Environmental Modelling & Software 59 (2014) 30–43

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

# Application of Probabilistic Neural Networks to microhabitat suitability modelling for adult brown trout (*Salmo trutta* L.) in Iberian rivers

R. Muñoz-Mas [a,*], F. Martínez-Capel [a], V. Garófano-Gómez [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 2 January 2014
Received in revised form
30 April 2014
Accepted 2 May 2014
Available online

*Keywords:*
Probabilistic Neural Networks
Brown trout
Microhabitat suitability
Prevalence
Spatially explicit evaluation
Mediterranean rivers

ABSTRACT

Probabilistic Neural Networks (PNN) have been tested for the first time in microhabitat suitability modelling for adult brown trout (*Salmo trutta* L.). The impact of data prevalence on PNN was studied. The PNN were evaluated in an independent river and the applicability of PNN to assess the environmental flow was analysed. Prevalence did not affect significantly the results. However PNN presented some limitations regarding the output range. Our results agreed previous studies because trout preferred deep microhabitats with medium-to-coarse substrate whereas velocity showed a wider suitable range. The 0.5 prevalence PNN showed similar classificatory capability than the 0.06 prevalence counterpart and the outputs covered the whole feasible range (from 0 to 1), but the 0.06 prevalence PNN showed higher generalisation because it performed better in the evaluation and it allowed a better modulation of the environmental flow. PNN has demonstrated to be a tool to be into consideration.



## 1. Introduction

The environmental impact of hydrological alteration is of major concern for researchers (Marsili-Libelli et al., 2013). Therefore, scientists and managers have developed a vast body of methodologies to assess the consequences of changes in running river flows (Acreman and Dunbar, 2004; Ahmadi-Nedushan et al., 2006). This concern overstepped the merely academic environment yielding the Water Framework Directive (WFD) (European Parliament and Council, 2000), a legislative framework that stated a set of environmental objectives for water bodies with the time frame of the year 2015. The WFD implicitly assumes an underlying link between ecological status and abiotic quality elements; thus, a key aspect is to identify and assess the links between the physical and biological components of streams (Conallin et al., 2010). Accordingly, freshwater fish are considered good indicators of water quality and biotic integrity in freshwater ecosystems (Pont et al., 2006).

The Instream Flow Incremental Methodology (IFIM) (Bovee et al., 1998) was the first methodological framework for the environmental impact assessment and negotiation in water allocation schemes (Paredes-Arquiola et al., 2013). Moreover it has been stressed as the most scientifically and legally defensible methodology available (Tharme, 2003). The physical habitat simulation is a part of the IFIM methodology that permits to understand the impact of flow alterations on stream habitat (Maddock, 1999). Consequently it has been considered in the transposition of the WFD to the Spanish norm for hydrological planning (MAGRAMA, 2008). Among the fish species considered in the physical habitat simulation, in the Iberian context brown trout (*Salmo trutta* L. 1758) has been specifically used as an indicator of ecological status (Ayllón et al., 2012). Therefore, insight into the habitat suitability of brown trout is crucial for the implementation of the WFD and for environmental flow (e-flow) assessments, especially in areas vulnerable to global change such as the Mediterranean streams (García-Ruiz et al., 2011).

The continuous univariate Habitat Suitability Curves (HSC) are a simple and common modelling approach in studies involving physical habitat simulation (Payne and Allen, 2009); hence several researchers have developed habitat suitability models in the form of HSC (Ayllón et al., 2010; Bovee, 1978; Hayes and Jowett, 1994; Raleigh et al., 1986; Vismara et al., 2001). The relationship between Weighted

* Corresponding author. Tel.: +34 962849458; fax: +34 962849309.
*E-mail addresses:* pitifleiter@hotmail.com, ramuoma@upvnet.upv.es (R. Muñoz-Mas), fmcapel@dihma.upv.es (F. Martínez-Capel), virgargm@upvnet.upv.es (V. Garófano-Gómez), ans.mouton@inbo.be (A.M. Mouton).

