biotic model also included the density of the accompanying fish species and several invertebrate predictors. The results showed that MLP Ensembles may outperform single MLPs. Moreover, active selection of MLP candidates to create an optimal subset of MLPs can further improve model performance. The physical habitat model confirmed the redfin barbel preference for middle-to-upper river segments whereas the importance of depth confirms that redfin barbel prefers pool-type habitats. Although the biotic model showed higher uncertainty, it suggested that redfin barbel, European eel and the considered cyprinid species have similar habitat requirements. Due to its high predictive performance and its ability to deal with model uncertainty, the MLP Ensemble is a promising tool for ecological modelling or habitat suitability prediction in environmental flow assessment.

1. Introduction

Ecological models for the quantitative prediction of species distributions are key to understanding the realised niche of species and its implication for species conservation in relation to global change (Austin, 2007). Therefore, ecological models have increasingly received attention due to their wide management applications in the context of biogeography, conservation biology and climate change studies (Mouton et al., 2010). Many studies on ecological modelling have focused on explanation rather than prediction (Elith and Leathwick, 2009); however, differences in the life-history or in the gene flow of fish assemblages could result in different realised niches (Mouton et al., 2010). Abiotic factors, together with dispersal and biotic interactions, are often considered the three elements that shape the ecological niche by determining species distribution and abundance (Barve et al., 2011). However, ecological models have usually focused on abiotic factors only (Boulangeat et al., 2012), and very few studies in freshwater fish ecology have explicitly included biotic variables (Elith and Leathwick, 2009) to explore biotic interactions and consumer-resource dynamics (Soberón, 2007). The consideration of these three elements (i.e. abiotic, biotic and dispersal factors) do not allow for simple statistical analysis because the data collected often exhibit non-linear and complex data structures (Crisci et al., 2012). Consequently, there is a need for new and innovative approaches to understand the complex structure of living systems (Larocque et al., 2011).

Several sophisticated modelling techniques have been applied in the ecological modelling of fish species, ranging from linear to multivariate and machine learning techniques such as Artificial Neural Networks (ANN) (Brosse and Lek, 2000, Muñoz-Mas et al., 2014, Palialexis et al., 2011). The most popular ANN architecture has been the Multilayer Perceptron (MLP) paradigm because it is considered to be able to approximate any continuous function (Olden et al., 2008). Formerly, MLP was referred to as a 'black box' because it provided little explanatory insight into the relative influence of variables in the prediction process (Olden and Jackson, 2002). To date, an enormous effort has been made to develop methods that clarify variable importance and interactions (Gevrey et al., 2006; Lek et al., 1996; Olden and Jackson, 2002), and consequently, MLPs

* Corresponding author. Tel.: +34 962849458; fax: +34 962849309.
E-mail addresses: pitifleiter@hotmail.com (R. Muñoz-Mas), fmcapel@dihma.upv.es (F. Martínez-Capel), jdalcaraz@gmail.com (J.D. Alcaraz-Hernández), ans.mouton@inbo.be (A.M. Mouton).

http://dx.doi.org/10.1016/j.ecolmodel.2015.04.025
0304-3800/

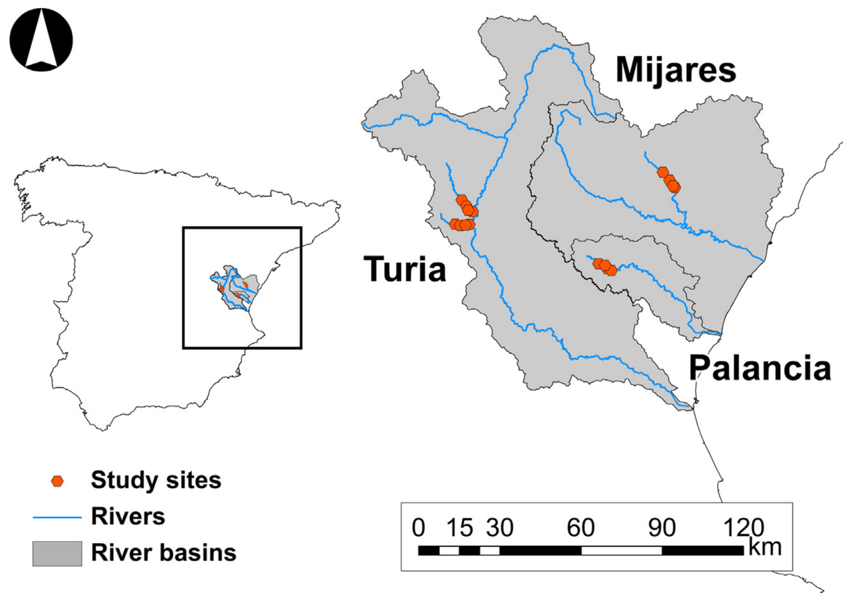

Fig. 1. Location of the target river basins in the Iberian Peninsula (left) and study sites in the Mijares, Palancia and Turia River basins.

should no longer be treated as 'black box' models (Özesmi et al., 2006).