http://dx.doi.org/10.1016/j.envsoft.2014.05.003
1364-8152/

Usable Area (WUA) and river flow (Bovee et al., 1998) derived from these models has been used extensively in e-flow assessments (Payne, 2003). However, several authors have suggested that considering each hydraulic variable independently may be questionable because ignoring significant interactions between variables may induce a bias (Orth and Maughan, 1982). As a consequence, the multivariate approach has recently increased in popularity (De Pauw et al., 2006). Several data-driven multivariate techniques have been applied in brown trout habitat suitability modelling. Specifically at the microhabitat scale, these studies ranged from simple bivariate polynomial functions (Lambert and Hanson, 1989; Vismara et al., 2001) to more complex fuzzy rule base models (Jorde et al., 2001). Thereby, logistic regression has been used by some researchers (Ayllón et al., 2010; Hayes and Jowett, 1994), as well as Generalized Additive Models (GAMs) (Jowett and Davey, 2007) to develop habitat suitability models for brown trout. Among the machine learning techniques, Artificial Neural Networks (ANN) and specifically the Multilayer Perceptron, have also been applied to model habitat suitability for brown trout (Reyjol et al., 2001). In the eastern Iberian Peninsula, the fuzzy logic approach has been applied to develop models for brown trout with the mesoscale as the central resolution (Mouton et al., 2011) whereas at the microscale Muñoz-Mas et al. (2012) developed fuzzy rule base models for middle-size brown trout (body length from 10 to 20 cm).

Overall, each approach for habitat modelling has advantages and disadvantages and due to their different model structures they are distinct in their data needs, transferability, user-friendliness and presentable outputs (Conallin et al., 2010). Therefore, the habitat simulation methodologies are in a permanent evolution driven by their imperfections and inherent constrains (Lamouroux et al., 1998) (Lamouroux et al., 1998). Probabilistic Neural Networks (PNN) (Specht, 1990) are a promising type of ANN. These were applied successfully in pattern classification in some areas related to fish (e.g. classification of sonar signals) (Moore et al., 2003). But to our knowledge, this technique has never been applied in habitat suitability modelling at the microscale before.

An important aspect in data-driven habitat suitability modelling is the prevalence (i.e., proportion of presence in the entire dataset) because it can have a strong effect on model performance (Fukuda, 2013; Manel et al., 2001). The decreasing trends in brown trout populations (Almodóvar et al., 2012) or the temporary absence of the species (Lütolf et al., 2006) in addition to the sampling protocols, can lead to low prevalence databases. The capability to deal with low prevalence has been the main focus in several studies. For instance, Mouton et al. (2009) developed a prevalence-adjusted method addressed to fuzzy rule base models, and Freeman et al. (2003) tested the ability of Random Forests (Breiman, 2001) to deal with low prevalence datasets. In this context, PNN are theoretically able to cope with low prevalence databases (Specht, 1990), thus suggesting its suitability to construct fish habitat models with unbalanced databases. Another remarkable issue in ecological modelling is the over-fitting. Some techniques are prone to that phenomenon in a different degree. Therefore, some authors highlighted the importance of making a successful evaluation (sensu Guisan and Zimmermann, 2000) using independent data to improve the reliability of the models (Bennett et al., 2013).

Our study was aimed at testing the suitability of PNN as a tool for brown trout habitat suitability modelling at the microscale. To achieve this general aim, (i) presence–absence PNN were generated and trained; (ii) the effect of prevalence on models performance and habitat assessment was analysed; (iii) the modelled brown trout habitat suitability was analysed in a multivariate way by checking how the PNN assess a synthetic database covering all the possible combinations of velocity, depth and substrate within the survey range; (iv) the PNN were evaluated in an independent river under similar ecological conditions to those where the training database was collected; and finally, (v) the applicability of the PNN models to assess minimum legal e-flows at the evaluation site was discussed by calculating the WUA - flow curve.

## 2. Materials and methods

### 2.1. Microhabitat data collection

The target species of this study at the microscale was the adult (body length > 20 cm) brown trout. The data samplings were carried out at low-flow conditions during late spring, summer and early autumn in the period 2007–2009 in the Guadiela and Cuervo Rivers (within the Tagus River Basin; TB) and in the Jucar and Senia Rivers (within the Jucar River Basin District; JRBD) (Fig. 1).

The microhabitat study was undertaken in complete and connected Hydro-Morphological Units (hereafter, HMUs) and classified as pool, glide, riffle and rapid. A sort of modification of the equal effort approach was applied (Bovee et al., 1998) with the selection of equal areas of slow and fast water HMUs, grouping pools with glides (slow) and riffles with rapids (fast). Each HMU was studied by underwater observation (snorkelling) during daylight with minimum disturbance to the fish according to common procedures (Heggenes et al., 1990; Martínez-Capel et al., 2009). This technique allows the observation of the fish behaviour, thus only adult brown trout that were 'feeding' or 'holding a feeding position' were considered because it is assumed that they are occupying such positions because are the most energetically profitable (Rincon and Lobon-Cervia, 1993). Microhabitat conditions, termed as training patterns, were measured along the HMU in cross-sections, classifying fish abundance into two groups as 'absence' (no fish observed) and 'presence' (at least one fish observed). The resulting sampled area per training pattern (measurement) ranged from 1.23 $m^2$ to 7.96 $m^2$. The high number of absence patterns *versus* presence patterns led to a low prevalence (average prevalence being 0.06) that ranged from 0.02 to 0.11 depending on the river (Table 1).

Depth was measured with a wading rod to the nearest cm and the mean flow velocity of the water column (hereafter velocity) was measured with an electromagnetic current meter (Valeport®). The percentage of each substrate class was visually estimated around the sampling point or fish location. The substrate classification was simplified from the American Geophysical Union size scale: bedrock, boulders (>256), cobbles (64–256 mm), gravel (8–64 mm), fine gravel (2–8 mm), sand (62 mm–2 mm), silt (<62 mm) and vegetated soil (i.e. substrate covered by macrophytes), similarly to a previous work in Iberian rivers (Martínez-Capel et al., 2009). Substrate composition was converted into a single value through the Substrate index (hereafter substrate), by summing weighted percentages of each substrate type as follows: Substrate index = 0.08 ·% bedrock + 0.07 ·% boulder + 0.06 ·% cobble + 0.05 ·% gravel + 0.04 ·% fine gravel + 0.03 ·% sand (Mouton et al., 2011) (Table 2). Finally, the three input variables (velocity, depth and substrate) were normalized *N(0,1)*.

### 2.2. Development of the Probabilistic Neural Network

#### 2.2.1. PNN theory

PNN are radial-basis neural networks based on a Bayes-Parzen classifier (Specht, 1990). PNN basically compare how close the input pattern is to the patterns of each category in the training database and assign the category that presents the highest

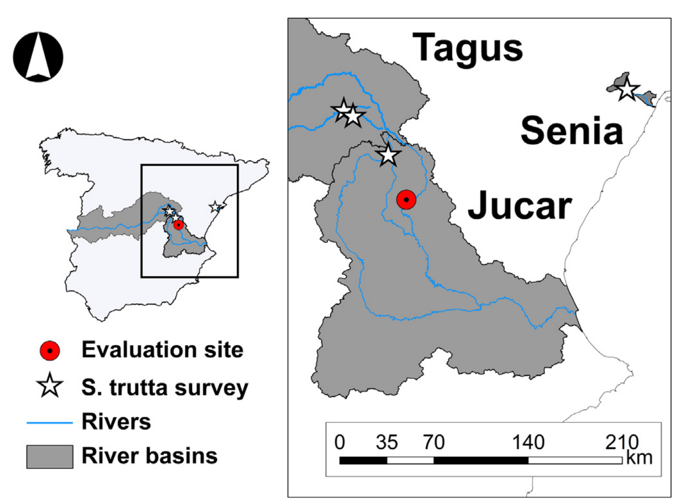


**Fig. 1.** In the Iberian Peninsula (left), location of the sites where microhabitat data of brown trout were collected in rivers within the Tagus River Basin and the Jucar River Basin District. Red circle shows the location where the models were evaluated in the Cabriel River. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)