There are several examples of single MLP applications in freshwater fish ecology (Park and Chon, 2007). For instance, MLPs have been successfully applied to model fish ecology through a broad range of ecosystems (Brosse and Lek, 2000; Gevrey et al., 2006; Kemp et al., 2007; Laffaille et al., 2003) and in some cases outperforming other statistical approaches (Baran et al., 1996; Lek et al., 1996). Despite those successful studies, it has been demonstrated that single models (e.g. a single MLP) do not necessarily perform consistently, resulting in divergent predictions (Buisson et al., 2010; Fukuda et al., 2011a, 2013). The use of model ensembles has been emphasised to overcome this phenomenon (Araújo and New, 2007). The Multilayer Perceptron Ensemble (MLP Ensemble, Hansen and Salamon, 1990) has proven to be proficient in several areas of ecology (Palialexis et al., 2011; Watts and Worner, 2008), but has rarely been applied in freshwater ecosystems (Muñoz-Mas et al., 2014).

Fish communities in Mediterranean rivers are an interesting targets to develop these novel statistical approaches (Hopkins II and Burr, 2009), particularly communities dominated by cyprinids, as they are characterised by a high number of endemic species for which there is insufficient knowledge about their ecology (Ferreira et al., 2007). Furthermore, endemic species tend to facilitate a more robust analysis of species–environment relationships. In this paper, we focused on the redfin barbel (*Barbus haasi* Mertens, 1925), a rheophilic small barbel (maximum body-length 30 cm) that is endemic to the Iberian Peninsula (Bianco, 1998) and categorised as vulnerable (Freyhof and Brooks, 2011). Their populations have decreased markedly, with pollution and the presence of exotic species being the main factors involved in the decline (Perea et al., 2011). Although redfin barbel has been the subject of numerous studies addressing its life-history, home-range, habitat preferences and the effects of pollutants (Aparicio and De Sostoa, 1999; Aparicio, 2002; Figuerola et al., 2012; Grossman and De Sostoa, 1994), a knowledge gap remains on the impact of biotic variables such as the density of accompanying fish species or invertebrate predictors in its ecological niche.

Therefore, the objective of this study was: (1) to test the proficiency of the MLP Ensembles to model the ecological niche of freshwater fish species and (2) to test whether biotic variables affect the distribution of redfin barbel. To achieve these aims using MLP Ensembles, two different models of redfin barbel were developed. The first considered only physical habitat variables, the second included biotic and physical habitat variables.

2. Materials and methods

2.1. Data collection

The study was conducted at the meso-scale in every summer, between 2003 and 2006. The study sites were located in the headwaters of the Ebron and Vallanca Rivers (Turia River tributaries), the Palancia River and the Villahermosa River (Mijares River Tributary) (Fig. 1) which approximately correspond to the southern limits of redfin barbel distribution (Perea et al., 2011). All the study sites were in unregulated streams and therefore a wide flow range was sampled (i.e. from 0.02 m^3/s to 1.84 m^3/s). For complete climatic description of the study area, see Alcaraz-Hernández et al. (2011) and Mouton et al. (2011).

2.1.1. Physical habitat survey

The physical habitat was assessed in every 300 m reach using an adaptation of the Basinwide Visual Estimation Technique (Dolloff et al., 1993). The approach stratifies the study site by HydroMorphological Units (hereafter called HMUs) classified as: pools, glides, riffles, and rapids (see Alcaraz-Hernández et al., 2011 for further details). Once an HMU was categorised, its physical attributes were recorded. They were, length, average width, obtained from three cross-sections corresponding to ¼, ½, and ¾ of the total length, mean depth (hereafter as depth), calculated from nine points corresponding to the measurements taken at each of the aforementioned cross-sections and the maximum depth, measured at the corresponding point. Percentage of shading over the channel, percentage of embeddedness, pieces of woody debris and percentage of the substrate types following a simplified classification from the American Geophysical Union (Martínez-Capel et al., 2009; Muñoz-Mas et al., 2012) were visually estimated and summarised in the substrate index (Mouton et al., 2011). In addition, the cover index (García de Jalón and Schmidt, 1995) was determined. This index characterises the available refuge due to caves, shading, substrate, submerged vegetation and water depth by assigning six scores from 0 (no refuge) to 5 (maximum score), and the weighted aggregation of these scores produces an index range from 0 to 10.